

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Main Contributions . . . . .	2
1.2	Task-Specific Pure Exploration . . . . .	2
1.3	Examples and Applications . . . . .	3
1.4	Related Works . . . . .	3
1.5	Organization . . . . .	4
<b>2</b>	<b>Summary of Results</b>	<b>5</b>
2.1	Optimality of Certainty Equivalence . . . . .	5
2.2	Task-Optimal Experiment Design . . . . .	6
<b>3</b>	<b>Interpreting the Results</b>	<b>7</b>
3.1	Instance-Optimal LQR Synthesis . . . . .	7
3.2	System Identification in Arbitrary Norms . . . . .	7
<b>4</b>	<b>Task-Guided Exploration yields Provable Gains</b>	<b>7</b>
4.1	Suboptimality of Low-Regret Algorithms . . . . .	8
<b>5</b>	<b>Numerical Experiments</b>	<b>8</b>
<b>6</b>	<b>Conclusion</b>	<b>9</b>
<b>A</b>	<b>Organization and Notation</b>	<b>16</b>
A.1	Organization . . . . .	16
A.2	Notation . . . . .	16
<b>I</b>	<b>Results and Extensions</b>	<b>19</b>
<b>B</b>	<b>Formal Results and Algorithm</b>	<b>19</b>
B.1	Martingale and Linear Dynamical Decision Making (MDM and LDDM) . . . . .	19
B.2	Lower Bound for Decision Making in MDM . . . . .	20
B.3	Lower Bound for Optimal Decision Making in LDDM . . . . .	21
B.4	Upper Bound for Certainty Equivalence Decision Making in MDM . . . . .	22
B.5	Efficient Experiment Design in Frequency Domain (LDDM) . . . . .	24
B.6	Optimal LDDM Decision-Making: Formal Statement and Guarantee for TOPLE . . . . .	27
<b>C</b>	<b>Remarks and Extensions</b>	<b>28</b>
C.1	Randomized Decisions . . . . .	28

C.2	Non-Identity Noise Covariance . . . . .	28
C.3	Unstable Systems and State Costs in Experiment Design . . . . .	29
C.4	Expectation v.s. High Probability . . . . .	30
C.5	Subspaces and Parametric Uncertainty . . . . .	30
C.6	Parametric Nonlinear Systems . . . . .	30
C.7	Suboptimality of Low Regret Algorithms . . . . .	31
<b>II Martingale Decision Making</b>		<b>34</b>
<b>D Optimal Rates for Martingale Regression in General Norms</b>		<b>34</b>
D.1	Upper and Lower Bounds on $M$ -norm Regression . . . . .	34
D.2	Proof of $M$ -norm Regression Lower Bound (Theorem D.1) . . . . .	35
D.3	Proof of $M$ -norm Regression Upper Bound (Theorem D.2) . . . . .	39
<b>E Lower Bounds on Martingale Decision Making</b>		<b>41</b>
E.1	Proof of General Decision Making Lower Bounds . . . . .	42
E.2	Proof of Theorem B.1 . . . . .	48
E.3	Proof of Proposition E.1 and Proposition E.2 . . . . .	48
<b>F Upper Bounds on Certainty Equivalence Decision Making</b>		<b>50</b>
F.1	Certainty Equivalence Upper Bound . . . . .	50
F.2	Euclidean Norm Estimation . . . . .	51
<b>III Linear Dynamical Decision Making</b>		<b>52</b>
<b>G Notation for Linear Dynamical Systems</b>		<b>52</b>
G.1	Covariance Notation . . . . .	52
G.2	Lifted Dynamical System . . . . .	53
G.3	Linear Dynamical Systems as Vector Regression . . . . .	53
G.4	Key Parameters in the Analysis . . . . .	53
G.5	Linear Dynamical Systems Notation Proofs . . . . .	54
<b>H Lower Bounds in Linear Dynamical Systems</b>		<b>57</b>
H.1	Regular Policies in Linear Dynamical Systems . . . . .	57
H.2	Proof of Theorem B.2 . . . . .	58
H.3	Periodicity of Optimal Inputs . . . . .	64
H.4	Frequency Domain Approximation . . . . .	68
H.5	Smoothness of Covariates . . . . .	70

H.6	Infinite-Horizon Approximation . . . . .	74
<b>I</b>	<b>Certainty Equivalence Decision Making with Sequential Open-Loop Policies</b>	<b>76</b>
I.1	Sequential Open-Loop Policies Satisfy Assumption 7 . . . . .	77
I.2	Concentration of Covariates . . . . .	79
I.3	State Norm Bounds . . . . .	83
<b>J</b>	<b>Experiment Design in Linear Dynamical Systems</b>	<b>86</b>
J.1	Proof of Theorem B.6 . . . . .	86
J.2	Certainty Equivalence Experiment Design . . . . .	88
J.3	Operator Norm Estimation . . . . .	89
J.4	Optimality of Inputs . . . . .	90
<b>IV</b>	<b>LQR and Further Examples</b>	<b>96</b>
<b>K</b>	<b>LQR as Linear Dynamical Decision Making</b>	<b>96</b>
K.1	LQR is an Instance of LDDM . . . . .	96
K.2	Norm Bounds on Gradients . . . . .	97
K.3	Norm Bounds on Control Theoretic Quantities . . . . .	104
<b>L</b>	<b>Provable Gains for Task-Optimal Design</b>	<b>106</b>
L.1	Preliminaries for Comparison of Designs . . . . .	106
L.2	Computation of Task Hessian . . . . .	108
L.3	Proof of Proposition 4.1 . . . . .	111
L.4	Proof of Proposition 4.2 . . . . .	116
<b>M</b>	<b>Details on Numerical Results</b>	<b>119</b>

## A. Organization and Notation

### A.1. Organization

We break the appendix up into four parts. In Part I, we provide a formal overview of our work. In particular, Appendix B formally defines our decision-making settings—martingale decision making (MDM) and linear dynamical decision making (LDDM)—formally states our results and defines TOPLE, and provides a more detailed explanation of our frequency-domain construction. This section also formally introduces the notion of a *sequential open-loop* policy, which figures prominently in the analysis of TOPLE. Appendix C sketches out various extensions to our results, provides additional remarks, and states and proves the formal version of the lower bound on low-regret algorithms.

Part II covers martingale decision making. Appendix D states and proves our upper and lower bounds for martingale regression in general norms. Next, Appendix E formally introduces our martingale decision-making setting, and, through a reduction to regression, proves a locally minimax lower bound on martingale decision making with smooth losses. Finally, Appendix F proves a general upper bound on certainty equivalence decision making under a certain regularity assumption. Of note, Part II does not assume we are operating in the setting of a linear dynamical system—the results here apply to the more general martingale decision making setting.

Part III covers our results in the setting of linear dynamical decision making setting. We begin in Appendix G by introducing additional notation specific to linear dynamical systems we will use throughout. In Appendix H, we apply the results of Part II to prove our lower bound on *optimal* decision-making in linear dynamical systems. This section also shows that our restricted policy class,  $\mathcal{U}_{\gamma^2}^p$ , the set of periodic signals, contains a near-optimal policy. Appendix I shows that sequential open-loop policies meet our regularity assumption and we therefore obtain a corollary on efficient certainty equivalence decision making in linear dynamical systems. In addition, Appendix I provides rates at which the covariates of linear dynamical systems concentrate, an important piece in our analysis. Finally, Appendix J proves the upper bound on the performance of TOPLE. Our proof relies on showing that certainty equivalence experiment design plays near-optimal inputs, and that TOPLE is itself a sequential open-loop policy, allowing us to apply our certainty equivalence bound proved in Appendix I.

Lastly, Part IV covers applications of our results. Appendix K shows that the LQR problem is an instance of our general smooth decision making setting, and that we can therefore apply all our results to this problem. Appendix L works out explicitly the rates obtained by TOPLE and other exploration approaches in several LQR examples. Finally, Appendix M provides additional details on our numerical experiments.

### A.2. Notation

Below we present notation used throughout this work. We define our signal notation in more detail in Appendix B.5.1 and additional details on our notation for linear dynamical systems is presented in Appendix G. We will overload notation somewhat throughout, using  $\theta$  to refer to a vector as well as the concatenation of matrices,  $\theta = (A, B)$ . In the latter case,  $\|\theta\|_{\text{op}}$  denotes the operator norm of the matrix  $(A, B)$  but we let  $\|\theta\|_M^2 = \text{vec}(\theta)^\top M \text{vec}(\theta)$ .

Mathematical Notation	Definition
$\ \cdot\ _{\text{op}}$	Matrix operator norm
$\ \cdot\ _F$	Matrix Frobenius norm
$\ \cdot\ _2$	Vector 2-norm
$\ \cdot\ _M$	Vector Mahalanobis norm, $\ x\ _M^2 = x^\top M x$
$\ \cdot\ _{\mathcal{H}_\infty}$	System $\mathcal{H}$ -infinity norm
$i$	Imaginary number, $\sqrt{-1}$
Policy Notation	
$\pi/\pi_{\text{exp}}$	Exploration policy
$\Pi_{\gamma^2}$	Policies with average expected power bounded by $\gamma^2$
$\Pi_{\gamma^2}^{\text{sol}}$	Sequential open-loop policies with average expected power bounded by $\gamma^2$
$\Pi_{\gamma^2}^p$	Periodic policies with average expected power bounded by $\gamma^2$
$\boldsymbol{\tau}$	Input-state trajectory, $\boldsymbol{\tau} = (x_{1:T+1}, u_{1:T})$
$\text{dec}(\boldsymbol{\tau})$	Decision rule
$\text{ce}(\boldsymbol{\tau})$	Certainty-equivalence decision rule

Complexity Notation	Definition
$\Phi_T(\pi; \theta_*)$ $\Phi_{\text{opt}}(\gamma^2; \theta_*)$ $\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_*)$ $\mathfrak{M}_{\pi_{\text{exp}}}(\mathcal{R}; \mathcal{B})$ $\mathfrak{M}_{\gamma^2}(\mathcal{R}; \mathcal{B})$	Idealized risk, $\Phi_T(\pi; \theta_*) := \text{tr}(\mathcal{H}(\theta_*)\Gamma_T(\pi; \theta_*)^{-1})$ Optimal risk, $\Phi_{\text{opt}}(\gamma^2; \theta_*) := \liminf_{T \rightarrow \infty} \inf_{\pi_{\text{exp}} \in \Pi_{\gamma^2}} \Phi_T(\pi_{\text{exp}}; \theta_*)$ Steady-state analogue of $\Phi_{\text{opt}}(\gamma^2; \theta_*)$ Local minimax risk Exploration local minimax risk, lower bound on optimal policy risk
MDM Notation	
$\theta \in \mathbb{R}^{d_\theta}$ $\mathbf{a} \in \mathbb{R}^{d_a}$ $\mathcal{J}_{\theta_*}(\mathbf{a})$ $\mathcal{R}_\theta(\mathbf{a})/\mathcal{R}(\mathbf{a}; \theta)$ $\mathbf{a}_{\text{opt}}(\theta)$ $\mathcal{H}(\theta_*)$ $L_{\mathcal{R}i}, i = 1, 2, 3$ $L_{\mathbf{a}i}, i = 1, 2, 3$ $L_{\text{hess}}$ $\mu$ $L_{\text{quad}}$ $L_{\mathcal{H}}$ $r_{\text{quad}}(\theta)$	Nominal instance Decision variable Loss function Excess risk Optimal decision for instance $\theta$ Hessian of certainty equivalence excess risk, $\mathcal{H}(\theta_*) = \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*) _{\theta=\theta_*}$ Upper bound on $\ \nabla_{\mathbf{a}}^{(i)} \mathcal{R}(\mathbf{a}; \theta)\ _{\text{op}}$ Upper bound on $\ \nabla_{\theta}^{(i)} \mathbf{a}_{\text{opt}}(\theta)\ _{\text{op}}$ Lipschitz constant of $\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)$ in $\theta$ Parameter for quadratic lower bound on $\mathcal{R}(\mathbf{a}; \theta_*)$ $\frac{1}{6}(L_{\mathcal{R}3}L_{\mathbf{a}1}^3 + 3L_{\mathcal{R}2}L_{\mathbf{a}2}L_{\mathbf{a}1} + L_{\mathcal{R}1}L_{\mathbf{a}3})$ $6L_{\text{quad}} + L_{\mathcal{R}2}L_{\mathbf{a}1} + L_{\text{hess}}L_{\mathbf{a}1}^2$ Radius in which gradient bounds hold
LDS Notation	
$\theta = (A, B)$ $d_x$ $d_u$ $d$ $\sigma_w^2$ $x_t^u$ $x_t^w$ $\rho(A)$ $\tilde{\theta} = (\tilde{A}, \tilde{B})$ $\tau(A, \rho)$ $\rho_*$ $\tau_*$	System parameters State dimension Input dimension $d_x + d_u$ Process noise variance Portion of state driven by input Portion of state drive by noise ( $x_t = x_t^u + x_t^w$ ) Spectral radius of $A$ Lifted dynamical system, $\tilde{A} = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}$ , $\tilde{B} = \begin{bmatrix} 0 \\ I \end{bmatrix}$ Scaling of upper bound on $\ A^k\ _{\text{op}}$ , $\ A^k\ _{\text{op}} \leq \tau(A, \rho)\rho^k$ $\max\left\{\frac{1}{2}, \frac{2\ A_*\ _{\mathcal{H}_\infty}\ A_*\ _{\text{op}}}{1+2\ A_*\ _{\mathcal{H}_\infty}\ A_*\ _{\text{op}}}\right\}$ $\tau(\tilde{A}_*, \rho_*)$ , scaling of upper bound on $\ \tilde{A}_*^k\ _{\text{op}}$
Signal Notation	
$\mathfrak{F}(\cdot)$ $\mathfrak{F}^{-1}(\cdot)$ $\mathbf{u}$ $\check{\mathbf{u}}$ $\check{u}_t$ $\mathbf{U}$ $U_t$ $\mathcal{U}_{\gamma^2, k}$	Discrete-time Fourier transform (DFT) Inverse discrete-time Fourier transform Vector signal, $(u_t)_{t=1}^k$ , $u_t \in \mathbb{C}^{d_u}$ Discrete-time Fourier transform of $\mathbf{u}$ Element of $\check{\mathbf{u}}$ , DFT of $(u_s)_{s=1}^k$ Matrix signal, $(U_t)_{t=1}^k$ , $U_t \in \mathbb{C}^{d_u \times d_u}$ , $U_t$ Hermitian Matrix signal element Set of length $k$ matrix signals with power bounded by $\gamma^2$

**Task-Optimal Exploration in Linear Dynamical Systems**

Covariance Notation	
$\Sigma_T$	Random covariates
$\Gamma$	Kronecker of covariates, $I_{d_x} \otimes \Gamma$
$\Gamma_t^{\text{noise}}(\theta, \Sigma_u)$	Expected $t$ -step noise covariance when $u_t \sim \mathcal{N}(0, \Sigma_u)$
$\Gamma_t^{\text{noise}}(\theta, \sigma_u)$	$\Gamma_t^{\text{noise}}(\theta, \sigma_u^2 I)$
$\Gamma_t^{\text{in}}(\theta, \mathbf{u}, x_0)$	Covariance obtained on noiseless system playing $\mathbf{u}$ starting from $x_0$
$\Gamma_T(\theta, \mathbf{u}, \sigma_u, x_0)$	Expected covariance on noisy system when playing $\mathbf{u}$ , input noise $\mathcal{N}(0, \sigma^2 I)$ , starting from $x_0$
$\Gamma_k^{\text{freq}}(\theta, \mathbf{U})$	Frequency-domain steady-state covariance for length- $k$ matrix input $\mathbf{U}$
$\Gamma_{t,k}^{\text{freq}}(\theta, \mathbf{U})$	$\frac{t}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U})$
$\Gamma_{t,k}^{\text{ss}}(\theta, \mathbf{U}, \sigma_u)$	Expected steady state covariance when playing length- $k$ input $\mathbf{U}$ and input noise $\mathcal{N}(0, \sigma_u^2 I)$
$\bar{\Gamma}_T$	High probability upper bound on covariates
$\bar{\gamma}$	High probability upper bound on covariates for sequential open-loop policies
$\underline{\lambda}$	Lower bound on minimum eigenvalue of covariates
$\lambda_{\text{noise}}^*(\sigma_u)$	Minimum eigenvalue of noise Grammian, minimum excitation due to noise
$\lambda_{\text{noise}}^*$	$\lambda_{\text{noise}}^*(\gamma/\sqrt{2d_u})$
$L_{\text{cov}}(\theta_*, \gamma^2)$	Smoothness of covariates with respect to $\theta$
$r_{\text{cov}}(\theta_*)$	Radius in which smoothness of covariates holds

## Part I

# Results and Extensions

## B. Formal Results and Algorithm

In this section, we formally state the results given in Section 2 and present the full definition of TOPLE. This section is organized as follows. We first formally define our decision-making settings, MDM and LDDM, in Appendix B.1. Next, we present a lower bound on decision-making in the MDM setting in Appendix B.2. In Appendix B.3, we assume we are in the stronger LDDM setting and present a lower bound on *optimal* decision-making. Appendix B.4 provides a sufficient condition on exploration policies and shows that, under this condition, certainty equivalence decision-making is optimal in the MDM setting. Appendix B.4.1 then introduces a restricted set of policies in the LDDM setting, *sequential-open loop* policies, which we show contains TOPLE and is sufficiently regular. Appendix B.5 provides an overview of frequency-domain representations of signals, an essential piece in our construction of TOPLE, and formally defines several sub-routines of TOPLE. Finally, in Appendix B.6 we formally state TOPLE and provide an upper bound on its performance.

### B.1. Martingale and Linear Dynamical Decision Making (MDM and LDDM)

Our decision making setting considers smooth loss functions parameterized by models  $\theta \in \mathbb{R}^{d_\theta}$ ,  $\mathcal{J}_\theta(\mathbf{a}) : \mathbb{R}^{d_a} \rightarrow \mathbb{R}$ . The loss function induces the *excess risk* function

$$\mathcal{R}(\hat{\mathbf{a}}; \theta_\star) := \mathcal{J}_{\theta_\star}(\hat{\mathbf{a}}) - \inf_{\mathbf{a}'} \mathcal{J}_{\theta_\star}(\mathbf{a}'),$$

We denote the plug-in optimal decision

$$\mathbf{a}_{\text{opt}}(\theta) := \arg \min_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta),$$

that is, the optimal decision when  $\theta$  is the nominal parameter. We are, in particular, interested in the case when  $\mathcal{R}$  and  $\mathbf{a}_{\text{opt}}$  are smooth functions. Formally, we will stipulate that the excess risk function  $\mathcal{R}(\cdot; \cdot)$ , and the plug-in optimal decision  $\mathbf{a}_{\text{opt}}(\theta)$  satisfy the following conditions:

**Assumption 3** (Smooth Decision-Making). *There exist  $r_{\text{quad}}(\theta_\star)$  and constants  $\mu > 0$ ,  $L_{\alpha i}, L_{\mathcal{R}i}$ ,  $i \in \{1, 2, 3\}$ , and  $L_{\text{hess}}$  such that for any  $\theta$  and  $\mathbf{a}$  satisfying*

$$\|\theta - \theta_\star\|_2 \leq r_{\text{quad}}(\theta_\star), \quad \|\mathbf{a} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 \leq L_{\alpha 1} r_{\text{quad}}(\theta_\star), \quad (\text{B.1})$$

the following conditions hold

- The optimal action  $\mathbf{a}_{\text{opt}}(\theta)$  is unique, and moreover, there is a parameter  $\mu$  such that  $\mathcal{R}(\mathbf{a}'; \theta) \geq \frac{\mu}{2} \|\mathbf{a}' - \mathbf{a}_{\text{opt}}(\theta)\|_2^2$  for all  $\mathbf{a}' \in \mathbb{R}^{d_a}$  (not restricted to  $\mathbf{a}'$  satisfying Eq. (B.1)).
- $\|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}1}$ ,  $\|\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}2}$ , and  $\|\nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}3}$ .
- $\|\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)\|_{\text{op}} \leq L_{\alpha 1}$ ,  $\|\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)\|_{\text{op}} \leq L_{\alpha 2}$ , and  $\|\nabla_{\theta}^3 \mathbf{a}_{\text{opt}}(\theta)[\delta, \delta, \delta]\|_{\text{op}} \leq L_{\alpha 3}$  for all  $\delta \in \mathbb{R}^{d_\theta}$  with  $\|\delta\|_2 = 1$ .
- $\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)$  is Lipschitz in  $\theta$  with Lipschitz constant  $L_{\text{hess}}$ .

We also define:

$$L_{\text{quad}} := \frac{1}{6} (L_{\mathcal{R}3} L_{\alpha 1}^3 + 3L_{\mathcal{R}2} L_{\alpha 2} L_{\alpha 1} + L_{\mathcal{R}1} L_{\alpha 3}), \quad L_{\mathcal{H}} := 6L_{\text{quad}} + L_{\mathcal{R}2} L_{\alpha 1} + L_{\text{hess}} L_{\alpha 1}^2.$$

In the most general case, which we will refer to as *martingale decision making*, we assume that we have observations of the form

$$y_t = \langle \theta_\star, z_t \rangle + w_t, \quad w_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_w^2), \quad z_t \text{ is } \mathcal{F}_{t-1}\text{-adapted.} \quad (\text{B.2})$$

for a filtration  $(\mathcal{F}_t)_{t \geq 1}$  and scalar observations  $y_t$ . We allow the distribution of the covariates  $z_t$  to be arbitrary: for example, there may be some function  $f(\dots)$  of appropriate shape such that  $z_t = f(t, z_{1:t-1}, y_{1:t-1}, w_{1:t}, u_{1:t}, \theta_\star)$ , for inputs of our choosing  $u_{1:t}$ . We are now ready to define our decision-making setting.

**Definition B.1** (Martingale Decision Making (MDM)). Assume our excess risk  $\mathcal{R}(\hat{\mathbf{a}}; \theta_*)$  satisfies Assumption 3 and our observations are generated by (B.2). Then we call the problem of choosing a decision  $\hat{\mathbf{a}}$  to minimize  $\mathcal{R}(\hat{\mathbf{a}}; \theta_*)$  *martingale decision making* (MDM).

Our goal in MDM is to estimate  $\theta_*$  from our observations well enough to find a decision rule  $\hat{\mathbf{a}}$  that minimizes  $\mathcal{R}(\hat{\mathbf{a}}; \theta_*)$ . We will be interested in particular in the special case when (B.2) is a linear dynamical system:

$$x_{t+1} = A_* x_t + B_* u_t + w_t \quad (\text{B.3})$$

for  $w_t \sim \mathcal{N}(0, \sigma_w^2 I)$ . As we show in Appendix G.3, linear dynamical systems are a special case of (B.2). This special case defines the following restriction of MDM.

**Definition B.2** (Linear Dynamical Decision Making (LDDM)). Assume our excess risk  $\mathcal{R}(\hat{\mathbf{a}}; \theta_*)$  satisfies Assumption 3 and that our observations are generated by a linear dynamical system, (B.3). Then we call the problem of choosing a decision  $\hat{\mathbf{a}}$  to minimize  $\mathcal{R}(\hat{\mathbf{a}}; \theta_*)$  *linear dynamical decision making* (LDDM).

Given these formalizations of our problem setting, we recall our interaction protocol:

**Task-Specific Pure Exploration Problem.** The learner’s behavior is specified by an exploration policy  $\pi_{\text{exp}} : (x_{1:t}, u_{1:t-1}) \rightarrow u_t$  and decision rule  $\text{dec}$  executed in the dynamics Eq. (B.2).

1. For steps  $t = 1, \dots, T$ , the learner executes  $\pi_{\text{exp}}$  and collects a *trajectory*  $\boldsymbol{\tau} = (y_{1:T}, z_{1:T}, u_{1:T})$ .
2. For a *budget*  $\gamma^2 \geq 0$ , the inputs  $u_{1:T}$  must satisfy the constraint  $\mathbb{E}_{\pi_{\text{exp}}} [\sum_{t=1}^T \|u_t\|^2] \leq T\gamma^2$ .
3. Finally, the learner proposes a decision  $\hat{\mathbf{a}} = \text{dec}(\boldsymbol{\tau})$  as a function of  $\boldsymbol{\tau}$ .

We emphasize the generality of this set of decision-making problems. While we will show that the LQR problem satisfies this assumption, many other decision-making problems can be cast as an instance of MDM or LDDM, as we discuss in Section 1.3.

Before stating our results, we remind the reader of our definition of power constrained policies:

**Definition 2.4** (Power-Constrained Policies). Let  $\Pi_{\gamma^2}$  denote the set of causal policies that have expected average power bounded as  $\gamma^2$ . That is, for any  $\pi \in \Pi_{\gamma^2}$ , we will have  $\mathbb{E}_{\theta, \pi} [\sum_{t=1}^T \|u_t\|_2^2] \leq T\gamma^2$  for all  $\theta$ .

Finally, recall that the  $\mathcal{H}_\infty$ -norm of  $A_*$  is defined as:

$$\|A_*\|_{\mathcal{H}_\infty} := \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}} \quad (\text{B.4})$$

where  $i$  denotes the imaginary number,  $\sqrt{-1}$ .

## B.2. Lower Bound for Decision Making in MDM

We first present a lower bound in the general MDM setting. We will assume we are playing a particular exploration policy,  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$ , and our goal is to derive lower bounds on decision-making given that our trajectory is generated by  $\pi_{\text{exp}}$ . Recall the definition of the local minimax risk:

$$\mathfrak{M}_{\pi_{\text{exp}}}(\mathcal{R}; \mathcal{B}) := \min_{\text{dec}} \max_{\theta \in \mathcal{B}} \mathbb{E}_{\boldsymbol{\tau} \sim \theta, \pi_{\text{exp}}} [\mathcal{R}(\text{dec}(\boldsymbol{\tau}); \theta)],$$

and the idealized risk:

$$\Phi_T(\pi; \theta_*) := \text{tr}(\mathcal{H}(\theta_*) \boldsymbol{\Gamma}_T(\pi; \theta_*)^{-1}).$$

Our argument will show that the local minimax risk is lower bounded by the estimation error of  $\theta$  in a relevant Mahalanobis norm, which yields the familiar “inverse-trace of the covariance” sample complexity. In the MDM setting, we denote our covariance as

$$\boldsymbol{\Sigma}_T := \sum_{t=1}^T z_t z_t^\top, \quad \boldsymbol{\Gamma}_T(\pi; \theta) := \frac{1}{T} \mathbb{E}_{\theta, \pi} [\boldsymbol{\Sigma}_T]$$

For our lower bound to hold, the covariance matrices in question must satisfy two rather mild regularity conditions.



**Assumption 4** (Sufficient Excitation). *For some  $\underline{\lambda} > 0$  independent of  $T$ , and under our exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$ :*

$$\lambda_{\min}(\mathbf{\Gamma}_T(\pi_{\text{exp}}; \theta_\star)) \geq \underline{\lambda}.$$

In the special case of linear dynamical systems, Assumption 4 can be enforced by adding a small amount of white noise to any exploration policy, and the budget constraint can still be met by scaling down inputs by a constant factor.

**Assumption 5** (Smooth Response). *There exist parameters  $r_{\text{cov}}(\theta_\star) > 0$ ,  $L_{\text{cov}}(\theta_\star, \gamma^2) > 0$ ,  $C_{\text{cov}} > 0$ ,  $c_{\text{cov}} > 0$ , and  $\alpha > 0$  such that, under our exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$ , for all  $\theta$  satisfying  $\|\theta - \theta_\star\|_2 \leq r_{\text{cov}}(\theta_\star)$ , we have:*

$$\mathbf{\Gamma}_T(\pi_{\text{exp}}; \theta) \preceq c_{\text{cov}} \mathbf{\Gamma}_T(\pi_{\text{exp}}; \theta_\star) + \left( L_{\text{cov}}(\theta_\star, \gamma^2) \cdot \|\theta - \theta_\star\|_2 + \frac{C_{\text{cov}}}{T^\alpha} \right) \cdot I.$$

Intuitively, Assumption 5 says that the covariance matrices do not vary too wildly in the ground truth instances. This will be true for any “reasonable” policy, and in fact, we can show that, without loss of generality, a comparable condition holds for the policies which perform near optimal experiment design in LDDM. Under these assumptions, we obtain the following lower bound.

**Theorem B.1** (Part 2 of Theorem 2.1). *Assume we are in the MDM setting, that  $\mathcal{R}$  satisfies Assumption 3, our exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  satisfies Assumption 4 and Assumption 5, and suppose that the time horizon  $T$  satisfies*

$$\underline{\lambda}T \geq \max \left\{ \left( \frac{80d_\theta}{r_{\text{quad}}(\theta_\star)^2} \right)^{6/5}, \left( \frac{\sigma_w^2 L_{\mathcal{R}2}}{5\mu} \right)^6, \left( \frac{L_{\text{cov}}(\theta_\star, \gamma^2) \sqrt{5d_\theta}}{c_{\text{cov}} \underline{\lambda}} \right)^{12/5}, \left( \frac{2C_{\text{cov}}}{c_{\text{cov}} \underline{\lambda}^{1-\alpha}} \right)^{1/\alpha}, \left( \frac{5d_\theta}{r_{\text{cov}}(\theta_\star)^2} \right)^{6/5} \right\}$$

*Then, defining the localizing ball  $\mathcal{B}_T := \{\theta : \|\theta - \theta_\star\|_2 \leq 5d_\theta/(\underline{\lambda}T)^{5/6}\}$ , and letting  $\tau$  denote a trajectory generated by  $\pi_{\text{exp}}$  on  $\theta$ , we have*

$$\mathfrak{M}_{\pi_{\text{exp}}}(\mathcal{R}; \mathcal{B}_T) = \min_{\text{dec}} \max_{\theta \in \mathcal{B}_T} \mathbb{E}_{\tau \sim \theta, \pi_{\text{exp}}} [\mathcal{R}(\text{dec}(\tau); \theta)] \geq \frac{\sigma_w^2}{1 + 2c_{\text{cov}}} \cdot \frac{\Phi_T(\pi_{\text{exp}}; \theta_\star)}{T} - \frac{C_{\text{lb}}}{(\underline{\lambda}T)^{5/4}}$$

where,

$$C_{\text{lb}} = c_1 \left( (L_{a1} L_{a2} L_{\mathcal{R}2} + L_{a1}^3 L_{\mathcal{R}3} + L_{\text{hess}}) d_\theta^{3/2} + L_{a1}^2 L_{\mathcal{R}2} \right)$$

for a universal constant  $c_1$ .

This result is itself a corollary of a more general result, Theorem E.3, which provides a lower bound without Assumption 4 or Assumption 5. We prove this result in Appendix E. We emphasize again that Theorem B.1 does not require that the data be generated from a linear dynamical system—it holds for *any* loss satisfying Assumption 3 so long as our observations follow Eq. (B.2). However, as we show in Appendix H.1, Assumption 4 and Assumption 5 are met for a fairly general set of policies in linear dynamical systems, and a clean corollary of this result may be stated in the LDDM setting.

### B.3. Lower Bound for Optimal Decision Making in LDDM

We turn now to the LDDM setting, and prove a lower bound that holds for *all* exploration policies  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$ . We first define the following:

$$\lambda_{\text{noise}}^*(\sigma_u) := \min \left\{ \lambda_{\min} \left( \sigma_w^2 \sum_{s=0}^{d_x-1} A_\star^s (A_\star^s)^\top + \sigma_u^2 \sum_{s=0}^{d_x-1} A_\star^s B_\star B_\star^\top (A_\star^s)^\top \right), \sigma_u^2 \right\}$$

and in particular set:

$$\lambda_{\text{noise}}^* := \lambda_{\text{noise}}^*(\gamma/\sqrt{2d_u})$$

Note that

$$\lambda_{\text{noise}}^*(\sigma_u) = \lambda_{\min} \left( \mathbb{E} \left[ \begin{bmatrix} x_{d_x} \\ u_{d_x} \end{bmatrix} \begin{bmatrix} x_{d_x} \\ u_{d_x} \end{bmatrix}^\top \mid u_s \sim \mathcal{N}(0, \sigma_u^2 I), s = 0, \dots, t \right] \right)$$

so it follows that  $\lambda_{\text{noise}}^*(\sigma_u)$  is the minimum eigenvalues of the covariates when we play isotropic noise, and can be thought of as a measure of how easily the system can be excited. We make the following assumption.

**Assumption 6.**  $\theta_*$  and  $\sigma_w$  are such that  $\lambda_{\text{noise}}^* > 0$ . In particular, it suffices that  $\sigma_w > 0$ , or the system is controllable.

Before stating our result, we recall the definition of the exploration local minimax risk:

$$\mathfrak{M}_{\gamma^2}(\mathcal{R}; \mathcal{B}) := \min_{\pi_{\text{exp}} \in \Pi_{\gamma^2}} \min_{\text{dec}} \max_{\theta \in \mathcal{B}} \mathbb{E}_{\tau \sim \theta, \pi_{\text{exp}}} [\mathcal{R}(\text{dec}(\tau); \theta)].$$

and the optimal risk:

$$\Phi_{\text{opt}}(\gamma^2; \theta_*) := \liminf_{T \rightarrow \infty} \inf_{\pi_{\text{exp}} \in \Pi_{\gamma^2}} \Phi_T(\pi_{\text{exp}}; \theta_*),$$

We then have the following.

**Theorem B.2** (Part 2 of Theorem 2.2). *Assume we are in the LDDM setting and consider a loss function  $\mathcal{J}_\theta(\mathbf{a}) : \mathbb{R}^{d_a} \rightarrow \mathbb{R}$  with induced excess risk  $\mathcal{R}(\mathbf{a}; \theta_*) := \mathcal{J}_{\theta_*}(\mathbf{a}) - \inf_{\mathbf{a}'} \mathcal{J}_{\theta_*}(\mathbf{a}')$ . Fix a model  $\theta_*$  and time horizon  $T$ . Suppose that*

- $\mathcal{R}$  satisfies the smoothness condition, Assumption 3.
- The model  $\theta_*$  satisfies the excitation assumption Assumption 6 with parameter  $\lambda_{\text{noise}}^* > 0$ .
- The time horizon satisfies  $T \geq \max \left\{ C_{\text{lb}}^{\text{init}}, \left( \frac{80(d_x^2 + d_x d_u)}{(\lambda_{\text{noise}}^*)^{5/6} r_{\text{quad}}(\theta_*)^2} \right)^{6/5}, \left( \frac{\sigma_w^2 L \mathcal{R}_2}{5\mu} \right)^6 \right\}$ .

Finally, define the localized ball of instances

$$\mathcal{B}_T := \{ \|\theta - \theta_*\|_{\mathbb{F}}^2 \leq 5(d_x^2 + d_x d_u) / (\lambda_{\text{noise}}^* T^{5/6}) \}$$

Then, any decision rule  $\text{dec}(\tau)$  suffers the following lower bound

$$\mathfrak{M}_{\gamma^2}(\mathcal{R}; \mathcal{B}_T) = \min_{\pi_{\text{exp}} \in \Pi_{\gamma^2}} \min_{\text{dec}} \max_{\theta \in \mathcal{B}_T} \mathbb{E}_{\tau \sim \theta, \pi_{\text{exp}}} [\mathcal{R}(\text{dec}(\tau); \theta)] \geq \frac{\sigma_w^2}{64} \cdot \frac{\Phi_{\text{opt}}(\gamma^2; \theta_*)}{T} - \frac{C_{\text{lb}}}{(\lambda_{\text{noise}}^* T)^{5/4}}$$

where above,

$$C_{\text{lb}}^{\text{init}} = \text{poly} \left( d_x, d_u, \|B_*\|_{\text{op}}, \|A_*\|_{\mathcal{H}_\infty}, \gamma^2, \sigma_w^2, \frac{1}{\lambda_{\text{noise}}^*}, \log T \right)$$

and  $C_{\text{lb}}$  is defined as in Theorem B.1 with  $d_\theta = d_x^2 + d_x d_u$ .

We emphasize that this result holds for any exploration policy with bounded power,  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$ . As such, it provides a lower bound on optimal decision-making. We prove Theorem B.2 in Appendix H.2.

#### B.4. Upper Bound for Certainty Equivalence Decision Making in MDM

We next consider upper bounds on decision making in the MDM setting when we are playing a fixed exploration policy  $\pi_{\text{exp}}$ . Given some data  $\{(y_t, z_t, u_t)\}_{t=1}^T$  generated by playing  $\pi_{\text{exp}}$  on Eq. (B.2), we define our estimator of  $\theta_*$  as

$$\hat{\theta}_{\text{ls}} = \min_{\theta} \sum_{t=1}^T \|y_t - \theta^\top z_t\|_2^2 \quad (\text{B.5})$$

The following is a sufficient assumption on  $\pi_{\text{exp}}$  to guarantee the efficiency of certainty equivalence decision making. Recall that  $\Sigma_T$  denotes the random covariates.

**Assumption 7** (Exploration Policy Regularity). *We assume that the true instance  $\theta_*$  and policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  satisfy the following regularity conditions:*

- There exists some time  $T_{\text{se}}(\pi_{\text{exp}})$  such that for any  $T \geq T_{\text{se}}(\pi_{\text{exp}})$  the system is sufficiently excited. That is, if  $T \geq T_{\text{se}}(\pi_{\text{exp}})$ :

$$\mathbb{P}_{\theta_*, \pi_{\text{exp}}} \left[ \lambda_{\min}(\Sigma_T) \geq \underline{\lambda} T, \Sigma_T \preceq T \bar{\Gamma}_T \right] \geq 1 - \delta$$

for deterministic  $\underline{\lambda} > 0$  and  $\bar{\Gamma}_T \succeq 0$ .

- There exists some time  $T_{\text{con}}(\pi_{\text{exp}})$  such that, for any  $T \geq T_{\text{con}}(\pi_{\text{exp}})$ , the covariates have concentrated to their mean. That is, if  $T \geq T_{\text{con}}(\pi_{\text{exp}})$ :

$$\mathbb{P}_{\theta_*, \pi_{\text{exp}}} \left[ \|\Sigma_T - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]\|_{\text{op}} \leq \frac{C_{\text{con}}}{T^\alpha} \lambda_{\min}(\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]) \right] \geq 1 - \delta$$

for deterministic  $C_{\text{con}} > 0$  and  $\alpha > 0$ .

The following result precisely quantifies the loss of the certainty equivalence decision-making rule under this assumption on the policy.

**Theorem B.3** (Part 1 of Theorem 2.1). *Assume we are in the MDM setting with some loss  $\mathcal{R}$  satisfying Assumption 3 and exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  which satisfies Assumption 7 with minimal times  $T_{\text{con}}(\pi_{\text{exp}})$  and  $T_{\text{se}}(\pi_{\text{exp}})$  and covariance lower bound  $\underline{\lambda} > 0$ . If*

$$T > \max \left\{ T_{\text{con}}(\pi_{\text{exp}}), T_{\text{se}}(\pi_{\text{exp}}), (4C_{\text{con}})^{1/\alpha}, \frac{c_1(\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T/\underline{\lambda} + I))}{\underline{\lambda} r_{\text{quad}}(\theta_*)^2} \right\} \quad (\text{B.6})$$

then for  $\delta \in (0, 1/2)$ , with probability  $1 - \delta$ , the certainty equivalence decision rule achieves the following rate,

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_{\text{ls}}); \theta_*) \leq 5\sigma_w^2 \log \frac{24d_\theta}{\delta} \cdot \frac{\Phi_T(\pi_{\text{exp}}; \theta_*)}{T} + \frac{C_{\text{ce},1}}{T^{3/2}} + \frac{C_{\text{ce},2}}{T^{1+2\alpha}}$$

where we let  $c_1, c_2, c_3$  be universal numerical constants and set

$$C_{\text{ce},1} := \frac{c_2 L_{\text{quad}}}{\underline{\lambda}^{3/2}} \left( \log \frac{1}{\delta} + d_\theta + \log \det(\bar{\Gamma}_T/\underline{\lambda} + I) \right)^{3/2}, \quad C_{\text{ce},2} := \frac{c_3 \sigma_w^2 C_{\text{con}}^2 d_\theta \text{tr}(\mathcal{H}(\theta_*))}{\underline{\lambda}} \log \frac{d_\theta}{\delta}.$$

We note that this upper bound matches the lower bound given in Theorem B.1. This shows that the certainty equivalence decision rule is instance optimal for *any* decision-making problem in the MDM setting.

The proof of this result is given in Appendix F.1. The burn-in time (B.6) and lower-order terms have transparent interpretations. For the burn-in, the requirement that  $T$  be larger than  $T_{\text{con}}(\pi_{\text{exp}})$  and  $T_{\text{se}}(\pi_{\text{exp}})$  is necessary to ensure that the concentration and excitation events stated in Assumption 7 hold with high probability. The requirement that  $T$  be larger than  $(4C_{\text{con}})^{1/\alpha}$  is necessary to ensure that the covariates have concentrated enough for our  $M$ -norm estimation bound, Theorem D.2, to hold. Finally, the last term in the burn-in ensures that our estimate  $\hat{\theta}_{\text{ls}}$  is in a ball of radius  $r_{\text{quad}}(\theta_*)$  around  $\theta_*$ , which allows us to approximate  $\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_{\text{ls}}); \theta_*)$  by a quadratic. The lower order terms similarly yield intuitive explanations.  $C_{\text{ce},1}/T^{3/2}$  quantifies the additional loss due to the error in our quadratic approximation of  $\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_{\text{ls}}); \theta_*)$ , while  $C_{\text{ce},2}/T^{1+2\alpha}$  is due to the lower order term given in our  $M$ -norm estimation bound, Theorem D.2.

#### B.4.1. COROLLARY: CERTAINTY-EQUIVALENCE DECISION MAKING IN LDDM

While Theorem B.3 holds in a very general setting, our optimal decision-making algorithm, TOPLE, applies only to the LDDM setting, and uses a highly structured set of policies. In order to facilitate the analysis of TOPLE, it is helpful to obtain a corollary of Theorem B.3 in this more restricted setting. Towards making this precise, we introduce a set of policies in the LDDM setting, *sequential-open loop* policies, which we show contains TOPLE. Before formally defining these policies, we need the following piece of notation:

**Definition B.3.** Let

$$\bar{\gamma} := c \left( (1 + \|B_\star\|_{\text{op}}^2) \|A_\star\|_{\mathcal{H}_\infty}^4 \right) \left( \sqrt{d_x} \log \frac{T}{\delta} + \gamma^2 \|A_\star\|_{\mathcal{H}_\infty}^2 \right).$$

By Lemma I.3,  $\bar{\Gamma}_T := T\bar{\gamma} \cdot I$  is a high probability upper bound on the covariates, assuming that  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$ , where we define  $\Pi_{\gamma^2}^{\text{sol}}$  below.

With this definition in place, we introduce the set of (sufficiently regular) sequential open loop policies satisfying the budgeted constraint, denoted  $\Pi_{\gamma^2}^{\text{sol}}$ .

**Definition B.4** (Sequential Open-Loop Policies). We define a *sequential open-loop policy* to be an exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  satisfying the following conditions:

- **(Open-Loop Gaussian)** There exist deterministic times  $\{\bar{t}_0, \bar{t}_1, \dots, \bar{t}_{n-1}, \bar{t}_n\} \subseteq [T]$  with  $\bar{t}_0 = 0, \bar{t}_n = T, \bar{t}_{i+1} \geq \bar{t}_i$ , such that, for  $t \in \{\bar{t}_i, \dots, \bar{t}_{i+1} - 1\}$ :

$$u_t | \mathcal{F}_{\bar{t}_i} \sim \mathcal{N}(\tilde{u}_t, \Lambda_{u,i})$$

for  $\mathcal{F}_{\bar{t}_i}$  measurable  $\tilde{u}_t$  and  $\Lambda_{u,i} \succeq 0$  satisfying:

$$\sum_{t=0}^{T-1} \tilde{u}_t^\top \tilde{u}_t \leq T\gamma^2, \quad \text{tr}(\Lambda_{u,i}) \leq \gamma^2, \quad \lambda_{\min}(\Lambda_{u,i}) \geq \sigma_u^2$$

almost surely, for deterministic  $\sigma_u$ .

- **(Low-Switching)** For any  $t$ , there exists some epoch  $i$  such that  $|\{t, \dots, t + T_{\text{se}}(\pi_{\text{exp}})\} \cap \{\bar{t}_i, \dots, \bar{t}_{i+1} - 1\}| \geq \frac{1}{2}T_{\text{se}}(\pi_{\text{exp}})$  where

$$T_{\text{se}}(\pi_{\text{exp}}) := c_1 d_x \left( (d_x + d_u) \log(\bar{\gamma}/\lambda_{\text{noise}}^*(\sigma_u) + 1) + \log \frac{n}{\delta} \right).$$

In words, at least half of any length  $T_{\text{se}}(\pi_{\text{exp}})$  interval is contained in a single epoch.

We make several comments on this definition.

- Any policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$  satisfies Assumption 7, which we prove in Appendix F.1.
- As we show in Appendix J, TOPLE, an optimal policy (up to constants), is in  $\Pi_{\gamma^2}^{\text{sol}}$ , with  $n = \mathcal{O}(\log T)$ .
- The assumption that  $u_t | \mathcal{F}_{T_i}$  be Gaussian is for simplicity of analysis, and in general is not necessary—the noise could take different sub-Gaussian distributions if desired.

The following result instantiates Theorem B.3 with any policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$ , and assuming we are in the LDDM setting.

**Corollary 2.** Assume we are in the LDDM setting and consider some loss  $\mathcal{R}$  satisfying Assumption 3, stable system  $\theta_*$ , and exploration policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$ . If

$$T \geq c_1 \left( \frac{C_{\text{sys}} n^2 (\sigma_w^4 + \gamma^4)}{\lambda_{\text{noise}}^*(\sigma_u)^2} + \frac{d_x}{\lambda_{\text{noise}}^*(\sigma_u) r_{\text{quad}}(\theta_*)^2} + \frac{\sqrt{d_x} \sigma_w^2}{\gamma^2} + n + d_x \right) \left( \log \frac{n}{\delta} + d \log(\bar{\gamma}/\lambda_{\text{noise}}^*(\sigma_u) + 3) \right) \quad (\text{B.7})$$

then for  $\delta \in (0, 1/3)$ , with probability  $1 - \delta$ :

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_{\text{ls}}); \theta_*) \leq 5\sigma_w^2 \log \frac{24(d_x^2 + d_x d_u)}{\delta} \cdot \frac{\Phi_T(\pi_{\text{exp}}; \theta_*)}{T} + \frac{C_{\text{ce},1}}{T^{3/2}} + \frac{C_{\text{ce},2}}{T^2}$$

where  $C_{\text{sys}} = \text{poly}(\|B_\star\|_{\text{op}}, \|A_\star\|_{\mathcal{H}_\infty})$ ,  $d := d_x + d_u$ , universal numerical constants  $c_1, c_2$ , and

$$C_{\text{ce},1} := \frac{c_2 L_{\text{quad}}}{\lambda_{\text{noise}}^*(\sigma_u)^{3/2}} \left( \log \frac{1}{\delta} + d_x d \log(\bar{\gamma}/\lambda_{\text{noise}}^*(\sigma_u) + 3) \right)^{3/2}, \quad C_{\text{ce},2} := \frac{C_{\text{sys}} \sigma_w^2 (\sigma_w^4 + \gamma^4) \text{tr}(\mathcal{H}(\theta_*)) d^3 n}{\lambda_{\text{noise}}^*(\sigma_u)^3} \log^2 \frac{dn}{\delta}.$$

We prove this result in Appendix I.

### B.5. Efficient Experiment Design in Frequency Domain (LDDM)

Before presenting the formal definition of TOPLE, we establish the relevant experiment design preliminaries, which are best stated in frequency domain. From this, we will specify TOPLE, and then show that it is indeed a sequential open loop policy, in the sense of Definition B.4. Recall that, in the LDDM setting, we set

$$\Gamma_T(\pi; \theta) := \frac{1}{T} \mathbb{E}_{\theta, \pi} \left[ \sum_{t=1}^T \begin{bmatrix} x_t \\ u_t \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top \right], \quad \mathbf{\Gamma}_T(\pi; \theta) := I_{d_x} \otimes \Gamma_T(\pi; \theta)$$

Through the remainder of this section, we will use the convention that in the LDDM setting bold matrices denote Kronecker products,  $\mathbf{\Gamma} := I_{d_x} \otimes \Gamma$ . Note that, by mapping our linear dynamical system (B.3) to our general regression setting (B.2) through the reduction given in Appendix G.3, this definition is consistent with our definition of  $\mathbf{\Gamma}$  in the MDM setting.

## B.5.1. FREQUENCY-DOMAIN REPRESENTATIONS

We let bold vectors  $\mathbf{u} = (u_i)_{i=1}^k \in \mathbb{C}^{kd_u}$  denote sequences of inputs, and denote their discrete-time Fourier transform (DFT)

$$\tilde{\mathbf{u}} = (\tilde{u}_s)_{s=1}^k = \mathfrak{F}(\mathbf{u}) \in \mathbb{C}^{kd_u}, \quad \text{where } \tilde{u}_s = \sum_{s=1}^k u_s \exp\left(\frac{2\pi i s}{k}\right) \quad (\text{B.8})$$

The mapping  $\mathfrak{F}$  is invertible, though in general  $\mathfrak{F}^{-1} : \mathbb{C}^{kd_u} \rightarrow \mathbb{C}^{kd_u}$ . However, if our frequency-domain representation is *symmetric*, we have that the inverse DFT is purely real.

**Definition B.5** (Symmetric Signal). We say that  $\tilde{\mathbf{u}} = (\tilde{u}_s)_{s=1}^k \in \mathbb{C}^{kd_u}$  is *symmetric* if  $\tilde{u}_s = \text{conj}(\tilde{u}_{k-s})$  for  $s < k$  and  $\tilde{u}_k$  is purely real, where  $\text{conj}(\cdot)$  denotes the complex conjugate.

**Fact B.1.**  $\mathfrak{F}^{-1}(\tilde{\mathbf{u}})$  is a vector with real coefficients if and only if  $\tilde{\mathbf{u}}$  is symmetric.

We now consider a convex relation of the outerproduct of this DFT. First, some preliminaries. For a complex vector  $z \in \mathbb{C}^d$  (resp. matrix  $A \in \mathbb{C}^{d_1 \times d_2}$ ), let  $z^H$  (resp.  $A^H$ ) denote its Hermitian adjoint; i.e., the complex conjugate of its transpose. We denote the set of *Hermitian* matrices as  $\mathbb{H}^d := \{A \in \mathbb{C}^{d \times d} : A^H = A\}$ , and the set of positive-semidefinite Hermitian matrices  $\mathbb{H}_+^d := \{A \in \mathbb{H}^d : z^H A z \geq 0, \quad \forall z \in \mathbb{C}^d\}$ . Given  $\tilde{\mathbf{u}} = (\tilde{u}_\ell)_{\ell=1}^k \in \mathbb{C}^{kd_u}$ , we define its outerproduct as the sequence of complex-rank one Hermitian matrices,  $\mathbf{U} = (U_\ell)_{\ell=1}^k$ , defined by

$$\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}} := \mathbf{U} = (U_\ell)_{\ell=1}^k, \quad \text{where } U_\ell = \tilde{u}_\ell \tilde{u}_\ell^H \in \mathbb{H}_+^d. \quad (\text{B.9})$$

We now define the following set, which relaxes outer products to matrix sequences of the above form, with a total power constraint on their trace:

$$\mathcal{U}_{\gamma^2, k} := \left\{ \mathbf{U} = (U_\ell)_{\ell=1}^k : U_\ell \in \mathbb{H}_+^d, \quad \mathbf{U} \text{ is symmetric, } \sum_{\ell=1}^k \text{tr}(U_\ell) \leq k^2 \gamma^2 \right\} \quad (\text{B.10})$$

Critically,  $\mathcal{U}_{\gamma^2, k}$  is convex. We generalize the definition of symmetric signals here to matrices, defining it identically as we have defined symmetric vector signals. The following class of sequences  $\mathbf{U}$  are of particular importance.

**Definition B.6** (Rank One Relaxation). We say that  $\mathbf{U} = \{U_\ell : 1 \leq \ell \leq k\} \in \mathcal{U}_{\gamma^2, k}$  is *rank one* if there exists a vector  $\tilde{\mathbf{u}} \in \mathbb{C}^{kd_u}$  such that  $\mathbf{U} = \tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}$ .

Lastly, we define a frequency-domain covariance operator defined on  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$ :

$$\Gamma_k^{\text{freq}}(\theta, \mathbf{U}) := \frac{1}{k} \sum_{\ell=1}^k (e^{t \frac{2\pi \ell}{k}} I - A)^{-1} B U_\ell B^H (e^{t \frac{2\pi \ell}{k}} I - A)^{-H}, \quad (\text{B.11})$$

$$\Gamma_{t,k}^{\text{freq}}(\theta, \mathbf{U}) := \frac{t}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) \quad (\text{B.12})$$

We will overload notation, defining

$$\Gamma_k^{\text{freq}}(\theta, \tilde{\mathbf{u}}) = \Gamma_k^{\text{freq}}(\theta, \tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}), \quad \Gamma_{t,k}^{\text{freq}}(\theta, \tilde{\mathbf{u}}) = \Gamma_{t,k}^{\text{freq}}(\theta, \tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}) \quad (\text{B.13})$$

$$\Gamma_k^{\text{freq}}(\theta, \mathbf{u}) = \Gamma_k^{\text{freq}}(\theta, \tilde{\mathbf{u}}), \quad \Gamma_{t,k}^{\text{freq}}(\theta, \mathbf{u}) = \Gamma_{t,k}^{\text{freq}}(\theta, \tilde{\mathbf{u}}) \quad (\text{B.14})$$

for  $\mathbf{u} = \mathfrak{F}^{-1}(\tilde{\mathbf{u}})$ . The following result shows that  $\Gamma_k^{\text{freq}}(\theta, \mathbf{u})$  corresponds to the steady-state covariates of our system when an input  $\mathbf{u}$  is played.

**Proposition B.4.** Let  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$  be rank one, with  $\mathbf{U} = \tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}$ ,  $\tilde{\mathbf{u}} \in \mathbb{C}^{kd_u}$ . Let  $\mathbf{u} = (u_t)_{t=1}^k = \mathfrak{F}^{-1}(\tilde{\mathbf{u}})$ . Define the extended inputs

$$\mathbf{u}_{1:T}^{\text{ext}} = (u_t^{\text{ext}})_{t \geq 1}, \quad \text{where } u_t^{\text{ext}} = u_{\text{mod}(t, k)}$$

Finally, let  $x^{\mathbf{u}}$  denote the evolution of the dynamical system obtained by starting at initial state  $x_0$  and executing the input  $u_t^{\text{ext}}$ . Then

1.  $\frac{1}{k}\Gamma_k^{\text{freq}}(\theta, \mathbf{U}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=0}^{T-1} x_s^{\mathbf{u}} (x_s^{\mathbf{u}})^\top = \lim_{T \rightarrow \infty} \frac{1}{T} \Gamma_T^{\text{in}}(\theta, \mathbf{u}^{\text{ext}}, x_0)$ .
2. Let  $k' \geq k$  be divisible by  $k$ . Let  $\mathbf{u}' = (u_s^{\text{ext}})_{s=1}^{k'}$ , and define the frequency domain quantities
 
$$\check{\mathbf{u}}' = (\check{u}_\ell)_{\ell=1}^{k'} = \mathfrak{F}(\mathbf{u}'), \quad \mathbf{U}' = \check{\mathbf{u}}' \otimes \check{\mathbf{u}}'. \quad (\text{B.15})$$

Then  $\Gamma_{k'}^{\text{freq}}(\theta, \mathbf{U}') = \frac{k'}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U})$ .

**Noise-Augmented Covariances.** We shall also study the covariance matrix that arises from exciting the system with white noise of covariance  $\Lambda_u$ , when the process noise has covariance  $\Lambda_w$ :

$$\Gamma_t^{\text{noise}}(\theta, \Lambda_u) := \sum_{s=0}^{t-1} A^s \Lambda_w (A^s)^\top + \sum_{s=0}^{t-1} A^s B \Lambda_u B^\top (A^s)^\top \quad (\text{B.16})$$

We will overload notation and set

$$\Gamma_t^{\text{noise}}(\theta, \sigma_u) := \Gamma_t^{\text{noise}}(\theta, \sigma_u^2 I)$$

Since the Fourier transform preserves Gaussianity, the relevant covariance matrices become:

$$\Gamma_{T,k}^{\text{ss}}(\theta, \mathbf{U}, \sigma_u) := \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) + \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \sigma_u) \quad (\text{B.17})$$

If  $\mathbf{U}$  is rank one, then Proposition B.4 implies that  $\Gamma_{T,k}^{\text{ss}}(\theta, \mathbf{U}, \sigma_u)$  corresponds to the expected steady-state covariates of the noisy system when playing inputs  $\mathbf{U}$ . If  $\mathbf{U}$  is not rank one, then  $\Gamma_{T,k}^{\text{ss}}(\theta, \mathbf{U}, \sigma_u)$  corresponds to the expected steady-state covariates of the noisy system when playing a sequence of inputs formed by decomposing  $\mathbf{U}$  into rank one inputs, as in Algorithm 3.

#### B.5.2. THE STEADYSTATEDESIGN SUBROUTINE

Using the preliminaries laid out in Appendix B.5.1, we define two subroutines of TOPLE before stating the full algorithm. We first state our experiment-design subroutine, which computes the certainty-equivalence task-optimal inputs.

---

#### Algorithm 2 SteadyStateDesign( $\hat{\theta}, t, k, \gamma$ )

---

- 1: Input time horizon  $t$ , signal length  $k$ , budget  $\gamma > 0$ , model estimate  $\hat{\theta}$
- 2: Let  $\mathcal{U}_{\gamma^2/2, k}$  be the lifted representation of inputs defined in (B.10)
- 3: Set  $\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k} \subset (\mathbb{H}_+^d)^k$

$$\mathbf{U} \leftarrow \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k}} \text{tr} \left( \mathcal{H}(\hat{\theta}) \cdot \Gamma_{t,t/d_u}^{\text{ss}}(\hat{\theta}, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) \quad (\text{B.18})$$

- 4: **return**  $\mathbf{U}$
- 

Algorithm 2 chooses the input  $\mathbf{U}$  to minimize a function of the steady-state covariates of the system  $\hat{\theta}$ . In particular, observe that the objective is the steady-state analogue of the lower bound given in Theorem B.2, and we can therefore interpret this routine as choosing the inputs that minimize the lower bound for our estimated system.

**Implementation via Projected Gradient Descent.** Note that the set  $\mathcal{U}_{\gamma^2, k}$  is convex, and that the objective is also convex, due to the convexity of  $\text{tr}(X^{-1})$  and since  $\Gamma_{t,t/d_u}^{\text{ss}}(\hat{\theta}, \mathbf{U}, \gamma/\sqrt{2d_u})$  is affine in  $\mathbf{U}$ . It follows that (B.18) can be efficiently solved with any SDP solver. The structure of  $\mathcal{U}_{\gamma^2, k}$ , however, allows for an even more efficient solution. Note that any  $\mathbf{U}$  can be projected onto  $\mathcal{U}_{\gamma^2, k}$  by computing the SVD of each  $U_\ell \in \mathbf{U}$ , an operation which takes time  $\mathcal{O}(kd_u^3)$ . Therefore, (B.18) can be efficiently solved by running the following projected gradient descent update:

$$\begin{aligned} \mathbf{U}_{i+1} &\leftarrow \mathbf{U}_i - \eta \nabla_{\mathbf{U}} \Gamma_{t,t/d_u}^{\text{ss}}(\hat{\theta}, \mathbf{U}_i, \gamma/\sqrt{2d_u}) \\ \mathbf{U}_{i+1} &\leftarrow \text{proj}(\mathbf{U}_{i+1}; \mathcal{U}_{\gamma^2/2, k}) \end{aligned}$$

where  $\text{proj}(\mathbf{U}_{i+1}; \mathcal{U}_{\gamma^2/2, k})$  denotes the projection of  $\mathbf{U}_{i+1}$  onto  $\mathcal{U}_{\gamma^2/2, k}$ .

## B.5.3. THE CONSTRUCTTIMEINPUT SUBROUTINE

In order to efficiently solve our experiment design problem, we allow our input set to contain inputs that are not rank one. While this relaxation ensures our input set is convex, for a given  $\mathbf{U} \in \mathcal{U}_{\gamma^2/2,k}$  that is not rank one, it is not clear if  $\mathbf{U}$  can be implemented in the time domain. Indeed, Proposition B.4 shows that, if  $\mathbf{U}$  is rank one, there exists some time domain input  $\mathbf{u} = (u_t)_{t=1}^k$  such that

$$\frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} x_s^{\mathbf{u}} (x_s^{\mathbf{u}})^\top$$

which implies that we can approximately realize the response covariates  $\Gamma_{T,k}^{\text{SS}}(\theta, \mathbf{U}, \gamma/\sqrt{2d_u})$  in the time domain, but this relationship no longer holds if  $\mathbf{U}$  is not rank one. To remedy this, we propose the following procedure, which decomposes an arbitrary, not necessarily rank one, input  $\mathbf{U}$  into a sequence of inputs that can be realized in the time domain.

---

**Algorithm 3** ConstructTimeInput( $\mathbf{U}, T, k$ )
 

---

- 1: Denote eigendecompositions  $U_\ell = \sum_{j=1}^{d_u} \lambda_{\ell,j} v_{\ell,j} v_{\ell,j}^H, \ell = 1, \dots, k, U_\ell \in \mathbf{U}$
  - 2:  $u_t = \mathbf{0} \in \mathbb{R}^{d_u}$  for  $t = 1, \dots, d_u T$
  - 3: **for**  $j = 1, \dots, d_u$  **do**
  - 4:      $\check{u}_{\ell,j} \leftarrow \sqrt{d_u \lambda_{\ell,j}} v_{\ell,j}$  for  $\ell = 1, \dots, k$
  - 5:     **for**  $n = 1, \dots, T/k$  **do**
  - 6:          $u_{(j-1)T+(n-1)k+1}, \dots, u_{(j-1)T+nk} \leftarrow \mathfrak{F}^{-1}(\check{u}_{1,j}, \dots, \check{u}_{k,j})$
  - 7: **return**  $u_1, \dots, u_{d_u T}$
- 

As the following result shows, ConstructTimeInput produces a time domain input which realizes the response covariates  $\Gamma_k^{\text{freq}}(\theta, \mathbf{U})$  for arbitrary  $\mathbf{U}$ .

**Proposition B.5.** Let  $\mathbf{U} \in \mathcal{U}_{\gamma^2,k}$  not necessarily rank one. Let  $\mathbf{u}_m = (u_t)_{t=1}^{d_u m k}$  denote the time-domain input returned by calling ConstructTimeInput( $\mathbf{U}, mk, k$ ) with  $m$  an integer. Then

$$\frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) = \lim_{m \rightarrow \infty} \frac{1}{d_u m k} \sum_{t=0}^{d_u m k} x_t^{\mathbf{u}_m} (x_t^{\mathbf{u}_m})^\top$$

and, furthermore, the input satisfies  $\sum_{t=1}^{d_u m k} u_t^\top u_t \leq d_u m k \gamma^2$ .

**B.6. Optimal LDDM Decision-Making: Formal Statement and Guarantee for TOPLE**

Finally, we provide a formal definition of TOPLE (Algorithm 4), and a formal guarantee for its performance. Note that TOPLE applies in the LDDM setting.

---

**Algorithm 4** Task OPTimal Experiment Design (TOPLE)
 

---

- 1: **Input:** Input power  $\gamma^2$ , initial epoch length  $C_{\text{init}} d_u$  ( $C_{\text{init}} \in \mathbb{N}$ )
  - 2:  $T_0 \leftarrow C_{\text{init}} d_u, k_0 \leftarrow C_{\text{init}}, T \leftarrow T_0$
  - 3: Run system for  $T_0$  steps with  $u_t \sim \mathcal{N}(0, \frac{\gamma^2}{d_u} I)$
  - 4: **for**  $i = 1, 2, 3, \dots$  **do**
  - 5:      $\hat{\theta}_{i-1} \leftarrow \arg \min_{\theta} \sum_{t=1}^T \|x_{t+1} - \theta[x_t; u_t]\|_2^2$
  - 6:      $T_i \leftarrow T_0 2^i, k_i \leftarrow k_0 2^{\lfloor i/4 \rfloor}, T \leftarrow T + T_i$
  - 7:      $\mathbf{U}_i \leftarrow \text{SteadyStateDesign}(\hat{\theta}_{i-1}, T_i, k_i, \gamma)$
  - 8:      $(\tilde{u}_t^i)_{t=1}^{T_i} \leftarrow \text{ConstructTimeInput}(\mathbf{U}_i, T_i/d_u, k_i)$
  - 9:     Run system for  $T_i$  steps with  $u_t = \tilde{u}_t^i + u_t^w, u_t^w \sim \mathcal{N}(0, \frac{\gamma^2}{2d_u} I)$
- 

TOPLE begins by injecting isotropic Gaussian noise into the system to achieve a minimum degree of excitation. It then solves a sequence of experiment design problems on the estimated system,  $\hat{\theta}_{i-1}$ , and then plays the inputs that would optimally

excite  $\hat{\theta}_{i-1}$ . Due to the computational efficiency of `SteadyStateDesign`, TOPLE is computationally efficient. Note that by construction we will always have that  $k_{i+1}$ ,  $T_{i+1}/d_u$ , and  $T_{i+1}/(d_u k_{i+1})$  are integers, so all quantities in the algorithm and subroutines are well-defined. The following assumption quantifies how large  $T$  must be to guarantee we achieve the optimal rate.

**Assumption 8** (Sufficiently Large  $T$ ).  $T$  is large enough that the burn-in time of Corollary 2, (B.7), is met with  $n = c_1 \log T$  and  $\lambda_{\text{noise}}^*(\sigma_u) = \lambda_{\text{noise}}^*$ , and

$$T \geq \max \left\{ \frac{C_{\text{TOPLE}}^{\text{init}} \sqrt{d_x} (\sigma_w^2 + 1)}{\lambda_{\text{noise}}^*}, \frac{c_2 (\log \frac{1}{\delta} + d \log(\bar{\gamma}/\lambda_{\text{noise}}^* + 1))}{\min\{r_{\text{cov}}(\theta_*)^2, d_x^{-1} r_{\text{quad}}(\theta_*)^2, (\lambda_{\text{noise}}^*)^2 L_{\text{cov}}(\theta_*, \gamma^2)^{-2}\} \lambda_{\text{noise}}^*} \right\} \quad (\text{B.19})$$

where

$$C_{\text{TOPLE}}^{\text{init}} = \text{poly} \left( \|A_\star\|_{\mathcal{H}_\infty}, \|B_\star\|_{\text{op}}, d_u, \gamma^2, \log \frac{1}{\delta}, C_{\text{init}} \right),$$

$r_{\text{cov}}(\theta_*) = C_{\text{sys}}^{-1}$  is defined as in Lemma H.5 for some  $C_{\text{sys}} = \text{poly}(\|A_\star\|_{\mathcal{H}_\infty}, \|B_\star\|_{\text{op}})$ ,  $d = d_x + d_u$ , and  $c_1, c_2$  are universal numerical constants.

Then we have the following theorem, upper bounding the loss achieved by TOPLE.

**Theorem B.6** (Part 1 of Theorem 2.2). Assume we are in the LDDM setting, that  $\mathcal{R}$  satisfies Assumption 3,  $\delta \in (0, 1/3)$ , and that  $T$  is large enough for Assumption 8 to hold. Then with probability at least  $1 - \delta$ , the estimate  $\hat{\theta}_T$  produced by Algorithm 4 satisfies:

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_T); \theta_*) \leq 480 \sigma_w^2 \log \frac{72(d_x^2 + d_x d_u)}{\delta} \cdot \frac{\Phi_{\text{opt}}(\gamma^2; \theta_*)}{T} + \frac{C_{\text{ce},1} + C_{\text{TOPLE},1}}{T^{3/2}} + \frac{C_{\text{ce},2} + C_{\text{TOPLE},2}}{T^2}$$

and, furthermore,  $\mathbb{E}[\sum_{t=1}^T u_t^\top u_t] \leq T \gamma^2$ . Here  $C_{\text{ce},1}$  and  $C_{\text{ce},2}$  are defined as in Corollary 2,

$$C_{\text{TOPLE},1} := c_1 \left( \frac{\sqrt{d_x} d^2 L_{\text{hess}}}{(\lambda_{\text{noise}}^*)^{3/2}} + \frac{L_{\text{cov}}(\theta_*, \gamma^2) \text{tr}(\mathcal{H}(\theta_*))}{(\lambda_{\text{noise}}^*)^{5/2}} \right) \sqrt{\log(1/\delta) + d \log(\bar{\gamma}/\lambda_{\text{noise}}^* + 1)},$$

$$C_{\text{TOPLE},2} := c_2 \frac{d^2 L_{\text{cov}}(\theta_*, \gamma^2) L_{\text{hess}}}{(\lambda_{\text{noise}}^*)^3} \left( \log(1/\delta) + d \log(\bar{\gamma}/\lambda_{\text{noise}}^* + 1) \right),$$

$L_{\text{cov}}(\theta_*, \gamma^2) = C_{\text{sys}}(\sigma_w^2 + \gamma^2)$  is defined as in Lemma H.5,  $d := d_x + d_u$ , and  $c_1, c_2$  are universal numerical constants.

We note that this upper bound matches the lower bound on LDDM decision making given in Theorem B.2, up to constants. We prove this result in Appendix J.1. The additional burn-in required and additional lower-order terms are required to quantify how close to optimal the inputs being played are. In particular, when  $T$  satisfies (B.19), we are able to show that the inputs being played achieve near-optimal performance. The additional lower order terms,  $C_{\text{TOPLE},1}/T^{3/2}$  and  $C_{\text{TOPLE},2}/T^2$ , both quantify the loss incurred by performing certainty equivalence experiment design with an estimate of  $\theta_*$ .

## C. Remarks and Extensions

### C.1. Randomized Decisions

Our framework extends to possibly randomized decisions  $\hat{\mathbf{a}}$ ; that is, decisions  $\hat{\mathbf{a}} = \text{synth}(\boldsymbol{\tau}, \boldsymbol{\xi})$ , where again  $\boldsymbol{\tau}$  is the observed trajectory, and  $\boldsymbol{\xi}$  is internal algorithmic randomness. Note that our upper bounds all hold for the *deterministic* certainty equivalence decision rule. Our lower bounds, however, encompass these randomized decision rules. This can be seen by examining the proof of our lower bound Theorem D.1, which proceeds by lower bounding the *Bayes risk* over a distribution supported on a ball of a given radius. Hence, the performance of any randomized decision rule  $\hat{\mathbf{a}}$  is no better than the performance of the deterministic decision rule which considers the value of the random seed  $\boldsymbol{\xi}$  attaining the least Bayes risk over the distribution considered in the lower bound Theorem D.1.

### C.2. Non-Identity Noise Covariance

A known non-identity noise covariance  $\Sigma_w$  can be addressed by a change of basis to whiten the noise. Unknown noise covariances can be estimated. One can show that one need only estimate  $\Sigma_w$  up to a constance accuracy, i.e.  $|\hat{\Sigma}_w - \Sigma_w| \leq c \lambda_{\min}(\Sigma_w)$  for a small constant  $c > 0$ , and use  $\hat{\Sigma}_w$  either for a change of basis. One can show that this will still yield optimal rates up to constant factors (determined by the magnitude of  $c$ ).



### C.3. Unstable Systems and State Costs in Experiment Design

In many cases, one may wish to perform experiment design on systems that are either unstable, i.e.  $\rho(A_*) > 1$ , or are systems which are marginally stable  $\rho(A_*) = 1$ , or which have a large mixing time,  $\rho(A_*) \approx 1$ . This poses two challenges:

- To show example optimality, our analysis requires concentration of the empirical covariance matrix around its expectation. For either unstable or marginally stable systems, existing analysis suggests this may not be true (Simchowitz et al., 2018; Sarkar & Rakhlin, 2019). Moreover, estimation with unstable systems requires additional nondegeneracy conditions (Sarkar & Rakhlin, 2019).
- Because the magnitude of the state, and thus eigenvalues of the covariance matrix grow rapidly in marginally stable and in unstable systems, they may constitute a somewhat unrealistic setting for experiment design: in practice, very large states/covariances are highly undesirable, whereas for estimation, they can be quite beneficial.

To address these concerns, we propose three settings which would yield meaningful extensions of experiment design to unstable/marginally unstable settings.

**Multiple Rollouts:** One can instead consider experiment design with, say,  $n$  independent rollouts of finite horizon  $H$ . By forcing the system to reset, this is sufficient to ensure concentration of the relevant covariance matrices, and obviate consistency issues that may arise in the unstable setting.

**Stabilizing Controller:** Another approach is to assume the existence of a stabilizing controller  $K_0$ , and select inputs  $u_t = K_0 x_t + \nu_t$ , where  $\nu_t$  is an additional input chosen to optimize the experiment design. We can then impose the total power constraint on the total square norm of the  $\nu_t$  inputs.

#### C.3.1. STATE COSTS

Imposing total power constraints on the additional inputs  $\nu_t$  in the above example may appear somewhat artificial. Instead, one may wish to explicitly encode the tradeoff between ensuring state magnitudes are small, and the rate of estimation (as determined by the eigenvalues of covariance matrix) is fast. To this end, we can consider control budgets of the form of LQR-like penalties

$$\sum_{t=1}^T x_t^\top Q_{\text{budg}} x_t + u_t^\top R_{\text{budg}} u_t \leq T\gamma^2. \quad (\text{C.1})$$

We stress that the cost matrices  $Q_{\text{budg}}$  and  $R_{\text{budg}}$  above pertain to the experiment design, and not to, say, an LQR synthesis task for which the experiment design is being considered.

We further note that satisfying the constraint (C.1) may be infeasible: indeed, this occurs whenever both (a) the optimal infinite LQR cost for with cost matrices  $(Q_{\text{budg}}, R_{\text{budg}})$  is strictly greater than  $\gamma^2$ , and (b) the the horizon  $T$  is sufficiently large (so that the finite horizon optimal costs approaches its limiting, infinite-horizon value).

Finally, the design for budgets of the form Eq. (C.1) may be *closed-loop*: that is, they may necessarily require a inputs  $u_t$  which are functions of past states  $\mathbf{x}_{1:t}$ . In contrast, when budget only constraints total *input power*, we have shown that *open-loop* inputs (i.e. those not depending on past states) suffice for optimality, up to constant factors. This raises the question of how to conduct efficient experiment design over such closed loop policies. In the interest of brevity, we sketch a promising approach to this problem, and omit the details for future work:

- Observe that the cost (C.1) is itself a linear form in the joint covariance matrices of the states and inputs. Hence, the experiment design roughly amounts to optimizing a convex function of the form  $\text{tr}(\mathcal{H} \cdot \Gamma^{-1})$  over feasible state-input covariance matrices  $\Gamma$ , subject to a linear constrain of the form  $\text{tr}(\mathcal{C} \cdot \Gamma)$ . This is a convex program in  $\Gamma$ .
- To characterize the set of feasible covariance matrices, we can observe that any feasible covariance matrix can be obtained by combining a linear feedback policy with an open loop policy (this can be verified using Gaussianity). One promising computation approach to perform this optimization is to use system level synthesis (Anderson et al., 2019), where the linear feedback term can be represented as a linear form in the noise variables  $w_t$ . Thus, the desired covariance matrices can be represented as outer-products of open-loop inputs and linear forms.

- The SLS representation then describes the set of feasible covariance matrices as outer-products of linear forms; this is not yet a convex representation. However, just as this paper operators on the convex hull of covariance matrices arising from open-loop inputs, a similar convex relaxation can yield a convex representation of covariance matrices with closed-loop feedback. This relaxation is not loose: the space of feasible expected covariance matrices over all policies is convex, since one can always interpolate between two covariance matrices via probabilistic interpolations between the policies which generated them (i.e. selected some policy  $\pi_1$  with probability  $p$ , and another with probability  $1 - p$ ). This means that, given a feasible expected covariance matrix recovered from this relaxation, we can produce a policy to generate it.

#### C.4. Expectation v.s. High Probability

Observe that our upper bounds are stated with high probability, whereas lower bounds are stated in expectation. This is because on lower bounds proceed (like most information theoretic lower bounds) via bounds on the Bayes-Risk, which regard expected performance; on the other hand, our upper bounds may not hold in expectation because, on a highly improbable failure event, the estimate may produce a decision which has *infinite* cost (e.g. a controller returned for an LQR task which fails to stabilize the system).

To close the gap between the two, we can make the following modifications:

- Our lower bounds on expected risk can be restated as lower bounds on a constant probability of error. To see this, we note that our lower bound holds over a localized set of instances,  $\mathcal{B} = \{\theta : \|\theta - \theta_0\|_2 \leq r\}$ . Hence, any decision rule / experiment design procedure can be modified to only return decisions which satisfy some minimum worst-case performance on  $\mathcal{B}$  (and, under the smoothness assumptions considered in this work, this can be done without harming the performance of the decision rule). Thus, the worst case suboptimality of the decision rule can be no more than a constant, and thus, the lower bound in expectation can be transformed into a lower bound holding with constant probability.
- Similarly, if the learner is given side information (e.g. a convex set  $\mathcal{A}$  of possible decisions known to contain an open ball around the optimal decision  $\alpha_{\text{opt}}(\theta_*)$ , and such that the cost  $\sup_{\alpha \in \mathcal{A}} \mathcal{J}_{\theta_*}(\alpha) < \infty$ ), then the learner can achieve upper bounds in expectation by projecting their decision  $\hat{\alpha}$  onto the set  $\mathcal{A}$ , namely

$$\hat{\alpha}' = \text{Proj}_{\mathcal{A}}(\hat{\alpha})$$

Then, whenever  $\hat{\alpha}'$  is sufficiently close to  $\alpha_{\text{opt}}(\theta_*)$ ,  $\hat{\alpha}' = \hat{\alpha}$  and the cost will be unaffected; however, on low-probability failure events, the projection step ensures the cost remains bounded.

#### C.5. Subspaces and Parametric Uncertainty

In many applications, one considers linear dynamical systems  $(A, B)$  where some coordinates, or more generally, subspaces of the dynamical matrices are known to the learner, and only some coordinates or subspaces must be learned. In this case, learning the matrices  $(A, B)$  with unconstrained least squares may be suboptimal.

However, the subspace-constrained learning setting can be easily re-written as an unconstrained learning problem restricted to an appropriate subspace, and this resulting structure obeys the general martingale least squares setting outlined in Part II. Hence, the same arguments given in that section demonstrate can be used to demonstrate optimality of certainty equivalence. The algorithm TOPLE can be similarly modified to optimize for the covariance matrix in the relevant restricted subspace.

A more general constrained setting is where  $(A, B) = (A(\theta), B(\theta))$  are smooth, possibly nonlinear functions of a hidden parameter  $\theta$ . In this case, we conjecture that one can achieve optimal rates by obtaining a course estimate  $\hat{\theta}$  of  $\theta$ , applying constrained least squares in the subspace defined by the image of the Jacobian  $\frac{d}{d\theta} A(\theta), B(\theta)$  at  $\hat{\theta} = \theta$ . We leave the details for future work.

#### C.6. Parametric Nonlinear Systems

Many of the results in this work can be extended to the parametric non-linear systems considered in the recent literature (Mania et al., 2020; Kakade et al., 2020):

$$x_t = \langle \theta_*, \phi(x_t, u_t) \rangle + w_t, \quad w_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2) \tag{C.2}$$

where  $\theta_\star \in \mathbb{R}^d$  is a linear parameter describing the dynamics, and  $\phi : \mathbb{R}^{d_x+d_u} \rightarrow \mathbb{R}^d$  is an embedding function *known* to the learner. Despite the nonlinearity, the dynamics (C.2) satisfy the martingale regression setting considered in Appendix D, and thus the MDM upper and lower bounds in that section extend (Theorem D.1 and Theorem D.2); similarly, the guarantees of Appendix E extend as well as long as the cost functional  $\mathcal{J}_\theta(a)$  satisfies the requisite regularity conditions.

Unfortunately, attempts to extend these guarantees to optimal experiment design encounter a number of difficulties:

1. The experiment-design lower bounds established for linear dynamical systems require verifying that we can consider, without loss of generality, exploration policies  $\pi_{\text{exp}}$  which produce sufficiently “regular” periodic inputs (see Appendix H); it is not clear how this argument would generalize to the nonlinear setting Eq. (C.2), where potentially highly pathological exploration policies may be preferable.
2. The certainty-equivalence upper bounds for linear systems require demonstrating concentration of the empirical covariance matrix around its mean; for linear systems, this can be verified as long as the dynamical matrix  $A_\star$  is stable. For nonlinear systems, further conditions need to be imposed.
3. The experiment design problem for nonlinear systems may be computationally intractable. In addition, the experiment design objective may be very sensitive to errors in the estimate of the parameter  $\theta_\star$ , so that solving the certainty equivalent experiment design objective (i.e. optimal design based on an estimate  $\hat{\theta}$ ) may be a poor proxy for the optimal design.
4. For nonlinear systems, controlling how the error in parameter estimation error translates into suboptimality in the decision  $\hat{a}$  for the given task may be quite challenging. Even for LQR synthesis in linear systems, verifying the smoothness conditions in Appendix E relies on subtle technical tools developed specifically for LQR (Simchowitz & Foster, 2020).

### C.7. Suboptimality of Low Regret Algorithms

Here, we state a formal lower bound about the suboptimality of low regret algorithms. Consider a nominal instance  $(A_\star, B_\star)$ . For simplicity, we consider a normalization where  $R_x \succeq I$  and  $R_u \succeq I$ , which can be enforced by suitable renormalization.

**Proposition C.1** (Formal statement of Proposition 4.3). *Fix a nominal instance  $\theta_\star = (A_\star, B_\star)$  with optimal value function  $P_\star$ , costs  $R_x, R_u \succeq I$ , and select a regret lower bound*

$$R \geq d_u^2 d_x \text{poly}(\|P_\star\|_{\text{op}}, \|B_\star\|_{\text{op}}) + \|P_\star\|_{\text{op}}^2 \sqrt{d_x T}/4.$$

Then, over the ball of instances  $\mathcal{B} = \mathcal{B}(R) := \{\theta : \|\theta_\star - \theta\|_{\text{F}}^2 \leq \frac{d_u^2 d_x}{16\|P_\star\|_{\text{op}} R}\}$ , the following lower bound for any low-regret exploration policy  $\pi_{\text{lr}}$  and controller  $K_{\text{lr}}$  synthesized from the trajectory it collects:

$$\max_{\theta \in \mathcal{B}} \mathbb{E}_{\theta, \pi_{\text{lr}}} [\text{Reg}_T] \geq R, \quad \text{or} \quad \max_{\theta \in \mathcal{B}} \mathbb{E}_{\theta, \pi_{\text{lr}}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{\text{lr}})] \geq \frac{d_u^2}{320R} \cdot \left( \max_{1 \leq m \leq d_x} m \cdot \sigma_m(A_\star + B_\star K_\star)^2 \right). \quad (\text{C.3})$$

where  $\sigma_m(\cdot)$  denotes the  $m$ -th largest singular value.

For many instances of interest,  $(\max_{1 \leq m \leq d_x} m \sigma_m(A_\star + B_\star K_\star)^2) > 0$  is a constant bounded away from 0, and even scales with dimension  $d_x$ . Hence, we find a strong tradeoff between low regret and optimal estimation. The key intuition behind the proof is that low regret algorithms converge to inputs  $\mathbf{u}_t \approx K_\star \mathbf{x}_t$  approaching the optimal control policy; in doing so, they under-explore directions *perpendicular* to the hyperplane  $\{(x, u) : u = K_\star x\}$ , which are necessary for identifying the optimal control policy. This idea, as well as the rigorous proof, draws heavily on the regret lower bound due to (Simchowitz & Foster, 2020).

*Proof of Proposition C.1.* Throughout, fix a low regret policy  $\pi_{\text{lr}}$ . The proof follows from the arguments of (Simchowitz & Foster, 2020). Fix a nominal instance  $\theta_\star = (A_\star, B_\star)$ , with optimal controller  $K_\star$ . Let  $m \in [d_x]$ , and adopt the shorthand  $n = d_u$ . For binary vectors  $e \in \{-1, +1\}^{nm}$  consider a packing

$$\theta_e := (A_e, B_e) = (A_\star - \Delta_e K_\star, B_\star + \Delta_e), \quad \Delta_e = \epsilon \sum_{i=1}^n \sum_{j=1}^m e_{i,j} v_i w_j^\top,$$

where  $\epsilon$  is a parameter to be chosen small than  $\epsilon_0 := \frac{1}{nm \cdot \text{poly}(\|P_\star\|_{\text{op}})}$  for a larger enough polynomial poly, and  $(v_i)$  and  $(w_j)$  are appropriately selected orthonormal basis vectors. These instance are constructed so that  $A_e + B_e K_\star$  are identical for all packing indices  $e$ ; in other words, by selecting the optimal controller for the nominal instance  $\theta_\star$ , all the instances are indistinguishable.

Let  $K_e := K_{\text{opt}}(\theta_e)$  denote the optimal controller for these instances. We let  $\mathbb{E}_e$  denote expectation under the uniform distribution over  $e \sim \{-1, 1\}^{nm}$  from the hypercube. We consider the term  $K$ -err from (Simchowitz & Foster, 2020), modified to include all  $T$  time steps (instead of  $T/2$ ). For any controller  $K \in \mathbb{R}^{d_u d_x}$ , define

$$K\text{-err} := \mathbb{E}_{\theta_e, \pi_{\text{lr}}} \left[ \sum_{t=1}^T \|u_t - Kx_t\|_2^2 \right].$$

The reason for considering  $T$  steps is because here we are concerned with the *offline* learning problem, where the learner is allowed to use all data from the trajectory to synthesize a controller.

The first claim lower bounds the regret by average deviation from the optimal control policy under the nominal instance:

**Claim C.1.** Let  $\gamma_{\text{err}} = d_x \cdot \text{poly}(\|P_\star\|_{\text{op}}, \|B_\star\|_{\text{op}})$ , where  $\mathcal{O}$  hides universal constants. Then,

$$\mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\text{Regret}_T] \geq \frac{1}{4} \mathbb{E}_e K\text{-err}_e[K_\star] - nmT \|P_\star\|_{\text{op}}^4 \epsilon^2 - \gamma_{\text{err}} \quad (\text{C.4})$$

*Proof.* A modification of (?)Lemma 4.3]simchowitz2020naive use all  $T$  steps (rather than  $T/2$ , and using a sum over the terms  $\eta_t$  in that proof rather than a bound by the maximum) shows that

$$\text{Reg}_e := \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\text{Regret}_T] \geq \frac{1}{2} K\text{-err}_e[K_e] - \gamma_{\text{err}}, \quad \text{where } \gamma_{\text{err}} = d_x \cdot \text{poly}(\|P_\star\|_{\text{op}}, \|B_\star\|_{\text{op}}), \quad (\text{C.5})$$

From Lemma 4.7 in (Simchowitz et al., 2017), we also have

$$\mathbb{E}_e K\text{-err}_e[K_e] \leq 2 \mathbb{E}_e K\text{-err}_e[K_\star] + 4nmT \|P_\star\|_{\text{op}}^4 \epsilon^2.$$

Combining the two displays gives the claim.  $\square$

Next, since the instances  $\theta_\star$  only differ along directions  $(x, u) \in \mathbb{R}^{d_x + d_u}$  perpendicular to the hyperplane  $u = K_\star x$ , samples collected perpendicular to this hyperplane essential for disambiguating between the instances  $\theta_e$ . This leads to the following lower bound.

**Claim C.2.** Set  $R = \frac{n}{48\epsilon^2}$ . Then either

$$\mathbb{E}_e [K\text{-err}_e[K_\star]] \geq 12R, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{\text{lr}})] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 n^2 m}{480R} \quad (\text{C.6})$$

*Proof.* Modifying Lemma 4.5 in (Simchowitz & Foster, 2020) shows that for any binary estimator  $\hat{e}$ , either  $K\text{-err}_e$  is small on  $e$  drawn from the hypercube, or else  $\hat{e}$  has large hamming error.

$$\text{either } \mathbb{E}_e [K\text{-err}_e[K_\star]] \geq \frac{n}{4\epsilon^2}, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [d_{\text{ham}}(\hat{e}, e)] \geq \frac{nm}{4}.$$

Combining with Lemma 4.6 in (Simchowitz & Foster, 2020), it follows that

$$\text{either } \mathbb{E}_e [K\text{-err}_e[K_\star]] \geq \frac{n}{4\epsilon^2}, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\|K_{\text{lr}} - K_e\|_{\text{F}}^2] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 nm \epsilon^2}{10}.$$

From (?)Lemma 3]mania2019certainty, we can bound  $\|K_{\text{lr}} - K_e\|_{\text{F}}^2 \geq \frac{\mathcal{R}_{\text{LQR}, \theta_e}(K_{\text{lr}})}{\sigma_{\min}(R_u)}$ . Thus, using  $R_u \succeq I$ ,

$$\mathbb{E}_e [K\text{-err}_e[K_\star]] \geq \frac{n}{4\epsilon^2}, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{\text{lr}})] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 nm \epsilon^2}{10}.$$

Reparameterizing  $R = \frac{n}{32\epsilon^2}$  gives either  $\mathbb{E}_{e \sim \{-1, 1\}} [K\text{-err}_e[K_\star]] \geq 12R$ , or else  $\mathbb{E}_{e \sim \{-1, 1\}} \mathbb{E}_{\theta_e, \pi_{\text{lr}}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{\text{lr}})] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 n^2 m}{480R}$ , as needed.  $\square$

Combining Claims C.1 and C.2 and taking  $R = \frac{n}{32\epsilon^2}$  gives

$$\mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{1r}} [\text{Reg}_T] \geq 3R - \underbrace{nmT \|P_\star\|_{\text{op}}^4 \epsilon^2}_{= \frac{mT \|P_\star\|_{\text{op}}^4}{32R}} - \gamma_{\text{err}}, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{1r}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{1r})] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 n^2 m}{320R} \quad (\text{C.7})$$

In particular, if we take  $\epsilon$  so that  $R \geq \max\{\gamma_{\text{err}}, \frac{mT \|P_\star\|_{\text{op}}^4}{32R}\}$ , then either

$$\mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{1r}} [\text{Reg}_T] \geq R, \quad \text{or} \quad \mathbb{E}_e \mathbb{E}_{\theta_e, \pi_{1r}} [\mathcal{R}_{\text{LQR}, \theta_e}(K_{1r})] \geq \frac{\sigma_m(A_\star + B_\star K_\star)^2 n^2 m}{320R} \quad (\text{C.8})$$

Let us conclude by verifying the requisite ranges for conditions on  $\epsilon$  and regret bound  $R$  for the above to hold. We require that  $R \geq \gamma_{\text{err}} = d \text{poly}(\|P_\star\|_{\text{op}}, \|B_\star\|_{\text{op}})$ . We also require  $R^2 \geq \frac{mT \|P_\star\|_{\text{op}}^4}{32}$ , so that  $R \geq \|P_\star\|_{\text{op}}^2 \sqrt{mT}/4$ . Finally, we require  $\epsilon^2 = \frac{n}{32R} \leq \frac{1}{nm \cdot \text{poly}(\|P_\star\|_{\text{op}})}$ , so  $R \geq n^2 m \cdot \text{poly}(\|P_\star\|_{\text{op}})$  for a possibly modified polynomial function. Concluding, and using  $n = d_u$  and  $m \leq d_x$ , it is enough to select

$$R \geq \text{poly}(\|P_\star\|_{\text{op}}, \|B_\star\|_{\text{op}}) d_u^2 d_x + \|P_\star\|_{\text{op}}^2 \sqrt{d_x T}/4.$$

Finally, we note that all the instances  $\theta_e$  have

$$\|\theta_e - \theta_\star\|_{\text{F}}^2 = \|K_\star \Delta_e\|_{\text{F}}^2 + \|\Delta_e\|_{\text{F}}^2 \leq (1 + \|K_\star\|_{\text{op}}^2) nm \epsilon^2 = \frac{n^2 m (1 + \|K_\star\|_{\text{op}}^2)}{32R}.$$

Concluding, we note that for  $R_{\mathbf{u}} \succeq I$ , one can bound  $\|K_\star\|_{\text{op}}^2 \leq \|P_\star\|_{\text{op}}$  (this follows since  $P_\star \succeq R_{\mathbf{u}} + K_\star^\top R_{\mathbf{u}} K_\star$  by a standard computation). Hence, taking  $n = d_u$  and  $m \leq d_x$  all instances lie in the ball  $\mathcal{B} = \{\theta : \|\theta_\star - \theta\|_{\text{F}}^2 \leq \frac{d_u^2 d_x}{16 \|P_\star\|_{\text{op}} R}\}$ . The bound follows. □

## Part II

# Martingale Decision Making

### D. Optimal Rates for Martingale Regression in General Norms

In this section, we establish upper and lower bounds in sequential linear regression setting with martingale structure. We consider a filtration  $(\mathcal{F}_t)_{t \geq 1}$ , and a well-specified linear model with true parameter  $\theta_* \in \mathbb{R}^{d_\theta}$ , covariates  $z_t \in \mathbb{R}^{d_\theta}$ , scalar observations  $y_t$ , and noise  $w_t$ . Specifically, for each  $\theta_* \in \mathbb{R}^{d_\theta}$ , we assume there is a distribution  $\mathcal{D}_{\theta_*}$  over  $(y_t, z_t, w_t)$  obeying

$$y_t = \langle \theta_*, z_t \rangle + w_t, \quad w_t \mid \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_w^2), \quad z_t \text{ is } \mathcal{F}_{t-1}\text{-adapted.}$$

As long as the above conditions hold, we allow the distribution of the covariates to be arbitrary: for example, that there is some function  $f(\dots)$  of appropriate shape such that  $z_t = f(t, z_{1:t-1}, y_{1:t-1}, w_{1:t}, \theta_*)$ . We let  $\mathbb{E}_\theta$  and  $\mathbb{P}_\theta$  denote probabilities and expectations with respect to the above law when  $\theta_* = \theta$ . Our aim is to produce an estimator  $\hat{\theta} \in \mathbb{R}^{d_\theta}$  so as to minimize the following weighted least-squares risk:

$$\mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) := \|\hat{\theta} - \theta\|_M^2, \quad M \succeq 0. \quad (\text{D.1})$$

#### D.1. Upper and Lower Bounds on $M$ -norm Regression

In this section, we show that the least squares estimator

$$\hat{\theta}_{\text{ls}} := \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1} \sum_{t=1}^T z_t y_t \quad (\text{D.2})$$

is the optimal estimator of  $\theta_*$  for the risk in Eq. (D.1), in a very strong, instance dependent sense. Throughout, the central object of our analysis is the random covariance matrix:

$$\Sigma_T := \sum_{t=1}^T z_t z_t^\top.$$

Let us start with the lower bound. We will call an estimator  $\hat{\theta}$  *measurable* if  $\hat{\theta}$  is a measurable function of the covariates and responses  $(y_t, z_t : 1 \leq t \leq T)$ , and possibly some internal randomness. We consider the localized risk in a Euclidean ball of radius  $r > 0$  around a nominal instance  $\theta_0$

**Theorem D.1** (Truncated van Trees). *Let  $\hat{\theta}$  be an arbitrary measurable estimator. Moreover, fix a covariance parameter  $\Gamma \in \mathbb{S}_{++}^{d_\theta}$ , nominal instance  $\theta_0 \in \mathbb{R}^{d_\theta}$ , and radius  $r \geq \sqrt{5 \text{tr}(\Gamma^{-1})}$ . Let  $\mathcal{B} := \{\theta : \|\theta - \theta_0\|_2 \leq r\}$  denote a Euclidean ball around  $\theta_0$ . Then, it holds that*

$$\inf_{\hat{\theta}} \max_{\theta \in \mathcal{B}} \mathbb{E}_\theta \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) \geq \sigma_w^2 \min_{\theta \in \mathcal{B}} \text{tr} \left( M \cdot (\mathbb{E}_\theta[\Sigma_T] + \Gamma)^{-1} \right) - \Psi(r; \Gamma, M),$$

where  $\Psi(r; \Gamma, M) := \frac{32 \|M\|_{\text{op}}}{\lambda_{\min}(\Gamma)} \exp(-\frac{r^2}{5} \lambda_{\min}(\Gamma))$ .

*Proof Sketch of Theorem D.1.* The proof is given shortly below in Appendix D.2; it is derived from a Bayes-risk lower bound deriving from an explicit computation of the conditional variance (and thus minimal mean square error in estimation) of a parameter  $\theta$  drawn from a normal distribution centered at  $\theta_0$ , with covariance  $\Lambda = \Gamma^{-1}$ . This distribution is then carefully truncated at radius  $r$  to ensure the local minimax bound holds when restricted to the ball  $\mathcal{B}$ .  $\square$

**Remark D.1** (Comparison to Previous Lower Bounds). Lower bounds for experiment design are typically derived from the Cramer-Rao bound (e.g. in (Chaudhuri et al., 2015)), which applies only to unbiased estimators, and does not rule out more efficient estimation by allowing bias. In contrast, our work provides an *unconditional* information theoretic lower bound, derived from a closed-form computation of an expected Bayes risk in linear regression with a Gaussian prior (Theorem D.1). This technique is similar in spirit to the Van Trees inequality (Gill et al., 1995) which was used in concurrent work to understand instance-optimal regret in LQR when  $A_*$  is known but  $B_*$  is not (Ziemann & Sandberg, 2020).

Another common technique for adaptive estimation lower bounds is Assouad’s method (Arias-Castro et al., 2012; Simchowitz & Foster, 2020), typically yielding worst-case (though not sharp, instance-dependent) lower bounds. The lower bounds for adaptive experiment design in linear systems due to (Wagenmaker & Jamieson, 2020; Jedra & Proutiere, 2019) hold in the asymptotic regime where the tolerated probability of failure  $\delta$  tends to 0, a regime pioneered by (Kaufmann et al., 2016) for pure-exploration multi-arm bandits, and extended to reinforcement learning in (Ok et al., 2018). For continuous parameter estimation (such as the linear control setting control), the  $\delta \rightarrow 0$  asymptotic lower bounds differs from non-asymptotic upper bounds by as much as a dimension factor, unless  $\delta$  is taken to be exponentially small in dimension (Simchowitz et al., 2017). In particular, taking  $\delta \rightarrow 0$  yields a qualitatively inaccurate picture of the expected error of the estimators in question. In contrast, this work achieves matching bounds in the (arguably more natural) “moderate  $\delta$ ” regime, where the tolerated failure probability is no smaller than inverse polynomial in the time horizon.

In our applications, we shall choose  $r$  sufficiently large and  $\Gamma$  sufficiently small so that the lower bound reads

$$\sigma_w^{-2} \inf_{\hat{\theta}} \max_{\theta \in \mathcal{B}} \mathbb{E}_{\theta} \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) \gtrsim \min_{\theta \in \mathcal{B}} \text{tr} (M \mathbb{E}_{\theta} [\Sigma_T]^{-1}); \quad (\text{D.3})$$

in other words, that the  $M$ -weighted trace of the inverse covariance matrix lower bounds the risk. Even though the right-hand side considers the minimum over  $\theta \in \mathcal{B}$ , the radius  $r$  of  $\mathcal{B}$  can be chosen small enough that this quantity does not vary significantly. For a sense of scaling  $\mathbb{E}_{\theta} [\Sigma_T]$  will typically scale like  $\Omega(T)$ , by choosing  $\Gamma \preceq o(T)$ , and  $r^2 \propto \text{tr}(\Gamma^{-1}) \log^2(T)$ , the term  $\Psi(r, \Gamma, M)$  vanishes as  $T^{-\omega(1)}$ , and the approximation Eq. (D.3) holds. Moreover, since this scaling of  $r$  vanishes at a rate of  $\log^2 T / \sqrt{T}$ ,  $r$  is small enough so as to ensure  $\mathbb{E}_{\theta} [\Sigma_T]$  does not vary significantly on  $\mathcal{B}$ .

**Theorem D.2.** Fix any matrices  $\Gamma \in \mathbb{S}_{++}^{d_{\theta}}$ ,  $M \in \mathbb{S}_{+}^{d_{\theta}}$ , with  $M \neq 0$ . Given a parameter  $\beta \in (0, 1/4)$ , define the event

$$\mathcal{E} := \{\|\Sigma_T - \Gamma\|_{\text{op}} \leq \beta \lambda_{\min}(\Gamma)\}$$

Then, if  $\mathcal{E}$  holds, the following holds with probability  $1 - \delta$ :

$$\begin{aligned} \|\hat{\theta}_{\text{ls}} - \theta_{\star}\|_M^2 &\leq 5(1 + \alpha) \cdot \sigma_w^2 \log \frac{6d}{\delta} \cdot \text{tr}(M\Gamma^{-1}), \text{ w.p. } 1 - \delta, \quad \text{where} \\ \alpha &:= 26\beta^2 \lambda_{\max}(\Gamma) \text{tr}(\Gamma^{-1}). \end{aligned}$$

*Proof Sketch.* Like many results of this flavor, the proof is based on the self-normalized martingale inequality (Abbasi-Yadkori et al., 2011). Unlike related results, however, our proof must relate in the error in the  $M$ -norm  $\|\hat{\theta}_{\text{ls}} - \theta_{\star}\|_M^2$  to the  $\Gamma$ -geometry so as to recover  $\text{tr}(M\Gamma^{-1})$ . It turns out that, due to the fact that matrix square does not preserve the Lowner order (i.e., it is possible to have  $0 \prec A \prec B$ , but  $A^2 \not\prec B^2$ ), we require the empirical matrix  $\Sigma_T$  to concentrated around  $\Gamma$  for this argument to go through. This forces us to require the above event  $\mathcal{E}$  to hold, and to suffer the error term  $\alpha$ . The complete proof is given in Appendix D.3.  $\square$

## D.2. Proof of $M$ -norm Regression Lower Bound (Theorem D.1)

Without loss of generality, set  $\sigma_w^2 = 1$ . The proof of the lower bound is a Gaussian-specialization of the Van Trees inequality (see, e.g. (Gill et al., 1995)), a Bayes-risk lower bound which considers the risk of estimating a quantity under a certain prior. For our prior, we use a normal distribution, which we truncate to a radius  $r$ . In what follows, we set

$$\Lambda := \Gamma^{-1} \in \mathbb{S}_{++}^{d_{\theta}}.$$

We let  $\mathcal{N}_{\text{tr}}$  denote the following *truncated normal* distribution: the distribution of  $Z \sim \mathcal{N}(\theta_0, \Lambda)$ , conditioned on the event  $\|Z - \theta_0\|^2 \leq r$ . We further define the full data  $\mathcal{D}_T := (y_{1:T}, z_{1:T})$ , and let

- $\mathcal{D}_{\text{full}}(\mathcal{D}_T)$  denote the posterior of  $\theta$  given  $\mathcal{D}_t$ , when  $\theta$  is drawn from  $\mathcal{N}(\theta_0, \Lambda)$ ;
- Let  $\mathcal{D}_{\text{trunc}}(\mathcal{D}_T)$  denote the distribution of  $\theta \mid \mathcal{D}_T$ .

Throughout, we assume that our posited estimator  $\hat{\theta}$  is a deterministic function of  $\mathcal{D}_T$ ; this is without loss of generality for a Bayes-risk lower bound. Then, since the distribution  $\mathcal{N}_{\text{tr}}$  is supported on the ball  $\mathcal{B} := \{\theta : \|\theta - \theta_0\|_2 \leq r\}$ ,

$$\inf_{\hat{\theta}} \max_{\theta \in \mathcal{B}} \mathbb{E}_{\theta} \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) \geq \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\theta} \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) = \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathcal{D}_T \sim \theta} \|\hat{\theta}(\mathcal{D}_T) - \theta\|_2^2.$$

**Lemma D.1** (Replica Lemma). *Let  $X, Y$  be abstract random variables, with  $X \sim \mathbb{P}_X$ ,  $Y \sim \mathbb{P}_{Y|X}(\cdot | X)$ , and let  $f(x, y)$  be an integrable function. Moreover, suppose that  $X | Y$  has density  $\mathbb{P}_{X|Y}(\cdot | Y)$ . Then*

$$\mathbb{E}_{X \sim \mathbb{P}_X} \mathbb{E}_{Y \sim \mathbb{P}_{Y|X}(\cdot | X)} [f(X, Y)] = \mathbb{E}_{X \sim \mathbb{P}_X} \mathbb{E}_{Y \sim \mathbb{P}_{Y|X}(\cdot | X)} \mathbb{E}_{X' \sim \mathbb{P}_{X|Y}(\cdot | Y)} [f(X', Y)].$$

*Proof.* By Fubini's theorem and Bayes' rule,

$$\begin{aligned} \mathbb{E}_{X \sim \mathbb{P}_X} \mathbb{E}_{Y \sim \mathbb{P}_{Y|X}(\cdot | X)} [f(X, Y)] &= \int \int f(x, y) \mathbb{P}_{Y|X}(y | x) \mathbb{P}_X(x) dy dx \\ &= \int \int \int f(x, y) \mathbb{P}_{X'|Y}(x' | y) \mathbb{P}_{Y|X}(y | x) \mathbb{P}_X(x) dx' dy dx \\ &= \int \int \int f(x, y) \frac{\mathbb{P}_{Y|X'}(y | x') \mathbb{P}_{X'}(x')}{\mathbb{P}_Y(y)} \mathbb{P}_{Y|X}(y | x) \mathbb{P}_X(x) dx' dy dx \\ &= \int \int \int f(x, y) \mathbb{P}_{X|Y}(x | y) \mathbb{P}_{Y|X}(y | x) \mathbb{P}_{X'}(x') dx' dy dx \\ &= \int \int \int f(x, y) \mathbb{P}_{X|Y}(x | y) \mathbb{P}_{Y|X}(y | x) \mathbb{P}_{X'}(x') dx dy dx' \\ &= \mathbb{E}_{X' \sim \mathbb{P}_{X'}} \mathbb{E}_{Y \sim \mathbb{P}_{Y|X}(\cdot | X')} \mathbb{E}_{X \sim \mathbb{P}_{X|Y}(\cdot | Y)} [f(X, Y)] \end{aligned}$$

Relabeling gives the result.  $\square$

By Lemma D.1, the above is equal to

$$\mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\theta} \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) = \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{trunc}}(\mathfrak{D}_T)} \|\hat{\theta}(\mathfrak{D}_T) - \theta'\|_M^2. \quad (\text{D.4})$$

For a random vector  $Z$  and any fixed  $a$ ,  $\mathbb{E}\|Z - a\|_M^2 \geq \mathbb{E}\|Z - \mathbb{E}Z\|_M^2$ ; that is, the Bayes estimator is optimal. Denoting the event  $\mathcal{E} := \{\|\theta' - \theta_0\|_2 \leq r\}$  (over the randomness of  $\theta'$ ), we lower bound Eq. (D.4) by

$$\begin{aligned} \text{Eq. (D.4)} &\geq \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{trunc}}(\mathfrak{D}_T)} \|\theta' - \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{trunc}}(\mathfrak{D}_T)} \theta'\|_M^2 \\ &= \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} [\theta' | \mathcal{E}]\|_M^2 \mid \mathcal{E} \right] \end{aligned} \quad (\text{D.5})$$

To handle this expression, we use the following technical lemma

**Lemma D.2.** *Consider a square-integrable random vector  $Z \in \mathbb{R}^{d_\theta}$ , fixed  $\mu \in \mathbb{R}^{d_\theta}$ ,  $r \geq 0$ . Defin the event  $\mathcal{E} := \{\|Z - \mu\|_2 \leq r\}$ . Then,*

$$\mathbb{E}[\|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2 \mid \mathcal{E}] \geq \mathbb{E}\|Z - \mathbb{E}[Z]\|_M^2 - 4\|M\|_{\text{op}} \mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \|Z - \mu\|_2^2].$$

*Proof.*

$$\begin{aligned} \mathbb{E}[\|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2 \mid \mathcal{E}] &\geq \mathbb{E}[\mathbb{I}\{\mathcal{E}\} \cdot \|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2] \\ &= \underbrace{\mathbb{E}\|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2}_{(i)} - \underbrace{\mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \cdot \|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2]}_{(ii)} \end{aligned}$$

Next, we lower bound (i) =  $\mathbb{E}\|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2 \geq \mathbb{E}\|Z - \mathbb{E}[Z]\|_M^2$ . Thus, it remains to upper bound (ii):

$$\begin{aligned} (ii) = \mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \cdot \|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2] &\leq 2\mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \cdot (\|Z - \mu\|_M^2 + \|\mu - \mathbb{E}[Z | \mathcal{E}]\|_M^2)] \\ &\leq 2\mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \cdot (\|Z - \mu\|_M^2 + r^2\|M\|_{\text{op}})] \\ &\leq 2\mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \cdot (\|Z - \mu\|_2^2\|M\|_{\text{op}} + r^2\|M\|_{\text{op}})], \end{aligned}$$

where in the second line, we use that, under  $\mathcal{E}$ ,  $\|Z - \mu\|_2 \leq r$ . Moreover, under  $\mathcal{E}^c$ ,  $\|Z - \mu\|_2^2 \geq r$ , so that  $\|Z - \mu\|_2^2 + r^2 \leq 2\|Z - \mu\|_2^2$ . Hence, (ii)  $\leq 4\|M\|_{\text{op}} \mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \|Z - \mu\|_2^2]$ . Thus,

$$\mathbb{E}[\|Z - \mathbb{E}[Z | \mathcal{E}]\|_M^2 \mid \mathcal{E}] \geq \mathbb{E}\|Z - \mathbb{E}[Z]\|_M^2 - 4\|M\|_{\text{op}} \mathbb{E}[\mathbb{I}\{\mathcal{E}^c\} \|Z - \mu\|_2^2],$$

as needed.  $\square$



Instantiating Lemma D.2,

$$\begin{aligned} \text{Eq. (D.4)} &\geq \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta'']\|_M^2 \right] \\ &\quad - 4\|M\|_{\text{op}}^2 \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \theta_0\|_2^2 \cdot \mathbb{I}\{\|\theta' - \theta_0\|_2^2 > r^2\} \right]. \end{aligned}$$

Hence, retracing our steps thus far,

$$\begin{aligned} \inf_{\hat{\theta}} \max_{\theta \in \mathcal{B}} \mathbb{E}_{\theta} \mathcal{R}_{\text{ls}}(\hat{\theta}; \theta) &\geq \underbrace{\mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta'']\|_M^2 \right]}_{(a)} \\ &\quad - 4\|M\|_{\text{op}}^2 \cdot \underbrace{\left( \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \theta_0\|_2^2 \cdot \mathbb{I}\{\|\theta' - \theta_0\|_2^2 > r^2\} \right] \right)}_{(b)} \end{aligned} \quad (\text{D.6})$$

Let us control the two resulting terms.

**Computing term (a):** First, we bound the dominant term (a):

**Lemma D.3.** *The following identity holds:*

$$\mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta'']\|_M^2 \right] = \sigma_w^2 \text{tr}(M^{1/2}(\Sigma_T + \Lambda^{-1})^{-1}M^{1/2}).$$

*Proof.* Due to the fact that we have gaussian likelihoods, we we

$$d\mathbb{P}(\mathfrak{D}_T | \theta) \propto \exp\left(-\frac{1}{2\sigma_w^2} \sum_{t=1}^T (y_t - \langle \theta, z_t \rangle)^2\right) \propto \exp\left(-\frac{1}{2\sigma_w^2} \theta^\top \Sigma_T \theta + \frac{1}{\sigma_w^2} \theta^\top \sum_{t=1}^T z_t y_t^\top\right)$$

On the other hand, for any given  $\theta$ ,  $d\mathbb{P}(\theta) \propto \exp(-\frac{1}{2}\theta^\top \Lambda^{-1}\theta)$ . Hence,

$$d\mathbb{P}(\theta | \mathfrak{D}_T) \propto \exp\left(-\frac{1}{2\sigma_w^2} \theta^\top (\Sigma_T + \Lambda^{-1})\theta + \frac{1}{\sigma_w^2} \theta^\top \sum_{t=1}^T z_t y_t^\top\right)$$

Thus,  $\theta | \mathfrak{D}_T$  is conditionally Gaussian with covariance  $\sigma_w^2(\Sigma_T + \Lambda^{-1})^{-1}$ . It follows that:

$$\begin{aligned} &\mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ \|\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta'']\|_M^2 \right] \\ &= \text{tr} \left( M^{1/2} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \left[ (\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta'']) (\theta' - \mathbb{E}_{\theta'' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)}[\theta''])^\top \right] M^{1/2} \right) \\ &= \sigma_w^2 \text{tr} \left( M^{1/2} (\Sigma_T + \Lambda^{-1})^{-1} M^{1/2} \right) \end{aligned}$$

□

As a direct consequence of the above lemma, we find that

$$\begin{aligned} \text{term (a)} &= \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathfrak{D}_T \sim \theta} \mathbb{E}_{\theta' \sim \mathcal{D}_{\text{full}}(\mathfrak{D}_T)} \mathbb{E}_{\theta''} \left[ \sigma_w^2 \text{tr}(M^{1/2}(\Sigma_T + \Lambda^{-1})^{-1}M^{1/2}) \right] \\ &= \mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\theta} \left[ \sigma_w^2 \text{tr}(M^{1/2}(\Sigma_T + \Lambda^{-1})^{-1}M^{1/2}) \right], \end{aligned} \quad (\text{D.7})$$

where in the last line, we have invoked Lemma D.1.

**Upper bounding term (b):** Let  $p_r := \mathbb{P}_{\theta \sim \mathcal{N}(\theta_0, \Lambda)}[\|\theta - \theta_0\| > r]$  denote the probability that  $\theta$  lies within the truncation region. Then, for any nonnegative function  $f(\theta) \geq 0$ ,

$$\mathbb{E}_{\theta \sim \mathcal{N}_{\text{tr}}}[f(\theta)] = \frac{1}{1 - p_r} \mathbb{E}_{\theta \sim \mathcal{N}(\theta_0, \Lambda)}[f(\theta) \cdot \mathbb{I}\{\|\theta - \theta_0\| \leq r\}] \leq \frac{1}{1 - p_r} \mathbb{E}_{\theta \sim \mathcal{N}(\theta_0, \Lambda)}[f(\theta)]$$

Thus,

$$\begin{aligned} \text{term (b)} &= \mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\mathcal{D}_T \sim \boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}' \sim \mathcal{D}(\mathcal{D}_T)} \left[ \|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2^2 \cdot \mathbb{I}\{\|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2^2 > r^2\} \right] \\ &\leq \frac{1}{1-p_r} \mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Lambda)} \mathbb{E}_{\mathcal{D}_T \sim \boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}' \sim \mathcal{D}(\mathcal{D}_T)} \left[ \|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2^2 \cdot \mathbb{I}\{\|\boldsymbol{\theta}' - \boldsymbol{\theta}_0\|_2^2 > r^2\} \right] \end{aligned} \quad (\text{D.8})$$

By the replica lemma (Lemma D.1), the second line is equal to

$$\text{Eq. (D.8)} = \frac{1}{1-p_r} \mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Lambda)} \left[ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2 \cdot \mathbb{I}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2 > r\} \right].$$

We now bound the above. Note that  $\boldsymbol{\theta} - \boldsymbol{\theta}_0$  has the same distribution as  $\Lambda^{1/2} \mathbf{g}$ , where  $\mathbf{g} \sim \mathcal{N}(0, I_d)$ . Hence,  $p_r = \mathbb{P}_{\mathbf{g} \sim \mathcal{N}(0, I_d)}[\mathbf{g}^\top \Lambda \mathbf{g} > r^2]$ . For  $r^2 \geq 2\text{tr}(\Lambda)$ , Markov's inequality therefore implies  $p_r \leq 1/2$ , so that  $\frac{1}{1-p_r} \leq 2$ . Moreover, by the same change of variables,

$$\mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Lambda)} \left[ \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2 \cdot \mathbb{I}\{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2^2 > r^2\} \right] = \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(0, I_d)} \left[ \mathbf{g}^\top \Lambda \mathbf{g} \cdot \mathbb{I}\{\mathbf{g}^\top \Lambda \mathbf{g} > r^2\} \right],$$

Thus, for  $r^2 \geq 2\text{tr}(\Lambda) = 2\text{tr}(\Gamma^{-1})$  (which follows from the condition of the theorem,  $r \geq \sqrt{5\text{tr}(\Gamma^{-1})}$ ), it holds that

$$\text{term (b)} \leq 2 \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(0, I_d)} \left[ \mathbf{g}^\top \Lambda \mathbf{g} \cdot \mathbb{I}\{\mathbf{g}^\top \Lambda \mathbf{g} > r^2\} \right] = 2 \int_{r^2}^{\infty} \mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > r^2] dr^2.$$

We now require invoke a coarse consequence of the Hanson-Wright inequality:

**Lemma D.4** (Consequence of Hanson-Wright). *For any  $u \geq 4\text{tr}(\Lambda)$ , we have*

$$\mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \text{tr}(\Lambda) + u] \leq e^{-\frac{u}{4\|\Lambda\|_{\text{op}}}}.$$

*Proof.* From Proposition 1.1. in (Hsu et al., 2012), we have  $t > 0$ , it holds that  $\mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \text{tr}(\Lambda) + 2\sqrt{t}\|\Lambda\|_{\text{F}} + 2\|\Lambda\|_{\text{op}}t] \leq e^{-t}$ . In particular, if  $\sqrt{t}\|\Lambda\|_{\text{op}} \geq \|\Lambda\|_{\text{F}}$ , then  $\mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \text{tr}(\Lambda) + 4t\|\Lambda\|_{\text{op}}] \leq e^{-t}$ . Reparametrizing  $u = 4t\|\Lambda\|_{\text{op}}$ , we have that if  $\sqrt{u}\|\Lambda\|_{\text{op}}/2 \geq \|\Lambda\|_{\text{F}}$ , then  $\mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \text{tr}(\Lambda) + u] \leq e^{-u/4\|\Lambda\|_{\text{op}}}$ . Lastly, the condition  $\sqrt{u}\|\Lambda\|_{\text{op}}/2 \leq \|\Lambda\|_{\text{F}}$  is equivalent to  $u \geq 4\|\Lambda\|_{\text{F}}^2/\|\Lambda\|_{\text{op}}$ . Since  $\|\Lambda\|_{\text{F}}^2 \leq \text{tr}(\Lambda)\|\Lambda\|_{\text{op}}$ , it suffices that  $u \geq 4\text{tr}(\Lambda)$ . This concludes the proof.  $\square$

Hence, under the assumption of the theorem,  $r^2 \geq 5\text{tr}(\Gamma^{-1}) = 5\text{tr}(\Lambda)$ , we may bound

$$\begin{aligned} \text{term (b)} &\leq 2 \int_{r^2}^{\infty} \mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \tau] d\tau \\ &= 2 \int_{r^2 - \text{tr}(\Lambda)}^{\infty} \mathbb{P}[\mathbf{g}^\top \Lambda \mathbf{g} > \text{tr}(\Lambda) + \tau] d\tau \\ &\leq 2 \int_{r^2 - \text{tr}(\Lambda)}^{\infty} e^{-\frac{u}{4\|\Lambda\|_{\text{op}}}} = 8\|\Lambda\|_{\text{op}} e^{-\frac{r^2 - \text{tr}(\Lambda)}{4\|\Lambda\|_{\text{op}}}} \\ &\leq 8\|\Lambda\|_{\text{op}} e^{-\frac{r^2}{5\|\Lambda\|_{\text{op}}}}. \end{aligned} \quad (\text{D.9})$$

**Concluding the Proof:** Combining Eqs. (D.6), (D.7) and (D.9), we have

$$\begin{aligned} \inf_{\hat{\boldsymbol{\theta}}} \max_{\boldsymbol{\theta} \in \mathcal{B}} \mathbb{E}_{\boldsymbol{\theta}} \mathcal{R}_{1s}(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}) &\geq \mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\boldsymbol{\theta}} \left[ \sigma_w^2 \text{tr}(M^{1/2}(\boldsymbol{\Sigma}_T + \Lambda^{-1})^{-1} M^{1/2}) \right] \\ &\quad - 32\|\Lambda\|_{\text{op}} \|M\|_{\text{op}} e^{-\frac{r^2}{5\|\Lambda\|_{\text{op}}}} \end{aligned}$$

Since  $\Lambda = \Gamma^{-1}$ , the last line of the above display as  $\Psi(r; \Gamma, M) := \frac{32\|M\|_{\text{op}}}{\lambda_{\min}(\Gamma)} \exp(-\frac{r^2}{5} \lambda_{\min}(\Gamma))$ . Finally, we lower bound the first line of the above display crudely via Jensen's inequality: indeed, since  $X \mapsto \text{tr}(X^{-1})$  is a convex function (on the domain of positive-definite matrices), and since convexity is preserved under affine transformation, we have

$$\mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}_{\text{tr}}} \mathbb{E}_{\boldsymbol{\theta}} \left[ \sigma_w^2 \text{tr}(M^{1/2}(\boldsymbol{\Sigma}_T + \Lambda^{-1})^{-1} M^{1/2}) \right] \geq \mathbb{E}_{\boldsymbol{\theta} \sim \mathcal{N}_{\text{tr}}} \left[ \sigma_w^2 \text{tr}(M^{1/2}(\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Sigma}_T] + \Lambda^{-1})^{-1} M^{1/2}) \right]$$

Substituting in  $\Gamma = \Lambda^{-1}$ , and noting that the distribution  $\mathcal{N}_{\text{tr}}$  is supported on the ball  $\mathcal{B}$  concludes the bound.  $\square$

### D.3. Proof of $M$ -norm Regression Upper Bound (Theorem D.2)

*Proof.* Let  $\beta$  be a parameter to be tuned, and set. Since  $M$  may not be full rank, we consider a perturbation

$$N := M + \zeta I, \quad \zeta = \text{tr}(M\Gamma^{-1})/\text{tr}(\Gamma^{-1}),$$

where  $\zeta > 0$  is to be chosen. We further define

- $\mathbf{Z} \in \mathbb{R}^{T \times d}$  denote the matrix whose rows are  $z_t^\top$
- $\mathbf{w} \in \mathbb{R}^T$  as the vector whose entries are  $w_t$ .
- $\hat{\theta}_{\text{ls}}$  the least squares estimate of  $\theta_*$  defined in Eq. (D.2)
- We let  $v_1, \dots, v_d$  be the eigenvectors of  $N^{1/2}\Gamma^{-1}N^{1/2}$  and  $\lambda_j := \lambda_j(N^{1/2}\Gamma^{-1}N^{1/2})$ , which we note are deterministic.

The error of the least-squares estimate is then,

$$\|\hat{\theta}_{\text{ls}} - \theta_*\|_M^2 = \|(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w}\|_M^2$$

Since  $\epsilon < \lambda_{\min}(\Gamma)/4$ , we can apply Lemma J.4 to get

$$\|(\mathbf{Z}^\top \mathbf{Z})^{-1} - \Gamma^{-1}\|_{\text{op}} \leq \frac{\epsilon}{\lambda_{\min}(\Gamma)(\lambda_{\min}(\Gamma) - \epsilon)} < \frac{2\epsilon}{\lambda_{\min}(\Gamma)^2}$$

We now invoke the following lemma, controlling the relation of (weighted) squares of matrices in the PSD order:

**Lemma D.5.** *Let  $A, B, M \succeq 0$  and  $C = A - B$ . Then,*

$$AMA + 7CMC \succeq BMB/2$$

The lemma is proven at the end of this section. Instantiating Lemma D.5 with  $A = \Gamma^{-1}$  and  $B = (\mathbf{Z}^\top \mathbf{Z})^{-1}$  and  $M = N$ , we have

$$\Gamma^{-1}N\Gamma^{-1} + \frac{13\|N\|_{\text{op}}\epsilon^2}{\lambda_{\min}(\Gamma)^4}I \succeq (\mathbf{Z}^\top \mathbf{Z})^{-1}N(\mathbf{Z}^\top \mathbf{Z})^{-1}$$

Suppose that  $\epsilon$  is chosen sufficiently small that, for a constant  $\alpha$  to be specified

$$\alpha\Gamma^{-1}N\Gamma^{-1} \succeq \frac{13\|N\|_{\text{op}}\epsilon^2}{\lambda_{\min}(\Gamma)^4}; \quad (\text{D.10})$$

we shall revisit this point at the end of the proof. Then,

$$\|(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w}\|_M^2 \leq \|(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w}\|_N^2 \quad (\text{D.11})$$

$$\begin{aligned} &= \mathbf{w}^\top \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1}N(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w} \\ &\leq (1 + \alpha) \mathbf{w}^\top \mathbf{Z}\Gamma^{-1}N\Gamma^{-1} \mathbf{Z}^\top \mathbf{w} \\ &= (1 + \alpha) \|N^{1/2}\Gamma^{-1} \mathbf{Z}^\top \mathbf{w}\|_2^2 \end{aligned}$$

$$= (1 + \alpha) \sum_{j=1}^{d_\theta} (v_j^\top N^{1/2}\Gamma^{-1}N^{1/2}N^{-1/2} \mathbf{Z}^\top \mathbf{w})^2$$

$$= (1 + \alpha) \sum_{j=1}^{d_\theta} \lambda_j^2 (v_j^\top N^{-1/2} \mathbf{Z}^\top \mathbf{w})^2$$

$$= (1 + \alpha) \sum_{j=1}^{d_\theta} \lambda_j^2 (v_j^\top \mathbf{Z}_j^\top \mathbf{w})^2, \quad \text{where } \mathbf{Z}_j = \mathbf{Z}N^{-1/2}v_j \in \mathbb{R}^T. \quad (\text{D.12})$$

We specialize the self-normalized martingale concentration inequality:

**Lemma D.6** (Theorem 1 of (Abbasi-Yadkori et al., 2011)). *Let  $\{\mathbf{e}_t\}_{t \geq 1} \in \mathbb{R}^N$  be a scalar,  $\mathcal{F}_t$ -adapted sequence such that  $\mathbf{e}_t | \mathcal{F}_{t-1}$  is  $\sigma_w^2$  sub-Gaussian. Let  $\{\mathbf{x}_t\}_{t \geq 1} \in (\mathbb{R}^{d_\theta})^N$  be a sequence of  $\mathcal{F}_t$ -adapted vectors. Fix a matrix  $V_0 \succeq 0$ . Then, with probability  $1 - \delta$ ,*

$$\left\| \sum_{t=1}^T \mathbf{x}_t \mathbf{e}_t \right\|_{(V_0 + \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^\top)^{-1}} \leq 2\sigma_w^2 \log \left( \frac{1}{\delta} \det \left( V_0^{-1/2} \left( V_0 + \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^\top \right) V_0^{-1/2} \right) \right).$$

In particular, if  $d = 1$ , then selecting a scalar  $V_0 = \tau$ ,

$$\left\| \sum_{t=1}^T \mathbf{x}_t \mathbf{e}_t \right\| \leq 2\sigma_w^2 \left( \tau + \sum_{t=1}^T \|\mathbf{x}_t\|^2 \right) \log \left( \frac{\tau + \sum_{t=1}^T \|\mathbf{x}_t\|^2}{\delta \tau} \right).$$

Applying Lemma D.6 with a union bound over indices  $j \in [d]$ , it holds with probability  $1 - \delta$  for all  $j \in [d]$  simultaneously for any fixed  $\tau > 0$

$$(v_j^\top \mathbf{Z}_j^\top \mathbf{w})^2 \leq 2\sigma_w^2 (\|\mathbf{Z}_j\|_2^2 + \tau \lambda_j) \log \frac{d(\|\mathbf{Z}_j\|_2^2 + \tau \lambda_j)}{\tau \lambda_j \delta} \quad (\text{D.13})$$

In addition, note that for any  $\tau \geq \beta$ ,

$$\begin{aligned} \|\mathbf{Z}_j\|_2^2 &= v_j^\top N^{-1/2} \mathbf{Z}^\top \mathbf{Z} N^{-1/2} v_j \leq v_j^\top N^{-1/2} (\Gamma + \epsilon I) N^{-1/2} v_j \\ &\leq (1 + \beta) v_j^\top N^{-1/2} \Gamma N^{-1/2} v_j = (1 + \beta) \lambda_j^{-1} \leq (1 + \tau) \lambda_j^{-1} \end{aligned}$$

Hence, with probability  $1 - \delta$  the following holds for all  $j \in [d]$  simultaneously

$$(v_j^\top \mathbf{Z}_j^\top \mathbf{w})^2 \leq \frac{2(1 + 2\tau)\sigma_w^2}{\lambda_j} \log \frac{d(\tau^{-1} + 2)}{\delta}$$

Hence, combining with Eq. (D.12), we have that with probability

$$\begin{aligned} \|(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w}\|_M^2 &\leq 2(1 + \alpha)(1 + \tau)\sigma_w^2 \sum_{j=1}^{d_\theta} \lambda_j \log \frac{d(\tau^{-1} + 2)}{\delta} \\ &\leq 2(1 + \alpha)(1 + \tau)\sigma_w^2 \log \frac{d(\tau^{-1} + 2)}{\delta} \cdot \text{tr}(N^{1/2} \Gamma^{-1} N^{1/2}). \end{aligned}$$

Finally, we can simplify  $\text{tr}(N^{1/2} \Gamma^{-1} N^{1/2}) = \text{tr}(N \Gamma^{-1}) = \text{tr}((M + \zeta I) \Gamma^{-1}) \leq 2\text{tr}(M \Gamma^{-1})$  for our choice of  $\zeta = \text{tr}(M \Gamma^{-1}) / \text{tr}(\Gamma^{-1})$ ; thus, choosing  $\tau \geq 1/4 \geq \beta$  (recall the assumption,  $\beta \leq 1/4$ ), we have

$$\|(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{w}\|_M^2 \leq 5\sigma_w^2 (1 + \alpha) \log \frac{6d}{\delta} \cdot \text{tr}(M \Gamma^{-1}), \text{ w.p. } 1 - \delta. \quad (\text{D.14})$$

To conclude, let us compute find a suitable constant  $\alpha$  satisfying Eq. (D.10). Recall that we want  $\alpha \Gamma^{-1} N \Gamma^{-1} \succeq \frac{13\|N\|_{\text{op}} \epsilon^2}{\lambda_{\min}(\Gamma)^4}$ . Since  $N \succeq \zeta$ , and  $\epsilon \leq \beta \lambda_{\min}(\Gamma)$ , we want

$$\alpha \zeta \geq 13\beta^2 \|N\|_{\text{op}} = 13\beta^2 (\|M\|_{\text{op}} + \zeta)$$

Recalling  $\zeta = \text{tr}(M \Gamma^{-1}) / \text{tr}(\Gamma^{-1}) \leq \|M\|_{\text{op}}$ , we can chose  $\alpha \geq \frac{26\beta^2 \|M\|_{\text{op}} \text{tr}(\Gamma^{-1})}{\text{tr}(M \Gamma^{-1})}$ . In particular, since  $\text{tr}(M \Gamma^{-1}) \geq \|M\|_{\text{op}} \lambda_{\min}(\Gamma^{-1})$ , we can take

$$\alpha = \frac{26\beta^2 \|M\|_{\text{op}} \text{tr}(\Gamma^{-1})}{\|M\|_{\text{op}} \lambda_{\min}(\Gamma)^{-1}} = 26\beta^2 \lambda_{\max}(\Gamma) \text{tr}(\Gamma^{-1})$$

□

*Proof of Lemma D.5.* Clearly,  $AMA = BMB + CMC + BMC + CMB$ . Since  $CMC, BMB \succeq 0$ :

$$\begin{aligned} BMC + CMB &\succeq BMC + CMB - 4CMC - BMB/4 \\ &= -(B/2 - 2C)M(B/2 - 2C) \\ &\succeq -2(BMB/4 + 4CMC) \\ &= -BMB/2 - 8CMC \end{aligned}$$

Thus,

$$AMA \succeq BMB + CMC - BMB/2 - 8CMC = BMB/2 - 7CMC.$$

□

## E. Lower Bounds on Martingale Decision Making

In this section, we convert the information theoretic bounds for  $M$ -norm estimation considered in Appendix D to the smooth decision making problem with decision  $\mathbf{a} \in \mathbb{R}^{d_a}$ . In particular, in this section we work in the MDM setting.

**Smoothness Assumptions and Consequences.** We first recall the smoothness assumption on our loss  $\mathcal{R}$  in the MDM setting.

**Assumption 3** (Smooth Decision-Making). *There exist  $r_{\text{quad}}(\theta_*)$  and constants  $\mu > 0$ ,  $L_{\mathbf{a}i}, L_{\mathcal{R}i}$ ,  $i \in \{1, 2, 3\}$ , and  $L_{\text{hess}}$  such that for any  $\theta$  and  $\mathbf{a}$  satisfying*

$$\|\theta - \theta_*\|_2 \leq r_{\text{quad}}(\theta_*), \quad \|\mathbf{a} - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \leq L_{\mathbf{a}1} r_{\text{quad}}(\theta_*), \quad (\text{B.1})$$

the following conditions hold

- The optimal action  $\mathbf{a}_{\text{opt}}(\theta)$  is unique, and moreover, there is a parameter  $\mu$  such that  $\mathcal{R}(\mathbf{a}'; \theta) \geq \frac{\mu}{2} \|\mathbf{a}' - \mathbf{a}_{\text{opt}}(\theta)\|_2^2$  for all  $\mathbf{a}' \in \mathbb{R}^{d_a}$  (not restricted to  $\mathbf{a}'$  satisfying Eq. (B.1)).
- $\|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}1}$ ,  $\|\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}2}$ , and  $\|\nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)\|_{\text{op}} \leq L_{\mathcal{R}3}$ .
- $\|\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)\|_{\text{op}} \leq L_{\mathbf{a}1}$ ,  $\|\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)\|_{\text{op}} \leq L_{\mathbf{a}2}$ , and  $\|\nabla_{\theta}^3 \mathbf{a}_{\text{opt}}(\theta)[\delta, \delta, \delta]\|_{\text{op}} \leq L_{\mathbf{a}3}$  for all  $\delta \in \mathbb{R}^{d_{\theta}}$  with  $\|\delta\|_2 = 1$ .
- $\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)$  is Lipschitz in  $\theta$  with Lipschitz constant  $L_{\text{hess}}$ .

The above assumption directly yields the following Lipschitz conditions.

**Proposition E.1.** *Assume that  $\mathcal{R}, \mathbf{a}_{\text{opt}}$  satisfy Assumption 3. Then for any model  $\theta \in \mathbb{R}^{d_{\theta}}$  and action  $\mathbf{a} \in \mathbb{R}^{d_a}$  satisfying (B.1), it holds that*

- $\nabla_{\mathbf{a}}^{(i)} \mathcal{R}(\mathbf{a}; \theta)$  is Lipschitz in the operator norm with Lipschitz constant  $L_{\mathcal{R}(i+1)}$  for  $i = 0, 1, 2$
- $\nabla_{\theta}^{(i)} \mathbf{a}_{\text{opt}}(\theta)$  is Lipschitz in the operator norm with Lipschitz constant  $L_{\mathbf{a}(i+1)}$ , for  $i = 0, 1$ .

In the above, we adopted the convention  $\nabla_x^{(0)} f(x) = f(x)$ .

**Relating Smooth Decision Making to  $M$ -norm Estimation.** The next step is to relate smooth decision making to  $M$ -norm estimation. We begin by introducing the relevant gradients and Hessians, and in particular, the *task Hessian*  $\mathcal{H}(\theta)$  introduced in Section 2.

**Definition E.1** (Key Gradients and Hessians). For some  $\theta_*$  and function  $\mathcal{R}, \mathbf{a}_{\text{opt}}$ , let:

- $M_{\mathbf{a}}(\theta_*) := \nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta_*)$  at  $\mathbf{a} = \mathbf{a}_{\text{opt}}(\theta_*)$ .
- $G_{\mathbf{a}}(\theta_*) := \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)$  at  $\theta = \theta_*$ .
- $\mathcal{H}(\theta_*) = \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)$  at  $\theta = \theta_*$ . In particular,  $\mathcal{H}(\theta_*) = G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*) G_{\mathbf{a}}(\theta_*)$ .

The following result utilizes Assumption 3 to guarantee that  $\mathcal{H}(\theta)$  is itself a smooth map, and that the norm induced by  $\mathcal{H}(\theta_*)$  can be used to approximate  $\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}); \theta_*)$ , both of which are critical pieces in our analysis.

**Proposition E.2.** *Assume that  $\mathcal{R}$ ,  $\mathbf{a}_{\text{opt}}$  satisfy Assumption 3 and that  $\hat{\theta}$  satisfies  $\|\hat{\theta} - \theta_*\|_2 \leq r_{\text{quad}}(\theta_*)$ . Then the following hold:*

$$\left| \mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}); \theta_*) - \|\hat{\theta} - \theta_*\|_{\mathcal{H}(\theta_*)}^2 \right| \leq L_{\text{quad}} \cdot \|\hat{\theta} - \theta_*\|_2^3, \quad \|\mathcal{H}(\theta_*) - \mathcal{H}(\hat{\theta})\|_{\text{op}} \leq L_{\mathcal{H}} \cdot \|\hat{\theta} - \theta_*\|_2.$$

where:

$$L_{\text{quad}} := \frac{1}{6} (L_{\mathcal{R}3} L_{\mathbf{a}1}^3 + 3L_{\mathcal{R}2} L_{\mathbf{a}2} L_{\mathbf{a}1} + L_{\mathcal{R}1} L_{\mathbf{a}3}), \quad L_{\mathcal{H}} := 6L_{\text{quad}} + L_{\mathcal{R}2} L_{\mathbf{a}1} + L_{\text{hess}} L_{\mathbf{a}1}^2.$$

**Lower Bound for Smooth Decision Making.** We now state our key lemma, which allows us to reduce smooth decision making to  $M$ -norm estimation and obtain a lower bound on the local minimax risk in terms of estimation in a particular norm.

**Lemma E.1.** *Assume that the excess risk  $\mathcal{R}$  and optimal decision function  $\mathbf{a}_{\text{opt}}$  satisfy Assumption 3 with smoothness parameters  $L_{\mathcal{R}i}$  and  $L_{\mathbf{a}i}$  and radius parameters  $r_{\text{quad}}$  dictating the region in which the smoothness holds. Let  $r > 0$  be a radius parameter satisfying*

$$r \leq r_{\text{quad}}(\theta_*)/4$$

and define the associated balls  $\mathcal{B}_T(\theta_*) := \{\theta : \|\theta - \theta_*\|_2 \leq r\}$ . Then,

$$\min_{\hat{\mathbf{a}}} \max_{\theta \in \mathcal{B}_T(\theta_*)} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\mathcal{R}(\hat{\mathbf{a}}; \theta)] \geq \min \left\{ \min_{\hat{\theta}} \max_{\theta \in \mathcal{B}_T(\theta_*)} \frac{1}{2} \mathbb{E}_{\theta, \pi_{\text{exp}}} \left[ \|\hat{\theta} - \theta\|_{\mathcal{H}(\theta_*)}^2 \right] - C_1 r^3, \mu L_{\mathbf{a}1}^2 r^2 \right\}$$

where we define the constant, for a universal numerical constant  $c_1$ ,

$$C_1 = c_1 \left( L_{\mathbf{a}1} L_{\mathbf{a}2} L_{\mathcal{R}2} + L_{\mathbf{a}1}^3 L_{\mathcal{R}3} + L_{\text{hess}} \right).$$

By tuning the radius parameter  $r(T)$  appropriately, we achieve the following general purpose lower bound on the excess risk:

**Theorem E.3.** *Suppose the smoothness assumption Assumption 3 holds with its stated smoothness parameters. In addition, fix a regularization parameter  $\lambda > 0$ , and suppose that  $T$  satisfies*

$$\lambda T \geq \max \left\{ \left( \frac{80d_{\theta}}{r_{\text{quad}}(\theta_*)^2} \right)^{6/5}, \left( \frac{\sigma_w^2 L_{\mathcal{R}2}}{5\mu} \right)^6 \right\}$$

Finally, define the localizing ball  $\mathcal{B}_T := \{\theta : \|\theta - \theta_*\|_2 \leq 5d_{\theta}/(\lambda T)^{5/6}\}$ . Then, for any  $\theta_0 \in \mathcal{B}_T(\theta_*)$ ,

$$\min_{\hat{\mathbf{a}}} \max_{\theta \in \mathcal{B}_T} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\mathcal{R}(\hat{\mathbf{a}}; \theta)] \geq \sigma_w^2 \min_{\theta \in \mathcal{B}_T} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbb{E}_{\theta, \pi_{\text{exp}}} [\mathbf{\Sigma}_T] + \lambda T \cdot I \right)^{-1} \right) - \frac{C_2}{(\lambda T)^{5/4}}.$$

where we have defined the constant, for a universal numerical constant  $c_2$ ,

$$C_2 = c_2 \left( (L_{\mathbf{a}1} L_{\mathbf{a}2} L_{\mathcal{R}2} + L_{\mathbf{a}1}^3 L_{\mathcal{R}3} + L_{\text{hess}}) d_{\theta}^{3/2} + L_{\mathbf{a}1}^2 L_{\mathcal{R}2} \right).$$

## E.1. Proof of General Decision Making Lower Bounds

For the remainder of Section E, unless otherwise stated we assume the expectation is taken with respect to  $\theta$  and some fixed exploration policy  $\pi_{\text{exp}}$ . Hence, we write  $\mathbb{E}[\cdot]$  in place of  $\mathbb{E}_{\theta, \pi_{\text{exp}}}[\cdot]$ .

## E.1.1. PROOF OF THEOREM E.3

For simplicity, we shall write  $r^2 = 5d_\theta/(\lambda T)^{5/6}$ . The result follows by instantiating Theorem D.1 to lower bound:

$$\min_{\hat{\theta}} \max_{\theta: \|\theta - \theta_\star\|_2 \leq r} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\|\hat{\theta} - \theta\|_{M(\theta_\star)}^2],$$

via the simplification provided by Lemma E.1. Apply Theorem D.1 with parameters  $\Gamma = \lambda T \cdot I$ , so that  $\text{tr}(\Gamma^{-1}) = \frac{d_\theta}{\lambda T}$  and  $\lambda_{\min}(\Gamma) = \lambda T$ . Then, the remainder term  $\Psi$  from that theorem is bounded by

$$\Psi(r; \Gamma, \mathcal{H}(\theta_\star)) \leq \frac{32\|\mathcal{H}(\theta_\star)\|_{\text{op}}}{\lambda T} \exp\left(-\frac{1}{5}r^2\lambda T\right)$$

Selecting  $r^2 = 5d_\theta/(\lambda T)^{5/6}$  yields

$$\Psi(r; \Gamma, \mathcal{H}(\theta_\star)) \leq \frac{32\|\mathcal{H}(\theta_\star)\|_{\text{op}}}{\lambda T} \exp\left(-d_\theta(\lambda T)^{1/6}\right)$$

Noting that  $r^2 \geq 5\text{tr}(\Gamma)$ , Theorem D.1 yields that

$$\begin{aligned} & \min_{\hat{\theta}} \max_{\theta: \|\theta - \theta_\star\|_2 \leq r} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\|\hat{\theta} - \theta\|_{\mathcal{H}(\theta_\star)}^2] \\ & \geq \sigma_w^2 \min_{\theta: \|\theta - \theta_\star\|_2 \leq r} \text{tr}\left(\mathcal{H}(\theta_\star) (\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] + \lambda T \cdot I)^{-1}\right) - \frac{32\|\mathcal{H}(\theta_\star)\|_{\text{op}}}{\lambda T} \exp\left(-d_\theta(\lambda T)^{1/6}\right) \\ & \geq \sigma_w^2 \min_{\theta: \|\theta - \theta_\star\|_2 \leq r} \text{tr}\left(\mathcal{H}(\theta_\star) (\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] + \lambda T \cdot I)^{-1}\right) - \frac{32L_{\text{a1}}^2 L_{\mathcal{R}2}}{\lambda T} \exp\left(-d_\theta(\lambda T)^{1/6}\right), \end{aligned}$$

where in the last line, we invoked Assumption 3 to obtain

$$\|\mathcal{H}(\theta_\star)\|_{\text{op}} = \|G_{\text{a}}(\theta_\star)^\top M_{\text{a}}(\theta_\star) G_{\text{a}}(\theta_\star)\|_{\text{op}} \leq L_{\text{a1}}^2 L_{\mathcal{R}2}$$

Now, observe that our condition on  $\lambda T$ , namely  $\lambda T \geq (80d_\theta/r_{\text{quad}}(\theta_\star)^2)^{6/5}$ , implies that  $r = (5d_\theta/(\lambda T)^{5/6})^{1/2} \leq \frac{1}{4}r_{\text{quad}}(\theta_\star)$ . Hence, we can apply Lemma E.1 to obtain

$$\begin{aligned} \min_{\hat{\mathbf{a}}} \max_{\theta: \|\theta - \theta_\star\|_2 \leq r} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] & \geq \min \left\{ 5\mu L_{\text{a1}}^2 r^2, \sigma_w^2 \min_{\theta: \|\theta - \theta_\star\|_2 \leq r} \text{tr}\left(\mathcal{H}(\theta_\star) (\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] + \lambda T I)^{-1}\right) \right. \\ & \left. - C_1 r^3 - \frac{32L_{\text{a1}}^2 L_{\mathcal{R}2}}{\lambda T} \exp\left(-d_\theta(\lambda T)^{1/6}\right) \right\} \end{aligned}$$

for  $C_1$  as in Lemma E.1 To conclude, we consolidate

$$C_1 r^3 + \frac{32L_{\text{a1}}^2 L_{\mathcal{R}2}}{\lambda T} \exp\left(-d_\theta(\lambda T)^{1/6}\right) = \frac{C_1(5d_\theta)^{3/2} + 32L_{\text{a1}}^2 L_{\mathcal{R}2}(\lambda T)^{1/4} \exp\left(-d_\theta(\lambda T)^{1/6}\right)}{(\lambda T)^{5/4}}$$

Observing that  $(\lambda T)^{1/4} \exp\left(-d_\theta(\lambda T)^{1/6}\right)$  is bounded above by a universal constant (since  $d_\theta \geq 1$ , and for all  $d \geq 1$ ,  $\max_{x \geq 0} x e^{-dx} \leq \max_{x \geq 0} x e^{-x}$  is bounded), the above is at most

$$\mathcal{O}\left(\frac{C_1 d_\theta^{3/2} + L_{\text{a1}}^2 L_{\mathcal{R}2}}{(\lambda T)^{5/4}}\right) = \frac{C_2}{(\lambda T)^{5/4}},$$

for  $C_2$  as in the statement of the lemma. To conclude, it suffices to show that for our choice of  $r$ , we have

$$5\mu L_{\text{a1}}^2 r^2 \geq \sigma_w^2 \min_{\theta: \|\theta - \theta_\star\|_2 \leq r} \text{tr}\left(\mathcal{H}(\theta_\star) (\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] + \lambda T I)^{-1}\right).$$

Lower bounding  $\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] + \lambda TI \succeq \lambda TI$ , and upper bounding  $\text{tr}(\mathcal{H}(\theta_\star)) \leq d_\theta \|\mathcal{H}(\theta_\star)\|_{\text{op}} \leq d_\theta L_{\mathcal{a}1}^2 L_{\mathcal{R}2}$ , and substituting in the choice of  $r^2$ , it is enough that

$$\mu L_{\mathcal{a}1}^2 \cdot (5d_\theta / (\lambda T))^{5/6} \geq \sigma_w^2 \frac{d_\theta L_{\mathcal{a}1}^2 L_{\mathcal{R}2}}{\lambda T}$$

Rearranging requires that

$$(\lambda T)^{1/6} \geq \frac{\sigma_w^2 L_{\mathcal{R}2}}{5\mu},$$

which is satisfied for our choice of  $\lambda$ .

### E.1.2. PROOF OF LEMMA E.1

Our strategy is to show that an action  $\mathbf{a}$  with low excess risk can be used to produce an estimate of a parameter  $\theta$  with low error in the task hessian norm  $\|\cdot\|_{\mathcal{H}(\theta)}^2$ . Specifically, we define the perturbation term

$$\begin{aligned} \delta_\star(\hat{\mathbf{a}}) &= \arg \min_{\delta} \|M_{\mathbf{a}}(\theta_\star)^{1/2} ((\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)) - G_{\mathbf{a}}(\theta_\star)\delta)\|_2 \\ &= (M_{\mathbf{a}}(\theta_\star)^{1/2} G_{\mathbf{a}}(\theta_\star)^\dagger M_{\mathbf{a}}(\theta_\star)^{1/2} (\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star))) \end{aligned}$$

and define the induced estimate

$$\hat{\theta}(\hat{\mathbf{a}}) = \theta_\star + \delta_\star(\hat{\mathbf{a}})$$

**Ensuring  $\hat{\mathbf{a}}$  close  $\mathbf{a}_{\text{opt}}(\theta_\star)$ :** We first want to restrict the lower bound to being only over  $\hat{\mathbf{a}}$  close  $\mathbf{a}_{\text{opt}}(\theta_\star)$ . To this end, note that

$$\min_{\hat{\mathbf{a}}} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r(T)} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] = \min \left\{ \begin{array}{l} \min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 \leq r_a^2} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r^2} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)], \\ \min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 > r_a^2} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r^2} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] \end{array} \right\}$$

where we are free to choose  $r_a$  as we wish but our choice will satisfy  $r_a \geq L_{\mathcal{a}1} r$ . Now:

$$\begin{aligned} \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 &\leq \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)\|_2 + \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 \\ &\leq \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)\|_2 + L_{\mathcal{a}1} \|\theta - \theta_\star\|_2 \\ &\leq \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)\|_2 + L_{\mathcal{a}1} r \end{aligned}$$

By the argument above,  $\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 > r_a(T)$  then implies that:

$$\mathcal{R}(\hat{\mathbf{a}}; \theta) \geq \frac{\mu}{2} \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)\|_2^2 \geq \frac{\mu}{2} (\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 - L_{\mathcal{a}1} r)^2 \geq \frac{\mu}{2} (r_a - L_{\mathcal{a}1} r)^2$$

Choosing  $r_a = (1 + \sqrt{2})L_{\mathcal{a}1} r$ ,

$$\min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 > r_a^2} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r^2} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] \geq \mu L_{\mathcal{a}1}^2 r^2$$

Thus, defining the constant

$$C_{\mathbf{a}} := (1 + \sqrt{2})L_{\mathcal{a}1},$$

So ultimately we have:

$$\min_{\hat{\mathbf{a}}} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r(T)} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] \geq \min \left\{ \begin{array}{l} \min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 \leq C_{\mathbf{a}}^2 r^2} \max_{\theta: \|\theta - \theta_\star\|_2^2 \leq r^2} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)], \mu L_{\mathcal{a}1}^2 r^2 \end{array} \right\} \quad (\text{E.1})$$

We now proceed to lower bound the first term in the above expression. In particular, throughout we assume that

$$\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 \leq C_{\mathbf{a}} r, \quad C_{\mathbf{a}} := (1 + \sqrt{2})L_{\mathcal{a}1}, \quad (\text{E.2})$$



**Taylor expansions:** Fix a  $t \in [0, 1]$ , parameter  $\theta$ , estimated action  $\hat{\mathbf{a}}$ , and define the interpolations

$$\hat{\mathbf{a}}_t = t\hat{\mathbf{a}} + (1-t)\mathbf{a}_{\text{opt}}(\theta), \quad \theta_t := t\theta + (1-t)\theta_t$$

Throughout, we will let  $\hat{\mathbf{a}}'$  and  $\theta'$  denote certain values of  $\hat{\mathbf{a}}_t$  and  $\theta_t$  for some interpolation parameters  $t', t'' \in [0, 1]$  chosen so as to satisfy the application of Taylor's theorem to follow.

First, by Taylor's theorem,

$$\begin{aligned} \mathcal{R}(\hat{\mathbf{a}}; \theta) &= \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta) + \frac{d}{dt} \mathcal{R}(\hat{\mathbf{a}}_t; \theta)|_{t=0} + \frac{1}{2} \frac{d^2}{dt^2} \mathcal{R}(\hat{\mathbf{a}}_t; \theta)|_{t=0} + \frac{1}{6} \frac{d^3}{dt^3} \mathcal{R}(\hat{\mathbf{a}}_t; \theta)|_{t=t'} \\ &= \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta) + (\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)})^\top (\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)) \\ &\quad + \frac{1}{2} (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}})^\top (\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)}) (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}}) \\ &\quad + \frac{1}{6} \nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}'} [\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)] \end{aligned}$$

where  $t' \in [0, 1]$ ,  $\mathbf{a}' = \hat{\mathbf{a}}_{t'}$ . The second equality follows by the chain rule and since  $\frac{d}{dt} \hat{\mathbf{a}}_t = \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)$ . Since  $\mathbf{a}_{\text{opt}}(\theta)$  minimizes the excess risk, we have

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta) = 0, \quad \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)} = 0$$

Thus, we may simplify

$$\begin{aligned} \mathcal{R}(\hat{\mathbf{a}}; \theta) &= \frac{1}{2} (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}})^\top (\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)}) (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}}) \\ &\quad + \frac{1}{6} \nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}'} [\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)] \end{aligned}$$

We can similarly Taylor expand  $\mathbf{a}_{\text{opt}}$  to get:

$$\mathbf{a}_{\text{opt}}(\theta) = \mathbf{a}_{\text{opt}}(\theta_*) + G_{\mathbf{a}}(\theta_*)(\theta - \theta_*) + \nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'} [\theta - \theta_*, \theta - \theta_*],$$

where again we set  $\theta' = t''\theta + (1-t'')\theta_*$  for some  $t'' \in [0, 1]$ . Recall the definitions

$$\begin{aligned} \delta_* &= \arg \min_{\delta} \|M_{\mathbf{a}}(\theta_*)^{1/2} ((\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*)) - G_{\mathbf{a}}(\theta_*)\delta)\|_2 \\ &= (M_{\mathbf{a}}(\theta_*)^{1/2} G_{\mathbf{a}}(\theta_*)^\dagger M_{\mathbf{a}}(\theta_*)^{1/2} (\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*))) \\ \hat{\theta}(\hat{\mathbf{a}}) &= \theta_* + \delta_* \end{aligned}$$

Writing  $\hat{\mathbf{a}} = \mathbf{a}_{\text{opt}}(\theta_*) + G_{\mathbf{a}}(\theta_*)\delta_* + \delta_{\hat{\mathbf{a}}}$  for some  $\delta_{\hat{\mathbf{a}}}$  and denoting  $\delta_{\theta} = \theta - \theta_*$ , we then have:

$$\begin{aligned} \mathcal{R}(\hat{\mathbf{a}}; \theta) &= \frac{1}{2} (\delta_{\theta} - \delta_*)^\top G_{\mathbf{a}}(\theta_*)^\top M_{\mathbf{a}}(\theta_*) G_{\mathbf{a}}(\theta_*) (\delta_{\theta} - \delta_*) \\ &\quad + \underbrace{\frac{1}{2} \delta_{\hat{\mathbf{a}}}^\top M_{\mathbf{a}}(\theta_*) \delta_{\hat{\mathbf{a}}}}_{(a1)} - \underbrace{\delta_{\hat{\mathbf{a}}}^\top M_{\mathbf{a}}(\theta_*) (\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'} [\delta_{\theta}, \delta_{\theta}])}_{(a2)} \\ &\quad + \underbrace{\frac{1}{2} (\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'} [\delta_{\theta}, \delta_{\theta}])^\top M_{\mathbf{a}}(\theta_*) (\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'} [\delta_{\theta}, \delta_{\theta}])}_{(a3)} \\ &\quad + \underbrace{(\delta_{\theta} - \delta_*)^\top G_{\mathbf{a}}(\theta_*)^\top M_{\mathbf{a}}(\theta_*) (\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'} [\delta_{\theta}, \delta_{\theta}])}_{(a4)} - \underbrace{(\delta_{\theta} - \delta_*)^\top G_{\mathbf{a}}(\theta_*)^\top M_{\mathbf{a}}(\theta_*) \delta_{\hat{\mathbf{a}}}}_{(a5)} \\ &\quad + \underbrace{\frac{1}{2} (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}})^\top (\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)} - M_{\mathbf{a}}(\theta_*)) (\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}})}_{(a6)} \\ &\quad + \underbrace{\frac{1}{6} \nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}'} [\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta), \hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta)]}_{(a7)} \end{aligned}$$

**Controlling the Taylor Expansion through norm bounds:** We verify that we are in the regime where Assumption 3 holds.

**Claim E.1.** For a satisfying Eq. (E.2), it holds that

$$\|\theta' - \theta_\star\|_2 \leq \|\theta - \theta_\star\|_2 \leq r_{\text{quad}}(\theta_\star) \quad (\text{E.3})$$

$$\max\{\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|, \|\mathbf{a}' - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2, \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_\star)\|\} \leq L_{\mathbf{a}1} r_{\text{quad}}(\theta_\star) \quad (\text{E.4})$$

*Proof of Claim E.1.* By assumption, we have,

$$r \leq \frac{1}{4} r_{\text{quad}}(\theta_\star) \quad (\text{E.5})$$

Now recall that  $\theta' = t'\theta + (1 - t')\theta_\star$ , for some  $t' \in [0, 1]$ , so

$$\|\theta' - \theta_\star\|_2 \leq \|\theta - \theta_\star\|_2 \leq r$$

From this and trivial manipulations of  $\|\theta - \theta_\star\|_{\text{op}}$ ,  $\|\theta_0 - \theta_\star\|_{\text{op}}$ , it follows that (E.5) implies (E.3).

To verify (E.4), recall that  $\mathbf{a}' = t''\hat{\mathbf{a}} + (1 - t'')\mathbf{a}_{\text{opt}}(\theta)$  for some  $t'' \in [0, 1]$ . Hence,

$$\begin{aligned} & \max\{\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|, \|\mathbf{a}' - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2, \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_\star)\|\} \\ & \leq \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\| + \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_\star)\| \end{aligned}$$

From Eq. (E.2), it holds that  $\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\| \leq (1 + \sqrt{2})rL_{\mathbf{a}1}$ ; moreover, since  $\|\theta - \theta_\star\| \leq r_{\text{quad}}(\theta_\star)$ , the smoothness condition, Assumption 3, implies that  $\|\mathbf{a}_{\text{opt}}(\theta_\star) - \mathbf{a}_{\text{opt}}(\theta)\| \leq L_{\mathbf{a}1}\|\theta_\star - \theta\| \leq rL_{\mathbf{a}1}$ . Hence,

$$\begin{aligned} & \max\{\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|, \|\mathbf{a}' - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2, \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_\star)\|\} \\ & \leq (2 + \sqrt{2})L_{\mathbf{a}1}r \leq 4L_{\mathbf{a}1}r \end{aligned}$$

Thus, (E.5) implies (E.4) holds. □

The following bounds will be useful.

- By assumption:  $\|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2^2 \leq C_{\mathbf{a}}r$ ,

$$\|\delta_\theta\|_2^2 = \|\theta - \theta_\star\|_2^2$$

- We have that

$$\|M_{\mathbf{a}}(\theta_\star)^{1/2}G_{\mathbf{a}}(\theta_\star)\delta_\star\|_2 \leq \|M_{\mathbf{a}}(\theta_\star)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star))\|_2 \leq \|M_{\mathbf{a}}(\theta_\star)^{1/2}\|_{\text{op}}C_{\mathbf{a}}r$$

This follows since, recalling the definition of  $\delta_\star$  and letting  $U\Sigma V^\top = M_{\mathbf{a}}(\theta_\star)^{1/2}G_{\mathbf{a}}(\theta_\star)$ , we have  $\|M_{\mathbf{a}}(\theta_\star)^{1/2}G_{\mathbf{a}}(\theta_\star)\delta_\star\|_2 = \|\Sigma\Sigma^\dagger U^\top M_{\mathbf{a}}(\theta_\star)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star))\|_2$  and since  $\|\Sigma\Sigma^\dagger\|_{\text{op}} \leq 1$ .

- $\|M_{\mathbf{a}}(\theta_\star)^{1/2}\delta_{\hat{\mathbf{a}}}\|_2 \leq \|M_{\mathbf{a}}(\theta_\star)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star))\|_2 + \|M_{\mathbf{a}}(\theta_\star)^{1/2}G_{\mathbf{a}}(\theta_\star)\delta_\star\|_2 \leq 2\|M_{\mathbf{a}}(\theta_\star)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_\star))\|_2$
- By Assumption 3, so long as (E.3) holds:  $\|G_{\mathbf{a}}(\theta_\star)\|_{\text{op}} \leq L_{\mathbf{a}1}$ ,  $\|\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\theta'}\|_{\text{op}} \leq L_{\mathbf{a}2}$ .
- By Assumption 3, so long as (E.4) holds:  $\|\nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}'}\|_{\text{op}} \leq L_{\mathcal{R}3}$ .
- $\|M_{\mathbf{a}}(\theta_\star)^{1/2}\|_{\text{op}} = \sqrt{\|M_{\mathbf{a}}(\theta_\star)\|_{\text{op}}} = \sqrt{\|\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta_\star)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_\star)}\|_{\text{op}}} \leq \sqrt{L_{\mathcal{R}2}}$ . To see why the first equality holds, note that for any PSD  $M = U\Sigma U^\top$ ,  $\|M^{1/2}\|_{\text{op}} = \|\Sigma^{1/2}\|_{\text{op}} = \max_i \sqrt{\sigma_i} = \sqrt{\max_i \sigma_i} = \sqrt{\|M\|_{\text{op}}}$ .

**Lower bounding the excess risk  $\mathcal{R}(\hat{\mathbf{a}}; \theta)$ :** Throughout the remainder of the proof, we let  $c$  denote a universal numerical constant which may change from line to line. From the above observations

$$(a2) = cL_{a2}L_{\mathcal{R}2}C_a r^3$$

By the bounds given above:

$$(a4) = cL_{a2}(L_{a1} + C_a)L_{\mathcal{R}2}r^3$$

To bound (a6), we can apply Proposition E.1 to get that, when (E.4) holds,

$$\|\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)} - \nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_*)}\|_{\text{op}} \leq L_{\mathcal{R}3} \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \leq L_{\mathcal{R}3}L_{a1}r$$

Using that  $\|\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}}\|_2 \leq \|\mathbf{a}_{\text{opt}}(\theta) - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 + \|\mathbf{a}_{\text{opt}}(\theta_*) - \hat{\mathbf{a}}\|_2 \leq (L_{a1} + C_a)r$ , we have:

$$(a6) = c(L_{a1} + C_a)^2(L_{\mathcal{R}3}L_{a1} + L_{\text{hess}})r^3$$

This same bound on  $\|\mathbf{a}_{\text{opt}}(\theta) - \hat{\mathbf{a}}\|_2$  gives:

$$(a7) = cL_{\mathcal{R}3}(L_{a1} + C_a)^3 r^{3/2}$$

It remains to bound (a5). Recall that

$$\delta_* = \arg \min_{\delta} \|M_{\mathbf{a}}(\theta_*)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*) - G_{\mathbf{a}}(\theta_*)\delta)\|_2$$

so  $\delta_*$  is the projection of  $M_{\mathbf{a}}(\theta_*)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*))$  onto the image of  $M_{\mathbf{a}}(\theta_*)^{1/2}G_{\mathbf{a}}(\theta_*)$ . It follows that:

$$M_{\mathbf{a}}(\theta_*)^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*) - G_{\mathbf{a}}(\theta_*)\delta_*) = M_{\mathbf{a}}(\theta_*)^{1/2}\delta_{\hat{\mathbf{a}}} \perp \text{image}(M_{\mathbf{a}}(\theta_*)^{1/2}G_{\mathbf{a}}(\theta_*))$$

which implies

$$(a5) = -(\delta_{\theta} - \delta_*)^{\top} G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*)\delta_{\hat{\mathbf{a}}} = 0$$

Combining everything, we've shown that:

$$\mathcal{R}(\hat{\mathbf{a}}; \theta) \geq \frac{1}{2}(\delta_{\theta} - \delta_*)^{\top} G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*)G_{\mathbf{a}}(\theta_*)(\delta_{\theta} - \delta_*) + (a1) + (a3) - \mathcal{O}(C_1 r^3)$$

for

$$C_1 = 2L_{a1}L_{a2}L_{\mathcal{R}2} + 8L_{a1}^3L_{\mathcal{R}3} + 4L_{\text{hess}}$$

However,  $M_{\mathbf{a}}(\theta_*)$  is PSD so (a1), (a3)  $\geq 0$ , giving:

$$\mathcal{R}(\hat{\mathbf{a}}; \theta) \geq \frac{1}{2}(\delta_{\theta} - \delta_*)^{\top} G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*)G_{\mathbf{a}}(\theta_*)(\delta_{\theta} - \delta_*) - cC_1 r^3$$

**Completing the proof:** By definition,  $\delta_{\theta} - \delta_* = \theta - \hat{\theta}(\hat{\mathbf{a}})$  and  $G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*)G_{\mathbf{a}}(\theta_*) = \mathcal{H}(\theta_*)$ , so

$$(\delta_{\theta} - \delta_*)^{\top} G_{\mathbf{a}}(\theta_*)^{\top} M_{\mathbf{a}}(\theta_*)G_{\mathbf{a}}(\theta_*)(\delta_{\theta} - \delta_*) = \|\theta - \hat{\theta}(\hat{\mathbf{a}})\|_{\mathcal{H}(\theta_*)}^2$$

Putting things together, we then have that

$$\begin{aligned} & \min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \leq C_a r} \max_{\theta: \|\theta - \theta_*\|_2 \leq r} \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] \\ & \geq \min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \leq C_a r} \max_{\theta: \|\theta - \theta_*\|_2 \leq r} \mathbb{E} \left[ \frac{1}{2} \|\theta - \hat{\theta}(\hat{\mathbf{a}})\|_{\mathcal{H}(\theta_*)}^2 \right] - cC_1 r^3 \end{aligned}$$

Given knowledge of  $\theta_*$ ,  $\hat{\theta}(\hat{\mathbf{a}})$  is simply an estimator of  $\theta$ , so it follows that from Eq. (E.1) that

$$\min_{\hat{\mathbf{a}}: \|\hat{\mathbf{a}} - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \leq C_a r(T)} \max_{\theta: \|\theta - \theta_*\|_2 \leq r(T)} \mathbb{E} \left[ \frac{1}{2} \|\theta - \hat{\theta}(\hat{\mathbf{a}})\|_{\mathcal{H}(\theta_*)}^2 \right] \geq \min_{\hat{\theta}} \max_{\theta: \|\theta - \theta_*\|_2 \leq r(T)} \mathbb{E} \left[ \frac{1}{2} \|\theta - \hat{\theta}\|_{\mathcal{H}(\theta_*)}^2 \right]$$

This concludes the proof.  $\square$

## E.2. Proof of Theorem B.1

We apply Theorem E.3 with  $\lambda = \underline{\lambda}$ , which is greater than 0 by Assumption 4. Then,

$$\text{tr} \left( \mathcal{H}(\theta_\star) \left( \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] + \underline{\lambda} T I \right)^{-1} \right) \geq \text{tr} \left( \mathcal{H}(\theta_\star) \left( \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] + \mathbb{E}_{\theta_\star, \pi_{\text{exp}}} [\Sigma_T] \right)^{-1} \right)$$

Under Assumption 5, for any  $\theta$  satisfying  $\|\theta - \theta_\star\|_2^2 \leq 5d_\theta/(\underline{\lambda}T)^{5/6}$  and as long as  $5d_\theta/(\underline{\lambda}T)^{5/6} \leq r_{\text{cov}}(\theta_\star)^2$ , we have

$$\mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] \preceq c_{\text{cov}} \mathbb{E}_{\theta_\star, \pi_{\text{exp}}} [\Sigma_T] + \left( \frac{L_{\text{cov}}(\theta_\star, \gamma^2) \sqrt{5d_\theta} T^{7/12}}{\underline{\lambda}^{5/12}} + C_{\text{cov}} T^{1-\alpha} \right) \cdot I$$

Therefore, since our alternate instances,  $\theta \in \mathcal{B}_T$ , do satisfy  $\|\theta - \theta_\star\|_2^2 \leq 5d_\theta/(\underline{\lambda}T)^{5/6}$ , if  $T$  is large enough that

$$\frac{L_{\text{cov}}(\theta_\star, \gamma^2) \sqrt{5d_\theta} T^{7/12}}{\underline{\lambda}^{5/12}} + C_{\text{cov}} T^{1-\alpha} \leq c_{\text{cov}} T \underline{\lambda}$$

we will have, for all  $\theta \in \mathcal{B}_T$ ,

$$\begin{aligned} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] &\preceq c_{\text{cov}} \mathbb{E}_{\theta_\star, \pi_{\text{exp}}} [\Sigma_T] + \left( \frac{L_{\text{cov}}(\theta_\star, \gamma^2) \sqrt{5d_\theta} T^{7/12}}{\underline{\lambda}^{5/12}} + C_{\text{cov}} T^{1-\alpha} \right) \cdot I \\ &\preceq c_{\text{cov}} \mathbb{E}_{\theta_\star, \pi_{\text{exp}}} [\Sigma_T] + c_{\text{cov}} T \underline{\lambda} \cdot I \preceq 2c_{\text{cov}} \mathbb{E}_{\theta_\star, \pi_{\text{exp}}} [\Sigma_T] \end{aligned}$$

The result then follows from Theorem E.3 and simple manipulations.  $\square$

## E.3. Proof of Proposition E.1 and Proposition E.2

*Proof of Proposition E.1.* We prove this for a generic function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Fix some  $x, y \in \mathbb{R}^n$  and let  $x_t = tx + (1-t)y$ . Then, by Taylor's Theorem,

$$f(x) = f(y) + \frac{d}{dt} f(x_t) \Big|_{t=t'}$$

for some  $t' \in [0, 1]$ . By the chain rule,  $\frac{d}{dt} f(x_t) = \nabla_x f(x) \Big|_{x=x_t} \cdot \frac{d}{dt} x_t = \nabla_x f(x) \Big|_{x=x_t} \cdot (x - y)$ . So:

$$\|f(x) - f(y)\|_{\text{op}} \leq \|\nabla_x f(x) \Big|_{x=x_{t'}}\|_{\text{op}} \cdot \|x - y\|_{\text{op}}$$

The result follows in our setting using the norm bounds given in Assumption 3.  $\square$

*Proof of Proposition E.2.* Let  $\theta_t = t\hat{\theta} + (1-t)\theta_\star$ . Note that for any  $t$ , by Proposition E.1,

$$\|\mathbf{a}_{\text{opt}}(\theta_t) - \mathbf{a}_{\text{opt}}(\theta_\star)\|_2 \leq L_{\mathbf{a}1} \|\theta_t - \theta_\star\|_2 \leq L_{\mathbf{a}1} \|\hat{\theta} - \theta_\star\|_2 \leq L_{\mathbf{a}1} r_{\text{quad}}(\theta_\star)$$

where the last inequality follows by Assumption 3. We are therefore in the regime where the norm bounds given in Assumption 3 hold, which we will make use of throughout the proof. By Taylor's Theorem:

$$\begin{aligned} \mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}); \theta_\star) &= \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_1); \theta_\star) = \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_0); \theta_\star) + \frac{d}{dt} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) \Big|_{t=0} + \frac{1}{2} \frac{d^2}{dt^2} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) \Big|_{t=0} \\ &\quad + \frac{1}{6} \frac{d^3}{dt^3} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) \Big|_{t=t'} \end{aligned}$$

where  $t' \in [0, 1]$ . Assumption 3 gives that  $\mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_0); \theta_\star) = \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_\star); \theta_\star) = 0$ . Furthermore,  $\frac{d}{dt} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) \Big|_{t=0} = \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_\star) \Big|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_0)} \cdot \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta) \Big|_{\theta=\theta_0} \cdot \frac{d}{dt} \theta_t \Big|_{t=0}$ , but by Assumption 3,  $\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_\star) \Big|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_0)} = 0$ . Finally, by the chain rule and since  $\frac{d}{dt} \theta_t = \hat{\theta} - \theta_\star$ :

$$\frac{d^2}{dt^2} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) \Big|_{t=0} = (\hat{\theta} - \theta_\star)^\top \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_\star) \Big|_{\theta=\theta_0} (\hat{\theta} - \theta_\star) = (\hat{\theta} - \theta_\star)^\top \mathcal{H}(\theta_\star) (\hat{\theta} - \theta_\star)$$

$$\frac{d^3}{dt^3} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_\star) = \nabla_{\theta}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_\star) \Big|_{\theta=\theta_{t'}} [\hat{\theta} - \theta_\star, \hat{\theta} - \theta_\star, \hat{\theta} - \theta_\star]$$

It then follows that,

$$\left| \mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}); \theta_*) - \frac{1}{2} \|\hat{\theta} - \theta_*\|_{\mathcal{H}(\theta_*)}^2 \right| \leq \frac{1}{6} \|\nabla_{\theta}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\theta_{t'}}[\hat{\theta} - \theta_*, \hat{\theta} - \theta_*, \hat{\theta} - \theta_*]\|_{\text{op}}$$

The chain rule gives,

$$\begin{aligned} & \nabla_{\theta}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)[\hat{\theta} - \theta_*, \hat{\theta} - \theta_*, \hat{\theta} - \theta_*] \\ &= \nabla_{\mathbf{a}}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)[\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*], \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*], \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*]] \\ & \quad + 3 \nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)[\nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*, \hat{\theta} - \theta_*], \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*]] \\ & \quad + \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)[\nabla_{\theta}^3 \mathbf{a}_{\text{opt}}(\theta)[\hat{\theta} - \theta_*, \hat{\theta} - \theta_*, \hat{\theta} - \theta_*]] \end{aligned}$$

so,

$$\|\nabla_{\theta}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\theta_{t'}}[\hat{\theta} - \theta_*, \hat{\theta} - \theta_*, \hat{\theta} - \theta_*]\|_{\text{op}} \leq (L_{\mathcal{R}3} L_{\mathbf{a}1}^3 + 3L_{\mathcal{R}2} L_{\mathbf{a}2} L_{\mathbf{a}1} + L_{\mathcal{R}1} L_{\mathbf{a}3}) \|\hat{\theta} - \theta_*\|_{\text{op}}^3$$

which proves the first inequality. For the second inequality, recall that by definition,

$$\mathcal{H}(\theta_*) = \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\theta_*}$$

so, by Taylor's Theorem,

$$\mathcal{H}(\theta_*) = \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\hat{\theta}} + \frac{d}{dt} \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_*)|_{t=t'}$$

for  $t' \in [0, 1]$ . However,

$$\frac{d}{dt} \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta_t); \theta_*)|_{t=t'} = \nabla_{\theta}^3 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\theta_{t'}} \cdot \frac{d}{dt} \theta_t$$

Thus,

$$\|\mathcal{H}(\theta_*) - \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\hat{\theta}}\|_{\text{op}} \leq (L_{\mathcal{R}3} L_{\mathbf{a}1}^3 + 3L_{\mathcal{R}2} L_{\mathbf{a}2} L_{\mathbf{a}1} + L_{\mathcal{R}1} L_{\mathbf{a}3}) \|\hat{\theta} - \theta_*\|_2$$

By the chain rule,

$$\nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta') = \nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta')|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)} [\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta), \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)] + \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta')|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta)} \cdot \nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)$$

So, by Definition E.1, since  $\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta')|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta')} = 0$ , we have:

$$\begin{aligned} & \|\nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \theta_*)|_{\theta=\hat{\theta}} - \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \hat{\theta})|_{\theta=\hat{\theta}}\|_{\text{op}} \\ &= \|\nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})} [\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}, \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}] + \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})} \cdot \nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}} \\ & \quad - \nabla_{\mathbf{a}}^2 \mathcal{R}(\mathbf{a}; \hat{\theta})|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})} [\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}, \nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}]\|_{\text{op}} \\ &\leq L_{\text{hess}} \|\nabla_{\theta} \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}\|_{\text{op}}^2 \|\hat{\theta} - \theta_*\|_2 + \|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})} \cdot \nabla_{\theta}^2 \mathbf{a}_{\text{opt}}(\theta)|_{\theta=\hat{\theta}}\|_{\text{op}} \\ &\leq L_{\text{hess}} L_{\mathbf{a}1}^2 \|\hat{\theta} - \theta_*\|_2 + L_{\mathbf{a}2} \|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})}\|_{\text{op}} \end{aligned}$$

However,  $\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})}$  is Lipschitz continuous so, since  $\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_*)} = 0$ ,

$$\begin{aligned} \|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})}\|_{\text{op}} &= \|\nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\hat{\theta})} - \nabla_{\mathbf{a}} \mathcal{R}(\mathbf{a}; \theta_*)|_{\mathbf{a}=\mathbf{a}_{\text{opt}}(\theta_*)}\|_{\text{op}} \\ &\leq L_{\mathcal{R}2} \|\mathbf{a}_{\text{opt}}(\hat{\theta}) - \mathbf{a}_{\text{opt}}(\theta_*)\|_2 \\ &\leq L_{\mathcal{R}2} L_{\mathbf{a}1} \|\hat{\theta} - \theta_*\|_2 \end{aligned}$$

Since  $\mathcal{H}(\hat{\theta}) = \nabla_{\theta}^2 \mathcal{R}(\mathbf{a}_{\text{opt}}(\theta); \hat{\theta})|_{\theta=\hat{\theta}}$ , we've shown that:

$$\|\mathcal{H}(\theta_*) - \mathcal{H}(\hat{\theta})\|_{\text{op}} \leq (L_{\mathcal{R}3} L_{\mathbf{a}1}^3 + 3L_{\mathcal{R}2} L_{\mathbf{a}2} L_{\mathbf{a}1} + L_{\mathcal{R}2} L_{\mathbf{a}1} + L_{\mathcal{R}1} L_{\mathbf{a}3} + L_{\text{hess}} L_{\mathbf{a}1}^2) \|\theta_* - \hat{\theta}\|_2$$

which proves the second inequality.  $\square$

## F. Upper Bounds on Certainty Equivalence Decision Making

### F.1. Certainty Equivalence Upper Bound

In this section we assume we are in the linear dynamical system setting of Section B and that we are playing an exploration policy  $\pi_{\text{exp}}$ .

*Proof of Theorem B.3.* We define the following events.

$$\begin{aligned} \mathcal{A} &= \left\{ \mathcal{R}(\mathbf{a}_{\text{opt}}(\widehat{\theta}_{\text{ls}}); \theta_*) \leq 5\sigma_w^2 \text{tr}(\mathcal{H}(\theta_*) \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]^{-1}) \log \frac{6d_\theta}{\delta} + \frac{C_1}{T^{3/2}} + \frac{C_2}{T^{1+2\alpha}} \right\} && \text{(Good event)} \\ \mathcal{E}_1 &= \{ \lambda_{\min}(\Sigma_T) \geq \lambda T, \Sigma_T \preceq T\bar{\Gamma}_T \} && \text{(Sufficient excitation)} \\ \mathcal{E}_2 &= \{ \|\widehat{\theta}_{\text{ls}} - \theta_*\|_2 \leq r_{\text{quad}}(\theta_*) \} && \text{(Quadratic approximation regime)} \\ \mathcal{E}_3 &= \{ \|\Sigma_T - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]\|_{\text{op}} \leq \frac{C_{\text{con}}}{T^\alpha} \lambda_{\min}(\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]) \} && \text{(Concentration of covariates)} \end{aligned}$$

We would like to show that  $\mathcal{A}$  holds with high probability. The following is trivial.

$$\mathbb{P}[\mathcal{A}^c] \leq \mathbb{P}[\mathcal{A}^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] + \mathbb{P}[\mathcal{E}_1^c] + \mathbb{P}[\mathcal{E}_2^c] + \mathbb{P}[\mathcal{E}_3^c]$$

**Events  $\mathcal{E}_i$  hold with high probability:** We now show that the events  $\mathcal{E}_1, \mathcal{E}_2,$  and  $\mathcal{E}_3$  hold with high probability. Since  $\pi_{\text{exp}}$  satisfies Assumption 7, we will have  $\mathbb{P}[\mathcal{E}_1^c] \leq \delta$  and  $\mathbb{P}[\mathcal{E}_3^c] \leq \delta$  as long as

$$T \geq T_{\text{se}}(\pi_{\text{exp}}), \quad T \geq T_{\text{con}}(\pi_{\text{exp}}) \tag{F.1}$$

By Lemma F.1, on the event  $\mathcal{E}_1$ , with probability at least  $1 - \delta$ ,

$$\|\widehat{\theta}_{\text{ls}} - \theta_*\|_2 \leq C \sqrt{\frac{\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T/\lambda + I)}{\lambda T}}$$

So as long as

$$T \geq \frac{C(\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T/\lambda + I))}{\lambda r_{\text{quad}}(\theta_*)^2} \tag{F.2}$$

we will have

$$\|\widehat{\theta}_{\text{ls}} - \theta_*\|_2 \leq C \sqrt{\frac{\log(1/\delta) + d_x + \log \det(\bar{\Gamma}_T/\lambda + I)}{\lambda T}} \leq r_{\text{quad}}(\theta_*)$$

Thus,  $\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2^c] \leq \delta$ .

**Events  $\mathcal{E}_i$  imply good event holds:** We now consider the event  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ . By Proposition E.2, since  $\mathcal{R}$  satisfies Assumption 3, on this event we have

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\widehat{\theta}_{\text{ls}}); \theta_*) \leq \|\widehat{\theta}_{\text{ls}} - \theta_*\|_{\mathcal{H}(\theta_*)}^2 + L_{\text{quad}} \|\widehat{\theta}_{\text{ls}} - \theta_*\|_2^3 \leq \|\widehat{\theta}_{\text{ls}} - \theta_*\|_{\mathcal{H}(\theta_*)}^2 + C_1/T^{3/2}$$

where the last inequality follows by the bound on  $\|\widehat{\theta}_{\text{ls}} - \theta_*\|_2$  shown above for

$$C_1 := CL_{\text{quad}} \frac{(\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T/\lambda + I))^{3/2}}{\lambda^{3/2}}$$

By Theorem D.2, on the event  $\mathcal{E}_3$  and if  $T$  is large enough so that

$$C_{\text{con}}/T^\alpha < 1/4 \tag{F.3}$$

and since  $\|\Sigma_T - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]\|_{\text{op}}$ , with probability at least  $1 - \delta$ ,

$$\|\widehat{\theta}_{\text{ls}} - \theta_*\|_{\mathcal{H}(\theta_*)}^2 \leq 5\sigma_w^2 \text{tr}(\mathcal{H}(\theta_*) \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]^{-1}) \log \frac{6d_\theta}{\delta}$$

$$\begin{aligned}
 & + \frac{130\sigma_w^2 C_{\text{con}}^2}{T^{2\alpha}} \lambda_{\min}(\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]) \text{tr}(\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]^{-1}) \text{tr}(\mathcal{H}(\theta_*) \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]^{-1}) \log \frac{6d_\theta}{\delta} \\
 & \leq 5\sigma_w^2 \text{tr}(\mathcal{H}(\theta_*) \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]^{-1}) \log \frac{6d_\theta}{\delta} + \frac{260\sigma_w^2 C_{\text{con}}^2 d_\theta \text{tr}(\mathcal{H}(\theta_*))}{\lambda T^{1+2\alpha}} \log \frac{6d_\theta}{\delta}
 \end{aligned}$$

where the final inequality follows since

$$\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T] \succeq \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\mathbb{I}\{\mathcal{E}_1\} \Sigma_T] \geq \mathbb{P}[\mathcal{E}_1] \lambda T \cdot I \geq \frac{1}{2} \lambda T \cdot I$$

Thus,  $\mathbb{P}[\mathcal{E}^c \cap \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \leq \delta$  with

$$C_2 := \frac{260\sigma_w^2 C_{\text{con}}^2 d_\theta \text{tr}(\mathcal{H}(\theta_*))}{\lambda} \log \frac{6d_\theta}{\delta}$$

so it follows that  $\mathbb{P}[\mathcal{A}^c] \leq 4\delta$ . The final result then follows by rescaling  $\delta$  and so long as  $T$  is large enough that (F.1), (F.2), and (F.3) hold, which will be the case if (B.6) holds.  $\square$

## F.2. Euclidean Norm Estimation

**Lemma F.1.** Assume our data is generated according to Eq. (B.2) and let

$$\hat{\theta}_{\text{ls}} := \min_{\theta} \sum_{t=1}^T \|y_t - \theta^\top z_t\|_2^2$$

Then on the event

$$\mathcal{E} := \left\{ \lambda_{\min}(\Sigma_T) \geq \lambda T, \Sigma_T \preceq T \bar{\Gamma}_T \right\}$$

with probability at least  $1 - \delta$ :

$$\|\hat{\theta}_{\text{ls}} - \theta_*\|_2 \leq C \sqrt{\frac{\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T / \lambda + I)}{\lambda T}}.$$

*Proof.* Define the following events:

$$\begin{aligned}
 \mathcal{A} &= \left\{ \|\hat{\theta}_{\text{ls}} - \theta_*\|_2 \leq C \sqrt{\frac{\log(1/\delta) + d_\theta + \log \det(\bar{\Gamma}_T / \lambda + I)}{\lambda T}} \right\} \\
 \mathcal{E}_1 &= \left\{ \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} \leq c_2 \sigma_w \sqrt{\log \frac{1}{\delta} + d_\theta + \log \det(\bar{\Gamma}_T / \lambda + I)} \right\}
 \end{aligned}$$

Our goal is to show that  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \delta$ . The following is trivial.

$$\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \mathbb{P}[\mathcal{A}^c \cap \mathcal{E} \cap \mathcal{E}_1] + \mathbb{P}[\mathcal{E} \cap \mathcal{E}_1^c]$$

As  $\hat{\theta}_{\text{ls}}$  is the least squares estimate, we will have that  $\hat{\theta}_{\text{ls}} = (\sum_{t=1}^T z_t z_t^\top)^{-1} \sum_{t=1}^T z_t y_t = \theta_* + (\sum_{t=1}^T z_t z_t^\top)^{-1} \sum_{t=1}^T z_t w_t^\top$ . Given this, the error can be decomposed as:

$$\begin{aligned}
 \|\hat{\theta}_{\text{ls}} - \theta_*\|_2 &= \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1} \sum_{t=1}^T z_t w_t^\top \right\|_2 \leq \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \right\|_{\text{op}} \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_2 \\
 &= \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_2 / \sqrt{\lambda_{\min} \left( \sum_{t=1}^T z_t z_t^\top \right)}
 \end{aligned}$$

It follows that, on the event  $\mathcal{E} \cap \mathcal{E}_1$ , the error bound given in  $\mathcal{A}$  holds. Thus,  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E} \cap \mathcal{E}_1] = 0$ . Lemma D.6 implies that  $\mathbb{P}[\mathcal{E} \cap \mathcal{E}_1^c] \leq \delta$ , so  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \delta$ .  $\square$

## Part III

# Linear Dynamical Decision Making

## G. Notation for Linear Dynamical Systems

We next introduce notation used throughout Part III. Throughout, we consider linear dynamical systems of the form

$$x_{t+1} = A_* x_t + B_* u_t + w_t \quad (\text{G.1})$$

where  $A_* \in \mathbb{R}^{d_x \times d_x}$ ,  $B_* \in \mathbb{R}^{d_x \times d_u}$ , and  $w_t \sim \mathcal{N}(0, \sigma_w^2 I)$ . We denote  $\theta = (A, B)$  and  $\theta_* := (A_*, B_*)$ . We will sometimes break up the state into the portion driven by the input,  $x_t^u$ , and the portion driven by the noise,  $x_t^w$ . In particular, we have

$$x_{t+1}^u = A_* x_t^u + B_* u_t, \quad x_{t+1}^w = A_* x_t^w + w_t$$

Due to linearity,  $x_t = x_t^u + x_t^w$ .

### G.1. Covariance Notation

At the center of our analysis are the covariance matrices that arise from excitation of the linear system with a certain input. For an input sequence  $\mathbf{u} := (u_1, \dots, u_t) \in \mathbb{R}^{td_u}$ , we define the open loop input covariance

$$\Gamma_t^{\text{in}}(\theta, \mathbf{u}, x_0) := \sum_{s=0}^{t-1} x_s^u (x_s^u)^\top \quad \text{where} \quad x_{s+1}^u = A x_s^u + B u_s, \quad x_0^u = x_0 \quad (\text{G.2})$$

We overload notation, so that the above is also defined when  $\mathbf{u} = (u_s)_{s=1}^{t'}$  for  $t' \geq t$ , or even infinite sequences  $\mathbf{u} = (u_s)_{s \geq 1}$ . In addition, if  $\mathbf{u} = (u_s)_{s=1}^{t'}$  for  $t' < t$ , we define  $\Gamma_t^{\text{in}}(\theta, \mathbf{u}, x_0)$  to be the open loop covariance when playing  $\mathbf{u}$  periodically: that is, the input  $u_t = u_{\text{mod}(t, t')}$ . Recall that:

$$\Gamma_t^{\text{noise}}(\theta, \Lambda_u) := \sum_{s=0}^{t-1} A^s \Lambda_w (A^s)^\top + \sum_{s=0}^{t-1} A^s B \Lambda_u B^\top (A^s)^\top$$

and observe that we can equivalently define

$$\Gamma_t^{\text{noise}}(\theta, \Lambda_u) = \mathbb{E} \left[ x_t x_t^\top \mid u_s \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Lambda_u), w_s \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Lambda_w), s \leq t, x_0 = 0 \right]$$

We also define the following, which corresponds to the total expected average covariates starting from some state  $x_0$  and playing any input  $u_t = \tilde{u}_t + u_t^w$ , where  $\mathbf{u} = (\tilde{u}_t)_{t=1}^k$  and  $u_t^w \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I)$ :

$$\Gamma_T(\theta, \mathbf{u}, \sigma_u, x_0) := \frac{1}{T} \Gamma_T^{\text{in}}(\theta, \mathbf{u}, x_0) + \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \sigma_u) \quad (\text{G.3})$$

We also set:

$$\tilde{\Gamma}_t^{\text{noise}} := \begin{bmatrix} \sigma_w^2 \sum_{s=0}^{t-1} A_*^s (A_*^s)^\top + \frac{\gamma^2}{2d_u} \sum_{s=0}^{t-1} A_*^s B_* B_*^\top (A_*^s)^\top & 0 \\ 0 & \frac{\gamma^2}{2d_u} I \end{bmatrix}$$

We briefly recall the following definitions stated in Section B. We will consider the set of inputs

$$\mathcal{U}_{\gamma^2, k} := \left\{ \mathbf{U} = (U_\ell)_{\ell=1}^k : U_\ell \in \mathbb{H}_+^d, \quad \mathbf{U} \text{ is symmetric,} \quad \sum_{\ell=1}^k \text{tr}(U_\ell) \leq k^2 \gamma^2 \right\}$$

For some  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$  we define

$$\Gamma_k^{\text{freq}}(\theta, \mathbf{U}) = \frac{1}{k} \sum_{\ell=1}^k (e^{i \frac{2\pi\ell}{k}} I - A)^{-1} B U_\ell B^\text{H} (e^{i \frac{2\pi\ell}{k}} I - A)^{-\text{H}}, \quad \Gamma_{t,k}^{\text{freq}}(\theta, \mathbf{U}) = \frac{t}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U})$$



which, as we noted in Section B, correspond to the steady-state covariates when the input  $U$  is played periodically. Finally, we set:

$$\Gamma_{T,k}^{\text{ss}}(\theta, U, \sigma_u) = \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, U) + \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \sigma_u)$$

which correspond to the expected steady-state covariates of the noisy system when playing inputs  $U$ .

## G.2. Lifted Dynamical System

We will set

$$\tilde{\theta} := (\tilde{A}, \tilde{B}), \quad \tilde{A} := \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}, \quad \tilde{B} := \begin{bmatrix} 0 \\ I \end{bmatrix} \quad (\text{G.4})$$

and in particular  $\tilde{\theta}_\star := (\tilde{A}_\star, \tilde{B}_\star)$ . Then consider the dynamical system:

$$z_{t+1} = \tilde{A}_\star z_t + \tilde{B}_\star u_{t+1} + \begin{bmatrix} w_t \\ 0 \end{bmatrix} \quad (\text{G.5})$$

We note that  $z_t = [x_t; u_t]$ , where  $x_t$  is the state of (G.1). It follows that

$$\Sigma_T := \sum_{t=1}^T z_t z_t^\top = \sum_{t=1}^T \begin{bmatrix} x_t \\ u_t \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix}^\top$$

so a bound on the covariates of the system (G.5) can be directly applied to the state-input covariates from (G.1). For subsequent results, we will use  $z_t := [x_t; u_t]$ .

## G.3. Linear Dynamical Systems as Vector Regression

We can write the system (G.1) in the form

$$y_s = \langle \phi_\star, v_s \rangle + \eta_s \quad (\text{G.6})$$

To obtain this mapping, we reindex time: for a fixed  $t$  of (G.1), define  $s = (d_x + d_u)t + i$  for some  $i \in \{1, \dots, d_x\}$ . Furthermore, we set  $\phi_\star = [A_{\star,1}; B_{\star,1}; \dots; A_{\star,d_x}; B_{\star,d_x}] \in \mathbb{R}^{d_x^2 + d_x d_u}$ , where  $A_{\star,j}, B_{\star,j}$  denote the  $j$ th row of  $A_\star$  and  $B_\star$ , respectively,  $\eta_s = [w_t]_i$ , and  $v_s = [\mathbf{0}, \dots, \mathbf{0}, x_t, u_t, \mathbf{0}, \dots, \mathbf{0}]$ , where  $x_t$  starts at index  $(d_x + d_u)(i - 1) + 1$ . With these definitions we will have  $y_s = [x_{t+1}]_i$ . It follows that if we run (G.6) from time  $s = 1$  to  $s = (d_x + d_u)T + d_x$  the set of observations obtained will be identical to those obtained from  $x_{t+1} = A_\star x_t + B_\star u_t + w_t$ . Thus, (G.6) is simply a vectorization of  $x_{t+1} = A_\star x_t + B_\star u_t + w_t$ . It is easy to see that, if  $\hat{\phi}$  denotes the least squares estimate of  $\phi_\star$  obtained from observations of (G.6) and  $\hat{A}, \hat{B}$  denote the least squares estimates of  $A_\star, B_\star$  obtained from observations of (G.1), we will have  $\hat{\phi} = [\hat{A}_1; \hat{B}_1; \dots; \hat{A}_{d_x}; \hat{B}_{d_x}]$ . Furthermore,<sup>3</sup>

$$\Sigma_T := \sum_{s=1}^{(d_x+d_u)T+d_x} z_s z_s^\top = I_{d_x} \otimes \sum_{t=1}^T x_t x_t^\top = I \otimes \Sigma_T$$

Thus,  $\mathbb{E} \Sigma_T = I_{d_x} \otimes \mathbb{E} \sum_{t=1}^T x_t x_t^\top$ ,  $\|\Sigma_T - \mathbb{E} \Sigma_T\|_{\text{op}} = \|\sum_{t=1}^T x_t x_t^\top - \mathbb{E} \sum_{t=1}^T x_t x_t^\top\|_{\text{op}}$ , and  $\lambda_{\min}(\mathbb{E} \sum_{t=1}^T x_t x_t^\top) = \lambda_{\min}(\mathbb{E} \Sigma_T)$ . This equivalence allows us to apply results from Section D and Section E in the dynamical system setting.

## G.4. Key Parameters in the Analysis

For any  $\theta = (A, B)$ , the  $\mathcal{H}_\infty$  norm of  $\theta$  is defined as:

$$\|\theta\|_{\mathcal{H}_\infty} := \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-1} B\|_{\text{op}}$$

<sup>3</sup>Note that we change notation slightly here. Previously  $\Sigma_T$  denoted the set of covariates in the general regression setting after  $T$  steps, while here  $\Sigma_T$  is the set of covariates after  $(d_x + d_u)T + d_x$  steps in the general regression setting, but corresponds to running our linear dynamical system for  $T$  steps. As subsequent results are concerned with the time scale of the linear dynamical system, this change in notation will simply further analysis.

To control the transient behavior, let:

$$\tau(A, \rho) := \sup\{\|A^k\|_{\text{op}}\rho^{-k} : k \geq 0\}$$

$\tau(A, \rho)$  is the smallest value such that  $\|A^k\|_{\text{op}} \leq \tau(A, \rho)\rho^k$  for all  $k$ . We will define

$$\rho_\star := \max\left\{\frac{1}{2}, \frac{2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}}{1 + 2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}}\right\}$$

and  $\tau_\star := \tau(\tilde{A}_\star, \rho_\star)$ . The following result relates  $\tau_\star$ ,  $\frac{1}{1-\rho_\star}$ , and  $\|\tilde{\theta}_\star\|_{\mathcal{H}_\infty}$  to  $\|A_\star\|_{\mathcal{H}_\infty}$  and  $\|B_\star\|_{\text{op}}$ , which will aid in simplifying our results.

**Lemma G.1.** *The following upper bounds hold:*

$$\frac{1}{1-\rho_\star} \leq 2 + 2\|A_\star\|_{\mathcal{H}_\infty}^2, \quad \tau_\star \leq 2(1 + 2\|B_\star\|_{\text{op}})\|A_\star\|_{\mathcal{H}_\infty}, \quad \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty} \leq 1 + (1 + \|B_\star\|_{\text{op}})\|A_\star\|_{\mathcal{H}_\infty}.$$

In addition, we can relate the value of  $\tau$  for a lifted system  $\tilde{\theta}$  to the original system  $\theta$ .

**Lemma G.2.** *Let  $\tilde{A}$  be defined as in (G.4). Then  $\tau(\tilde{A}, \rho) \leq (1 + \rho^{-1}\|B\|_{\text{op}})\tau(A, \rho)$ .*

We introduce the following constants to control the smoothness of the covariates:

$$r_{\text{cov}}(\theta_\star) := \min\left\{\frac{1-\rho_\star}{2\tau_\star}, \frac{1}{2\|A_\star\|_{\mathcal{H}_\infty}}, 1\right\}$$

$$L_{\text{cov}}(\theta_\star, \gamma^2) := \frac{8(\sigma_w^2 + \sigma_u^2\|B_\star\|_{\text{op}}^2)\tau_\star^3}{(1-\rho_\star^2)^2} + \frac{4\sigma_u^2(\|B_\star\|_{\text{op}} + 1)\tau_\star}{1-\rho_\star^2} + 34\gamma^2\|A_\star\|_{\mathcal{H}_\infty}^3(\|B_\star\|_{\text{op}} + 1)^2$$

Lemma H.5 implies that, if

$$\|\theta - \theta_\star\|_{\text{op}} \leq r_{\text{cov}}(\theta_\star)$$

then for any  $u \in \mathcal{U}_{\gamma^2, k}$ , if  $T_2$  is divisible by  $k$ ,

$$\|\mathbf{\Gamma}_{T_1, T_2}^{\text{ss}}(\theta, u, \sigma_u) - \mathbf{\Gamma}_{T_1, T_2}^{\text{ss}}(\theta_\star, u, \sigma_u)\|_{\text{op}} \leq L_{\text{cov}}(\theta_\star, \gamma^2) \cdot \|\theta - \theta_\star\|_{\text{op}}.$$

This holds regardless of the loss  $\mathcal{R}$ .

Finally, in our analysis it will be convenient to work with a slightly different definition of the optimal risk, which we define as:

$$\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star) := \liminf_{T \rightarrow \infty} \min_{u \in \mathcal{U}_{\gamma^2, T}} \text{tr}\left(\mathcal{H}(\theta_\star)\mathbf{\Gamma}_{T, T}^{\text{ss}}(\tilde{\theta}_\star, u, 0)^{-1}\right)$$

As the following result shows,  $\Phi_{\text{opt}}$  and  $\Phi_{\text{opt}}^{\text{ss}}$  are equal up to absolute constants.

**Lemma G.3.**  *$\Phi_{\text{opt}}(\gamma^2; \theta_\star)$  and  $\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star)$  are equal up to constants:*

$$\frac{1}{4}\Phi_{\text{opt}}(\gamma^2; \theta_\star) \leq \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star) \leq 16\Phi_{\text{opt}}(\gamma^2; \theta_\star).$$

## G.5. Linear Dynamical Systems Notation Proofs

*Proof of Lemma G.4.* We have that

$$\begin{aligned} \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty} &= \max_{\omega \in [0, 2\pi]} \left\| \left( e^{i\omega} I - \begin{bmatrix} A_\star & B_\star \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix} \right\|_{\text{op}} \\ &\leq \max_{\omega \in [0, 2\pi]} \left\| \left( e^{i\omega} I - \begin{bmatrix} A_\star & B_\star \\ 0 & 0 \end{bmatrix} \right)^{-1} \right\|_{\text{op}} \end{aligned}$$

For each  $\omega$ , set  $A(\omega) := (e^{\iota\omega}I - A_\star)^{-1}$ . Then, using the block matrix inverse formula,

$$\left( e^{\iota\omega}I - \begin{bmatrix} A_\star & B_\star \\ 0 & 0 \end{bmatrix} \right)^{-1} = \begin{bmatrix} A(\omega) & -A(\omega)e^{-\iota\omega}B_\star \\ 0 & e^{-\iota\omega}I \end{bmatrix}$$

Thus,

$$\begin{aligned} \max_{\omega \in [0, 2\pi]} \left\| \left( e^{\iota\omega}I - \begin{bmatrix} A_\star & B_\star \\ 0 & 0 \end{bmatrix} \right)^{-1} \right\|_{\text{op}} &\leq \max_{\omega \in [0, 2\pi]} \left\| \begin{bmatrix} A(\omega) & -A(\omega)e^{-\iota\omega}B_\star \\ 0 & e^{-\iota\omega}I \end{bmatrix} \right\| \\ &\leq 1 + (1 + \|B_\star\|_{\text{op}}) \max_{\omega \in [0, 2\pi]} \|A(\omega)\|_{\text{op}} \\ &= 1 + (1 + \|B_\star\|_{\text{op}}) \|A_\star\|_{\mathcal{H}_\infty}. \end{aligned}$$

In the case of scalar  $A_\star$ , Lemma 4.1 (Tu et al., 2017) shows that  $\|A_\star^k\|_{\text{op}} \leq \|\frac{1}{\rho}A_\star\|_{\mathcal{H}_\infty}\rho^k$ . In the case when  $d_x > 1$ , we can apply their proof to the sequence  $u^\top A_\star^k v$  for some  $u, v$  with  $\|u\|_2 = \|v\|_2 = 1$ . Doing so, we obtain

$$u^\top A_\star^k v \leq \|\frac{1}{\rho}A_\star\|_{\mathcal{H}_\infty}\rho^k$$

As this holds for all  $u$  and  $v$ , we have  $\|A_\star^k\|_{\text{op}} \leq \|\frac{1}{\rho}A_\star\|_{\mathcal{H}_\infty}\rho^k$ . As  $\tau(A_\star, \rho)$  is the smallest value satisfying  $\|A_\star^k\|_{\text{op}} \leq \tau(A_\star, \rho)\rho^k$  for all  $k$ , it follows that  $\tau(A_\star, \rho) \leq \|\frac{1}{\rho}A_\star\|_{\mathcal{H}_\infty}$ . We next wish to upper bound  $\|\frac{1}{\rho}A_\star\|_{\mathcal{H}_\infty}$  by  $\|A_\star\|_{\mathcal{H}_\infty}$  for some choice of  $\rho$ . Lemma F.9 of (Wagenmaker & Jamieson, 2020) gives that

$$\|A_\star - \frac{1}{\rho}A_\star\|_{\text{op}} \leq \frac{1}{2\|A_\star\|_{\mathcal{H}_\infty}} \quad \text{implies} \quad \|\frac{1}{\rho}A_\star\|_{\text{op}} \leq 2\|A_\star\|_{\mathcal{H}_\infty}$$

A sufficient condition to meet this is

$$\rho \geq \frac{2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}}{1 + 2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}}$$

As  $\rho_\star$  satisfies this, it follows that  $\tau(A_\star, \rho_\star) \leq \|\frac{1}{\rho_\star}A_\star\|_{\mathcal{H}_\infty} \leq 2\|A_\star\|_{\mathcal{H}_\infty}$ . Combining this with Lemma G.2, we conclude that

$$\tau_\star \leq (1 + \rho_\star^{-1}\|B_\star\|_{\text{op}})\tau(A_\star, \rho_\star) \leq 2(1 + \rho_\star^{-1}\|B_\star\|_{\text{op}})\|A_\star\|_{\mathcal{H}_\infty} \leq 2(1 + 2\|B_\star\|_{\text{op}})\|A_\star\|_{\mathcal{H}_\infty}$$

Finally, by definition of  $\rho_\star$  it follows

$$\frac{1}{1 - \rho_\star} \leq \max\{1 + 2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}, 2\} \leq 2 + 2\|A_\star\|_{\mathcal{H}_\infty}\|A_\star\|_{\text{op}}$$

We then upper bound  $\|A_\star\|_{\text{op}} \leq \|A_\star\|_{\mathcal{H}_\infty}$  to obtain the final result.  $\square$

*Proof of Lemma G.2.* Note that:

$$\tilde{A}^k = \begin{bmatrix} A^k & A^{k-1}B \\ 0 & 0 \end{bmatrix}$$

Thus,

$$\|\tilde{A}^k\|_{\text{op}} = \sup_{v \in \mathcal{S}^{d+p-1}} \left\| \begin{bmatrix} A^k & A^{k-1}B \\ 0 & 0 \end{bmatrix} v \right\|_{\text{op}} = \sup_{v \in \mathcal{S}^{d+p-1}} \|A^k v_1 + A^{k-1}B v_2\|_{\text{op}} \leq \|A^k\|_{\text{op}} + \|A^{k-1}B\|_{\text{op}}$$

so, for any  $\rho > 0$ ,

$$\|\tilde{A}^k\|_{\text{op}}\rho^{-k} \leq \|A^k\|_{\text{op}}\rho^{-k} + \|B\|_{\text{op}}\|A^{k-1}\|_{\text{op}}\rho^{-k} \leq \tau(A, \rho) + \rho^{-1}\|B\|_{\text{op}}\|A^{k-1}\|_{\text{op}}\rho^{-(k-1)} \leq (1 + \rho^{-1}\|B\|_{\text{op}})\tau(A, \rho)$$

$\square$

*Proof of Proposition B.4.* 1. Follows by Parseval's Theorem and simple manipulations. For 2., take some  $\ell$  such that  $\frac{\ell}{k'} = \frac{n}{k}$  for some integer  $n$ . Then,

$$\tilde{u}'_\ell = \sum_{s=1}^{k'} u_s e^{-i\frac{2\pi\ell s}{k'}} = \sum_{s=1}^{k'} u_s e^{-i\frac{2\pi n s}{k}} = \sum_{s=1}^{k'} \tilde{u}_{\text{mod}(s,k)} e^{-i\frac{2\pi n \text{mod}(s,k)}{k}} = \frac{k'}{k} \sum_{s=1}^k \tilde{u}_s e^{-i\frac{2\pi n s}{k}} = \frac{k'}{k} \tilde{u}_n$$

Furthermore, if  $\frac{\ell}{k'} \neq \frac{n}{k}$  for all integers  $n$ , we will have

$$\tilde{u}'_\ell = \sum_{s=1}^{k'} u_s e^{-i\frac{2\pi\ell s}{k'}} = \sum_{r=1}^k \tilde{u}_r \sum_{s=0}^{k'/k-1} e^{-i\frac{2\pi\ell(ks+r)}{k'}} = \sum_{r=1}^k \tilde{u}_r e^{-i\frac{2\pi\ell r}{k'}} \sum_{s=0}^{k'/k-1} e^{-i\frac{2\pi\ell ks}{k'}} = 0$$

Plugging this into the expression for  $\Gamma_{k'}^{\text{freq}}(\theta, \mathbf{u}')$ , the conclusion follows.  $\square$

*Proof of Proposition B.5.* Fix some  $m$  and  $j \in [d_u]$  and consider the segment of  $\mathbf{u}_m$ ,  $\mathbf{u}_m^j := (u_t)_{t=(j-1)km+1}^{jkm}$ . By construction, this is a signal with period  $k$ . Assume we play this input starting from some state  $x_0$  not necessarily equal to 0. Let  $x_t^{\mathbf{u}_m^j}$  denote the response generated on the noiseless system. By Parseval's Theorem and Proposition B.4, it follows that

$$\frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{u}_1^j) = \lim_{m \rightarrow \infty} \frac{1}{km} \sum_{t=0}^{km} x_t^{\mathbf{u}_m^j} (x_t^{\mathbf{u}_m^j})^\top$$

Furthermore, by the construction of  $\mathbf{u}_1^j$  given in `ConstructTimeInput`, we have

$$\frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{u}_1^j) = \frac{d_u}{k^2} \sum_{\ell=1}^k \lambda_{\ell,j} (e^{i\frac{2\pi\ell}{k}} I - A)^{-1} B v_{\ell,j} v_{\ell,j}^H B^H (e^{i\frac{2\pi\ell}{k}} I - A)^{-H}$$

Now note that, if we play the entire sequence of inputs  $\mathbf{u}_m$ , we will have

$$\sum_{t=0}^{d_u km} x_t^{\mathbf{u}_m} (x_t^{\mathbf{u}_m})^\top = \sum_{j=1}^{d_u} \sum_{t=(j-1)km+1}^{jkm} x_t^{\mathbf{u}_m^j} (x_t^{\mathbf{u}_m^j})^\top$$

where the starting state,  $x_{(j-1)km+1}^{\mathbf{u}_m^j}$ , is equal to the final state produced when playing the previous input,  $x_{(j-1)km}^{\mathbf{u}_m^{j-1}}$ . Note that, as we assume the system is stable and the input has bounded energy and is of period  $k$ , the norm of  $x_{(j-1)km}^{\mathbf{u}_m^{j-1}}$  will scale sublinearly  $m$  (see Appendix I.3). It follows that,

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{d_u mk} \sum_{t=0}^{d_u mk} x_t^{\mathbf{u}_m} (x_t^{\mathbf{u}_m})^\top &= \lim_{m \rightarrow \infty} \frac{1}{d_u mk} \sum_{j=1}^{d_u} \sum_{t=(j-1)km+1}^{jkm} x_t^{\mathbf{u}_m^j} (x_t^{\mathbf{u}_m^j})^\top \\ &= \frac{1}{d_u} \sum_{j=1}^{d_u} \lim_{m \rightarrow \infty} \frac{1}{mk} \sum_{t=(j-1)km+1}^{jkm} x_t^{\mathbf{u}_m^j} (x_t^{\mathbf{u}_m^j})^\top \\ &= \frac{1}{d_u} \sum_{j=1}^{d_u} \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{u}_1^j) \\ &= \frac{1}{k^2} \sum_{\ell=1}^k \sum_{j=1}^{d_u} (e^{i\frac{2\pi\ell}{k}} I - A)^{-1} B (\lambda_{\ell,j} v_{\ell,j} v_{\ell,j}^H) B^H (e^{i\frac{2\pi\ell}{k}} I - A)^{-H} \\ &= \frac{1}{k^2} \sum_{\ell=1}^k (e^{i\frac{2\pi\ell}{k}} I - A)^{-1} B U_\ell B^H (e^{i\frac{2\pi\ell}{k}} I - A)^{-H} \\ &= \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) \end{aligned}$$

where the second to last inequality follows by the definition of  $\lambda_{\ell,j}, v_{\ell,j}$  given in `ConstructTimeInput`. To see that the power constraint holds, note that, by Parseval's Theorem and the construction of the input,

$$\sum_{t=1}^{d_u m k} u_t^\top u_t = \sum_{j=1}^{d_u} \sum_{t=(j-1)km+1}^{jkm} u_t^\top u_t = \sum_{j=1}^{d_u} \frac{m}{k} \sum_{\ell=1}^k d_u \lambda_{j,\ell} v_{j,\ell}^\top v_{j,\ell} = \sum_{j=1}^{d_u} \frac{m}{k} \sum_{\ell=1}^k d_u \text{tr}(\lambda_{j,\ell} v_{j,\ell} v_{j,\ell}^\top) = \frac{m d_u}{k} \sum_{\ell=1}^k \text{tr}(U_\ell) \leq m d_u k \gamma^2$$

where the final inequality holds since  $U \in \mathcal{U}_{\gamma^2, k}$ .  $\square$

*Proof of Lemma G.3.* Fix some  $T$ . In the proof of Theorem B.2 we showed that, for some  $\theta_0$  satisfying  $\|\theta_\star - \theta_0\|_F^2 \leq 5(d_x^2 + d_x d_u)/(\lambda_{\min, \infty}^* T^{5/6})$ ,

$$\min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_0) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I)^{-1} \right) \geq \frac{1}{16T} \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star) - \mathcal{O} \left( \frac{1}{T^{17/12}} \right)$$

Following the proof of Theorem B.2, we can use Proposition E.2 to show that

$$\begin{aligned} \min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_0) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I)^{-1} \right) &\leq \min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_\star) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I)^{-1} \right) + \mathcal{O} \left( \frac{1}{T^{17/12}} \right) \\ &\leq \min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_\star) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T])^{-1} \right) + \mathcal{O} \left( \frac{1}{T^{17/12}} \right) \end{aligned}$$

Renormalizing by  $T$ , it follows that for any  $T$

$$\inf_{\pi \in \Pi_{\gamma^2}} \Phi_T(\theta_\star; \pi) = \min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_\star) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T/T])^{-1} \right) \geq \frac{1}{16} \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star) - \mathcal{O} \left( \frac{1}{T^{5/12}} \right)$$

Taking  $\liminf_{T \rightarrow \infty}$  of both sides proves the first inequality.

For the second inequality, some trivial manipulations of (J.8) in the proof of Lemma J.7 shows that, for sufficiently large  $T$ ,

$$\min_{\pi \in \Pi_{\gamma^2}} \text{tr} \left( \mathcal{H}(\theta_0) (\mathbb{E}_{\theta_\star, \pi} [\Sigma_T])^{-1} \right) \leq \frac{4}{T} \min_{u \in \mathcal{U}_{\gamma^2, T}} \text{tr} \left( \mathcal{H}(\theta_\star) \Gamma_{T, T}^{\text{ss}}(\tilde{\theta}_\star, u, 0)^{-1} \right)$$

Renormalizing by  $T$  and taking  $\liminf_{T \rightarrow \infty}$  of both sides gives the result.  $\square$

## H. Lower Bounds in Linear Dynamical Systems

### H.1. Regular Policies in Linear Dynamical Systems

The following result shows that Assumption 4 and 5 are met if we assume the dynamics are linear, and that  $\pi_{\text{exp}}$  satisfies a certain regularity condition. This implies that Theorem B.1 holds for a fairly general set of policies in linear dynamical systems.

**Lemma H.1.** *Assume that  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  plays input  $u_t \sim \mathcal{N}(\tilde{u}_t, \sigma_u^2 I)$  where  $\tilde{u}_t$  and  $\sigma_u$  are chosen deterministically at time 0, and  $\tilde{u}_t$  is periodic with period  $k$ . Then, if our dynamics are*

$$x_{t+1} = A_\star x_t + B_\star u_t + w_t$$

$\pi_{\text{exp}}$  and  $\theta_\star$  satisfy Assumption 4 and 5 with  $\alpha = 1$ ,  $c_{\text{cov}} = 1$ , and

$$\underline{\lambda} = \min\{\sigma_w^2, \sigma_u^2\}, \quad L_{\text{cov}}(\theta_\star; \gamma^2) = C_{\text{sys}}(\sigma_w^2 + \gamma^2), \quad r_{\text{cov}}(\theta_\star) = C_{\text{sys}}^{-1}, \quad C_{\text{cov}} = C_{\text{sys}} \gamma^2 (\sqrt{T}k + k^2)$$

for some constant  $C_{\text{sys}} = \text{poly}(\|B_\star\|_{\text{op}}, \|A_\star\|_{\mathcal{H}_\infty})$  and with  $d_\theta = d_x^2 + d_x d_u$ .

*Proof.* That this policy satisfies Assumption 4 with  $\underline{\lambda} = \min\{\sigma_w^2, \sigma_u^2\}$  is trivial.

To see that Assumption 5 is satisfied, fix some  $\theta$  and denote  $\tilde{u} = (\tilde{u}_t)_{t=1}^k$ . Then, by definition,

$$\frac{1}{T} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] = \frac{1}{T} \Gamma_T^{\text{in}}(\tilde{\theta}, \tilde{u}, 0) + \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, \sigma_u)$$

Note that since  $\pi_{\text{exp}} \in \Pi_{\gamma^2}$  and  $\tilde{u}_t$  is periodic we have  $\sum_{t=0}^{k-1} \tilde{u}_t^\top \tilde{u}_t \leq k\gamma^2$ . By Lemma H.4, it follows that

$$\left\| \frac{1}{T} \Gamma_T^{\text{in}}(\tilde{\theta}, \tilde{\mathbf{u}}, 0) - \frac{1}{k} \Gamma_k^{\text{freq}}(\tilde{\theta}, \tilde{\mathbf{u}}) \right\|_{\text{op}} \leq \frac{\tau(\tilde{A}, \rho) \|\tilde{\theta}\|_{\mathcal{H}_\infty}^2 \sqrt{T+1} k \gamma^2}{(1-\rho^k)T} + \frac{\tau(\tilde{A}, \rho)^2 \|\tilde{\theta}\|_{\mathcal{H}_\infty}^2 k^2 \gamma^2}{(1-\rho^k)^2 T} =: \Delta(\theta)$$

which implies that

$$\left\| \frac{1}{T} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] - \Gamma_{T,k}^{\text{ss}}(\tilde{\theta}, \tilde{\mathbf{u}}, \sigma_u) \right\|_{\text{op}} \leq \Delta(\theta)$$

Applying this same bound to  $\theta_*$ , the triangle inequality gives

$$\left\| \frac{1}{T} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] - \frac{1}{T} \mathbb{E}_{\theta_*, \pi_{\text{exp}}} [\Sigma_T] \right\|_{\text{op}} \leq \|\Gamma_{T,k}^{\text{ss}}(\tilde{\theta}, \tilde{\mathbf{u}}, \sigma_u) - \Gamma_{T,k}^{\text{ss}}(\tilde{\theta}_*, \tilde{\mathbf{u}}, \sigma_u)\|_{\text{op}} + \Delta(\theta) + \Delta(\theta_*)$$

Lemma H.5 gives that, as long as  $\|\theta - \theta_*\|_{\text{op}} \leq r_{\text{cov}}(\tilde{\theta}_*)/2$ ,

$$\|\Gamma_{T,k}^{\text{ss}}(\tilde{\theta}, \tilde{\mathbf{u}}, \Lambda_u) - \Gamma_{T,k}^{\text{ss}}(\tilde{\theta}_*, \tilde{\mathbf{u}}, \Lambda_u)\|_{\text{op}} \leq 2L_{\text{cov}}(\tilde{\theta}_*, \gamma^2) \|\theta - \theta_*\|_{\text{op}}$$

where  $r_{\text{cov}}(\tilde{\theta}_*)$  and  $L_{\text{cov}}(\tilde{\theta}_*)$  are defined here as in Lemma H.5. Note here that we use that  $\|\theta - \theta_*\|_{\text{op}}$  and  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}}$  are within a factor of 2 of each other since

$$\max\{\|A - A_*\|_{\text{op}}, \|B - B_*\|_{\text{op}}\} \leq \|\theta - \theta_*\|_{\text{op}}, \|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq \|A - A_*\|_{\text{op}} + \|B - B_*\|_{\text{op}}$$

It remains to simplify  $\Delta(\theta)$ . By Lemma F.9 of (Wagenmaker & Jamieson, 2020), as long as  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq c/\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}$ , we will have that  $\|\tilde{\theta}\|_{\mathcal{H}_\infty}$  and  $\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}$  are within a constant factor of each other. Next, note that Lemma H.1 implies that, so long as  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq \epsilon$ ,  $\|\tilde{A}^k\|_{\text{op}} \leq \tau(\tilde{A}_*, \rho)(\rho + \tau(\tilde{A}_*, \rho)\epsilon)^k$ . This implies that

$$\tau(\tilde{A}, \rho + \tau(\tilde{A}_*, \rho)\epsilon) = \sup_k \|\tilde{A}^k\|_{\text{op}} (\rho + \tau(\tilde{A}_*, \rho)\epsilon)^{-k} \leq \tau(\tilde{A}_*, \rho)$$

As long as  $\epsilon < (1 - \rho_*)/(2\tau(\tilde{A}_*, \rho))$  we can then choose  $\rho = \rho_* + \tau(\tilde{A}_*, \rho_*)\epsilon$  which will allow us to upper bound

$$\frac{\tau(\tilde{A}, \rho)}{(1-\rho^k)} \leq \frac{c\tau(\tilde{A}_*, \rho_*)}{(1-\rho_*^k)}$$

It follows that  $\Delta(\theta) \leq c\Delta(\theta_*)$ . By Lemma G.4 and some algebra, the assumptions then hold with

$$\begin{aligned} \lambda &= \min\{\sigma_w^2, \sigma_u^2\}, \quad L_{\text{cov}}(\theta_*; \gamma^2) = \frac{c_1(\sigma_w^2 + \sigma_u^2)\tau_*^3}{(1-\rho_*)^2} + c_2\gamma^2(1 + \|B_*\|_{\text{op}})^3 \|A_*\|_{\mathcal{H}_\infty}^3, \\ r_{\text{cov}}(\theta_*) &= c_3 \min\left\{ \frac{1-\rho_*}{\tau_*}, \frac{1/2}{1 + (1 + \|B_*\|_{\text{op}})\|A_*\|_{\mathcal{H}_\infty}}, 1 \right\}, \\ C_{\text{cov}} &= \frac{c_4\tau_*(1 + \|B_*\|_{\text{op}})^2 \|A_*\|_{\mathcal{H}_\infty}^2 \sqrt{T} k \gamma^2}{1-\rho_*^k} + \frac{c_5\tau_*^2(1 + \|B_*\|_{\text{op}})^2 \|A_*\|_{\mathcal{H}_\infty}^2 k^2 \gamma^2}{(1-\rho_*^k)^2}, \quad \alpha = 1, \quad c_{\text{cov}} = 1 \end{aligned}$$

For the final statement we simplify all expressions involving problem-dependent constants by simply upper bounding them by constants that are poly( $\|A_*\|_{\mathcal{H}_\infty}, \|B_*\|_{\text{op}}$ ), and noting that  $\sigma_u^2 \leq \gamma^2$ .  $\square$

## H.2. Proof of Theorem B.2

*Proof.* The outline of the proof is as follows.

1. Apply Theorem E.3 to show that

$$\min_{\hat{\mathbf{a}}} \max_{\theta \in \mathcal{B}_T} \mathbb{E}_{\theta, \pi_{\text{exp}}} [\mathcal{R}(\hat{\mathbf{a}}_T; \theta)] \geq \min_{\theta: \|\theta - \theta_0\|_F^2 \leq 5(d_x^2 + d_x d_u)/(\lambda T^{5/6})} \mathbb{E} \left[ \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbb{E}_{\theta, \pi_{\text{exp}}} [\Sigma_T] + \lambda T \cdot I \right)^{-1} \right) \right]$$

for a particular choice of  $\lambda$ .

2. Apply Lemma H.3 to show that, for any policy  $\pi_{\text{exp}}$  and any  $\theta$ , there exists a *periodic* policy  $\pi'_{\text{exp}}$  such that

$$\mathbb{E}_{\theta, \pi_{\text{exp}}}[\Sigma_T] \preceq \mathbb{E}_{\theta, \pi'_{\text{exp}}}[\Sigma_{c_1 T}] + c_2$$

3. Given that  $\pi'_{\text{exp}}$  is periodic, apply Lemma H.4 to show that we can upper bound the expected covariates by the expected *steady-state* covariates:

$$\mathbb{E}_{\theta, \pi'_{\text{exp}}}[\Sigma_{c_1 T}] \preceq \mathbb{E}_{\theta, \pi'_{\text{exp}}}[\Gamma_{c_1 T}^{\text{freq}}(\tilde{\theta}, \mathbf{U})] + c_3$$

4. Use the frequency-domain representation to show that there exists a non-random input  $\mathbf{U}'$  that meets the power constraint and achieves the same steady state covariates:

$$\mathbb{E}_{\theta, \pi'_{\text{exp}}}[\Gamma_{c_1 T}^{\text{freq}}(\tilde{\theta}, \mathbf{U})] = \Gamma_{c_1 T}^{\text{freq}}(\tilde{\theta}, \mathbf{U}')$$

5. Apply the perturbation bound for the steady state covariates given in Lemma H.5 to show that, for any  $\theta$  in our set, we can upper bound the covariates on  $\theta$  by the covariates on  $\theta_*$ :

$$\Gamma_{c_1 T}^{\text{freq}}(\tilde{\theta}, \mathbf{U}') \preceq \Gamma_{c_1 T}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}') + c_4$$

6. Finally, we conclude the proof by optimizing over  $\pi'_{\text{exp}}$  to obtain a lower bound scaling as  $\Phi_{\text{opt}}(\gamma^2; \theta_*)$ .

Throughout the proof, we assume expectations are taken with respect to  $\theta$  and  $\pi_{\text{exp}}$ , and therefore write  $\mathbb{E}[\cdot]$  in place of  $\mathbb{E}_{\theta, \pi_{\text{exp}}}[\cdot]$ . As stated in Section G.3, linear dynamical systems are simply an instance of vector regression and we can therefore apply the results of Section E in this setting.

**Applying Theorem E.3:** Define

$$\lambda_{\min, \infty}^* := \limsup_{T' \rightarrow \infty} \max_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \frac{1}{10T'} \lambda_{\min}(\Gamma_{T'}^{\text{SS}}(\tilde{\theta}_*, \mathbf{U}, 0))$$

Under Assumption 6 and by Lemma H.2, we will have that  $\lambda_{\min, \infty}^* > 0$ . Then the first conclusion of Theorem E.3 holds with  $\lambda = \lambda_{\min, \infty}^*$ . That is, if  $T$  is large enough that the burn-in of Theorem E.3 is met, we will have

$$\min_{\hat{\mathbf{a}}} \max_{\theta \in \mathcal{B}_T} \mathbb{E}_{\theta, \pi_{\text{exp}}}[\mathcal{R}(\hat{\mathbf{a}}_T; \theta)] \geq \min_{\theta: \|\theta - \theta_0\|_F^2 \leq 5(d_x^2 + d_x d_u) / (\lambda_{\min, \infty}^* T^{5/6})} \mathbb{E} \left[ \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbb{E}[\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right) \right]$$

**Sufficiency of periodic policies:** Our goal is to lower bound

$$\min_{\theta: \|\theta - \theta_0\|_F^2 \leq 5(d_x^2 + d_x d_u) / (\lambda_{\min, \infty}^* T^{5/6})} \mathbb{E} \left[ \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbb{E}[\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right) \right]$$

Fix some  $\theta$  such that  $\|\theta_0 - \theta\|_F^2 \leq 5(d_x^2 + d_x d_u) / (\lambda_{\min, \infty}^* T^{5/6})$ , and consider the extended system  $\tilde{\theta}$ , as defined in (G.4). Let  $z_t^u$  denote the component of the state of  $\tilde{\theta}$  driven by both the random and deterministic components of the input and  $z_t^w$  the component driven by the process noise. Then  $z_t = z_t^u + z_t^w$ , so

$$\sum_{t=1}^T z_t z_t^\top = \sum_{t=1}^T (z_t^u + z_t^w)(z_t^u + z_t^w)^\top \preceq 2 \sum_{t=1}^T (z_t^u (z_t^u)^\top + z_t^w (z_t^w)^\top)$$

Therefore,

$$\mathbb{E}[\Sigma_T] \preceq 2 \mathbb{E}[\sum_{t=1}^T z_t^u (z_t^u)^\top] + 2 \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0)$$

By the power constraint on  $u_t$  and Lemma H.3,

$$\mathbb{E} \sum_{t=1}^T z_t^u (z_t^u)^\top \preceq \mathbb{E} \sum_{t=1}^{2T+T_\epsilon} z_t^{\tilde{u}} (z_t^{\tilde{u}})^\top + 5\epsilon I \preceq \mathbb{E} \sum_{t=1}^{4T} z_t^{\tilde{u}} (z_t^{\tilde{u}})^\top + 5\epsilon I$$

for some input  $\tilde{\mathbf{u}}$  with period  $k_{\tilde{\mathbf{u}}} := T_\epsilon = 2H((d_x^2 + d_x)/2 + 1)$  satisfying  $\mathbb{E}[\sum_{t=0}^{k_{\tilde{\mathbf{u}}}-1} \tilde{\mathbf{u}}_t^\top \tilde{\mathbf{u}}_t] \leq k_{\tilde{\mathbf{u}}} \gamma^2$ , and

$$H = \left\lceil \log \left( \frac{\epsilon(1 - \rho^2)}{8\tau(\tilde{A}, \rho)^3 \gamma^2 T^2} \right) / \log \rho \right\rceil$$

The final inequality follows by upper bounding  $T_\epsilon \leq 2T$ , which will hold by our definition of  $H$  and assumption on the size of  $T$ . Choosing  $\epsilon = \lambda_{\min, \infty}^* T / 5$ , we can upper bound

$$\mathbb{E}[\sum_{t=1}^T z_t^{\tilde{\mathbf{u}}} (z_t^{\tilde{\mathbf{u}}})^\top] + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) \leq \mathbb{E} \sum_{t=1}^{4T} z_t^{\tilde{\mathbf{u}}} (z_t^{\tilde{\mathbf{u}}})^\top + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + \lambda_{\min, \infty}^* T \cdot I$$

**From time domain to frequency domain:** The conditions of Lemma H.4 are met for this  $\tilde{\mathbf{u}}$ , so it follows that

$$\begin{aligned} \mathbb{E} \sum_{t=1}^{4T} z_t^{\tilde{\mathbf{u}}} (z_t^{\tilde{\mathbf{u}}})^\top &\leq \mathbb{E} \Gamma_{4T}^{\text{freq}}(\tilde{\theta}, \tilde{\mathbf{u}}) \\ &+ \left( \frac{\tau(\tilde{A}, \rho)(2H((d_x^2 + d_x)/2 + 1))\sqrt{4T+1}}{1 - \rho^k} + \frac{\tau(\tilde{A}, \rho)^2(2H((d_x^2 + d_x)/2 + 1))^2}{(1 - \rho^k)^2} \right) \|\tilde{\theta}\|_{\mathcal{H}_\infty}^2 \gamma^2 \cdot I \end{aligned}$$

Then if

$$T \geq \frac{1}{\lambda_{\min, \infty}^*} \left( \frac{\tau(\tilde{A}, \rho)(2H((d_x^2 + d_x)/2 + 1))\sqrt{4T+1}}{1 - \rho^k} + \frac{\tau(\tilde{A}, \rho)^2(2H((d_x^2 + d_x)/2 + 1))^2}{(1 - \rho^k)^2} \right) \|\tilde{\theta}\|_{\mathcal{H}_\infty}^2 \gamma^2 \quad (\text{H.1})$$

we can upper bound

$$\mathbb{E} \sum_{t=1}^{4T} z_t^{\tilde{\mathbf{u}}} (z_t^{\tilde{\mathbf{u}}})^\top + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + \lambda_{\min, \infty}^* T \cdot I \leq \mathbb{E} \Gamma_{4T}^{\text{freq}}(\tilde{\theta}, \tilde{\mathbf{u}}) + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + 2\lambda_{\min, \infty}^* T \cdot I$$

**Sufficiency of deterministic inputs:** Let  $\tilde{\mathcal{U}}_{\gamma^2, k_{\tilde{\mathbf{u}}}}$  denote the set of inputs with average expected power bounded by  $\gamma^2$  and period  $k_{\tilde{\mathbf{u}}}$ . Then we have shown that

$$\begin{aligned} \text{tr} \left( \mathcal{H}(\theta_\star) (\mathbb{E}[\Sigma_T] + \lambda_{\min, \infty}^* T \cdot I)^{-1} \right) &\geq \frac{1}{2} \text{tr} \left( \mathcal{H}(\theta_\star) \left( \mathbb{E} \Gamma_{4T}^{\text{in,ss}}(\tilde{\theta}, \tilde{\mathbf{u}}) + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + 3\lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right) \\ &\geq \frac{1}{2} \min_{\mathbf{u} \in \tilde{\mathcal{U}}_{\gamma^2, k_{\tilde{\mathbf{u}}}}} \text{tr} \left( \mathcal{H}(\theta_\star) \left( \mathbb{E} \Gamma_{4T}^{\text{in,ss}}(\tilde{\theta}, \mathbf{u}) + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + 3\lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right) \end{aligned}$$

By definition of  $\Gamma^{\text{freq}}$  and for any  $\mathbf{u} \in \tilde{\mathcal{U}}_{\gamma^2, k_{\tilde{\mathbf{u}}}}$ , using that  $\tilde{\mathbf{u}} = \mathfrak{F}(\mathbf{u})$ ,

$$\begin{aligned} \mathbb{E} \Gamma_{4T}^{\text{freq}}(\tilde{\theta}, \mathbf{u}) &= \mathbb{E} \frac{4T}{k_{\tilde{\mathbf{u}}}} \frac{1}{k_{\tilde{\mathbf{u}}}} \sum_{t=1}^{k_{\tilde{\mathbf{u}}}} (e^{i \frac{2\pi t}{k_{\tilde{\mathbf{u}}}}} I - \tilde{A})^{-1} \tilde{B} \tilde{\mathbf{u}}_t \tilde{\mathbf{u}}_t^H \tilde{B}^H (e^{i \frac{2\pi t}{k_{\tilde{\mathbf{u}}}}} I - \tilde{A})^{-H} \\ &= \frac{4T}{k_{\tilde{\mathbf{u}}}} \frac{1}{k_{\tilde{\mathbf{u}}}} \sum_{t=1}^{k_{\tilde{\mathbf{u}}}} (e^{i \frac{2\pi t}{k_{\tilde{\mathbf{u}}}}} I - \tilde{A})^{-1} \tilde{B} \mathbb{E}[\tilde{\mathbf{u}}_t \tilde{\mathbf{u}}_t^H] \tilde{B}^H (e^{i \frac{2\pi t}{k_{\tilde{\mathbf{u}}}}} I - \tilde{A})^{-H} \end{aligned}$$

Define  $U_t := \mathbb{E}[\tilde{\mathbf{u}}_t \tilde{\mathbf{u}}_t^H]$ . By Parseval's Theorem, and the power constraint on  $\mathbf{u}$ , we have

$$\sum_{t=1}^{k_{\tilde{\mathbf{u}}}} \text{tr}(U_t) = \mathbb{E}[\sum_{t=1}^{k_{\tilde{\mathbf{u}}}} \tilde{\mathbf{u}}_t \tilde{\mathbf{u}}_t^H] = \mathbb{E}[k_{\tilde{\mathbf{u}}} \sum_{t=0}^{k_{\tilde{\mathbf{u}}}-1} \mathbf{u}_t^\top \mathbf{u}_t] \leq k_{\tilde{\mathbf{u}}}^2 \gamma^2$$

Thus, optimizing over over the (possibly random) input  $\mathbf{u}$ , is equivalent to optimizing over PSD matrices  $U_t$  that satisfy this trace constraint. Therefore,

$$\frac{1}{2} \min_{\mathbf{u} \in \tilde{\mathcal{U}}_{\gamma^2, k_{\tilde{\mathbf{u}}}}} \text{tr} \left( \mathcal{H}(\theta_\star) \left( \mathbb{E} \Gamma_{4T}^{\text{in,ss}}(\tilde{\theta}, \mathbf{u}) + \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}, 0) + 3\lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right)$$



$$\begin{aligned}
 &\geq \frac{1}{2} \min_{\mathbf{u} \in \mathcal{U}_{\gamma^2, k_{\tilde{\mathbf{u}}}}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{in,ss}}(\tilde{\theta}, \mathbf{u}) + \sum_{t=1}^T \mathbf{\Gamma}_t^{\text{noise}}(\tilde{\theta}, 0) + 3\lambda_{\min, \infty}^* T \cdot I \right)^{-1} \right) \\
 &\geq \frac{1}{8T} \min_{\mathbf{u} \in \mathcal{U}_{\gamma^2, k_{\tilde{\mathbf{u}}}}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}, \mathbf{u}, 0) + 3\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) \\
 &\geq \frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, 4T}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}, \mathbf{U}, 0) + 3\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right)
 \end{aligned}$$

where the constraint set in the second minimization is simply the set defined in (B.10), and we can thus drop the expectation.

**From  $\theta$  to  $\theta_*$ :** By assumption

$$\|\theta - \theta_*\|_{\text{op}} \leq \|\theta - \theta_*\|_F \leq \|\theta - \theta_0\|_F + \|\theta_0 - \theta_*\|_F \leq 2\sqrt{5(d_x^2 + d_x d_u)} / (\sqrt{\lambda_{\min, \infty}^*} T^{5/12})$$

so if

$$2\sqrt{5(d_x^2 + d_x d_u)} / (\sqrt{\lambda_{\min, \infty}^*} T^{5/12}) \leq r_{\text{cov}}(\theta_*) \quad (\text{H.2})$$

we are in the domain of Lemma H.5 and

$$\begin{aligned}
 \|\mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}, \mathbf{U}, 0) - \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0)\|_{\text{op}} &\leq L_{\text{cov}}(\theta_*, \gamma^2) \cdot \|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} = L_{\text{cov}}(\theta_*, \gamma^2) \cdot \|\theta - \theta_*\|_{\text{op}} \\
 &\leq \frac{L_{\text{cov}}(\theta_*, \gamma^2) 2\sqrt{5(d_x^2 + d_x d_u)}}{\sqrt{\lambda_{\min, \infty}^*} T^{5/12}}
 \end{aligned}$$

It follows that as long as

$$\frac{L_{\text{cov}}(\theta_*, \gamma^2) 2\sqrt{5(d_x^2 + d_x d_u)}}{\sqrt{\lambda_{\min, \infty}^*} T^{5/12}} \leq \lambda_{\min, \infty}^* \quad (\text{H.3})$$

then

$$\mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}, \mathbf{U}, 0) \preceq \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + \lambda_{\min, \infty}^* \cdot I$$

and thus,

$$\frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, 4T}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}, \mathbf{U}, 0) + 3\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) \geq \frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, 4T}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 4\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right)$$

**Concluding the lower bound:** Next, by Lemma H.8, so long as

$$4T \geq \max \left\{ \frac{8\pi \|\tilde{\theta}_*\|_{\mathcal{H}_\infty} \gamma^2}{\lambda_{\min, \infty}^*}, \frac{\pi}{2\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}} \right\} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_*)^{-2} \tilde{B}_*\|_{\text{op}} \right) \quad (\text{H.4})$$

then for any  $T' \geq 4T$  and  $\mathbf{U}^* \in \mathcal{U}_{\gamma^2, 4T}$ , there exists a  $\mathbf{U}' \in \mathcal{U}_{\gamma^2, T'}$  such that

$$\left\| \frac{1}{4T} \mathbf{\Gamma}_{4T}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}') - \frac{1}{T'} \mathbf{\Gamma}_{T'}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}^*) \right\|_{\text{op}} \leq \frac{1}{2} \lambda_{\min, \infty}^*$$

Furthermore, by Lemma H.9, if

$$4T \geq \max \left\{ \frac{16\tau(\tilde{A}_*, \rho)^2(\sigma_w^2 + \gamma^2/d_u)}{(1 - \rho^2)^2 \lambda_{\min, \infty}^*}, \log \left( \frac{(1 - \rho^2)^2 \lambda_{\min, \infty}^*}{16\tau(\tilde{A}_*, \rho)^2(\sigma_w^2 + \gamma^2/d_u)} \right) \frac{1}{2 \log \rho} \right\} \quad (\text{H.5})$$

then, for any  $T' \geq 4T$ ,

$$\left\| \frac{1}{4T} \sum_{t=1}^{4T} \mathbf{\Gamma}_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{d_u}) - \frac{1}{T'} \sum_{t=1}^{T'} \mathbf{\Gamma}_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{d_u}) \right\|_{\text{op}} \leq \frac{1}{2} \lambda_{\min, \infty}^*$$

By what we've just shown, for any  $T' \geq 4T$

$$\mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}', 0) \preceq \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}^*, 0) + \lambda_{\min, \infty}^* \cdot I$$

Thus,

$$\begin{aligned} & \frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, 4T}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{4T}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 4\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) \\ & \geq \liminf_{T' \rightarrow \infty} \frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 5\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) \end{aligned}$$

Let

$$\lambda_{\min, T}^* := \inf_{T' \geq T} \max_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \frac{1}{10T'} \lambda_{\min}(\mathbf{\Gamma}_{T'}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}))$$

Note that by Lemma H.8, so long as

$$T \geq \max \left\{ \frac{8\pi \|\tilde{\theta}_*\|_{\mathcal{H}_\infty} \gamma^2}{\lambda_{\min, \infty}^*}, \frac{\pi}{2\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}} \right\} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_*)^{-2} \tilde{B}_*\|_{\text{op}} \right) \quad (\text{H.6})$$

then for any  $T' \geq T$  and  $\mathbf{U}^* \in \mathcal{U}_{\gamma^2, T}$ , there exists a  $\mathbf{U}' \in \mathcal{U}_{\gamma^2, T'}$  such that

$$\left\| \frac{1}{T'} \mathbf{\Gamma}_{T'}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}') - \frac{1}{T'} \mathbf{\Gamma}_{T'}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}^*) \right\|_{\text{op}} \leq \frac{1}{2} \lambda_{\min, \infty}^*$$

This implies that so long as  $T$  satisfies (H.6), we will have  $\lambda_{\min, T}^* \geq \frac{1}{2} \lambda_{\min, \infty}^*$ . By definition of  $\lambda_{\min, T}^*$ , for any  $T' \geq T$  there exists some input  $\mathbf{U}'' \in \mathcal{U}_{\gamma^2, T'}$  such that  $\lambda_{\min}(\mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}'', 0)) \geq 10\lambda_{\min, T}^*$ . It follows that for any  $T' \geq T$ ,

$$\begin{aligned} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 5\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) & \geq \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 10\lambda_{\min, T}^* \cdot I \right)^{-1} \right) \\ & \geq \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}'', 0) \right)^{-1} \right) \\ & \geq \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) \right)^{-1} \right) \\ & \geq \frac{1}{2} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) \right)^{-1} \right) \end{aligned}$$

This implies that

$$\begin{aligned} & \liminf_{T' \rightarrow \infty} \frac{1}{8T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) + 5\lambda_{\min, \infty}^* \cdot I \right)^{-1} \right) \\ & \geq \liminf_{T' \rightarrow \infty} \frac{1}{16T} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_*) \left( \mathbf{\Gamma}_{T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0) \right)^{-1} \right) = \frac{1}{16T} \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_*). \end{aligned}$$

Putting everything together, Theorem E.3 and what we have shown imply that as long as  $T$  is large enough so that the burn-in of Theorem E.3 is met,  $T \geq H((d_x^2 + d_x)/2 + 1)$  and (H.1), (H.2), (H.3), (H.4), (H.5), and (H.6) hold, we will have

$$\min_{\hat{\mathbf{a}} \mid \|\theta - \theta_0\|_2^2 \leq 5(d_x^2 + d_x d_u)/(\lambda_{\min, \infty}^* T^{5/6})} \max \mathbb{E}[\mathcal{R}(\hat{\mathbf{a}}; \theta)] \geq \frac{\sigma_w^2}{16T} \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_*) - \frac{C_1}{(\lambda_{\min, \infty}^* T)^{5/4}}$$

where  $C_2 = \mathcal{O}\left((L_{\text{a1}} L_{\text{a2}} L_{\mathcal{R}2} + L_{\text{a1}}^3 L_{\mathcal{R}3} + L_{\text{hess}})(d_x^2 + d_x d_u)^{3/2} + L_{\text{a1}}^2 L_{\mathcal{R}2}\right)$ . Finally we can lower bound  $\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_*)$  with  $\Phi_{\text{opt}}(\gamma^2; \theta_*)/4$  by Lemma G.3.

**Simplifying the Burn-In Time:** It remains to simplify the bound. First, note that by Lemma F.9 of (Wagenmaker & Jamieson, 2020), as long as  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq c/\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}$ , we will have that  $\|\tilde{\theta}\|_{\mathcal{H}_\infty}$  and  $\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}$  are within a constant factor of each other. Next, note that Lemma H.1 implies that, so long as  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq \epsilon$ ,  $\|\tilde{A}^k\|_{\text{op}} \leq \tau(\tilde{A}_*, \rho)(\rho + \tau(\tilde{A}_*, \rho)\epsilon)^k$ . This implies that

$$\tau(\tilde{A}, \rho + \tau(\tilde{A}_*, \rho)\epsilon) = \sup_k \|\tilde{A}^k\|_{\text{op}} (\rho + \tau(\tilde{A}_*, \rho)\epsilon)^{-k} \leq \tau(\tilde{A}_*, \rho)$$

As long as  $\epsilon < (1 - \rho_*)/(2\tau(\tilde{A}_*, \rho))$  we can then choose  $\rho = \rho_* + \tau(\tilde{A}_*, \rho_*)\epsilon$  which will allow us to upper bound

$$\frac{\tau(\tilde{A}, \rho)^n}{(1 - \rho^m)^p} \leq \frac{c\tau(\tilde{A}_*, \rho_*)^n}{(1 - \rho_*^m)^p}$$

As we have assumed  $\|\theta - \theta_0\|_F, \|\theta_* - \theta_0\|_F \leq \sqrt{5(d_x^2 + d_x d_u)}/(\sqrt{\lambda_{\min, \infty}^* T^{5/12}})$ , we can upper bound  $\|\tilde{\theta} - \tilde{\theta}_*\|_{\text{op}} \leq \|\theta - \theta_*\|_F \leq 2\sqrt{5(d_x^2 + d_x d_u)}/(\sqrt{\lambda_{\min, \infty}^* T^{5/12}})$ . Some algebra, Lemma G.2, and the definition of  $L_{\text{cov}}(\theta_*, \gamma^2)$  and  $r_{\text{cov}}(\theta_*)$  then gives that as long as

$$T \geq \text{poly} \left( \frac{1}{1 - \rho_*}, \tau_*, \|B_*\|_{\text{op}}, d_x, d_u, \|\tilde{\theta}_*\|_{\mathcal{H}_\infty}, \gamma^2, \sigma_w^2, \frac{1}{\lambda_{\min, \infty}^*}, \log T \right)$$

these bounds on  $\|\tilde{\theta}\|_{\mathcal{H}_\infty}$  and  $\tau(\tilde{A}, \rho)$  will hold,  $T \geq H((d_x^2 + d_x)/2 + 1)$  and (H.1), (H.2), (H.3), (H.4), (H.5), and (H.6) hold. Finally, we use Lemma H.2 to replace  $\lambda_{\min, \infty}^*$  with  $\lambda_{\text{noise}}^*$ , and Lemma G.4 to upper bound  $\tau_*$ ,  $\frac{1}{1 - \rho_*}$  and  $\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}$  by  $\text{poly}(\|B_*\|_{\text{op}}, \|A_*\|_{\mathcal{H}_\infty})$ .  $\square$

**Lemma H.2.**

$$\min \left\{ \lambda_{\min} \left( \sigma_w^2 \sum_{t=0}^{d_x} A_*^t (A_*^t)^\top + \sigma_u^2 \sum_{t=0}^{d_x} A_*^t B_* B_*^\top (A_*^t)^\top \right), \sigma_u^2 \right\} \leq \limsup_{T \rightarrow \infty} \sup_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T}} 2\lambda_{\min}(\Gamma_T^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}))$$

*Proof.* Fix  $T$  and consider playing the input  $u_t \sim \mathcal{N}(0, \sigma_u^2 \cdot I)$ . By definition,

$$\sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \sigma_u) = \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0) + \mathbb{E} \left[ \sum_{t=1}^T x_t^{\mathbf{u}} (x_t^{\mathbf{u}})^\top \right]$$

By Lemma H.3, it follows that there exists some input  $\tilde{\mathbf{u}} = (\tilde{u}_t)_{t=0}^{k-1}$  with average expected power bounded by  $\gamma^2$  such that

$$\mathbb{E} \left[ \sum_{t=1}^T x_t^{\mathbf{u}} (x_t^{\mathbf{u}})^\top \right] \preceq \mathbb{E} \left[ \sum_{t=1}^{2T+k} x_t^{\tilde{\mathbf{u}}} (x_t^{\tilde{\mathbf{u}}})^\top \right] + 5I$$

where  $k := T_1 = 2H((d_x^2 + d_x)/2 + 1)$  and

$$H = \left\lceil \log \left( \frac{1 - \rho^2}{8\tau(\tilde{A}_*, \rho)^3 \gamma^2 T^2} \right) / \log \rho \right\rceil = \mathcal{O}(\log T)$$

By Lemma H.4,

$$\mathbb{E} \left[ \sum_{t=1}^{2T+k} x_t^{\tilde{\mathbf{u}}} (x_t^{\tilde{\mathbf{u}}})^\top \right] \preceq \mathbb{E} \Gamma_{2T+k}^{\text{freq}}(\tilde{\theta}_*, \tilde{\mathbf{u}}) + \left( \frac{\tau(\tilde{A}_*, \rho) \|\tilde{\theta}_*\|_{\mathcal{H}_\infty}^2 \sqrt{2T+k+1} k \gamma^2}{1 - \rho^k} + \frac{\tau(\tilde{A}_*, \rho)^2 \|\tilde{\theta}_*\|_{\mathcal{H}_\infty}^2 k^2 \gamma^2}{(1 - \rho^k)^2} \right) \cdot I$$

By definition of  $\Gamma^{\text{freq}}$  and for any  $\mathbf{U} \in \tilde{\mathcal{U}}_{\gamma^2, k}$ ,

$$\mathbb{E} \Gamma_{2T+k}^{\text{freq}}(\tilde{\theta}, \mathbf{U}) = \mathbb{E} \frac{4T}{k} \frac{1}{k} \sum_{t=1}^k (e^{i\frac{2\pi t}{k}} I - \tilde{A})^{-1} \tilde{B} \tilde{u}_t \tilde{u}_t^H \tilde{B}^H (e^{i\frac{2\pi t}{k}} I - \tilde{A})^{-H}$$

$$= \frac{4T}{k} \frac{1}{k} \sum_{t=1}^k (e^{t \frac{2\pi t}{k}} I - \tilde{A})^{-1} \tilde{B} \mathbb{E}[\tilde{u}_t \tilde{u}_t^H] \tilde{B}^H (e^{t \frac{2\pi t}{k}} I - \tilde{A})^{-H}$$

Define  $U_t := \mathbb{E}[\tilde{u}_t \tilde{u}_t^H]$ . By Parseval's Theorem, and the power constraint on  $\tilde{u}$ , we have

$$\sum_{t=1}^k \text{tr}(U_t) = \mathbb{E}[\sum_{t=1}^k \tilde{u}_t^H \tilde{u}_t] = \mathbb{E}[k \sum_{t=0}^{k-1} u_t^T u_t] \leq k^2 \gamma^2$$

It follows that there exists some  $\tilde{U} \in \mathcal{U}_{\gamma^2, k}$  such that  $\mathbb{E}\Gamma_{2T+k}^{\text{freq}}(\tilde{\theta}_*, \tilde{u}) = \Gamma_{2T+k}^{\text{freq}}(\tilde{\theta}_*, \tilde{U})$ . Putting this together, we have that

$$\begin{aligned} \lambda_{\min} \left( \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{d_u}) \right) &\leq \lambda_{\min} \left( (2T+k) \Gamma_{2T+k}^{\text{ss}}(\tilde{\theta}_*, \tilde{U}) \right) + \mathcal{O}(\sqrt{T} \log T + \text{poly} \log T) \\ &\leq \sup_{U \in \mathcal{U}_{\gamma^2, 2T+k}} \lambda_{\min} \left( (2T+k) \Gamma_{2T+k}^{\text{ss}}(\tilde{\theta}_*, U) \right) + \mathcal{O}(\sqrt{T} \log T + \text{poly} \log T) \end{aligned}$$

Dividing through by  $T$  and taking the  $\limsup_{T \rightarrow \infty}$ , we have

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \lambda_{\min} \left( \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \sigma_u) \right) \leq \limsup_{T \rightarrow \infty} \sup_{U \in \mathcal{U}_{\gamma^2, T}} 2 \lambda_{\min}(\Gamma_T^{\text{ss}}(\tilde{\theta}_*, U))$$

Finally, we see that by definition and some algebra that

$$\begin{aligned} \limsup_{T \rightarrow \infty} \frac{1}{T} \lambda_{\min} \left( \sum_{t=1}^T \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \sigma_u) \right) &= \lambda_{\min} \left( \sigma_w^2 \sum_{t=0}^{\infty} \tilde{A}_*^t (\tilde{A}_*^t)^T + \sigma_u^2 \sum_{t=0}^{\infty} \tilde{A}_*^t \tilde{B}_* \tilde{B}_*^T (\tilde{A}_*^t)^T \right) \\ &= \min \left\{ \lambda_{\min} \left( \sigma_w^2 \sum_{t=0}^{\infty} A_*^t (A_*^t)^T + \sigma_u^2 \sum_{t=0}^{\infty} A_*^t B_* B_*^T (A_*^t)^T \right), \sigma_u^2 \right\} \end{aligned}$$

Noting that

$$\lambda_{\min} \left( \sigma_w^2 \sum_{t=0}^{\infty} A_*^t (A_*^t)^T + \sigma_u^2 \sum_{t=0}^{\infty} A_*^t B_* B_*^T (A_*^t)^T \right) \geq \lambda_{\min} \left( \sigma_w^2 \sum_{t=0}^{d_x} A_*^t (A_*^t)^T + \sigma_u^2 \sum_{t=0}^{d_x} A_*^t B_* B_*^T (A_*^t)^T \right)$$

completes the proof.  $\square$

### H.3. Periodicity of Optimal Inputs

In what follows, consider an arbitrary system  $(A, B)$ , with  $A$  stable. Let  $\rho \geq \rho(A)$  be less than 1, and recall  $\tau(A, \rho) := \sup_{n \geq 0} \rho^{-n} \|A^n\|_{\text{op}}$ , finally for any error parameter  $\epsilon > 0$ ,

$$H_\epsilon := \left\lceil \log \left( \frac{\epsilon(1-\rho^2)}{8\|B\|_{\text{op}}^2 \tau(A, \rho)^3 \gamma^2 T^2} \right) / \log \rho \right\rceil. \quad (\text{H.7})$$

and define the effective time horizon

$$T_\epsilon := 2H_\epsilon((d_x^2 + d_x)/2 + 1). \quad (\text{H.8})$$

**Lemma H.3.** *Consider some input  $\{u_t\}_{t=0}^{T-1}$  satisfying  $\mathbb{E}[\sum_{t=0}^{T-1} u_t^T u_t] \leq T\gamma^2$ . Then there exists an input  $\{\tilde{u}_t\}_{t=0}^{T_\epsilon-1}$  such that*

$$\mathbb{E} \left[ \sum_{t=0}^{T_\epsilon-1} \tilde{u}_t^T \tilde{u}_t \right] \leq T_\epsilon \gamma^2$$

and extending to times  $t \geq T_\epsilon - 1$  via a periodic signal  $\tilde{u}_t = \tilde{u}_{\text{mod}(t, T_\epsilon)}$ , where equality here holds almost surely, satisfies

$$\mathbb{E} \left[ \sum_{t=1}^T x_t^u (x_t^u)^T \right] \preceq \mathbb{E} \left[ \sum_{t=1}^{2T+3T_\epsilon/2} x_t^{\tilde{u}} (x_t^{\tilde{u}})^T \right] + 5\epsilon I,$$

where above,  $x^u$  are the states under the initial inputs  $(u_t)$ , and  $x^{\tilde{u}}$  are the iterates under  $(\tilde{u}_t)$ , and where we take  $x_0^u = 0$  in both.

*Proof.* In what follows, we regard  $\epsilon > 0$  as fixed, and write  $H \leftarrow H_\epsilon$ .

We consider the response on the system with no process noise starting from  $x_0^u = 0$ . Given some input  $\{u_t\}_{t=0}^{T-1} \in \mathcal{U}_{\gamma^2}$ , the state evolves as

$$x_t^u = \sum_{s=0}^{t-1} A^{t-s-1} B u_s$$

We will use  $G_t$  to denote the Markov parameters for this system:

$$G_t := [B, AB, \dots, A^{t-2}B, A^{t-1}B]$$

and will define the extended input,  $\mathbf{u}_t$ , and truncated extended input,  $\mathbf{u}_{t;H}$  as:

$$\mathbf{u}_t := [u_t; u_{t-1}; \dots; u_1; u_0], \quad \mathbf{u}_{t;H} := [u_t; u_{t-1}; \dots; u_{t-H+2}; u_{t-H+1}]$$

If  $t < H - 1$ , we define  $u_{-s} = 0$  for all  $s > 0$ . Then the state can be written as  $x_t^u = G_t \mathbf{u}_{t-1}$ . We can approximate the state using the last  $H$  inputs as  $x_{t;H}^u = G_H \mathbf{u}_{t-1;H}$ . The following result bounds the error in such an approximation.

**Claim H.1.** Fix some input  $u_t$  with  $\mathbb{E}[\sum_{t=0}^{T-1} u_t^\top u_t] \leq T\gamma^2$  and assume we start from  $x_0^u = 0$ . Then, under our choice of  $H \geq \log(\frac{\epsilon(1-\rho^2)}{2\|B\|_{\text{op}}^2 \tau(A, \rho)^3 \gamma^2 T}) / \log \rho$ , we will have

$$\mathbb{E}\|x_t^u (x_t^u)^\top - x_{t;H}^u (x_{t;H}^u)^\top\|_{\text{op}} \leq \epsilon$$

*Proof of Claim H.1.* We first bound the state difference:

$$\|x_{t;H}^u - x_t^u\|_2 = \left\| \sum_{s=0}^{t-H-1} A^{t-s-1} B u_s \right\|_2 \leq \|A^H\|_{\text{op}} \left\| \sum_{s=0}^{t-H-1} A^{t-H-s-1} B u_s \right\|_2 \leq \tau(A, \rho) \rho^H \|x_{t-H}^u\|_2$$

By Jensen's inequality, we can bound  $\mathbb{E}\|x_t^u\|_2$  as

$$\begin{aligned} \mathbb{E}\|x_t^u\|_2 &= \mathbb{E} \left\| \sum_{s=0}^{t-1} A^{t-s-1} B u_s \right\|_2 \leq \|B\|_{\text{op}} \tau(A, \rho) \mathbb{E} \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 \right) \\ &\leq \|B\|_{\text{op}} \tau(A, \rho) \sqrt{\sum_{s=0}^{t-1} \rho^{2(t-s-1)}} \sqrt{\mathbb{E} \sum_{s=0}^{t-1} \|u_s\|_2^2} \\ &\leq \|B\|_{\text{op}} \tau(A, \rho) \sqrt{\sum_{s=0}^{t-1} \rho^{2(t-s-1)}} \sqrt{\mathbb{E} \sum_{s=0}^{T-1} \|u_s\|_2^2} \\ &\leq \frac{\|B\|_{\text{op}} \tau(A, \rho) \gamma \sqrt{T}}{\sqrt{1-\rho^2}} \end{aligned}$$

Thus,  $\mathbb{E}\|x_{t;H}^u - x_t^u\|_2 \leq \frac{\|B\|_{\text{op}} \tau(A, \rho)^2 \gamma \sqrt{2T}}{\sqrt{1-\rho^2}} \rho^H$ . Thus, by the triangle inequality and what we have just shown,

$$\mathbb{E}\|x_t^u (x_t^u)^\top - x_{t;H}^u (x_{t;H}^u)^\top\|_{\text{op}} \leq \mathbb{E}[(\|x_t^u\|_2 + \|x_{t;H}^u\|_2) \|x_t^u - x_{t;H}^u\|_2] \leq \frac{2\|B\|_{\text{op}}^2 \tau(A, \rho)^3 \gamma^2 T}{1-\rho^2} \rho^H$$

where the last inequality follows by noting that our above argument also applies to bounding  $\mathbb{E}\|x_{t;H}^u\|_2$ . The conclusion follows by some algebra.  $\square$

Fix  $H = \lceil \log(\frac{\epsilon(1-\rho^2)}{8\|B\|_{\text{op}}^2 \tau(A, \rho)^3 \gamma^2 T^2}) / \log \rho \rceil$ , then, by Claim H.1,

$$\mathbb{E} \sum_{t=1}^T x_t^u (x_t^u)^\top \preceq \mathbb{E} \sum_{j=1}^{\lceil T/H \rceil} \sum_{t=H(j-1)+1}^{Hj} x_t^u (x_t^u)^\top \preceq \mathbb{E} \sum_{j=1}^{\lceil T/H \rceil} \sum_{t=H(j-1)+1}^{Hj} x_{t;H}^u (x_{t;H}^u)^\top + \epsilon I$$

Now consider a realization of  $(u_t)$  in our probability space.

Defining

$$\mathbf{U}_{j;H} := \sum_{t=H(j-1)+1}^{Hj} \mathbf{u}_{t-1;H} \mathbf{u}_{t-1;H}^\top \quad (\text{H.9})$$

we can rewrite the covariates in terms of the Markov parameters as

$$\sum_{j=1}^{\lceil T/H \rceil} \sum_{t=H(j-1)+1}^{Hj} x_{t;H}^{\mathbf{u}} (x_{t;H}^{\mathbf{u}})^\top = \sum_{j=1}^{\lceil T/H \rceil} G_H \left( \sum_{t=H(j-1)+1}^{Hj} \mathbf{u}_{t-1;H} \mathbf{u}_{t-1;H}^\top \right) G_H^\top = \sum_{j=1}^{\lceil T/H \rceil} G_H \mathbf{U}_{j;H} G_H^\top$$

We will define the set of normalized covariance matrices as

$$\mathcal{M}_H := \left\{ G_H \mathbf{U} G_H^\top : \mathbf{U} \text{ of form (H.9) for input } \{\bar{u}_t\}_{t=1}^{2H}, \sum_{t=1}^{2H} \bar{u}_t^\top \bar{u}_t = 1 \right\}$$

The following result will allow us to express this in a more convenient form.

**Claim H.2.** Consider any  $n$  and some  $\{\mathbf{U}'_{j;H}\}_{j=1}^n$ ,  $\mathbf{U}'_{j;H} \in \mathcal{M}_H$ . Let  $p_j \in [0, 1]$ ,  $\sum_{j=1}^n p_j = 1$ . Then there exists some set of matrix inputs  $\{\mathbf{U}''_{j;H}\}_{j=1}^{(d_x^2+d_x)/2+1}$ ,  $\mathbf{U}''_{j;H} \in \mathcal{M}_H$ , and some set of weights  $q_j \in [0, 1]$ ,  $\sum_{j=1}^{(d_x^2+d_x)/2+1} q_j = 1$  such that

$$\sum_{j=1}^n p_j G_H \mathbf{U}'_{j;H} G_H^\top = \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j G_H \mathbf{U}''_{j;H} G_H^\top.$$

*Proof.* This is a direct consequence of Caratheodory's Theorem. By definition,  $\mathcal{M} \subseteq \mathcal{S}_+^{d_x}$ . The dimension of  $\mathcal{S}_+^{d_x}$  is  $(d_x^2 + d_x)/2$  so the points in  $\mathcal{M}$  can be thought of as living in a  $(d_x^2 + d_x)/2$ -dimensional space. Caratheodory's Theorem then gives that, for any point,  $x$ , that is a convex combination of elements of  $\mathcal{M}$ ,  $x$  can also be written as a convex combination of at most  $\dim(\mathcal{M}) + 1$  points in  $\mathcal{M}$ . Taking  $x = \sum_{j=1}^n p_j G_H \mathbf{U}'_{j;H} G_H^\top$ , it follows that there exists  $(d_x^2 + d_x)/2 + 1$  points  $\mathbf{U}''_{j;H} \in \mathcal{M}$  and set of weights  $q_j \in [0, 1]$ ,  $\sum_{j=1}^m q_j = 1$  such that  $x = \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j G_H \mathbf{U}''_{j;H} G_H^\top$ .  $\square$

We shall use the following definition going forward:

**Definition H.1.** For a given  $\mathbf{U}_{j;H}$ , let  $\gamma^2[\mathbf{U}_{j;H}]$  denote the power of the input corresponding to  $\mathbf{U}_{j;H}$ . That is, if  $\mathbf{U}_{j;H}$  is formed according to (H.9),

$$\gamma^2[\mathbf{U}_{j;H}] := \sum_{t=H(j-2)+1}^{Hj-1} u_t^\top u_t \quad (\text{H.10})$$

Note then that  $\mathbf{U}_{j;H} = \gamma^2[\mathbf{U}_{j;H}] \cdot \mathbf{U}'_{j;H}$  for some  $\mathbf{U}'_{j;H} \in \mathcal{M}_H$ .

Instantiating Claim H.2 with

$$\mathbf{U}'_{j;H} \leftarrow \mathbf{U}_{j;H} / \gamma^2[\mathbf{U}_{j;H}], \quad \text{and} \quad p_j = \gamma^2[\mathbf{U}_{j;H}] / \left( \sum_{i=1}^{\lceil T/H \rceil} \gamma^2[\mathbf{U}_{i;H}] \right),$$

we have that there exists some set of matrices  $\{\tilde{\mathbf{U}}_{j;H}\}_{j=1}^{(d_x^2+d_x)/2+1} \subseteq \mathcal{M}_H$  and some set of weights  $q_j$  such that

$$\sum_{j=1}^{\lceil T/H \rceil} G_H \mathbf{U}_{j;H} G_H^\top = \left( \sum_{i=1}^{\lceil T/H \rceil} \gamma^2[\mathbf{U}_{i;H}] \right) \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j G_H \tilde{\mathbf{U}}_{j;H} G_H^\top \quad (\text{H.11})$$

For future reference, we denote

$$\tilde{\gamma}^2 := \sum_{i=1}^{\lceil T/H \rceil} \gamma^2[\mathbf{U}_{i;H}]$$

Note that while  $\tilde{\mathbf{U}}_{j;H} \in \mathcal{M}_H$ , the associate covariates may not be realizable in only  $H((d_x^2 + d_x)/2 + 1)$  steps, because  $\tilde{u}_t$  for a given  $t$  will be present in both blocks  $\mathbf{U}_{j;H}$  and  $\mathbf{U}_{j+1;H}$ —these blocks cannot be chosen independently. However, this response can be realized in  $2H((d_x^2 + d_x)/2 + 1)$  steps, which we recall is precisely our definition of  $T_\epsilon$ .

For a given  $j \in \{1, \dots, (d_x^2 + d_x)/2 + 1\}$ , let  $\{\tilde{u}_{t;j}\}_{t=0}^{2H-1}$  be the set of inputs for which (H.9) is satisfied for  $\tilde{\mathbf{U}}_{j;H}$ , and such that  $\sum_{t=0}^{2H-1} \tilde{u}_{t;j}^\top \tilde{u}_{t;j} = 1$ . Let  $\{\tilde{u}_t\}_{t=0}^{T_\epsilon}$  denote the sequence of inputs formed by concatenating  $\{\sqrt{\frac{1}{T} \tilde{\gamma}^2 q_j T_\epsilon} \tilde{u}_{t;j}\}_{t=0}^{2H-1}$  for all  $j$ . That is, set

$$\tilde{u}_t = \sqrt{\frac{1}{T} \tilde{\gamma}^2 q_j T_\epsilon} \tilde{u}_{t',j'} \quad \text{where } j' = \lfloor t/j \rfloor + 1, t' = t - (j' - 1)2H$$

Finally, extend  $\tilde{u}_t$  to all  $t$  via  $\tilde{u}_t = \tilde{u}_{\text{mod}(t, T_\epsilon)}$ , where the equality holds almost surely. Then,

$$\frac{1}{T} \frac{\tilde{\gamma}^2 q_j T_\epsilon}{2} G_H \tilde{\mathbf{U}}_{j;H} G_H^\top = \sum_{t=2Hj-H+1}^{2Hj} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \quad (\text{H.12})$$

so  $\sum_{j=1}^{(d_x^2+d_x)/2+1} \frac{T_\epsilon}{2T} \tilde{\gamma}^2 q_j G_H \tilde{\mathbf{U}}_{j;H} G_H^\top$  corresponds to half of the input response, and thus

$$\begin{aligned} \tilde{\gamma}^2 \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j G_H \tilde{\mathbf{U}}_{j;H} G_H^\top &= \frac{T}{T_\epsilon/2} \sum_{j=1}^{(d_x^2+d_x)/2+1} \sum_{t=2Hj-H+1}^{2Hj} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \\ &\preceq \left[ \frac{T}{T_\epsilon/2} \right] \sum_{j=1}^{(d_x^2+d_x)/2+1} \sum_{t=2Hj-H+1}^{2Hj} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \\ &\stackrel{(a)}{=} \sum_{j=1}^{\lceil \frac{T}{T_\epsilon/2} \rceil ((d_x^2+d_x)/2+1)} \sum_{t=2Hj-H+1}^{2Hj} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \\ &\stackrel{(b)}{\preceq} \sum_{t=1}^{2H \lceil \frac{T}{T_\epsilon/2} \rceil ((d_x^2+d_x)/2+1)} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \\ &\preceq \sum_{t=1}^{2T+T_\epsilon} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top \end{aligned}$$

where (a) holds because, by construction, we will have  $x_{t;H}^{\tilde{u}} = x_{t+j2H((d_x^2+d_x)/2+1);H}^{\tilde{u}}$  for any  $j$ , since the input is  $T_\epsilon$ -periodic, and (b) follows as we are simply including more PSD terms in the sum. As this holds pointwise in our probability space, it follows that

$$\mathbb{E} \sum_{j=1}^{\lceil T/H \rceil} \sum_{t=H(j-1)+1}^{Hj} x_{t;H}^{\mathbf{u}}(x_{t;H}^{\mathbf{u}})^\top = \mathbb{E} \tilde{\gamma}^2 \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j G_H \tilde{\mathbf{U}}_{j;H} G_H^\top \preceq \mathbb{E} \sum_{t=1}^{2T+T_\epsilon} x_{t;H}^{\tilde{u}}(x_{t;H}^{\tilde{u}})^\top$$

The input sequence  $\{\tilde{u}_t\}_{t=0}^{T_\epsilon}$  satisfies

$$\begin{aligned} \mathbb{E} \sum_{t=0}^{T_\epsilon-1} \tilde{u}_t^\top \tilde{u}_t &= \mathbb{E} \frac{T_\epsilon/2}{T} \cdot \tilde{\gamma}^2 \cdot \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j \sum_{t=0}^{2H-1} \tilde{u}_{t;j}^\top \tilde{u}_{t;j} \\ &\stackrel{(a)}{=} \mathbb{E} \frac{T_\epsilon/2}{T} \cdot \tilde{\gamma}^2 \cdot \sum_{j=1}^{(d_x^2+d_x)/2+1} q_j \end{aligned}$$

$$\begin{aligned}
 &= \frac{T_\epsilon/2}{T} \mathbb{E} \tilde{\gamma}^2 \\
 &\stackrel{(b)}{\leq} T_\epsilon \gamma^2
 \end{aligned}$$

where (a) follows since  $\sum_{t=0}^{2H-1} \tilde{u}_{t;j}^\top \tilde{u}_{t;j} = 1$  almost surely, and (b) follows since

$$\mathbb{E} \tilde{\gamma}^2 = \mathbb{E} \sum_{i=1}^{\lceil T/H \rceil} \gamma^2(\mathbf{u}_{i;H}) = \mathbb{E} \sum_{i=1}^{\lceil T/H \rceil} \sum_{t=H(i-2)+1}^{Hi-1} u_t^\top u_t \leq 2T\gamma^2$$

Thus,  $\tilde{u}_t$  satisfies the power constraint  $\mathbb{E} \sum_{t=0}^{T_\epsilon-1} \tilde{u}_t^\top \tilde{u}_t \leq T_\epsilon \gamma^2$ , which implies it also satisfies the constraint  $\mathbb{E} \sum_{t=1}^T \tilde{u}_t^\top \tilde{u}_t \leq T\gamma^2$ . By Lemma H.1, given our choice of  $H$  and this power constraint,

$$\mathbb{E} \sum_{t=1}^{2T+T_\epsilon} x_{t;H}^{\tilde{u}} (x_{t;H}^{\tilde{u}})^\top \leq \mathbb{E} \sum_{t=1}^{2T+T_\epsilon} x_t^{\tilde{u}} (x_t^{\tilde{u}})^\top + 4\epsilon I$$

Finally, note that for any  $s$ , we can bound

$$\mathbb{E} \sum_{t=s}^{T_\epsilon+s-1} \tilde{u}_t^\top \tilde{u}_t \leq 2T_\epsilon \gamma^2$$

since the sum can be contained by at most two periods of the input. The conclusion follows.  $\square$

#### H.4. Frequency Domain Approximation

**Lemma H.4.** *Let  $\{u_t\}_{t=0}^{k-1}$  be a signal with  $\mathbb{E}[\sum_{t=0}^{k-1} u_t^\top u_t] \leq k\gamma^2$ . Consider playing  $u_t$  periodically for  $T$  steps on system  $\theta = (A, B)$  with no process noise, where we assume  $x_0 = 0$  and set  $u_t = u_{\text{mod}(t,k)}$  almost surely. Then,*

$$\begin{aligned}
 &\left\| \mathbb{E} \sum_{t=0}^T x_t x_t^\top - \mathbb{E} \frac{1}{T} \sum_{t=1}^T (e^{t\frac{2\pi t}{T}} I - A)^{-1} B \check{u}_t \check{u}_t^\top B^\top (e^{t\frac{2\pi t}{T}} I - A)^{-\top} \right\|_{\text{op}} \\
 &\leq \frac{\tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}} \sqrt{T+1} k \gamma^2}{1 - \rho^k} + \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k^2 \gamma^2}{(1 - \rho^k)^2}.
 \end{aligned}$$

where  $(\check{u}_t)_{t=1}^T = \mathfrak{F}^{-1}((u_t)_{t=1}^T)$ .

*Proof.* Define  $G(e^{i\omega}) := (e^{i\omega} I - A)^{-1} B$  and let  $(\check{x}_t)_{t=1}^T = \mathfrak{F}^{-1}((x_t)_{t=1}^T)$  denote the  $T$  point DFT of  $x_t$ . Then, by Parseval's Theorem,

$$\begin{aligned}
 &\left\| \sum_{t=0}^T x_t x_t^\top - \frac{1}{T} \sum_{t=1}^T G(e^{i\frac{2\pi t}{T}}) \check{u}_t \check{u}_t^\top G(e^{i\frac{2\pi t}{T}})^\top \right\|_{\text{op}} = \left\| \frac{1}{T} \sum_{t=1}^T \check{x}_t \check{x}_t^\top - \frac{1}{T} \sum_{t=1}^T G(e^{i\frac{2\pi t}{T}}) \check{u}_t \check{u}_t^\top G(e^{i\frac{2\pi t}{T}})^\top \right\|_{\text{op}} \\
 &\leq \frac{1}{T} \sum_{t=1}^T \|G(e^{i\frac{2\pi t}{T}}) \check{u}_t - \check{x}_t\|_2 (\|\check{x}_t\|_2 + \|G(e^{i\frac{2\pi t}{T}}) \check{u}_t\|_2)
 \end{aligned}$$

By Taylor expanding,

$$G(e^{i\omega}) = \sum_{s=0}^{\infty} e^{-i\omega(s+1)} A^s B$$

By definition of a DFT, and since  $x_0 = 0$ ,

$$\check{x}_\ell = \sum_{t=0}^{T-1} e^{-i\frac{2\pi \ell}{T} t} x_t = \sum_{t=1}^{T-1} \sum_{s=0}^{t-1} e^{-i\frac{2\pi \ell}{T} t} A^{t-s-1} B u_s = \sum_{s=0}^{T-1} \left( \sum_{t=0}^{T-s-2} e^{-i\frac{2\pi \ell}{T} (t+1)} A^t \right) e^{-i\frac{2\pi \ell}{T} s} B u_s$$



$$\tilde{u}_s = \sum_{t=0}^{T-1} e^{-i\frac{2\pi\ell}{T}t} u_t$$

Therefore,

$$\begin{aligned} G(e^{i\frac{2\pi\ell}{T}})\tilde{u}_\ell - \tilde{x}_\ell &= \sum_{s=0}^{T-1} \left( \sum_{t=0}^{\infty} e^{-i\frac{2\pi\ell}{T}(t+1)} A^t \right) e^{-i\frac{2\pi\ell}{T}s} B u_s - \sum_{s=0}^{T-1} \left( \sum_{t=0}^{T-s-2} e^{-i\frac{2\pi\ell}{T}(t+1)} A^t \right) e^{-i\frac{2\pi\ell}{T}s} B u_s \\ &= \sum_{s=0}^{T-1} \left( \sum_{t=T-s-1}^{\infty} e^{-i\frac{2\pi\ell}{T}(t+1)} A^t \right) e^{-i\frac{2\pi\ell}{T}s} B u_s \\ &= \sum_{s=0}^{T-1} e^{-i\frac{2\pi\ell}{T}(T-s-1)} A^{T-s-1} \left( \sum_{t=0}^{\infty} e^{-i\frac{2\pi\ell}{T}(t+1)} A^t \right) e^{-i\frac{2\pi\ell}{T}s} B u_s \\ &= e^{-i\frac{2\pi\ell}{T}(T-1)} \sum_{s=0}^{T-1} A^{T-s-1} G(e^{i\frac{2\pi\ell}{T}}) B u_s \end{aligned}$$

Thus, since  $u_s = u_{s+k}$  by assumption,

$$\begin{aligned} \|G(e^{i\frac{2\pi\ell}{T}})\tilde{u}_\ell - \tilde{x}_\ell\|_2 &\leq \tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty} \|B\|_{\text{op}} \sum_{s=0}^{T-1} \rho^{T-s-1} \|u_s\|_2 \\ &\leq \tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty} \|B\|_{\text{op}} \sum_{j=0}^{\lceil T/k \rceil} \rho^{kj} \sum_{s=0}^{k-1} \|u_s\|_2 \\ &\leq \tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty} \|B\|_{\text{op}} \sqrt{k} \sqrt{\sum_{s=0}^{k-1} \|u_s\|_2^2} \sum_{j=0}^{\lceil T/k \rceil} \rho^{kj} \\ &\leq \frac{\tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty} \|B\|_{\text{op}} \sqrt{k}}{1 - \rho^k} \sqrt{\sum_{s=0}^{k-1} \|u_s\|_2^2} \end{aligned}$$

By Parseval's Theorem, and again since  $u_s = u_{s+k}$ ,

$$\sum_{t=1}^T \|\tilde{u}_t\|_2^2 = T \sum_{t=0}^{T-1} \|u_t\|_2^2 \leq T \lceil T/k \rceil \sum_{t=0}^{k-1} \|u_t\|_2^2$$

So,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \|G(e^{i\frac{2\pi\ell}{T}})\tilde{u}_t - \tilde{x}_t\|_2 \|G(e^{i\frac{2\pi\ell}{T}})\tilde{u}_t\|_2 &\leq \frac{\|\theta\|_{\mathcal{H}_\infty}}{T} \sqrt{\sum_{t=1}^T \|G(e^{i\frac{2\pi\ell}{T}})\tilde{u}_t - \tilde{x}_t\|_2^2} \sqrt{\sum_{t=1}^T \|\tilde{u}_t\|_2^2} \\ &\leq \frac{\|\theta\|_{\mathcal{H}_\infty}}{T} \sqrt{T \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k}{(1 - \rho^k)^2} \sum_{s=0}^{k-1} \|u_s\|_2^2} \sqrt{T \lceil T/k \rceil \sum_{t=0}^{k-1} \|u_t\|_2^2} \\ &= \frac{\tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty} \|B\|_{\text{op}} \sqrt{k \lceil T/k \rceil}}{1 - \rho^k} \sum_{t=0}^{k-1} \|u_t\|_2^2 \end{aligned}$$

Again by Parseval's theorem, and by the same calculation as was performed above,

$$\begin{aligned} \sum_{t=1}^T \|\tilde{x}_t\|_2^2 &= T \sum_{t=1}^{T-1} \|x_t\|_2^2 = T \sum_{t=1}^{T-1} \left\| \sum_{s=0}^{t-1} A^{t-s-1} B u_s \right\|_2^2 \leq T \sum_{t=1}^{T-1} \tau(A, \rho)^2 \|B\|_{\text{op}}^2 \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 \right)^2 \\ &\leq \frac{T \tau(A, \rho)^2 \|B\|_{\text{op}}^2 k}{(1 - \rho^k)^2} \sum_{s=0}^{k-1} \|u_s\|_2^2 \end{aligned}$$

So,

$$\begin{aligned}
 \frac{1}{T} \sum_{t=1}^T \|G(e^{i\frac{2\pi t}{T}})\check{u}_t - \check{x}_t\|_2 \|X(e^{i\frac{2\pi t}{T}})\|_2 &\leq \frac{1}{T} \sqrt{\sum_{t=1}^T \|G(e^{i\frac{2\pi t}{T}})\check{u}_t - \check{x}_t\|_2^2} \sqrt{\sum_{t=1}^T \|\check{x}_t\|_2^2} \\
 &\leq \frac{\|\theta\|_{\mathcal{H}_\infty}}{T} \sqrt{T \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k \sum_{s=0}^{k-1} \|u_s\|_2^2}{(1-\rho^k)^2}} \sqrt{\frac{T \tau(A, \rho)^2 \|B\|_{\text{op}}^2 k \sum_{s=0}^{k-1} \|u_s\|_2^2}{(1-\rho^k)^2}} \\
 &= \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k \sum_{t=0}^{k-1} \|u_t\|_2^2}{(1-\rho^k)^2}
 \end{aligned}$$

It follows that

$$\begin{aligned}
 \mathbb{E} \left\| \sum_{t=0}^T x_t x_t^\top - \frac{1}{T} \sum_{t=1}^T G(e^{i\frac{2\pi t}{T}})\check{u}_t \check{u}_t^\top G(e^{i\frac{2\pi t}{T}})^\top \right\|_{\text{op}} &\leq \frac{\tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}} \sqrt{k \lceil T/k \rceil}}{1-\rho^k} \mathbb{E} \sum_{t=0}^{k-1} \|u_t\|_2^2 + \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k \sum_{t=0}^{k-1} \|u_t\|_2^2}{(1-\rho^k)^2} \\
 &\leq \frac{\tau(A, \rho) \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}} \sqrt{T+1} k \gamma^2}{1-\rho^k} + \frac{\tau(A, \rho)^2 \|\theta\|_{\mathcal{H}_\infty}^2 \|B\|_{\text{op}}^2 k^2 \gamma^2}{(1-\rho^k)^2}
 \end{aligned}$$

and the conclusion follows.  $\square$

## H.5. Smoothness of Covariates

**Lemma H.5.** For all  $U \in \mathcal{U}_{\gamma^2, k}$  and all  $\theta$  with

$$\|\theta - \theta_\star\|_{\text{op}} \leq \min \left\{ \frac{1-\rho}{2\tau(A_\star, \rho)}, \frac{1}{2\|A_\star\|_{\mathcal{H}_\infty}}, 1 \right\} =: r_{\text{cov}}(\theta_\star) \quad (\text{H.13})$$

if  $T_2$  is divisible by  $k$ ,

$$\begin{aligned}
 \|\Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u) - \Gamma_{T_1, T_2}^{\text{ss}}(\theta_\star, U, \sigma_u)\|_{\text{op}} &\leq \left( \frac{8(\sigma_w^2 + \sigma_u^2 \|B_\star\|_{\text{op}}^2) \tau(A_\star, \rho)^3}{(1-\rho^2)^2} + \frac{4\sigma_u^2 (\|B_\star\|_{\text{op}} + 1) \tau(A_\star, \rho)^2}{1-\rho^2} \right. \\
 &\quad \left. + 34\gamma^2 \|A_\star\|_{\mathcal{H}_\infty}^3 (\|B_\star\|_{\text{op}} + 1)^2 \right) \|\theta - \theta_\star\|_{\text{op}} \\
 &=: L_{\text{cov}}(\theta_\star, \gamma^2) \cdot \|\theta - \theta_\star\|_{\text{op}}.
 \end{aligned} \quad (\text{H.14})$$

*Proof.* For convenience denote  $\epsilon = \|\theta - \theta_\star\|_{\text{op}}$ . As  $\Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u) = I_{d_x} \otimes \Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u)$ ,

$$\|\Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u) - \Gamma_{T_1, T_2}^{\text{ss}}(\theta_\star, U, \sigma_u)\|_{\text{op}} = \|\Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u) - \Gamma_{T_1, T_2}^{\text{ss}}(\theta_\star, U, \sigma_u)\|_{\text{op}}$$

By definition, when  $T_2$  is divisible by  $k$ ,

$$\begin{aligned}
 \Gamma_{T_1, T_2}^{\text{ss}}(\theta, U, \sigma_u) &= \frac{1}{T_2} \Gamma_{T_2}^{\text{freq}}(\theta, U) + \frac{1}{T_1} \sum_{t=1}^{T_1} \Gamma_t^{\text{noise}}(\theta, \sigma_u) \\
 \Gamma_{T_2}^{\text{freq}}(\theta, U) &= \frac{T_2}{k} \frac{1}{k} \sum_{\ell=1}^k (e^{i\frac{2\pi \ell}{k}} I - A)^{-1} B U_\ell B^\top (e^{i\frac{2\pi \ell}{k}} I - A)^{-\text{H}} \\
 \Gamma_t^{\text{noise}}(\theta, \sigma_u) &= \sigma_w^2 \sum_{s=0}^{t-1} A^s (A^s)^\top + \sigma_u^2 \sum_{s=0}^{t-1} A^s B B^\top (A^s)^\top
 \end{aligned}$$

Thus,

$$\begin{aligned} & \|\Gamma_{T_1, T_2}^{\text{ss}}(\theta, \mathbf{U}, \sigma_u) - \Gamma_{T_1, T_2}^{\text{ss}}(\theta_*, \mathbf{U}, \sigma_u)\|_{\text{op}} \\ & \leq \underbrace{\frac{1}{T_2} \|\Gamma_{T_2}^{\text{freq}}(\theta, \mathbf{U}) - \Gamma_{T_2}^{\text{freq}}(\theta_*, \mathbf{U})\|_{\text{op}}}_{(a)} + \underbrace{\frac{1}{T_1} \sum_{t=1}^{T_1} \|\Gamma_t^{\text{noise}}(\theta, \sigma_u) - \Gamma_t^{\text{noise}}(\theta_*, \sigma_u)\|_{\text{op}}}_{(b)} \end{aligned}$$

Lemma H.7 gives that, when (H.13) holds,

$$\begin{aligned} (a) & \leq \left( \max_{\omega \in [0, 2\pi]} \gamma^2 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}}^2 \|B_*\|_{\text{op}} (16 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}} \|B_*\|_{\text{op}} + 2) \right) \epsilon \\ & \quad + \left( \max_{\omega \in [0, 2\pi]} \gamma^2 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}}^2 (32 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}} \|B_*\|_{\text{op}} + 2) \right) \epsilon^2 \\ & \quad + \left( \max_{\omega \in [0, 2\pi]} 16 \gamma^2 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}}^3 \right) \epsilon^3 \\ & \leq \left( \max_{\omega \in [0, 2\pi]} 34 \gamma^2 \|(e^{i\omega} I - A_*)^{-1}\|_{\text{op}}^3 (\|B_*\|_{\text{op}} + 1)^2 \right) \epsilon \end{aligned}$$

while Lemma H.6 gives that, when (H.13) holds,

$$\begin{aligned} (b) & \leq \left( \frac{8(\sigma_w^2 + \sigma_u^2 \|B_*\|_{\text{op}}^2) \tau(A_*, \rho)^3}{(1 - \rho^2)^2} + \frac{4\sigma_u^2 \|B_*\|_{\text{op}} \tau(A_*, \rho)^2}{1 - \rho^2} \right) \epsilon + \frac{2\sigma_u^2 \tau(A_*, \rho)^2}{1 - \rho^2} \epsilon^2 \\ & \leq \left( \frac{8(\sigma_w^2 + \sigma_u^2 \|B_*\|_{\text{op}}^2) \tau(A_*, \rho)^3}{(1 - \rho^2)^2} + \frac{4\sigma_u^2 (\|B_*\|_{\text{op}} + 1) \tau(A_*, \rho)^2}{1 - \rho^2} \right) \epsilon \end{aligned}$$

The result follows.  $\square$

**Lemma H.6.** Assume that  $\|[A, B] - [\hat{A}, \hat{B}]\|_{\text{op}} \leq \epsilon$  and that  $\epsilon \leq \frac{1-\rho}{2\tau(A, \rho)}$ , then

$$\begin{aligned} & \left\| \sum_{k=0}^t (\sigma_w^2 A^k (A^k)^\top + \sigma_u^2 A^k B B^\top (A^k)^\top) - \sum_{k=0}^t (\sigma_w^2 \hat{A}^k (\hat{A}^k)^\top + \sigma_u^2 \hat{A}^k \hat{B} \hat{B}^\top (\hat{A}^k)^\top) \right\|_{\text{op}} \\ & \leq \left( \frac{8(\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2) \tau(A, \rho)^3}{(1 - \rho^2)^2} + \frac{4\sigma_u^2 \|B\|_{\text{op}} \tau(A, \rho)^2}{1 - \rho^2} \right) \epsilon + \frac{2\sigma_u^2 \tau(A, \rho)^2}{1 - \rho^2} \epsilon^2. \end{aligned}$$

*Proof.* First note that  $\|[A, B] - [\hat{A}, \hat{B}]\|_{\text{op}} \leq \epsilon$  implies  $\|A - \hat{A}\|_{\text{op}} \leq \epsilon, \|B - \hat{B}\|_{\text{op}} \leq \epsilon$ . We will denote  $\hat{A} = A + \Delta_A$ , where  $\|\Delta_A\|_{\text{op}} \leq \epsilon$ . We can upper bound

$$\begin{aligned} & \left\| \sum_{k=0}^t (\sigma_w^2 A^k (A^k)^\top + \sigma_u^2 A^k B B^\top (A^k)^\top) - \sum_{k=0}^t (\sigma_w^2 \hat{A}^k (\hat{A}^k)^\top + \sigma_u^2 \hat{A}^k \hat{B} \hat{B}^\top (\hat{A}^k)^\top) \right\|_{\text{op}} \\ & \leq \sigma_w^2 \sum_{k=0}^t \|A^k (A^k)^\top - (A + \Delta_A)^k ((A + \Delta_A)^k)^\top\|_{\text{op}} + \sigma_u^2 \sum_{k=0}^t \|A^k B B^\top (A^k)^\top - (A + \Delta_A)^k \hat{B} \hat{B}^\top ((A + \Delta_A)^k)^\top\|_{\text{op}} \end{aligned}$$

By the triangle inequality,

$$\begin{aligned} & \|A^k (A^k)^\top - (A + \Delta_A)^k ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ & \leq \|A^k (A^k)^\top - A^k ((A + \Delta_A)^k)^\top + A^k ((A + \Delta_A)^k)^\top - (A + \Delta_A)^k ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ & \leq (\|A^k\|_{\text{op}} + \|(A + \Delta_A)^k\|_{\text{op}}) \|A^k - (A + \Delta_A)^k\|_{\text{op}} \end{aligned}$$

By Proposition H.1,

$$\|(A + \Delta_A)^k\|_{\text{op}} \leq \tau(A, \rho) (\rho + \tau(A, \rho) \epsilon)^k$$

and

$$\|(A + \Delta_A)^k - A^k\|_{\text{op}} \leq k\tau(A, \rho)^2(\rho + \tau(A, \rho)\epsilon)^{k-1}\epsilon$$

Combining all of this we have

$$\|A^k(A^k)^\top - (A + \Delta_A)^k((A + \Delta_A)^k)^\top\|_{\text{op}} \leq k\tau(A, \rho)^3(\rho^k + (\rho + \tau(A, \rho)\epsilon)^k)(\rho + \tau(A, \rho)\epsilon)^{k-1}\epsilon$$

Denote  $\rho_2 := \rho + \tau(A, \rho)\epsilon$ . Since we have assumed that  $\epsilon \leq \frac{1-\rho}{2\tau(A, \rho)}$ ,  $\rho_2 \leq \frac{1}{2} + \frac{1}{2}\rho$ . Then it follows:

$$\begin{aligned} \sigma_w^2 \sum_{k=0}^t \|A^k(A^k)^\top - (A + \Delta_A)^k((A + \Delta_A)^k)^\top\|_{\text{op}} &\leq \frac{\sigma_w^2 \tau(A, \rho)^3 \epsilon}{\rho_2} \sum_{k=0}^t k((\rho\rho_2)^k + \rho_2^{2k}) \\ &\leq \sigma_w^2 \tau(A, \rho)^3 \epsilon \left( \frac{\rho}{(1 - \rho\rho_2)^2} + \frac{\rho_2}{(1 - \rho_2^2)^2} \right) \\ &\leq \frac{2\sigma_w^2 \tau(A, \rho)^3 \epsilon}{(1 - \rho_2^2)^2} \\ &\leq \frac{8\sigma_w^2 \tau(A, \rho)^3 \epsilon}{(1 - \rho^2)^2} \end{aligned}$$

where the last inequality follows since  $1 - \rho_2^2 \geq \frac{1}{2}(1 - \rho^2)$ . Denoting  $\widehat{B} = B + \Delta_B$ , and using what we have already shown, we have

$$\begin{aligned} &\|A^k B B^\top (A^k)^\top - (A + \Delta_A)^k \widehat{B} \widehat{B}^\top ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ &= \|A^k B B^\top (A^k)^\top - (A + \Delta_A)^k B B^\top ((A + \Delta_A)^k)^\top - (A + \Delta_A)^k \Delta_B B^\top ((A + \Delta_A)^k)^\top \\ &\quad - (A + \Delta_A)^k B \Delta_B^\top ((A + \Delta_A)^k)^\top - (A + \Delta_A)^k \Delta_B \Delta_B^\top ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ &\leq \|A^k B B^\top (A^k)^\top - A^k B B^\top ((A + \Delta_A)^k)^\top + A^k B B^\top ((A + \Delta_A)^k)^\top - (A + \Delta_A)^k B B^\top ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ &\quad + 2\|B\|_{\text{op}} \tau(A, \rho)^2 (\rho + \tau(A, \rho)\epsilon)^{2k} \epsilon + \tau(A, \rho)^2 (\rho + \tau(A, \rho)\epsilon)^{2k} \epsilon^2 \\ &\leq \|B\|_{\text{op}}^2 (\|A^k\|_{\text{op}} + \|(A + \Delta_A)^k\|_{\text{op}}) \|A^k - (A + \Delta_A)^k\|_{\text{op}} + 2\|B\|_{\text{op}} \tau(A, \rho)^2 (\rho + \tau(A, \rho)\epsilon)^{2k} \epsilon \\ &\quad + \tau(A, \rho)^2 (\rho + \tau(A, \rho)\epsilon)^{2k} \epsilon^2 \\ &\leq \|B\|_{\text{op}}^2 k\tau(A, \rho)^3 (\rho^k + \rho_2^k) \rho_2^{k-1} \epsilon + 2\|B\|_{\text{op}} \tau(A, \rho)^2 \rho_2^{2k} \epsilon + \tau(A, \rho)^2 \rho_2^{2k} \epsilon^2 \end{aligned}$$

Thus,

$$\begin{aligned} &\sigma_u^2 \sum_{k=0}^t \|A^k B B^\top (A^k)^\top - (A + \Delta_A)^k \widehat{B} \widehat{B}^\top ((A + \Delta_A)^k)^\top\|_{\text{op}} \\ &\leq \sigma_u^2 \|B\|_{\text{op}}^2 \tau(A, \rho)^3 \epsilon \sum_{k=0}^t k(\rho^k + \rho_2^k) \rho_2^{k-1} + 2\sigma_u^2 \|B\|_{\text{op}} \tau(A, \rho)^2 \epsilon \sum_{k=0}^t \rho_2^{2k} + \sigma_u^2 \tau(A, \rho)^2 \epsilon^2 \sum_{k=0}^t \rho_2^{2k} \\ &\leq \frac{8\sigma_u^2 \|B\|_{\text{op}}^2 \tau(A, \rho)^3 \epsilon}{(1 - \rho^2)^2} + \frac{4\sigma_u^2 \|B\|_{\text{op}} \tau(A, \rho)^2 \epsilon}{1 - \rho^2} + \frac{2\sigma_u^2 \tau(A, \rho)^2 \epsilon^2}{1 - \rho^2} \end{aligned}$$

The conclusion follows.  $\square$

**Lemma H.7.** Assume that  $\|[\widehat{A}, \widehat{B}] - [A, B]\|_{\text{op}} \leq \epsilon$ ,  $\epsilon \leq (\max_{\omega \in [0, 2\pi]} 2\|(e^{i\omega} I - A)^{-1}\|_{\text{op}})^{-1}$ , and  $U_\ell \succeq 0$ ,  $\sum_{\ell=1}^k \text{tr}(U_\ell) \leq k^2 \gamma^2$ , then:

$$\begin{aligned} &\frac{1}{k} \left\| \sum_{\ell=1}^k (e^{i\omega_\ell} I - \widehat{A})^{-1} \widehat{B} U_\ell \widehat{B}^\top (e^{i\omega_\ell} I - \widehat{A})^{-\text{H}} - \sum_{\ell=1}^k (e^{i\omega_\ell} I - A)^{-1} B U_\ell B^\top (e^{i\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}} \\ &\leq \left( \max_{\omega \in [0, 2\pi]} k\gamma^2 \|(e^{i\omega} I - A)^{-1}\|_{\text{op}}^2 \|B\|_{\text{op}} (16\|(e^{i\omega} I - A)^{-1}\|_{\text{op}} \|B\|_{\text{op}} + 2) \right) \epsilon \end{aligned}$$

$$\begin{aligned}
 & + \left( \max_{\omega \in [0, 2\pi]} k\gamma^2 \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}^2 (32\|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}\|B\|_{\text{op}} + 2) \right) \epsilon^2 \\
 & + \left( \max_{\omega \in [0, 2\pi]} 16k\gamma^2 \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}^3 \right) \epsilon^3
 \end{aligned}$$

*Proof.* Note first that  $\|[\widehat{A}, \widehat{B}] - [A, B]\|_{\text{op}} \leq \epsilon$  implies  $\|\widehat{A} - A\|_{\text{op}} \leq \epsilon$ ,  $\|\widehat{B} - B\|_{\text{op}} \leq \epsilon$  since:

$$\|[\widehat{A}, \widehat{B}] - [A, B]\|_{\text{op}} = \max_{u \in \mathcal{S}^{2d}, v \in \mathcal{S}^{d+p}} u^\top ([\widehat{A}, \widehat{B}] - [A, B])v \geq \max_{\substack{u \in \mathcal{S}^{2d}, u_{d+1:2d}=0 \\ v \in \mathcal{S}^{d+p}, v_{d+1:d+p}=0}} u^\top ([\widehat{A}, \widehat{B}] - [A, B])v = \|\widehat{A} - A\|_{\text{op}}$$

If we denote  $\widehat{A} = A + \Delta_A$ ,  $\widehat{B} = B + \Delta_B$ , then:

$$\begin{aligned}
 & \left\| \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - \widehat{A})^{-1} \widehat{B} U_\ell \widehat{B}^\text{H} (e^{\iota\omega_\ell} I - \widehat{A})^{-\text{H}} - \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} B U_\ell B^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}} \\
 & \leq \underbrace{\left\| \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - \widehat{A})^{-1} \widehat{B} U_\ell \widehat{B}^\text{H} (e^{\iota\omega_\ell} I - \widehat{A})^{-\text{H}} - \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} \widehat{B} U_\ell \widehat{B}^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}}}_{(i)} \\
 & \quad + \underbrace{\left\| \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} \widehat{B} U_\ell \widehat{B}^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} - \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} B U_\ell B^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}}}_{(ii)}
 \end{aligned}$$

By Lemma F.4 and F.7 of (Wagenmaker & Jamieson, 2020):

$$\begin{aligned}
 (i) & \leq \max_{\omega \in [0, 2\pi]} 16k^2\gamma^2 \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}} \|(e^{\iota\omega} I - A)^{-1} \widehat{B}\|_{\text{op}}^2 \epsilon \\
 & \leq \max_{\omega \in [0, 2\pi]} 16k^2\gamma^2 \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}^3 (\|B\|_{\text{op}} + \epsilon)^2 \epsilon
 \end{aligned}$$

Note that these lemmas assume that  $U_\ell$  are rank 1, but a trivial modification extends the results to arbitrary  $U_\ell \succeq 0$  satisfying  $\sum_{\ell=1}^k \text{tr}(U_\ell) \leq k^2\gamma^2$ . Further:

$$\begin{aligned}
 (ii) & \leq 2 \left\| \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} \Delta_B U_\ell B^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}} + \left\| \sum_{\ell=1}^k (e^{\iota\omega_\ell} I - A)^{-1} \Delta_B U_\ell \Delta_B^\text{H} (e^{\iota\omega_\ell} I - A)^{-\text{H}} \right\|_{\text{op}} \\
 & \leq (2\|\Delta_B\|_{\text{op}}\|B\|_{\text{op}} + \|\Delta_B\|_{\text{op}}^2) \left( \max_{\omega \in [0, 2\pi]} \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}^2 \right) \sum_{\ell=1}^k \|U_\ell\|_{\text{op}} \\
 & \leq k^2\gamma^2 (2\epsilon\|B\|_{\text{op}} + \epsilon^2) \left( \max_{\omega \in [0, 2\pi]} \|(e^{\iota\omega} I - A)^{-1}\|_{\text{op}}^2 \right)
 \end{aligned}$$

The result is then immediate.  $\square$

**Proposition H.1.** Assume that  $\|A^k\|_{\text{op}} \leq \tau\rho^k$  for all  $k \geq 0$  and that  $\|\Delta\|_{\text{op}} \leq \epsilon$ . Then, for  $k \geq 1$ ,

$$\|(A + \Delta)^k\|_{\text{op}} \leq \tau(\rho + \tau\epsilon)^k, \quad \|(A + \Delta)^k - A^k\|_{\text{op}} \leq \tau(\rho + \tau\epsilon)^k - \tau\rho^k \leq k\tau^2(\rho + \tau\epsilon)^{k-1}\epsilon$$

*Proof.* If  $A$  and  $\Delta$  were scalars, the Binomial Theorem would give:

$$(A + \Delta)^k = \sum_{s=0}^k \binom{k}{s} A^{k-s} \Delta^s$$

As matrix multiplication does not commute, we cannot simply apply the Binomial Theorem. However, we note that, for a fixed  $s$ , we will have  $\binom{k}{s}$  terms of the form

$$A^{n_1} \Delta^{m_1} A^{n_2} \Delta^{m_2} \dots A^{n_s} \Delta^{m_s} A^{n_{s+1}}$$

where  $\sum_{i=1}^{s+1} n_i = k - s$ ,  $\sum_{i=1}^s m_i = s$ . Critically, there will be at most  $s$   $\Delta$  terms in this product. Then, using our assumption on  $\|A^k\|_{\text{op}}$ , we have

$$\begin{aligned} \|A^{n_1} \Delta^{m_1} A^{n_2} \Delta^{m_2} \dots A^{n_s} \Delta^{m_s} A^{n_{s+1}}\|_{\text{op}} &\leq \left( \prod_{i=1}^{s+1} \|A^{n_i}\|_{\text{op}} \right) \left( \prod_{i=1}^s \|\Delta^{m_i}\|_{\text{op}} \right) \\ &\leq \tau^{s+1} \left( \prod_{i=1}^{s+1} \rho^{n_i} \right) \left( \prod_{i=1}^s \epsilon^{m_i} \right) \\ &= \tau^{s+1} \rho^{k-s} \epsilon^s \end{aligned}$$

As this bound does not depend on the specific values of  $n_i, m_i$ , we will have

$$\|(A + \Delta)^k\|_{\text{op}} \leq \sum_{s=0}^k \binom{k}{s} \tau^{s+1} \rho^{k-s} \epsilon^s = \tau(\rho + \tau\epsilon)^k$$

where the final equality holds by the Binomial Theorem. Similarly, note that  $\|(A + \Delta)^k - A^k\|_{\text{op}}$  will behave identically, except that the  $A^k$  term will be removed from the expansion of  $(A + \Delta)^k$ . Thus,

$$\|(A + \Delta)^k - A^k\|_{\text{op}} \leq \sum_{s=1}^k \binom{k}{s} \tau^{s+1} \rho^{k-s} \epsilon^s = \tau(\rho + \tau\epsilon)^k - \tau\rho^k$$

To show the final conclusion, note that, for  $k \geq 1$ , the derivative of  $(a + x)^k$  is  $\frac{d}{dx}(a + x)^k = k(a + x)^{k-1}$ . By the Mean Value Theorem,

$$|(a + x)^k - (a + y)^k| \leq \left( \max_{z \in [x, y]} k|(a + z)^{k-1}| \right) |x - y|$$

Applying this observation in our setting gives that

$$\tau(\rho + \tau\epsilon)^k - \tau\rho^k \leq \tau \left( \max_{z \in [0, \tau\epsilon]} k(\rho + z)^{k-1} \right) \tau\epsilon \leq k\tau^2(\rho + \tau\epsilon)^{k-1} \epsilon$$

□

## H.6. Infinite-Horizon Approximation

**Lemma H.8.** Fix any  $\bar{k}$  and input  $\mathbf{U}^* \in \mathcal{U}_{\gamma^2, \bar{k}}$ . Then for any:

$$k \geq \max \left\{ \frac{4\pi\|\theta\|_{\mathcal{H}_\infty} \gamma^2}{\epsilon}, \frac{\pi}{2\|\theta\|_{\mathcal{H}_\infty}} \right\} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$$

there exists an input  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$  such that:

$$\left\| \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) - \frac{1}{\bar{k}} \Gamma_{\bar{k}}^{\text{freq}}(\theta, \mathbf{U}^*) \right\|_{\text{op}} \leq \epsilon.$$

*Proof.* For simplicity denote  $G_{k, \ell} = (e^{i2\pi\ell/k} I - A)^{-1} B$ . Consider some  $k$  and, given  $\ell \in [1, \bar{k}]$ , let  $\ell_k(\ell) \in [1, k]$  be the index such that  $|\ell_k(\ell)/k - \ell/\bar{k}|$  is minimized. Let  $\ell_k^{-1}(\ell) : \{1, \dots, k\} \rightarrow 2^{\{1, \dots, \bar{k}\}}$  return the set of indices that map to  $\ell$ . Then:

$$\begin{aligned} \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}^*) &= \frac{1}{k^2} \sum_{\ell=1}^{\bar{k}} G_{\bar{k}, \ell} U_{\bar{k}, \ell}^* G_{\bar{k}, \ell}^H = \frac{1}{k^2} \sum_{\ell=1}^{\bar{k}} (G_{k, \ell_k(\ell)} + \Delta_\ell) U_{\bar{k}, \ell}^* (G_{k, \ell_k(\ell)} + \Delta_\ell)^H \\ &= \sum_{\ell=1}^k G_{k, \ell} \left( \frac{1}{k^2} \sum_{\ell' \in \ell_k^{-1}(\ell)} U_{\bar{k}, \ell'}^* \right) G_{k, \ell}^H + \frac{1}{k^2} \sum_{\ell=1}^{\bar{k}} \left( G_{k, \ell_k(\ell)} U_{\bar{k}, \ell}^* \Delta_\ell^H + \Delta_\ell U_{\bar{k}, \ell}^* G_{k, \ell_k(\ell)}^H + \Delta_\ell U_{\bar{k}, \ell}^* \Delta_\ell^H \right) \end{aligned}$$

Set  $U_{k,\ell} = \frac{1}{\bar{k}^2} \sum_{\ell' \in \ell_k^{-1}(\ell)} U_{\bar{k},\ell'}^*$ , denote  $\delta := \max_{\ell \in [1, \dots, \bar{k}]} \|\Delta_\ell\|_2$ , and note that  $\|\theta\|_{\mathcal{H}_\infty} \geq \|G_{k,\ell}$  for all  $k, \ell$ . Then:

$$\begin{aligned} \left\| \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) - \frac{1}{\bar{k}} \Gamma_{\bar{k}}^{\text{freq}}(\theta, \mathbf{U}^*) \right\|_{\text{op}} &\leq \left\| \frac{1}{\bar{k}^2} \sum_{\ell=1}^{\bar{k}} \left( G_{k,\ell_k(\ell)} U_{\bar{k},\ell}^* \Delta_\ell^H + \Delta_\ell U_{\bar{k},\ell}^* G_{k,\ell_k(\ell)}^H + \Delta_\ell U_{\bar{k},\ell}^* \Delta_\ell^H \right) \right\|_{\text{op}} \\ &\leq \frac{2\delta \|\theta\|_{\mathcal{H}_\infty} + \delta^2}{k^2} \sum_{\ell=1}^{\bar{k}} \|U_{\bar{k},\ell}^*\|_{\text{op}} \\ &\leq \frac{2\delta \|\theta\|_{\mathcal{H}_\infty} + \delta^2}{k^2} \sum_{\ell=1}^{\bar{k}} \text{tr}(U_{\bar{k},\ell}^*) \\ &\leq (2\delta \|\theta\|_{\mathcal{H}_\infty} + \delta^2) \gamma^2 \end{aligned}$$

where the final equality holds since  $\tilde{U}_{\bar{k},\ell}^*$  is a feasible input and so must meet the power constraint. By Lemma H.1 of (Wagenmaker & Jamieson, 2020):

$$\|G_{k,\ell} - G_{k',\ell'}\|_2 \leq 2\pi |\ell/k - \ell'/k'| \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$$

Using this we can bound:

$$\|G_{\bar{k},\ell} - G_{k,\ell_k(\ell)}\|_2 \leq 2\pi |\ell/\bar{k} - \ell_k(\ell)/k| \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right) \leq \pi/k \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$$

since any  $x \in [0, 1]$  is at most  $1/(2k)$  from the nearest fraction  $i/k$ . This implies

$$\delta \leq \pi/k \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$$

so:

$$\begin{aligned} \left\| \frac{1}{k} \Gamma_k^{\text{freq}}(\theta, \mathbf{U}) - \frac{1}{\bar{k}} \Gamma_{\bar{k}}^{\text{freq}}(\theta, \mathbf{U}^*) \right\|_{\text{op}} &\leq \frac{2\pi \|\theta\|_{\mathcal{H}_\infty} \gamma^2}{k} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right) \\ &\quad + \frac{\pi^2 \gamma^2}{k^2} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)^2 \\ &\leq \frac{4\pi \|\theta\|_{\mathcal{H}_\infty} \gamma^2}{k} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right) \end{aligned}$$

where the last inequality holds so long as  $k \geq \frac{\pi}{2\|\theta\|_{\mathcal{H}_\infty}} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$ . To make this less than  $\epsilon$ , we must choose:

$$k \geq \frac{4\pi \|\theta\|_{\mathcal{H}_\infty} \gamma^2}{\epsilon} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-2} B\|_{\text{op}} \right)$$

Finally, note that the input  $\tilde{U}_{k,\ell}$  is feasible since:

$$\sum_{\ell=1}^k \text{tr}(U_{k,\ell}) = \frac{1}{\bar{k}^2} \sum_{\ell=1}^k \sum_{\ell' \in \ell_k^{-1}(\ell)} \text{tr}(U_{\bar{k},\ell'}^*) \leq \frac{1}{\bar{k}^2} \sum_{\ell=1}^{\bar{k}} \text{tr}(U_{\bar{k},\ell}^*) \leq \gamma^2$$

The conclusion follows immediately.  $\square$

**Lemma H.9.** *If*

$$T \geq \max \left\{ \frac{8\tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2)}{(1 - \rho^2)^2} \frac{1}{\epsilon}, \log \left( \frac{(1 - \rho^2)^2 \epsilon}{8\tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2)} \right) \frac{1}{2 \log \rho} \right\}$$

then, for any  $T' \geq T$ ,

$$\left\| \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \sigma_u) - \frac{1}{T'} \sum_{t=1}^{T'} \Gamma_t^{\text{noise}}(\theta, \sigma_u) \right\|_{\text{op}} \leq \epsilon.$$

*Proof.* By definition of  $\Gamma_t^{\text{noise}}$ ,

$$\begin{aligned}
 & \left\| \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \sigma_u) - \frac{1}{T'} \sum_{t=1}^{T'} \Gamma_t^{\text{noise}}(\theta, \sigma_u) \right\|_{\text{op}} \\
 &= \left\| \frac{1}{T} \sum_{t=1}^T \sum_{s=0}^{t-1} (\sigma_w^2 A^s (A^s)^\top + \sigma_u^2 A^s B B^\top (A^s)^\top) - \frac{1}{T'} \sum_{t=1}^{T'} \sum_{s=0}^{t-1} (\sigma_w^2 A^s (A^s)^\top + \sigma_u^2 A^s B B^\top (A^s)^\top) \right\|_{\text{op}} \\
 &= \left\| \frac{1}{T} \sum_{s=0}^{T-1} (\sigma_w^2 (T-s) A^s (A^s)^\top + \sigma_u^2 (T-s) A^s B B^\top (A^s)^\top) \right. \\
 &\quad \left. - \frac{1}{T'} \sum_{s=0}^{T'-1} (\sigma_w^2 (T'-s) A^s (A^s)^\top + \sigma_u^2 (T'-s) A^s B B^\top (A^s)^\top) \right\|_{\text{op}} \\
 &= \left\| \sum_{s=0}^{T-1} \left( \sigma_w^2 \left(1 - \frac{s}{T}\right) A^s (A^s)^\top + \sigma_u^2 \left(1 - \frac{s}{T}\right) A^s B B^\top (A^s)^\top \right) \right. \\
 &\quad \left. - \sum_{s=0}^{T'-1} \left( \sigma_w^2 \left(1 - \frac{s}{T'}\right) A^s (A^s)^\top + \sigma_u^2 \left(1 - \frac{s}{T'}\right) A^s B B^\top (A^s)^\top \right) \right\|_{\text{op}} \\
 &= \left\| \sum_{s=0}^{T-1} \left( \left( \frac{s}{T'} - \frac{s}{T} \right) (\sigma_w^2 A^s (A^s)^\top + \sigma_u^2 A^s B B^\top (A^s)^\top) \right) \right. \\
 &\quad \left. - \sum_{s=T}^{T'-1} \left( \sigma_w^2 \left(1 - \frac{s}{T'}\right) A^s (A^s)^\top + \sigma_u^2 \left(1 - \frac{s}{T'}\right) A^s B B^\top (A^s)^\top \right) \right\|_{\text{op}} \\
 &\leq |T'-1 - T^{-1}| \tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2) \sum_{s=0}^{T-1} s \rho^{2s} + (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2) \tau(A, \rho)^2 \sum_{s=T}^{T'-1} (1 - s/T') \rho^{2s} \\
 &\leq |T'-1 - T^{-1}| \tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2) \frac{\rho^2 + \rho^{2T+2T}}{(1 - \rho^2)^2} \\
 &\quad + \tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2) \frac{\rho^{2T'+2} + \rho^{2T+2T} + \rho^{2T} T'}{(1 - \rho^2)^2 T'} \\
 &\leq \frac{4\tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2)}{(1 - \rho^2)^2} (T^{-1} + \rho^{2T})
 \end{aligned}$$

If we set  $T$  large enough such that

$$\frac{4\tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2)}{(1 - \rho^2)^2} T^{-1} \leq \epsilon/2, \quad \frac{4\tau(A, \rho)^2 (\sigma_w^2 + \sigma_u^2 \|B\|_{\text{op}}^2)}{(1 - \rho^2)^2} \rho^{2T} \leq \epsilon/2$$

the desired bound will hold. Rearranging these gives the result.  $\square$

## I. Certainty Equivalence Decision Making with Sequential Open-Loop Policies

In this section we assume we are in the linear dynamical system setting of Section B and that we are playing an exploration policy  $\pi_{\text{exp}}$ .

*Proof of Corollary 2.* This is a direct consequence of Theorem B.3 and Lemma I.1. The stated results follows from some algebra to simplify terms and using Lemma G.4, to upper bound  $\text{poly}(\tau_*, \frac{1}{1-\rho_*})$  terms by  $\text{poly}(\|B_*\|_{\text{op}}, \|A_*\|_{\mathcal{H}_\infty})$ , and setting  $d_\theta = d_x^2 + d_x d_u$ , the dimensionality of  $(A, B)$ .  $\square$



### I.1. Sequential Open-Loop Policies Satisfy Assumption 7

**Lemma I.1.** Any policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$  meets Assumption 7 on some system  $\theta_*$  with

$$\begin{aligned} T_{\text{se}}(\pi_{\text{exp}}) &= c_1 d_x \left( (d_x + d_u) \log(\bar{\gamma} / \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)) + 1) + \log \frac{n}{\delta} \right) \\ \underline{\lambda} &= c_2 \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)) \\ \bar{\Gamma}_T &= \bar{\gamma} \cdot I \\ T_{\text{con}}(\pi_{\text{exp}}) &= \max \left\{ T_{\text{se}}(\pi_{\text{exp}}), (n + \sqrt{d_x} \sigma_w^2 / \gamma^2) \log \frac{2(n+1)^2}{\delta} \right\} \\ C_{\text{con}} &= n \frac{c_3 \tau_*^3 (\sigma_w^2 + \gamma^2)}{(1 - \rho_*)^{5/2} \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u))} \sqrt{\log \frac{n}{\delta} + d_x + d_u} \\ \alpha &= 1/2 \end{aligned}$$

for universal constants  $c_1, c_2, c_3$ .

*Proof of Theorem B.3.* Let:

$$\Sigma_i := \sum_{t=\bar{t}_{i-1}}^{\bar{t}_i-1} z_t z_t^\top, i = 1, \dots, n, \quad \Sigma_{t:t'} := \sum_{s=t}^{t'} z_s z_s^\top$$

**Sufficient Excitation:** We first show that the sufficient excitation condition is met by  $\pi_{\text{exp}}$ . First, note that by Lemma I.3 and some algebra, for any time  $t$ ,  $\Sigma_t \preceq T\bar{\gamma}I$  with probability at least  $1 - \delta/n$ . Fix a time  $t \geq T_{\text{se}}(\pi_{\text{exp}})$  where  $T_{\text{se}}(\pi_{\text{exp}})$  is defined as above. Since  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$ , the low-switching condition implies that there exists some set of epochs  $\{i_1, \dots, i_m\} \subseteq [n]$ , such that for  $j \in [m]$ ,  $\bar{t}_{i_j+1} - \bar{t}_{i_j} \geq \frac{1}{2} T_{\text{se}}(\pi_{\text{exp}})$ ,  $\bar{t}_{i_j+1} \leq t$ , and

$$\sum_{j=1}^m (\bar{t}_{i_j+1} - \bar{t}_{i_j}) \geq \frac{1}{2} t$$

This follows directly from the fact that, for any  $t_0$ , there exists some epoch  $i \in [n]$  such that  $|\{t_0, t_0 + T_{\text{se}}(\pi_{\text{exp}}) - 1\} \cap \{\bar{t}_i, \dots, \bar{t}_{i+1} - 1\}| \geq \frac{1}{2} T_{\text{se}}(\pi_{\text{exp}})$ . Now consider some  $j \in [m]$ . By Lemma I.2, if

$$\bar{t}_{i_j+1} - \bar{t}_{i_j} \geq c_1 d_x \left( (d_x + d_u) \log(\bar{\gamma} / \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \Lambda_{u, i_j})) + 1) + \log \frac{n}{\delta} \right) \quad (\text{I.1})$$

we will have that, with probability at least  $1 - \delta/n$ , for some  $c_2$ ,

$$\Sigma_{\bar{t}_{i_j}: \bar{t}_{i_j+1}} \succeq c_2 (\bar{t}_{i_j+1} - \bar{t}_{i_j}) \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \Lambda_{u, i_j}))$$

However, by definition of  $\Gamma^{\text{noise}}$  and since by assumption  $\lambda_{\min}(\Lambda_{u, i_j}) \geq \sigma_u^2$ , we will have  $\lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \Lambda_{u, i_j})) \geq \lambda_{\text{noise}}^*(\sigma_u)$ , which implies that  $\log(\bar{\gamma} / \lambda_{\text{noise}}^*(\sigma_u) + 1) \geq \log(\bar{\gamma} / \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \Lambda_{u, i_j})) + 1)$ . As we know that  $\bar{t}_{i_j+1} - \bar{t}_{i_j} \geq \frac{1}{2} T_{\text{se}}(\pi_{\text{exp}})$ , it follows that (I.1) is met, so we conclude that with probability at least  $1 - \delta/n$ ,

$$\Sigma_{\bar{t}_{i_j}: \bar{t}_{i_j+1}} \succeq c_2 (\bar{t}_{i_j+1} - \bar{t}_{i_j}) \lambda_{\text{noise}}^*(\sigma_u)$$

Union bounding over this event holding for each  $j \in [m]$ , it follows that with probability at least  $1 - \delta$ ,

$$\Sigma_t \succeq \sum_{j=1}^m \Sigma_{\bar{t}_{i_j}: \bar{t}_{i_j+1}} \succeq \sum_{j=1}^m c_2 (\bar{t}_{i_j+1} - \bar{t}_{i_j}) \lambda_{\text{noise}}^*(\sigma_u) \geq \frac{c_2}{2} t \lambda_{\text{noise}}^*(\sigma_u)$$

By proper choice of constants, we will then have that the sufficient excitation condition of Assumption 7 is met with

$$\begin{aligned} T_{\text{se}}(\pi_{\text{exp}}) &= c_1 d_x \left( (d_x + d_u) \log(\bar{\gamma} / \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)) + 1) + \log \frac{n}{\delta} \right) \\ \underline{\lambda} &= c_2 \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)). \end{aligned}$$

**Concentration of Covariates:** We define the following events for some  $\epsilon_i$  to be specified.

$$\begin{aligned} \mathcal{E} &= \{\|\Sigma_T - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]\|_{\text{op}} \leq \frac{C_{\text{con}}}{T^\alpha} \lambda_{\min}(\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T])\} && \text{(Good event)} \\ \mathcal{E}_1 &= \{\lambda_{\min}(\Sigma_T) \geq c_2 \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u))T\} && \text{(Sufficient excitation)} \\ \mathcal{E}_{2,i} &= \{\|\Sigma_i - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_i]\|_{\text{op}} \leq \epsilon_i\} && \text{(Concentration of covariates)} \\ \mathcal{E}_{3,i} &= \left\{ \|z_{\tilde{t}_{i-1}}\|_2 \leq \sqrt{\frac{c\tau_*^2 \gamma^2 T}{1-\rho_*^2}} \right\} && \text{(Bounded states)} \end{aligned}$$

We would like to show that  $\mathcal{E}$  holds with high probability. The following is trivial.

$$\mathbb{P}[\mathcal{E}^c] \leq \mathbb{P}[\mathcal{E}^c \cap \mathcal{E}_1 \cap (\cap_{i=1}^n \mathcal{E}_{2,i}) \cap (\cap_{i=1}^n \mathcal{E}_{3,i})] + \mathbb{P}[\mathcal{E}_1^c] + \sum_{i=1}^n \mathbb{P}[\mathcal{E}_{3,i} \cap \mathcal{E}_{2,i}^c] + \sum_{i=1}^n \mathbb{P}[\mathcal{E}_{3,i}^c]$$

We first show that  $\mathcal{E}_1, \mathcal{E}_{2,i}, \mathcal{E}_{3,i}$  hold with high probability.

Note first that, by what we have just shown,  $\mathbb{P}[\mathcal{E}_1^c] \leq \delta$  as long as  $T \geq T_{\text{se}}(\pi_{\text{exp}})$ . By Lemma I.8, since  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$  B.4, we will have that  $\mathbb{P}[\mathcal{E}_{3,i}^c] \leq \delta$  as long as

$$T \geq (n + \sqrt{d_x} \sigma_w^2 / \gamma^2) \log \frac{2(n+1)}{\delta} \quad (\text{I.2})$$

Next, we show that  $\mathbb{P}[\mathcal{E}_{3,i} \cap \mathcal{E}_{2,i}^c] \leq \delta$ . Setting

$$\begin{aligned} \epsilon_i &= \left( \frac{c_3 \tau_* \|\Gamma_T^{\text{noise}}(\tilde{\theta}_*, \Lambda_{u,i})\|_{\text{op}} \sqrt{\bar{t}_i - \bar{t}_{i-1}} + c_4 \tau_*^2 (\sqrt{T} \gamma + \sqrt{(c\tau_*^2 \gamma^2 T)(1-\rho^2)})}{1-\rho_*} \max\{\sigma_w, \sqrt{\|\Lambda_u\|_{\text{op}}}\} \right) \\ &\quad \cdot \sqrt{\log \frac{1}{\delta} + d_x + d_u} + \frac{c_5 \max\{\sigma_w^2, \|\Lambda_u\|_{\text{op}}\} \tau_*^2}{(1-\rho_*)^2} (\log \frac{1}{\delta} + d_x + d_u) \end{aligned}$$

Lemma I.4 with  $\rho = \rho_*$  implies directly that  $\mathbb{P}[\mathcal{E}_{3,i} \cap \mathcal{E}_{2,i}^c] \leq \delta$ . For future convenience, we can upper bound  $\epsilon_i$  by

$$\frac{c_3 \tau_*^3 (\sigma_w^2 + \gamma^2) \sqrt{T}}{(1-\rho_*)^{5/2}} \sqrt{\log \frac{1}{\delta} + d_x + d_u}$$

as long as

$$T \geq \log \frac{1}{\delta} + d_x + d_u \quad (\text{I.3})$$

On the event  $\mathcal{E}_1 \cap (\cap_{i=1}^n \mathcal{E}_{2,i}) \cap (\cap_{i=1}^n \mathcal{E}_{3,i})$ , we have

$$\begin{aligned} \|\Sigma_T - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T]\|_{\text{op}} &= \left\| \sum_{i=1}^n \Sigma_i - \sum_{i=1}^n \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_i] \right\|_{\text{op}} \leq \sum_{i=1}^n \|\Sigma_i - \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_i]\|_{\text{op}} \leq \sum_{i=1}^n \epsilon_i \\ &\leq n \frac{c_3 \tau_*^3 (\sigma_w^2 + \gamma^2) \sqrt{T}}{(1-\rho_*)^{5/2}} \sqrt{\log \frac{1}{\delta} + d_x + d_u} \end{aligned}$$

and

$$\mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\Sigma_T] \succeq \mathbb{E}_{\theta_*, \pi_{\text{exp}}}[\mathbb{I}\{\mathcal{E}_1\} \Sigma_T] \geq \mathbb{P}[\mathcal{E}_1] c_2 \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)) T \cdot I \geq \frac{1}{2} \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u)) T \cdot I$$

Thus,  $\mathbb{P}[\mathcal{E}^c \cap \mathcal{E}_1 \cap (\cap_{i=1}^n \mathcal{E}_{2,i}) \cap (\cap_{i=1}^n \mathcal{E}_{3,i})] = 0$  with  $\alpha = 1/2$  and

$$C_{\text{con}} := n \frac{c_3 \tau_*^3 (\sigma_w^2 + \gamma^2)}{(1-\rho_*)^{5/2} \lambda_{\min}(\Gamma_{d_x}^{\text{noise}}(\tilde{\theta}_*, \sigma_u))} \sqrt{\log \frac{1}{\delta} + d_x + d_u}$$

Thus,  $\mathbb{P}[\mathcal{E}^c] \leq 2n + 1$ . Rescaling  $\delta$  and setting  $T_{\text{con}}(\pi_{\text{exp}})$  to guarantee (I.2) and (I.3) hold gives the result.  $\square$

## I.2. Concentration of Covariates

**Lemma I.2** (Lemma E.3 of (Wagenmaker & Jamieson, 2020)). *Assume that our system  $\theta$  is driven by some input  $u_t = \tilde{u}_t + u_t^w$  where  $\tilde{u}_t$  is deterministic and  $u_t^w \sim \mathcal{N}(0, \Lambda_u)$ . Then on the event that  $\sum_{t=1}^T z_t z_t^\top \preceq T\bar{\Gamma}_T$ , for some  $\bar{\Gamma}_T$ , choosing  $k$  so that:*

$$T \geq \frac{25600}{27} k \left( 2(d_x + d_u) \log \frac{200}{3} + \log \det(\bar{\Gamma}_T (\Gamma_k^{\text{noise}})^{-1}) + \log \frac{1}{\delta} \right) \quad (\text{I.4})$$

we will have with probability less than  $\delta$ :

$$\sum_{t=1}^T z_t z_t^\top \not\preceq \frac{27}{25600} T \Gamma_k^{\text{noise}}(\theta, \Lambda_u).$$

**Lemma I.3.** *Assume that we are playing a policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$  and that  $\delta \in (0, 1/3)$ . Then with probability at least  $1 - \delta$ , and assuming we start from some state  $x_0 = 0$ ,*

$$\begin{aligned} \sum_{t=1}^T z_t z_t^\top &\preceq T \frac{c\tau(\tilde{A}_*, \rho)^2}{1-\rho} \left( \sqrt{d_x} \log \frac{T}{\delta} + \frac{\gamma^2}{1-\rho} \right) \cdot I \\ &\preceq cT \left( (1 + \|B_*\|_{\text{op}}^2) \|A_*\|_{\mathcal{H}_\infty}^4 \right) \left( \sqrt{d_x} \log \frac{T}{\delta} + \gamma^2 \|A_*\|_{\mathcal{H}_\infty}^2 \right) \cdot I. \end{aligned}$$

*Proof.* Note that we can break the state into the portion driven by the conditionally non-random input,  $\tilde{u}_t$ , and the process noise and random input. We denote these components as  $z_t^u$  and  $z_t^w$ . Then

$$\left\| \sum_{t=1}^T z_t z_t^\top \right\|_{\text{op}} \leq \sum_{t=1}^T \|z_t\|_2^2 \leq 2 \sum_{t=1}^T (\|z_t^u\|_2^2 + \|z_t^w\|_2^2)$$

Note that the input  $\tilde{u}_t$  will almost surely satisfy  $\sum_{t=0}^{T-1} \tilde{u}_t^\top \tilde{u}_t \leq T\gamma^2$ . We can then apply Lemma I.7 to get that

$$\sum_{t=1}^T \|z_t^u\|_2^2 \leq \frac{4\tau(\tilde{A}_*, \rho)^2 \gamma^2 T}{(1-\rho)^2}$$

To bound the component  $\sum_{t=1}^T \|z_t^w\|_2^2$ , we apply Lemma I.8 with  $\gamma^2 = 0$ , union bounding over all  $T$  steps. We simplify the bound by upper bounding  $n$  by  $T$  and using that  $\delta \in (0, 1/3)$  implies  $\log T/\delta \geq 1$ . The second bound follows by Appendix G.4.  $\square$

**Lemma I.4.** *Consider the system*

$$x_{t+1} = Ax_t + Bu_t + \Lambda_w^{1/2} w_t$$

where  $A \in \mathbb{R}^{d \times d}$ ,  $w_t \sim \mathcal{N}(0, I)$  and  $u_t$  is deterministic and satisfies  $\sum_{t=1}^T u_t^\top u_t \leq T\gamma^2$ . Assume that we start from some state  $x_0$ . Then we will have that, with probability at least  $1 - \delta$

$$\begin{aligned} \left\| \sum_{t=0}^T x_t x_t^\top - \mathbb{E} \sum_{t=0}^T x_t x_t^\top \right\|_{\text{op}} &\leq \frac{c_3 \|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2}{(1-\rho)^2} \left( \log \frac{1}{\delta} + d \right) \\ &+ \left( \frac{c_1 \tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}} \sqrt{T}}{1-\rho} + \frac{c_2 \tau(A, \rho)^2 (\sqrt{T} \gamma \|B\|_{\text{op}} + \|x_0\|_2)}{(1-\rho)^2} \sqrt{\|\Lambda_w\|_{\text{op}}} \right) \sqrt{\log \frac{1}{\delta} + d} \end{aligned}$$

for universal constants  $c_1, c_2, c_3$ .

*Proof.* Consider the systems

$$x_{t+1}^u = Ax_t^u + Bu_t, \quad x_{t+1}^w = Ax_t^w + \Lambda_w^{1/2} w_t$$

and note that  $x_t = x_t^u + x_t^w$ . Therefore,

$$\sum_{t=0}^T x_t x_t^\top = \sum_{t=0}^T x_t^u x_t^u{}^\top + \sum_{t=0}^T \left( x_t^u x_t^w{}^\top + x_t^w x_t^u{}^\top \right) + \sum_{t=0}^T x_t^w x_t^w{}^\top, \quad \mathbb{E} \sum_{t=0}^T x_t x_t^\top = \sum_{t=0}^T x_t^u x_t^u{}^\top + \mathbb{E} \sum_{t=0}^T x_t^w x_t^w{}^\top$$

The second equality is true as  $x_t^u$  is deterministic and  $x_t^w$  is mean 0. Fix some  $v \in \mathcal{S}^{d-1}$ . By Lemma I.5 and Lemma I.6, we'll have, simultaneously with probability  $1 - \delta$ :

$$\left| \sum_{t=1}^T (v^\top x_t^w)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t^w)^2 \right| \leq \frac{2\tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}}{1 - \rho^2} \sqrt{T \log \frac{4}{\delta}} + \frac{8\|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2} \log \frac{4}{\delta}$$

$$\left| \sum_{t=1}^T v^\top x_t^u x_t^w v \right| \leq \frac{\tau(A, \rho)^2 (4\sqrt{T}\gamma \|B\|_{\text{op}} + \|x_0^u\|_2)}{(1 - \rho)^2} \sqrt{2\|\Lambda_w\|_{\text{op}} \log \frac{4}{\delta}}$$

Which implies that

$$\left| \sum_{t=0}^T (v^\top x_t)^2 - \mathbb{E} \sum_{t=0}^T (v^\top x_t)^2 \right| \leq \left| \sum_{t=1}^T (v^\top x_t^w)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t^w)^2 \right| + 2 \left| \sum_{t=1}^T v^\top x_t^u x_t^w v \right|$$

$$\leq \frac{2\tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}}{1 - \rho^2} \sqrt{T \log \frac{4}{\delta}} + \frac{8\|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2} \log \frac{4}{\delta}$$

$$+ \frac{2\tau(A, \rho)^2 (4\sqrt{T}\gamma \|B\|_{\text{op}} + \|x_0\|_2)}{(1 - \rho)^2} \sqrt{2\|\Lambda_w\|_{\text{op}} \log \frac{4}{\delta}}$$

Note that if  $M$  is symmetric  $\|M\|_{\text{op}} = \sup_{v \in \mathcal{S}^{d-1}} |v^\top M v|$ . Fix  $v$  to be a vector for which this equality is attained. Let  $\mathcal{T}$  be an  $\epsilon$ -net of  $\mathcal{S}^{d-1}$ . Then we can then find some  $v_0 \in \mathcal{T}$  such that  $\|v - v_0\|_2 \leq \epsilon$ , and thus,

$$|v^\top M v - v_0^\top M v_0| \leq |v^\top M v - v_0^\top M v| + |v_0^\top M v - v_0^\top M v_0| \leq 2\|M\|_{\text{op}} \|v_0 - v\|_2 \leq 2\epsilon \|M\|_{\text{op}}$$

Therefore,

$$|v_0^\top M v_0| \geq |v^\top M v| - |v^\top M v - v_0^\top M v_0| \leq (1 - 2\epsilon) \|M\|_{\text{op}}$$

so  $\|M\|_{\text{op}} \leq \frac{1}{1-2\epsilon} \max_{v \in \mathcal{T}} |v^\top M v|$ . Applying this in our setting and choosing  $\epsilon = 1/2$ , gives

$$\left\| \sum_{t=0}^T x_t^w x_t^w{}^\top - \mathbb{E} \sum_{t=0}^T x_t^w x_t^w{}^\top \right\|_{\text{op}} \leq 2 \max_{v \in \mathcal{T}} \left| \sum_{t=1}^T (v^\top x_t^w)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t^w)^2 \right|$$

By Corollary 4.2.13 of (Vershynin, 2018), we will have  $|\mathcal{T}| \leq 5^d$ . Using our high probability bound on

$\left| \sum_{t=1}^T (v^\top x_t^w)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t^w)^2 \right|$  given above, and union bounding over  $\mathcal{T}$ , we conclude that, with probability at least  $1 - \delta$

$$\left\| \sum_{t=0}^T x_t x_t{}^\top - \mathbb{E} \sum_{t=0}^T x_t x_t{}^\top \right\|_{\text{op}} \leq \frac{2\tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}}{1 - \rho^2} \sqrt{T \left( \log \frac{4}{\delta} + d \log 5 \right)} + \frac{8\|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2} \left( \log \frac{4}{\delta} + d \log 5 \right)$$

$$+ \frac{2\tau(A, \rho)^2 (4\sqrt{T}\gamma \|B\|_{\text{op}} + \|x_0\|_2)}{(1 - \rho)^2} \sqrt{2\|\Lambda_w\|_{\text{op}} \left( \log \frac{4}{\delta} + d \log 5 \right)}$$

□

**Lemma I.5.** Consider the system

$$x_{t+1} = Ax_t + \Lambda_w^{1/2} w_t$$

where  $A \in \mathbb{R}^{d \times d}$ ,  $w_t \sim \mathcal{N}(0, I)$ , and assume  $x_0 = 0$ . Then, for any  $v \in \mathcal{S}^{d-1}$ , we'll have that, with probability at least  $1 - \delta$

$$\left| \sum_{t=1}^T (v^\top x_t)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t)^2 \right| \leq \frac{2\tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}}{1 - \rho^2} \sqrt{T \log \frac{2}{\delta}} + \frac{8\|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2} \log \frac{2}{\delta}$$

*Proof.* This is a direct consequence of the Hanson-Wright Inequality. Note that,

$$x_t = \sum_{s=0}^{t-1} A^{t-s-1} \Lambda_w^{1/2} w_s = G_t \mathbf{w}_t$$

where we have defined  $G_t := [A^{t-1} \Lambda_w^{1/2}, A^{t-2} \Lambda_w^{1/2}, \dots, A \Lambda_w^{1/2}, \Lambda_w^{1/2}, 0, \dots, 0] \in \mathbb{R}^{d \times dT}$ ,  $\mathbf{w}_T := [w_0^\top, w_1^\top, \dots, w_{T-2}^\top, w_{T-1}^\top]^\top$ . So,

$$(v^\top x_t)^2 = \mathbf{w}_T^\top G_t^\top v v^\top G_t \mathbf{w}_T$$

and,

$$\sum_{t=1}^T (v^\top x_t)^2 = \sum_{t=1}^T \mathbf{w}_T^\top G_t^\top v v^\top G_t \mathbf{w}_T = \mathbf{w}_T^\top \mathbf{G}_T \mathbf{w}_T$$

where  $\mathbf{G}_T := \sum_{t=1}^T G_t^\top v v^\top G_t$ . The Hanson-Wright inequality then immediately gives that,

$$\mathbb{P} \left[ \left| \sum_{t=1}^T (v^\top x_t)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t)^2 \right| \geq \epsilon \right] \leq 2 \exp \left( -c \min \left\{ \frac{\epsilon^2}{\|\mathbf{G}_T\|_F^2}, \frac{\epsilon}{\|\mathbf{G}_T\|_{\text{op}}} \right\} \right)$$

For a fixed  $\delta$ , rearranging gives

$$\mathbb{P} \left[ \left| \sum_{t=1}^T (v^\top x_t)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t)^2 \right| \geq 2\sqrt{\|\mathbf{G}_T\|_F^2 \log(2/\delta)} + 2\|\mathbf{G}_T\|_{\text{op}} \log(2/\delta) \right] \leq \delta$$

We proceed to bound  $\|\mathbf{G}_T\|_{\text{op}}$  and  $\|\mathbf{G}_T\|_F^2$ . Consider some  $u \in \mathcal{S}^{dT-1}$  and note that, if we write  $u = [u_0^\top, u_1^\top, \dots, u_{T-2}^\top, u_{T-1}^\top]^\top$ , where  $u_i \in \mathbb{R}^d$ , using the definition of  $G_t$  given above, we have:

$$G_t u = \sum_{s=0}^{t-1} A^{t-s-1} \Lambda_w^{1/2} u_s = x_t^u$$

where  $x_t^u$  is the state of the system with matrix  $A$  when the input  $u$  is played and there is no noise. Thus,

$$u^\top \mathbf{G}_T u = \sum_{t=1}^T (v^\top G_t u)^2 = \sum_{t=1}^T (v^\top x_t^u)^2 \leq \lambda_{\max} \left( \sum_{t=1}^T x_t^u x_t^{u\top} \right)$$

Then, invoking Lemma I.7 with  $\gamma^2 = \|\Lambda_w\|_{\text{op}}$  and  $B = I$ , we can bound,

$$\lambda_{\max} \left( \sum_{t=1}^T x_t^u x_t^{u\top} \right) \leq \sum_{t=1}^T \|x_t^u\|_2^2 \leq \frac{4\tau(A, \rho)^2 \|\Lambda_w\|_{\text{op}}}{(1-\rho)^2}$$

As this does not depend on  $u$ , it is a valid bound on  $\|\mathbf{G}_T\|_{\text{op}}$ :

$$\|\mathbf{G}_T\|_{\text{op}} \leq \frac{4\tau(A, \rho)^2 \|\Lambda_w\|_{\text{op}}}{(1-\rho)^2} \quad (\text{I.5})$$

To bound  $\|\mathbf{G}_T\|_F^2$ , we can write,

$$\|\mathbf{G}_T\|_F^2 = \text{tr}(\mathbf{G}_T^\top \mathbf{G}_T) = \sum_{t=1}^T \sum_{s=1}^T \text{tr}(G_t^\top v v^\top G_t G_s^\top v v^\top G_s) = \sum_{t=1}^T \sum_{s=1}^T (v^\top G_t G_s^\top v)^2 \leq \sum_{t=1}^T \sum_{s=1}^T \|G_t G_s^\top\|_{\text{op}}^2$$

From the definition of  $G$ , we have,

$$G_t G_s^\top = A^{\max\{t-s, 0\}} \left( \sum_{k=0}^{\min\{t,s\}} A^k \Lambda_w A^{k\top} \right) A^{\max\{s-t, 0\}} = A^{\max\{t-s, 0\}} \Gamma_{\min\{t,s\}}^{\text{noise}}(\theta, 0) A^{\max\{s-t, 0\}}^\top$$

so,

$$\|G_t G_s^\top\|_{\text{op}} \leq \|A^{\max\{t-s, 0\}}\|_{\text{op}} \|A^{\max\{s-t, 0\}}\|_{\text{op}} \|\Gamma_{\min\{t, s\}}^{\text{noise}}(\theta, 0)\|_{\text{op}} \leq \tau(A, \rho) \rho^{|t-s|} \|\Gamma_{\min\{t, s\}}^{\text{noise}}(\theta, 0)\|_{\text{op}}$$

which implies

$$\begin{aligned} \sum_{t=1}^T \sum_{s=1}^T \|G_t G_s^\top\|_{\text{op}}^2 &\leq \tau(A, \rho)^2 \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}^2 \sum_{t=1}^T \sum_{s=1}^T \rho^{2|t-s|} \\ &= \tau(A, \rho)^2 \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}^2 \frac{(1-\rho^4)T + 2\rho^{2(T+1)} - 2\rho^2}{(1-\rho^2)^2} \leq \frac{\tau(A, \rho)^2 \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}^2 T}{(1-\rho^2)^2} \end{aligned}$$

Combining everything, we have shown that, for any  $v \in \mathcal{S}^{d-1}$ ,

$$\mathbb{P} \left[ \left| \sum_{t=1}^T (v^\top x_t)^2 - \mathbb{E} \sum_{t=1}^T (v^\top x_t)^2 \right| \geq \frac{2\tau(A, \rho) \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}} \sqrt{T \log \frac{2}{\delta}}}{1-\rho^2} + \frac{8\|\Lambda_w\|_{\text{op}} \tau(A, \rho)^2 \log \frac{2}{\delta}}{(1-\rho)^2} \right] \leq \delta$$

□

**Lemma I.6.** Consider the systems

$$x_{t+1}^u = Ax_t^u + Bu_t, \quad x_{t+1}^w = Ax_t^w + \Lambda_w^{1/2} w_t$$

where  $A \in \mathbb{R}^{d \times d}$ ,  $w_t \sim \mathcal{N}(0, I)$  and  $u_t$  a deterministic signal with  $\sum_{t=1}^T u_t^\top u_t \leq T\gamma^2$ . Assume that  $x_0^w = 0$ . Then, for any  $v \in \mathcal{S}^{d-1}$ , we will have that, with probability at least  $1 - \delta$

$$\left| \sum_{t=1}^T v^\top x_t^u x_t^w v \right| \leq \frac{\tau(A, \rho)^2 (4\sqrt{T}\gamma \|B\|_{\text{op}} + \|x_0^u\|_2)}{(1-\rho)^2} \sqrt{2\|\Lambda_w\|_{\text{op}} \log \frac{2}{\delta}}.$$

*Proof.* We adopt the same notation as in the proof of Lemma I.5. Defining

$$G_t^u := [A^{t-1}B, A^{t-2}B, \dots, AB, B, 0, \dots, 0] \in \mathbb{R}^{d \times d_u T}, \quad \mathbf{u}_T := [u_0^\top, u_1^\top, \dots, u_{T-2}^\top, u_{T-1}^\top]^\top \in \mathbb{R}^{d_u T}$$

we have

$$x_t^u = G_t^u \mathbf{u}_T + A^t x_0^u, \quad x_t^w = G_t \mathbf{w}_T$$

which implies

$$\sum_{t=1}^T v^\top x_t^u x_t^w v = \left( \mathbf{u}_T^\top \sum_{t=1}^T (G_t^u)^\top v v^\top G_t + x_0^u{}^\top \sum_{t=1}^T A^{t^\top} v v^\top G_t \right) \mathbf{w}_T =: \mathbf{g}^\top \mathbf{w}_T \sim \mathcal{N}(0, \mathbf{g}^\top \mathbf{g})$$

By standard Gaussian concentration results, we then have that

$$\mathbb{P} \left[ \left| \sum_{t=1}^T v^\top x_t^u x_t^w v \right| \geq \sqrt{2\mathbf{g}^\top \mathbf{g} \log \frac{2}{\delta}} \right] \leq \delta$$

It remains to bound  $\mathbf{g}^\top \mathbf{g}$ . To this end, note that

$$\mathbf{g}^\top \mathbf{g} \leq \left( \left\| \mathbf{u}_T^\top \sum_{t=1}^T (G_t^u)^\top v v^\top G_t \right\|_2 + \left\| x_0^u{}^\top \sum_{t=1}^T A^{t^\top} v v^\top G_t \right\|_2 \right)^2$$

We can bound  $\|A^t\|_{\text{op}} \leq \tau(A, \rho) \rho^t$  and,

$$\|G_t\|_{\text{op}} \leq \|\Lambda_w^{1/2}\|_{\text{op}} \sum_{s=0}^{t-1} \|A^s\|_{\text{op}} \leq \|\Lambda_w^{1/2}\|_{\text{op}} \tau(A, \rho) \sum_{s=0}^{t-1} \rho^s \leq \frac{\|\Lambda_w^{1/2}\|_{\text{op}} \tau(A, \rho)}{1-\rho}$$

so,

$$\left\| x_0^{\mathbf{u}\top} \sum_{t=1}^T A^{t\top} v v^\top G_t \right\|_2 \leq \frac{\|x_0^{\mathbf{u}}\|_2 \|\Lambda_w^{1/2}\|_{\text{op}} \tau(A, \rho)^2}{1 - \rho} \sum_{t=1}^T \rho^t \leq \frac{\|x_0^{\mathbf{u}}\|_2 \|\Lambda_w^{1/2}\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2}$$

Furthermore, letting  $G'_t = [A^{t-1}, A^{t-2}, \dots, A, I, 0, \dots, 0]$ , we have

$$\left\| \mathbf{u}_T^\top \sum_{t=1}^T (G'_t)^\top v v^\top G'_t \right\|_2 \leq \|\mathbf{u}_T\|_2 \|B\|_{\text{op}} \|\Lambda_w^{1/2}\|_{\text{op}} \left\| \sum_{t=1}^T (G'_t)^\top v v^\top G'_t \right\|_2 \leq \|\mathbf{u}_T\|_2 \|B\|_{\text{op}} \|\Lambda_w^{1/2}\|_{\text{op}} \|\mathbf{G}'_T\|_{\text{op}}$$

where  $\mathbf{G}'_T = \sum_{t=1}^T (G'_t)^\top v v^\top G'_t$ . By (I.5),  $\|\mathbf{G}'_T\|_{\text{op}} \leq 4\tau(A, \rho)^2 / (1 - \rho)^2$ . Since we have assumed that  $\sum_{t=1}^T u_t^\top u_t \leq T\gamma^2$ , we also have  $\|\mathbf{u}_T\|_2 \leq \sqrt{T}\gamma$ . Combining everything, we have shown that

$$\mathbf{g}^\top \mathbf{g} \leq \|\Lambda_w\|_{\text{op}} \left( \frac{4\sqrt{T}\gamma \|B\|_{\text{op}} \tau(A, \rho)^2}{(1 - \rho)^2} + \frac{\|x_0^{\mathbf{u}}\|_2 \tau(A, \rho)^2}{(1 - \rho)^2} \right)^2$$

Thus,

$$\mathbb{P} \left[ \left| \sum_{t=1}^T v^\top x_t^{\mathbf{u}} x_t^{\mathbf{w}\top} v \right| \geq \frac{\tau(A, \rho)^2 (4\sqrt{T}\gamma \|B\|_{\text{op}} + \|x_0^{\mathbf{u}}\|_2)}{(1 - \rho)^2} \sqrt{2\|\Lambda_w\|_{\text{op}} \log \frac{2}{\delta}} \right] \leq \delta$$

□

### I.3. State Norm Bounds

**Lemma I.7.** Consider the system

$$x_{t+1} = Ax_t + Bu_t$$

and assume that we start at state  $x_0 = 0$ . Then if  $\sum_{t=0}^{T-1} u_t^\top u_t \leq T\gamma^2$ , we will have

$$\sum_{t=1}^T \|x_t\|_2^2 \leq \frac{4\tau(A, \rho)^2 \|B\|_{\text{op}}^2 \gamma^2 T}{(1 - \rho)^2}.$$

*Proof.* By definition  $x_T = \sum_{s=0}^{T-1} A^{T-s-1} B u_s$ , so

$$\sum_{t=1}^T \|x_t\|_2^2 \leq \sum_{t=1}^T \left( \sum_{s=0}^{t-1} \|A^{t-s-1}\|_{\text{op}} \|B\|_{\text{op}} \|u_s\|_2 \right)^2 \leq \tau(A, \rho)^2 \|B\|_{\text{op}}^2 \sum_{t=1}^T \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 \right)^2$$

Letting  $\rho_t = \rho^t$ ,  $v_t = \|u_t\|_2$ , we define  $y_t = \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 = (\rho * v)[t]$ , where  $*$  denotes convolution. By Parseval's Theorem,

$$\sum_{t=1}^T \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 \right)^2 = \sum_{t=1}^T y_t^2 = \frac{1}{T} \sum_{k=1}^T |Y_k|^2$$

where  $Y_k$  denotes the DFT of  $y_t$ . As convolution in the time domain is multiplication in the frequency domain, we will have  $Y_k = P_k V_k$  where  $P_k$  is the DFT of  $\rho_t$  and  $V_k$  is the DFT of  $v_t$ . We can explicitly calculate  $P_k$  as:

$$P_k = \sum_{t=0}^{T-1} \rho^t e^{-i\frac{2\pi kt}{T}} = \frac{1 - e^{-i2\pi k} \rho^T}{1 - e^{-i\frac{2\pi k}{T}} \rho}$$

Thus,

$$\frac{1}{T} \sum_{k=1}^T |Y_k|^2 = \frac{1}{T} \sum_{k=1}^T \left| \frac{1 - e^{-i2\pi k} \rho^T}{1 - e^{-i\frac{2\pi k}{T}} \rho} \right|^2 |V_k|^2$$

Note that, also by Parseval's Theorem, the constraint  $\sum_{t=1}^T \|u_t\|_2^2 \leq \gamma^2$  translates to  $\frac{1}{T} \sum_{k=1}^T |V_k|^2 \leq \gamma^2$ . So,

$$\begin{aligned} \frac{1}{T} \sum_{k=1}^T \left| \frac{1 - e^{-\iota 2\pi k} \rho^T}{1 - e^{-\iota \frac{2\pi k}{T}} \rho} \right|^2 |V_k|^2 &\leq \max_{z: \|z\|_1 \leq T^2 \gamma^2} \frac{1}{T} \sum_{k=1}^T \left| \frac{1 - e^{-\iota 2\pi k} \rho^T}{1 - e^{-\iota \frac{2\pi k}{T}} \rho} \right|^2 z_k \\ &= \gamma^2 T \max_{k \in \{1, \dots, T\}} \left| \frac{1 - e^{-\iota 2\pi k} \rho^T}{1 - e^{-\iota \frac{2\pi k}{T}} \rho} \right|^2 \leq \frac{4\gamma^2 T}{(1 - \rho)^2} \end{aligned}$$

The conclusion follows.  $\square$

**Lemma I.8.** Assume that we are playing a policy  $\pi_{\text{exp}} \in \Pi_{\gamma^2}^{\text{sol}}$ . Then with probability at least  $1 - \delta$ , assuming  $z_0 = 0$ ,

$$\|z_t\|_2^2 \leq \frac{c\tau(\tilde{A}_*, \rho)^2}{1 - \rho^2} \left( \gamma^2 T + (\sqrt{d_x} \sigma_w^2 + \sqrt{nT} \gamma^2) \sqrt{\log \frac{2(n+1)}{\delta}} + (\sigma_w^2 + n\gamma^2) \log \frac{2(n+1)}{\delta} \right)$$

and if

$$T \geq (n + \sqrt{d_x} \sigma_w^2 / \gamma^2) \log \frac{2(n+1)}{\delta}$$

this bound can be simplified to

$$\|z_t\|_2^2 \leq \frac{c\tau(\tilde{A}_*, \rho)^2 \gamma^2 T}{1 - \rho^2}.$$

*Proof.* By Assumption B.4, for  $s$  in epoch  $j$ , we can always write the input  $u_s$  as  $u_s = \tilde{u}_s + u_s^w$ , where  $\tilde{u}_s$  is  $\mathcal{F}_{\tilde{t}_j}$  measurable and  $u_s^w \sim \mathcal{N}(0, \Lambda_{u,j})$ . Given this, we break the state up into the component driven by  $\tilde{u}_s$ , which we denote as  $z_t^u$ , and the component driven by the process noise and  $u_s^w$ , which we denote as  $z_t^w$ . By linearity, we will have that  $z_t = z_t^u + z_t^w$ , so

$$\|z_t\|_2^2 \leq 2\|z_t^u\|_2^2 + 2\|z_t^w\|_2^2$$

We can easily bound  $\|z_t^u\|_2^2$  as:

$$\begin{aligned} \|z_t^u\|_2^2 &= \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} \tilde{B}_* u_s \right\|_2^2 \leq \tau(\tilde{A}_*, \rho)^2 \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s\|_2 \right)^2 \\ &\leq \tau(\tilde{A}_*, \rho)^2 \left( \sum_{s=0}^{t-1} \rho^{2(t-s-1)} \right) \left( \sum_{s=0}^{t-1} \|u_s\|_2^2 \right) \leq \frac{\tau(\tilde{A}_*, \rho)^2 \gamma^2 T}{1 - \rho^2} \end{aligned}$$

where the final inequality follows since, by assumption,  $\sum_{s=0}^{t-1} \|u_s\|_2^2 \leq T\gamma^2$  almost surely. We now bound  $\|z_t^w\|_2^2$ . Note that due to the possible correlations between  $\Lambda_{u,j}$  and previous epochs, we cannot naively apply Gaussian concentration. We first upper bound  $\|z_t^w\|_2^2$  as

$$\begin{aligned} \|z_t^w\|_2^2 &= \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} (\tilde{B}_* u_s^w + w_s) \right\|_2^2 \leq 2 \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} \tilde{B}_* u_s^w \right\|_2^2 + 2 \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} w_s \right\|_2^2 \\ &\leq 2\tau(\tilde{A}_*, \rho)^2 \left( \sum_{s=0}^{t-1} \rho^{t-s-1} \|u_s^w\|_2 \right)^2 + 2 \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} w_s \right\|_2^2 \\ &\leq 2\tau(\tilde{A}_*, \rho)^2 \left( \sum_{s=0}^{t-1} \rho^{2(t-s-1)} \right) \left( \sum_{s=0}^{t-1} \|u_s^w\|_2^2 \right) + 2 \left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} w_s \right\|_2^2 \end{aligned}$$

We note that  $\left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} w_s \right\|_2^2$  is simply the norm of the the state of a dynamical system driven by noise  $w_s$ . We can therefore apply Lemma I.9 to get that with probability at least  $1 - \delta$ ,

$$\left\| \sum_{s=0}^{t-1} \tilde{A}_*^{t-s-1} w_s \right\|_2^2 \leq 4 \sqrt{\|\Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0)\|_F^2 \log \frac{2}{\delta}} + 4 \|\Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0)\|_{\text{op}} \log \frac{2}{\delta}$$



We can upper bound

$$\begin{aligned}\|\Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0)\|_{\text{op}} &= \sigma_w^2 \left\| \sum_{s=0}^{t-1} \tilde{A}_*^t (\tilde{A}_*^s)^\top \right\|_{\text{op}} \leq \sigma_w^2 \tau(\tilde{A}_*, \rho)^2 \sum_{s=0}^{t-1} \rho^{2s} \leq \frac{\sigma_w^2 \tau(\tilde{A}_*, \rho)^2}{1 - \rho^2} \\ \|\Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0)\|_F^2 &\leq d_x \|\Gamma_t^{\text{noise}}(\tilde{\theta}_*, 0)\|_{\text{op}}^2\end{aligned}$$

To bound  $\sum_{s=0}^{t-1} \|u_s^w\|_2^2$  we can apply Hanson-Wright to some epoch  $j$  to get that

$$\begin{aligned}\sum_{s=\bar{t}_j}^{\bar{t}_{j+1}-1} \|u_s^w\|_2^2 &\leq \mathbb{E} \sum_{s=\bar{t}_j}^{\bar{t}_{j+1}-1} \|u_s^w\|_2^2 + 2\sqrt{(\bar{t}_{j+1} - \bar{t}_j) \|\Lambda_{u,j}\|_F^2 \log \frac{2}{\delta}} + 2\|\Lambda_{u,j}\|_{\text{op}} \log \frac{2}{\delta} \\ &= \sum_{s=\bar{t}_j}^{\bar{t}_{j+1}-1} \text{tr}(\Lambda_{u,j}) + 2\sqrt{(\bar{t}_{j+1} - \bar{t}_j) \|\Lambda_{u,j}\|_F^2 \log \frac{2}{\delta}} + 2\|\Lambda_{u,j}\|_{\text{op}} \log \frac{2}{\delta} \\ &\leq (\bar{t}_{j+1} - \bar{t}_j) \gamma^2 + 2\sqrt{(\bar{t}_{j+1} - \bar{t}_j) \gamma^4 \log \frac{2}{\delta}} + 2\gamma^2 \log \frac{2}{\delta}\end{aligned}$$

Assume that  $t$  occurs in epoch  $i$ . Then if this bound holds for all epoch  $j \leq i$ , we can bound

$$\sum_{s=0}^{t-1} \|u_s^w\|_2^2 \leq t\gamma^2 + \sum_{j=0}^i \sqrt{(\bar{t}_{j+1} - \bar{t}_j) \gamma^4 \log \frac{2}{\delta}} + 2i\gamma^2 \log \frac{2}{\delta} \leq t\gamma^2 + \sqrt{it\gamma^4 \log \frac{2}{\delta}} + 2i\gamma^2 \log \frac{2}{\delta}$$

Union bounding over all  $n$  epochs and the bound on the process noise, we then have that with probability at least  $1 - \delta$ ,

$$\|z_t^w\|_2^2 \leq \frac{c\tau(\tilde{A}_*, \rho)^2}{1 - \rho^2} \left( \gamma^2 t + (\sqrt{d_x} \sigma_w^2 + \sqrt{it} \gamma^2) \sqrt{\log \frac{2(n+1)}{\delta}} + (\sigma_w^2 + i\gamma^2) \log \frac{2(n+1)}{\delta} \right)$$

The result follows by combining this with our bound on  $\|z_t^u\|_2^2$ , and upper bounding  $t$  by  $T$  and  $i$  by  $n$ . The simplified bound holds by noting that for large enough  $T$ , we can upper bound the two lower order terms in the bound on  $\|z_t^w\|_2^2$  by  $(d_x \sigma_w^2 + \gamma)T$ .  $\square$

**Lemma I.9.** Consider the system

$$x_{t+1} = Ax_t + \Lambda_w^{1/2} w_t$$

where  $w_t \sim \mathcal{N}(0, I)$  and assume that we start at state  $x_0 = 0$ . Then, with probability at least  $1 - \delta$

$$\|x_T\|_2^2 \leq 2\sqrt{\|\Gamma_T^{\text{noise}}(\theta, 0)\|_F^2 \log \frac{2}{\delta}} + 2\|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}} \log \frac{2}{\delta}.$$

*Proof.* Using the same notation as in the proof of Lemma I.5 and I.6, we will have that  $x_T = G_T \mathfrak{w}_T$ . Applying Hanson-Wright then gives that, with probability at least  $1 - \delta$ ,

$$|x_T^\top x_T - \text{tr}(G_T^\top G_T)| \leq 2\sqrt{\|G_T^\top G_T\|_F^2 \log \frac{2}{\delta}} + 2\|G_T^\top G_T\|_{\text{op}} \log \frac{2}{\delta}$$

By definition of  $G_T$  we have that

$$G_T^\top G_T = \sum_{s=0}^{T-1} \Lambda_w^{1/2} (A^s)^\top A^s \Lambda_w^{1/2}, \quad G_T G_T^\top = \sum_{s=0}^{T-1} A^s \Lambda_w (A^s)^\top = \Gamma_T^{\text{noise}}(\theta, 0)$$

So,

$$\|G_T^\top G_T\|_{\text{op}} = \|G_T G_T^\top\|_{\text{op}} = \|\Gamma_T^{\text{noise}}(\theta, 0)\|_{\text{op}}, \quad \|G_T^\top G_T\|_F^2 = \text{tr}(G_T G_T^\top G_T G_T^\top) = \|\Gamma_T^{\text{noise}}(\theta, 0)\|_F^2$$

This concludes the proof.  $\square$

**Theorem I.1** (Hanson-Wright Inequality, (Vershynin, 2018)). *Let  $X \in \mathbb{R}^d$  be a random vector with independent, mean-zero, sub-Gaussian coordinates. Let  $A \in \mathbb{R}^{d \times d}$ . Then, for every  $\epsilon \geq 0$ , we have*

$$\mathbb{P} [ |X^\top A X - \mathbb{E} X^\top A X| \geq \epsilon ] \leq 2 \exp \left( -c \min \left\{ \frac{\epsilon^2}{K^4 \|A\|_F^2}, \frac{\epsilon}{K^2 \|A\|_{\text{op}}} \right\} \right)$$

where  $K = \max_i \|X_i\|_{\psi_2}$ .

Recall that, if  $X_i$  is gaussian with variance  $\sigma_w^2$ ,  $\|X_i\|_{\psi_2} \leq C \sigma_w$ .

## J. Experiment Design in Linear Dynamical Systems

### J.1. Proof of Theorem B.6

*Proof.* Fix an epoch  $i$  and let  $T = \sum_{j=0}^i T_j$ . Note that, by the definition of  $T_i$ , we will have  $T_i = \frac{1}{2}(T + T_0)$ . Similarly,  $T_{i-1} = \frac{1}{4}(T + T_0)$ . Define the following events.

$$\begin{aligned} \mathcal{E}_1 &= \left\{ \|\hat{\theta}_{i-1} - \theta_\star\|_{\text{op}} \leq C \sqrt{\frac{\log(1/\delta) + (d_x + d_u) \log(\bar{\gamma}/\lambda_{\text{noise}}^\star + 1)}{T_{i-1} \lambda_{\text{noise}}^\star}} =: \epsilon_{\text{op}, i-1} \right\} \\ \mathcal{E}_2 &= \left\{ \mathcal{R}(\mathfrak{a}_{\text{opt}}(\hat{\theta}_i); \theta_\star) \leq 5\sigma_w^2 (\mathcal{H}(\theta_\star) \mathbb{E}_{\theta_\star} [\Sigma_T]^{-1}) \log \frac{24(d_x^2 + d_x d_u)}{\delta} + \frac{C_1}{T^{3/2}} + \frac{C_2}{T^2} \right\} \\ \mathcal{E}_3 &= \left\{ \|z_{T-T_i}\|_2^2 \leq \frac{4\tau_\star^2 \gamma^2 k_i^2}{1 - \rho_\star^{k_i}} + 8\sqrt{\|\tilde{\Gamma}_T^{\text{noise}}\|_F^2 \log \frac{2}{\delta}} + 8\|\tilde{\Gamma}_T^{\text{noise}}\|_{\text{op}} \log \frac{2}{\delta} \right\} \end{aligned}$$

for  $C_1, C_2$  as defined in Corollary 2.

**Events  $\mathcal{E}_i$  hold:** By Lemma J.1, we know that  $\text{TOPLE} \in \Pi_{\gamma^2}^{\text{sol}}$ . By Lemma I.1, this implies that  $\text{TOPLE}$  satisfies Assumption 7 with

$$T_{\text{se}}(\text{TOPLE}) = c_1 d_x \left( (d_x + d_u) \log(\bar{\gamma}/\lambda_{\text{noise}}^\star + 1) + \log \frac{n}{\delta} \right), \quad \underline{\lambda} = c_2 \lambda_{\text{noise}}^\star, \quad \bar{\Gamma}_T = \bar{\gamma} \cdot I$$

Thus, as long as

$$T \geq T_{\text{se}}(\text{TOPLE}) \tag{J.1}$$

we will have that with probability at least  $1 - \delta$ ,  $\lambda_{\min}(\Sigma_T) \geq c_2 \lambda_{\text{noise}}^\star T$  and  $\Sigma_T \preceq T \bar{\gamma} I$ . We can therefore apply Lemma J.5, our operator norm estimation bound<sup>4</sup>, to get that  $\mathbb{P}[\mathcal{E}_1^c] \leq \delta$ . Furthermore, by Corollary 2, we will have, as long as  $T$  is large enough for the burn-in, (B.7) to be met, that  $\mathbb{P}[\mathcal{E}_2^c] \leq \delta$ . To show that  $\mathcal{E}_3$  occurs with high probability, we break up the state into two components:  $z_t^u$ , the portion of the state driven by  $\tilde{u}_t$ , and  $z_t^w$ , the portion of the state driven by the input noise and process noise. As the structure of  $\text{TOPLE}$  is identical to that of the algorithm considered in (Wagenmaker & Jamieson, 2020), Lemma D.7 of (Wagenmaker & Jamieson, 2020) gives that

$$\|z_{T-T_i}^u\|_2^2 \leq \frac{4\tau(\tilde{A}_\star, \rho)^2 k_i^2 \gamma^2}{(1 - \rho^{k_i})^2}$$

and we choose  $\rho = \rho_\star$ . Note that while this result is stated as a high-probability bound, since we are only considering the non-random portion of the input, it will hold deterministically. Crucially for subsequent steps, this scales as  $k_i^2$  instead of  $T$ , which is the scaling we would obtain applying Lemma I.8 would scale. Next, applying Lemma I.9 gives that, with probability  $1 - \delta$ ,

$$\|z_{T-T_i}^w\|_2^2 \leq 4\sqrt{\|\tilde{\Gamma}_T^{\text{noise}}\|_F^2 \log \frac{2}{\delta}} + 4\|\tilde{\Gamma}_T^{\text{noise}}\|_{\text{op}} \log \frac{2}{\delta}$$

Note that we can apply Lemma I.9 since the input noise variance is deterministically fixed for all epochs, and by upper bounding the state bound for epochs  $i \geq 1$  by the state bound that would hold if we always set the input noise to have variance  $\gamma^2/d_u$ . This implies  $\mathbb{P}[\mathcal{E}_3^c] \leq \delta$ . Altogether then, we have that  $\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \geq 1 - 3\delta$ .

<sup>4</sup>Note that we could have instead employed Lemma F.1 to upper bound  $\|\hat{\theta}_{i-1} - \theta_\star\|_F$ . By exploiting the matrix structure of  $\theta_\star$  and using an operator norm bound instead, we are able to save a factor of dimensionality in the burn-in time.

**Events  $\mathcal{E}_i$  imply optimal inputs:** We now assume that  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$  holds. Assume that  $T_i$  is large enough that

$$\epsilon_{\text{op},i-1} \leq \min\{r_{\text{cov}}(\theta_\star), r_{\text{quad}}(\theta_\star)/\sqrt{d_x}, \lambda_{\text{noise}}^\star/(4L_{\text{cov}}(\theta_\star, \gamma^2))\} \quad (\text{J.2})$$

Then, as long as,

$$T_i \geq \text{poly}\left(\frac{1}{1-\rho_\star}, \tau_\star, \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty}\right) \frac{d_u \|z_{T-T_i}\|_2^2 + d_u^{3/2} \gamma^2 k_i \sqrt{T_i} + d_u^2 \gamma^2 k_i^2 + \sigma_w^2}{\lambda_{\text{noise}}^\star} \quad (\text{J.3})$$

$$k_i \geq \max\left\{\frac{80\pi \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty} \gamma^2}{\lambda_{\text{noise}}^\star}, \frac{\pi}{2\|\tilde{\theta}_\star\|_{\mathcal{H}_\infty}}\right\} \left(\max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_\star)^{-2} \tilde{B}\|_{\text{op}}\right) \quad (\text{J.4})$$

we can apply Lemma J.7, which gives that the performance achieved by  $U_i$  is nearly optimal. That is, for any  $T' \geq T_i$ ,

$$\text{tr}(\mathcal{H}(\theta_\star)(\mathbb{E}_{\theta_\star}[\Sigma_T])^{-1}) \leq \text{tr}(\mathcal{H}(\theta_\star)(\mathbb{E}_{\theta_\star}[I_{d_x} \otimes \sum_{t=T-T_i}^T z_t z_t^\top])^{-1}) \leq \min_{\mathbf{u} \in \mathcal{U}_{\gamma^2, T'}} \frac{3\text{tr}\left(\mathcal{H}(\theta_\star)\Gamma_{T', T'}^{\text{ss}}(\tilde{\theta}_\star, \mathbf{u}, 0)^{-1}\right)}{T_i} + 2C_{\text{exp}}$$

where

$$C_{\text{exp}} = \left(\frac{4\sqrt{d_x}(d_x^2 + d_x d_u)L_{\text{hess}}}{\lambda_{\text{noise}}^\star} + \frac{8L_{\text{cov}}(\theta_\star, \gamma^2)\text{tr}(\mathcal{H}(\theta_\star))}{(\lambda_{\text{noise}}^\star)^2}\right) \frac{\epsilon_{\text{op},i-1}}{T_i} + \frac{8(d_x^2 + d_x d_u)L_{\text{cov}}(\theta_\star, \gamma^2)L_{\text{hess}}}{(\lambda_{\text{noise}}^\star)^2} \frac{\epsilon_{\text{op},i-1}^2}{T_i}.$$

and we have chosen  $\epsilon$  such that  $1/(1-\epsilon)^3 = 3/2$ . Recall that Algorithm 4 uses  $T_i = C_{\text{init}} d_u 2^i$  and  $k_i = C_{\text{init}} 2^{\lfloor i/4 \rfloor}$ . We then have that,

$$\frac{C_{\text{init}}^{3/4}}{d_u^{1/4}} T_i^{1/4} \geq C_{\text{init}} 2^{i/4} \geq k_i \geq C_{\text{init}} 2^{i/4-1} = \frac{C_{\text{init}}^{3/4}}{2d_u^{1/4}} T_i^{1/4}$$

On event  $\mathcal{E}_3$ , which upper bounds  $\|z_{T-T_i}\|_2^2$ , it follows that (J.3) and (J.4) hold as long as

$$T_i \geq \frac{\text{poly}\left(\frac{1}{1-\rho_\star}, \tau_\star, \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty}\right)}{\lambda_{\text{noise}}^\star} \left(d_u^{5/4} \gamma^2 C_{\text{init}}^{3/4} T_i^{3/4} + d_u^{3/2} \gamma^2 C_{\text{init}}^{3/2} \sqrt{T_i} + \sigma_w^2 + \sqrt{\|\Gamma_{T-T_i}^{\text{noise}}\|_F^2 \log \frac{2}{\delta}} + \|\Gamma_{T-T_i}^{\text{noise}}\|_{\text{op}} \log \frac{2}{\delta}\right) \quad (\text{J.5})$$

$$T_i^{1/4} \geq \max\left\{\frac{80\pi \|\tilde{\theta}_\star\|_{\mathcal{H}_\infty} \gamma^2}{\lambda_{\text{noise}}^\star}, \frac{\pi}{2\|\tilde{\theta}_\star\|_{\mathcal{H}_\infty}}\right\} \frac{2d_u^{1/4} \left(\max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_\star)^{-2} \tilde{B}\|_{\text{op}}\right)}{C_{\text{init}}^{3/4}} \quad (\text{J.6})$$

We have then shown that, on the event  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$  and assuming  $T$  is large enough to meet the burn-ins stated above, we have, for any  $T'$ ,

$$\begin{aligned} \mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_i); \theta_\star) &\leq 5\sigma_w^2 (\mathcal{H}(\theta_\star)\mathbb{E}_{\theta_\star}[\Sigma_T])^{-1} \log \frac{18(d_x^2 + d_x d_u)}{\delta} + \frac{C_1}{T^{3/2}} + \frac{C_2}{T^2} \\ &\leq 15\sigma_w^2 \min_{\mathbf{u} \in \mathcal{U}_{\gamma^2, T'}} \frac{\text{tr}\left(\mathcal{H}(\theta_\star)\Gamma_{T', T'}^{\text{ss}}(\tilde{\theta}_\star, \mathbf{u}, 0)^{-1}\right)}{T_i} \log \frac{24(d_x^2 + d_x d_u)}{\delta} + \frac{C_1}{T^{3/2}} + \frac{C_2}{T^2} \\ &\quad + 10\sigma_w^2 C_{\text{exp}} \log \frac{24(d_x^2 + d_x d_u)}{\delta} \end{aligned}$$

As this bounds hold for any  $T' \geq T_i$ , we take  $\liminf_{T' \rightarrow \infty}$ , to obtain

$$\mathcal{R}(\mathbf{a}_{\text{opt}}(\hat{\theta}_i); \theta_\star) \leq \frac{15\sigma_w^2 \Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star)}{T_i} \log \frac{24(d_x^2 + d_x d_u)}{\delta} + \frac{C_1}{T^{3/2}} + \frac{C_2}{T^2} + 10\sigma_w^2 C_{\text{exp}} \log \frac{24(d_x^2 + d_x d_u)}{\delta}$$

We can also upper bound  $\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star)$  by  $16\Phi_{\text{opt}}(\gamma^2; \theta_\star)$  via Lemma G.3. On  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_4$ , it is easy to see  $C_{\text{exp}} = C_4/T^{3/2} + C_5/T^2$  for some  $C_4, C_5$ . The conclusion then follows by rescaling  $\delta$  by a factor of 3, and since  $T_i \geq T/2$ . The fact that the average expected power of the inputs is bounded by  $\gamma^2$  follows by Lemma J.2. Finally, some algebra shows that the burn-in times stated above are all met as long as Assumption 8 holds.  $\square$

**Lemma J.1.**  $\text{TOPLE} \in \Pi_{\gamma^2}^{\text{sol}}$  with  $\sigma_u = \frac{\gamma}{\sqrt{2d_u}}$  and  $n = \mathcal{O}(\log T)$ .

*Proof.* This follows directly by the formal definition of TOPLE, Algorithm 4. In particular, we see that at each epoch  $i$ , TOPLE plays open-loop inputs  $\tilde{u}_t$  that are  $\mathcal{F}_{\bar{t}_i}$  measurable. Furthermore,  $\bar{t}_i$  and  $\Lambda_{u,i}$  are deterministically specified at the start of the algorithm,  $\text{tr}(\Lambda_{u,i}) \leq \gamma^2$ ,  $\lambda_{\min}(\Lambda_{u,i}) \geq \gamma^2/(2d_u)$ , and Lemma J.2 gives  $\sum_{t=0}^{T-1} \tilde{u}_t^\top \tilde{u}_t \leq \gamma^2$  deterministically. The fact that  $n = \mathcal{O}(\log T)$  follows since we increase the epoch length exponentially. Finally, the low-switching condition follows since the length of the epochs increase exponentially—once  $T$  is large enough that  $T \geq T_{\text{se}}(\pi_{\text{exp}})$ , we will have that at least half the initial interval is contained in the final epoch. Then for subsequent epochs, any interval of length  $T_{\text{se}}(\pi_{\text{exp}})$  will contain at most one epoch boundary.  $\square$

**Lemma J.2.** Running Algorithm 4, we will have  $\frac{1}{T} \sum_{t=1}^T \mathbb{E}[u_t^\top u_t] \leq \gamma^2$ .

*Proof.* By Proposition B.5, we have that  $\sum_{t=1}^{T_i} (\tilde{u}_t^i)^\top \tilde{u}_t^i \leq T_i \gamma^2/2$ . Thus,

$$\sum_{t=1}^{T_i} \mathbb{E}[u_t^\top u_t] \leq T_i \gamma^2/2 + \sum_{t=1}^{T_i} \mathbb{E}[(u_t^w)^\top u_t^w] = T_i \gamma^2/2 + \sum_{t=1}^{T_i} \gamma^2/2 = T_i \gamma^2$$

Thus, the average expected input power for a given epoch is bounded by  $\gamma^2$ . It follows then that, after running for  $i$  epochs,

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}[u_t^\top u_t] = \frac{1}{T} \sum_{j=1}^i \sum_{t=1}^{T_j} \mathbb{E}[u_{t,j}^\top u_{t,j}] \leq \frac{1}{T} \sum_{j=1}^i T_j \gamma^2 = \gamma^2$$

where the last equality follows since, by definition,  $T = \sum_{j=1}^i T_j$ .  $\square$

## J.2. Certainty Equivalence Experiment Design

**Lemma J.3.** Fix a nominal instance  $\theta_\star$  and let  $\hat{\theta}$  be some instance such that  $\|\theta_\star - \hat{\theta}\|_\circ \leq \epsilon_\circ$ , for  $\circ \in \{\text{op}, 2\}$ . Let  $\Gamma(\theta, \mathbf{U}) \in \mathcal{S}_+^{d_\theta}$  be a map that satisfies, for all  $\mathbf{U} \in \mathcal{U}_{\gamma^2}$  and all  $\theta$  with  $\|\theta - \theta_\star\|_\circ \leq r_{\text{cov}}(\theta_\star)$ ,

$$\|\Gamma(\theta, \mathbf{U}) - \Gamma(\theta_\star, \mathbf{U})\|_{\text{op}} \leq L_{\text{cov}}(\theta_\star, \gamma^2) \cdot \|\theta - \theta_\star\|_\circ \quad (\text{J.7})$$

Assume that for all  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$ ,  $\lambda_{\min}(\Gamma(\theta_\star, \mathbf{U})) \geq \underline{\lambda}$  and that  $\epsilon_\circ < \min\{r_{\text{cov}}(\theta_\star), \underline{\lambda}/(2L_{\text{cov}}(\theta_\star, \gamma^2))\}$ ,  $\epsilon_2 \leq r_{\text{quad}}(\theta_\star)$ . Let:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, k}} \text{tr} \left( \mathcal{H}(\hat{\theta}) \Gamma(\hat{\theta}, \mathbf{U})^{-1} \right), \quad \mathbf{U}^\star = \arg \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, k}} \text{tr} \left( \mathcal{H}(\theta_\star) \Gamma(\theta_\star, \mathbf{U})^{-1} \right)$$

Then, under Assumption 3, we have:

$$\left| \text{tr} \left( \mathcal{H}(\theta_\star) \Gamma(\theta_\star, \mathbf{U}^\star)^{-1} \right) - \text{tr} \left( \mathcal{H}(\theta_\star) \Gamma(\theta_\star, \hat{\mathbf{U}})^{-1} \right) \right| \leq \frac{2d_\theta L_{\text{hess}}}{\underline{\lambda}} \epsilon_2 + \frac{4L_{\text{cov}}(\theta_\star, \gamma^2) \text{tr}(\mathcal{H}(\theta_\star))}{\underline{\lambda}^2} \epsilon_\circ + \frac{4d_\theta L_{\text{cov}}(\theta_\star, \gamma^2) L_{\text{hess}}}{\underline{\lambda}^2} \epsilon_\circ \epsilon_2.$$

*Proof.* By Proposition E.2, under Assumption 3 and since  $\|\theta_\star - \hat{\theta}\|_2 \leq r_{\text{quad}}(\theta_\star)$ , we have

$$\|\mathcal{H}(\theta_\star) - \mathcal{H}(\hat{\theta})\|_{\text{op}} \leq L_{\text{hess}} \|\theta_\star - \hat{\theta}\|_2$$

Furthermore, by Lemma J.4, (J.7), and since  $\epsilon_\circ < \underline{\lambda}/(2L_{\text{cov}}(\theta_\star, \gamma^2))$ ,

$$\|\Gamma(\hat{\theta}, \mathbf{U})^{-1} - \Gamma(\theta_\star, \mathbf{U})^{-1}\|_{\text{op}} \leq \frac{\|\Gamma(\hat{\theta}, \mathbf{U}) - \Gamma(\theta_\star, \mathbf{U})\|_{\text{op}}}{\underline{\lambda}(\underline{\lambda} - L_{\text{cov}}(\theta_\star, \gamma^2)\epsilon_\circ)} \leq \frac{2L_{\text{cov}}(\theta_\star, \gamma^2)\epsilon_\circ}{\underline{\lambda}^2}$$

Thus, denoting  $\Delta_{\mathcal{H}} = \mathcal{H}(\hat{\theta}) - \mathcal{H}(\theta_\star)$  and  $\Delta_{\Sigma^{-1}} = \Gamma(\theta_\star, \mathbf{U})^{-1} - \Gamma(\hat{\theta}, \mathbf{U})^{-1}$ , the above bounds and Von Neumann's trace inequality imply:

$$\left| \text{tr} \left( \mathcal{H}(\hat{\theta}) \Gamma(\hat{\theta}, \mathbf{U})^{-1} \right) - \text{tr} \left( \mathcal{H}(\theta_\star) \Gamma(\theta_\star, \mathbf{U})^{-1} \right) \right| \leq |\text{tr}(\Delta_{\mathcal{H}} \Gamma(\theta_\star, \mathbf{U})^{-1})| + |\text{tr}(\mathcal{H}(\theta_\star) \Delta_{\Sigma^{-1}})| + |\text{tr}(\Delta_{\mathcal{H}} \Delta_{\Sigma^{-1}})|$$

$$\begin{aligned}
 &\leq L_{\text{hess}} \text{tr}(\Gamma(\theta_*, \mathbf{U})^{-1}) \epsilon_2 + \frac{2L_{\text{cov}}(\theta_*, \gamma^2) \text{tr}(\mathcal{H}(\theta_*))}{\lambda^2} \epsilon_o + \frac{2d_\theta L_{\text{cov}}(\theta_*, \gamma^2) L_{\text{hess}}}{\lambda^2} \epsilon_o \epsilon_2 \\
 &\leq \frac{d_\theta L_{\text{hess}}}{\lambda} \epsilon_2 + \frac{2L_{\text{cov}}(\theta_*, \gamma^2) \text{tr}(\mathcal{H}(\theta_*))}{\lambda^2} \epsilon_o + \frac{2d_\theta L_{\text{cov}}(\theta_*, \gamma^2) L_{\text{hess}}}{\lambda^2} \epsilon_o \epsilon_2 \\
 &=: f(\epsilon)
 \end{aligned}$$

Assume that  $\text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \hat{\mathbf{U}})^{-1}) > \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1})$ , then:

$$\begin{aligned}
 \left| \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1}) - \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \hat{\mathbf{U}})^{-1}) \right| &\leq \left| \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1}) - \text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \hat{\mathbf{U}})^{-1}) \right| \\
 &\quad + \left| \text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \hat{\mathbf{U}})^{-1}) - \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \hat{\mathbf{U}})^{-1}) \right| \\
 &\leq \left| \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1}) - \text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \mathbf{U}^*)^{-1}) \right| \\
 &\quad + \left| \text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \hat{\mathbf{U}})^{-1}) - \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \hat{\mathbf{U}})^{-1}) \right| \\
 &\leq 2f(\epsilon)
 \end{aligned}$$

where the second inequality holds because  $\hat{\mathbf{U}}$  is the minimizer of  $\text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \mathbf{U})^{-1})$ . If instead  $\text{tr}(\mathcal{H}(\hat{\theta})\Gamma(\hat{\theta}, \hat{\mathbf{U}})^{-1}) \leq \text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1})$ , we can replace  $\text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \mathbf{U}^*)^{-1})$  with  $\text{tr}(\mathcal{H}(\theta_*)\Gamma(\theta_*, \hat{\mathbf{U}})^{-1})$  in the above calculation to get the same result. The conclusion follows.  $\square$

**Lemma J.4** (Matrix Perturbation Bound). *Assume  $A, B \in \mathbb{S}_{++}^d$ ,  $\|A - B\|_{\text{op}} \leq \epsilon$ , and  $\epsilon < \lambda_{\min}(B)$ . Then*

$$\|A^{-1} - B^{-1}\|_{\text{op}} \leq \frac{\epsilon}{\lambda_{\min}(B)(\lambda_{\min}(B) - \epsilon)}$$

*Proof.* Denote  $\Delta = A - B$ . By the matrix inversion lemma:

$$A^{-1} = (B + \Delta)^{-1} = B^{-1} - B^{-1}(B^{-1} + \Delta^{-1})^{-1}B^{-1}$$

so:

$$\begin{aligned}
 \|A^{-1} - B^{-1}\|_{\text{op}} &= \|B^{-1}(B^{-1} + \Delta^{-1})^{-1}B^{-1}\|_{\text{op}} \\
 &\leq \|B^{-1}\|_{\text{op}}^2 \|(B^{-1} + \Delta^{-1})^{-1}\|_{\text{op}} \\
 &= \frac{1}{\lambda_{\min}(B)^2 \sigma_d(B^{-1} + \Delta^{-1})}
 \end{aligned}$$

However:

$$\sigma_d(B^{-1} + \Delta^{-1}) \geq \sigma_d(\Delta^{-1}) - \sigma_1(B^{-1}) = \frac{1}{\|\Delta\|_{\text{op}}} - \frac{1}{\lambda_{\min}(B)} = \frac{\lambda_{\min}(B) - \|\Delta\|_{\text{op}}}{\lambda_{\min}(B)\|\Delta\|_{\text{op}}}$$

Since we have assumed  $\epsilon < \lambda_{\min}(B)$  and since  $\|\Delta\|_{\text{op}} \leq \epsilon$ , this lower bound on  $\sigma_d(B^{-1} + \Delta^{-1})$  will be positive, so:

$$\frac{1}{\lambda_{\min}(B)^2 \sigma_d(B^{-1} + \Delta^{-1})} \leq \frac{\lambda_{\min}(B)\|\Delta\|_{\text{op}}}{\lambda_{\min}(B)^2(\lambda_{\min}(B) - \|\Delta\|_{\text{op}})} \leq \frac{\|\Delta\|_{\text{op}}}{\lambda_{\min}(B)(\lambda_{\min}(B) - \|\Delta\|_{\text{op}})}$$

The result follows since  $\|\Delta\|_{\text{op}} \leq \epsilon$ .  $\square$

### J.3. Operator Norm Estimation

**Lemma J.5.** *Let*

$$\hat{\theta}_{\text{ls}} = \min_{A, B} \sum_{t=1}^T \|x_{t+1} - Ax_t - Bu_t\|_2^2$$

Then on the event

$$\mathcal{E} := \left\{ \lambda_{\min}(\Sigma_T) \geq \underline{\lambda}T, \Sigma_T \preceq T\bar{\Gamma}_T \right\}$$

with probability at least  $1 - \delta$ :

$$\|\hat{\theta}_{\text{ls}} - \theta_\star\|_{\text{op}} \leq C \sqrt{\frac{\log(1/\delta) + d_x + \log \det(\bar{\Gamma}_T/\underline{\lambda} + I)}{\underline{\lambda}T}}.$$

*Proof.* Define the following events:

$$\begin{aligned} \mathcal{A} &= \left\{ \|\hat{\theta}_i - \theta_\star\|_{\text{op}} \leq C \sqrt{\frac{\log(1/\delta) + d_x + \log \det(\bar{\Gamma}_T/\underline{\lambda} + I)}{\underline{\lambda}T}} \right\} \\ \mathcal{E}_1 &= \left\{ \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} \leq c_2 \sigma_w \sqrt{\log \frac{1}{\delta} + d_x + \log \det(\bar{\Gamma}_T/\underline{\lambda} + I)} \right\} \end{aligned}$$

Our goal is to show that  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \delta$ . The following is trivial.

$$\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \mathbb{P}[\mathcal{A}^c \cap \mathcal{E} \cap \mathcal{E}_1] + \mathbb{P}[\mathcal{E} \cap \mathcal{E}_1^c]$$

As  $\hat{\theta}_{\text{ls}}$  is the least squares estimate, we will have that  $\hat{\theta}_{\text{ls}}^\top = (\sum_{t=1}^T z_t z_t^\top)^{-1} \sum_{t=1}^T z_t x_{t+1}^\top = \theta_\star^\top + (\sum_{t=T-T_i}^T z_t z_t^\top)^{-1} \sum_{t=1}^T z_t w_t^\top$ . Given this, the error can be decomposed as:

$$\begin{aligned} \|\hat{\theta}_{\text{ls}} - \theta_\star\|_{\text{op}} &= \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} \leq \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \right\|_{\text{op}} \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} \\ &= \left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} / \sqrt{\lambda_{\min} \left( \sum_{t=1}^T z_t z_t^\top \right)} \end{aligned}$$

It follows that, on the event  $\mathcal{E} \cap \mathcal{E}_1$ , the error bound given in  $\mathcal{A}$  holds. Thus,  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E} \cap \mathcal{E}_1] = 0$ . Lemma J.6 implies that  $\mathbb{P}[\mathcal{E} \cap \mathcal{E}_1^c] \leq \delta$ , so  $\mathbb{P}[\mathcal{A}^c \cap \mathcal{E}] \leq \delta$ .  $\square$

**Lemma J.6** (Lemma E.6 of (Wagenmaker & Jamieson, 2020), see also (Abbasi-Yadkori et al., 2011)). *Assume that  $z_t$  is generated by (G.5) with  $w_t \sim \mathcal{N}(0, \sigma_w^2 I)$  and input  $u_t = \tilde{u}_t + u_t^w$ , where  $\tilde{u}_t$  is  $\mathcal{F}_{t-1}$  measurable and  $u_t^w \sim \mathcal{N}(0, \Lambda_u)$ . On the event that  $V_+ \succeq \sum_{t=1}^T z_t z_t^\top \succeq V_-$ , we will have that, with probability less than  $\delta$ :*

$$\left\| \left( \sum_{t=1}^T z_t z_t^\top \right)^{-1/2} \sum_{t=1}^T z_t w_t^\top \right\|_{\text{op}} > \sigma_w \sqrt{16 \log \frac{1}{\delta} + 8 \log \det(V_+ V_-^{-1} + I) + 16(d_x + d_u) \log 5}.$$

#### J.4. Optimality of Inputs

**Lemma J.7.** *Fix an epoch  $i$  of Algorithm 4 and let  $k_i$  denote the discretization level of the input set at that epoch,  $\mathcal{U}_{\gamma^2/2, k_i}$ , and assume that  $T_i/(d_u k_i)$  is an integer. Let  $k$  be any value satisfying  $d_x \leq k \leq T_i/2$  and take  $\epsilon \in (0, 1)$ . Then, as long as*

$$T_i \geq \max \left\{ \frac{2d_u \tau_\star^3 (2k_i \gamma + \|z_{T-T_i}\|_2) ((4 + 2\tau_\star) k_i \gamma + \tau_\star \|z_{T-T_i}\|_2)}{(1 - \rho_\star^{k_i})^2 (1 - \rho_\star) \lambda_{\text{noise}}^\star \epsilon}, \frac{16\tau_\star^2 (\sigma_w^2 + \gamma^2 / (2d_u))}{(1 - \rho_\star^2)^2 \lambda_{\text{noise}}^\star \epsilon}, \left( \frac{\tau_\star^2 k_i^2 d_u \gamma^2}{(1 - \rho_\star^{k_i})^2} + \frac{\tau_\star k_i \sqrt{d_u} \gamma^2 \sqrt{T_i}}{1 - \rho_\star^{k_i}} \right) \frac{16d_u \|\tilde{\theta}_\star\|_{\tilde{\mathcal{H}}_\infty}^2}{\lambda_{\text{noise}}^\star \epsilon}, \log \left( \frac{(1 - \rho_\star^2)^2 \lambda_{\text{noise}}^\star \epsilon}{16\tau_\star^2 (\sigma_w^2 + \gamma^2 / (2d_u))} \right) \frac{1}{2 \log \rho_\star} \right\},$$

$$k_i \geq \max \left\{ \frac{8\pi \|\tilde{\theta}_*\|_{\mathcal{H}_\infty} \gamma^2}{\lambda_{\text{noise}}^* \epsilon}, \frac{\pi}{2\|\tilde{\theta}_*\|_{\mathcal{H}_\infty}} \right\} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_*)^{-2} \tilde{B}\|_{\text{op}} \right),$$

$$\|\hat{\theta}_{i-1} - \theta_*\|_{\text{op}} \leq \epsilon_{\text{op}, i-1} \leq \min \left\{ r_{\text{cov}}(\theta_*), \frac{r_{\text{quad}}(\theta_*)}{\sqrt{d_x}}, \frac{\lambda_{\text{noise}}^*}{4L_{\text{cov}}(\theta_*, \gamma^2)} \right\}$$

we have, for any  $T' \geq T_i$ ,

$$\text{tr} \left( \mathcal{H}(\theta_*) (\mathbb{E}_{\theta_*, \mathbf{U}_i, T_i} [I_{d_x} \otimes \sum_{t=T-T_i}^T z_t z_t^\top])^{-1} \right) \leq \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \frac{2\text{tr} \left( \mathcal{H}(\theta_*) \mathbf{\Gamma}_{T', T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0)^{-1} \right)}{(1-\epsilon)^3 T_i} + \frac{C_{\text{exp}}}{(1-\epsilon)^2}$$

where

$$C_{\text{exp}} = \left( \frac{4\sqrt{d_x}(d_x^2 + d_x d_u) L_{\text{hess}}}{\lambda_{\text{noise}}^*} + \frac{8L_{\text{cov}}(\theta_*, \gamma^2) \text{tr}(\mathcal{H}(\theta_*))}{(\lambda_{\text{noise}}^*)^2} \right) \frac{\epsilon_{\text{op}, i-1}}{T_i} + \frac{8(d_x^2 + d_x d_u) L_{\text{cov}}(\theta_*, \gamma^2) L_{\text{hess}} \epsilon_{\text{op}, i-1}^2}{(\lambda_{\text{noise}}^*)^2 T_i}.$$

*Proof.* We will show that the following set of inequalities hold for large enough  $T$  and arbitrary  $T' \geq T$ :

$$\begin{aligned} & \text{tr} \left( \mathcal{H}(\theta_*) (\mathbb{E}_{\theta_*, \mathbf{U}_i, T_i} [I_{d_x} \otimes \sum_{t=T-T_i}^T z_t z_t^\top])^{-1} \right) && \text{(J.8)} \\ & \stackrel{(a)}{\leq} \frac{1}{(1-\epsilon)^2 T_i} \text{tr} \left( \mathcal{H}(\theta_*) \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u})^{-1} \right) && \text{(Steady-state)} \\ & \stackrel{(b)}{\leq} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k_i}} \frac{1}{(1-\epsilon)^2 T_i} \text{tr} \left( \mathcal{H}(\theta_*) \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) + \frac{C_{\text{exp}}}{(1-\epsilon)^2} && \text{(Optimal inputs)} \\ & \stackrel{(c)}{\leq} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, T'}} \frac{1}{(1-\epsilon)^3 T_i} \text{tr} \left( \mathcal{H}(\theta_*) \mathbf{\Gamma}_{T', T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) + \frac{C_{\text{exp}}}{(1-\epsilon)^2} && \text{(Infinite horizon)} \\ & \stackrel{(d)}{\leq} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \frac{2}{(1-\epsilon)^3 T_i} \text{tr} \left( \mathcal{H}(\theta_*) \mathbf{\Gamma}_{T', T'}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}, 0)^{-1} \right) + \frac{C_{\text{exp}}}{(1-\epsilon)^2} && \text{(Noiseless inputs)} \end{aligned}$$

**Steady-state:** By definition,

$$\mathbb{E}_{\theta_*, \mathbf{U}_i, T_i} [I_{d_x} \otimes \sum_{t=T-T_i}^T z_t z_t^\top] = T_i \mathbf{\Gamma}_{T_i}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u}, z_{T-T_i})$$

Note that the inputs played by Algorithm 4 are constructed as in Lemma J.8. It follows then that, by Lemma J.8, as long as,

$$T_i \geq \max \left\{ \frac{2d_u \tau(\tilde{A}_*, \rho)^3 (2k_i \gamma + \|z_{T-T_i}\|_2) ((4 + 2\tau(\tilde{A}_*, \rho)) k_i \gamma + \tau(\tilde{A}_*, \rho) \|z_{T-T_i}\|_2)}{(1-\rho^{k_i})^2 (1-\rho) \lambda_{\min}(\mathbf{\Gamma}_{T_i/2}^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) \epsilon}, \right. \\ \left. \left( \frac{\tau(\tilde{A}_*, \rho)^2 k_i^2 d_u \gamma^2}{(1-\rho^{k_i})^2} + \frac{\tau(\tilde{A}_*, \rho) k_i d_u \gamma^2 \sqrt{T_i/d_u}}{1-\rho^{k_i}} \right) \frac{16d_u \|\tilde{\theta}_*\|_{\mathcal{H}_\infty}^2}{\lambda_{\min}(\mathbf{\Gamma}_{T_i/2}^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) \epsilon} \right\}$$

we will have

$$\mathbf{\Gamma}_{T_i}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u}, z_{T-T_i}) \succeq (1-\epsilon)^2 \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u})$$

This implies that

$$\mathbf{\Gamma}_{T_i}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u}, z_{T-T_i}) \succeq (1-\epsilon)^2 \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{ss}}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u})$$

from which (a) follows.

**Optimal inputs:** We next apply Lemma J.3, which bounds the suboptimality of certainty equivalent experiment design, to show (b). We instantiate Lemma J.3 with

$$\underline{\lambda} = \lambda_{\min}(\mathbf{\Gamma}_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))/2, \quad \Gamma(\theta, \mathbf{U}) = \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{ss}}(\theta, \mathbf{U}, \gamma/\sqrt{2d_u})$$

Note that Lemma H.5 gives that the smoothness condition (J.7) holds for  $L_{\text{cov}}(\theta_*, \gamma^2)$  as defined in (H.14). Furthermore, it is clear that, as long as  $T_i \geq 2k$ , we will have  $\lambda_{\min}(\Gamma(\tilde{\theta}_*, \mathbf{U})) \geq \underline{\lambda}$  for all  $\mathbf{U}$ . To apply Lemma J.3, we need

$$\|\widehat{\theta}_{i-1} - \theta_*\|_{\text{op}} \leq \min\{r_{\text{cov}}(\theta_*), \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))/ (4L_{\text{cov}}(\theta_*, \gamma^2))\}, \quad \|\widehat{\theta}_{i-1} - \theta_*\|_F \leq r_{\text{quad}}(\theta_*)$$

where we choose to instantiate Lemma J.3 in the operator norm, and since the matrix Frobenius norm coincides with the vector 2-norm. The condition on  $\|\widehat{\theta}_{i-1} - \theta_*\|_{\text{op}}$  will hold as long as our assumption on  $\epsilon_{\text{op}, i-1}$  holds. Since  $\widehat{\theta}_i - \theta_*$  is at most rank  $d_x$ , we have

$$\|\widehat{\theta}_{i-1} - \theta_*\|_F \leq \sqrt{d_x} \|\widehat{\theta}_{i-1} - \theta_*\|_{\text{op}} \leq r_{\text{quad}}(\theta_*)$$

where the last inequality again holds so long as our assumption on  $\epsilon_{\text{op}, i-1}$  holds. Then, since we design the input  $\mathbf{U}_i$  on the estimate  $\widehat{\theta}_{i-1}$ , the conditions of Lemma J.3 are met for  $\mathbf{U}_i$ , so

$$\begin{aligned} & \frac{1}{T_i} \left| \text{tr} \left( \mathcal{H}(\theta_*) \Gamma_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_*, \mathbf{U}_i, \gamma/\sqrt{2d_u})^{-1} \right) - \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k_i}} \text{tr} \left( \mathcal{H}(\theta_*) \Gamma_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_*, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) \right| \\ & \leq \left( \frac{4\sqrt{d_x}(d_x^2 + d_x d_u) L_{\text{hess}} \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) + 8L_{\text{cov}}(\theta_*, \gamma^2) \text{tr}(\mathcal{H}(\theta_*))}{\lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))^2} \right) \frac{\epsilon_{\text{op}, i-1}}{T_i} \\ & \quad + \frac{8(d_x^2 + d_x d_u) L_{\text{cov}}(\theta_*, \gamma^2) L_{\text{hess}} \epsilon_{\text{op}, i-1}^2}{\lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))^2 T_i} \\ & =: C_{\text{exp}} \end{aligned}$$

and thus (b) holds.

**Infinite horizon:** Now,

$$\Gamma_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_*, \mathbf{U}, \gamma/\sqrt{2d_u}) = \frac{d_u}{T_i} \Gamma_{T_i/d_u}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}) + \frac{1}{T_i} \sum_{t=1}^{T_i} \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})$$

If  $\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k_i}$ , since  $T_i/(d_u k_i)$  is an integer,

$$\Gamma_{T_i/d_u}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}) = \frac{T_i}{d_u k_i^2} \sum_{\ell=1}^{k_i} (e^{i \frac{2\pi \ell}{k_i}} I - \tilde{A}_*)^{-1} \tilde{B}_* U_{\ell} \tilde{B}_*^{\top} (e^{i \frac{2\pi \ell}{k_i}} I - \tilde{A}_*)^{-H}$$

Then, by Lemma H.8, as long as

$$k_i \geq \max \left\{ \frac{8\pi \|\tilde{\theta}_*\|_{\mathcal{H}_{\infty}} \gamma^2}{\lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) \epsilon}, \frac{\pi}{2 \|\tilde{\theta}_*\|_{\mathcal{H}_{\infty}}} \right\} \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - \tilde{A}_*)^{-2} \tilde{B}\|_{\text{op}} \right)$$

then for any  $T' \geq k_i$  and  $\mathbf{U}^* \in \mathcal{U}_{\gamma^2/2, T'}$ , there exists a feasible  $\mathbf{U}' \in \mathcal{U}_{\gamma^2/2, k_i}$  such that

$$\left\| \frac{d_u}{T_i} \Gamma_{T_i/d_u}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}') - \frac{1}{T'} \Gamma_{T'}^{\text{freq}}(\tilde{\theta}_*, \mathbf{U}^*) \right\|_{\text{op}} \leq \frac{\epsilon}{2} \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))$$

Furthermore, by Lemma H.9, if

$$T_i \geq \max \left\{ \frac{16\tau(\tilde{A}_*, \rho)^2 (\sigma_w^2 + \gamma^2/(2d_u))}{(1 - \rho^2)^2 \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) \epsilon}, \log \left( \frac{(1 - \rho^2)^2 \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u})) \epsilon}{16\tau(\tilde{A}_*, \rho)^2 (\sigma_w^2 + \gamma^2/(2d_u))} \right) \frac{1}{2 \log \rho} \right\}$$

then, for any  $T' \geq T_i$ ,

$$\left\| \frac{1}{T_i} \sum_{t=1}^{T_i} \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}) - \frac{1}{T'} \sum_{t=1}^{T'} \Gamma_t^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}) \right\|_{\text{op}} \leq \frac{\epsilon}{2} \lambda_{\min}(\Gamma_k^{\text{noise}}(\tilde{\theta}_*, \gamma/\sqrt{2d_u}))$$



Now note that, by definition,

$$\min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, k_i}} \text{tr} \left( \mathcal{H}(\theta_\star) \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) \leq \text{tr} \left( \mathcal{H}(\theta_\star) \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}', \gamma/\sqrt{2d_u})^{-1} \right)$$

By what we've just shown, for any  $T' \geq T_i$

$$\begin{aligned} & \mathbf{\Gamma}_{T_i, T_i/d_u}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}', \gamma/\sqrt{2d_u}) \\ & \succeq \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, \gamma/\sqrt{2d_u}) - \frac{\epsilon}{2} \lambda_{\min}(\mathbf{\Gamma}_k^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u})) \cdot I \\ & \succeq (1 - \epsilon) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, \gamma/\sqrt{2d_u}) + \epsilon \frac{1}{2T'} \sum_{t=1}^{T'} \mathbf{\Gamma}_t^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u}) - \frac{\epsilon}{2} \lambda_{\min}(\mathbf{\Gamma}_k^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u})) \cdot I \\ & \succeq (1 - \epsilon) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, \gamma/\sqrt{2d_u}) + \frac{\epsilon}{2} \lambda_{\min}(\mathbf{\Gamma}_{T'/2}^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u})) \cdot I - \frac{\epsilon}{2} \lambda_{\min}(\mathbf{\Gamma}_k^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u})) \cdot I \\ & \succeq (1 - \epsilon) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, \gamma/\sqrt{2d_u}) \end{aligned}$$

From this (c) follows directly.

**Noiseless inputs:** Finally, since

$$\mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, \gamma/\sqrt{2d_u}) \succeq \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}^\star, 0)$$

we can bound

$$\begin{aligned} \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, T'}} \text{tr} \left( \mathcal{H}(\theta_\star) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}, \gamma/\sqrt{2d_u})^{-1} \right) & \leq \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2/2, T'}} \text{tr} \left( \mathcal{H}(\theta_\star) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}, 0)^{-1} \right) \\ & \leq 2 \min_{\mathbf{U} \in \mathcal{U}_{\gamma^2, T'}} \text{tr} \left( \mathcal{H}(\theta_\star) \mathbf{\Gamma}_{T', T'}^{\text{SS}}(\tilde{\theta}_\star, \mathbf{U}, 0)^{-1} \right) \end{aligned}$$

which proves (d). Finally, to simplify the bound, we note that  $\lambda_{\min}(\mathbf{\Gamma}_k^{\text{noise}}(\tilde{\theta}_\star, \gamma/\sqrt{2d_u})) \geq \lambda_{\text{noise}}^\star$ , and we choose  $\rho = \rho_\star$ .  $\square$

**Lemma J.8.** Fix an input  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$  and consider a system  $\theta = (A, B)$ . If we start from some state  $x_0$  and play the time domain input  $u_t = \text{ConstructTimeInput}(\mathbf{U}, T/d_u, k)$  where *ConstructTimeInput* is defined in Algorithm 3 then, so long as  $T/d_u$  is divisible by  $k$ , and

$$T \geq \max \left\{ \frac{2d_u \tau(A, \rho)^3 (2\|B\|_{\text{op}} k \gamma + \|x_0\|_2) ((4 + 2\tau(A, \rho)) \|B\|_{\text{op}} k \gamma + \tau(A, \rho) \|x_0\|_2)}{(1 - \rho^k)^2 (1 - \rho) \lambda_{\min}(\mathbf{\Gamma}_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \epsilon}, \left( \frac{\tau(A, \rho)^2 k^2 d_u \gamma^2}{(1 - \rho^k)^2} + \frac{\tau(A, \rho) k d_u \gamma^2 \sqrt{T/d_u}}{1 - \rho^k} \right) \frac{16d_u \|\theta\|_{\mathcal{H}_\infty}^2}{\lambda_{\min}(\mathbf{\Gamma}_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \epsilon} \right\}$$

we will have

$$\mathbf{\Gamma}_T(\theta, \mathbf{U}, \gamma/\sqrt{2d_u}, x_0) \succeq (1 - \epsilon)^2 \mathbf{\Gamma}_{T, T/d_u}^{\text{SS}}(\theta, \mathbf{U}, \gamma/\sqrt{2d_u}).$$

*Proof.* Let  $\tilde{u}_{\ell, j}$  be defined as in Algorithm 3 for this  $\mathbf{U}$ , and denote  $\mathbf{U}_j = (\tilde{u}_{\ell, j} \tilde{u}_{\ell, j}^{\text{H}})_{\ell=1}^k$ . Let  $\tilde{u}_{j, t}$  denote the time domain version of  $\{\tilde{u}_{\ell, j}\}_{\ell=1}^k$ , as is specified in Algorithm 3. Since  $\mathbf{U} \in \mathcal{U}_{\gamma^2, k}$ , some algebra shows that

$$\frac{1}{k} \sum_{t=1}^k \tilde{u}_{j, t+s}^\top \tilde{u}_{j, t+s} \leq 2d_u \gamma^2$$

for any  $s \geq 0$ . We break up the sum of the response based on which input is being played:

$$\sum_{t=1}^T x_t^{\mathbf{u}} (x_t^{\mathbf{u}})^\top = \sum_{j=1}^{d_u} \sum_{t=(j-1)T/d_u+1}^{jT/d_u} x_t^{\mathbf{u}} (x_t^{\mathbf{u}})^\top$$

which gives:

$$\begin{aligned}\Gamma_T(\theta, \mathbf{U}, \gamma/\sqrt{2d_u}, 0) &= \frac{1}{T} \sum_{j=1}^{d_u} \sum_{t=(j-1)T/d_u+1}^{jT/d_u} x_t^{\mathbf{u}}(x_t^{\mathbf{u}})^\top + \frac{1}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \\ &= \frac{1}{T} \sum_{j=1}^{d_u} \left( \Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, x_{(j-1)T/d_u}^{\mathbf{u}}) + \frac{1}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \right)\end{aligned}$$

If we start from some initial state  $x'$ ,

$$x_t^{\mathbf{u}} = A^t x' + \sum_{s=0}^{t-1} A^{t-s-1} B u_s =: x_t^{\mathbf{u},0} + \tilde{x}_t^{\mathbf{u}}$$

then,

$$\Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, x_{(j-1)T/d_u}^{\mathbf{u}}) = \sum_{t=0}^T x_t^{\mathbf{u}}(x_t^{\mathbf{u}})^\top = \sum_{t=0}^T \left[ \tilde{x}_t^{\mathbf{u}}(\tilde{x}_t^{\mathbf{u}})^\top + \tilde{x}_t^{\mathbf{u}}(x_t^{\mathbf{u},0})^\top + x_t^{\mathbf{u},0}(\tilde{x}_t^{\mathbf{u}})^\top + x_t^{\mathbf{u},0}(x_t^{\mathbf{u},0})^\top \right]$$

Now,

$$\begin{aligned}\left\| \sum_{t=0}^T x_t^{\mathbf{u},0}(\tilde{x}_t^{\mathbf{u}})^\top \right\|_{\text{op}} &\leq \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 \sum_{t=0}^T \|A^t\|_{\text{op}} \|\tilde{x}_t^{\mathbf{u}}\|_2 \leq \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 \tau(A, \rho) \sum_{t=0}^T \rho^t \|\tilde{x}_t^{\mathbf{u}}\|_2 \\ &\leq \frac{2\tau(A, \rho)^2 \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 \|B\|_{\text{op}} k \gamma}{(1 - \rho^k)(1 - \rho)}\end{aligned}$$

where the last inequality follows by Lemma D.7 of (Wagenmaker & Jamieson, 2020), which gives:

$$\|\tilde{x}_t^{\mathbf{u}}\|_2 \leq \frac{2\tau(A, \rho) \|B\|_{\text{op}} k \gamma}{1 - \rho^k}$$

Furthermore,

$$\left\| \sum_{t=0}^T x_t^{\mathbf{u},0}(x_t^{\mathbf{u},0})^\top \right\|_{\text{op}} \leq \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2^2 \sum_{t=0}^T \|A^t\|_{\text{op}}^2 \leq \frac{\|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2^2 \tau(A, \rho)^2}{1 - \rho}$$

Therefore,

$$\begin{aligned}\Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, x_{(j-1)T/d_u}^{\mathbf{u}}) &+ \frac{1}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \\ &\succeq \Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, 0) + \frac{1}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) - \frac{\tau(A, \rho)^2 \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 (4\|B\|_{\text{op}} k \gamma + (1 - \rho^k) \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2)}{(1 - \rho^k)(1 - \rho)} \\ &\succeq (1 - \epsilon) \Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, 0) + \frac{1 - \epsilon}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) + \frac{\epsilon T}{2d_u} \lambda_{\min}(\Gamma_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \cdot I \\ &\quad - \frac{\tau(A, \rho)^2 \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 (4\|B\|_{\text{op}} k \gamma + (1 - \rho^k) \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2)}{(1 - \rho^k)(1 - \rho)} \\ &\succeq (1 - \epsilon) \Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, 0) + \frac{1 - \epsilon}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})\end{aligned}$$

where the last inequality holds as long as

$$T \geq \frac{2d_u \tau(A, \rho)^2 \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 (4\|B\|_{\text{op}} k \gamma + (1 - \rho^k) \|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2)}{(1 - \rho^k)(1 - \rho) \lambda_{\min}(\Gamma_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \epsilon} \quad (\text{J.9})$$

By Proposition B.4, since  $T/d_u$  is divisible by  $k$ , we will have that

$$\Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}_j) = \frac{d_u}{T} \sum_{\ell=1}^T (e^{i\frac{2\pi\ell}{T}} I - A)^{-1} B U_{j,\ell} U_{j,\ell}^H B^H (e^{i\frac{2\pi\ell}{T}} I - A)^{-H}$$

where  $\{U_{j,\ell}\}_{\ell=1}^T = \mathfrak{F}(\tilde{u}_{j,0}, \dots, \tilde{u}_{j,T/d_u})$ . Then, by Lemma H.4, we will have,

$$\begin{aligned} \|\Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, 0) - \Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}_j)\|_{\text{op}} &\leq \left( \frac{8\tau(A, \rho)^2 k^2 d_u \gamma^2}{(1 - \rho^k)^2} + \frac{8\tau(A, \rho) k d_u \gamma^2 \sqrt{T/d_u}}{1 - \rho^k} \right) \left( \max_{\omega \in [0, 2\pi]} \|(e^{-i\omega} I - A)^{-1} B\|_{\text{op}}^2 \right) \\ &\leq \frac{\epsilon T}{2d_u} \lambda_{\min}(\Gamma_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \end{aligned}$$

where the last inequality is true so long as

$$T \geq \left( \frac{8\tau(A, \rho)^2 k^2 d_u \gamma^2}{(1 - \rho^k)^2} + \frac{8\tau(A, \rho) k d_u \gamma^2 \sqrt{T/d_u}}{1 - \rho^k} \right) \left( \max_{\omega \in [0, 2\pi]} \|(e^{i\omega} I - A)^{-1} B\|_{\text{op}}^2 \right) \frac{2d_u}{\lambda_{\min}(\Gamma_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \epsilon}$$

Thus,

$$\begin{aligned} \Gamma_{T/d_u}^{\text{in}}(\theta, \mathbf{U}_j, 0) &+ \frac{1}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \\ &\succeq \Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}_j) + \frac{1}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) - \frac{\epsilon T}{2d_u} \lambda_{\min}(\Gamma_{T/2}^{\text{noise}}(\theta, \gamma/\sqrt{2d_u})) \cdot I \\ &\succeq (1 - \epsilon) \Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}_j) + \frac{1 - \epsilon}{d_u} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \end{aligned}$$

It follows that if (J.9) holds for each  $j$ ,

$$\begin{aligned} \Gamma_T(\theta, \mathbf{U}, \gamma/\sqrt{2d_u}, 0) &\succeq \frac{(1 - \epsilon)^2}{T} \sum_{j=1}^{d_u} \Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}_j) + \frac{(1 - \epsilon)^2}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \\ &= \frac{d_u (1 - \epsilon)^2}{T} \Gamma_{T/d_u}^{\text{freq}}(\theta, \mathbf{U}) + \frac{(1 - \epsilon)^2}{T} \sum_{t=1}^T \Gamma_t^{\text{noise}}(\theta, \gamma/\sqrt{2d_u}) \\ &= (1 - \epsilon)^2 \Gamma_{T, T/d_u}^{\text{ss}}(\theta, \mathbf{U}, \gamma/\sqrt{2d_u}) \end{aligned}$$

It remains to ensure that (J.9) holds by bounding  $\|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2$ . Again by Lemma D.7 of (Wagenmaker & Jamieson, 2020), we have

$$\|x_{(j-1)T/d_u}^{\mathbf{u}}\|_2 \leq \frac{2\tau(A, \rho) \|B\|_{\text{op}} k \gamma}{1 - \rho^k} \mathbb{I}\{j > 1\} + \tau(A, \rho) \rho^{(j-1)T/d_u} \|x_0\|_2 \leq \frac{\tau(A, \rho) (2\|B\|_{\text{op}} k \gamma + \|x_0\|_2)}{1 - \rho^k}$$

Some algebra gives the result. □

## Part IV

# LQR and Further Examples

## K. LQR as Linear Dynamical Decision Making

### K.1. LQR is an Instance of LDDM

Throughout this section we will assume that  $R_x, R_u \succeq I$  and that  $\theta_*$  is stabilizable. Note that, by Lemma 3.1 of (Simchowitz & Foster, 2020), the assumption that  $\theta_*$  is stabilizable implies that in a neighborhood of  $\theta_*$ ,  $P_\infty(\theta)$  and  $K_{\text{opt}}(\theta)$  are infinitely differentiable. We will make use of this fact throughout this section, freely taking derivatives of both quantities. Define:

$$\Psi_{P_*} := \|P_*\|_{\text{op}}, \quad \Psi_{B_*} := \|B_*\|_{\text{op}}, \quad \Psi_{R_u} := \|R_u\|_{\text{op}}.$$

**Theorem K.1.** *If  $\theta_*$  is stabilizable and  $R_u, R_x \succeq I$ , Assumption 3 is satisfied for  $\mathcal{R}_{\text{LQR}}$  with:*

1.  $\mu = 2$ .
2.  $r_{\text{quad}}(\theta_*) = \min \left\{ \frac{1}{150\Psi_{P_*}^5}, \frac{1}{240\Psi_{B_*}\Psi_{P_*}^5}, \frac{\Psi_{B_*}}{2} \right\}$ .
3.  $L_{\mathcal{R}1} = c_1 d_x \Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{c_2 d_x \Psi_{R_u} (1+1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}$ ,  $L_{\mathcal{R}2} = c_3 d_x \left( \Psi_{B_*}^2 \Psi_{P_*}^2 + \Psi_{R_u} (1 + \Psi_{B_*}) \Psi_{P_*} \right)$ ,  $L_{\mathcal{R}3} = c_4 d_x \left( \Psi_{B_*}^3 \Psi_{P_*}^{7/2} + \Psi_{R_u} \Psi_{B_*} (1 + \Psi_{B_*}) \Psi_{P_*}^{5/2} \right)$ .
4.  $L_{a1} = 8\Psi_{P_*}^{7/2}$ ,  $L_{a2} = \text{poly}(\Psi_{P_*})$ ,  $L_{a3} = \text{poly}(\Psi_{P_*}, \Psi_{B_*})$ .
5.  $L_{\text{hess}} = d_x \text{poly}(\Psi_{P_*}, \Psi_{R_u}, \Psi_{B_*}, 1/\Psi_{B_*}) + d_x^2 \left( \Psi_{B_*}^2 \Psi_{P_*}^4 + (1 + \Psi_{B_*}) \Psi_{R_u} \Psi_{P_*}^3 + \frac{\Psi_{R_u} \Psi_{P_*}}{\Psi_{B_*}} \right)$ .

for universal constants  $c_1, c_2, c_3, c_4$ .

*Proof.* From Lemma B.9 of (Simchowitz & Foster, 2020), we have that:

$$\mathcal{R}_{\text{LQR}}(K; \theta_*) = \text{tr} \left( \text{dlyap} \left( A_* + B_* K, (K - K_*)^\top (R_u + B_*^\top P_* B_*) (K - K_*) \right) \right)$$

From this and the definition of  $\text{dlyap}$  it follows that  $\mathcal{R}_{\text{LQR}}(K_*; \theta_*) = 0$ . Furthermore, if we define  $K(t) = K_* + t\Delta_K$  for some  $\Delta_K$ , by the chain rule we have that

$$\frac{d}{dt} \mathcal{R}_{\text{LQR}}(K(t); \theta_*) = \nabla_K \mathcal{R}_{\text{LQR}}(K; \theta_*)|_{K=K(t)} [\Delta_K]$$

Using the expression for  $\frac{d}{dt} \mathcal{R}_{\text{LQR}}(K(t); \theta_*)$  given in the proof of Lemma K.2, we see that  $\frac{d}{dt} \mathcal{R}_{\text{LQR}}(K(t); \theta_*)|_{t=0} = 0$  for  $\Delta_K$ , from which it follows that  $\nabla_K \mathcal{R}_{\text{LQR}}(K; \theta_*)|_{K=K_*} = 0$ . Under the assumption that  $A_* + B_* K$  stable, Lemma B.5 of (Simchowitz & Foster, 2020) gives that:

$$\text{dlyap} \left( A_* + B_* K, (K - K_*)^\top (R_u + B_*^\top P_* B_*) (K - K_*) \right) \succeq (K - K_*)^\top (R_u + B_*^\top P_* B_*) (K - K_*)$$

If  $R_u \succeq I$ , then  $(K - K_*)^\top (R_u + B_*^\top P_* B_*) (K - K_*) \succeq (K - K_*)^\top (K - K_*)$ , and thus, under these conditions, we have

$$\mathcal{R}_{\text{LQR}}(K; \theta_*) \geq \text{tr} \left( (K - K_*)^\top (K - K_*) \right) = \|K - K_*\|_F^2$$

If  $A_* + B_* K$  is not stable but  $(A_*, B_*)$  is a stabilizable system, then the LQR cost is infinite but the optimal LQR cost is finite so  $\mathcal{R}(K; \theta_*) = \infty \geq \|K - K_*\|_F^2$ . Thus, we can choose  $\mu = 2$ . The gradient norm bounds follow directly from Lemmas K.1, K.2, K.3, and K.4. Note that these bounds hold in the domain

$$\|\theta - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\}, \quad \|K - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2}) \quad (\text{K.1})$$

and that  $L_{\alpha 1} = 8\Psi_{P_\star}^{7/2}$ . Since

$$\|\theta - \theta_\star\|_{\text{op}} \leq \|\theta - \theta_\star\|_F, \quad \|K - K_\star\|_{\text{op}} \leq \|K - K_\star\|_F$$

choosing

$$r_{\text{quad}}(\theta_\star) = \min \left\{ \frac{1}{150\Psi_{P_\star}^5}, \frac{1}{240\Psi_{B_\star}\Psi_{P_\star}^5}, \frac{\Psi_{B_\star}}{2} \right\}.$$

we will have that any  $\theta$  satisfying  $\|\theta - \theta_\star\|_F \leq r_{\text{quad}}(\theta_\star)$  also satisfies (K.1) and that any  $K$  satisfying  $\|K - K_\star\|_F \leq L_{\alpha 1}r_{\text{quad}}(\theta_\star)$  also satisfies (K.1).  $\square$

## K.2. Norm Bounds on Gradients

**Lemma K.1.** *Assume that  $\theta_\star$  is stabilizable and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ . Consider some alternate instance  $\theta_0 = (A_0, B_0)$  with  $\|\theta_0 - \theta_\star\|_{\text{op}} \leq \min\{1/(150\Psi_{P_\star}^5), \Psi_{B_\star}/2\}$ . Then, for any  $\delta$  with  $\|\delta\|_{\text{op}} = 1$ ,*

$$\begin{aligned} \|\nabla_\theta K_{\text{opt}}(\theta)|_{\theta=\theta_0}\|_{\text{op}} &\leq 8\Psi_{P_\star}^{7/2} \\ \|\nabla_\theta^2 K_{\text{opt}}(\theta)|_{\theta=\theta_0}\|_{\text{op}} &\leq \text{poly}(\Psi_{P_\star}) \\ \|\nabla_\theta^3 K_{\text{opt}}(\theta)|_{\theta=\theta_0}[\delta, \delta, \delta]\|_{\text{op}} &\leq \text{poly}(\Psi_{P_\star}, \Psi_{B_\star}) \end{aligned}$$

where  $c$  is a universal constant.

*Proof.* Fix some  $\Delta_\theta$  with  $\|\Delta_\theta\|_{\text{op}} = 1$  and let  $\tilde{\theta}(s) = \theta_0 + s\Delta_\theta$ . By the chain rule,

$$\frac{d}{ds} K_{\text{opt}}(\tilde{\theta}(s)) = \nabla_\theta K_{\text{opt}}(\theta)|_{\theta=\tilde{\theta}(s)}[\Delta_\theta]$$

so to bound  $\|\nabla_\theta K_{\text{opt}}(\theta)|_{\theta=\theta_0}\|_{\text{op}}$ , it suffices to bound  $\|\frac{d}{ds} K_{\text{opt}}(\tilde{\theta}(s))|_{s=0}\|_{\text{op}}$  for all unit norm  $\Delta_\theta$ . Lemma 3.2 of (Simchowitz & Foster, 2020) gives that, for  $s$  where  $\tilde{\theta}(s)$  is stabilizable, and any unit norm  $\Delta_\theta$ ,

$$\left\| \frac{d}{ds} K_{\text{opt}}(\tilde{\theta}(s)) \right\|_{\text{op}} \leq 7\|P(s)\|_{\text{op}}^{7/2}$$

By Lemma K.6,  $\tilde{\theta}(0) = \theta_0$  will be stabilizable, and  $\|P(0)\|_{\text{op}} \leq 5\sqrt{\frac{3}{71}}\Psi_{P_\star}$ . Immediately, then, we have

$$\|\nabla_\theta K_{\text{opt}}(\theta)|_{\theta=\theta_0}\|_{\text{op}} \leq 8\Psi_{P_\star}^{7/2}$$

For the second bound, we note that

$$\frac{d^2}{ds^2} K_{\text{opt}}(\tilde{\theta}(s)) = \nabla_\theta^2 K_{\text{opt}}(\theta)|_{\theta=\tilde{\theta}(s)}[\Delta_\theta, \Delta_\theta]$$

and, since the Hessian is symmetric, to obtain a bound on  $\|\nabla_\theta^2 K_{\text{opt}}(\theta)|_{\theta=\theta_0}\|_{\text{op}}$  we can simply bound  $\|\frac{d^2}{ds^2} K_{\text{opt}}(\tilde{\theta}(s))|_{s=0}\|_{\text{op}}$  for all unit norm  $\Delta_\theta$ . However, Lemma B.3 of (Simchowitz & Foster, 2020), and the argument made above give that

$$\left\| \frac{d^2}{ds^2} K_{\text{opt}}(\tilde{\theta}(s)) \right\|_{\text{op}} \leq \text{poly}(\Psi_{P_\star})$$

from which the second conclusion follows. Finally, for the third result, note that

$$\frac{d^3}{ds^3} K_{\text{opt}}(\tilde{\theta}(s)) = \nabla_\theta^3 K_{\text{opt}}(\theta)|_{\theta=\tilde{\theta}(s)}[\Delta_\theta, \Delta_\theta, \Delta_\theta]$$

As before, it is sufficient to simply bound  $\|\frac{d^3}{ds^3} K_{\text{opt}}(\tilde{\theta}(s))|_{s=0}\|_{\text{op}}$ . Since  $\tilde{\theta}(0)$  is stabilizable by Lemma K.6, Lemma K.8 gives

$$\left\| \frac{d^3}{ds^3} K_{\text{opt}}(\tilde{\theta}(s)) \right\|_{\text{op}} \leq \text{poly}(\|P(0)\|_{\text{op}}, \|B(0)\|_{\text{op}}, \|A_{\text{cl}}(0)\|_{\text{op}})$$

By Lemma K.6, and Lemma B.8 of (Simchowitz & Foster, 2020), we can upper bound this by  $\text{poly}(\Psi_{P_\star}, \Psi_{B_\star})$ , which gives the final conclusion.  $\square$

**Lemma K.2.** Assume that  $\theta_*$  is stabilizable and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ . For any  $\theta$  satisfying  $\|\theta - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\}$  and  $K_0$  satisfying  $\|K_0 - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2})$ :

$$\|\nabla_K \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}} \leq c_1 d_x \Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{c_2 d_x \Psi_{R_{\mathbf{u}}}(1 + 1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}$$

for universal constant  $c_1, c_2$ .

*Proof.* Throughout this proof we will assume that  $P_* \succeq I, P_\theta \succeq I$ , which holds by Lemma 4.2 of (Simchowitz & Foster, 2020) so long as  $R_{\mathbf{x}} \succeq I$ . Fix  $\theta$  and let  $\tilde{K}(t) = K_0 + t\Delta_K$  for  $\Delta_K$  satisfying  $\|\Delta_K\|_{\text{op}} = 1$ . By the chain rule

$$\frac{d}{dt} \mathcal{R}(\tilde{K}(t); \theta)|_{t=0} = \nabla_K \mathcal{R}(K; \theta)|_{K=K_0}[\Delta_K]$$

so to bound  $\|\nabla_K \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}}$ , it is sufficient to bound  $\frac{d}{dt} \mathcal{R}(\tilde{K}(t); \theta)|_{t=0}$  over all unit norm  $\Delta_K$ .

For a given  $\theta$ , we'll denote  $K_\theta := K_{\text{opt}}(\theta)$ ,  $P_\theta := P_\infty(\theta)$ , and use  $A, B$  to refer to the system matrices associated with  $\theta$ . Then by Lemma B.9 of (Simchowitz & Foster, 2020), we have

$$\mathcal{R}(K; \theta) = \text{tr}(\text{dlyap}(A + BK, (K - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)(K - K_\theta)))$$

Define  $Q(t) := \text{dlyap}(A + B\tilde{K}(t), (\tilde{K}(t) - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)(\tilde{K}(t) - K_\theta))$ . It follows that  $\frac{d}{dt} \mathcal{R}(\tilde{K}(t); \theta) = \text{tr}(\frac{d}{dt} Q(t))$ . By definition of dlyap,

$$Q(t) = (A + B\tilde{K}(t))^\top Q(t)(A + B\tilde{K}(t)) + (\tilde{K}(t) - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)(\tilde{K}(t) - K_\theta)$$

Differentiating  $Q(t)$  (and hiding  $t$  dependence for simplicity), and since  $\tilde{K}' = \Delta_K$ , we have

$$\begin{aligned} Q' &= (A + B\tilde{K})^\top Q'(A + B\tilde{K}) + (B\Delta_K)^\top Q(A + B\tilde{K}) + (A + B\tilde{K})^\top Q(B\Delta_K) \\ &\quad \Delta_K^\top (R_{\mathbf{u}} + B^\top P_\theta B)(\tilde{K} - K_\theta) + (\tilde{K} - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)\Delta_K \\ &= \text{dlyap}(A + B\tilde{K}, (B\Delta_K)^\top Q(A + B\tilde{K}) + (A + B\tilde{K})^\top Q(B\Delta_K) \\ &\quad \Delta_K^\top (R_{\mathbf{u}} + B^\top P_\theta B)(\tilde{K} - K_\theta) + (\tilde{K} - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)\Delta_K) \end{aligned}$$

By Lemma B.5 of (Simchowitz & Foster, 2020), we can upper bound this dlyap expression as,

$$\begin{aligned} \|Q'(0)\|_{\text{op}} &\leq 2\|\text{dlyap}(A + B\tilde{K}(0), I)\|_{\text{op}} \left( \|(A + B\tilde{K}(0))^\top Q(0)B\Delta_K\|_{\text{op}} \right. \\ &\quad \left. + \|(\tilde{K}(0) - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)\Delta_K\|_{\text{op}} \right) \end{aligned}$$

Since  $\|\theta - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\}$  and  $\|\tilde{K}(0) - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2})$ , we can apply the norm bounds in Lemma K.6 to upper bound this as

$$\|Q'(0)\|_{\text{op}} \leq c\Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{c\Psi_{R_{\mathbf{u}}}(1 + 1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}$$

As this holds independent of  $\Delta_K$  and since  $|\frac{d}{dt} \mathcal{R}(K(t); \theta)|_{t=0}| \leq d_x \|Q'(0)\|$ , it follows that

$$\|\nabla_K \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}} \leq cd_x \Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{cd_x \Psi_{R_{\mathbf{u}}}(1 + 1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}.$$

□

**Lemma K.3.** Assume that  $\theta_*$  is stabilizable and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ . For any  $\theta$  satisfying  $\|\theta - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\}$  and  $K_0$  satisfying  $\|K_0 - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2})$ :

$$\|\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}} \leq c_1 d_x \left( \Psi_{B_*}^2 \Psi_{P_*}^2 + \Psi_{R_{\mathbf{u}}}(1 + \Psi_{B_*})\Psi_{P_*} \right)$$

$$\|\nabla_K^3 \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}} \leq c_2 d_x \left( \Psi_{B_*}^3 \Psi_{P_*}^{7/2} + \Psi_{R_{\mathbf{u}}}\Psi_{B_*}(1 + \Psi_{B_*})\Psi_{P_*}^{5/2} \right)$$

for some universal constants  $c_1, c_2$ .

*Proof.* Throughout, unless otherwise specified, we adopt the same notation as is used in the proof of Lemma K.2. Let  $\tilde{K}(t_1, t_2) = K_0 + t_1 \Delta_{K_1} + t_2 \Delta_{K_2}$ . By the chain rule, and since  $\frac{d}{dt_1} \tilde{K}(t_1, t_2) = \Delta_{K_1}$ ,  $\frac{d}{dt_2} \tilde{K}(t_1, t_2) = \Delta_{K_2}$ ,

$$\frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2); \theta)|_{t_1=t_2=0} = \nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0} [\Delta_{K_1}, \Delta_{K_2}]$$

To bound  $\|\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0}\|_{\text{op}}$ , it then suffices to bound  $\|\frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2); \theta)|_{t_1=t_2=0}\|_{\text{op}}$  over all unit norm  $\Delta_{K_1}, \Delta_{K_2}$ . From the proof of Lemma K.2, we have that

$$\frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2); \theta) = \text{tr} \left( \frac{d}{dt_2} \frac{d}{dt_1} Q(t_1, t_2) \right)$$

where

$$Q(t_1, t_2) = (A + B\tilde{K}(t_1, t_2))^\top Q(t_1, t_2) (A + B\tilde{K}(t_1, t_2)) + (\tilde{K}(t_1, t_2) - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B) (\tilde{K}(t_1, t_2) - K_\theta)$$

Using our expression for the first derivate of  $Q$  from the proof of Lemma K.2, dropping the explicit  $t_1, t_2$  dependence, and adopting the notation  $Q_{t_i} = \frac{d}{dt_i} Q$ ,

$$\begin{aligned} Q_{t_1} &= (A + B\tilde{K})^\top (Q_{t_1}) (A + B\tilde{K}) + (B\Delta_{K_1})^\top Q (A + B\tilde{K}) + (A + B\tilde{K})^\top Q (B\Delta_{K_1}) \\ &\quad \Delta_{K_1}^\top (R_{\mathbf{u}} + B^\top P_\theta B) (\tilde{K} - K_\theta) + (\tilde{K} - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_1} \\ &= \text{dlyap} \left( A + B\tilde{K}, (B\Delta_{K_1})^\top Q (A + B\tilde{K}) + (A + B\tilde{K})^\top Q (B\Delta_{K_1}) \right. \\ &\quad \left. \Delta_{K_1}^\top (R_{\mathbf{u}} + B^\top P_\theta B) (\tilde{K} - K_\theta) + (\tilde{K} - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_1} \right) \end{aligned}$$

then taking the derivative of this with respect to  $t_2$  gives

$$\begin{aligned} Q_{t_1, t_2} &= (A + B\tilde{K})^\top (Q_{t_1, t_2}) (A + B\tilde{K}) + (B\Delta_{K_2})^\top (Q_{t_1}) (A + B\tilde{K}) + (A + B\tilde{K})^\top (Q_{t_1}) (B\Delta_{K_2}) \\ &\quad + (B\Delta_{K_1})^\top (Q_{t_2}) (A + B\tilde{K}) + (A + B\tilde{K})^\top (Q_{t_2}) (B\Delta_{K_1}) + (B\Delta_{K_1})^\top Q (B\Delta_{K_2}) \\ &\quad + (B\Delta_{K_2})^\top Q (B\Delta_{K_1}) + \Delta_{K_1}^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_2} + \Delta_{K_2}^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_1} \\ &= \text{dlyap} \left( A + B\tilde{K}, (B\Delta_{K_2})^\top (Q_{t_1}) (A + B\tilde{K}) + (A + B\tilde{K})^\top (Q_{t_1}) (B\Delta_{K_2}) + (B\Delta_{K_1})^\top (Q_{t_2}) (A + B\tilde{K}) \right. \\ &\quad + (A + B\tilde{K})^\top (Q_{t_2}) (B\Delta_{K_1}) + (B\Delta_{K_1})^\top Q (B\Delta_{K_2}) + (B\Delta_{K_2})^\top Q (B\Delta_{K_1}) \\ &\quad \left. + \Delta_{K_1}^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_2} + \Delta_{K_2}^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_{K_1} \right) \end{aligned}$$

We would like to bound the operator norm of  $Q_{t_1, t_2}(0)$ . Note that the bound on  $\|Q'(0)\|_{\text{op}}$  given in Lemma K.2 still applies in this setting due to our restriction that  $\|\tilde{K}(0, 0) - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2})$ , so

$$\|Q(0)\|_{\text{op}} \leq \frac{c}{\Psi_{P_*}} + \frac{c\Psi_{R_{\mathbf{u}}}}{\Psi_{B_*}^2 \Psi_{P_*}^2}$$

$$\|Q_{t_1}(0)\|_{\text{op}}, \|Q_{t_2}(0)\|_{\text{op}} \leq c_2 \Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{c_3 \Psi_{R_{\mathbf{u}}}(1 + 1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}$$

Furthermore, we are in the domain where Lemma K.6 holds so,

$$\begin{aligned} \|Q_{t_1, t_2}(0)\|_{\text{op}} &\leq \|\text{dlyap}(A + B\tilde{K}(0), I)\|_{\text{op}} \left( 2\|B\|_{\text{op}} \|A + B\tilde{K}(0)\|_{\text{op}} (\|Q_{t_1}(0)\|_{\text{op}} + \|Q_{t_2}(0)\|_{\text{op}}) \right. \\ &\quad \left. + 2\|B\|_{\text{op}}^2 \|Q(0)\|_{\text{op}} + 2\|R_{\mathbf{u}} + B^\top P_\theta B\|_{\text{op}} \right) \\ &\leq c\Psi_{B_\star}^2 \Psi_{P_\star}^2 + c\Psi_{R_{\mathbf{u}}}(1 + \Psi_{B_\star})\Psi_{P_\star} \end{aligned}$$

Since  $|\frac{d}{dt_1} \frac{d}{dt_2} \mathcal{R}(\tilde{K}(t_1, t_2); \theta)| \leq d_x \|Q_{t_1, t_2}\|_{\text{op}}$ , the first bound follows.

To bound  $\|\nabla_K^3 \mathcal{R}(K; \theta)\|_{\text{op}}$ , we define  $\tilde{K}(t_1, t_2, t_3) = K_0 + t_1 \Delta_{K1} + t_2 \Delta_{K2} + t_3 \Delta_{K3}$ , for

$$\|\Delta_{K1}\|_{\text{op}}, \|\Delta_{K2}\|_{\text{op}}, \|\Delta_{K3}\|_{\text{op}} = 1$$

and note that by the chain rule

$$\frac{d}{dt_3} \frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2, t_3); \theta)|_{t_1=t_2=t_3=0} = \nabla_K^3 \mathcal{R}(K; \theta)|_{K=K_0} [\Delta_{K1}, \Delta_{K2}, \Delta_{K3}]$$

so, as before, it suffices to bound  $\|\frac{d}{dt_3} \frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2, t_3); \theta)|_{t_1=t_2=t_3=0}\|_{\text{op}}$  over all unit norm  $\Delta_{K1}, \Delta_{K2}, \Delta_{K3}$ . Again we have that

$$\frac{d}{dt_3} \frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(\tilde{K}(t_1, t_2, t_3); \theta) = \text{tr} \left( \frac{d}{dt_3} \frac{d}{dt_2} \frac{d}{dt_1} Q(t_1, t_2, t_3) \right)$$

To bound this, we can differentiate the expression for  $Q_{t_1, t_2}$  given above with respect to  $t_3$ :

$$\begin{aligned} Q_{t_1, t_2, t_3} &= (A + B\tilde{K})^\top (Q_{t_1, t_2, t_3})(A + B\tilde{K}) + (B\Delta_{K3})^\top (Q_{t_1, t_2})(A + B\tilde{K}) + (A + B\tilde{K})^\top (Q_{t_1, t_2})(B\Delta_{K3}) \\ &\quad + (B\Delta_{K2})^\top (Q_{t_1, t_3})(A + B\tilde{K}) + (B\Delta_{K2})^\top (Q_{t_1})(B\Delta_{K3}) + (A + B\tilde{K})^\top (Q_{t_1, t_3})(B\Delta_{K2}) \\ &\quad + (B\Delta_{K3})^\top (Q_{t_1})(B\Delta_{K2}) + (B\Delta_{K1})^\top (Q_{t_2, t_3})(A + B\tilde{K}) + (B\Delta_{K1})^\top (Q_{t_2})(B\Delta_{K3}) \\ &\quad + (A + B\tilde{K})^\top (Q_{t_2, 3})(B\Delta_{K1}) + (B\Delta_{K3})^\top (Q_{t_2})(B\Delta_{K1}) + (B\Delta_{K1})^\top Q_{t_3}(B\Delta_{K2}) \\ &\quad + (B\Delta_{K2})^\top Q_{t_3}(B\Delta_{K1}) \\ &= \text{dlyap} \left( A + B\tilde{K}, (B\Delta_{K3})^\top (Q_{t_1, t_2})(A + B\tilde{K}) + (A + B\tilde{K})^\top (Q_{t_1, t_2})(B\Delta_{K3}) \right. \\ &\quad + (B\Delta_{K2})^\top (Q_{t_1, t_3})(A + B\tilde{K}) + (B\Delta_{K2})^\top (Q_{t_1})(B\Delta_{K3}) + (A + B\tilde{K})^\top (Q_{t_1, t_3})(B\Delta_{K2}) \\ &\quad + (B\Delta_{K3})^\top (Q_{t_1})(B\Delta_{K2}) + (B\Delta_{K1})^\top (Q_{t_2, t_3})(A + B\tilde{K}) + (B\Delta_{K1})^\top (Q_{t_2})(B\Delta_{K3}) \\ &\quad + (A + B\tilde{K})^\top (Q_{t_2, 3})(B\Delta_{K1}) + (B\Delta_{K3})^\top (Q_{t_2})(B\Delta_{K1}) + (B\Delta_{K1})^\top Q_{t_3}(B\Delta_{K2}) \\ &\quad \left. + (B\Delta_{K2})^\top Q_{t_3}(B\Delta_{K1}) \right) \end{aligned}$$

Thus,

$$\begin{aligned} \|Q_{t_1, t_2, t_3}(0)\|_{\text{op}} &\leq \|\text{dlyap}(A + B\tilde{K}(0), I)\|_{\text{op}} \left( 2\|B\|_{\text{op}} \|A + B\tilde{K}(0)\|_{\text{op}} (\|Q_{t_1, t_2}(0)\|_{\text{op}} + \|Q_{t_1, t_3}(0)\|_{\text{op}} \right. \\ &\quad \left. + \|Q_{t_2, t_3}(0)\|_{\text{op}}) + 2\|B\|_{\text{op}}^2 (\|Q_{t_1}(0)\|_{\text{op}} + \|Q_{t_2}(0)\|_{\text{op}} + \|Q_{t_3}(0)\|_{\text{op}}) \right) \end{aligned}$$

Note that the norm bounds proved on  $\|Q'(0)\|_{\text{op}}$  given in Lemma K.2 still applies in this setting due to our restriction that  $\|K_0 - K_\star\|_{\text{op}} \leq 1/(30\Psi_{B_\star} \Psi_{P_\star}^{3/2})$ , and similarly our bound proved above on  $\|Q_{t_1, t_2}(0)\|_{\text{op}}$  can be used to bound each of the second derivatives. Combining these results, and using that  $\Psi_{P_\star} \geq 1$ , gives

$$\|Q_{t_1, t_2, t_3}(0)\|_{\text{op}} \leq c\Psi_{B_\star}^3 \Psi_{P_\star}^{7/2} + c\Psi_{R_{\mathbf{u}}} \Psi_{B_\star} (1 + \Psi_{B_\star}) \Psi_{P_\star}^{5/2}$$

The second bound then follows directly.  $\square$



**Lemma K.4.** Assume that  $\theta_*$  is stabilizable and  $R_x, R_u \succeq I$ . For  $\theta_0, \theta_1$  satisfying  $\|\theta_0 - \theta_*\|_{\text{op}}, \|\theta_1 - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\}$  and  $K_0$  satisfying  $\|K_0 - K_*\|_{\text{op}} \leq 1/(30\Psi_{B_*}\Psi_{P_*}^{3/2})$ , we have

$$\begin{aligned} \|\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0} - \nabla_K^2 \mathcal{R}(K; \theta')|_{K=K_0}\|_{\text{op}} &\leq \left( d_x \text{poly}(\Psi_{P_*}, \Psi_{R_u}, \Psi_{B_*}, 1/\Psi_{B_*}) \right. \\ &\quad \left. + d_x^2 \left( \Psi_{B_*}^2 \Psi_{P_*}^4 + (1 + \Psi_{B_*}) \Psi_{R_u} \Psi_{P_*}^3 + \frac{\Psi_{R_u} \Psi_{P_*}}{\Psi_{B_*}} \right) \right) \cdot \|\theta_0 - \theta_1\|_{\text{op}}. \end{aligned}$$

*Proof.* Note that, since the Hessian is symmetric,

$$\begin{aligned} &\|\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0} - \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0}\|_{\text{op}} \\ &= \max_{\Delta_K: \|\Delta_K\|_{\text{op}}=1} |\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0}[\Delta_K, \Delta_K] - \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0}[\Delta_K, \Delta_K]| \end{aligned}$$

so it suffices to bound  $|\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0}[\Delta_K, \Delta_K] - \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0}[\Delta_K, \Delta_K]|$  over all unit norm  $\Delta_K$ . Let  $\Delta_\theta = (\Delta_A, \Delta_B)$  satisfy  $\|\Delta_\theta\|_{\text{op}} = 1$  and denote  $A(s) = A_* + s\Delta_A$ ,  $B(s) = B_* + s\Delta_B$ , and  $\tilde{\theta}(s) = (A(s), B(s))$ . Assume that  $\theta_1 = \theta_0 + s_1\Delta_\theta$  for some  $s_1$ . Fix  $\Delta_K$  with  $\|\Delta_K\|_{\text{op}} = 1$ . By the chain rule,

$$\frac{d}{ds} \nabla_K^2 \mathcal{R}(K; \tilde{\theta}(s))|_{K=K_0}[\Delta_K, \Delta_K] = \nabla_\theta (\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0}[\Delta_K, \Delta_K])|_{\theta=\tilde{\theta}(s)}[\Delta_\theta]$$

So by Taylor's Theorem,

$$\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0}[\Delta_K, \Delta_K] = \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0}[\Delta_K, \Delta_K] - \nabla_\theta (\nabla_K^2 \mathcal{R}(K; \tilde{\theta}(s))|_{K=K_0}[\Delta_K, \Delta_K])|_{s=s_2}[\Delta_\theta] s_1$$

for some  $s_2 \in [0, s_1]$ . Thus, since  $\|\theta_0 - \theta_1\|_{\text{op}} = \|s_1\Delta_\theta\|_{\text{op}} = s_1$ , denoting  $\theta_2 := \tilde{\theta}(s_2)$ ,

$$\begin{aligned} &|\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0}[\Delta_K, \Delta_K] - \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0}[\Delta_K, \Delta_K]| \\ &\leq |\nabla_\theta (\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0}[\Delta_K, \Delta_K])|_{\theta=\theta_2}[\Delta_\theta]| \cdot \|\theta_0 - \theta_1\|_{\text{op}} \end{aligned}$$

So we can simply bound  $|\nabla_\theta (\nabla_K^2 \mathcal{R}(K; \theta)|_{K=K_0}[\Delta_K, \Delta_K])|_{\theta=\theta_2}[\Delta_\theta]|$  over all unit norm  $\Delta_\theta$ , and all  $\theta_2 = \theta + s_2\Delta_\theta$ ,  $s_2 \in [0, s_1]$ . Note that  $\Delta_\theta = \frac{1}{s_1}(\theta_1 - \theta_0)$ , so

$$\begin{aligned} \|\theta_2 - \theta_*\|_{\text{op}} &= \|\theta_0 + s_2\Delta_\theta - \theta_*\|_{\text{op}} = \|(1 - s_2/s_1)\theta_0 + (s_2/s_1)\theta_1 - \theta_*\|_{\text{op}} \\ &\leq (1 - s_2/s_1)\|\theta_0 - \theta_*\|_{\text{op}} + s_2/s_1\|\theta_1 - \theta_*\|_{\text{op}} \leq \min\{1/(150\Psi_{P_*}^5), \Psi_{B_*}/2\} \end{aligned}$$

so we are in the domain where the bounds given in Lemmas K.2, K.3, and K.6 hold. By Lemma K.3, we know that (where we drop the dependence on  $s$  for brevity)

$$\begin{aligned} \nabla_K^2 \mathcal{R}(K; \tilde{\theta}(s))|_{K=K_0}[\Delta_K, \Delta_K] &= \text{tr} \left( \text{dlyap} \left( A + BK_0, 2(B\Delta_K)^\top (Q_1)(A + BK_0) + 2(A + BK_0)^\top (Q_1)(B\Delta_K) \right. \right. \\ &\quad \left. \left. + 2(B\Delta_K)^\top Q(B\Delta_K) + 2\Delta_K^\top (R_u + B^\top P_{\tilde{\theta}} B) \Delta_K \right) \right) \\ &=: \text{tr}(M(s)) \end{aligned}$$

where  $Q$  and  $Q_1$  satisfy

$$Q = (A + BK_0)^\top Q(A + BK_0) + (K_0 - K_{\tilde{\theta}})^\top (R_u + B^\top P_{\tilde{\theta}} B)(K_0 - K_{\tilde{\theta}})$$

$$\begin{aligned} Q_1 &= (A + BK_0)^\top (Q_1)(A + BK_0) + (B\Delta_K)^\top Q(A + BK_0) + (A + BK_0)^\top Q(B\Delta_K) \\ &\quad + \Delta_K^\top (R_u + B^\top P_{\tilde{\theta}} B)(K_0 - K_{\tilde{\theta}}) + (K_0 - K_{\tilde{\theta}})^\top (R_u + B^\top P_{\tilde{\theta}} B) \Delta_K \end{aligned}$$

and  $P_{\tilde{\theta}} = P_\infty(\tilde{\theta})$ . It follows, by the definition of  $\text{dlyap}$ , that  $M$  satisfies,

$$M = (A + BK_0)^\top M(A + BK_0) + 2(B\Delta_K)^\top (Q_1)(A + BK_0) + 2(A + BK_0)^\top (Q_1)(B\Delta_K)$$

$$+ 2(B\Delta_K)^\top Q(B\Delta_K) + 2\Delta_K^\top (R_u + B^\top P_\theta B)\Delta_K$$

and that

$$\frac{d}{ds} \nabla_K^2 \mathcal{R}(K; \tilde{\theta})|_{K=K_0}[\Delta_K, \Delta_K] = \text{tr}\left(\frac{d}{ds} M\right)$$

Differentiating this expression for  $M$  with respect to  $s$  gives (where here we let  $(\cdot)'$  denote the derivative with respect to  $s$ )

$$\begin{aligned} \text{tr}(M') &= \text{tr}\left((A + BK_0)^\top M'(A + BK_0) + (\Delta_A + \Delta_B K_0)^\top M(A + BK_0) + (A + BK_0)^\top M(\Delta_A + \Delta_B K_0)\right. \\ &\quad + 2(\Delta_B \Delta_K)^\top (Q_1)(A + BK_0) + 2(B\Delta_K)^\top (Q'_1)(A + BK_0) + 2(B\Delta_K)^\top (Q_1)(\Delta_A + \Delta_B K_0) \\ &\quad + 2(\Delta_A + \Delta_B K_0)^\top (Q_1)(B\Delta_K) + 2(A + BK_0)^\top (Q'_1)(B\Delta_K) + 2(A + BK_0)^\top (Q_1)(\Delta_B \Delta_K) \\ &\quad + 2(\Delta_B \Delta_K)^\top Q(B\Delta_K) + 2(B\Delta_K)^\top Q'(B\Delta_K) + 2(B\Delta_K)^\top Q(\Delta_B \Delta_K) \\ &\quad \left. + 2\Delta_K^\top (\Delta_B^\top P_\theta B + B^\top P'_\theta B + B^\top P_\theta \Delta_B)\Delta_K\right) \\ &= \text{tr}\left(\text{dlyap}\left(A + BK_0, (\Delta_A + \Delta_B K_0)^\top M(A + BK_0) + (A + BK_0)^\top M(\Delta_A + \Delta_B K_0)\right.\right. \\ &\quad + 2(\Delta_B \Delta_K)^\top (Q_1)(A + BK_0) + 2(B\Delta_K)^\top (Q'_1)(A + BK_0) + 2(B\Delta_K)^\top (Q_1)(\Delta_A + \Delta_B K_0) \\ &\quad + 2(\Delta_A + \Delta_B K_0)^\top (Q_1)(B\Delta_K) + 2(A + BK_0)^\top (Q'_1)(B\Delta_K) + 2(A + BK_0)^\top (Q_1)(\Delta_B \Delta_K) \\ &\quad + 2(\Delta_B \Delta_K)^\top Q(B\Delta_K) + 2(B\Delta_K)^\top Q'(B\Delta_K) + 2(B\Delta_K)^\top Q(\Delta_B \Delta_K) \\ &\quad \left. + 2\Delta_K^\top (\Delta_B^\top P_\theta B + B^\top P'_\theta B + B^\top P_\theta \Delta_B)\Delta_K\right) \\ &\leq d_x \|\text{dlyap}(A + BK_0, I)\|_{\text{op}} \left(2(1 + \|K_0\|_{\text{op}})\|A + BK_0\|_{\text{op}}\|M\|_{\text{op}} + 4\|A + BK_0\|_{\text{op}}\|Q_1\|_{\text{op}}\right. \\ &\quad + 4\|B\|_{\text{op}}(1 + \|K_0\|_{\text{op}})\|Q_1\|_{\text{op}} + 4\|A + BK_0\|_{\text{op}}\|B\|_{\text{op}}\|Q'_1\|_{\text{op}} + 4\|B\|_{\text{op}}\|Q\|_{\text{op}} + 2\|B\|_{\text{op}}^2\|Q'\|_{\text{op}} \\ &\quad \left. + 4\|B\|_{\text{op}}\|P'_\theta\|_{\text{op}} + 2\|B\|_{\text{op}}^2\|P'_\theta\|_{\text{op}}\right) \end{aligned}$$

Then, using bounds proved in Lemmas K.2 and K.3 to upper bound this at  $s = s_2$ ,

$$\begin{aligned} \text{tr}(M'(s_2)) &\leq cd_x \Psi_{P_*} \left( (1 + \|K_0\|_{\text{op}}) \sqrt{\Psi_{P_*}} \|M(s_2)\|_{\text{op}} + \sqrt{\Psi_{P_*}} \|Q_1(s_2)\|_{\text{op}} \right. \\ &\quad + \Psi_{B_*} (1 + \|K_0\|_{\text{op}}) \|Q_1(s_2)\|_{\text{op}} + \Psi_{B_*} \sqrt{\Psi_{P_*}} \|Q'_1(s_2)\|_{\text{op}} + \Psi_{B_*} \|Q(s_2)\|_{\text{op}} \\ &\quad \left. + \Psi_{B_*}^2 \|Q'(s_2)\|_{\text{op}} + \Psi_{B_*} \Psi_{P_*} + \Psi_{B_*}^2 \|P'_\theta(s_2)\|_{\text{op}} \right) \end{aligned}$$

By Lemma K.2 we can bound

$$\|Q(s_2)\|_{\text{op}} \leq \frac{c}{\Psi_{P_*}} + \frac{c\Psi_{R_u}}{\Psi_{B_*}^2 \Psi_{P_*}^2}, \quad \|Q_1(s_2)\|_{\text{op}} \leq c\Psi_{B_*} \sqrt{\Psi_{P_*}} + \frac{c\Psi_{R_u}(1 + 1/\Psi_{B_*})}{\sqrt{\Psi_{P_*}}}$$

and by Lemma K.3 we can bound

$$\|M(s_2)\|_{\text{op}} \leq \|\nabla_K^2 \mathcal{R}(K; \theta_2)|_{K=K_0}\|_{\text{op}} \leq cd_x \left( \Psi_{B_*}^2 \Psi_{P_*}^2 + \Psi_{R_u}(1 + \Psi_{B_*})\Psi_{P_*} \right)$$

Furthermore, by our assumption on  $K_0$  and Lemma B.8 of (Simchowitz & Foster, 2020),

$$\|K_0\|_{\text{op}} \leq \|K_*\|_{\text{op}} + 1/(30\Psi_{B_*} \Psi_{P_*}^{3/2}) \leq \sqrt{\Psi_{P_*}} + 1/(30\Psi_{B_*} \Psi_{P_*}^{3/2})$$

and by Lemma 3.2 of (Simchowitz & Foster, 2020) and Lemma K.6

$$\|P'_\theta(s_2)\|_{\text{op}} \leq 4\|P_\theta(s_2)\|_{\text{op}}^3 \leq c\Psi_{P_*}^3$$

It remains to bound  $\|Q'(s_2)\|_{\text{op}}$  and  $\|Q'_1(s_2)\|_{\text{op}}$ . Given the expression for  $Q$ , we can differentiate it to get

$$Q' = \text{dlyap}\left(A + BK_0, (\Delta_A + \Delta_B K_0)^\top Q(A + BK_0) + (A + BK_0)^\top Q(\Delta_A + \Delta_B K_0)\right)$$

$$\begin{aligned}
 & - (K'_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)(K_0 - K_\theta) - (K_0 - K_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B)(K_\theta) \\
 & + (K_0 - K_\theta)^\top (\Delta_B^\top P_\theta B + B^\top P'_\theta B + B^\top P_\theta \Delta_B)(K_0 - K_\theta)
 \end{aligned}$$

so,

$$\begin{aligned}
 \|Q'\|_{\text{op}} & \leq \|\text{dlyap}(A + BK_0, I)\|_{\text{op}} \left( 2(1 + \|K_0\|_{\text{op}}) \|A + BK_0\|_{\text{op}} \|Q\|_{\text{op}} + 2\|(R_{\mathbf{u}} + B^\top P_\theta B)(K_0 - K_\theta)\|_{\text{op}} \|K'_\theta\|_{\text{op}} \right. \\
 & \quad \left. + 2\|K_0 - K_\theta\|_{\text{op}} \|B(K_0 - K_\theta)\|_{\text{op}} \|P_\theta\|_{\text{op}} + \|B(K_0 - K_\theta)\|_{\text{op}}^2 \|P'_\theta\|_{\text{op}} \right)
 \end{aligned}$$

Lemma 3.2 of (Simchowitz & Foster, 2020) gives  $\|K'_{\tilde{\theta}(s_2)}\|_{\text{op}} \leq 7\|P_{\tilde{\theta}(s_2)}\|_{\text{op}}^{7/2} \leq c\Psi_{P_\star}^{7/2}$ . Then by the Mean Value Theorem,

$$\begin{aligned}
 \|K_0 - K_{\tilde{\theta}(s_2)}\|_{\text{op}} & \leq \|K_0 - K_\star\|_{\text{op}} + \|K_\star - K_{\tilde{\theta}(s_2)}\|_{\text{op}} \\
 & \leq 1/(30\Psi_{B_\star}\Psi_{P_\star}^{3/2}) + \max_{s: \|\tilde{\theta}(s) - \theta_\star\|_{\text{op}} \leq \min\{1/(150\Psi_{P_\star}^5), \Psi_{B_\star}/2\}} \|K'_{\tilde{\theta}(s_2)}\|_{\text{op}} \|\theta - \theta_\star\|_{\text{op}} \\
 & \leq c(1 + 1/\Psi_{B_\star})/\Psi_{P_\star}^{3/2}
 \end{aligned}$$

Using this, Lemma K.6, and what we have shown above, we can then bound

$$\begin{aligned}
 \|Q'(s_2)\|_{\text{op}} & \leq c \left( \Psi_{B_\star} \Psi_{P_\star}^4 + (\Psi_{R_{\mathbf{u}}}^{1/2} + \Psi_{R_{\mathbf{u}}}/\Psi_{B_\star}) \Psi_{P_\star}^3 + \Psi_{P_\star} + \Psi_{R_{\mathbf{u}}}/\Psi_{B_\star}^2 \right. \\
 & \quad \left. + 1/(\Psi_{B_\star} \Psi_{P_\star}) + \Psi_{R_{\mathbf{u}}}/(\Psi_{B_\star}^3 \Psi_{P_\star}^2) \right)
 \end{aligned}$$

We now bound  $\|Q'_1(s_2)\|_{\text{op}}$ . Differentiating the expression for  $Q_1$  given above yields:

$$\begin{aligned}
 Q'_1 & = \text{dlyap} \left( A + BK_0, (\Delta_A + \Delta_B K_0)^\top Q_1 (A + BK_0) + (A + BK_0)^\top Q_1 (\Delta_A + \Delta_B K_0) \right. \\
 & \quad + (\Delta_B \Delta_K)^\top Q (A + BK_0) + (B \Delta_K)^\top Q' (A + BK_0) + (B \Delta_K)^\top Q (\Delta_A + \Delta_B K_0) \\
 & \quad + (\Delta_A + \Delta_B K_0)^\top Q (B \Delta_K) + (A + BK_0)^\top Q' (B \Delta_K) + (A + BK_0)^\top Q (\Delta_B \Delta_K) \\
 & \quad + \Delta_K^\top (\Delta_B^\top P_\theta B + B^\top P'_\theta B + B^\top P_\theta \Delta_B)(K_0 - K_\theta) - \Delta_K^\top (R_{\mathbf{u}} + B^\top P_\theta B)(K'_\theta) \\
 & \quad \left. - (K'_\theta)^\top (R_{\mathbf{u}} + B^\top P_\theta B) \Delta_K + (K_0 - K_\theta)^\top (\Delta_B^\top P_\theta B + B^\top P'_\theta B + B^\top P_\theta \Delta_B) \Delta_K \right)
 \end{aligned}$$

Since we have already shown that the operator norms of all terms in this expression are polynomial in problem parameters, we can bound

$$\|Q'_1(s_2)\|_{\text{op}} \leq \text{poly}(\Psi_{P_\star}, \Psi_{R_{\mathbf{u}}}, \Psi_{B_\star}, 1/\Psi_{B_\star})$$

Plugging these quantities into our bound on  $\text{tr}(M'(s_2))$ , it follows that

$$\begin{aligned}
 \text{tr}(M'(s_2)) & \leq d_x^2 \left( \Psi_{B_\star}^2 \Psi_{P_\star}^4 + (1 + \Psi_{B_\star}) \Psi_{R_{\mathbf{u}}} \Psi_{P_\star}^3 + \frac{\Psi_{R_{\mathbf{u}}} \Psi_{P_\star}}{\Psi_{B_\star}} \right) \\
 & \quad + d_x \text{poly}(\Psi_{P_\star}, \Psi_{R_{\mathbf{u}}}, \Psi_{B_\star}, 1/\Psi_{B_\star})
 \end{aligned}$$

As this holds regardless of  $\Delta_K, \Delta_\theta$ , and for all  $\theta_2$  with  $\|\theta_2 - \theta_\star\|_{\text{op}} \leq \min\{1/(150\Psi_{P_\star}^5), \Psi_{B_\star}/2\}$ , we have shown that

$$\begin{aligned}
 & |\nabla_K^2 \mathcal{R}(K; \theta_0)|_{K=K_0} [\Delta_K, \Delta_K] - \nabla_K^2 \mathcal{R}(K; \theta_1)|_{K=K_0} [\Delta_K, \Delta_K] \\
 & \leq |\nabla_\theta (\nabla_K^2 \mathcal{R}(K; \theta))|_{K=K_0} [\Delta_K, \Delta_K] |_{\theta=\theta_2} [\Delta_\theta] \cdot \|\theta_0 - \theta_1\|_{\text{op}} \\
 & \leq d_x^2 \left( \Psi_{B_\star}^2 \Psi_{P_\star}^4 + (1 + \Psi_{B_\star}) \Psi_{R_{\mathbf{u}}} \Psi_{P_\star}^3 + \frac{\Psi_{R_{\mathbf{u}}} \Psi_{P_\star}}{\Psi_{B_\star}} \right) \cdot \|\theta_0 - \theta_1\|_{\text{op}} \\
 & \quad + d_x \text{poly}(\Psi_{P_\star}, \Psi_{R_{\mathbf{u}}}, \Psi_{B_\star}, 1/\Psi_{B_\star}) \cdot \|\theta_0 - \theta_1\|_{\text{op}}
 \end{aligned}$$

from which the desired result follows.  $\square$

### K.3. Norm Bounds on Control Theoretic Quantities

**Lemma K.5** (Lemmas 3.2, B.3, B.8, and C.5 of (Simchowitz & Foster, 2020)). *Let  $\theta(t) = (A_\star + t\Delta_A, B_\star + t\Delta_B)$  and  $P(t) := P_\infty(\theta(t)), K(t) := K_{\text{opt}}(\theta(t))$ . If  $\max\{\|\Delta_A\|_{\text{op}}, \|\Delta_B\|_{\text{op}}\} \leq \epsilon$  and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ , then, for  $t$  where  $(A(t), B(t))$  is stabilizable,*

1.  $\|P'(t)\|_{\text{op}} \leq 4\|P(t)\|_{\text{op}}^3 \epsilon.$
2.  $\|P''(t)\|_{\text{op}} \leq \text{poly}(\|P(t)\|_{\text{op}}) \epsilon^2.$
3.  $\|K(t)\|_{\text{op}} \leq \sqrt{\|P(t)\|_{\text{op}}}.$
4.  $\|K'(t)\|_{\text{op}} \leq 7\|P(t)\|_{\text{op}}^{7/2} \epsilon.$
5.  $\|K''(t)\|_{\text{op}} \leq \text{poly}(\|P(t)\|_{\text{op}}) \epsilon^2.$

**Lemma K.6.** *Assume that  $\theta_\star$  is stabilizable and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ . Consider some alternate  $\theta = (A, B)$  with  $\|\theta - \theta_\star\|_{\text{op}} \leq \min\{1/(150\Psi_{P_\star}^5), \Psi_{B_\star}/2\}$  and a controller  $K_0$  with  $\|K_0 - K_{\text{opt}}(\theta_\star)\|_{\text{op}} \leq 1/(30\Psi_{B_\star}\Psi_{P_\star}^{3/2})$ . Denote  $P_\theta := P_\infty(\theta), K_\theta := K_{\text{opt}}(\theta)$ . Then the following are true.*

1.  $\Psi_{P_\star}, \|P_\theta\|_{\text{op}} \geq 1.$
2.  $\theta$  is stabilizable.
3.  $\Psi_{P_\star} \cong \|P_\theta\|_{\text{op}}$  and  $\|P_\theta\|_{\text{op}} \leq 5\sqrt{\frac{3}{71}}\Psi_{P_\star}.$
4.  $\|B\|_{\text{op}} \leq \frac{3}{2}\Psi_{B_\star}.$
5.  $\|(R_{\mathbf{u}} + B^\top P_\theta B)(K_\star - K_\theta)\|_{\text{op}} \leq c_1(\Psi_{R_{\mathbf{u}}}^{1/2} + \Psi_{P_\star}^{1/2}\Psi_{B_\star})/(\Psi_{P_\star}^{3/2}).$
6.  $\|(R_{\mathbf{u}} + B^\top P_\theta B)(K_0 - K_\star)\|_{\text{op}} \leq c_2(\Psi_{R_{\mathbf{u}}} + \Psi_{B_\star}^2\Psi_{P_\star})/(\Psi_{B_\star}\Psi_{P_\star}^{3/2}).$
7.  $\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}, \|\text{dlyap}(A + BK_0, I)\|_{\text{op}} \leq c_3\Psi_{P_\star}.$
8.  $\|A + BK_0\|_{\text{op}} \leq c_4\sqrt{\Psi_{P_\star}}.$

for absolute constants  $c_1, c_2, c_3, c_4$  and where  $\cong$  denotes equality up to absolute constants.

*Proof.* First, note that  $R_{\mathbf{x}} \succeq I$  implies  $\Psi_{P_\star}, \|P_\theta\|_{\text{op}} \geq 1$  by Lemma 4.2 of (Simchowitz & Foster, 2020), and that  $\|A_\star - A\|_{\text{op}}, \|B_\star - B\|_{\text{op}} \leq \|\theta - \theta_\star\|_{\text{op}}$ . Given our assumption on  $\|\theta - \theta_\star\|_{\text{op}}$ , it follows that  $8\Psi_{P_\star}^2\|\theta - \theta_\star\|_{\text{op}} \leq 8/(150\Psi_{P_\star}^3) \leq \frac{4}{75} < 1$ . Proposition 6 of (Simchowitz & Foster, 2020) then implies that  $\theta$  is stabilizable and that

$$\|P_\theta\|_{\text{op}} \leq (1 - 4/75)^{-1/2}\Psi_{P_\star} \leq 5\sqrt{\frac{3}{71}}\Psi_{P_\star}$$

Applying the same bound in the opposite direction, we have that  $8\|P_\theta\|_{\text{op}}^2\|\theta - \theta_\star\|_{\text{op}} \leq 8\frac{75}{71}\Psi_{P_\star}^2\|\theta - \theta_\star\|_{\text{op}} \leq 8\frac{75}{71}\frac{1}{150} = \frac{4}{71} < 1$ , so

$$\Psi_{P_\star} \leq (1 - 4/71)^{-1/2}\|P_\theta\|_{\text{op}} \leq \sqrt{\frac{71}{67}}\|P_\theta\|_{\text{op}}$$

From which 3. follows.

If  $\|B(K_0 - K_\theta)\|_{\text{op}} \leq 1/(5\|P_\theta\|_{\text{op}})^{3/2}$ ,  $A + BK_0$  is stable, by Proposition 7 of (Simchowitz & Foster, 2020). By Proposition 6 of (Simchowitz & Foster, 2020),  $\|B(K_0 - K_\theta)\|_{\text{op}} \leq \|B(K_0 - K_\star)\|_{\text{op}} + \|B(K_\star - K_\theta)\|_{\text{op}} \leq \|B\|_{\text{op}}\|K_0 - K_\star\|_{\text{op}} + c\|P_\theta\|_{\text{op}}^{7/2}\|\theta - \theta_\star\|_{\text{op}} \leq 1/(5\|P_\theta\|_{\text{op}}^{3/2})$ , where the last inequality holds by our bounds on  $\|K_0 - K_\star\|_{\text{op}}$  and  $\|\theta - \theta_\star\|_{\text{op}}$ , and plugging in the appropriate constants. Thus,  $A + BK_0$  is stable. Then the following hold.

- $\|B\|_{\text{op}} \leq \Psi_{B_\star} + \|\theta - \theta_\star\|_{\text{op}} \leq \frac{3}{2}\Psi_{B_\star}$ , where the last inequality holds since  $\|\theta - \theta_\star\|_{\text{op}} \leq \Psi_{B_\star}/2$ .

- Let  $\zeta = 8\|P_\theta\|_{\text{op}}^2\|\theta - \theta_\star\|_{\text{op}}$ . Since  $\zeta < 1/2$  and  $\|\theta - \theta_\star\|_{\text{op}} \leq 1/(32\|P_\theta\|_{\text{op}}^3)$ , then

$$\begin{aligned} \|(K_\star - K_\theta)^\top(R_{\mathbf{u}} + B^\top P_\theta B)\|_{\text{op}} &\leq \|R_{\mathbf{u}}^{1/2}\|_{\text{op}}\|R_{\mathbf{u}}^{1/2}(K_\star - K_\theta)\|_{\text{op}} + \|P_\theta^{1/2}B\|_{\text{op}}\|P_\theta^{1/2}B(K_\star - K_\theta)\|_{\text{op}} \\ &\stackrel{(a)}{\leq} 4(7\|R_{\mathbf{u}}^{1/2}\|_{\text{op}} + 9\|P_\theta^{1/2}B\|_{\text{op}})\|P_\theta\|_{\text{op}}^{7/2}\|\theta - \theta_\star\|_{\text{op}} \\ &\leq \frac{c(\Psi_{R_{\mathbf{u}}}^{1/2} + \|P_\theta\|_{\text{op}}^{1/2}\|B\|_{\text{op}})}{\|P_\theta\|_{\text{op}}^{3/2}} \\ &\leq \frac{c(\Psi_{R_{\mathbf{u}}}^{1/2} + \Psi_{P_\star}^{1/2}\Psi_{B_\star})}{\Psi_{P_\star}^{3/2}} \end{aligned}$$

(a) holds by Proposition 6 of (Simchowitz & Foster, 2020).

- Since  $\|K_0 - K_\star\|_{\text{op}} \leq 1/(c\|B\|_{\text{op}}\|P_\theta\|_{\text{op}}^{3/2})$

$$\begin{aligned} \|(K_0 - K_\star)^\top(R_{\mathbf{u}} + B^\top P_\theta B)\|_{\text{op}} &= \|K_0 - K_\star\|_{\text{op}}\|R_{\mathbf{u}} + B^\top P_\theta B\|_{\text{op}} \\ &\leq \frac{\|R_{\mathbf{u}} + B^\top P_\theta B\|_{\text{op}}}{c\|B\|_{\text{op}}\|P_\theta\|_{\text{op}}^{3/2}} \leq \frac{c(\Psi_{R_{\mathbf{u}}} + \Psi_{B_\star}^2\Psi_{P_\star})}{\Psi_{B_\star}\Psi_{P_\star}^{3/2}} \end{aligned}$$

- By Lemma B.5,  $\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}} \leq \|P_\theta\|_{\text{op}} \leq c\Psi_{P_\star}$ , so long as  $R_{\mathbf{x}} \succeq I$ .
- By Lemma B.12 of (Simchowitz & Foster, 2020), if  $\|B(K_0 - K_\theta)\|_{\text{op}} \leq 1/(5\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}^{3/2})$ , then  $\|\text{dlyap}(A + BK_0, I)\|_{\text{op}} \leq 2\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}$ . Note that  $\|B(K_0 - K_\theta)\|_{\text{op}} \leq 1/(5\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}^{3/2})$  holds since  $\|B(K_0 - K_\theta)\|_{\text{op}} \leq 1/(5\|P_\theta\|_{\text{op}}^{3/2}) \leq 1/(5\|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}^{3/2})$ .
- Note that  $\|A + BK_\theta\|_{\text{op}}^2 \leq \|\text{dlyap}(A + BK_\theta, I)\|_{\text{op}}$ , so, as long as  $\zeta \leq 1/2$ ,

$$\begin{aligned} \|A + BK_0\|_{\text{op}} &\leq \|A + BK_\theta\|_{\text{op}} + \|B(K_0 - K_\theta)\|_{\text{op}} \leq \sqrt{\|P_\theta\|_{\text{op}} + t\|B\|_{\text{op}}} + 32\|P_\theta\|_{\text{op}}^{7/2}\|\theta - \theta_\star\|_{\text{op}} \\ &\leq c\sqrt{\Psi_{P_\star}} + \frac{c}{\Psi_{P_\star}^{3/2}} \leq c\sqrt{\Psi_{P_\star}}. \end{aligned}$$

□

**Lemma K.7.** Let  $\theta(t) = (A_\star + t\Delta_A, B_\star + t\Delta_B)$  and  $P(t) := P_\infty(\theta(t))$ ,  $K(t) := K_{\text{opt}}(\theta(t))$ ,  $A_{\text{cl}}(t) := A(t) + B(t)K(t)$ . If  $\max\{\|A_\star - \hat{A}\|_{\text{op}}, \|B_\star - \hat{B}\|_{\text{op}}\} \leq \epsilon$  and  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ , then for  $t$  such that  $(A(t), B(t))$  is stabilizable:

$$\|P'''(t)\|_{\text{op}} \leq (1 + \|A_{\text{cl}}(t)\|_{\text{op}})(1 + \|B(t)\|_{\text{op}})\text{poly}(\|P(t)\|_{\text{op}})\epsilon^3.$$

*Proof.* For simplicity, we drop the  $t$  throughout the remainder of the proof. By Lemma C.2 of (Simchowitz & Foster, 2020),  $P'' = \text{dlyap}(A_{\text{cl}}, Q_2)$  where  $Q_2 = A_{\text{cl}}^\top P' A_{\text{cl}} + A_{\text{cl}}^\top P' A_{\text{cl}}' + Q_1'$ ,  $Q_1' = A_{\text{cl}}^\top P \Delta_{A_{\text{cl}}} + A_{\text{cl}}^\top P' \Delta_{A_{\text{cl}}} + A_{\text{cl}}^\top P B' K' + (B' K')^\top P A_{\text{cl}} + \Delta_{A_{\text{cl}}}^\top P' A_{\text{cl}} + \Delta_{A_{\text{cl}}}^\top P A_{\text{cl}}'$ , and  $\Delta_{A_{\text{cl}}}(t) = A'(t) + B'(t)K_{\text{opt}}(A(t), B(t)) = \Delta_A + \Delta_B K_{\text{opt}}(A(t), B(t))$ . By the definition of  $\text{dlyap}$ :

$$P'' = A_{\text{cl}}^\top P'' A_{\text{cl}} + Q_2$$

so:

$$P''' = A_{\text{cl}}^\top P'' A_{\text{cl}} + A_{\text{cl}}^\top P'' A_{\text{cl}}' + A_{\text{cl}}^\top P''' A_{\text{cl}} + Q_2' = \text{dlyap}(A_{\text{cl}}, A_{\text{cl}}^\top P'' A_{\text{cl}} + A_{\text{cl}}^\top P'' A_{\text{cl}}' + Q_2')$$

By Lemma B.5 of (Simchowitz & Foster, 2020),  $\|P'''\|_{\text{op}} \leq \|P\|_{\text{op}}\|A_{\text{cl}}^\top P'' A_{\text{cl}} + A_{\text{cl}}^\top P'' A_{\text{cl}}' + Q_2'\|_{\text{op}} \leq \|P\|_{\text{op}}(2\|A_{\text{cl}}^\top P'' A_{\text{cl}}\|_{\text{op}} + \|Q_2'\|_{\text{op}})$ .  $A_{\text{cl}}' = \Delta_A + \Delta_B K + BK'$  so  $\|A_{\text{cl}}'\|_{\text{op}} \leq (\sqrt{\|P\|_{\text{op}}} + 7\|B\|_{\text{op}}\|P\|_{\text{op}}^{7/2})\epsilon$ . By Lemma K.5,  $\|A_{\text{cl}}^\top P'' A_{\text{cl}}\|_{\text{op}} \leq \|A_{\text{cl}}\|_{\text{op}}(1 + \|B\|_{\text{op}})\text{poly}(\|P\|_{\text{op}})\epsilon^3$ .

It remains to upper bound the operator norm of  $Q_2'$ . By definition of  $Q_2$ , we see, for small absolute constant  $c$ :

$$\|Q_2'\|_{\text{op}} \leq c\left(\|A_{\text{cl}}^\top P' A_{\text{cl}}'\|_{\text{op}} + \|A_{\text{cl}}^\top P' A_{\text{cl}}'\|_{\text{op}} + \|A_{\text{cl}} P'' A_{\text{cl}}'\|_{\text{op}} + \|A_{\text{cl}}''^\top P \Delta_{A_{\text{cl}}}\|_{\text{op}} + \|A_{\text{cl}}^\top P' \Delta_{A_{\text{cl}}}\|_{\text{op}}\right)$$

$$\begin{aligned}
 & + \|A'_{\text{cl}}{}^\top P \Delta'_{A_{\text{cl}}}\|_{\text{op}} + \|A'_{\text{cl}}{}^\top P'' \Delta_{A_{\text{cl}}}\|_{\text{op}} + \|A'_{\text{cl}}{}^\top P' \Delta'_{A_{\text{cl}}}\|_{\text{op}} + \|A'_{\text{cl}}{}^\top P B' K'\|_{\text{op}} \\
 & + \|A_{\text{cl}}{}^\top P' B' K'\|_{\text{op}} + \|A_{\text{cl}}{}^\top P B' K''\|_{\text{op}}
 \end{aligned}$$

By Lemma K.5,  $\|A'_{\text{cl}}\|_{\text{op}} = \|\Delta_B K' + B K''\|_{\text{op}} \leq 7\|P\|_{\text{op}}^{7/2} \epsilon^2 + \|B\|_{\text{op}} \text{poly}(\|P\|_{\text{op}}) \epsilon^2$ . By definition,  $\|\Delta_{A_{\text{cl}}}\|_{\text{op}} \leq \epsilon$ ,  $\|B'\|_{\text{op}} \leq \epsilon$ . Finally,  $\Delta'_{A_{\text{cl}}} = \Delta_B K'$  so  $\|\Delta'_{A_{\text{cl}}}\|_{\text{op}} \leq 7\|P\|_{\text{op}}^{7/2} \epsilon^2$ . We see then that, by Lemma K.5, every term in the above sum is order  $\epsilon^3$  so, combining everything, we have:

$$\|P''\|_{\text{op}} \leq (1 + \|A_{\text{cl}}\|_{\text{op}})(1 + \|B\|_{\text{op}}) \text{poly}(\|P\|_{\text{op}}) \epsilon^3$$

□

**Lemma K.8.** *Let  $\theta(t) = (A_\star + t\Delta_A, B_\star + t\Delta_B)$  and  $K(t) := K_{\text{opt}}(\theta(t))$ ,  $A_{\text{cl}}(t) := A(t) + B(t)K(t)$ . If  $\max\{\|A_\star - \widehat{A}\|_{\text{op}}, \|B_\star - \widehat{B}\|_{\text{op}}\} \leq \epsilon$ ,  $R_{\mathbf{x}}, R_{\mathbf{u}} \succeq I$ , then for  $t$  such that  $(A(t), B(t))$  is stabilizable:*

$$\|K'''(t)\|_{\text{op}} \leq \text{poly}(\|P(t)\|_{\text{op}}, \|B(t)\|_{\text{op}}, \|A_{\text{cl}}(t)\|_{\text{op}}) \epsilon^3.$$

*Proof.* As before, we drop the  $t$  throughout the remainder of the proof. By Lemma B.3 of (Simchowitz & Foster, 2020):

$$K'' = R_0^{-1} Q'_3(t) + R_0^{-1} (R_{\mathbf{u}} + B^\top P B)' K'$$

where  $Q_3 = \Delta_B^\top P A_{\text{cl}} + B^\top P \Delta_{A_{\text{cl}}} + B^\top P' A_{\text{cl}}$  and  $R_0 = R_{\mathbf{u}} + B^\top P B$ . So, using the identity  $(X^{-1})' = -X^{-1} X' X^{-1}$ :

$$\begin{aligned}
 K''' &= R_0^{-1} Q_3'' + R_0^{-1} (R_{\mathbf{u}} + B^\top P B)' R_0^{-1} Q'_3 + R_0^{-1} (R_{\mathbf{u}} + B^\top P B)' R_0^{-1} (R_{\mathbf{u}} + B^\top P B)' K' \\
 &+ R_0^{-1} (R_{\mathbf{u}} + B^\top P B)'' K' + R_0^{-1} (R_{\mathbf{u}} + B^\top P B)' K''
 \end{aligned}$$

By Lemma C.3 of (Simchowitz & Foster, 2020),  $\|R_0^{-1}\|_{\text{op}} \leq 1$ .  $(R_{\mathbf{u}} + B^\top P B)' = \Delta_B^\top P B + B^\top P' B + B^\top P' \Delta_B$  so by Lemma K.5,  $\|(R_{\mathbf{u}} + B^\top P B)'\|_{\text{op}} \leq \text{poly}(\|B\|_{\text{op}}, \|P\|_{\text{op}}) \epsilon$ . Similarly, we see that  $\|(R_{\mathbf{u}} + B^\top P B)''\|_{\text{op}} \leq \text{poly}(\|B\|_{\text{op}}, \|P\|_{\text{op}}) \epsilon^2$ . Using Lemma K.5 to bound  $\|K'\|_{\text{op}}$  and  $\|K''\|_{\text{op}}$ , we have:

$$\|K'''\|_{\text{op}} \leq \|Q_3''\|_{\text{op}} + \|Q'_3\|_{\text{op}} \text{poly}(\|B\|_{\text{op}}, \|P\|_{\text{op}}) \epsilon + \text{poly}(\|B\|_{\text{op}}, \|P\|_{\text{op}}) \epsilon^3$$

It remains to bound  $\|Q_3''\|_{\text{op}}$  and  $\|Q'_3\|_{\text{op}}$ . By definition:

$$Q'_3 = 2\Delta_B^\top P' A_{\text{cl}} + \Delta_B^\top P A'_{\text{cl}} + \Delta_B^\top P \Delta_{A_{\text{cl}}} + B^\top P' \Delta_{A_{\text{cl}}} + B^\top P \Delta'_{A_{\text{cl}}} + B^\top P'' A_{\text{cl}} + B^\top P' A'_{\text{cl}}$$

Using previously computed norm bounds, we have  $\|Q'_3\|_{\text{op}} \leq \text{poly}(\|B\|_{\text{op}}, \|P\|_{\text{op}}) \epsilon^2$ . To bound  $Q_3''$ , we can differentiate the above, obtaining that, for a small absolute constant  $c$ :

$$\begin{aligned}
 \|Q_3''\|_{\text{op}} &\leq c \left( \|\Delta_B^\top P'' A_{\text{cl}}\|_{\text{op}} + \|\Delta_B^\top P' A'_{\text{cl}}\|_{\text{op}} + \|\Delta_B^\top P A''_{\text{cl}}\|_{\text{op}} + \|\Delta_B^\top P' \Delta_{A_{\text{cl}}}\|_{\text{op}} + \|\Delta_B^\top P \Delta'_{A_{\text{cl}}}\|_{\text{op}} \right. \\
 &+ \|B^\top P'' \Delta_{A_{\text{cl}}}\|_{\text{op}} + \|B^\top P' \Delta'_{A_{\text{cl}}}\|_{\text{op}} + \|B^\top P \Delta''_{A_{\text{cl}}}\|_{\text{op}} + \|B^\top P''' A_{\text{cl}}\|_{\text{op}} \\
 &\left. + \|B^\top P'' A'_{\text{cl}}\|_{\text{op}} + \|B^\top P' A''_{\text{cl}}\|_{\text{op}} \right)
 \end{aligned}$$

$\|\Delta'_{A_{\text{cl}}}\|_{\text{op}} \leq \|\Delta_B K''\|_{\text{op}} \leq \text{poly}(\|P\|_{\text{op}}) \epsilon^3$ . Then, using Lemma K.7 and previously computed norm bounds, we have:

$$\|Q_3''\|_{\text{op}} \leq \text{poly}(\|P\|_{\text{op}}, \|B\|_{\text{op}}, \|A_{\text{cl}}\|_{\text{op}}) \epsilon^3$$

Combining everything gives the stated result. □

## L. Provable Gains for Task-Optimal Design

### L.1. Preliminaries for Comparison of Designs

**Asymptotic Notation.** We assume that  $\rho$  is close to 1, and are primarily concerned with the scaling in  $d_x$  and  $\frac{1}{1-\rho}$ . As such, we let  $\Theta_\rho(\cdot)$  hide numerical constants and terms lower order in  $\frac{1}{1-\rho}$ . That is, we write  $x = \Theta_\rho(\frac{1}{(1-\rho)^n})$  if  $x = \frac{c_n}{(1-\rho)^n} +$

$\sum_{j=1}^{n-1} \frac{c_j}{(1-\rho)^j}$ , for numerical constants  $c_1, \dots, c_n$ . Similarly, we write  $x = \Theta_\rho(\frac{-1}{(1-\rho)^n})$  if  $x = \frac{-c_n}{(1-\rho)^n} + \sum_{j=1}^{n-1} \frac{c_j}{(1-\rho)^j}$  for  $c_n > 0$ . In addition, we let  $\approx_\rho$  denote that two quantities have the same scaling in  $1 - \rho$ , up to absolute constants.

$\Theta_\rho(\cdot)$  does not suppress dimension dependence, but in some cases it will be convenient to hide dimension dependence that is lower order in  $\frac{1}{1-\rho}$ . In such cases, to make clear that we are doing this, we will use  $\Theta_{\rho, d_x}(\cdot)$ . Formally,  $x = \Theta_{\rho, d_x}(\frac{1}{(1-\rho)^n} + \frac{d_x^m}{(1-\rho)^p})$  if  $x = \frac{c_n}{(1-\rho)^n} + \sum_{j=1}^{n-1} \frac{c_j}{(1-\rho)^j} + \frac{d_x^m}{(1-\rho)^p} + \sum_{j=1}^{p-1} \frac{b_j d_x^m}{(1-\rho)^j}$ . In general we will only apply this notation to the final sample complexities when we are concerned with identifying the leading terms. We will also use  $\mathcal{O}(\cdot)$  in its standard form, suppressing lower order dependence on  $d_x$  and  $\frac{1}{1-\rho}$ .

**Convex Representation of Inputs.** Recall that Theorem 2.1 show the task complexity achieves by any policy scales as  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1})$  (where, throughout this section, we let  $\Gamma_T(\pi; \theta_\star)$  denote the expected covariates under policy  $\pi$ ). In particular, if we are playing periodic policies, as Lemma H.4 shows we can approximate  $\Gamma_T(\pi; \theta_\star)$  with  $\Gamma_T^{\text{ss}}(\pi; \theta_\star)$  so the complexity scales instead as  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{ss}}(\pi; \theta_\star)^{-1})$ . Similarly, if we are playing only noise, the complexity scales as  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{noise}}(\pi; \theta_\star)^{-1})/T$ .

Throughout, we will assume that  $\gamma^2 \gg \sigma^2$  so we ignore the contribution of process noise to  $\Gamma_T^{\text{ss}}(\pi; \theta_\star)$ . If our policy plays a periodic input  $U = (U_\ell)_{\ell=1}^k \in \mathcal{U}_{\gamma^2, k}$ , recall that

$$\Gamma_T^{\text{ss}}(\pi; \theta_\star) = \sum_{\ell=1}^k (e^{t \frac{2\pi\ell}{k}} I - \tilde{A}_\star)^{-1} \tilde{B}_\star U_\ell \tilde{B}_\star^H (e^{t \frac{2\pi\ell}{k}} I - \tilde{A}_\star)^{-H}$$

Some algebra shows that

$$(e^{t \frac{2\pi\ell}{k}} I - \tilde{A}_\star)^{-1} \tilde{B}_\star = \begin{bmatrix} (e^{t \frac{2\pi\ell}{k}} I - A_\star)^{-1} B_\star \\ I \end{bmatrix}$$

As we have already shown, a general matrix signal  $U$  can be realized in the time domain via a certain decomposition (see Appendix B.5.3) so in the following we will consider optimizing over  $U_\ell$  so that  $\Gamma_T^{\text{ss}}(\pi; \theta_\star)$  satisfies our objective.

We will consider the performance of TOPLE, optimal operator-norm identification, optimal Frobenius norm identification, and optimal noise excitation. By construction, TOPLE plays periodic inputs. Similarly, note that the optimal Frobenius norm identification algorithm is itself an instance of TOPLE—with  $\mathcal{H}_{\text{fro}}(\theta_\star) = I$ —so we can assume that the optimal Frobenius norm identification algorithm also plays periodic inputs. As (Wagenmaker & Jamieson, 2020) show, the optimal operator-norm identification also plays periodic inputs. In all cases, then, we will consider the complexity  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{ss}}(\pi; \theta_\star)^{-1})$ . For optimal noise excitation, we consider instead the complexity  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{noise}}(\pi; \theta_\star)^{-1})/T$ .

**Simplifications.** Note that due to the structure of  $\Gamma_T$ , we have

$$\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1}) = \sum_{i=1}^d \text{tr}(\mathcal{H}_i \Gamma_T(\pi; \theta_\star)^{-1})$$

where we let  $\mathcal{H}_i := [\mathcal{H}(\theta_\star)]_{(i-1)(d+p)+1:i(d+p), (i-1)(d+p)+1:i(d+p)}$ , the  $i$ th  $(d+p) \times (d+p)$  block diagonal element of  $\mathcal{H}(\theta_\star)$ . We will denote elements of  $\mathcal{H}(\theta_\star)$  with  $[\mathcal{H}(\theta_\star)]_{A_{ij}, A_{nm}}$ , where

$$[\mathcal{H}(\theta_\star)]_{A_{ij}, A_{nm}} = \text{vec}(\theta_{ij})^\top \mathcal{H}(\theta_\star) \text{vec}(\theta_{nm})$$

and  $\theta_{ij} = (e_i e_j^\top, 0)$  (and similarly for  $B$ ). See Appendix G.3 for a more in-depth discussion of the vectorization of linear dynamical systems. By our construction of  $\mathcal{H}(\theta_\star)$ , the elements  $[\mathcal{H}(\theta_\star)]_{A_{nm}, A_{nm}}$  and  $[\mathcal{H}(\theta_\star)]_{B_{nm}, B_{nm}}$  will lie on the diagonal of  $\mathcal{H}(\theta_\star)$ , while other elements will not. Furthermore, elements of the form  $[\mathcal{H}(\theta_\star)]_{A_{nm}, A_{mn}}$  (and replacing  $A$  with  $B$ ) with  $m \neq n$  will not be contained in any  $\mathcal{H}_i$ , as these will lie off the block-diagonal. It follows that the expressions given above include all entries of  $\mathcal{H}(\theta_\star)$  that will appear in our calculations.

**Computation of Inputs.** For each exploration strategy, our goal will be to compute the inputs  $\pi$  optimal for a given exploration criteria, and then compute the value of  $\Phi_T(\pi; \theta_\star) = \text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1})$  for this input. As Theorem 2.1 shows, the task complexity of any given exploration strategy  $\pi$  scales as  $\Phi_T(\pi; \theta_\star) = \text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1})$ .

Note that the optimal Frobenius norm identification algorithm is simply TOPLE, but with  $\mathcal{H}(\theta_\star) = I$ . Thus, for the optimal task-specific strategy and the Frobenius norm identification strategy, we can compute the optimal inputs by choosing those inputs which minimize  $\Phi_T(\pi; \theta_\star) = \text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1})$ , for each  $\mathcal{H}(\theta_\star)$ . By Lemma G.3 we have that  $\Phi_{\text{opt}}(\gamma^2; \theta_\star)$  and  $\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star)$  are equivalent up to constants, where

$$\Phi_{\text{opt}}^{\text{ss}}(\gamma^2; \theta_\star) := \liminf_{T \rightarrow \infty} \min_{u \in \mathcal{U}_{\gamma^2, T}} \text{tr}(\mathcal{H}(\theta_\star)\Gamma_{T,T}^{\text{ss}}(\tilde{\theta}_\star, u, 0)^{-1})$$

Note that this corresponds to the covariates obtained when playing an input that is only sinusoidal and has no noise component. Furthermore, as TOPLE is optimal and itself plays periodic inputs, it suffices to consider only periodic inputs. Putting this together, for large enough  $T$ , for the task-optimal and Frobenius norm strategies, we simply analyze  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{ss}}(\pi; \theta_\star)^{-1})$  and only consider periodic, non-noise inputs.

The case of operator norm identification is similar. As is shown in (Wagenmaker & Jamieson, 2020), the optimal inputs here are also periodic, so it follows that  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1}) \approx_\rho \text{tr}(\mathcal{H}(\theta_\star)\Gamma_T^{\text{ss}}(\pi; \theta_\star)^{-1})$ , for  $\pi$  the optimal operator norm inputs. Furthermore, an argument similar to that used in the proof of Theorem B.2 can be used to show that the optimal inputs are sinusoidal with no noise component. It follows that we can, in this case, also simply analyze the steady-state covariates with no noise component.

Note that TOPLE itself does mix the sinusoidal input with a noise component to ensure sufficient excitation. While the above argument shows that this does not improve the complexity of Frobenius or operator norm identification evaluated in the Frobenius or operator norm, one might hope that the inclusion of noise would help these exploration strategies more easily transfer to the actual task of interest. We make two remarks on this. First, as we are concerned with the inputs optimal on Frobenius and operator norm identification, and as these inputs do not require this noise component, the result we obtain would still hold even if this was the case. Second, our analysis shows that if we play the noise that optimally excites the system for completing the task of interest, the complexity obtained is still suboptimal. Thus, even if we were to mix the operator or Frobenius norm identification inputs with the optimal noise, the resulting strategy would still be suboptimal, so our conclusion holds regardless of whether noise is played or not.

Finally, to simplify the analysis further we ignore the contribution from the excitation due to the process noise when computing the covariates. This is reasonable for small values of  $\sigma_w^2$ , which will make this contribution lower order.

## L.2. Computation of Task Hessian

**Lemma L.1** (Computation of Task Hessian). *Consider the following instance of the LQR problem:*

$$A_\star = \rho_1 e_1 e_1^\top + \rho_2 (I - e_1 e_1^\top), \quad B_\star = bI, \quad R_x = \kappa_1 e_1 e_1^\top + \kappa_2 (I - e_1 e_1^\top), \quad R_u = \mu I$$

Then, if  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nm}, A_{nm}} = \frac{1}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( \frac{b p_n a_n^2}{1 - a_n a_m} \right)^2$$

and if  $n = m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nn}, A_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( \frac{b p_n (1 + a_n^2)}{1 - a_n^2} \right)^2$$

If  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{B_{nm}, B_{nm}} = \frac{1}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n k_m}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n a_n^2 k_m}{1 - a_n a_m} \right)^2$$

and if  $n = m$ ,

$$[\mathcal{H}(\theta_\star)]_{B_{nn}, B_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n k_n (1 + a_n^2)}{1 - a_n^2} \right)^2$$

Finally, if  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nm}, B_{nm}} = \frac{-k_m}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n a_n^2 k_m}{1 - a_n a_m} \right) \frac{b p_n a_n^2}{1 - a_n a_m}$$



and if  $n = m$ ,

$$[\mathcal{H}(\theta_*)]_{A_{nn}, B_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( \frac{b p_n (1 + a_n^2)}{1 - a_n^2} \right) \left( p_n a_n - \frac{b p_n k_n (1 + a_n^2)}{1 - a_n^2} \right).$$

Furthermore, all other elements of  $\mathcal{H}(\theta_*)$  which appear in  $\sum_{i=1}^d \text{tr}(\mathcal{H}_i \Gamma_T(\pi; \theta_*)^{-1})$  are 0.

*Proof.* Note that here  $P_*$  is diagonal with diagonal elements  $p_i = [P_*]_{ii}$  satisfying

$$p_i = \rho_i^2 p_i - \frac{\rho_i^2 b^2 p_i^2}{\mu + b^2 p_i} + \kappa_i \implies p_i = \frac{1}{2b^2} \left( b^2 \kappa_i - \mu + \mu \rho_i^2 + \sqrt{4b^2 \mu \kappa_i + (\mu - b^2 \kappa_i - \mu \rho_i^2)^2} \right)$$

where we set  $\rho_i = \rho_2, \kappa_i = \kappa_2, i \geq 2$ . It follows that  $K_*$  is diagonal and that  $k_i := [K_*]_{ii} = \frac{\rho_i b p_i}{\mu + b^2 p_i}$ , so  $A_{\text{cl},*}$  is also diagonal. Let  $a_i := [A_{\text{cl},*}]_{ii}$ .

By definition,  $\mathcal{H}(\theta_*) = (\nabla_{\theta} K_{\text{opt}}(\theta)|_{\theta=\theta_*})^{\top} (\nabla_K^2 \mathcal{R}(K; \theta_*)|_{K=K_{\text{opt}}(\theta_*)}) (\nabla_{\theta} K_{\text{opt}}(\theta)|_{\theta=\theta_*})$ . Our goal is to calculate how  $\mathcal{H}(\theta_*)$  scales with the problem parameters, and from this determine the error rate of different exploration approaches. To this end, let  $A(t_1, t_2) = A_* + t_1 \Delta_A^1 + t_2 \Delta_A^2$ ,  $B(t_1, t_2) = B_* + t_1 \Delta_B^1 + t_2 \Delta_B^2$ ,  $\Delta_{\theta}^1 = (\Delta_A^1, \Delta_B^2)$ ,  $\Delta_{\theta}^2 = (\Delta_A^2, \Delta_B^2)$ , and  $K(t_1, t_2) = K_{\text{opt}}(A(t_1, t_2), B(t_1, t_2))$ . Then by the chain rule,

$$\begin{aligned} \frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(K(t_1, t_2); \theta_*)|_{t_1=t_2=0} &= (\nabla_{\theta} K_{\text{opt}}(\theta)|_{\theta=\theta_*} [\Delta_{\theta}^1])^{\top} (\nabla_K^2 \mathcal{R}(K; \theta_*)|_{K=K_*}) (\nabla_{\theta} K_{\text{opt}}(\theta)|_{\theta=\theta_*} [\Delta_{\theta}^2]) \\ &= \text{vec}(\Delta_{\theta}^1)^{\top} \mathcal{H}(\theta_*) \text{vec}(\Delta_{\theta}^2) \end{aligned}$$

Thus, to determine the value of  $\mathcal{H}(\theta_*)$ , we can simply evaluate  $\frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(K(t_1, t_2); \theta_*)|_{t_1=t_2=0}$  for different values of  $\Delta_{\theta}^1, \Delta_{\theta}^2$ . Now recall that,

$$\begin{aligned} \mathcal{R}(K; \theta_*) &= \text{tr}(\text{dlyap}(A_* + B_* K, (K - K_*)^{\top} (R_{\mathbf{u}} + B_*^{\top} P_* B_*) (K - K_*))) \\ &= \text{tr} \left( \sum_{s=0}^{\infty} (A_* + B_* K)^s (K - K_*)^{\top} (R_{\mathbf{u}} + B_*^{\top} P_* B_*) (K - K_*) ((A_* + B_* K)^{\top})^s \right) \end{aligned}$$

Setting  $K = K(t_1, t_2)$  and differentiating this with respect to  $t_1, t_2$ , we find that

$$\begin{aligned} \text{vec}(\Delta_{\theta}^1)^{\top} \mathcal{H}(\theta_*) \text{vec}(\Delta_{\theta}^2) &= \frac{d}{dt_2} \frac{d}{dt_1} \mathcal{R}(K(t_1, t_2); \theta_*)|_{t_1=t_2=0} \\ &= \text{tr} \left( \sum_{s=0}^{\infty} A_{\text{cl},*}^s (K^{t_1}(0, 0))^{\top} (\mu I + b^2 P_*) (K^{t_2}(0, 0)) A_{\text{cl},*}^s \right) \\ &= \sum_{s=0}^{\infty} \text{tr} \left( A_{\text{cl},*}^{2s} (K^{t_1}(0, 0))^{\top} (\mu I + b^2 P_*) (K^{t_2}(0, 0)) \right) \\ &= \sum_{s=0}^{\infty} \sum_{i=1}^{d_x} a_i^{2s} [K^{t_1}(0, 0)^{\top} (\mu I + b^2 P_*) K^{t_2}(0, 0)]_{ii} \\ &= \sum_{i=1}^{d_x} \frac{[K^{t_1}(0, 0)^{\top} (\mu I + b^2 P_*) K^{t_2}(0, 0)]_{ii}}{1 - a_i^2} \end{aligned}$$

Recall that

$$K^{t_i}(0, 0) = -(\mu I + b^2 P_*)^{-1} \left( (\Delta_B^i)^{\top} P_* A_{\text{cl},*} + b P_* \Delta_{A_{\text{cl}}}^i + b P^{t_i}(0, 0) A_{\text{cl},*} \right)$$

where  $\Delta_{A_{\text{cl}}}^i = \Delta_A^i - \Delta_B^i K_*$ , and

$$P^{t_i}(0, 0) = \text{dlyap}(A_{\text{cl},*}, Q_i) = \sum_{s=0}^{\infty} A_{\text{cl},*}^s Q_i A_{\text{cl},*}^s, \quad Q_i = A_{\text{cl},*}^{\top} P_* \Delta_{A_{\text{cl}}}^i + (\Delta_{A_{\text{cl}}}^i)^{\top} P_* A_{\text{cl},*}$$

Thus,

$$\begin{aligned} [K^{t_1}(0, 0)^{\top} (\mu I + b^2 P_*) K^{t_2}(0, 0)]_{ii} &= \left[ \left( (\Delta_B^1)^{\top} P_* A_{\text{cl},*} + b P_* \Delta_{A_{\text{cl}}}^1 + b P^{t_1}(0, 0) A_{\text{cl},*} \right)^{\top} \cdot (\mu I + b^2 P_*)^{-1} \right. \\ &\quad \left. \cdot \left( (\Delta_B^2)^{\top} P_* A_{\text{cl},*} + b P_* \Delta_{A_{\text{cl}}}^2 + b P^{t_2}(0, 0) A_{\text{cl},*} \right) \right]_{ii} \end{aligned}$$

$$\begin{aligned}
 &= [(\Delta_B^1)^\top P_\star A_{cl,\star} + bP_\star \Delta_{A_{cl}}^1 + bP^{t_1}(0,0)A_{cl,\star}]^\top \cdot (\mu I + b^2 P_\star)^{-1} \\
 &\quad \cdot [(\Delta_B^2)^\top P_\star A_{cl,\star} + bP_\star \Delta_{A_{cl}}^2 + bP^{t_2}(0,0)A_{cl,\star}]_{:,i} \\
 &= (p_i a_i [\Delta_B^1]_{i,:} + bP_\star [\Delta_{A_{cl}}^1]_{:,i} + b a_i [P^{t_1}(0,0)]_{:,i})^\top \cdot (\mu I + b^2 P_\star)^{-1} \\
 &\quad \cdot (p_i a_i [\Delta_B^2]_{i,:} + bP_\star [\Delta_{A_{cl}}^2]_{:,i} + b a_i [P^{t_2}(0,0)]_{:,i})
 \end{aligned}$$

and

$$\begin{aligned}
 [P^{t_1}(0,0)]_{:,i} &= \sum_{s=0}^{\infty} [A_{cl,\star}^s Q_1 A_{cl,\star}^s]_{:,i} = \sum_{s=0}^{\infty} a_i^s A_{cl,\star}^s [Q_1]_{:,i} = (I - a_i A_{cl,\star})^{-1} [Q_1]_{:,i} \\
 &= (I - a_i A_{cl,\star})^{-1} (A_{cl,\star} P_\star [\Delta_{A_{cl}}^1]_{:,i} + p_i a_i [\Delta_{A_{cl}}^1]_{i,:})
 \end{aligned}$$

Putting this together, we have that

$$\begin{aligned}
 &\text{vec}(\Delta_\theta^1)^\top \mathcal{H}(\theta_\star) \text{vec}(\Delta_\theta^2) \\
 &= \sum_{i=1}^{d_x} \frac{1}{(1 - a_i^2)} \left( p_i a_i [\Delta_B^1]_{i,:} + b(I + a_i (I - a_i A_{cl,\star})^{-1} A_{cl,\star}) P_\star [\Delta_{A_{cl}}^1]_{:,i} + b p_i a_i^2 (I - a_i A_{cl,\star})^{-1} [\Delta_{A_{cl}}^1]_{i,:} \right)^\top \\
 &\quad \cdot (\mu I + b^2 P_\star)^{-1} \left( p_i a_i [\Delta_B^2]_{i,:} + b(I + a_i (I - a_i A_{cl,\star})^{-1} A_{cl,\star}) P_\star [\Delta_{A_{cl}}^2]_{:,i} + b p_i a_i^2 (I - a_i A_{cl,\star})^{-1} [\Delta_{A_{cl}}^2]_{i,:} \right) \\
 &= \sum_{i=1}^{d_x} \sum_{j=1}^{d_x} \frac{1}{(\mu + b^2 p_j)(1 - a_i^2)} \left( p_i a_i [\Delta_B^1]_{i,j} + \frac{b p_j}{1 - a_i a_j} ([\Delta_A^1]_{j,i} - k_i [\Delta_B^1]_{j,i}) + \frac{b p_i a_i^2}{1 - a_i a_j} ([\Delta_A^1]_{i,j} - k_j [\Delta_B^1]_{i,j}) \right) \\
 &\quad \cdot \left( p_i a_i [\Delta_B^2]_{i,j} + \frac{b p_j}{1 - a_i a_j} ([\Delta_A^2]_{j,i} - k_i [\Delta_B^2]_{j,i}) + \frac{b p_i a_i^2}{1 - a_i a_j} ([\Delta_A^2]_{i,j} - k_j [\Delta_B^2]_{i,j}) \right) \quad (\text{L.1})
 \end{aligned}$$

We now evaluate the above when  $\Delta_B^1 = \Delta_B^2 = 0$  and  $\Delta_A^1 = e_\ell e_o^\top$ ,  $\Delta_A^2 = e_n e_m^\top$ . For this to be non-zero, we must have that either  $\ell = n, o = m$  or  $\ell = m, o = n$  and, as noted previously, we can ignore the case when  $\ell = m, o = n$ . Therefore, if  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nm}, A_{nm}} = \frac{1}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( \frac{b p_n a_n^2}{1 - a_n a_m} \right)^2$$

and if  $n = m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nn}, A_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( \frac{b p_n (1 + a_n^2)}{1 - a_n^2} \right)^2$$

Now consider  $\Delta_A^1 = \Delta_A^2 = 0$  and  $\Delta_B^1 = e_\ell e_o^\top$ ,  $\Delta_B^2 = e_n e_m^\top$ . As before, for  $[\mathcal{H}(\theta_\star)]_{B_{\ell o}, B_{nm}}$  to be non-zero, we need either  $\ell = n, o = m$  or  $\ell = m, o = n$ . Therefore, if  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{B_{nm}, B_{nm}} = \frac{1}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n k_m}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n a_n^2 k_m}{1 - a_n a_m} \right)^2$$

and if  $n = m$ ,

$$[\mathcal{H}(\theta_\star)]_{B_{nn}, B_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n k_n (1 + a_n^2)}{1 - a_n^2} \right)^2$$

Finally, we consider the case where  $\Delta_A^1 = e_\ell e_o^\top$ ,  $\Delta_A^2 = 0$  and  $\Delta_B^1 = 0$ ,  $\Delta_B^2 = e_n e_m^\top$ . Again, we must have that either  $\ell = n, o = m$  or  $\ell = m, o = n$  for  $[\mathcal{H}(\theta_\star)]_{A_{\ell o}, B_{nm}}$  to be non-zero. Therefore, if  $n \neq m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nm}, B_{nm}} = \frac{-k_m}{(\mu + b^2 p_n)(1 - a_m^2)} \left( \frac{b p_n}{1 - a_n a_m} \right)^2 + \frac{1}{(\mu + b^2 p_m)(1 - a_n^2)} \left( p_n a_n - \frac{b p_n a_n^2 k_m}{1 - a_n a_m} \right) \frac{b p_n a_n^2}{1 - a_n a_m}$$

and if  $n = m$ ,

$$[\mathcal{H}(\theta_\star)]_{A_{nn}, B_{nn}} = \frac{1}{(\mu + b^2 p_n)(1 - a_n^2)} \left( \frac{b p_n (1 + a_n^2)}{1 - a_n^2} \right) \left( p_n a_n - \frac{b p_n k_n (1 + a_n^2)}{1 - a_n^2} \right).$$

□

### L.3. Proof of Proposition 4.1

Here we choose  $\rho_1 = \rho, \rho_2 = 0, b = \sqrt{1-\rho}, \kappa_1 = \kappa_2 = \frac{1}{\sqrt{1-\rho}}$ , and  $\mu = \frac{1}{(1-\rho)^2}$ . With these constants, some algebra shows that

$$\begin{aligned} a_1 &= \mathcal{O}(\rho), \quad \frac{1}{1-a_1} = \Theta_\rho\left(\frac{1}{1-\rho}\right), \quad a_i = 0, i \geq 2 \\ k_1 &= \mathcal{O}(1-\rho), \quad k_i = 0, i \geq 2 \\ p_1 &= \Theta_\rho\left(\frac{1}{(1-\rho)^{3/2}}\right), \quad p_i = \Theta_\rho\left(\frac{1}{\sqrt{1-\rho}}\right), i \geq 2 \end{aligned}$$

Plugging these values into the expression given for  $\mathcal{H}(\theta_*)$  in Lemma L.1, we have

$$\begin{aligned} [\mathcal{H}(\theta_*)]_{A_{11}, A_{11}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^3}\right), \quad [\mathcal{H}(\theta_*)]_{A_{1m}, A_{1m}} = \Theta_\rho\left(\frac{1}{1-\rho}\right), m \geq 2 \\ [\mathcal{H}(\theta_*)]_{B_{11}, B_{11}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^2}\right), \quad [\mathcal{H}(\theta_*)]_{B_{1m}, B_{1m}} = \Theta_\rho\left(\frac{1}{(1-\rho)^2}\right), m \geq 2 \\ [\mathcal{H}(\theta_*)]_{A_{11}, B_{11}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^{5/2}}\right), \quad [\mathcal{H}(\theta_*)]_{A_{1m}, B_{1m}} = \Theta_\rho\left(\frac{1}{(1-\rho)^{3/2}}\right), m \geq 2 \end{aligned}$$

All other terms are 0 or do not scale with  $\frac{1}{1-\rho}$  and can therefore be ignored. It follows that the sample complexity will scale as

$$\text{tr}(\mathcal{H}(\theta_*)\Gamma_T(\pi; \theta_*)^{-1}) \approx_\rho \text{tr}(\mathcal{H}_1\Gamma_T(\pi; \theta_*)^{-1})$$

where here

$$\begin{aligned} \mathcal{H}_1 \approx_\rho & \frac{1}{(1-\rho)^3} e_1 e_1^\top + \frac{1}{1-\rho} \sum_{j=2}^{d_x} e_j e_j^\top + \frac{1}{(1-\rho)^2} \sum_{j=1}^{d_x} e_{d_x+j} e_{d_x+j}^\top + \frac{1}{(1-\rho)^{5/2}} (e_1 e_{d_x+1}^\top + e_{d_x+1} e_1^\top) \\ & + \frac{1}{(1-\rho)^{3/2}} \sum_{j=2}^{d_x} (e_j e_{d_x+j}^\top + e_{d_x+j} e_j^\top) \end{aligned}$$

**Sample Complexity of TOPLE.** Our results show that the sample complexity of TOPLE scale with steady state covariates, and we can therefore analyze  $\text{tr}(\mathcal{H}_1\Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1})$ . As TOPLE plays the optimal inputs and we are concerned with obtaining an upper bound on its performance, we will simply construct a feasible input, which will then upper bound the actual performance.

In particular, we will set  $U_\ell = 0$  for all but two  $\ell$  (and their conjugate partners), and for those  $\ell$  will set  $U_\ell = U' := \begin{bmatrix} u_1 & 0 \\ 0 & u_2 I_{d_x-1} \end{bmatrix}$ . In that case, we will have

$$\begin{aligned} \Gamma_T^{\text{ss}}(\pi; \theta_*) &\propto 2\text{real}\left(\begin{bmatrix} (e^{\ell\omega_1} I - A_*)^{-1} B_* \\ I \end{bmatrix} U' \begin{bmatrix} (e^{\ell\omega_1} I - A_*)^{-1} B_* \\ I \end{bmatrix}^H + \begin{bmatrix} (e^{\ell\omega_2} I - A_*)^{-1} B_* \\ I \end{bmatrix} U' \begin{bmatrix} (e^{\ell\omega_2} I - A_*)^{-1} B_* \\ I \end{bmatrix}^H\right) \\ &= 2 \begin{bmatrix} |e^{\ell\omega_1} - \rho|^{-2} (1-\rho) u_1 & 0 & \text{real}((e^{\ell\omega_1} - \rho)^{-1}) \sqrt{1-\rho} u_1 & 0 \\ 0 & 2(1-\rho) u_2 I & 0 & \text{real}((e^{\ell\omega_1})^{-1}) \sqrt{1-\rho} u_2 I \\ \text{real}((e^{\ell\omega_1} - \rho)^{-1}) \sqrt{1-\rho} u_1 & 0 & 2u_1 & 0 \\ 0 & \text{real}((e^{\ell\omega_1})^{-1}) \sqrt{1-\rho} u_2 I & 0 & 2u_2 I \end{bmatrix} \\ &+ 2 \begin{bmatrix} |e^{\ell\omega_2} - \rho|^{-2} (1-\rho) u_1 & 0 & \text{real}((e^{\ell\omega_2} - \rho)^{-1}) \sqrt{1-\rho} u_1 & 0 \\ 0 & 0 & 0 & \text{real}((e^{\ell\omega_2})^{-1}) \sqrt{1-\rho} u_2 I \\ \text{real}((e^{\ell\omega_2} - \rho)^{-1}) \sqrt{1-\rho} u_1 & 0 & 0 & 0 \\ 0 & \text{real}((e^{\ell\omega_2})^{-1}) \sqrt{1-\rho} u_2 I & 0 & 0 \end{bmatrix} \end{aligned}$$

where we simplify using the values of  $A_*, B_*, \omega_1$  and  $\omega_2$  are the input frequencies we choose, and the real comes from the conjugate symmetry. We write “ $\propto$ ” instead of “ $=$ ” as additional normalization by factors of  $T$  and  $k$  are necessary to yield

equality, and for simplicity we currently ignore. We will handle these factors later. For large enough  $T$ ,  $\omega_1$  and  $\omega_2$  can be chosen essentially as desired, so we set  $\omega_1 = 1 - \rho$  and  $\omega_2 = \pi + 1 - \rho$ . As we take  $\rho$  close to 1, we have

$$e^{i\omega_1} = \cos(\omega_1) + i \sin(\omega_1) = 1 + i(1 - \rho) + o(1 - \rho), \quad e^{i\omega_2} = -1 - i(1 - \rho) + o(1 - \rho)$$

It follows that

$$\begin{aligned} \text{real}((e^{i\omega_1})^{-1}) &= -\text{real}((e^{i\omega_2})^{-1}) \\ \text{real}((e^{i\omega_1} - \rho)^{-1}) &= \frac{1 - \rho}{2(1 - \rho)^2} + o((1 - \rho)^2), \quad |e^{i\omega_1} - \rho|^{-2} = \frac{1}{2(1 - \rho)^2} + o((1 - \rho)^2) \\ \text{real}((e^{i\omega_2} - \rho)^{-1}) &= \frac{-1 - \rho}{(1 + \rho)^2} + o(1 - \rho), \quad |e^{i\omega_2} - \rho|^{-2} = \mathcal{O}(1) \end{aligned}$$

Plugging these in, we get that the above is equal to:

$$\begin{bmatrix} \frac{u_1}{1-\rho} + \mathcal{O}((1-\rho)u_1) & 0 & \frac{u_1}{\sqrt{1-\rho}} - \mathcal{O}((1-\rho)^{3/2}u_1) & 0 \\ 0 & 4(1-\rho)u_2 I & 0 & 0 \\ \frac{u_1}{\sqrt{1-\rho}} - \mathcal{O}((1-\rho)^{3/2}u_1) & 0 & 4u_1 & 0 \\ 0 & 0 & 0 & 4u_2 I \end{bmatrix}$$

This has the form given in Lemma L.2 so, applying this result and approximating  $\frac{u_1}{1-\rho} + \mathcal{O}((1-\rho)u_1)$  as  $\frac{u_1}{1-\rho}$  and  $\frac{u_1}{\sqrt{1-\rho}} - \mathcal{O}((1-\rho)^{3/2}u_1)$  as  $\frac{u_1}{\sqrt{1-\rho}}$ , (we note that this approximation will not affect the leading terms in the inverse due to the form of the inverse given in Lemma L.2) we have that the inverse of this matrix will be

$$\begin{bmatrix} \frac{4(1-\rho)}{3u_1} & 0 & \frac{-\sqrt{1-\rho}}{3u_1} & 0 \\ 0 & \frac{1}{4(1-\rho)u_2} I & 0 & 0 \\ \frac{-\sqrt{1-\rho}}{3u_1} & 0 & \frac{1}{3u_1} & 0 \\ 0 & 0 & 0 & \frac{1}{4u_2} I \end{bmatrix} \quad (\text{L.2})$$

Plugging this into our expression for  $\text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1})$  gives

$$T \text{tr}(\mathcal{H}(\theta_\star)\Gamma_T(\pi; \theta_\star)^{-1}) \approx_\rho \frac{1}{(1-\rho)^2 u'_1} + \frac{d_x}{(1-\rho)^2 u'_2} + \frac{1}{(1-\rho)^2 u'_1} + \frac{d_x}{(1-\rho)^2 u'_2} - \frac{1}{(1-\rho)^2 u'_1}$$

where  $u'_i$  denotes  $u_i/k^2$ . As we are interested in obtaining an upper bound on the sample complexity of TOPLE, we upper bound this by

$$\mathcal{O}\left(\frac{1}{(1-\rho)^2 u'_1} + \frac{d_x}{(1-\rho)^2 u'_2}\right)$$

It remains to choose  $u'_1, u'_2$  that satisfy  $u'_1 + (d_x - 1)u'_2 \leq \gamma^2/2$ . Choosing the values that minimize the above gives the final complexity:

$$\mathcal{O}\left(\frac{d_x^2}{(1-\rho)^2 \gamma^2}\right).$$

**Sample Complexity of Operator Norm Identification.** As was shown in (Wagenmaker & Jamieson, 2020), the optimal operator norm identification algorithm will choose the inputs that maximize  $\lambda_{\min}(\Gamma_T^{\text{ss}}(\pi; \theta_\star))$ . We will first construct an input with diagonal  $\mathbf{U}_\ell$  and, as in the previous section, with  $u_2 = \dots = u_{d_x}$ , and will then show that this input is in fact optimal.

Intuitively, the optimal operator norm identification algorithm seeks to input energy at frequencies which best excite the system (maximize  $\lambda_{\min}(\Gamma_T^{\text{ss}}(\pi; \theta_\star))$ ), and that balance the gain in each direction. Note that, regardless of the input frequencies, the computation in the previous section shows that, other than the first element, the diagonal components of  $\Gamma_T^{\text{ss}}(\pi; \theta_\star)$  will scale as  $(1-\rho)u_2, u_1$ , and  $u_2$ , respectively. Ignoring for a minute the contribution of the off-diagonal

terms, the value of the first coordinate will be maximized if energy is input at the frequency  $\omega_1$  which maximizes  $|e^{j\omega_1}I - \rho|^{-2}(1 - \rho)u_1$ . It is easy to see that this is maximized at  $\omega_1 = 0$ , which gives the value to the first coordinate of  $\frac{u_1}{1-\rho}$ , and, ignoring off diagonal entries, a minimum eigenvalue of

$$\lambda_{\min}(\Gamma_T^{\text{SS}}(\pi; \theta_\star)) = 2T \min\left\{\frac{u_1}{1-\rho}, (1-\rho)u_2, u_1, u_2\right\} = 2T \min\{(1-\rho)u_2, u_1\}$$

To maximize this while respecting the power constraint, operator norm-identification will choose  $u_2 \approx_\rho \gamma^2/d_x$  and  $u_1 \approx_\rho (1-\rho)\gamma^2/d_x$ , which will yield the minimum eigenvalue of

$$\lambda_{\min}(\Gamma_T^{\text{SS}}(\pi; \theta_\star)) \approx_\rho T(1-\rho)\gamma^2/d_x$$

Now taking into account the off-diagonal terms and applying Lemma L.3, using the expression for  $\Gamma_T^{\text{SS}}(\pi; \theta_\star)$  given in the analysis of TOPLE, we see that the contribution of the off-diagonal entries causes the minimum eigenvalue to instead be 0. However, if we instead set the inputs to those chosen in the previous section, as we saw there we have

$$\Gamma_T^{\text{SS}}(\pi; \theta_\star) = 2T \begin{bmatrix} \frac{u_1}{1-\rho} + \mathcal{O}((1-\rho)u_1) & 0 & \frac{u_1}{\sqrt{1-\rho}} - \mathcal{O}((1-\rho)^{3/2}u_1) & 0 \\ 0 & 4(1-\rho)u_2I & 0 & 0 \\ \frac{u_1}{\sqrt{1-\rho}} - \mathcal{O}((1-\rho)^{3/2}u_1) & 0 & 4u_1 & 0 \\ 0 & 0 & 0 & 4u_2I \end{bmatrix} \quad (\text{L.3})$$

Applying Lemma L.3 to this, after some algebra we see that

$$\lambda_{\min}(\Gamma_T^{\text{SS}}(\pi; \theta_\star)) \approx_\rho T \min\left\{\frac{u_1}{1-\rho}, u_1, u_2, (1-\rho)u_2\right\} = T \min\{u_1, (1-\rho)u_2\}$$

Choosing  $u_2 \approx_\rho \gamma^2/d_x$  and  $u_1 \approx_\rho (1-\rho)\gamma^2/d_x$  to balance this as before, we have that this input yields

$$\lambda_{\min}(\Gamma_T^{\text{SS}}(\pi; \theta_\star)) \approx_\rho T(1-\rho)\gamma^2/d_x$$

Observe that this achieves the same minimum eigenvalue as that achieved ignoring off-diagonal terms (up to constants) and that, furthermore, the form of the minimum eigenvalue given in Lemma L.3 implies that the off-diagonal terms will only decrease the minimum eigenvalue. It follows that  $T(1-\rho)\gamma^2/d_x$  is an upper bound on the minimum achievable eigenvalue when the inputs are diagonal so, since this input achieves this value, this is the near-optimal diagonal input for operator norm identification. It follows that the optimal covariance with diagonal inputs will take the form given in (L.3).

We now show that the globally optimal inputs are diagonal. We have just shown that the optimal covariance, when playing a diagonal input, will take the form

$$\Gamma_T^{\text{SS}}(\pi^\star; \theta_\star) \approx_\rho 2T \begin{bmatrix} \frac{\gamma^2}{d_x} & 0 & \frac{\sqrt{1-\rho}\gamma^2}{d_x} & 0 \\ 0 & \frac{4(1-\rho)\gamma^2}{d_x}I & 0 & 0 \\ \frac{\sqrt{1-\rho}\gamma^2}{d_x} & 0 & \frac{4(1-\rho)\gamma^2}{d_x} & 0 \\ 0 & 0 & 0 & \frac{4\gamma^2}{d_x}I \end{bmatrix} \quad (\text{L.4})$$

Now consider some  $\Delta \in \mathbb{C}^{d_u \times d_u}$ , and consider perturbing our optimal diagonal input at some frequency  $\omega$  by  $\Delta$  to form the new input  $U_\omega + \Delta$ . For our new input to be in our feasible set, we must have that  $\text{tr}(\Delta) \leq 0$  and that  $\Delta$  is symmetric. We want to show that, for every such perturbation,  $\lambda_{\min}(\Gamma_T^{\text{SS}}(\pi'; \theta_\star)) \leq \lambda_{\min}(\Gamma_T^{\text{SS}}(\pi^\star; \theta_\star))$  (where  $\pi'$  denotes the perturbed input). By first-order optimality conditions, this will imply that the diagonal input is optimal.

The resulting perturbation to the input will perturb  $\Gamma_T^{\text{SS}}(\pi^\star; \theta_\star)$  as

$$\Gamma_T^{\text{SS}}(\pi'; \theta_\star) = \Gamma_T^{\text{SS}}(\pi^\star; \theta_\star) + 2T \text{real} \left( \begin{bmatrix} G\Delta G^H & G\Delta \\ \Delta G^H & \Delta \end{bmatrix} \right), \quad G := \begin{bmatrix} \sqrt{1-\rho}(e^{j\omega} - \rho)^{-1} & 0 \\ 0 & \sqrt{1-\rho}e^{-j\omega}I \end{bmatrix}$$

where here  $G$  is equal to  $(e^{j\omega}I - A_\star)^{-1}B_\star$ . Note that the eigenvectors corresponding to the minimum eigenvalues of  $\Gamma_T^{\text{SS}}(\pi^\star; \theta_\star)$  are  $v_2, \dots, v_{d_x} = e_2, \dots, e_{d_x}$  and, some algebra shows,

$$v_1 = \left[ \frac{1 - 4(1-\rho) - \sqrt{1 - 4(1-\rho) + 16(1-\rho)^2}}{2\sqrt{1-\rho}}, 0, \dots, 0, 1, 0, \dots, 0 \right] = \left[ -\sqrt{1-\rho} + o(1-\rho), 0, \dots, 0, 1, 0, \dots, 0 \right]$$

where the 1 as at the index  $d_x + 1$ . Some algebra shows that

$$v_i^\top \text{real} \left( \begin{bmatrix} G\Delta G^H & G\Delta \\ \Delta G^H & \Delta \end{bmatrix} \right) v_i = (1 - \rho)\Delta_{ii}, \quad i \geq 2$$

where, since  $v_i = e_i$  the inner products select only the diagonal elements, and

$$\begin{aligned} v_1^\top \text{real} \left( \begin{bmatrix} G\Delta G^H & G\Delta \\ \Delta G^H & \Delta \end{bmatrix} \right) v_1 &\approx_\rho \left( (1 - \rho)^2 |e^{j\omega} - \rho|^{-2} - 2(1 - \rho)\text{real}((e^{j\omega} - \rho)^{-1}) + 1 \right) \Delta_{11} \\ &= \left( \frac{(1 - \rho)^2}{(\cos \omega - \rho)^2 + \sin^2 \omega} + 1 - \frac{2(1 - \rho)(\cos \omega - \rho)}{(\cos \omega - \rho)^2 + \sin^2 \omega} \right) \Delta_{11} \\ &\geq \left( 1 - \frac{(1 - \rho)(\cos \omega - \rho)}{(\cos \omega - \rho)^2 + \sin^2 \omega} \right) \Delta_{11} \\ &= \left( 1 - \frac{(1 - \rho)(\cos \omega - \rho)}{1 + \rho^2 - 2\rho \cos \omega} \right) \Delta_{11} \end{aligned}$$

and note that  $\left( 1 - \frac{(1 - \rho)(\cos \omega - \rho)}{1 + \rho^2 - 2\rho \cos \omega} \right) \geq 0$  for all  $\omega$ . We must have that  $\sum_{i=1}^{d_x} \Delta_{ii} \leq 0$  to meet our constraint. This implies that either all  $\Delta_{ii} = 0$ , or there exists  $i'$  such that  $\Delta_{i'i'} < 0$ . By the above expressions, it follows that, if  $i' \geq 2$ , the latter case will cause the minimum eigenvalue to decrease, and if  $i' = 1$ , the minimum eigenvalue cannot increase. It follows that our perturbation  $\Delta$  cannot increase  $\lambda_{\min}(\Gamma_T^{\text{ss}}(\pi^*; \theta_\star))$ , which implies that the optimal input is in fact diagonal.

Returning to the optimal covariates obtained with diagonal inputs, using the inverse expression for  $\Gamma_T^{\text{ss}}(\pi^*; \theta_\star)$  given in the analysis of TOPLE (L.2), we then have that

$$T \text{tr}(\mathcal{H}(\theta_\star) \Gamma_T(\pi^*; \theta_\star)^{-1}) = [\mathcal{H}(\theta_\star)]_{A_{11}, A_{11}} \frac{4(1 - \rho)}{3u_1} + 2[\mathcal{H}(\theta_\star)]_{A_{11}, B_{11}} \frac{-\sqrt{1 - \rho}}{3u_1} + [\mathcal{H}(\theta_\star)]_{B_{11}, B_{11}} \frac{1}{3u_1} + \Theta_{\rho, d_x} \left( \frac{d_x}{(1 - \rho)^2 u_2} \right)$$

From the expression for  $\mathcal{H}(\theta_\star)$  given in (L.1), we see that we can express

$$[\mathcal{H}(\theta_\star)]_{A_{11}, A_{11}} = c_1^2, \quad [\mathcal{H}(\theta_\star)]_{B_{11}, B_{11}} = c_2^2, \quad [\mathcal{H}(\theta_\star)]_{A_{11}, B_{11}} = c_1 c_2$$

for some values  $c_1, c_2$ . It follows that,

$$[\mathcal{H}(\theta_\star)]_{A_{11}, A_{11}} \frac{4(1 - \rho)}{3u_1} + 2[\mathcal{H}(\theta_\star)]_{A_{11}, B_{11}} \frac{-\sqrt{1 - \rho}}{3u_1} + [\mathcal{H}(\theta_\star)]_{B_{11}, B_{11}} \frac{3}{u_1} = \frac{1}{3u_1} \left( \sqrt{1 - \rho} c_1 - c_2 \right)^2 + \frac{1}{u_1} (1 - \rho) c_1^2$$

Now plugging in values of  $c_1, c_2$ , we have

$$\frac{1}{3u_1} \left( \sqrt{1 - \rho} c_1 - c_2 \right)^2 + \frac{1}{u_1} (1 - \rho) c_1^2 = \Theta_\rho \left( \frac{1}{(1 - \rho)^2 u_1} \right)$$

Which gives

$$T \text{tr}(\mathcal{H}(\theta_\star) \Gamma_T(\pi^*; \theta_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{1}{(1 - \rho)^2 u_1} + \frac{d_x}{(1 - \rho)^2 u_2} \right) \quad (\text{L.5})$$

Plugging in our values for  $u_1, u_2$  gives the complexity:

$$\Theta_{\rho, d_x} \left( \frac{d_x}{(1 - \rho)^3 \gamma^2} + \frac{d_x^2}{(1 - \rho)^2 \gamma^2} \right).$$

Note that our analysis is somewhat sensitive to the constants present in the entries of  $\text{tr}(\mathcal{H}(\theta_\star) \Gamma_T(\pi; \theta_\star)^{-1})$  that correspond to  $u_1$ . It is difficult to determine the precise constants that will appear in actual operator norm identification allocation. However, we note that any increase to the value of the constant in the off-diagonal term,  $\frac{u_1}{\sqrt{1 - \rho}}$ , will cause the minimum eigenvalue to decrease, by Lemma L.3, and we can therefore expect the constants to be no larger than their stated values. If we use constants smaller than what is stated here, this will only cause the magnitude of the off-diagonal terms in the inverse,  $\frac{-\sqrt{1 - \rho}}{3u_1}$ , to decrease, which will further reduce the contribution of the term  $2[\mathcal{H}(\theta_\star)]_{A_{11}, B_{11}} \frac{-\sqrt{1 - \rho}}{3u_1}$ , causing the final complexity to be larger. However, this will not change the fact that

$$[\mathcal{H}(\theta_\star)]_{A_{11}, A_{11}} \frac{4(1 - \rho)}{3u_1} + 2[\mathcal{H}(\theta_\star)]_{A_{11}, B_{11}} \frac{-\sqrt{1 - \rho}}{3u_1} + [\mathcal{H}(\theta_\star)]_{B_{11}, B_{11}} \frac{3}{u_1} = \Theta_\rho \left( \frac{1}{(1 - \rho)^2 u_1} \right)$$

So it follows the true complexity is as stated.

**Sample Complexity of Frobenius Norm Identification.** Note that TOPLE is the optimal Frobenius norm identification algorithm. In this case,  $\mathcal{H}_{\text{fro}}(\theta_*) = I$ , so the optimal Frobenius norm identification algorithm minimizes  $\text{tr}(\Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1})$ . A similar argument to that used in determining the optimal operator norm identification inputs can be used to show that here the optimal covariance is again of the form (L.3). Then using our inverse expression from the previous sections (L.2), we have that

$$T \text{tr}(\Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1}) \approx_{\rho} \frac{1-\rho}{u_1} + \frac{d_x}{(1-\rho)u_2} + \frac{1}{u_1} + \frac{d_x}{u_2} \approx_{\rho} \frac{d_x}{(1-\rho)u_2} + \frac{1}{u_1}$$

$u_1, u_2$  will be chosen to minimize this while respecting the constraint  $u_1 + (d_x - 1)u_2 \leq \gamma^2$ . Some algebra shows that this is minimized for values

$$u_2 \approx_{\rho} \frac{d_x \gamma^2 - \gamma^2 \sqrt{1-\rho}}{d_x^2}, \quad u_1 \approx_{\rho} \frac{\gamma^2 \sqrt{1-\rho}}{d_x}$$

Plugging these into the complexity expression for operator norm identification, (L.5), gives the complexity

$$\Theta_{\rho, d_x} \left( \frac{d_x}{(1-\rho)^{5/2} \gamma^2} + \frac{d_x^2}{(1-\rho)^2 \gamma^2} \right).$$

**Sample Complexity of Optimal Noise Identification.** Finally, we turn to the policy which plays the inputs  $u_t \sim \mathcal{N}(0, \Lambda_*)$  for optimal  $\Lambda_*$  satisfying  $\text{tr}(\Lambda_*) \leq \gamma^2$ . In this case, our results show that the sample complexity will scale as  $\text{tr}(\mathcal{H}(\theta_*) \Gamma_T^{\text{noise}}(\theta_*, \Lambda_*)^{-1})/T$ . Our goal is then to determine the optimal choice of  $\Lambda_*$ . A simple application of the KKT conditions shows that the optimal  $\Lambda_*$  is diagonal (see the proof of Proposition 4.2 for this stated explicitly in a similar setting), and will take the form  $\Lambda_* = \text{diag}([u_1, u_2, \dots, u_2])$ . In this case, some algebra shows that

$$\Gamma_T^{\text{noise}}(\theta_*, \Lambda_*) \approx_{\rho} \begin{bmatrix} u_1 & 0 & 0 & 0 \\ 0 & (1-\rho)u_2 I & 0 & 0 \\ 0 & 0 & u_1 & 0 \\ 0 & 0 & 0 & u_2 I \end{bmatrix}$$

so

$$\text{tr}(\mathcal{H}(\theta_*) \Gamma_T^{\text{noise}}(\theta_*, \Lambda_*)^{-1}) = \Theta_{\rho} \left( \frac{1}{(1-\rho)^3 u_1} + \frac{d_x}{(1-\rho)^2 u_2} \right)$$

Choosing  $u_1, u_2$  that minimizes this gives the complexity

$$\Theta_{\rho, d_x} \left( \frac{1}{(1-\rho)^3 \gamma^2} + \frac{d_x^4}{(1-\rho) \gamma^2} \right).$$

□

**Lemma L.2.** Consider diagonal matrices  $D_1, D_2, D_3 \in \mathbb{R}^{d \times d}$ . Then,

$$D = \begin{bmatrix} D_1 & D_2 \\ D_2 & D_3 \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2 \\ C_2 & C_3 \end{bmatrix}$$

where  $C_1, C_2$ , and  $C_3$  are diagonal and

$$[C_1]_{ii} = \frac{[D_3]_{ii}}{[D_1]_{ii}[D_3]_{ii} - [D_2]_{ii}^2}, \quad [C_3]_{ii} = \frac{[D_1]_{ii}}{[D_1]_{ii}[D_3]_{ii} - [D_2]_{ii}^2}, \quad [C_2]_{ii} = \frac{-[D_2]_{ii}}{[D_1]_{ii}[D_3]_{ii} - [D_2]_{ii}^2}$$

provided these quantities are well-defined.

*Proof.* Note that we can permute the columns and rows of  $D$  with some permutation  $P$  such to create a block diagonal matrix  $D'$ :

$$PDP^{\top} = D' = \begin{bmatrix} D'_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D'_d \end{bmatrix}, \quad D'_i = \begin{bmatrix} [D_1]_{ii} & [D_2]_{ii} \\ [D_2]_{ii} & [D_3]_{ii} \end{bmatrix}, i = 1, \dots, d$$

From the block diagonal structure and the inverse of  $2 \times 2$  matrices, we have

$$(D')^{-1} = \begin{bmatrix} (D'_1)^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (D'_d)^{-1} \end{bmatrix}, \quad (D'_i)^{-1} = \frac{1}{[D_1]_{ii}[D_3]_{ii} - [D_2]_{ii}^2} \begin{bmatrix} [D_3]_{ii} & -[D_2]_{ii} \\ -[D_2]_{ii} & [D_1]_{ii} \end{bmatrix}$$

The orthogonality of a permutation gives that  $D = P^\top D' P$ , so  $D^{-1} = P^\top (D')^{-1} P$ . Permuting the form of  $(D')^{-1}$  gives the expression in the statement of the result.  $\square$

**Lemma L.3.** Consider diagonal matrices  $D_1, D_2, D_3 \in \mathbb{R}^{d \times d}$ . Then,

$$\lambda_{\min} \left( \begin{bmatrix} D_1 & D_2 \\ D_2 & D_3 \end{bmatrix} \right) = \min_{i \in \{1, \dots, d\}} \frac{1}{2} \left( [D_1]_{ii} + [D_3]_{ii} - \sqrt{([D_1]_{ii} + [D_3]_{ii})^2 - 4([D_1]_{ii}[D_3]_{ii} - [D_2]_{ii}^2)} \right).$$

*Proof.* Let  $P$  be the permutation described in the proof of Lemma L.2. Note that the eigenvalues of  $PDP^\top$  are the same as those of  $D$  since, if we write the eigendecomposition of  $D$  as  $V\Lambda V^\top$ , we see that  $PDP^\top = (PV)\Lambda(PV)^\top$ , and that  $PV$  is orthogonal, so this is the eigendecomposition of  $PDP^\top$ . Given this, we have that  $\lambda_{\min}(D) = \min_{i \in \{1, \dots, d\}} \lambda_{\min}(D'_i)$ . The eigenvalues,  $\lambda$ , of  $D'_i$  satisfy

$$([D_1]_{ii} - \lambda)([D_3]_{ii} - \lambda) - [D_2]_{ii}^2 = 0$$

Solving this for  $\lambda$  and taking the minimum solution gives the result.  $\square$

#### L.4. Proof of Proposition 4.2

We now choose  $\rho_1 = \rho_2 = 1$ ,  $b = 1$ ,  $\kappa_1 = \frac{1}{(1-\rho)^4}$ ,  $\kappa_2 = 1$ ,  $\mu = \frac{1}{(1-\rho)^2}$ . With this choice, some algebra shows that

$$\begin{aligned} a_1 &= \mathcal{O}((1-\rho)^2), & a_i &= \mathcal{O}(1/(2-\rho)), \quad i \geq 2 \\ \frac{1}{1-a_1} &= \mathcal{O}(1), & \frac{1}{1-a_i} &= \Theta_\rho\left(\frac{1}{1-\rho}\right), \quad i \geq 2 \\ k_1 &= \mathcal{O}(1), & k_i &= \mathcal{O}(1-\rho), \quad i \geq 2 \\ p_1 &= \Theta_\rho\left(\frac{1}{(1-\rho)^4}\right), & p_i &= \Theta_\rho\left(\frac{1}{1-\rho}\right), \quad i \geq 2 \end{aligned} \tag{L.6}$$

Plugging these values into the expression given for  $\mathcal{H}(\theta_*)$  in Lemma L.1 gives:

$$\begin{aligned} [\mathcal{H}(\theta_*)]_{A_{11}, A_{11}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^4}\right), & [\mathcal{H}(\theta_*)]_{A_{1m}, A_{1m}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^5}\right), \quad m > 1 \\ [\mathcal{H}(\theta_*)]_{A_{n1}, A_{n1}} &= \Theta_\rho(1), \quad n > 1, & [\mathcal{H}(\theta_*)]_{A_{nm}, A_{nm}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^3}\right), \quad n > 1, m > 1 \\ [\mathcal{H}(\theta_*)]_{B_{11}, B_{11}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^4}\right), & [\mathcal{H}(\theta_*)]_{B_{n1}, B_{n1}} &= \Theta_\rho(1), \quad n > 1 \\ [\mathcal{H}(\theta_*)]_{B_{1m}, B_{1m}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^3}\right), \quad m > 1, & [\mathcal{H}(\theta_*)]_{B_{nm}, B_{nm}} &= \Theta_\rho\left(\frac{1}{1-\rho}\right), \quad n > 1, m > 1 \\ [\mathcal{H}(\theta_*)]_{A_{11}, B_{11}} &= \Theta_\rho\left(\frac{-1}{(1-\rho)^4}\right), & [\mathcal{H}(\theta_*)]_{A_{1m}, B_{1m}} &= \Theta_\rho\left(\frac{-1}{(1-\rho)^4}\right), \quad m > 1 \\ [\mathcal{H}(\theta_*)]_{A_{n1}, B_{n1}} &= \Theta_\rho(1), \quad n > 1, & [\mathcal{H}(\theta_*)]_{A_{nm}, B_{nm}} &= \Theta_\rho\left(\frac{1}{(1-\rho)^2}\right), \quad n > 1, m > 1 \end{aligned}$$

**Sample Complexity of TOPLE.** As we will consider  $\rho$  close to 1, elements scaling as  $\Theta_\rho\left(\frac{1}{(1-\rho)^n}\right)$  for  $n \leq 2$  will be dominated by elements scaling as  $\Theta_\rho\left(\frac{1}{(1-\rho)^n}\right)$  for  $n > 2$ . For simplicity, we henceforth ignore these elements. Given these approximations, we see that  $\mathcal{H}_i$  for  $i > 1$  is approximately diagonal and therefore:

$$\text{tr}(\mathcal{H}_i \Gamma_T(\pi; \theta_*)^{-1}) \approx_\rho \sum_{j=1}^{d_x+d_u} [\mathcal{H}_i]_{jj} [\Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1}]_{jj}$$



$\mathcal{H}_i$ , however, contains non-negligible off-diagonal elements,  $[\mathcal{H}(\theta_*)]_{A_{1m}, B_{1m}}$ , and will take the form:

$$\begin{aligned} \mathcal{H}_1 &\approx_\rho \frac{1}{(1-\rho)^4} e_1 e_1^\top + \frac{1}{(1-\rho)^5} \sum_{j=2}^{d_x} e_j e_j^\top + \frac{1}{(1-\rho)^4} e_{d_x+1} e_{d_x+1}^\top + \frac{1}{(1-\rho)^3} \sum_{j=2}^{d_x} e_{d_x+j} e_{d_x+j}^\top \\ &\quad - \frac{1}{(1-\rho)^4} \sum_{j=1}^{d_x} (e_j e_{d_x+j}^\top + e_{d_x+j} e_j^\top) \\ &= \frac{1}{(1-\rho)^4} (e_1 - e_{d_x+1})(e_1 - e_{d_x+1})^\top + \sum_{j=2}^{d_x} \frac{1}{(1-\rho)^5} (e_j - (1-\rho)e_{d_x+j})(e_j - (1-\rho)e_{d_x+j})^\top \end{aligned}$$

As we are concerned with showing an upper bound on the performance of TOPLE, we can simply choose a feasible set of inputs and compute the sample complexity obtained by them. Since TOPLE obtains the optimal sample complexity, it follows that this will be a valid upper bound on performance. Given this, let  $U_\ell = 0$  for all but a single  $\ell$  to be chosen (and it's conjugate partner), and assume that the nonzero  $U_\ell = \text{diag}([u_1, \dots, u_{d_x}])$  is real and diagonal. We will choose  $u_2 = \dots = u_{d_x}$ . In that case, we will have

$$\begin{aligned} \Gamma_T^{\text{ss}}(\pi; \theta_*) &\propto 2 \text{real} \left( \begin{bmatrix} (e^{\iota\omega_\ell} I - A_\star)^{-1} B_\star & \\ & I \end{bmatrix} U_\ell \begin{bmatrix} (e^{\iota\omega_\ell} I - A_\star)^{-1} B_\star \\ I \end{bmatrix}^H \right) \\ &= 2 \begin{bmatrix} |e^{\iota\omega_\ell} - \rho|^{-2} U_\ell & \text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) U_\ell \\ \text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) U_\ell & U_\ell \end{bmatrix} \end{aligned}$$

where we simplify using the values of  $A_\star, B_\star, \omega_\ell = \iota 2\pi\ell/k$ , and the real comes from the conjugate symmetry. Since  $U_\ell$  is diagonal, we can apply Lemma L.2 to invert this:

$$\begin{bmatrix} |e^{\iota\omega_\ell} - \rho|^{-2} U_\ell & \text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) U_\ell \\ \text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) U_\ell & U_\ell \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2 \\ C_2 & C_3 \end{bmatrix}$$

where

$$\begin{aligned} [C_1]_{ii} &= \frac{u_i^{-1}}{|e^{\iota\omega_\ell} - \rho|^{-2} - \text{real}((e^{\iota\omega_\ell} - \rho)^{-1})^2}, & [C_2]_{ii} &= \frac{-\text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) u_i^{-1}}{|e^{\iota\omega_\ell} - \rho|^{-2} - \text{real}((e^{\iota\omega_\ell} - \rho)^{-1})^2} \\ [C_3]_{ii} &= \frac{|e^{\iota\omega_\ell} - \rho|^{-2} u_i^{-1}}{|e^{\iota\omega_\ell} - \rho|^{-2} - \text{real}((e^{\iota\omega_\ell} - \rho)^{-1})^2} \end{aligned}$$

We choose  $\omega_\ell = 1 - \rho$ . Then,

$$e^{\iota\omega_\ell} - \rho = \cos(\omega_\ell) - \rho + \iota \sin(\omega_\ell) = 1 - \rho + \iota(1 - \rho) + o(1 - \rho)$$

and

$$\text{real}((e^{\iota\omega_\ell} - \rho)^{-1}) = \frac{1}{2(1-\rho)} + o((1-\rho)^2), \quad |e^{\iota\omega_\ell} - \rho|^{-2} = \frac{1}{2(1-\rho)^2} + o((1-\rho)^2)$$

So it follows that

$$[C_1]_{ii} \approx_\rho \frac{4(1-\rho)^2}{u_i}, \quad [C_2]_{ii} \approx_\rho \frac{-2(1-\rho)}{u_i}, \quad [C_3]_{ii} \approx_\rho \frac{2}{u_i}$$

Using our expressions for  $\mathcal{H}_i$  given above and this expression, we have that

$$T \text{tr}(\mathcal{H}_1 \Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1}) = \Theta_{\rho, d_x} \left( \frac{1}{(1-\rho)^2 u'_1} + \frac{d_x}{(1-\rho)^3 u'_2} + \frac{1}{(1-\rho)^4 u'_1} + \frac{d_x}{(1-\rho)^3 u'_2} + \frac{1}{(1-\rho)^3 u'_1} + \frac{d_x}{(1-\rho)^3 u'_2} \right)$$

$$T \text{tr}(\mathcal{H}_i \Gamma_T^{\text{ss}}(\pi; \theta_*)^{-1}) = \Theta_{\rho, d_x} \left( \frac{d_x}{(1-\rho) u'_2} + \frac{d_x}{(1-\rho) u'_2} \right), \quad i > 1$$

where  $u'_i$  denotes  $u_i/k^2$ . Thus,

$$T \text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{SS}}(\pi; \theta_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{1}{(1-\rho)^4 u'_1} + \frac{d_x}{(1-\rho)^3 u'_2} \right)$$

We can choose  $u_1$  and  $u_2$  as we wish as long as they meet the power constraint  $u'_1 + (d_x - 1)u'_2 \leq \gamma^2$ . Choosing the values that minimize the complexity yields:

$$T \text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{SS}}(\pi; \theta_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{1}{(1-\rho)^4 \gamma^2} + \frac{d_x^2}{(1-\rho)^3 \gamma^2} \right)$$

By our construction, this is an upper bound on the performance of TOPLE.

**Sample Complexity of Operator Norm and Frobenius Norm Identification.** We can follow a similar argument as that used in the proof of Proposition 4.1 to show that the optimal inputs will be diagonal for both operator norm and Frobenius norm identification. Furthermore, as both  $A_\star$  and  $B_\star$  are scalings of the identity, the optimal operator norm and Frobenius norm identification algorithms will allocate the same energy to each coordinate. Note that the input constructed in the previous section will yield the maximum gain, so it follows that both operator and Frobenius norm identification will play inputs at similar frequencies, and the analysis in the preceding section can be applied here. However, given that the inputs are isotropic, we will have  $u_1 = u_2 = \gamma^2/d_x$ , which will yield a sample complexity of

$$T \text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{SS}}(\pi; \theta_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{d_x}{(1-\rho)^4 \gamma^2} + \frac{d_x^2}{(1-\rho)^3 \gamma^2} \right).$$

**Sample Complexity of Optimal Noise Identification.** When playing noise, the complexity will scale as  $\text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{noise}}(\theta_\star, \Lambda_u)^{-1})/T$ . To analyze the sample complexity of this approach, we must first determine the  $\Lambda_u \succeq 0$  that minimizes this and satisfies  $\text{tr}(\Lambda_u) \leq \gamma^2$ . In our setting, we will have

$$\mathbf{\Gamma}_T^{\text{noise}}(\theta_\star, \Lambda_u) = \begin{bmatrix} \sum_{s=0}^T \rho^{2s} \Lambda_u & 0 \\ 0 & \Lambda_u \end{bmatrix} \approx_\rho \begin{bmatrix} \frac{1}{1-\rho} \Lambda_u & 0 \\ 0 & \Lambda_u \end{bmatrix}$$

So it follows that

$$\text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{noise}}(\theta_\star, \Lambda_u)^{-1}) \approx_\rho \sum_{i=1}^{d_x} [(1-\rho) \text{tr}(\mathcal{H}_{i,1} \Lambda_u^{-1}) + \text{tr}(\mathcal{H}_{i,2} \Lambda_u^{-1})]$$

where  $\mathcal{H}_{i,1}, \mathcal{H}_{i,2}$  denote the first and second  $d_x \times d_x$  block diagonals of  $\mathcal{H}_i$ , respectively. Computing the gradient of this expression and the constraint  $\text{tr}(\Lambda_u) \leq \gamma^2$  and applying the KKT conditions gives that the optimal  $\Lambda_\star$  must satisfy:

$$-\sum_{i=1}^{d_x} [(1-\rho)(\Lambda_\star)^{-1} \mathcal{H}_{i,1} (\Lambda_\star)^{-1} + (\Lambda_\star)^{-1} \mathcal{H}_{i,2} (\Lambda_\star)^{-1}] + \mu I = 0$$

As  $\mathcal{H}_{i,1}, \mathcal{H}_{i,2}$  are diagonal, up to lower order terms, it follows that a diagonal  $\Lambda_\star$  will satisfy this expression. Furthermore, given the symmetry of  $A_\star, B_\star$ , it is clear that  $\Lambda_\star$  will then take the form  $\text{diag}([u_1, u_2, \dots, u_2])$  for  $u_1 + (d_x - 1)u_2 \leq \gamma^2$ . Plugging this into the expressions given above, we have that the complexity scales as

$$\text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{noise}}(\theta_\star, \Lambda_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{d_x}{(1-\rho)^4 u_2} + \frac{1}{(1-\rho)^4 u_1} \right)$$

Choosing  $u_1, u_2$  that minimize this yields the complexity

$$\text{tr}(\mathcal{H}(\theta_\star) \mathbf{\Gamma}_T^{\text{noise}}(\theta_\star, \Lambda_\star)^{-1}) = \Theta_{\rho, d_x} \left( \frac{d_x^2}{(1-\rho)^4 \gamma^2} \right).$$

□

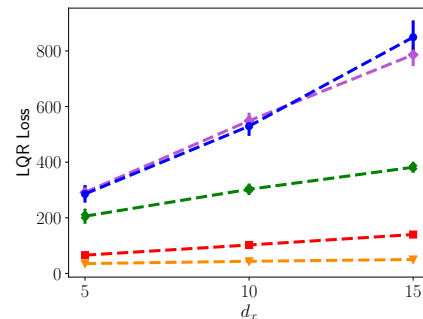
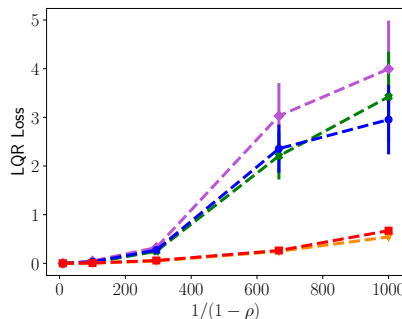
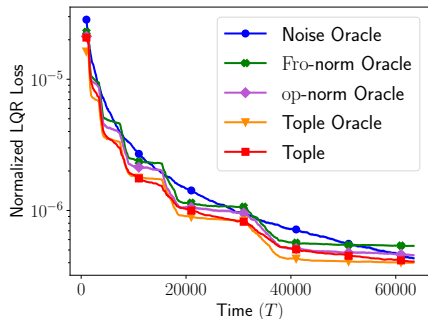


Figure 4. LQR loss versus time on  $A_*$ , a Jordan block and  $B_*$ ,  $R_x$ ,  $R_u$  randomly generated. Figure 5. LQR loss when varying  $\rho$  on example stated in Proposition 4.1. Figure 6. LQR loss when varying  $d_x$  on example stated in Proposition 4.2.

## M. Details on Numerical Results

In Figures 4, 5, and 6, we plot Figures 1, 2, and 3 with error bars. In all cases the error bars indicate a standard error. We make several additional remarks on the experiments. For Figure 4, we chose  $d_x = d_u = 5$  and chose  $\rho = 0.8$ , which gave us

$$A_* = \begin{bmatrix} 0.8 & 1 & 0 & 0 & 0 \\ 0 & 0.8 & 1 & 0 & 0 \\ 0 & 0 & 0.8 & 1 & 0 \\ 0 & 0 & 0 & 0.8 & 1 \\ 0 & 0 & 0 & 0 & 0.8 \end{bmatrix}$$

As was stated in the main text, we generated  $B_*$ ,  $R_x$ , and  $R_u$  randomly. For each realization, we ran 15 trials, so Figure 4 is, in total, the average over 225 trials. As different  $R_x$  and  $R_u$  would cause  $\|\mathcal{H}(\theta_*)\|_{\text{op}}$  to vary widely, we divided the loss of each realization by  $\|\mathcal{H}(\theta_*)\|_{\text{op}}$  to ensure they were on the same scale. The reader may wonder why the error decays in a stepwise fashion. This is due to the convex relaxation of the inputs. In this example, the majority of the energy is concentrated in the first eigenvalue of the input, and thus, when the matrix input is decomposed, the majority of the energy is played in only a fraction of  $1/d_u$  of the time. We therefore see a much steeper decrease in this time. As we show, however, our convex relaxation is tight and nothing is lost by playing inputs in this way.

For Figure 5, we chose the values of  $R_x$  and  $R_u$  as given in the proof of Proposition 4.1 and set  $d_x = d_u = 5$ . For Figure 6, we chose  $R_x$  and  $R_u$  as given in the proof of Proposition 4.2 and set  $\rho = 0.99$ .

Our implementation of TOPLE uses the convex relaxation and projected gradient descent solution given in Appendix B.5. While (Wagenmaker & Jamieson, 2020) does not provide a computationally efficient solution to their proposed operator norm identification algorithm, we note that the convex relaxation given in Appendix B.5 can be applied to this problem as well, yielding a computationally efficient version of the algorithm given in (Wagenmaker & Jamieson, 2020). We rely on this computationally efficient relaxation for our implementation of the operator norm identification algorithm.