Improper Learning for Non-Stochastic Control

Max Simchowitz MSIMCHOW@BERKELEY.EDU
UC Berkeley

Karan Singh KARANS@PRINCETON.EDU
Princeton University and Google AI Princeton

Elad Hazan EHAZAN@PRINCETON.EDU
Princeton University and Google AI Princeton

Editors: Jacob Abernethy and Shivani Agarwal

Abstract

We consider the problem of controlling a possibly unknown linear dynamical system with adversarial perturbations, adversarially chosen convex loss functions, and partially observed states, known as non-stochastic control. We introduce a controller parametrization based on the denoised observations, and prove that applying online gradient descent to this parametrization yields a new controller which attains sublinear regret vs. a large class of closed-loop policies. In the fully-adversarial setting, our controller attains an optimal regret bound of $\sqrt{T}$-when the system is known, and, when combined with an initial stage of least-squares estimation, $T^{2/3}$ when the system is unknown; both yield the first sublinear regret for the partially observed setting.

Our bounds are the first in the non-stochastic control setting that compete with *all* stabilizing linear dynamical controllers, not just state feedback. Moreover, in the presence of semi-adversarial noise containing both stochastic and adversarial components, our controller attains the optimal regret bounds of $\mathrm{poly}(\log T)$ when the system is known, and $\sqrt{T}$ when unknown. To our knowledge, this gives the first end-to-end $\sqrt{T}$ regret for online Linear Quadratic Gaussian controller, and applies in a more general setting with adversarial losses and semi-adversarial noise.

1. Introduction

In recent years, the machine learning community has produced a great body of work applying modern statistical and algorithmic techniques to classical control problems. Subsequently, recent work has turned to a more general paradigm termed the *non-stochastic control problem*: a model for dynamics that replaces stochastic noise with adversarial perturbations in the dynamics.

In this non-stochastic model, it is impossible to pre-compute an instance-wise optimal controller. Instead, the metric of performance is regret, or total cost compared to the best in hindsight given the realization of the noise. Previous work has introduced new adaptive controllers that are learned using iterative optimization methods, as a function of the noise, and are able to compete with the best controller in hindsight.

This paper presents a novel approach to non-stochastic control which unifies, generalizes, and improves upon existing results in the literature. Notably, we provide the first sublinear regret guarantees for non-stochastic control with partial observation for both *known* and *unknown* systems. Our non-stochastic framework also leads to new results for classical stochastic settings: e.g., the first tight regret bound for linear quadratic gaussian control (LQG) with an unknown system.

The non-stochastic linear control problem is defined using the following dynamical equations:

$$\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t, \qquad \mathbf{y}_t = C_\star \mathbf{x}_t + \mathbf{e}_t, \tag{1.1}$$

where $\mathbf{x}_t$ is the system state, $\mathbf{u}_t$ the control, $\mathbf{w}_t, \mathbf{e}_t$ adversarially-chosen noise terms, and $\mathbf{y}_t$ the observation. A learner iteratively chooses a control $\mathbf{u}_t$ upon observing $\mathbf{y}_t$, and suffers a loss $c_t(\mathbf{x}_t, \mathbf{u}_t)$ according to an adversarially-chosen loss function. Regret is defined as the difference between the sum of costs and that of the best controller in hindsight among a possible controller class.

Our technique a classical formulation based on the Youla parametrization Youla et al. (1976) for optimal control, which rewrites the state in terms of what we term "Nature's y's ', observations $\{\mathbf{y}_t^{\mathrm{nat}}\}$ *that would have resulted had we entered zero control at all times*. This yields a convex parametrization approximating possible stabilizing controllers we call *Disturbance Response Control*, or DRC. By applying online gradient descent to losses induced by this convex controller parametrization, we obtain a new controller we call the *Gradient Response Controller via Gradient Descent*, or DRC-GD. We show that DRC-GD attains wide array of results for stochastic and nonstochastic control, described in Section 1.2. Among the highlights:

1. $\widetilde{\mathcal{O}}(\sqrt{T})$ for controlling a known system with partial observation in the non-stochastic control model (Theorem 2), and $\widetilde{\mathcal{O}}(T^{2/3})$ regret (Theorem 3) when this system is *unknown*. This is the first sublinear regret bound for either setting, and the former rate is tight (Theorem 6).
2. The first $\widetilde{\mathcal{O}}(\sqrt{T})$ regret bound for the LQG unknown system (Theorem 5). This bound is tight, even when the state is observed (Simchowitz and Foster, 2020), and extends to mixed stochastic and adversarial perturbations (*semi-adversarial*). We also give $\mathrm{poly}\log T$ regret for semi-adversarial control with partial observation when the system is known (Theorem 4).
3. Regret bounds hold against the class of linear dynamical controllers (Definition 2.1), a much richer class than static feedback controllers previously considered for the non-stochastic control problem. This class is necessary to encompass $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal controllers under partial observation, and is ubiquitous in practical control applications.

Organization: The remainder of this section formally defines the setting, describes our results, and surveys the related literature. Section 2 expounds the relevant assumptions and describes our regret bound, and Section 3 describes our controller parametrization. Section 4 presents our algorithm and main results. All proofs are deferred to the appendix, whose organization and notation is detailed in Appendix A. Appendix B states lower bounds and provides extended comparison to past work, Part I gives high-level proofs of main results, and Part II supplies additional proof details, and allows for possibly unstable systems placed in feedback with stabilizing controllers.

1.1. Problem Setting

Dynamical Model: We consider *partially observed* linear dynamical system (PO-LDS), a continuous state-action, partially observable Markov decision process (POMDP) described by Eq. (1.1), with linear state-transition dynamics, where the observations are linear functions of the state. Here, $\mathbf{x}_t, \mathbf{w}_t \in \mathbb{R}^{d_x}$, $\mathbf{y}_t, \mathbf{e}_t \in \mathbb{R}^{d_y}$, $\mathbf{u}_t \in \mathbb{R}^{d_u}$ and $A_\star, B_\star, C_\star$ are of appropriate dimensions. We denote by $\mathbf{x}_t$ the state, $\mathbf{u}_t$ the control input, $\mathbf{y}_t$ is the output sequence, and $\mathbf{w}_t, \mathbf{e}_t$ are perturbations that the system is subject to. A *fully observed* linear dynamical system (FO-LDS) corresponds to the setting where $C_\star = I$ and $\mathbf{e}_t \equiv 0$, yielding a (fully observed) MDP where $\mathbf{x}_t \equiv \mathbf{y}_t$. We consider both the setting where $A_\star$ is stable (Assumption 1), and in Appendix G, unstable systems where the controller is put in feedback with stabilizing controllers

Interaction Model: A control policy (or learning algorithm) alg iteratively chooses an adaptive control input $\mathbf{u}_t = \mathsf{alg}_t(\mathbf{y}_{1:t}, \mathbf{u}_{1:t-1}, \ell_{1:t-1})$ upon the observation of the output sequence $(\mathbf{y}_1, \dots \mathbf{y}_t)$, and the sequence of loss functions $(\ell_1, \dots \ell_{t-1})$, past inputs, and possibly internal random coins. Let $(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})$ be the observation-action sequence from this resultant interaction. The cost of executing this controller is $J_T(\mathsf{alg}) = \sum_{t=1}^T \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})$. Notice that the learning algorithm alg does not observe the state sequence $\mathbf{x}_t$. Furthermore, it is unaware of the perturbation sequence $(\mathbf{w}_t, \mathbf{e}_t)$, except as may be inferred from observing the outputs $\mathbf{y}_t$. Lastly, the loss function ℓ_t is only made known to alg once the control input $\mathbf{u}_t^{\mathsf{alg}}$ is chosen. Morever generally, our results extend to achieving low regret on loss functions that depend on a finite history of inputs and outputs, namely $\ell_t(\mathbf{y}_{t:t-h}, \mathbf{u}_{t:t-h})$.

Policy Regret: Given a *benchmark class* of comparator control policies $\pi \in \Pi$, our aim is to minimize the cumulative regret with respect to the best policy in hindsight:

$$\mathrm{Regret}_T(\Pi) := J_T(\mathsf{alg}) - \min_{\pi\in\Pi} J_T(\pi) = \sum_{t=1}^T \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \min_{\pi\in\Pi} \sum_{t=1}^T \ell_t(\mathbf{y}_t^{\pi}, \mathbf{u}_t^{\pi}) \tag{1.2}$$

Note that the choice of the controller in Π may be made with the complete foreknowledge of the perturbations and the loss functions that the controller π (and the algorithm alg) is subject to. In this work, we compete with benchmark class of stabilizing linear dynamic controllers (LDC's) with internal state (see Definition 2.1 and Section 2). This generalizes the state-feedback class $\mathbf{u}_t = K\mathbf{x}_t$ considered in prior work.

Loss and Noise Regimes: We consider both the *known system* setting where alg has foreknowledge of the system Eq. (1.1), and the *unknown system* setting where alg does not (in either case, the comparator is selected with knownledge of the system). We also consider two loss and noise regimes: the *Lipschitz loss & non-stochastic noise* regime where the losses are Lipschitz over bounded sets (Assumption 2) and noises bounded and adversarial (Assumption 3), and the *strongly convex loss & semi-adversarial* regime where the losses are smooth and strongly convex (Assumption 5, and noise has a well-conditioned stochastic component, as well as an oblivious, possibly adversarial one (Assumption 6). We term this new noise model *semi-adversarial*; it is analogous to smoothed-adversarial and semi-random models considered in other domains (Spielman and Teng, 2004; Moitra et al., 2016; Bhaskara et al., 2014). In the first noise regime, the losses are selected by an adaptive adversary; in the second, an oblivious one.

Relation to LQR, LQG, $\mathcal{H}_2$ and $\mathcal{H}_\infty$: The online *LQG* problem corresponds to the problem where the system is driven by well-conditioned, independent Gaussian noise, and the losses $\ell_t(y, u) = y^\top Q y + u^\top R u$ are fixed quadratic functions. *LQR* is the fully observed analogue of LQG. The solution to the LQR (resp. LQG) problems are known as the $\mathcal{H}_2$-optimal controllers, which are well-approximated by a fixed state feedback controller (resp. LDC). In context of worst-case control, the $\mathcal{H}_\infty$ program can be used to compute a minimax controller that is optimal for the worst-case noise, and is also well-approximated by a LDC. In contrast to worst-case optimal control methods, low regret algorithms offer significantly stronger guarantees of instance-wise optimality on each noise sequence. We stress that the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ optimal control for partially observed systems are LDCs controllers; state feedback suffices only for full observations.

1.2. Contributions

We present Disturbance Response Controller via Gradient Descent, or DRC-GD, a unified algorithm which achieves sublinear regret for online control of a partially observed LDS with *both* adversarial losses and noises, even when the true system is *unknown* to the learner. In comparison to past work, this consitutes the first regret guarantee for partially observed systems (known or unknown to the learner) with *either* adversarial losses or adversarial noises. Furthermore, our bounds are the first in the online control literature which demonstrate low regret with respect to the broader class of linear dynamic controllers or *LDCs* described above (see also Definition 2.1); we stress that LDCs are necessary to capture the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control laws under partial observation, and yield strict improvements under full observation for certain non-stochastic noise sequences. In addition, all regret guarantees are non-asymptotic, and have polynomial dependence on other relevant problem parameters. Our guarantees hold in four different regimes of interest, summarized in Table 1, and described below. Tables 2 and 3 in the appendix gives a detailed comparison to past work.

For *known* systems, Algorithm 1 attains $\widetilde{\mathcal{O}}(\sqrt{T})$ regret for Lipschitz losses and adversarial noise (Theorem 2), which we show in Theorem 6 is optimal up to logarithmic factors, even when the state is observed and the noises/losses satisfy quite restrictive conditions. For strongly convex losses and semi-adversarial noise, we achieve $\mathrm{poly}\log T$ regret (Theorem 4). This result strengthens the prior art even for full observation due to Agarwal et al. (2019b) by removing extraneous assumptions on the gradient oracle, handling semi-adversarial noise, and ensuring bounded regret (rather than pseudo-regret). We do so via a regret bounds for "conditionally-strongly convex loses" (Appendix E), which may be of broader interest to the online learning community.

For *unknown* systems, Algorithm 3 attains $\widetilde{\mathcal{O}}(T^{2/3})$ regret for Lipschitz losses and adversarial noise, and $\widetilde{\mathcal{O}}(\sqrt{T})$-regret for strongly convex losses and semi-adversarial noise (Theorems 3 and 5). The former result has been established under full observation but required "strong controllability" (Hazan et al., 2019); the latter $\widetilde{\mathcal{O}}(\sqrt{T})$ bound is novel even for full observation. This latter model subsumes both LQG (partial observation) and LQR (full observation); concurrent work demonstrates that $\sqrt{T}$ regret is optimal for the LQR setting (Simchowitz and Foster, 2020). As a special case, we obtain the first (to our knowledge) $\widetilde{\mathcal{O}}(\sqrt{T})$ end-to-end regret guarantee for the problem of online LQG with an unknown system, even the stochastic setting[1]. Even with full-observation LQR setting, this is the first algorithm to attain $\sqrt{T}$ regret for either adversarial losses or semi-adversarial noise. This is also the first algorithm to obtain $\sqrt{T}$ regret *without* computing a state-space representation, demonstrating that learning methods based on improper, convex controller parametrizations can obtain this optimal rate. Adopting quite a different proof strategy than prior work (outlined in Appendix F), our bound hinges in part on a simple, useful and, to our knowledge, novel fact[2]: strongly convex online gradient descent has a *quadratic* (rather than linear) sensitivity to adversarial perturbations of the gradients (Proposition F.3).

Disturbance Response Control Our results are based on novel perspective on the classical Youla parametrization, called Disturbance Response Control (DRC). DRC affords seemless generalization to partially-observed system, competes with linear dynamic controllers, and, by avoiding state space representations, drastically simplifies our treatment of the unknown system setting. Our regret

1. An optimal $\sqrt{T}$-regret for this setting can be derived by combining Mania et al. (2019) with careful state-space system identification results of either Sarkar et al. (2019) or Tsiamis and Pappas (2019); we are unaware of work in the literature which presents this result. The DRC parametrization obviates the system identification subtleties required for this argument.
2. Robustness for the batch (fixed-objective) setting was demonstrated by Devolder et al. (2014)

guarantees are achieved by a *remarkably simple* online learning algorithm we term Disturbance Response Controller via Gradient Descent (DRC-GD): estimate the system using least squares (if it is unknown), and then run online gradient descent on surrogate losses defined by this parametrization.

In Appendix G, we present a generalization called DRC-EX, or Disturbance Response Control with Exogenous Inputs, which combines *exogenous* dictated by the DRC parametrization with a nominal stabilizing controller. This allows us to leverage the full strength of the Youla parametrization, and extend our results to arbitrary stabilizable and detectable systems. As we explain, the classical Youla parametrization requires precise system knowledge to implement. In constrast, our Nature's Y's perspective allows yields a novel formulation which is implementable under *inexact* system knowledge.

Regret Rate		
Setting	Known	Unknown
General Convex Loss Adversarial Noise	$\widetilde{\mathcal{O}}(\sqrt{T})$ (Theorem 2)	$\widetilde{\mathcal{O}}(T^{2/3})$ (Theorem 3)
Strongly Convex & Smooth Loss Adversarial + Stochastic Noise	poly $\log T$ (Theorem 4)	$\widetilde{\mathcal{O}}(\sqrt{T})$ (Theorem 5)

Table 1: Summary of our results for online control. Tables 2 and 3 in appendix compare with prior work.

1.3. Prior Work

Online Control. The field of online and adaptive control is vast and spans decades of research, see for example Sastry and Bodson (2011); Ioannou and Sun (2012) for survey. Here we restrict our discussion to online control with low *regret*, which measures the total cost incurred by the learner compared to the loss she would have incurred by instead following the best policy in some prescribed class; comparison between our results and prior art is summarized in Tables 2 and 3 in the appendix. To our knowledge, all prior end-to-end regret bounds are for the *fully observed* setting; a strength of our approach is tackling the more challenging partial observation case.

Regret for classical control models. We first survey relevant work that assume either no perturbation in the dynamics at all, or i.i.d. Gaussian perturbations. Much of this work has considered obtaining low regret in the online LQR setting (Abbasi-Yadkori and Szepesvári, 2011; Dean et al., 2018; Mania et al., 2019; Cohen et al., 2019) where a fully-observed linear dynamic system is drive by i.i.d. Gaussian noise via $\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t$, and the learner incurs constant quadratic state and input cost $\ell(x,u) = \frac{1}{2}x^\top Q x + \frac{1}{2}u^\top R u$. The optimal policy for this setting is well-approximated by a *state feedback controller* $\mathbf{u}_t = K_\star \mathbf{u}_t$, where $K_\star$ is the solution to the Discrete Algebraic Ricatti Equation (DARE), and thus regret amounts to competing with this controller. Recent algorithms Mania et al. (2019); Cohen et al. (2019) attain $\sqrt{T}$ regret for this setting, with polynomial runtime and polynomial regret dependence on relevant problem parameters. Further, Mania et al. (2019) present technical results can be used to establish $\sqrt{T}$-regret for the partially observed *LQG* setting (see Footnote 1). A parallel line by Cohen et al. (2018) establish $\sqrt{T}$ in a variant of online LQR where the system is known to the learner, noise is stochastic, but an adversary selects quadratic loss functions ℓ_t at each time t. Again, the regret is measured with respect to a best-in-hindsight state feedback controller. Provable control in the Gaussian noise setting via the policy gradient method was studied in Fazel et al. (2018). Other relevant work from the machine learning literature includes the technique of spectral filtering for learning and open-loop control of partially observable systems (Hazan et al., 2017; Arora et al., 2018; Hazan et al., 2018), as well as prior work on tracking adversarial targets (Abbasi-Yadkori et al., 2014).

The non-stochastic control problem. The setting we consider in this paper was established in Agarwal et al. (2019a), who obtain $\sqrt{T}$-regret in the more general and challenging setting where the Lipschitz loss function *and the perturbations* are adversarially chosen. The key insight behind this result is combining an improper controller parametrization know as disturbance-based control with recent advances in online convex optimization with memory due to Anava et al. (2015). Follow up work by Agarwal et al. (2019b) achieves logarithmic pseudo-regret for strongly convex, adversarially selected losses and well-conditioned stochastic noise. Under the considerably stronger condition of controllability, the recent work by Hazan et al. (2019) attains $T^{2/3}$ regret for adversarial noise/losses when the system is *unknown*. Analogous problems have also been studied in the tabular MDP setting (Even-Dar et al., 2009; Zimin and Neu, 2013; Dekel and Hazan, 2013).

Convex Parameterization of Linear Controllers There is a rich history of convex or *lifted* parameterizations of controllers. Nature's y's is equivalent to input-ouput parametrizations Zames (1981); Rotkowitz and Lall (2005); Furieri et al. (2019), and in Appendix G, we extend to more general parametrizations encompassing the classical Youla or Youla-Kucěra parametrization (Youla et al., 1976; Kučera, 1975), and approximations to the Youla parametrization which require only *approximate* knowledge of the system. More recently, Goulart et al. (2006) propose a parametrization over state-feedback policies, and Wang et al. (2019) introduce a generalization of Youla called system level synthesis (SLS); SLS is equivalent to the parametrizations adopted by Agarwal et al. (2019a) et seq., and underpins the $T^{2/3}$-regret algorithm of Dean et al. (2018) for online LQR with an unknown system; one consequence of our work is that convex parametrizations can achieve the optimal $\sqrt{T}$ in this setting. However, it is unclear if SLS (as opposed to input-output or Youla) can be used to attain sublinear regret under partial observation and adversarial noise.

Online learning and online convex optimization. We make extensive use of techniques from the field of online learning and regret minimization in games (Cesa-Bianchi and Lugosi, 2006; Shalev-Shwartz et al., 2012; Hazan, 2016). Of particular interest are techniques for coping with policy regret and online convex optimization for loss functions with memory (Anava et al., 2015).

Linear System Identification: To adress unknown systems, we make use of tools from the decades-old field linear system identification (Ljung, 1999). To handle partial observation and ensure robustness to biased and non-stochastic noise, we take up the approach in Simchowitz et al. (2019); other recent approaches include (Oymak and Ozay, 2019; Sarkar et al., 2019; Tsiamis and Pappas, 2019; Simchowitz et al., 2018).

2. Assumptions and Regret Benchmark

In the main text, we assume the system is stable:

Assumption 1 *We assume that is $\rho(A_\star) < 1$, where $\rho(\cdot)$ denotes the spectral radius.*

In Appendix G, we detail generalizations which apply to stabilizable and detectable, but potentially unstable systems. For simplicty, we assume $\mathbf{x}_0 = 0$; further, we assume:

Assumption 2 (Sub-quadratic Lipschitz Loss) *There exists a constant $L > 0$ such that non-negative convex loss functions ℓ_t obey that for all $(\mathbf{y}, \mathbf{u}), (\mathbf{y}'\mathbf{u}') \in \mathbb{R}^{d_y+d_u}$, and for the choice $R = \max\{\|(\mathbf{y}, \mathbf{u})\|_2, \|(\mathbf{y}', \mathbf{u}')\|_2, 1\}$,*[3]

$$|\ell_t(\mathbf{y}', \mathbf{u}') - \ell_t(\mathbf{y}, \mathbf{u})| \le LR \left\| \begin{bmatrix} \mathbf{y} - \mathbf{y}' \\ \mathbf{u} - \mathbf{u}' \end{bmatrix} \right\|_2 \quad \textit{and} \quad 0 \le \ell_t(\mathbf{y}, \mathbf{u}) \le LR^2.$$

Linear Dynamic Controllers Previous works on fully observable LDS consider a policy class of linear controllers, where $u_t = -Kx_t$ for some K. Here, for partially observable systems, we consider a richer class of controller with an internal notion of state. Such a policy class is necessary to capture the optimal control law in presence of i.i.d. perturbations (the LQG setting), as well as, the H_∞ control law for partially observable LDSs (Başar and Bernhard, 2008).

Definition 2.1 (Linear Dynamic Controllers) *A* linear dynamic controller, *or* LDC, *π is a linear dynamical system $(A_\pi, B_\pi, C_\pi, D_\pi)$, with internal state $\mathring{\mathbf{s}}_t \in \mathbb{R}^{d_\pi}$, input $\mathbf{y}_t^{\mathrm{in}} \in \mathbb{R}^{d_y}, \mathbf{u}_t^{\mathrm{in}} \in \mathbb{R}^{d_u}$, output $\mathbf{u}_t^{\mathrm{out}} \in \mathbb{R}^{d_u}$, equipped with the dynamical equations:*

$$\mathring{\mathbf{s}}_{t+1} = A_\pi \mathring{\mathbf{s}}_t + B_\pi \mathbf{y}_t^{\mathrm{in}} \quad \textit{and} \quad \mathbf{u}_t^{\mathrm{out}} := C_\pi \mathring{\mathbf{s}}_t + D_\pi \mathbf{y}_t^{\mathrm{in}}. \tag{2.1}$$

The closed loop *iterates $(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ are the unique sequence of iterates satisfying both the LDS dynamical equations Eq. (1.1) with $(\mathbf{y}_t, \mathbf{u}_t) = (\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ and LDC dynamical equations Eq. (2.1) with $\mathbf{y}_t^{\mathrm{in}} = \mathbf{y}_t^\pi$ and $\mathbf{u}_t^{\mathrm{out}} = \mathbf{u}_t^{\mathrm{in}} = \mathbf{u}_t^\pi$.*

The dynamics governing $(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ are described by an augmented LDS, detailed in detailed in Lemma G.2. Note that the optimal LQR and LQG controllers take the above form. The class of policies that our proposed algorithm competes is defined in terms of the Markov operators of these induced dynamical systems.

Definition 2.2 (Markov Operator) *The associated Markov operator of a linear system (A, B, C, D) is the sequence of matrices $G = (G^{[i]})_{i \ge 0} \in (\mathbb{R}^{d_\mathbf{y} \times d_\mathbf{u}})^{\mathbb{N}}$, where $G^{[0]} = D$ and $G^{[i]} = CA^{i-1}B$ for $i \ge 1$. Let $G_\star$ (resp. $G_{\pi,\mathrm{cl},e\to u}$) be the Markov operator of the nominal system $(A_\star, B_\star, C_\star, 0)$ (resp. of $(A_{\pi,\mathrm{cl}}, B_{\pi,\mathrm{cl},e}, C_{\pi,\mathrm{cl},u}, D_\pi)$, given explicity by Lemma G.2). We let $\|G\|_{\ell_1,\mathrm{op}} := \sum_{i\ge 0} \|G^{[i]}\|_{\mathrm{op}}$.*

Definition 2.3 (Decay Functions & Policy Class) *We say $\psi : \mathbb{N} \to \mathbb{R}_{>0}$ is a* proper decay function *if ψ is non-increasing and $\lim_{n\to\infty} \psi(n) = 0$. Given a Markov operator G, we define its* induced decay function *$\psi_G(n) := \sum_{i \ge n} \|G_\star^{[i]}\|_{\mathrm{op}}$. For proper decay funciton ψ, the class of all controllers whose induced* closed-loop system *has decay bounded by ψ is denoted as follows:*

$$\Pi(\psi) := \left\{\pi : \forall n \ge 0,\ \psi_{G_{\pi,\mathrm{cl},e\to u}}(n) \le \psi(n)\right\}.$$

We define $R_\psi := 1 \vee \psi(0)$ and $R_{G_\star} := 1 + \psi_{G_\star}(0)$, where $\psi_{G_\star}(0) = \sum_{i\ge 0} \|G_\star^{[i]}\|_{\mathrm{op}} = \|G_\star\|_{\ell_1,\mathrm{op}}$.

Note that the class $\Pi(\psi)$ *does not* require that the controllers be internally stable ($\rho(A_\star) < 1$), only that they induce stable closed-loop dynamics. The decay function captures the decay of the response of the system to past inputs, and is invariant to state-space representation. For stable systems G, we can always bound the decay functions by $\psi_G(m) \le C\rho^m$ for some constants $C > 0$, $\rho \in (0, 1)$; this can be made quantitative for strongly-stable systems (Cohen et al., 2018). While we assume $G_\star$ exhibits this decay in the main text, our results naturally extend to the stabilized systems via the DRC-EX parametrization (Appendix G).

3. This characterization captures, without loss of generality, any Lipschitz loss function. The $L \cdot R$ scaling of the Lipschitz constant captures, e.g. quadratic functions whose lipschitz constant scales with radius.

Regret with LDC Benchmark We are concerned with regret accumulated by an algorithm alg as the excess loss it suffers in comparison to that of the best choice of a LDC with decay ψ, specializing Eq. (1.2) with $\Pi \leftarrow \Pi(\psi)$:

$$\mathrm{Regret}_T(\psi) := J_T(\mathsf{alg}) - \min_{\pi \in \Pi(\psi)} J_T(\pi) = \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \min_{\pi \in \Pi(\psi)} \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^{\pi}, \mathbf{u}_t^{\pi}). \tag{2.2}$$

Note that the choice of the LDC in Π may be made with the complete foreknowledge of the perturbations and the loss functions that the controller π (and the algorithm alg) is subject to. We remark that the result in this paper can be easily extended to compete with controllers that have fixed affine terms (known as a *DC offset*), or periodic (time-varying) affine terms with bounded period.

3. Disturbance Response Control

The induced closed-loop dynamics for a LDC π involves feedback between the controller π and LDS, which makes the cost $J(\pi)$ non-convex in π, even in the fully observed LQR setting (Fazel et al., 2018).[4] We propose representing our controllers with the classical Youla parametrization, which both ensures convexity and is ammenable to partial observation. Our formulation emphasizes a novel perspective we call "Nature's Y's", which allows us to execute these Youla controllers in the non-stochastic setting.

Nature's y's Define $\mathbf{y}_t^{\mathrm{nat}}$ as the corresponding output of the system in the absence of any controller. Note that the sequence *does not* depend on the choice of control inputs $\mathbf{u}_t$. In the analysis, we shall assume that $1 \vee \max_t \|\mathbf{y}_t^{\mathrm{nat}}\| \le R_{\mathrm{nat}}$.

Definition 3.1 (Nature's y's) *Given a sequence of disturbances* $(\mathbf{w}_t, \mathbf{e}_t)_{t \ge 1}$, *we define the* natures $\mathbf{y}$'s *as the sequence* $\mathbf{y}_t^{\mathrm{nat}} := \mathbf{e}_t + \sum_{s=1}^{t-1} C_\star A_\star^{t-s-1} \mathbf{w}_s$.

Throughout, we assume that the noises selected by the adversary ensure $\mathbf{y}_t^{\mathrm{nat}}$ are bounded

Assumption 3 (Bounded Nature's y) *We assume that that $\mathbf{w}_t$ and $\mathbf{e}_t$ are chosen by an oblivious adversary, and that $\|\mathbf{y}_t^{\mathrm{nat}}\|_2 \le R_{\mathrm{nat}}$ for all t.* [5]

The next lemma shows for any fixed system with known control inputs the output is completely determined given Nature's y's, even if $\mathbf{w}_t, \mathbf{e}_t$ are not known. In particular, this implies that the one of the central observations of this work: *Nature's y's can be computed exactly given just control inputs and the corresponding outputs of a system.* More precisely:

Lemma 3.2 *For any LDS $(A_\star, B_\star, C_\star)$ subject to (possibly adaptive) control inputs $\mathbf{u}_1, \mathbf{u}_2, \dots \in \mathbb{R}^{d_u}$, the following relation holds for the output sequence: $\mathbf{y}_t = \mathbf{y}_t^{\mathrm{nat}} + \sum_{i=1}^{t-1} G_\star^{[i]} \mathbf{u}_{t-i}$.*

Proof This is an immediate consequence of the definitions of Nature's y's and that of a LDS. ■

4. This has motivated a long line of work to consider control parameterizations for which $J(\pi)$ is convex (Youla et al., 1976; Zames, 1981). For non-stochastic control, Agarwal et al. (2019a) consider a parametrization which selects inputs as linear functions of the disturbances $\mathbf{w}_t$, which can be exactly recovered under a full state observation. But under partial observation, the disturbances $\mathbf{w}_t$ cannot in general be recovered (e.g. whenever $C_\star$ does not possess a left inverse).
5. Note that, if the system is stable and perturbations bounded, that $\mathbf{y}_t^{\mathrm{nat}}$ will be bounded for all t.

Disturbance Response Control In the spirit of Zames (1981), we show that any linear controller can be represented by its action on Nature's y's, and that this leads to a convex parametrization of controllers which approximates the performance of any LDC controller.

Definition 3.3 (Distrubance Response Controller) *A Disturbance Response Controller (*DRC*), parameterized by a m-length sequence of matrices $M = (M^{[i]})_{i=0}^{m-1}$, chooses the control input as $\mathbf{u}_t^M = \sum_{s=0}^{m-1} M^{[s]}\mathbf{y}_{t-s}^{\mathrm{nat}}$. We let $\mathbf{y}_t^M$ denote the associated output sequence, and $J_T(M)$ the loss functional.*

Define a class of Distrubance Response Controllers with bound length and norm $\mathcal{M}(m, R) = \{M = (M^{[i]})_{i=0}^{m-1} : \|M\|_{\ell_1,\mathrm{op}} \le R\}$. Under full observation, the state-feedback policy $\mathbf{u}_t = K\mathbf{x}_t$ lies in the set of DRCs $\mathcal{M}(1, \|K\|_{\mathrm{op}})$. The following theorem, proven in Appendix C.1, states that *all stabilizing LDCs can be approximated by* DRC*s*:

Theorem 1 *For proper decay function ψ, $\pi \in \Pi(\psi)$, and any $m \ge 1$, there exists $M \in \mathcal{M}(m, R_{\mathcal{M}})$ such that $J_T(M) - J_T(\pi) \le 2LTR_{\mathcal{M}}R_{G_\star}^2 R_{\mathrm{nat}}^2\, \psi(m)$*

As $\psi(m)$ typically decays exponentially in m, we find that for any stabilizing LDC, there exists a DRC that approximately emulates its behavior. This observation ensures it sufficient for the regret guarantee to hold against an appropriately defined Disturbance Response class, as opposed to the class of LDCs. Note that the fidelity of the approximation in Theorem 1 depends only on the magnitude of the true system response $G_\star$, and *decay* of the comparator system $G_{\pi,\mathrm{cl}}$, but not on the order of a state-space realization. Theorem 1b in the appendix extends Theorem 1 to the setting where $G_\star$ may be unstable, but is placed in feedback with a stabilizing controller.

4. Algorithmic Description & Main Result

OCO with Memory: Our regret bounds are built on reductions to the online convex optimization (OCO) with memory setting as defined by Anava et al. (2015): at every time step t, an online algorithm makes a decision $x_t \in \mathcal{K}$, after which it is revealed a loss function $F_t : \mathcal{K}^{h+1} \to \mathbb{R}$, and suffers a loss of $F_t[x_t, \dots, x_{t-h}]$. The *policy regret* is defined as $\sum_{t=h+1}^{T} F_t(x, \dots, x_{t-h}) - \min_{x\in\mathcal{K}} F_t(x, \dots x)$. Anava et al. (2015) show that Online Gradient Descent on the *unary specialization* $f_t(x) := F_t(x, \dots, x)$ achieves a sub-linear policy regret bound (Proposition C.4).

Algorithm: Non-bold letters $M_0, M_1, \dots$ denote function arguments, and bold letters $\mathbf{M}_0, \mathbf{M}_1, \dots$ denote the iterates produced by the learner. We first introduce a notion of *counterfactual cost* that measures the cost incurred at the t^{th} timestep had a non-stationary distrubance feedback controller $M_{t:t-h} = (M_t, \dots, M_{t-h})$ been executed in the last h steps: This cost is entirely defined by Markov operators and Nature's y's, without reference to an explicit realization of system parameters.

Definition 4.1 (Counterfactual Costs and Dyamics) *Given $M_{t:t-h} \in \mathcal{M}(m, R_{\mathcal{M}})^{h+1}$, we define $\mathbf{u}_t\left(M_t \mid \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right) := \sum_{i=0}^{m-1} M_t^{[i]} \cdot \widehat{\mathbf{y}}_{t-i}^{\mathrm{nat}}$, $\mathbf{y}_t\left[M_{t:t-h} \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right] := \widehat{\mathbf{y}}_t^{\mathrm{nat}} + \sum_{i=1}^{h} \widehat{G}^{[i]} \cdot \mathbf{u}_{t-i}\left(M_{t-i} \mid \widehat{\mathbf{y}}_{1:t-i}^{\mathrm{nat}}\right)$, and $F_t\left[M_{t:t-h} \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right] := \ell_t\left(\mathbf{y}_t\left[M_{t:t-h} \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right], \mathbf{u}_t\left(M_t \mid \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right)\right)$. Overloading notation, for a given $M \in (\mathcal{M}, \mathcal{R}_{\mathcal{M}})$, we let $\mathbf{y}_t(M \mid \cdot) := \mathbf{y}_t[M, \dots, M \mid \cdot]$ denote the unary (single-M) specialization of $\mathbf{y}_t$, and lower case $f_t(M|\cdot) = F_t[M, \dots, M|\cdot]$ the specialization of F_t. Throughout, we use paranthesis for unary functions of $M_t \in \mathcal{M}(m, R_{\mathcal{M}})$, and brackets for functions of $M_{t:t-h} \in \mathcal{M}(m, R_{\mathcal{M}})^{h+1}$.*

For known $G_\star$, Algorithm 1 compute $\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ exactly, and we simply run online gradient descent on the costs $f_t(\cdot \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$. When $G_\star$ is unknown, we invoke Algorithm 3, which first dedicates N steps to estimating $G_\star$ via least squares (Algorithm 2), and then executes online gradient descent (Algorithm 1) with the resulting estimate $\widehat{G}$. The following algorithms are intended for *stable* $G_\star$. Unstable $G_\star$ can be handled by incorporating a nominal stabilizing controller (Appendix G).

Algorithm 1: Disturbance Response Control via Gradient Descent (DRC-GD)

Input: Stepsize $(\eta_t)_{t\geq 1}$, radius $R_{\mathcal{M}}$, memory m, Markov operator $\widehat{G}$.
Define $\mathcal{M} = \mathcal{M}(m, R_{\mathcal{M}}) = \{M = (M^{[i]})_{i=0}^{m-1} : \|M\|_{\ell_1,\mathrm{op}} \leq R_{\mathcal{M}}\}$.
Initialize $\mathbf{M}_1 \in \mathcal{M}$ arbitrarily.
for $t = 1, \dots, T$ **do**
  Observe $\mathbf{y}_t^{\mathsf{alg}}$ and determine $\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ as $\widehat{\mathbf{y}}_t^{\mathrm{nat}} \leftarrow \mathbf{y}_t^{\mathsf{alg}} - \sum_{i=1}^{t-1} \widehat{G}^{[i]} \mathbf{u}_{t-i}^{\mathsf{alg}}$. [6]
  Choose the control input as $\mathbf{u}_t^{\mathsf{alg}} \leftarrow \mathbf{u}_t\left(\mathbf{M}_t \mid \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right) = \sum_{i=0}^{m-1} \mathbf{M}_t^{[i]} \widehat{\mathbf{y}}_{t-i}^{\mathrm{nat}}$.
  Observe the loss function ℓ_t and suffer a loss of $\ell_t(y_t, u_t)$.
  Recalling $f_t(\cdot|\cdot)$ from Definition 4.1,update the disturbance feedback controller as
  $\mathbf{M}_{t+1} = \Pi_{\mathcal{M}}\left(\mathbf{M}_t - \eta_t \partial f_t\left(\mathbf{M}_t \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right)\right)$, where $\Pi_{\mathcal{M}}$ denotes projection onto $\mathcal{M}$.[7]

Algorithm 2: Estimation of Unknown System

Input: Number of samples N, system length h.
Initialize $\widehat{G}^{[i]} = 0$ for $i \notin [h]$.
For $t = 1, 2, \dots, N$, play $\mathbf{u}_t^{\mathsf{alg}} \sim \mathcal{N}(0, I_{d_u})$.
Estimate $\widehat{G}^{[1:h]} \leftarrow \arg\min \sum_{t=h+1}^{N} \|\mathbf{y}_t^{\mathsf{alg}} - \sum_{i=1}^{h} \widehat{G}^{[i]} \mathbf{u}_{t-i}^{\mathsf{alg}}\|_2^2$ via least squares, and return $\widehat{G}$.

Algorithm 3: DRC-GD for Unknown System

Input: Stepsizes $(\eta_t)_{t\geq 1}$, radius $R_{\mathcal{M}}$, memory m, rollout h, Exploration length N,
Run the estimation procedure (Algorithm 2) for N steps with system length h to estimate $\widehat{G}$
Run the regret minimizing algorithm (Algorithm 1) for $T - N$ remaining steps with estimated Markov operators $\widehat{G}$, stepsizes $(\eta_{t+N})_{t\geq 1}$, radius $R_{\mathcal{M}}$, memory m, rollout parameter h.

4.1. Main Results for Non-Stochastic Control

For simplicity, we assume a finite horizon T; extensions to infinite horizon can be obtained by a doubling trick. To simplify presentation, we will also assume the learner has foreknowledge of relevant decay parameters system norms. Throughout, let $d_{\min} = \min\{d_y, d_u\}$ and $d_{\max} =$

6. This step may be truncated to $\widehat{\mathbf{y}}_t^{\mathrm{nat}} \leftarrow \mathbf{y}_t^{\mathsf{alg}} - \sum_{i=t-h}^{t-1} \widehat{G}^{[i]} \mathbf{u}_{t-i}^{\mathsf{alg}}$; these are identical when $\widehat{G}$ is estimated from Algorithm 2, and the analysis can be extended to accomodate this truncation in the known system case
7. To simplify analysis, we project onto the ℓ_1, op-ball $\mathcal{M}(m, R) := \{M = (M^{[i]})_{i=0}^{m-1} : \|M\|_{\ell_1,\mathrm{op}} \leq R\}$. While this admits an efficient implementation (Appendix A.5), in practice one can instead project onto outer-approximations of the set, just as a Frobenius norm ball containing $\mathcal{M}(m, R)$, at the expense of a greater dependence on m.

$\max\{d_y, d_u\}$. We shall present all our results for general decay-functions, and further specialize our bounds to when the system and comparator exhibit explicity geometric decay, and where the noise satisfies subgaussian magnitude bound:

Assumption 4 (Typical Decay and Noise Bounds) *Let $C > 0$, $\rho \in (0,1)$ and $\delta \in (0,1)$. We assume that $R_{\mathrm{nat}}^2 \le d_{\max} R_{G_\star}^2 \sigma_{\mathrm{noise}}^2 \log(T/\delta)$[8], that the system decay $\psi_{G_\star}$ satisfies $\sum_{i \ge n} \|C_\star A_\star^i\|_{\mathrm{op}} \le \psi_{G_\star}$, and that $\psi_{G_\star}$ and the comparator ψ satisfies $\psi(n), \psi_{G_\star}(n) \le C\rho^n$.*

We explain the above assumption, relations between parameters, and analogues for the strong-stabilized setting adressed in Appendix A.4. For known systems, our main theorem is proved in Appendix C:

Theorem 2 (Main Result for Known System) *Suppose Assumptions 1 to 3 hold, and fix a decay function ψ. When Algorithm 1 is run with exact knowledge of Markov parameters (ie. $\hat{G} = G_\star$), radius $R_{\mathcal{M}} \ge R_\psi$, parameters $m, h \ge 1$ such that $\psi_{G_\star}(h+1) \lesssim R_{G_\star}/T$ and $\psi(m) \lesssim R_{\mathcal{M}}/T$, and step size $\eta_t = \eta = \sqrt{d_{\min}}/4LhR_{\mathrm{nat}}^2 R_{G_\star}^2 \sqrt{2mT}$, we have*[9]

$$\mathrm{Regret}_T(\psi) \lesssim L R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}^2 \sqrt{h^2 d_{\min} m} \cdot \sqrt{T}.$$

In particular, under Assumption 4, we obtain $\mathrm{Regret}_T(\psi) \lesssim \mathrm{poly}(C, \frac{1}{1-\rho}, \log\frac{T}{\delta}) \cdot \sigma_{\mathrm{noise}}^2 \sqrt{d_{\min} d_{\max}^2 T}$.

Theorem 6 in the appendix shows that $\sqrt{T}$ is the optimal rate for the above setting. For unknown systems, we prove in Appendix D:

Theorem 3 (Main Result for Unknown System) *Fix a decay function ψ, time horizon T, and confidence $\delta \in (e^{-T}, T^{-1})$. Let m, h satisfy $\psi(m) \le R_{\mathcal{M}}/\sqrt{T}$ and $\psi_{G_\star}(h+1) \le 1/10\sqrt{T}$, and suppose $R_{\mathcal{M}} \ge R_\psi$ and $R_{\mathrm{nat}} R_{\mathcal{M}} \ge \sqrt{d_u + \log(1/\delta)}$. Define the parameters*

$$C_\delta := \sqrt{d_{\max} + \log\tfrac{1}{\delta} + \log(1 + R_{\mathrm{nat}})}. \tag{4.1}$$

Then, if Assumptions 1 to 3 hold, and Algorithm 3 is run with estimation length $N = (Th^2 R_{\mathcal{M}} R_{\mathrm{nat}} C_\delta)^{2/3}$ and parameters $m, h, R_{\mathcal{M}}$, step size $\eta_t = \eta = \sqrt{d_{\min}}/4LhR_{\mathrm{nat}}^2 R_{G_\star}^2 \sqrt{2mT}$, and if $T \ge c' h^4 C_\delta^5 R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 + d_{\min} m^3$ for a universal constant c', then with probability $1 - \delta - T^{-\Omega(\log^2 T)}$,

$$\mathrm{Regret}_T(\psi) \lesssim L R_{G_\star}^2 R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 \left(h^2 R_{\mathcal{M}} R_{\mathrm{nat}} C_\delta\right)^{2/3} \cdot T^{2/3}.$$

In particular, under Assumption 4, we obtain $\mathrm{Regret}_T(\psi) \lesssim \mathrm{poly}(C, \sigma_{\mathrm{noise}}^2, \frac{1}{1-\rho}, \log\frac{T}{\delta}) \cdot d_{\max}^{5/3} L T^{2/3}$.

4.2. Fast rates under strong convexity & semi-adversarial noise

We show that OCO-with-memory obtains improved regret the losses are strongly convex and smooth, and when system is excited by persistent noise. We begin with a strong convexity assumption:

Assumption 5 (Smoothness and Strong Convexity) *For all t, $\alpha_{\mathrm{loss}} \preceq \nabla^2 \ell_t(\cdot,\cdot) \preceq \beta_{\mathrm{loss}} I$.*

8. For typical noise models, the magnitude of the covariates scales with output dimension, not internal dimension
9. If the loss is assumed to be *globably Lipschitz*, then the term $R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}^2$ can be improved to $R_{\mathrm{nat}} R_{G_\star} R_{\mathcal{M}}$.

The necessity of the smoothness assumption is explained further in Appendix E. Unfortunately, strongly convex losses are not sufficient to ensure strong convexity of the unary functions $f_t(M \mid \cdot)$. Generalizing Agarwal et al. (2019b), we assume an *semi-adversarial* noise model where disturbances decomposes as $\mathbf{w}_t = \mathbf{w}_t^{\mathrm{adv}} + \mathbf{w}_t^{\mathrm{stoch}}$ and $\mathbf{e}_t = \mathbf{e}_t^{\mathrm{adv}} + \mathbf{e}_t^{\mathrm{stoch}}$, where $\mathbf{w}_t^{\mathrm{adv}}$ and $\mathbf{e}_t^{\mathrm{adv}}$ are an adversarial sequence of disturbances, and $\mathbf{w}_t^{\mathrm{stoch}}$ and $\mathbf{e}_t^{\mathrm{stoch}}$ are stochastic disturbances which provide persistent excitation. We make the following assumption:

Assumption 6 (Semi-Adversarial Noise) *The sequences $\mathbf{w}_t^{\mathrm{adv}}$ and $\mathbf{e}_t^{\mathrm{adv}}$ and losses ℓ_t are selected by an oblivious adverary. Moreover, $\mathbf{w}_1^{\mathrm{stoch}}, \dots, \mathbf{w}_T^{\mathrm{stoch}}$ and $\mathbf{e}_1^{\mathrm{stoch}}, \dots, \mathbf{e}_T^{\mathrm{stoch}}$ are independent random variables, with $\mathbb{E}[\mathbf{w}_t^{\mathrm{stoch}}] = 0$, $\mathbb{E}[\mathbf{e}_t^{\mathrm{stoch}}] = 0$ and*

$$\mathbb{E}[\mathbf{w}_t^{\mathrm{stoch}}(\mathbf{w}_t^{\mathrm{stoch}})^\top] \succeq \sigma_{\mathbf{w}}^2 I, \quad \text{and} \quad \mathbb{E}[\mathbf{e}_t^{\mathrm{stoch}}(\mathbf{e}_t^{\mathrm{stoch}})^\top] \succeq \sigma_{\mathbf{e}}^2 I.$$

This assumption can be generalized slightly to require only a martingale structure (see Assumption 6b). Throughout, we shall also assume bounded noise. Via truncation arguments, this can easily be extended to light-tailed excitations (e.g. Gaussian) at the expense of additional logarithmic factors, as in Assumption 4. For known systems, we obtain the following bound, which we prove in Appendix E:

Theorem 4 (Logarithmic Regret for Known System) *Define the effective strong convexity parameter $\alpha_f := \alpha_{\mathrm{loss}} \cdot \left(\sigma_{\mathbf{e}}^2 + \sigma_{\mathbf{w}}^2 \left(\frac{\sigma_{\min}(C_\star)}{1+\|A_\star\|_{\mathrm{op}}^2}\right)^2\right)$ and assume Assumptions 1 to 3, 5 and 6 hold. For a decay function ψ, if Algorithm 1 is run with $\widehat{G} = G_\star$, radius $R_{\mathcal{M}} \ge R_\psi$, parameters $1 \le h \le m$ satisfying $\psi_{G_\star}(h+1) \le R_{G_\star}/T$, $\alpha \le \alpha_f$, $\psi(m) \le R_{\mathcal{M}}/T$, $T \ge \alpha m R_{\mathcal{M}}^2$, and step size $\eta_t = \frac{3}{\alpha t}$, we have that with probability $1-\delta$,*

$$\mathrm{Regret}_T(\psi) \lesssim \frac{L^2 m^3 d_{\min} R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\min\left\{\alpha, L R_{\mathrm{nat}}^2 R_{G_\star}^2\right\}} \left(1 + \frac{\beta_{\mathrm{loss}}}{L R_{\mathcal{M}}}\right) \cdot \log\frac{T}{\delta}. \tag{4.2}$$

Under Assumption 4, we have $\mathrm{Regret}_T(\psi) \lesssim \frac{L^2}{\alpha} d_{\max}^3 (1 + \beta_{\mathrm{loss}}/L) \mathrm{poly}(C, \sigma_{\mathrm{noise}}^2, \frac{1}{1-\rho}, \log\frac{T}{\delta})$.

Finally, for unknown systems, we show in Appendix F that Algorithm 3 attains optimal $\sqrt{T}$ regret:

Theorem 5 ($\sqrt{T}$-regret for Unknown System) *Fix a decay function ψ, time horizon T, and confidence $\delta \in (e^{-T}, T^{-1})$. Let $m \ge 3h \ge 1$ satisfy $\psi(\lfloor \frac{m}{2} \rfloor - h) \le R_{\mathcal{M}}/T$ and $\psi_{G_\star}(h+1) \le 1/10T$, and suppose $R_{\mathcal{M}} \ge 2R_\psi$ and $R_{\mathrm{nat}} R_{\mathcal{M}} \ge (d_u + \log(1/\delta))^{1/2}$. Finally, let $\alpha \ge \alpha_f$ for α_f as in Theorem 4, and C_δ as in Theorem 3. Then, if Assumptions 1 to 3, 5 and 6 hold, and Algorithm 3 is run with parameters $m, h, R_{\mathcal{M}}$, step sizes $\eta_t = \frac{12}{\alpha t}$, appropriate N and T sufficiently large* (Eq. (F.4))*, we have with probability $1 - \delta - T^{-\Omega(\log^2 T)}$,*

$$\mathrm{Regret}_T(\psi) \lesssim (\text{RHS of Eq. (4.2)}) + L m^{1/2} h^2 R_{G_\star}^3 R_{\mathcal{M}}^3 R_{\mathrm{nat}}^3 C_\delta \left(1 + \tfrac{L}{\alpha} + \tfrac{\beta_{\mathrm{loss}}^2}{\alpha L}\right)^{1/2} \cdot \sqrt{T}$$

Under Assumption 4, $\mathrm{Regret}_T(\psi) \lesssim \mathrm{poly}(C, L, \beta_{\mathrm{loss}}, \frac{1}{\alpha}, \sigma_{\mathrm{noise}}^2, \frac{1}{1-\rho}, \log\frac{T}{\delta}) \cdot (d_{\max}^2 \sqrt{T} + d_{\max}^3)$.

Due to space limitiations, examples for the LQR and LQG settings are deferred to Appendix A.6, and concluding remarks are provided in Appendix A.7.

Acknowledgements The authors acknowledge Tuhin Sarkar and Nikolai Matni for their insightful feedback. Elad Hazan acknowledges funding from NSF grant 1704860. Max Simchowitz is generously supported by an Open Philanthropy graduate student fellowship.

References

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

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Online least squares estimation with self-normalized processes: An application to bandit problems. *arXiv preprint arXiv:1102.2670*, 2011.

Yasin Abbasi-Yadkori, Peter Bartlett, and Varun Kanade. Tracking adversarial targets. In *International Conference on Machine Learning*, pages 369–377, 2014.

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

Naman Agarwal, Elad Hazan, and Karan Singh. Logarithmic regret for online control. In *Advances in Neural Information Processing Systems 32*, pages 10175–10184. Curran Associates, Inc., 2019b.

Oren Anava, Elad Hazan, and Shie Mannor. Online learning for adversaries with memory: price of past mistakes. In *Advances in Neural Information Processing Systems*, pages 784–792, 2015.

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

Sanjeev Arora, Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. Towards provable control for unknown linear dynamical systems. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=BygpQlbA-. rejected: invited to workshop track.

Tamer Başar and Pierre Bernhard. *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

Dimitri Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 2005.

Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603. ACM, 2014.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038, 2018.

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

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. In *Advances in Neural Information Processing Systems*, pages 4188–4197, 2018.

Ofer Dekel and Elad Hazan. Better rates for any adversarial deterministic MDP. In *International Conference on Machine Learning*, pages 675–683, 2013.

Olivier Devolder, François Glineur, and Yurii Nesterov. First-order methods of smooth convex optimization with inexact oracle. *Mathematical Programming*, 146(1-2):37–75, 2014.

Eyal Even-Dar, Sham M Kakade, and Yishay Mansour. Online Markov decision processes. *Mathematics of Operations Research*, 34(3):726–736, 2009.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Input perturbations for adaptive regulation and learning. *arXiv preprint arXiv:1811.04258*, 2018.

Maryam Fazel, Rong Ge, Sham M Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. *arXiv preprint arXiv:1801.05039*, 2018.

Luca Furieri, Yang Zheng, Antonis Papachristodoulou, and Maryam Kamgarpour. An input–output parametrization of stabilizing controllers: Amidst youla and system level synthesis. *IEEE Control Systems Letters*, 3(4):1014–1019, 2019.

Paul J Goulart, Eric C Kerrigan, and Jan M Maciejowski. Optimization over state feedback policies for robust control with constraints. *Automatica*, 42(4):523–533, 2006.

Yoram Halevi. Stable lqg controllers. *IEEE Transactions on Automatic Control*, 39(10):2104–2106, 1994.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016. ISSN 2167-3888. doi: 10.1561/2400000013. URL http://dx.doi.org/10.1561/2400000013.

Elad Hazan, Karan Singh, and Cyril Zhang. Learning linear dynamical systems via spectral filtering. In *Advances in Neural Information Processing Systems*, pages 6702–6712, 2017.

Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. Spectral filtering for general linear dynamical systems. In *Advances in Neural Information Processing Systems*, pages 4634–4643, 2018.

Elad Hazan, Sham M. Kakade, and Karan Singh. The nonstochastic control problem. *arXiv preprint arXiv:1911.12178*, 2019.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Petros A Ioannou and Jing Sun. *Robust adaptive control*. Courier Corporation, 2012.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Vladimír Kučera. Stability of discrete linear feedback systems. *IFAC Proceedings Volumes*, 8(1): 573–578, 1975.

Lennart Ljung. System identification. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–19, 1999.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalent control of lqr is efficient. *arXiv preprint arXiv:1902.07826*, 2019.

Alexander Megretski. Lecture 10: Q-parametrization. *6.245: Multivariable Control Systems*, 2004.

Ankur Moitra, William Perry, and Alexander S Wein. How robust are reconstruction thresholds for community detection? In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 828–841. ACM, 2016.

Samet Oymak and Necmiye Ozay. Non-asymptotic identification of lti systems from a single trajectory. In *2019 American Control Conference (ACC)*, pages 5655–5661. IEEE, 2019.

Ariadna Quattoni, Xavier Carreras, Michael Collins, and Trevor Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 857–864. ACM, 2009.

Michael Rotkowitz and Sanjay Lall. A characterization of convex problems in decentralized control. *IEEE transactions on Automatic Control*, 50(12):1984–1996, 2005.

Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. Finite-time system identification for partially observed lti systems of unknown order. *arXiv preprint arXiv:1902.01848*, 2019.

Shankar Sastry and Marc Bodson. *Adaptive control: stability, convergence and robustness*. Courier Corporation, 2011.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Max Simchowitz and Dylan J. Foster. Naive exploration is optimal for online lqr. *arXiv preprint arXiv:2001.09576*, 2020.

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

Max Simchowitz, Ross Boczar, and Benjamin Recht. Learning linear dynamical systems with semi-parametric least squares. In *Conference on Learning Theory*, pages 2714–2802, 2019.

Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. *arXiv preprint arXiv:1903.09122*, 2019.

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

Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. A system level approach to controller synthesis. *IEEE Transactions on Automatic Control*, 2019.

Dante Youla, Hamid Jabr, and Jr Bongiorno. Modern wiener-hopf design of optimal controllers–part ii: The multivariable case. *IEEE Transactions on Automatic Control*, 21(3):319–338, 1976.

George Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Transactions on automatic control*, 26(2):301–320, 1981.

Alexander Zimin and Gergely Neu. Online learning in episodic markovian decision processes by relative entropy policy search. In *Advances in neural information processing systems*, pages 1583–1591, 2013.

Contents

Appendix A. Appendix Organization and Notation

A.1. Organization

This appendix presents notation and organization. Appendix B presents a $\Omega(\sqrt{T})$ lower bound for the online non-stochastic control problem, even with partial observant and benign conditions. It also gives a detailed comparison with prior work, detailed in Tables 2 and 3.

Part I: Sections C, D, E, and F provide the main arguments in the proofs of our main theorems in the order in which they are presented in Section 4.

Part II: Appendix G introduces the general *stabilized setting*, where the nominal system need not be stable, but is placed in feedback with a stabilizing controller. All results from the stable setting (Assumption 1) extend to the stabilized setting (Assumption 1b) with appropriate modifications. Statement which apply to the apply specifically to stabilized setting are denoted by the number of their corresponding statment for the stable setting, with the suffix 'b'. For example, Assumption 1 stipulates the stable-system setting, and Assumption 1b the stabilized setting.

Appendix H adresses ommited proofs and stabilized-system generalizations of Theorems 2 and 3, which give regret bounds for nonstochastic control for known and unknown systems respectively. Appendix I does the same for the strongly-convex, semi-adversarial setting, namely Theorems 4 and 5. This section relies on two technical appendices: Appendix J verifies strong convexity of the induced losses under semi-adversarial noise, and Appendix K derives the regret bounds for conditionally-strongly convex losses (see Condition E.1), and under deterministic errors in the gradients(see Condition F.1).

A.2. Notation for Stable Setting

We first present the relevant notation for the *stable* setting, where the transfer function $G_\star$ of the nominal system is assumed to be stable. This is the setting assumed in the body of the text.

Transfer Operators	**Definition (Stable Case)**
$G_\star$ (stable case)	$G_\star^{[i]} = \mathbb{I}_{i \geq 0} C_\star A_\star^{i-1} B_\star$ is nomimal system (Definition 2.2)
π	refers to an LDC (Definition 2.1)
$G_{\pi,\mathrm{cl},e\to u}$	transfer function of closed loop system (Lemma G.2)
$M = (M^{[i]})$	disturbance response controller or DRC (Definition 3.3)
Transfer Classes	
ψ	proper decay function if $\sum_{n\geq 0}\psi(n) < \infty$, $\psi(n) \geq 0$
$\psi_G(n)$	$\sum_{i\geq n}\|G^{[i]}\|_{\mathrm{op}}$ (e.g. $\psi_{G_\star}$).
$\Pi(\psi)$	Policy Class $\{\pi : \forall n, \psi_{G_{\pi,\mathrm{cl}}}(n) \leq \psi(n)\}$, assuming ψ is proper
$\mathcal{M}(m,R)$	$\{M = (M^{[i]})_{i=0}^{m-1} : \|M\|_{\ell_1,\mathrm{op}} \leq R\}$ (calls of DRCs)
Input/Output Sequence	
ℓ_t	Loss function
$\mathbf{e}_t, \mathbf{w}_t$	output and state disturbances (do not depend on control policy)
$\mathbf{y}_t^{\mathrm{nat}}$	Nature's y (see Definition 3.1, also does not depend on control policy)
$\mathbf{y}_t^\pi, \mathbf{u}_t^\pi$	output, input induced by LDC policy π
$\mathbf{y}_t^M, \mathbf{u}_t^M$	output, input induced by DRC policy M
$\mathbf{y}_t^{\mathsf{alg}}$	output seen by the algorithm

$\mathbf{u}_t^{\mathsf{alg}}$	input introduced by the algorithm
$\mathbf{M}_t$	DRC selected by algorithm at step t
$\mathbf{u}_t\left(M_t \mid \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right)$	counterfactual input (Definition 4.1)
$\mathbf{y}_t(M \mid \widehat{G},\, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})$	unary counterfactual output (Definition 4.1)
$\mathbf{y}_t\left[M_{t:t-h} \mid \widehat{G},\, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right]$	non-unary counterfactual output (Definition 4.1)
$f_t(M \mid \widehat{G},\, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})$	unary counterfactual cost (Definition 4.1)
$F_t\left[M_{t:t-h} \mid \widehat{G},\, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right]$	non-unary counterfactual cost (Definition 4.1)
Radius Terms and Alg Parameters	
m	length of DRC
h	memory off approximation to transfer function
R_{nat}	$\\|\mathbf{y}_t^{\mathrm{nat}}\\| \le R_{\mathrm{nat}}$
$R_{G_\star}$	stable case: $1 \vee \\|G_\star\\|_{\ell_1,\mathrm{op}} \le R_{G_\star}$
R_ψ	$1 \vee \sum_{n\ge 0} \psi(n) \le R_\psi$
$R_{\mathcal{M}}$	$\\|\mathbf{M}_t\\|_{\ell_1,\mathrm{op}} \le R_{\mathcal{M}}$ (algorithm parameter)
$R_{\mathbf{u},\mathrm{est}} = R_{\mathbf{u},\mathrm{est}}(\delta)$	$5\sqrt{d_u + \log\frac{3}{\delta}}$
$\overline{R}_{\mathbf{u}} = \overline{R}_{\mathbf{u}}(\delta)$	$\overline{R}_{\mathbf{u}}(\delta) := 2\max\\{R_{\mathbf{u},\mathrm{est}}(\delta), R_{\mathcal{M}}R_{\mathrm{nat}}\\}$
C_δ	$\sqrt{d_{\max} + \log\frac{1}{\delta} + \log(1 + R_{\mathrm{nat}})}$ (least squares estimation constant)

A.3. Notation for Stabilized Setting

In general, we do not require that $G_\star$ be a stable matrix, but instead that $G_\star$ is placed in feedback with a stabilizing controller π_0. In this case, we let $G_\star$ denote the dynamics introduced by the feedback between the nominal system and π_0; details are given in Appendix G; at present, we summarize the relevant notation.

A.4. Relationship Between Parameters (Assumption 4)

- In typical settings, we might imagine that (w_t, e_t) are a sequence of noise which are possibly biased, by have mean say at most σ and subGaussian proxy σ^2. Then, with probability $1-\delta$, the $\max\{\|Xe_t\|, \|Xw_t\|\} \le \|X\|_{\mathrm{op}}\,\mathcal{O}\left(1\right)\cdot \sigma d \log(T/\delta)$ for any matrix X of rank at most d.

- By inflating $R_{G_\star}, \psi_{G_\star}$ if necessary, we can take $\psi_{G_\star}(n)$ to be an upper bound on

$$\mathcal{O}\left(1\right)\cdot \sum_{i\ge n} \max\{\|BA^{i-1}C\|_{\mathrm{op}}, \|A^{i-1}C\|_{\mathrm{op}}\}.$$

  This together with the previous statement yields a bound of $R_{\mathrm{nat}} \le R_{G_\star}\sigma d \log(T/\delta)$, where $d = \max\{d_u, d_y\}$.

- While this parameter regimes suggest suggests that R_{nat} is large relative to $R_{G_\star}$, we recal that $R_{\mathrm{nat}}, R_{G_\star}$ are upper bounds on various system norms, rather than exact characterizations any given norm. Thus, we can satisfy the relation in the previous bullet by inflating R_{nat} appropriately. Note that our bound degrade gracefully in R_{nat}, so this does not force an undue increase in regret.

Transfer Operators	**Definition (Stabilized Case)**
π_0	nomimal stabilizing controller
$\boldsymbol{\eta}_t$	"control-output" produced by nomimal controller reduces to $\mathbf{y}_t$ in stable case
$\mathbf{u}_t^{\mathrm{ex}}$	Exogenous input to controller reduces to $\mathbf{u}_t$ in stable case
$G_{\mathrm{ex}\to(y,u)}$	transfer function from exogenous inputs to system outputs and inputs (almost equivalent to $G_{\pi_0,\mathrm{cl}}$, see Definition 2.2b)
$G_{\mathrm{ex}\to\eta}$	transfer function from exogenous inputs to control-output $\boldsymbol{\eta}_t$ (Definition 2.2b)
Input/Output Sequence	
$\mathbf{v}_t = (\mathbf{y}_t, \mathbf{u}_t)$ $\mathbf{v}_t^{\mathsf{alg}} = (\mathbf{y}_t, \mathbf{u}_t)$ $\mathbf{v}_t^{\mathrm{nat}} = (\mathbf{y}_t^{\mathrm{nat}}, \mathbf{u}_t^{\mathrm{nat}})$	Output-Input Pair Output-Input Pair produced by algorithm Output-Input Pair with no exogenous input
$\boldsymbol{\eta}_t^{\mathrm{nat}}$	Nature's "control-output" under zero output reduces to $\mathbf{y}_t^{\mathrm{nat}}$ in stable case
$\mathbf{u}_t^{\mathrm{ex,alg}}$	exogenous input introduced by algorithm (not including nominal controller)
$\mathbf{u}_t^{\mathrm{ex}}\left(M_t \mid \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}\right)$	Exogenous input from estimates of $\boldsymbol{\eta}^{\mathrm{nat}}$ See Definition 4.1b for expression, and for below
$\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right]$	Prediction of $\mathbf{v}_t$ under estimated dynamics and $\boldsymbol{\eta}^{\mathrm{nat}}$.
$F_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right]$	Prediction of loss under estimated dynamics and $\boldsymbol{\eta}^{\mathrm{nat}}$.
$f_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}})$	Unary specialization of the above.
Radius Terms	
$R_{G_\star}$	$1 \vee \|G_{\star,u\to u}\|_{\ell_1,\mathrm{op}} \vee \|G_{\star,(\mathbf{y}\to\mathbf{u})}\|_{\ell_1,\mathrm{op}} \le R_{G_\star}$

- The reason for the geometric decay is as follows: any stable matrix A with $\rho(A) < 1$ admits a positive definite Lyapunov matrix $P \succeq I$, for which $A^\top PA \preceq (1-\epsilon)P$, for some appropriate ϵ. This implies that, for a suitable constant $\kappa > 0$ depending on $\|C\|_{\mathrm{op}}, \|B\|_{\mathrm{op}}$, and $\|P\|$, $\max\{\|CA^{i-1}B\|, \|CA^{i-1}B\|\} \le \kappa(1-\epsilon)^{i-1}$, Thus, using our inflated definition $\psi_{G_\star} = \mathcal{O}(1) \cdot \sum_{i \ge n} \max\{\|BA^{i-1}C\|_{\mathrm{op}}, \|A^{i-1}C\|_{\mathrm{op}}\}$, we have that $\psi_{G_\star}(n) = \mathcal{O}(\kappa) \sum_{i \ge n} K(1-\epsilon)^{n-1} \le \kappa(1-\epsilon)^{n-1}/\epsilon$. Absorbing these other factors int oκ gives the desired geometric decrease.

- Most conditions can be relaxed up to constant factors, because online learning methods degrade gracefully when parameters are misspecified. The main exceptions are: (a) one needs to still choose h, m so that $\psi_{G_\star}(h) \lesssim 1/T$, and similarly $\psi(m) \lesssim 1/T$. If the decay parameters are not known exactly, then the learner must choose a larger h to be conservative. (b), for strongly convex losses, the effective strong convexity parameter used must be *less* than the true strong convexity modulus. Lastly, (c), parameters out to be selected so as to ensure stability in the unknown system setting (see Lemma D.1)

A.5. Efficient ℓ_1, op Projection

We describe an efficient implemtation of the ℓ_1, op projection step in the algorithms above. As with other spectral norms, it suffices to diagonalize and compute a projection of the singular values onto the corresponding vector-ball, which in this case is the ball: $\{(z^{[i]}) : z^{[i]} = 0, i > m, \sum_{i=0}^{m-1} \|z^{[i]}\|_\infty \le R\}$; an efficient algorithm for this projection step is given by Quattoni et al. (2009).

A.6. Examples for LQR and LQG

We now demonstrate how our results specialize to the LQR and LQG settings:

Example 1 (LQR) *In the LQR setting, the observable state* $\mathbf{x}_t$ *evolves as* $\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t$*, where* $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_{\mathbf{w}}^2)$*, and the associated losses are fixed quadratics* $l(\mathbf{x}, \mathbf{u}) = \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u}$*. The optimal control*[10] *is expressible as* $\mathbf{u}_t = -K\mathbf{x}_t$ *(trivially an LDC). Our framework realizes this setting by choosing* $C_\star = I$ *and* $\sigma_{\mathbf{e}}^2 = 0$ *(observations are noiselss). The strong convexity parameter is then* $\alpha_f = \frac{\sigma_{\mathbf{w}}^2}{(1+\|A_\star\|_{\mathrm{op}}^2)^2}\alpha_{loss}$*, which degrades with the norm of* $A_\star$*, but does not vanish even as* $A_\star$ *becomes unstable. For LQR, Theorem 5 guarantees a regret of* $\tilde{O}(\sqrt{T})$ *matching the previous results (Cohen et al., 2019; Mania et al., 2019).*

Example 2 (LQG) *In the LQG setting, the state evolves as Equation 1.1, where* $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_{\mathbf{w}}^2), \mathbf{e}_t \sim \mathcal{N}(0, \sigma_{\mathbf{e}}^2)$*, and the associated losses are fixed quadratics* $l(\mathbf{y}, \mathbf{u}) = \mathbf{y}^\top Q \mathbf{y} + \mathbf{u}^\top R \mathbf{u}$*. The optimal control for a known system may be obtained via the separation principle (Bertsekas, 2005), which involves the applying the LQR controller on a latent-state estimate* z_t *obtained via Kalman filtering.*

$$\widehat{\mathbf{x}}_{t+1} = A_\star \widehat{\mathbf{x}}_{t+1} + B_\star \mathbf{u}_t + L(\mathbf{y}_{t+1} - C_\star(A_\star \widehat{\mathbf{x}}_t + B_\star \mathbf{u}_t)); \qquad \mathbf{u}_t = K\mathbf{z}_t$$

By setting $\mathbf{z}_t = \widehat{\mathbf{x}}_{t-1}$ *and* $A_\pi = (I - LC_\star)(A_\star + B_\star K)$*, we get* $\mathbf{z}_{t+1} = A_\pi \mathbf{z}_t + L\mathbf{y}_t$*, and* $\mathbf{u}_t = KA_\pi \mathbf{z}_t + KL\mathbf{y}_t$ *therefore, implying that such a filtering-enabled policy is a LDC too. For an unknown LQG system, Theorem 5 guarantees a regret of* $\tilde{O}(\sqrt{T})$*.*

We remark both of the above examples can be extended to the setting where $G_\star$ may be unstable, but is placed in feedback with a known stabilizing controller (Assumption 1b) via Theorems Theorems 4b and 5b; assumption of such a stabilizing control is standard in the LQR setting. We note that for general partially-observed stabilized settings, the strong convexity modulus is somewhat more opaque, but still yields $\widetilde{\mathcal{O}}\left(\sqrt{T}\right)$ regret asymptotically.

10. In strict terms, this is only true for the infinite horizon case. However, even in the finite horizon setting, such a control law (utilizing the infinite horizon controller) is at most $\log T$ sub-optimal additively.

A.7. Concluding Remarks

This work presented a new adaptive controller we termed Disturbance Response Control via Gradient Descent (DRC-GD), inspired by a Youla's parametrization. This method is particularly suitable for controlling system with partial observation, where we show an efficient algorithm that attains the first sublinear regret bounds under adversarial noise for both known and unknown systems.

This technique attains optimal regret rates for many regimes of interest. Notably, this is the only technique which attains $\sqrt{T}$-regret for partially observed, non-stochastic control model with general convex losses. Our bound is also the first technique to attain $\sqrt{T}$-regret for the classical LQG problem, and extends this bound to a more general semi-adversarial setting.

In future work we intend to implement these methods and benchmark them against recent novel methods for online control, including the gradient pertrubation controller Agarwal et al. (2019a). We also intend to compare our guarantees to techniques tailored to the stochastic setting, including Certainty Equivalence Control Mania et al. (2019), Robust System Level System Dean et al. (2018), and SDP-based relaxations Cohen et al. (2019). We also hope to design variants of DRC-GD which adaptively select algorithm parameters to optimize algorithm performance, and remove the need for prior knowledge about system properties (e.g. decay of the nominal system). Lastly, we hope to understand how to use these convex parametrizations for related problem formulations, such as robustness to system mispecification, safety contraints, and distributed control.

Appendix B. Comparison with Past Work & Lower Bounds

B.1. Comparison to Prior Work

Tables 2 and 3 describe regret rates for existing alorithms for known system and unknown system settings, respectively. Within each table, bold lines further divide the results into nonstochastic and stochastic/semi-adversarial regimes. Specifically, stochastic noise means well conditioned noise that is bounded or light-tailed, non-stochastic noise means noise selected by an arbitrary adversary, and semi-adversarial noise is an intermediate regime described formally by Assumption 6/ 6b. Compared to past work in non-stochastic control, we compete with stabilizing LDCs, which strictly generalize state feedback control. We note however that for stochastic linear control with fixed quadratic costs, state feedback is optimal, up to additive constants that do not grow with horizon T.

Table 2: Comparison with prior work for known system. See above for explanation of relevant settings.

Comparison with Past Work: Known System

Work	Rate	Obs.	Loss Type	Noise Type	Comparator
Agarwal et al. (2019a)	$\sqrt{T}$	Full	Adversarial Lipschitz	Nonstochastic	Disturbance & State Feedback
Theorem 2	$\sqrt{T}$	Partial	Adversarial Lipschitz	Nonstochastic	Stabilizing LDC
Cohen et al. (2018) (Known System & Noise)[(a)]	$\sqrt{T}$	Full	Adversarial Quadratic	Stochastic	State Feedback (Pseudo-regret)
Agarwal et al. (2019b) (Known System & Noise)[(a)]	$\mathrm{poly}\log T$	Full	Adversarial Strongly Convex	Stochastic	Disturbace & State Feedback (Pseudo-regret)[(b)]
Theorem 4	$\mathrm{poly}\log T$	Partial	Adversarial Strongly Convex & Smooth[(c)]	Semi-Adversarial	Stabilizing LDC

[(a)] Agarwal et al. (2019b), Cohen et al. (2018) assume the knowledge of the noise model making the assumption stronger than simply knowing the system

[(b)] Pseudo-regret refers to the best comparator "outside the expectation". It is strictly weaker than regret.

[(c)] The smoothness assumption is necessary to remove the need for the expected-gradient oracle, and can be removed if such a stronger oracle is provided.

Table 3: Comparison with prior work for unknown system. See above for explanation of relevant settings.

Comparison with Past Work: Unknown System					
Hazan et al. (2019) [(d)]	$T^{2/3}$	Full	Adversarial Lipschitz	Nonstochastic	Disturbance & State Feedback
Theorem 3	$T^{2/3}$	Partial	Adversarial Lipschitz	Nonstochastic	Stabilizing LDC
Abbasi-Yadkori & Szepesvári [2011]	$e^d \cdot \sqrt{T}$ [(e)]	Full	Fixed Quadratic	Stochastic	State Feedback
Dean et al. (2018)	$T^{2/3}$	Full	Fixed Quadratic	Stochastic	State Feedback
Cohen et al. (2019) Faradonbeh et al. (2018) Mania et al. (2019)[(f)]	$\sqrt{T}$	Full	Fixed Quadratic	Stochastic	State Feedback
Theorem 5	$\sqrt{T}$	Partial	Adversarial Strongly Convex & Smooth[(g)]	Semi-Adversarial	Stabilizing LDC

[(d)] To identify the system, Hazan et al. (2019) assumes that the pair $(A_\star, B_\star)$ satisfies a strong contrability assumption. Our Nature's y's formulation dispenses with this assumption.
[(e)] This bound is exponential in dimension d.
[(f)] The authors in Mania et al. (2019) present technical guarantees that can be used to imply $T^{2/3}$ regret for the partially observed setting when combined with concurrent results. Since the paper was released, stronger system identification guarantees can be used to establish $\sqrt{T}$ regret for this setting (Sarkar et al., 2019; Tsiamis and Pappas, 2019). To our knowledge, this complete end-to-end result does not yet exist in the literature.
[(g)] Unlike Theorem 5, smoothness is still necessary even when given access to the stronger oracle. Alernatively, certain noise distributions (e.g. Gaussian) can be used to induce smoothness.

B.2. Regret Lower Bounds for Known Systems

We formally prove our lower bound in the following interaction model:

Definition B.1 (Lower Bound Interaction Model) *We assume that* $\mathbf{x}_{t+1} = A_\star \mathbf{x}_t + B_\star \mathbf{u}_t + \mathbf{w}_t$, *where* $\mathbf{w}_t$ *are drawn i.i.d. from a fixed distribution. We assume that the learners controlers* $\mathbf{u}_t^{\mathsf{alg}}$ *may depend arbitrarily on* $\mathbf{x}_t$ *and* $(\mathbf{u}_s, \mathbf{x}_s)_{1 \le s < t}$*. For a policy class* Π *and joint distribution* $\mathcal{D}$ *over losses and disturbances, we define*

$$\mathrm{PseudoRegret}_T^{\mathsf{alg}}(\Pi, \mathcal{D}) := \mathbb{E}_{\mathcal{D}}\left[\sum_{t=1}^{T} \ell_t(\mathbf{x}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})\right] - \inf_{\pi \in \Pi} \mathbb{E}_{\mathcal{D}}\left[\sum_{t=1}^{T} \ell_t(\mathbf{x}_t^{\pi}, \mathbf{u}_t^{\pi})\right] \le \mathbb{E}\left[\mathrm{Regret}_T(\Pi)\right],$$

Informally, our lower bound states that $\sqrt{T}$ regret is necessary to compete with the optimal *state feedback* controller for pseudo-regret in the fully observed regime, either when the noises are stochastic and loss is known to the learner, or the the noises are constant and deterministic, and the losses stochastic. Formally:

Theorem 6 *Let $d_x = 2$, and $d_u = 1$, $A_\star = \mathbf{0}_{d_x \times d_x}$, $B_\star = -e_1$, and Π denote the set of all state-feedback controllers of the form $\mathbf{u}_t = K_v \mathbf{x}_t$, for $K_v = v \cdot e_1^\top$. Then for the interaction Model of Definition B.1, the following hold*

1. **Fixed Lipschitz Loss & Unknown i.i.d Noise:** *Fix a loss $\ell(x, u) = |x[1]|$, and a family of distributions $\mathscr{P} = \{\mathcal{P}\}$ over i.i.d. sequences of $\mathbf{w}_t$ with $\|\mathbf{w}_t\| \le 2$ for $T \ge 2$*

$$\inf_{\mathsf{alg}} \max_{\mathcal{D} \in \{\mathbb{I}_{\ell_t = \ell}\} \otimes \mathscr{P}} \mathrm{PseudoRegret}_T^{\mathsf{alg}}(\Pi, \mathcal{D}) \ge -1 + \Omega(T^{1/2}),$$

where $\{\mathbb{I}_{\ell_t = \ell}\} \otimes \mathscr{P}$ is the set of joint loss and noise distribution induced $\ell_t = \ell$ and $\mathbf{w}_t \overset{\text{i.i.d}}{\sim} \mathcal{P}$.

2. **I.i.d Lipschitz Loss & Known Deterministic Noise** *Then there exists a family of distributions $\mathscr{P} = \{\mathcal{P}\}$ over i.i.d sequences of 1-Lipschitz loss functions with $0 \le \ell_t(0) \le 1$ almost surely such that*

$$\inf_{\mathsf{alg}} \max_{\mathcal{D} \in \mathscr{P} \otimes \{I_{\mathbf{w}_t = (1,0)}\}} \mathrm{PseudoRegret}_T^{\mathsf{alg}}(\Pi, \mathcal{D}) \ge -1 + \Omega(T^{1/2})$$

where $\mathscr{P} \otimes \{I_{\mathbf{w}_t = (0,1)}\}\}$ is the set of joint loss and noise distribution induced $\ell_1, \ell_2, \ldots \overset{\text{i.i.d}}{\sim} \mathcal{P}$ and $\mathbf{w}_t = (1, 0)$ for all t.

Proof Let us begin by proving Part 1. Let $\mathscr{P}$ denote the set of distributions $\mathcal{P}_p$ where $\mathbf{w}_t[2] = 1$ for all t, and $\mathbf{w}_t[1] \overset{\text{i.i.d}}{\sim} \mathrm{Bernoulli}(p)$ for $t \ge 1$. Let $\mathbb{E}_p$ denote the corresponding expectation operator, and let $\mathrm{PseudoRegret}_T^{\mathsf{alg}}(T, p)$ denote the associated PseudoRegret. We can verify

$$\mathbf{x}_{t+1}^{\mathsf{alg}} = \begin{bmatrix} \mathbf{w}_t[1] - \mathbf{u}_t^{\mathsf{alg}}[1] \\ 1 \end{bmatrix}, \mathbf{u}_t^{K_v} = v, \mathbf{x}_t^{K_v} = \begin{bmatrix} \mathbf{w}_t[1] - v \\ 1 \end{bmatrix}$$

For $\ell(x, u) = |x[1]|$, we have

$$\mathbb{E}_p[\sum_{t=1}^T \ell(\mathbf{x}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})] = \mathbb{E}_p[\sum_{t=1}^T |\mathbf{w}_t[1] - \mathbf{u}_t^{\mathsf{alg}}[1]|] \ .$$

Since $\mathbf{w}_t[1] \in \{0, 1\}$, we can assume $\mathbf{u}_t^{\mathsf{alg}}[1] \in [0, 1]$, since projecting into this interval always decreases the regret. In this case, given the interaction model, $\mathbf{w}_t[1] \mid \mathbf{u}_t^{\mathsf{alg}}[1]$ is still $\mathrm{Bernoulli}(p)$ distributed. Therefore,

$$\begin{aligned} \mathbb{E}_p[\sum_{t=1}^T |\mathbf{w}_t[1] - \mathbf{u}_t^{\mathsf{alg}}[1]|] &= \mathbb{E}_p[\sum_{t=1}^T (1-p)\mathbf{u}_t^{\mathsf{alg}} + p(1 - \mathbf{u}_t^{\mathsf{alg}})] \\ &= pT + (1 - 2p)\mathbb{E}_p[\sum_{t=1}^T \mathbf{u}_t^{\mathsf{alg}}] \\ &= pT + (1 - 2p)T\mathbb{E}_p[Z_t], \end{aligned}$$

where we let $Z_t := \frac{1}{T}\sum_{t=1}^{T} \mathbf{u}_t^{\mathsf{alg}} \in [0,1]$. On ther other hand, for any $v \in [0,1]$,

$$\mathbb{E}_p[\sum_{t=1}^{T} \ell(\mathbf{x}_t^{K_v}, \mathbf{u}^{K_v})] \le 1 + T\{(1-p)v + p(1-v)\} = 1 + pT + (1-2p)Tv,$$

where the additive 1 accounts for the initial time step. Hence,

$$\mathrm{PseudoRegret}_T^{\mathsf{alg}}(T,p) \ge -1 + \max_{v\in[0,1]} (1-2p)T(\mathbb{E}_p(Z_t) - v) = -1 + T\,|1-2p| \cdot |\mathbb{E}_p(Z_t) - \mathbb{I}(1-2p \ge 0)|\,.$$

The lower bound now follows from a hypothesis testing argument. Since $Z_t \in [0,1]$, it follows that there exists an $\epsilon = \Omega(T^{-1/2})$ such that (see e.g. Kaufmann et al. (2016))

$$|\mathbb{E}_{p=1/2+\epsilon}(Z_t) - \mathbb{E}_{p=1/2-\epsilon}(Z_t)| \le \frac{7}{8}.$$

Combining with the previous display, this shows that for $p \in \{1/2-\epsilon, 1/2+\epsilon\}$,

$$\mathrm{PseudoRegret}_T^{\mathsf{alg}}(T,p) \ge -1 + T\,|1-2p|\,\Omega(1) \ge -1 + \Omega(T\epsilon) = -1 + \Omega(T^{1/2}).$$

This proves part 1. Part 2 follows by observing that the above analysis goes through by moving the disturbacnce into the loss, namely $\ell_t(x,u) = |x[1] - \mathbf{e}_t|$ where $\mathbf{e}_t \overset{\text{i.i.d}}{\sim} \mathrm{Bernoulli}(p)$ and $\mathbf{w}_t[1] = 0$ for zero. ■

Part I
High Level Proofs

Appendix C. Analysis for Known System

In this section, we prove Theorem 2. We begin with the following regret decomposition, for simplicity, we abbreviate $\mathcal{M} \leftarrow \mathcal{M}(m, R_{\mathcal{M}})$:

$$\begin{aligned}\mathrm{Regret}_T(\Pi(\psi)) &= \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \inf_{\pi\in\Pi(\psi)} \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^{\pi}, \mathbf{u}_t^{\pi}) \qquad \text{(C.1)}\\ &\le \underbrace{\left(\sum_{t=1}^{m+h} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})\right)}_{\text{burn-in loss}} + \underbrace{\left(\sum_{t=m+h+1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \sum_{t=m+h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}}]\right)}_{\text{algorithm truncation error}} \\ &+ \underbrace{\left(\sum_{t=m+h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}}] - \inf_{M\in\mathcal{M}} \sum_{t=m+h+1}^{T} f_t(M \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}})\right)}_{f \text{ policy regret}}\end{aligned}$$

$$+ \underbrace{\left(\inf_{M \in \mathcal{M}} \sum_{t=m+h+1}^{T} f_t(M \mid G_\star,\, \mathbf{y}^{\mathrm{nat}}_{1:t}) - \inf_{M \in \mathcal{M}} \sum_{t=m+h+1}^{T} \ell_t(\mathbf{y}^M_t, \mathbf{u}^M_t) \right)}_{\text{comparator truncation error}} \tag{C.2}$$

$$+ \underbrace{\inf_{M \in \mathcal{M}} \sum_{t=1}^{T} \ell_t(\mathbf{y}^M_t, \mathbf{u}^M_t) - \inf_{\pi \in \Pi(\psi)} \ell_t(\mathbf{y}^\pi_t, \mathbf{u}^\pi_t)}_{\text{policy approximation} := J(M) - J(\pi)} \tag{C.3}$$

Here, the burn-in captures rounds before the algorithm attains meaningful regret guarantees, the truncation errors represent how closely the counterfactual losses track the losses suffered by the algorithm (algorithm truncation error), or those suffered by the algorithm selecting policy $\pi = M$. The dominant term in the above bound in the f policy regret, which we bound using the OCO-with-Memory bound from Proposition C.4. Lastly, the policy approximation error measures how well finite-history policies $M \in \mathcal{M}$ approximate LDC's $\pi \in \Pi(\psi)$, and is adressed by Theorem 1; this demonstrates the power of the nature's y parametrization.

We shall now bound the regret term-by-term. All subsequent bounds hold in the more general setting of stabilized-systems (defined in Appendix G), and all ommited proofs are given in Appendix H.1. Before beginning, we shall need a uniform bound on the magnitude of $\mathbf{u}^{\mathsf{alg}}_t$ and $\mathbf{y}^{\mathsf{alg}}_t$. This is crucial because the magnitudes and Lipschitz constants of the losses ℓ_t depend on the magnitudes of their arguments:

Lemma C.1 (Magnitude Bound) *For all t, and $M, M_1, M_2, \dots \in \mathcal{M}$, we ahve*

$$\max\left\{ \left\|\mathbf{u}^{\mathsf{alg}}_t\right\|_2, \left\|\mathbf{u}^M_t\right\|_2, \left\|\mathbf{u}_t(M_t \mid \mathbf{y}^{\mathrm{nat}}_{1:t})\right\|_2 \right\} \le R_{\mathcal{M}} R_{\mathrm{nat}}$$

$$\max\left\{ \left\| \begin{bmatrix} \mathbf{y}^{\mathsf{alg}}_t \\ \mathbf{u}^{\mathsf{alg}}_t \end{bmatrix} \right\|_2, \left\| \begin{bmatrix} \mathbf{y}^M_t \\ \mathbf{u}^M_t \end{bmatrix} \right\|_2, \left\| \begin{bmatrix} \mathbf{y}_t[M_{t:t-h} \mid G_\star,\, \mathbf{y}^{\mathrm{nat}}_{1:t}] \\ \mathbf{u}_t[M_{t:t-h} \mid G_\star,\, \mathbf{y}^{\mathrm{nat}}_{1:t}] \end{bmatrix} \right\|_2 \right\} \le 2 R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}$$

Proof The proof is a special case of Lemma C.1b in the appendix. ■

The above lemma directly yields a bound on the first term of the regret decomposition (C.3):

Lemma C.2 *We have that* (burn-in loss) $\le 4 L R^2_{G_\star} R^2_{\mathcal{M}} R^2_{\mathrm{nat}} (m+h)$

Proof Combine Assumption 2 on the loss, Lemma C.1, and the fact $R_{G_\star}, R_{\mathcal{M}}, R_{\mathrm{nat}} \ge 1$. ■

The algorithm and comparator truncation errors represent the extent to which the h-step truncation differs from the true losses induced by the algorithm:

Lemma C.3 (Bound on Truncation Errors) *We can bound*

$$\text{(algorithm truncation error)} + \text{(comparator truncation error)} \le 4 L T R_{G_\star} R^2_{\mathcal{M}} R^2_{\mathrm{nat}} \psi_{G_\star}(h+1).$$

Now, we turn the the f-regret. We begin by quoting a result of Anava et al. (2015):

Proposition C.4 *For any a sequence of $(h+1)$-variate F_t, define $f_t(x) = F_t(x, \dots x)$. Let L_c be an upper bound on the coordinate-wise Lipschitz constant of F_t, L_f be an upper bound on the*

Lipschitz constant of f_t, and D be an upper bound on the diameter of $\mathcal{K}$. Then, the sequence $\{x_t\}_{t=1}^T$ produced by executing OGD on the unary loss functions f_t with learning rate η guarantees

$$\text{PolicyRegret} := \sum_{t=h+1}^{T} F_t(x, \ldots, x_{t-h}) - \min_{x \in \mathcal{K}} F_t(x, \ldots x) \ \leq \ \frac{D^2}{\eta} + \eta T \cdot (L_f^2 + h^2 L_c L_f).$$

In order to apply the OCO reduction, we need to bound the appropriate Lipschitz constants. Notice that, in order to apply projected gradient descent, we require that the functions f_t are Lipschitz in the Euclidean (i.e., Frobenius) norm:

Lemma C.5 (Lipschitz/Diameter Bounds) *Define $L_f := L\sqrt{m} R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}$. Then,*

- *The functions $f_t(\cdot \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$ are L_f-Lipschitz*
- *The functions $F_t[M_{t:t-h} \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}]$ are L_f-coordinate-wise Lipschitz on $\mathcal{M}$ in the Frobenius norm $\|M\|_{\mathrm{F}} = \|[M^{[0]}, \ldots, M^{[m-1]}]\|_{\mathrm{F}}$.*
- *the Euclidean diameter of $\mathcal{M}$ is at most $D \leq 2\sqrt{d_{\min}} R_{\mathcal{M}}$.*

We now bound the policy regret by appealing to the OCO-with-Memory guarante, Proposition C.4:

Lemma C.6 (Bound on the f-policy regret) *Let $d_{\min} = \min\{d_u, d_y\}$, and $\eta_t = \eta = \sqrt{d_{\min}}/4LhR_{\mathrm{nat}}^2 R_{G_\star}^2 \sqrt{2mT}$ for all t. Then,*

$$(f\text{-policy regret}) \ \leq \ 2L\sqrt{T d_{\min} m} h R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}^2$$

Proof From Proposition C.4 with $L_f = L_c$ as in Lemma C.5, and diameter $D \leq 2\sqrt{d_{\min}} R_{\mathcal{M}}$ from the same lemma, Lemma H.2, we

$$(f\text{-policy regret}) \ \leq \ \frac{D^2}{\eta} + TL_f^2(h^2+1)\eta \ \leq \ \frac{R_{\mathcal{M}}^2 d \min}{\eta} + T(LR_{\mathrm{nat}}^2 R_{G_\star}^2)^2 R_{\mathcal{M}}^2 m(h^2+1)\eta.$$

Selecting $\eta = \sqrt{d_{\min}}/2LhR_{\mathrm{nat}}^2 R_{G_\star}^2 \sqrt{2mT}$ and bounding $\sqrt{h^2+1} \leq \sqrt{2}h$ concludes the proof. ■

Recalling the bound on policy approximation from Theorem 1, we combine all the relevant bounds above to prove our regret guarantee:

Proof [Proof of Theorem 2] Summing up bounds in Lemmas C.2, C.3 and C.6, and Theorem 1, and using $R_{\mathcal{M}} \geq 1$,

$$\begin{aligned}
\mathrm{Regret}_T(\psi) &\lesssim LR_{\mathrm{nat}}^2 \left(R_{\mathcal{M}}^2 R_{G_\star}^2 (m+h) + \sqrt{T} h \sqrt{m d_{\min}} R_{\mathcal{M}}^2 R_{G_\star}^2 + R_{\mathcal{M}}^2 R_{G_\star} \psi_{G_\star}(h+1)T + R_{\mathcal{M}} R_{G_\star}^2 \psi(m) T \right) \\
&\lesssim LR_{\mathrm{nat}}^2 \left(\sqrt{T} h \sqrt{m d_{\min}} R_{\mathcal{M}}^2 R_{G_\star}^2 + R_{\mathcal{M}}^2 R_{G_\star} \psi_{G_\star}(h+1)T + R_{\mathcal{M}} R_{G_\star}^2 \psi(m) T \right) \\
&= LR_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}^2 \sqrt{T} \left(h\sqrt{m d_{\min}} + \sqrt{T} \left(\frac{\psi_{G_\star}(h+1)}{R_{G_\star}} + \frac{\psi(m)}{R_{\mathcal{M}}} \right) \right).
\end{aligned}$$

■

C.1. Proof of Theorem 1

Proof Let $(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ be the output-input sequence produced on the execution of a LDC π on a LDS $(A_\star, B_\star, C_\star)$, and $(\mathbf{y}_t^M, \mathbf{u}_t^M)$ be the output-input sequence produced by the execution of an Disturbance Feedback Controller M on the same LDS. By Lemma G.2, the closed-loop dynamics are given by

$$\begin{bmatrix} \mathbf{x}_{t+1}^\pi \\ \mathbf{z}_{t+1}^\pi \end{bmatrix} = \underbrace{\begin{bmatrix} A_\star + B_\star D_\pi C_\star & B_\star C_\pi \\ B_\pi C_\star & A_\pi \end{bmatrix}}_{A_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{x}_t^\pi \\ \mathbf{z}_t^\pi \end{bmatrix} + \underbrace{\begin{bmatrix} I & B_\star D_\pi \\ 0 & B_\pi \end{bmatrix}}_{B_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{e}_t \end{bmatrix}, \tag{C.4}$$

$$\begin{bmatrix} \mathbf{y}_t^\pi \\ \mathbf{u}_t^\pi \end{bmatrix} = \underbrace{\begin{bmatrix} C_\star & 0 \\ D_\pi C_\star & C_\pi \end{bmatrix}}_{C_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{x}_t^\pi \\ \mathbf{z}_t^\pi \end{bmatrix} + \begin{bmatrix} I \\ D_\pi \end{bmatrix} \mathbf{e}_t. \tag{C.5}$$

Further, define $C_{\pi,\mathrm{cl},u}$ as the second row of $C_{\pi,\mathrm{cl}}$, and $B_{\pi,\mathrm{cl},w}$ and $B_{\pi,\mathrm{cl},e}$ as the first and second columns of $B_{\pi,\mathrm{cl}}$. We then have

$$\mathbf{u}_t^\pi = D_\pi \mathbf{e}_t + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},e} \mathbf{e}_{t-s} + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},w} \mathbf{w}_{t-s}.$$

Our argument hinges on the following claim which we establish shortly below:

Claim C.7 (Control Approximation Identity) *Define the matrices* $M^{[0]} = D_\pi$*, and* $M^{[i]} = C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{[i-1]} B_{\pi,\mathrm{cl},e}$ *for* $i \geq 1$*. Then,*

$$\mathbf{u}_t^\pi = \sum_{i=0}^{t-1} M^{[i]} \mathbf{y}_{t-i}^{\mathrm{nat}}.$$

As a consequence, we have that, for $\mathbf{u}_t^M = \sum_{i=0}^{m-1} M^{[i]} \mathbf{u}_{t-i}$, we find

$$\mathbf{u}_t^\pi = \mathbf{u}_t^M + \sum_{i=m}^{t-1} M^{[i]} \mathbf{y}_{t-i}^{\mathrm{nat}} \tag{C.6}$$

This, in particular, implies the following bounds.

$$\begin{aligned} \|\mathbf{u}_t^\pi - \mathbf{u}_t^M\| &\leq \left(\sum_{i=m}^{t-1} \|C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{i-1} B_{\pi,\mathrm{cl},e}\| \right) \max_i \|\mathbf{y}_i^{\mathrm{nat}}\| \leq \psi(m) R_{\mathrm{nat}} \\ \|\mathbf{y}_t^\pi - \mathbf{y}_t^M\| &\leq \|G_\star\|_{\ell_1,\mathrm{op}} \psi(m) R_{\mathrm{nat}} \end{aligned}$$

Hence,

$$\left\| \begin{bmatrix} \mathbf{u}_t^\pi - \mathbf{u}_t^M \\ \mathbf{y}_t^\pi - \mathbf{y}_t^M \end{bmatrix} \right\|_2 \leq (1 + \|G_\star\|_{\ell_1,\mathrm{op}}) \psi(m) R_{\mathrm{nat}} := R_{G_\star} \psi(m) R_{\mathrm{nat}}.$$

Moreover, from Eq. (C.6), we can show that $\|(\mathbf{u}_t^\pi, \mathbf{y}_t^\pi)\| \leq 2R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}$, as per Lemma C.1. Thus, from the sub-quadratic assumption (Assumption 2),

$$|\ell_t(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi) - \ell_t(\mathbf{y}_t^M, \mathbf{u}_t^M)| \leq 2LR_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} \left\| \begin{bmatrix} \mathbf{u}_t^\pi - \mathbf{u}_t^M \\ \mathbf{y}_t^\pi - \mathbf{y}_t^M \end{bmatrix} \right\|_2 \leq 2LR_{G_\star}^2 R_{\mathrm{nat}}^2 R_{\mathcal{M}} \psi(m).$$

■

Proof [Proof of Claim C.7]

$$\begin{aligned}
&\sum_{i=0}^{t-1} M^{[i]} \mathbf{y}_{t-i}^{\mathrm{nat}} = \sum_{i=0}^{t-1} M^{[i]} \left(\mathbf{e}_{t-i} + \sum_{s=1}^{t-i-1} C_\star A_\star^{t-i-s-1} \mathbf{w}_s \right) \\
&= \sum_{i=0}^{t-1} M^{[i]} \mathbf{e}_{t-i} + \sum_{s=1}^{t-1} \sum_{i=0}^{s-1} (M^{[i]} C_\star A_\star^{s-1-i}) \mathbf{w}_{t-s} \\
&= D_\pi \mathbf{e}_t + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},e} \mathbf{e}_{t-s} + D_\pi \sum_{s=1}^{t-1} C_\star A_\star^{t-s} \mathbf{w}_t + \sum_{s=1}^{t-1} \sum_{i=1}^{s-1} (M^{[i]} C_\star A_\star^{s-1-i}) \mathbf{w}_{t-s}
\end{aligned} \tag{C.7}$$

Let us unpack the last line:

$$\begin{aligned}
M^{[i]} C_\star A_\star^{s-1-i} &= C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{i-1} B_{\pi,\mathrm{cl},e} C_\star A_\star^{s-1-i} \mathbf{w}_{t-s} \\
&= C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{i-1} \begin{bmatrix} B_\star D_\pi C_\star \\ B_\pi C_\star \end{bmatrix} A_\star^{s-1-i} \mathbf{w}_{t-s} \\
&= C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{i-1} \underbrace{\begin{bmatrix} B_\star D_\pi C_\star & B_\star C_\pi \\ B_\pi C_\star & A_\pi \end{bmatrix}}_{(i)} \begin{bmatrix} A_\star^{s-1-i} \mathbf{w}_{t-s} \\ 0 \end{bmatrix},
\end{aligned}$$

where we fill the last two columns of the matrix (i) arbitrarily, since $\begin{bmatrix} A_\star^{s-1-i} \mathbf{w}_t \\ 0 \end{bmatrix}$ has a zero in its second block component. Define the matrices $(X, Y) = \left(\begin{bmatrix} A_\star & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} B_\star D_\pi C_\star & B_\star C_\pi \\ B_\pi C_\star & A_\pi \end{bmatrix} \right)$. We then recognize $A_{\pi,\mathrm{cl}} = X + Y$, and can thus express

$$M^{[i]} C_\star A_\star^{s-i-1} = C_{\pi,\mathrm{cl},u} (X+Y)^{i-1} Y X^{s-i-1} \begin{bmatrix} \mathbf{w}_{t-s} \\ 0 \end{bmatrix}$$

Before proceeding, observe the following identity for any positive integer n.

$$X^n = Y^n + \sum_{i=1}^{n} X^{i-1} (X - Y) Y^{n-i}.$$

Thus,

$$\begin{aligned}
\sum_{i=1}^{s-1} (M^{[i]} C_\star A_\star^{s-1-i}) \mathbf{w}_{t-s} &= -C_{\pi,\mathrm{cl},u} X^{s-1} \begin{bmatrix} \mathbf{w}_{t-s} \\ 0 \end{bmatrix} + C_{\pi,\mathrm{cl},u} (X+Y)^{s-1} \begin{bmatrix} \mathbf{w}_{t-s} \\ 0 \end{bmatrix} \\
&= -D_\pi C_\star A_\star^{s-1} \mathbf{w}_{t-s} + C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},w} \mathbf{w}_{t-s}
\end{aligned}$$

Picking up where we left off,

$$\begin{aligned}
&\sum_{i=0}^{t-1} M^{[i]}\mathbf{y}^{\mathrm{nat}}_{t-i} \\
&= D_\pi \mathbf{e}_t + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},e}\mathbf{e}_{t-s} + D_\pi \sum_{s=1}^{t-1} C_\star A_\star^{t-s}\mathbf{w}_t + \sum_{s=1}^{t-1}\sum_{i=1}^{s-1}(M^{[i]} C_\star A_\star^{s-1-i})\mathbf{w}_{t-s} \\
&= D_\pi \mathbf{e}_t + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},e}\mathbf{e}_{t-s} + \sum_{s=1}^{t-1}\sum_{i=0}^{s-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},w}\mathbf{w}_{t-s} \\
&= D_\pi \mathbf{e}_t + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},e}\mathbf{e}_{t-s} + \sum_{s=1}^{t-1} C_{\pi,\mathrm{cl},u} A_{\pi,\mathrm{cl}}^{s-1} B_{\pi,\mathrm{cl},w}\mathbf{w}_{t-s} = \mathbf{u}_t^\pi. \qquad \text{(C.8)}
\end{aligned}$$

■

Appendix D. Analysis for Unknown System

D.1. Estimation of Markov Operators

In this section, we describe how to estimate the hidden system. We prove the following theorem assuming some knowledge about the decay of $G_\star$. We defer the setting where the learner does not have this knowledge to later work. The proof of the following guarantee applies results from Simchowitz et al. (2019), and is given in Appendix Appendix H.2:

Theorem 7 (Guarantee for Algorithm 2) *Let $\delta \in (e^{-T}, T^{-1})$, $N, d_u \le T$, and $\psi_{G_\star}(h{+}1) \le \frac{1}{10}$. For universal constants c, C_{est}, define*

$$\epsilon_G(N,\delta) = C_{\mathrm{est}}\frac{h^2 R_{\mathrm{nat}}}{\sqrt{N}} C_\delta, \qquad \text{where } C_\delta := \sqrt{d_{\max} + \log\tfrac{1}{\delta} + \log(1+R_{\mathrm{nat}})}.$$

and suppose that $N \ge c h^4 C_\delta^4 R_{\mathcal{M}}^2 R_{G_\star}^2$. Then with probability $1-\delta-N^{-\log^2 N}$, Algorithm Algorithm 2 satisfies the following bounds

1. *For all $t \in [N]$, $\|\mathbf{u}_t\| \le R_{\mathbf{u},\mathrm{est}}(\delta) := 5\sqrt{d_u + 2\log(3/\delta)}$*
2. *The estimation error is bounded as*

$$\|\widehat{G} - G_\star\|_{\ell_1,\mathrm{op}} \le \|\widehat{G}^{[0:h]} - G_\star^{[0:h]}\|_{\ell_1,\mathrm{op}} + R_{\mathbf{u},\mathrm{est}}\psi_{G_\star}(h+1) \le \epsilon_G(N,\delta) \le 1/2\max\{R_{\mathcal{M}}R_{G_\star}, R_{\mathbf{u},\mathrm{est}}\}.$$

For simplicity, we suppress the dependence of ϵ_G on N and δ when clear from context. Throughout, we shall assume the following condition

Condition D.1 (Estimation Condition) *We assume that the event of Theorem 7 holds.*

D.2. Stability of Estimated Nature's y

One technical challenge in the analyis of the unknown $G_\star$ setting is that the estimates $\widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}$ *depend on the history of the algorithm*, because subtracting off the contribution of the inexact estimate $\widehat{G}$ *does not* entirely mitigate the effects of past inputs. Hence, our the first step of the analysis is to show that if $\widehat{G}$ is sufficiently close to $G_\star$, then this dependence on history does not lead to an unstable feedback. Note that the assumption of the following lemma holds under Condition D.1:

Lemma D.1 (Stability of $\widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}$) *Introduce the notation $\overline{R}_{\mathbf{u}}(\delta) := 2\max\{R_{\mathbf{u},\mathrm{est}}(\delta), R_{\mathcal{M}}R_{\mathrm{nat}}\}$, assume that $\epsilon_G(N,\delta) \le 1/2\max\{R_{\mathcal{M}}R_{G_\star}\}$. Then, for $t \in [T]$, we have the bounds*

$$\|\mathbf{u}^{\mathsf{alg}}_t\|_2 \le \overline{R}_{\mathbf{u}}(\delta), \quad \|\widehat{\mathbf{y}}^{\mathrm{nat}}_t\|_2 \le 2R_{\mathrm{nat}}, \quad \|\mathbf{y}^{\mathsf{alg}}_t\| \le R_{\mathrm{nat}} + R_{G_\star}\overline{R}_{\mathbf{u}}(\delta)$$

Proof Introduce $\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty} := \max_{s\le t}\|\mathbf{u}^{\mathsf{alg}}_s\|_2$. We then have

$$\|\mathbf{y}^{\mathrm{nat}}_t - \widehat{\mathbf{y}}^{\mathrm{nat}}_t\|_2 = \left\|\sum_{s=1}^{t} G_\star^{[t-s]}\mathbf{u}^{\mathsf{alg}}_s - \widehat{G}^{[t-s]}\mathbf{u}^{\mathsf{alg}}_s\right\|_2 \le \|\widehat{G} - G_\star\|_{\ell_1,\mathrm{op}} \max_{s\le t}\left\|\mathbf{u}^{\mathsf{alg}}_s\right\|_2 \le \epsilon_G\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty}. \tag{D.1}$$

We now have that

$$\begin{aligned}
\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty} &\le \max\left\{R_{\mathbf{u},\mathrm{est}}, \left\|\max_{s\le t}\sum_{j=t-m+1}^{j-1}\mathbf{M}_s^{[t-s]}\widehat{\mathbf{y}}^{\mathrm{nat}}_j\right\|\right\}\\
&\le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}\max_{s\le t-1}\|\widehat{\mathbf{y}}^{\mathrm{nat}}_s\|_2\}\\
&\le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}(R_{\mathrm{nat}} + \|G_\star\|_{\ell_1,\mathrm{op}}\epsilon_G\|\mathbf{u}^{\mathsf{alg}}_{1:t-1}\|_{2,\infty})\}\\
&\le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\} + R_{\mathcal{M}}\|G_\star\|_{\ell_1,\mathrm{op}}\epsilon_G\|\mathbf{u}^{\mathsf{alg}}_{1:t-1}\|_{2,\infty}
\end{aligned}$$

Moreover, by assumption, we have $\epsilon_G \le 1/2R_{\mathcal{M}}R_{G_\star}$, so that

$$\begin{aligned}
\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty} &\le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\} + \|\mathbf{u}^{\mathsf{alg}}_{1:t-1}\|_{2,\infty}/2\\
&\le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\} + \frac{1}{2}(R_{\mathbf{u},\mathrm{est}} + R_{\mathcal{M}}R_{\mathrm{nat}}) + \|\mathbf{u}^{\mathsf{alg}}_{1:t-2}\|_{2,\infty}/4\\
&\le \cdots \le\ 2\max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\} := \overline{R}_{\mathbf{u}}.
\end{aligned}$$

The bound $\|\widehat{\mathbf{y}}^{\mathrm{nat}}_t\|_2 \le 2R_{\mathrm{nat}}$ follows by plugging the above into Eq. (D.1), the the final bound from $\|(\mathbf{y}^{\mathsf{alg}}_t, \mathbf{u}^{\mathsf{alg}})\| \le R_{\mathrm{nat}} + \|G_\star\|_{\ell_1,\mathrm{op}}\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty} + \|\mathbf{u}^{\mathsf{alg}}_t\|_2 \le R_{\mathrm{nat}} + (1 + \|G_\star\|_{\ell_1,\mathrm{op}})\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty} := R_{\mathrm{nat}} + R_{G_\star}\|\mathbf{u}^{\mathsf{alg}}_{1:t}\|_{2,\infty}$. ■

D.3. Regret Analysis

We apply an analogous regret decomposition to the proof of Theorem 2, again abbreviating $\mathcal{M} \leftarrow \mathcal{M}(m, R_{\mathcal{M}})$:

$$
\begin{aligned}
\mathrm{Regret}_T(\Pi(\psi)) \le\ & \underbrace{\left(\sum_{t=1}^{m+2h+N} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})\right)}_{\text{estimation \& burn-in loss}} \\
& + \underbrace{\left(\sum_{t=N+m+2h+1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \sum_{t=N+m+2h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid \widehat{G},\, \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}]\right)}_{\text{loss approximation-error}}, \\
& + \underbrace{\left(\sum_{t=N+m+2h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid \widehat{G},\, \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}] - \inf_{M\in\mathcal{M}} \sum_{t=N+m+2h+1}^{T} f_t(M \mid \widehat{G},\, \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t})\right)}_{\widehat{f}\text{ policy regret}} \\
& + \underbrace{\left(\inf_{M\in\mathcal{M}} \sum_{t=N+m+2h+1}^{T} f_t(M \mid \widehat{G},\, \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}) - \inf_{M\in\mathcal{M}} \sum_{t=N+m+2h+1}^{T} \ell_t(\mathbf{y}_t^{M}, \mathbf{u}_t^{M})\right)}_{\text{comparator approximation-error}} \\
& + \underbrace{\inf_{M\in\mathcal{M}} \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^{M}, \mathbf{u}_t^{M}) - \inf_{\pi\in\Pi(\psi)} \ell_t(\mathbf{y}_t^{\pi}, \mathbf{u}_t^{\pi})}_{\text{policy approximation } := J(M) - J(\pi)}
\end{aligned} \tag{D.2}
$$

Let us draw our attention two the main differences between the present decomposition and that in Eq. (C.3): first, the burn-in phase contains the initial estimation stage N. During this phase, the system is excited by the Guassian inputs before estimation takes place. Second, the truncation costs are replaced with approximation errors, which measure the discrepancy between using $\widehat{G}$ and $\widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}$ and using $G_\star$ and $\mathbf{y}^{\mathrm{nat}}_{1:t}$.

Observe that the policy approximation error is exactly the same as that from the known-system regret bound, and is adressed by Theorem 1. Moreover, the policy regret can be bounded by a black-box reduction to the policy regret in the known-system cases:

Lemma D.2 *Assume Condition D.1. Then, for $\eta_t = \eta \propto \sqrt{d_{\min}}/LhR_{\mathrm{nat}}^2 R_{G_\star}^2 \sqrt{mT}$, we have*

$$
(\widehat{f}\text{-policy regret}) \lesssim \sqrt{T} L \sqrt{d_{\min} m} h R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}^2
$$

Proof Observe that the $\widehat{f}$-regret depends only on the $\widehat{\mathbf{y}}^{\mathrm{nat}}_t$ and $\widehat{G}$ sequence, but not on any other latent dynamics of the system. Hence, from the proof Lemma C.6, we see can see more generally that if $\max_t \|\widehat{\mathbf{y}}^{\mathrm{nat}}_t\| \le R'_{\mathrm{nat}}$ and $\|\widehat{G}\|_{\ell_1,\mathrm{op}} \le R'_{G_\star}$, then $\eta \propto \sqrt{d_{\min}}/Lh(R'_{\mathrm{nat}})^2 (R'_{G_\star})^2 \sqrt{mT}$,

$$
(\widehat{f}\text{-policy regret}) \lesssim L\sqrt{d_{\min} m h^2 T} (R'_{\mathrm{nat}})^2 (R'_{G_\star})^2 R_{\mathcal{M}}^2
$$

In particular, under Condition D.1 and by Lemma D.1, we can take $R'_{\mathrm{nat}} \le 2R_{\mathrm{nat}}$ and $R'_{G_\star} \le 2R_{G_\star}$. ■

From Lemma D.1, $\|\mathbf{y}_t^{\mathsf{alg}}\|_2 \le R_{\mathrm{nat}} + R_{G_\star}\overline{R}_{\mathbf{u}} \le 2R_{G_\star}\overline{R}_{\mathbf{u}}$. Assumption 2 then yields

Lemma D.3 *Under Condition D.1, we have that* (estimation & burn-in) $\le 4L(m+2h+N)R_{G_\star}^2\overline{R}_{\mathbf{u}}^2$.

To conclude, it remains to bound the approximation errors. We begin with the following bound on the accuracy of estimated nature's y, proven in Appendix H.3.1:

Lemma D.4 (Accuracy of Estimated Nature's y) *Assume Condition D.1, and let $t \ge N+h+1$, we have that $\|\widehat{\mathbf{y}}_t^{\mathrm{nat}} - \mathbf{y}_t^{\mathrm{nat}}\| \le 2R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$.*

We then use this to show that the estimation error is linear in T, but also decays linearly in ϵ_G:

Lemma D.5 (Approximation Error Bounds) *Under Condition D.1,*

$$\text{(loss approximation error)} + \text{(comparator approximation error)} \lesssim LTR_{G_\star}R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\epsilon_G$$

Proof [Proof Sketch] For the "loss approximation error", we must control the error introduce by predicting using $\widehat{G}$ instead of $G_\star$, and by the difference from affine term $\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ in $\mathbf{y}(\cdot \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})$ from the true natures y $\mathbf{y}_t^{\mathrm{nat}}$. For the "comparator approximation error", we must also adress the mismatch between using the estimated $\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ sequence of the controls in the functions $f_t(\cdot \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})$, and the true natures y's $\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ for the sequence $(\mathbf{y}_t^M, \mathbf{u}_t^M)$. A complete proof is given in Appendix H.3.2 ■

Proof [Proof of Theorem 3] Assuming Condition D.1, taking $N \ge m+h$ and combining Lemma D.3 with the substituting $\overline{R}_{\mathbf{u}}(\delta) \lesssim \max\{\sqrt{d_u + \log(1/\delta)}, R_{\mathcal{M}}R_{G_\star}\}$, and with Lemmas D.2 and D.5 and theorem 1,

$$\mathrm{Regret}_T(\psi) \lesssim LR_{G_\star}^2R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\left(\left(\frac{d_u + \log(1/\delta)}{R_{\mathrm{nat}}^2R_{\mathcal{M}}^2} \vee 1\right)N + T\epsilon_G(N,\delta)R_{\mathcal{M}} + \sqrt{dmh^2T} + \frac{T\psi(m)}{R_{\mathcal{M}}}\right).$$

For $\psi(m) \le R_{\mathcal{M}}/\sqrt{T}$, the last term is dominated by the second-to-last. Now, for the constant $C(\delta)$ as in the theorem statment, and for $R_{\mathrm{nat}}R_{\mathcal{M}} \ge d_u + \log(1/\delta)$, we have $(\frac{d_u+\log(1/\delta)}{R_{\mathrm{nat}}^2R_{\mathcal{M}}^2} \vee 1)N + T\epsilon_G(N,\delta)R_{\mathcal{M}} \le (N + C_\delta Th^2R_{\mathcal{M}}R_{\mathrm{nat}}/\sqrt{N})$. We see that if we have $N = (Th^2R_{\mathcal{M}}R_{\mathrm{nat}}C_\delta)^{2/3}$, then the above is at most

$$\mathrm{Regret}_T(\psi) \lesssim LR_{G_\star}^2R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\left((h^2TR_{\mathcal{M}}R_{\mathrm{nat}}C_\delta)^{2/3} + \sqrt{d_{\min}h^2mT}\right).$$

Finally, since $\psi_{G_\star}(h+1) \le 1/10\sqrt{T}$, one can check that Condition D.1 holds with probability $1 - 3\delta - N^{-\log^2 N} = 1 - \delta - T^{-\Omega(\log^2 T)}$ as soon as $T \ge c'h^4C_\delta^5R_{\mathcal{M}}^2R_{\mathrm{nat}}^2$ for a universal constant c'. When $T \ge d_{\min}m^3$, we can bound the above by $\lesssim LR_{G_\star}^2R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\left(h^2R_{\mathcal{M}}R_{\mathrm{nat}}C_\delta\right)^{2/3} \cdot T^{2/3}$. ■

Appendix E. Logarithmic Regret for Known System

In this section, we prove Theorem 4. The analoguous result for the strongly-stabilized setting is proved in Appendix I.

Theorem 4 applies the same regret decomposition as Theorem 2; the key difference is in bounding the f-policy regret in Eq. (C.3). Following the strategy of Agarwal et al. (2019a), we first show that the persistent excitation induces strongly convex losses (in expectation). Unlike this work, we *do not assume* access to gradients of expected functions, but only the based on losses and outputs revelead to the learner. We therefore reason about losses conditional on $k \ge m$ steps in the past:

Definition E.1 (Filtration and Conditional Functions) *Let $\mathcal{F}_t$ denote the filtration generated by the stochastic sequences $\{(\mathbf{e}_s^{\mathrm{stoch}}, \mathbf{w}_t^{\mathrm{stoch}})\}_{s \le t}$, and define the conditional losses*

$$f_{t;k}\left(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}\right) := \mathbb{E}\left[f_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}) \mid \mathcal{F}_{t-k}\right].$$

A key technical component is to show that $f_{t;k}$ are strongly convex:

Proposition E.2 *For α_f as in Theorem 4 and $t \ge k \ge m$, $f_{t;k}(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$ is α_f-strongly convex.*

The above proposition is proven in Appendix J.4, with the Appendix J devoted to establishing a more general bound for strongly-stabilized (but not necessarily stable) systems.

Typically, one expects strong-convex losses to yield $\log T$-regret. However, only the condition expectations of the loss are strongly convex; the losses themselves are not. Agarwal et al. (2019a) assume that access to a gradient oracle for expected losses, which circumvents this discrepancy. In this work, we show that such an assumption is not necessary if the unary losses are also β-smooth.

We now set up regularity conditions and state a regret bound (Theorem 8) under which conditionally-strong convex functions yield logarithmic regret. The proof of Theorem 4 follows by specializing these conditions to the problem at hand.

Condition E.1 (Unary Regularity Condition (uRC) for Conditionally-Strongly Convex Losses) *Suppose that $\mathcal{K} \subset \mathbb{R}^d$. Let $f_t := \mathcal{K} \to \mathbb{R}$ denote a sequence of functions and $(\mathcal{F}_t)_{t \ge 1}$ a filtration. We suppose f_t is L_{f}-Lipschitz, and $\max_{x \in \mathcal{K}} \|\nabla^2 f_t(x)\|_{\mathrm{op}} \le \beta$, and that $f_{t;k}(x) := \mathbb{E}[f_t(x) \mid \mathcal{F}_{t-k}]$ is α-strongly convex on $\mathcal{K}$.*

Observe that Proposition E.2 precisely establishes the strong convexity requirement for uRC, and we can verify the remaining conditions below. In the Appendix, we prove a generic high-probability regret bounds for applying online gradient descent to uRC functions (Theorem 12). Because we require bounds on policy regret, here we shall focus on a consequence of that bound for the "with-memory" setting:

Condition E.2 (With-Memory Regularity Condition (wmRC)) *Suppose that $\mathcal{K} \subset \mathbb{R}^d$ and $h \ge 1$. We let $F_t := \mathcal{K}^{h+1} \to \mathbb{R}$ be a sequence of L_{c} coordinatewise-Lipschitz functions with the induced unary functions $f_t(x) := F_t(x, \dots, x)$ satisfying Condition E.1.*

Our main regret bound in the with-memory setting is as follows:

Theorem 8 *Let $\mathcal{K} \subset \mathbb{R}^d$ have Euclidean diameter D, consider functions F_t and f_t satisfying Condition E.2 with $k \ge h \ge 1$. Consider gradient descent updates $z_{t+1} \leftarrow \Pi_{\mathcal{K}}(z_t - \eta_{t+1} \nabla f_t(z_t))$, with $\eta_t = \frac{3}{\alpha t}$ applied for $t \ge t_0$ for some $t_0 \le k$, with $z_0 = z_1 = \cdots = z_{t_0} \in \mathcal{K}$. Then, with probability $1 - \delta$,*

$$\sum_{t=k+1}^{T} F_t(z_t, z_{t-1}, \dots, z_{t-h}) - \inf_{z \in \mathcal{K}} \sum_{t=k+1}^{T} f_t(z)$$
$$\lesssim \alpha k D^2 + \frac{(k + h^2) L_{\mathrm{f}} L_{\mathrm{c}} + k d L_{\mathrm{f}}^2 + k \beta L_{\mathrm{f}}}{\alpha} \log(T) + \frac{k L_{\mathrm{f}}^2}{\alpha} \log\left(\frac{1 + \log(e + \alpha D^2)}{\delta}\right).$$

The above bound is a special case of Theorem 9 below, a more general result that adresses complications that arise when $G_\star$ is unknown. Our regret bound incurs a dimension factor d due to a uniform convergence argument[11], which can be refined for structured $\mathcal{K}$.

To conclude the proof of Theorem 4, it we simply apply Theorem 8 with the appropriate parameters.

Proof [Proof of Theorem 4] From Lemma C.5, we can take $L_{\mathrm{f}} := 4L\sqrt{m}R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}$ and $D \le 2\sqrt{d}R_{\mathcal{M}}$. We bound the smoothness in Appendix I.1:

Lemma E.3 (Smoothness) *The functions $f_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$ are β_f-smooth, where we define $\beta_f := mR_{\mathrm{nat}}^2 R_{G_\star}^2 \beta_{\mathrm{loss}}$.*

This yields that, for $\alpha \le \alpha_f$ and step sizes $\eta_t = \frac{3}{\alpha t}$, the f-policy regret is bounded by

$$
\begin{aligned}
&\lesssim \alpha m d R_{\mathcal{M}}^2 + \frac{(md+h^2)L_{\mathrm{f}}^2 + m\beta_f L_{\mathrm{f}}}{\alpha}\log(T) + \frac{mL_{\mathrm{f}}^2}{\alpha}\log\left(\frac{1+\log(e+\alpha_f d R_{\mathcal{M}}^2)}{\delta}\right) \\
&\lesssim \alpha m d R_{\mathcal{M}}^2 + \frac{L^2 m^2 R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\alpha}\left(\left(d + \frac{h^2}{m}\right) + \frac{\beta_{\mathrm{loss}} m^{1/2}}{L d R_{\mathcal{M}}}\right) \\
&\qquad \cdot \max\left\{\log T, \log\left(\frac{1+\log(e+\alpha_f d R_{\mathcal{M}}^2)}{\delta}\right)\right\} \\
&\lesssim \alpha m d R_{\mathcal{M}}^2 + \frac{L^2 m^3 d R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\alpha} \\
&\qquad \cdot \max\left\{1, \frac{\beta_{\mathrm{loss}}}{L d R_{\mathcal{M}}}\right\}\log\left(\frac{T+\log(e+\alpha d R_{\mathcal{M}}^2)}{\delta}\right) && (1 \le h \le m) \\
&\lesssim \alpha m d R_{\mathcal{M}}^2 + \frac{L^2 m^3 d R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\alpha}\max\left\{1, \frac{\beta_{\mathrm{loss}}}{L d R_{\mathcal{M}}}\right\}\log\left(\frac{T}{\delta}\right) && (T \ge \log(e+\alpha d R_{\mathcal{M}}^2)) \\
&\lesssim \frac{L^2 m^3 d R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\alpha \wedge L R_{\mathrm{nat}}^2 R_{G_\star}^2}\max\left\{1, \frac{\beta_{\mathrm{loss}}}{L d R_{\mathcal{M}}}\right\}\log\left(\frac{T}{\delta}\right), && \text{(E.1)}
\end{aligned}
$$

Therefore, combining the above with Lemmas C.2 and C.3, and Theorem 1,

$$
L R_{\mathrm{nat}}^2 R_{\mathcal{M}}^2 R_{G_\star}^2 \left(m + h + \frac{\psi_{G_\star}(h+1)T}{R_{G_\star}} + \frac{\psi(m)T}{R_{\mathcal{M}}} + \text{(Eq. (E.1))}\right).
$$

Applying $\psi_{G_\star}(h+1) \le R_{G_\star}/T$, $\psi(m) \le R_{\mathcal{M}}/T$, and $h \le m$, the term in Eq. (E.1) dominates. ■

Appendix F. $\sqrt{T}$-regret for unknown system under strong convexity

In this section, we prove of Theorem 5, which requires the most subtle argument of the four settings considered in the paper. We begin with a high level overview, and defer the precise steps to Appendix F.1. The core difficulty in proving this result is demonstrating that the error ϵ_G in estimating

11. This is because we consider best comparator $z_\star$ for the realized losses, rather than a pseudoregret comparator defined in terms of expectations

the system propagates *quadratically* as $T\epsilon_G^2/\alpha$ (for appropriate $\alpha = \alpha_f/4$), rather than as $T\epsilon_G$ in the weakly convex case. By setting $N = \sqrt{T/\alpha}$, we obtain regret bounds of roughly

$$\mathrm{Regret}_T(\psi) \lessapprox N + \frac{T}{\alpha}\epsilon_G(N,\delta)^2 + \frac{1}{\alpha}\log T \lessapprox \sqrt{\frac{T}{\alpha}} + \frac{\log T}{\alpha}\,, \tag{F.1}$$

where we let $\lessapprox$ denote an informal inequality, possibly suppressing problem-dependent quantities and logarithmic factors, and use $\epsilon_G(N,\delta) \lessapprox 1/\sqrt{N}$. To achieve this bound, we modify our regret decomposition by introducing the following a *hypothetical* "true prediction" sequence:

Definition F.1 (True Prediction Losses) *We define the true prediction losses as*

$$\mathbf{y}_t^{\mathrm{pred}}[M_{t:t-h}] := \mathbf{y}_t^{\mathrm{nat}} + \sum_{i=1}^{h} G_\star^{[i]} \cdot \mathbf{u}_{t-i}\left(M_{t-i} \mid \widehat{\mathbf{y}}_{1:t-i}^{\mathrm{nat}}\right)$$
$$F_t^{\mathrm{pred}}\left[M_{t:t-h}\right] := \ell_t\left(\mathbf{y}_t^{\mathrm{pred}}\left[M_{t:t-h}\right], \mathbf{u}_t\left(M_t \mid \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}\right)\right),$$

and let $f_t^{\mathrm{pred}}(M) = F_t^{\mathrm{pred}}(M,\ldots,M)$ denote the unary specialization, and define the conditional unary functions $f_{t;k}^{\mathrm{pred}}(M) := \mathbb{E}[f_t^{\mathrm{pred}}(M) \mid \mathcal{F}_{t-k}]$.

Note that the affine term of $\mathbf{y}_t^{\mathrm{pred}}$ is the true nature's y[12], and the inputs $\mathbf{u}_{t-i}\left(M_{t-i} \mid \widehat{\mathbf{y}}_{1:t-i}^{\mathrm{nat}}\right)$ are multiplied by the true transfer function $G_\star$. Thus, up to a truncation by h, $\mathbf{y}_t^{\mathrm{pred}}$ describes the *true* counterfactual output of system due to the control inputs $\mathbf{u}_{t-i}\left(M_{t-i} \mid \widehat{\mathbf{y}}_{1:t-i}^{\mathrm{nat}}\right)$ selected based on the *estimated* nature's y's. F_t^{pred} and f_t^{pred} then correspond to the counterfactual loss functions induced by these true counterfactuals.

While the algorithm does not access the unary losses f_t^{pred} directly (it would need to know $\mathbf{y}_t^{\mathrm{nat}}$ and $G_\star$ to do so), we show in Appendix I.2.2 that the gradients of f_t^{pred} and $f_t(M \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})$ are $\mathcal{O}(\epsilon_G)$ apart:

Lemma F.2 *For any $M \in \mathcal{M}$, we have that*

$$\left\|\nabla f_t(M \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}) - \nabla f_t^{\mathrm{pred}}(M)\right\|_{\mathrm{F}} \le C_{\mathrm{approx}}\,\epsilon_G,$$

where we define $C_{\mathrm{approx}} := \sqrt{m} R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}^2 (8\beta_{\mathrm{loss}} + 12L)$.

As a consequence, we can view Algorithm 1 as performing gradient descent with respect to the sequence f_t^{pred}, but with non-stochastic errors $\boldsymbol{\epsilon}_t := \nabla f(M \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}) - \nabla f_t^{\mathrm{pred}}(M)$. The key observation is that online gradient descent with strongly convex losses is robust in that the regret grows *quadraticaly* in the errors via $\frac{1}{\alpha}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2$. By modifying the step size slightly, we also enjoy a negative regret term. The following bound applies to the standard strongly convex online learning setup:

Proposition F.3 (Robustness of Strongly Convex OGD) *Let $\mathcal{K} \subset \mathbb{R}^d$ be convex with diameter D, and let f_t denote a sequence of α-strongly convex functions on $\mathcal{K}$. Consider the gradient update*

12. Note that $\mathbf{y}_t^{\mathrm{pred}}[M_{t:t-h}]$ differs from $\mathbf{y}_t[M_{t:t-h} \mid G_\star, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}]$ (the counterfactual loss given the true function $G_\star$ and estimates $\widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}$) precisely in this affine term

rules $z_{t+1} = \Pi_{\mathcal{K}}(z_t - \eta_{t+1}(\nabla f_t(z_t) + \boldsymbol{\epsilon}_t))$*, where* $\boldsymbol{\epsilon}_t$ *is an arbitrary error sequence. Then, for step size* $\eta_t = \frac{3}{\alpha t}$*,*

$$\forall z_\star \in \mathcal{K}, \ \sum_{t=1}^{T} f_t(z_t) - f_t(z_\star) \le \frac{6L^2}{\alpha}\log(T+1) + \alpha D^2 + \frac{6}{\alpha}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2 - \frac{\alpha}{6}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2$$

For our setting, we shall need a strengthing of the above theorem to the conditionally strongly convex with memory setting of Condition E.2. But for the present sketch, the above proposition captures the essential elements of the regret bound: (1) logarithmic regret, (2) quadratic sensitivity to $\boldsymbol{\epsilon}_t$, yielding a dependence of $T\epsilon_G^2/\alpha$, and (3) negative regret relative to arbitrary comparators. With this observation in hand, we present our regret decomposition in Eq. (F.2), which is described in terms of a comparator $M_{\mathrm{apprx}} \in \mathcal{M}$, and restricted set $\mathcal{M}_0 = \mathcal{M}(R_{\mathcal{M}}/2, m_0) \subset \mathcal{M}$, where $m_0 = \lfloor \frac{m}{2} \rfloor - h$:

$$\begin{aligned}
\mathrm{Regret}_T(\Pi(\psi)) \le\ & \underbrace{\left(\sum_{t=1}^{m+2h+N} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})\right)}_{\text{burn-in loss}} \\
& + \underbrace{\left(\sum_{t=m+2h+N+1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \sum_{t=m+2h+N+1}^{T} F_t^{\mathrm{pred}}[\mathbf{M}_{t:t-h}]\right)}_{\text{algorithm truncation error}}, \\
& + \underbrace{\left(\sum_{t=m+2h+N+1}^{T} F_t^{\mathrm{pred}}[\mathbf{M}_{t:t-h}] - \sum_{t=m+2h+N+1}^{T} f_t^{\mathrm{pred}}(M_{\mathrm{apprx}})\right)}_{f^{\mathrm{pred}} \text{ policy regret}} \\
& + \underbrace{\sum_{t=N+m+2h+1}^{T} f_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - \inf_{M \in \mathcal{M}_0} \sum_{t=N+m+2h+1}^{T} f_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})}_{\widehat{\mathbf{y}}_t^{\mathrm{nat}} \text{ control approximation error}} \\
& + \underbrace{\left(\inf_{M \in \mathcal{M}_0} \sum_{t=N+m+2h+1}^{T} f_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}) - \inf_{M \in \mathcal{M}_0} \sum_{t=N+m+2h+1}^{T} \ell_t(\mathbf{y}_t^M, \mathbf{u}_t^M)\right)}_{\text{comparator truncation error}} \\
& + \underbrace{\inf_{M \in \mathcal{M}_0} \sum_{t=1}^{T} \ell_t(\mathbf{y}_t^M, \mathbf{u}_t^M) - \inf_{\pi \in \Pi(\psi)} \ell_t(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)}_{\text{policy approximation } := J(M) - J(\pi)}
\end{aligned} \tag{F.2}$$

The novelty in this regret decomposition are the "f^{pred} policy regret" and "$\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ control approximation error" terms, which are coupled by a common choice of comparator $M_{\mathrm{apprx}} \in \mathcal{M}$. The first is precisely the policy regret on the F_t^{pred}, f_t^{pred} sequence, which (as decribed above) we bound via viewing descent on $f_t(\cdot \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$ as a running OGD on the former sequence, corrupted with nonstochastic error.

The term "$\widehat{\mathbf{y}}_t^{\mathrm{nat}}$ control approximation error" arises from the fact that, even though f_t^{pred} describes (up to truncation) the true response of the system to the controls, it considers controls based on *estimates* of natures y's, and not on nature's y's themselves. To bound this term, we show that there exists a specific comparator M_{apprx} which competes with the best controller in in the restricted class $\mathcal{M}_0$ that access the *true* natures y's. Proposition F.8 constructs a controller M_{apprx} which builds in a correction for the discrepancy between $\widehat{\mathbf{y}}^{\mathrm{nat}}$ and $\mathbf{y}^{\mathrm{nat}}$. We show that this controller satisfies for any $c > 0$,:

$$\text{“}\widehat{\mathbf{y}}_t^{\mathrm{nat}} \text{ control approximation error”} \lessapprox \frac{T\epsilon_G^2}{c} + c\sum_t \|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2. \tag{F.3}$$

A proof sketch is given in Appendix F.1.2, which highlights how we use that $\mathcal{M}$ *overparametrizes* $\mathcal{M}_0$. Unlike the coarse argument in the weakly convex case, the first term Eq. (F.3) has the desired quadratic sensitivity to ϵ_G^2. However, the second term is a movement cost which may scale linearly in T in the worst case.

Surprisingly, the proof of Eq. (F.3) *does not require strong convexity*. However, in the presence of strong convexity, we can cancel the movement cost term with the negative-regret term from the f^{pred} policy regret. As decribed above, the f^{pred} policy regret can bounded using a strengthening of Proposition F.3, to

$$\text{“}f^{\mathrm{pred}} \text{ policy regret”} \lessapprox \frac{1}{\alpha}\log T + \frac{T\epsilon_G^2}{\alpha} - \Omega(\alpha)\sum_t \|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2,$$

for appropriate strong convexity parameter α. By taking c to be a sufficiently small multiple of α,

$$\text{“}\widehat{\mathbf{y}}_t^{\mathrm{nat}} \text{ control approximation error”} + \text{“}f^{\mathrm{pred}} \text{ policy regret”} \lessapprox \frac{1}{\alpha}\log T + \frac{T\epsilon_G^2}{\alpha}.$$

In light of Eq. (F.1), we obtain the desired regret bound by setting $N = \sqrt{T/\alpha}$.

F.1. Rigorous Proof of Theorem 5

F.1.1. f^{pred}-POLICY REGRET

We begin with by stating our general result for conditionally-strongly convex gradient descent with erroneous gradients. Our setup is as follows:

Condition F.1 *We suppose that $z_{t+1} = \Pi_{\mathcal{K}}(z_t - \eta \boldsymbol{g}_t)$, where $\boldsymbol{g}_t = \nabla f_t(z_t) + \boldsymbol{\epsilon}_t$. We further assume that the gradient descent iterates applied for $t \geq t_0$ for some $t_0 \leq k$, with $z_0 = z_1 = \cdots = z_{t_0} \in \mathcal{K}$. We assume that $\|\boldsymbol{g}_t\|_2 \leq L_{\mathbf{g}}$, and $\mathrm{Diam}(\mathcal{K}) \leq D$.*

The following theorem is proven in Appendix K.

Theorem 9 *Consider the setting of Condition E.2 and F.1, with $k \geq 1$. Then with step size $\eta_t = \frac{3}{\alpha t}$, the following bound holds with probability $1 - \delta$ for all comparators $z_\star \in \mathcal{K}$ simultaenously:*

$$\begin{aligned}&\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) - \left(\frac{6}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2 - \frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2\right) \\ &\quad\lesssim \alpha k D^2 + \frac{(kL_{\mathrm{f}} + h^2 L_{\mathrm{c}})L_{\mathbf{g}} + kdL_{\mathrm{f}}^2 + k\beta L_{\mathbf{g}}}{\alpha}\log(T) + \frac{kL_{\mathrm{f}}^2}{\alpha}\log\left(\frac{1 + \log(e + \alpha D^2)}{\delta}\right)\end{aligned}$$

Observe that when $\boldsymbol{\epsilon}_t = 0$, we can take $L_{\mathbf{g}} = L_{\mathrm{f}}$ and discard the second on third terms on the first line, yielding Theorem 8. Let us now specialize the above to bound the f^{pred}-policy regret. First, we verify appropriate smoothness, strong convexity and Lipschitz condiitons; the following three lemmas in this section are all proven in Appendix I.2.3.

Lemma F.4 *Under Condition D.1, $f_t^{\mathrm{pred}}(M)$ are $4\beta_f$-smooth, for β_f as in Lemma E.3.*

Lemma F.5 *Let α_f as in Proposition E.2 and suppose that*

$$\epsilon_G \le \frac{1}{9R_{\mathrm{nat}}R_{\mathcal{M}}R_{G_\star}}\sqrt{\frac{\alpha_f}{m\alpha_{\mathrm{loss}}}}$$

Then under Condition D.1, the losses $f_{t;k}^{\mathrm{pred}}(M)$ are $\alpha_f/4$ strongly convex.

Lemma F.6 (Lipschitzness: Unknown & Strongly Convex) *Recall the Lipschitz constant L_f from Lemma C.5. Then under Condition D.1, $f_t^{\mathrm{pred}}(M)$ is $4L_f$-Lipschitz, $f_t^{\mathrm{pred}}[M_{t:t-h}]$ is $4L_f$ coordinate Lipschitz. Moreover, $\max_{M\in\mathcal{M}} \|\nabla \widehat{f}_t(M;\widehat{G},\widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})\|_2 \le 4L_f$.*

Specializing the above theorem to our setting, we obtain the following:

Lemma F.7 (Strongly Convex Policy Regret: Unknown System) *For the step size choose $\eta_t = \frac{12}{\alpha t}$, the following bound holds with probability $1-\delta$:*

$$(f^{\mathrm{pred}}\text{-policy regret}) + \frac{\alpha}{48}\sum_{t=N+h+m}^{T}\|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2 \lesssim (\text{Eq. (E.1)}) + \frac{TC_{\mathrm{approx}}^2\epsilon_G(N,\delta)^2}{\alpha}$$

Proof In our setting, $z_t \leftarrow \mathbf{M}_t$, $z_\star \leftarrow M_{\mathrm{apprx}}$, $\boldsymbol{\epsilon}_t$, and $\boldsymbol{\epsilon}_t \leftarrow \nabla f(M \mid \widehat{G},\, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}) - \nabla f_t^{\mathrm{pred}}(M)$. Moreover, we can can bound the smoothness $\beta \lesssim \beta_f$, the strong convexity $\alpha \gtrsim \alpha_f$, and all Lipschitz constants $\lesssim L_f$, where L_f was as in Lemma C.5. Using the same diameter bound as in that lemma, we see that the term on the right hand side of Theorem 9 can be bounded as in Eq. (E.1)), up to constant factors. Moreover, in light of Lemma F.2, we can bound the term $\frac{6}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2$ from Theorem 9 by $\lesssim \frac{1}{\alpha}T(C_{\mathrm{approx}}\epsilon_G)^2$. This concludes the proof. Lastly, we lower bound $-\frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2$ by $-\frac{\alpha}{48}\sum_{t=N+h+m}^{T}\|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2$. ■

F.1.2. $\widehat{\mathbf{y}}^{\mathrm{nat}}$-COMPARATOR APPROXIMATION ERROR

We prove the following theorem in Appendix I.3:

Proposition F.8 *Let $\mathcal{M}_0 := \mathcal{M}(m_0, R_{\mathcal{M}}/2)$, suppose that $m \ge 2m_0 - 1 + h$, $\psi_{G_\star}(h+1) \le R_{G_\star}/T$, and that Condition D.1 holds. Them there exists a universal constant $C > 0$ such that, for all $\tau > 0$,*

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 36m^2R_{G_\star}^4R_{\mathrm{nat}}^4R_{\mathcal{M}}^3(m + T\epsilon_G^2)\max\{L, L^2/\tau\}$$
$$+\,\tau\sum_{t=N+m+h+1}^{T}\|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2\,.$$

Proof [Proof Sketch] Let $M_\star$ denote the optimal $M \in \mathcal{M}_0$ for the loss sequence defined in terms of the true $\mathbf{y}^{\mathrm{nat}}$ and $G_\star$. First, consider what happens when the learner selects the controller $\mathbf{M}_t = M_\star$ for each $t \in [T]$. By expanding appropriate terms, one can show that (up to truncation terms), the contoller $M_\star$ operating on $\mathbf{y}^{\mathrm{nat}}_{1:t}$ produces the same inputs as the controller $M_{\mathrm{apprx}} = M_\star + M_\star * (\widehat{G} - G_\star) * M_\star$ operating on $\widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}$, where '$*$' denotes the convolution operator. Since $\mathcal{M}$ overparametrizes $\mathcal{M}_0$, we ensure that $M_{\mathrm{apprx}} \in \mathcal{M}$.

Realistically, the learner does not play $\mathbf{M}_t = M_\star$ at each round. However, we can show that the quality in the approximation for playing $\mathbf{M}_t$ instead of $M_\star$ degrades as $\sum_t \epsilon_G \cdot \|\mathbf{M}_t - M_\star\|_{\mathrm{F}}$. By construction, $\|M_{\mathrm{apprx}} - M_\star\|_{\mathrm{F}}$ scales as ϵ_G, so the triangle inequality gives $\epsilon_G \cdot \|\mathbf{M}_t - M_\star\|_{\mathrm{F}} \lessapprox \epsilon_G^2 + \epsilon_G \|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}$. By the elementary inequality $ab \le a^2/\tau + \tau b^2$, we find that the total penalty for the movement cost scales as $\sum_t \epsilon_G^2 + \epsilon_G^2/\tau + \tau\|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2 = T\epsilon_G^2(1 + 1/\tau) + \tau \sum_t \|\mathbf{M}_t - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2$; this argument gives rise quadratic dependence on ϵ_G^2, at the expense of the movement cost penalty. ∎

F.1.3. CONCLUDING THE PROOF OF THEOREM 5

We assume that N and T satisfy, for an appropriately large universal constant c',

$$N = mh^2 C_\delta R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} \sqrt{mT(1 + \frac{L}{\alpha} + \frac{\beta_{\mathrm{loss}}^2}{L\alpha})} \quad \text{and} \quad T \ge c' h^4 C_\delta^6 R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2, \tag{F.4}$$

Since we also have $\psi_{G_\star}(h+1) \le 1/10T$, our choice of N ensures Condition D.1 holds with probability $1 - \delta - N^{-\log^2 N} = 1 - \delta - T^{-\Omega(\log^2 T)}$. Combining Lemma F.7 and Proposition F.8 with $\tau = \frac{\alpha}{48}$, we can cancel the movement cost in the second bound with the negative regret in the first:

$$\begin{aligned}
&(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) + (f^{\mathrm{pred}}\text{-policy regret}) \\
&\lesssim (\text{Eq. (E.1)}) + (T\epsilon_G^2 + m)\left(m^2 R_{G_\star}^4 R_{\mathcal{M}}^4 R_{\mathrm{nat}}^4 \max\{L, \tfrac{L^2}{\alpha}\} + \tfrac{C_{\mathrm{approx}}^2}{\alpha}\right) \\
&\lesssim (\text{Eq. (E.1)}) + (\frac{TC_\delta^2}{N} + m) m^2 h^4 R_{G_\star}^4 R_{\mathcal{M}}^4 R_{\mathrm{nat}}^4 \,(L + \tfrac{(L+\beta_{\mathrm{loss}})^2}{\alpha}) \\
&\lesssim (\text{Eq. (E.1)}) + \frac{TC_\delta^2}{N} m^2 h^4 R_{G_\star}^4 R_{\mathcal{M}}^4 R_{\mathrm{nat}}^4 \,(L + \tfrac{(L+\beta_{\mathrm{loss}})^2}{\alpha}) \,.
\end{aligned}$$

where in the second line we recall from Lemma F.2 the bound $C_{\mathrm{approx}} \lesssim \sqrt{m} R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}^2 (\beta_{\mathrm{loss}} + L)$, and use $\epsilon_G(N, \delta) = C_{\mathrm{est}} \frac{h^2 R_{\mathrm{nat}}}{\sqrt{N}} C_\delta$ for $C_\delta := \sqrt{d_{\max} + \log\frac{1}{\delta} + \log(1 + R_{\mathrm{nat}})}$. In the third line, we use the assumption that $T \ge m^2$ from the Theorem.

From Lemma D.3, we can bound

$$(\text{estimation \& burn-in loss}) \le 4L(m + 2h + N) R_{G_\star}^2 \overline{R}_{\mathbf{u}}^2 \lesssim LN R_{G_\star}^2 \overline{R}_{\mathbf{u}}^2,$$

where $\overline{R}_{\mathbf{u}}$ as in Lemma D.1. By assumption $\sqrt{d_u + \log(1/\delta)} \le R_{\mathcal{M}} R_{\mathrm{nat}}$, we have $\overline{R}_{\mathbf{u}} \lesssim R_{\mathcal{M}} R_{\mathrm{nat}}$. Thus,

$$\begin{aligned}
&(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) + (f^{\mathrm{pred}}\text{-policy regret}) + (\text{estimation \& burn-in loss}) \\
&\lesssim (\text{Eq. (E.1)}) + L\left(\frac{TC_\delta^2}{N} m^2 h^4 R_{G_\star}^4 R_{\mathcal{M}}^4 R_{\mathrm{nat}}^4 \left(1 + \frac{L + \beta_{\mathrm{loss}}^2/L}{\alpha}\right) + N R_{\mathrm{nat}}^2 R_{\mathcal{M}}^2 R_{G_\star}^2\right).
\end{aligned}$$

For our choice of N, the above is at most

$$\lesssim \text{(Eq. (E.1))} + mh^2 L R_{G_\star}^3 R_{\mathcal{M}}^3 R_{\text{nat}}^3 C_\delta \sqrt{m(1 + \frac{L}{\alpha} + \frac{\beta_{\text{loss}}^2}{\alpha L})} \cdot \sqrt{T}.$$

Finally, similar arguments as those in previous bounds show that the truncation costs and policy approximation error are dominated by the above regret contribution under the assumptions $\psi(\lfloor \frac{m}{2} \rfloor - h) = \psi(m_0) \le R_{G_\star}/T$ and $R_\psi \le R_{\mathcal{M}}/2$ (note that the policy approximation is for the class $\mathcal{M}_0 = \mathcal{M}(m_0, R_{\mathcal{M}}/2)$), and $\psi_{G_\star}(h+1) \le R_{G_\star}/T$. □

Part II
Proof Details

Appendix G. Generalization to Stabilized Systems

In this section, we consider a generalization to settings where the system may not be internally stable; that is, where $\rho(A_\star) \geq 1$. Throughout, we assume the system is stabilizable and detectable: a linear system is said to be *stabilizable* if, in the absence of perturbations, there is a state-feedback controller that drives the state of the system asymptotically to zero; a *detectable* system is one where, in absence of perturbations, the state asymptotically tends to zero as long as the observations are all zeros. Relaxing the notions notions of controllability and observability respectively, these requirements do not impose any conditions on the stable modes of the system. In particular, we will employ these assumptions to guarantee the existence of a stabilizing observer-feedback control. See Anderson and Moore (2007) for an extensive discussion.

Our general recipe is as follows:

1. We assume access to a stabilizing *nominal controller* π_0. This induces an *dynamical* system with exogenous inputs, or LDC-EX (Definition G.4).

2. The LDC-EX produces a control ouput, $\boldsymbol{\eta}_t$. It's "natural" version $\boldsymbol{\eta}_t^{\mathrm{nat}}$ (Definition 3.1b) can be computed from input output data, and is what is used to parametrize the controller. This formulation is described in Appendix G.2.

3. In Appendix G.3, we formalally detail our controller parametrization for this framework, which we call DRC-EX, or Disturbance Response Control with Exogenous inputs. We then provide the generalization of our main algorithm, which we term DRC-GD-EX.

4. In Appendix G.4, we detail various examples of LDC-EX parametrizations.

   (a) We show that the stable setting can be recovered as a special case, as well as the static-feedback control, and control with nominal stabilizing controllers which are themselves internally-stable (Examples 3 to 5).

   (b) In general, unstable systems may require *internally-unstable* controllers to yield stable closed-loop dynamics. To this end, we describe an LDC-EX parametrization based on *exact observer feedback* (Example 6), which yields the classical Youla parametrization (Youla et al., 1976), and allows us extend our results to arbitrary stabilizable and detectable systems.

   (c) The exact Youla parametrization requires *full system knowledge* to construct an exact observer-feedback controller. To circumvent this, we demonstrate a convex parametrization based on approximate observer feedback, Example 7. This combines the classical Youla parametrization with a perspective based on Nature's η's, which affords convex parametrization without an exact observer-feedback controller.

5. Finally, in Appendix G.5, we demonstrate that all above examples of DRC-EXparametrizations are *fully expressive*, in that they can approximate the dynamics of any stabilizing linear dynamic controller to arbitrary degrees of accuracy (Theorems 10 and 1b).

G.1. Preliminaries

Going forward, it will be useful to slightly formalize our notion of Markov operators, which we shall interchangably refer to as *transfer operators*. We define

Definition G.1 (Markov Operator) *Let $\mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the set of Markov operators $G = (G^{[i]})_{i \geq 0}$ with $G^{[i]} \in \mathbb{R}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, such that $\|G\|_{\ell_1, \mathrm{op}} < \infty$. Given a system (A, B, C, D) with input dimension d_{in} and output dimension d_{o}, we let $G = \mathsf{Transfer}(A, B, C, D) \in \mathscr{G}_{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the system $G^{[0]} = D$ and $G^{[i]} = CA^{i-1}B$.*

Next, we state a computation of the joint evolution of a system under an LDC π:

Lemma G.2 *Let $(\mathbf{y}^\pi_t, \mathbf{x}^\pi_t)$ be the observation-state sequence produced on the execution of a LDC π on the LDS parameterized via $(A_\star, B_\star, C_\star)$. For a given sequence of disturbances $(\mathbf{e}_t, \mathbf{w}_t)$, the joint evolution of the system may be described as*

$$\begin{bmatrix} \mathbf{x}^\pi_{t+1} \\ \mathbf{v}^\pi_{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A_\star + B_\star D_\pi C_\star & B_\star C_\pi \\ B_\pi C_\star & A_\pi \end{bmatrix}}_{A_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{x}^\pi_t \\ \mathbf{v}^\pi_t \end{bmatrix} + \underbrace{\begin{bmatrix} I & B_\star D_\pi \\ 0 & B_\pi \end{bmatrix}}_{B_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{e}_t \end{bmatrix}, \tag{G.1}$$

$$\begin{bmatrix} \mathbf{y}^\pi_t \\ \mathbf{u}^\pi_t \end{bmatrix} = \underbrace{\begin{bmatrix} C_\star & 0 \\ D_\pi C_\star & C_\pi \end{bmatrix}}_{C_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{x}^\pi_t \\ \mathbf{v}^\pi_t \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & I \\ 0 & D_\pi \end{bmatrix}}_{D_{\pi,\mathrm{cl}}} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{e}_t \end{bmatrix}. \tag{G.2}$$

We refer to this dynamical system as the closed-loop *system in the main paper. Finally, we define*

$$C_{\pi,\mathrm{cl},u} = \begin{bmatrix} D_\pi C_\star & C_\pi \end{bmatrix}, \quad B_{\pi,\mathrm{cl},e} := \begin{bmatrix} B_\star D_\pi \\ B_\pi \end{bmatrix},$$

and let $G_{\pi,\mathrm{cl},e \to u} := \mathsf{Transfer}(A_{\pi,\mathrm{cl}}, B_{\pi,\mathrm{cl},e}, C_{\pi,\mathrm{cl},u}, D_\pi)$.

Proof The dynamical equations may be verified as an immediate consequence of Equation 2.1. ∎

Definition G.3 (Markov Operators for closed loop systems) *Given an LDC π, we define the systems,*

$$G^{[i]}_{\pi,\mathrm{cl},u \to (y,u)} := \mathbb{I}_{i=0} D_{\pi,\mathrm{cl}} + \mathbb{I}_{i>0} C_{\pi,\mathrm{cl}} A^{i-1}_{\pi,\mathrm{cl}} B_{\pi,\mathrm{cl,in}}, \quad G^{[i]}_{\pi,\mathrm{cl},u \to (y,u)} = \begin{bmatrix} G^{[i]}_{\pi,\mathrm{cl},u \to y} \\ G^{[i]}_{\pi,\mathrm{cl},u \to u} \end{bmatrix}, \quad B_{\pi,\mathrm{cl,in}} := \begin{bmatrix} B_\star \\ 0 \end{bmatrix}$$

where $(A_{\pi,\mathrm{cl}}, B_{\pi,\mathrm{cl}}, C_{\pi,\mathrm{cl}}, D_{\pi,\mathrm{cl}})$ are given by Lemma G.2. Furthernote, we define $\psi_{\pi,\mathrm{cl}} = \psi_{G_{\pi,\mathrm{cl}}}$ as the decay function of $G_{\pi,\mathrm{cl}}$, namely, $\psi_{\pi,\mathrm{cl}}(n) = \sum_{i \geq n} G_{\pi,\mathrm{cl}}$.

G.2. Linear Dynamic Controllers with Exogenous Inputs (LDC-EX)

In this section, let us set up a general stabilized parametrization. First, let us define the notion of an internal stabilizing controller:

Definition G.4 *An* linear dynamic controller with exogenous inputs *or* LDC-EX*, denoted by a policy* $\pi_0 = (A_{\pi_0}, B_{\pi_0}, C_{\pi_0}, D_{\pi_0})$*, as well as matrices* $B_{\pi_0,u}, C_{\pi_0,\eta}, D_{\pi_0,\eta}$*, which selects inputs* $\mathbf{u}_t^{\mathsf{alg}}$ *according to the following dynamics:*

$$\begin{aligned}
\mathring{\mathbf{s}}_{t+1} &= A_{\pi_0}\mathring{\mathbf{s}}_t + B_{\pi_0}\mathbf{y}_t^{\mathsf{alg}} + B_{\pi_0,u}\mathbf{u}_t^{\mathrm{ex}} \\
\mathring{\mathbf{u}}_t &= C_{\pi_0}\mathring{\mathbf{s}}_t + D_{\pi_0}\mathbf{y}_t^{\mathsf{alg}} \\
\boldsymbol{\eta}_t^{\mathsf{alg}} &= C_{\pi_0,\eta}\mathring{\mathbf{s}}_t + D_{\pi_0,\eta}\mathbf{y}_t^{\mathsf{alg}} \\
\mathbf{u}_t^{\mathsf{alg}} &= \mathbf{u}_t^{\mathrm{ex}} + \mathring{\mathbf{u}}_t
\end{aligned}$$

We refer to $\mathbf{u}_t^{\mathrm{ex}}$ *as the* exogenous input*,* $\mathring{\mathbf{u}}_t$ *as the* internal input*, and* $\mathbf{u}_t^{\mathsf{alg}}$ *as the* total input*. We refer to* $\boldsymbol{\eta}_t^{\mathsf{alg}} \in \mathbb{R}^{d_\eta}$ *as the control-output. The control policy* π_0 *is called the* nominal controller*. Lastly, we also define* $\mathbf{v}_t^{\mathsf{alg}} := (\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) \in \mathbb{R}^{d_y+d_u}$*, which we call the* total output*.*

Overloading notation, we will alternatively use π_0 to refer to the policy control $(A_{\pi_0}, B_{\pi_0}, C_{\pi_0}, D_{\pi_0})$, and to index objects associated with both π_0 and the additional matrices $B_{\pi_0,u}, C_{\pi_0,\eta}, D_{\pi_0,\eta}$.

In an LDC-EX, the endogenous input $\mathring{\mathbf{u}}_t$ is chosen so that, in the absence of inputs – i.e. $\mathbf{u}_t^{\mathrm{ex}} \equiv 0$ – the joint dynamics of the system remain stable. This allows us to generalize to settings where the dynamics of the nominal system may not be stable. For somewhat sophisticated reasons, in stabilized systems, one can be restricted by using Nature's y's for inputs. Instead, we will base our inputs on Nature's η's, defining $\boldsymbol{\eta}_t^{\mathrm{nat}}$ to be the control-output in the absence of exogenous inputs:

Definition 3.1b (Natures u's, y's, η's) *We define* $\mathbf{u}_t^{\mathrm{nat}}$*,* $\mathbf{y}_t^{\mathrm{nat}}$*, and* $\boldsymbol{\eta}_t^{\mathrm{nat}}$ *as the sequence that arises when, for all s,* $\mathbf{u}_t^{\mathrm{ex}} = 0$*. We set* $\mathbf{v}_t^{\mathrm{nat}} = (\mathbf{y}_t^{\mathrm{nat}}, \mathbf{u}_t^{\mathrm{nat}})$*. We note that* $\mathbf{u}_t^{\mathrm{nat}}, \mathbf{y}_t^{\mathrm{nat}}$ *coincide with* $\mathbf{u}_t^{\pi_0}, \mathbf{y}_t^{\pi_0}$*, whose dynamics are given by Lemma G.2 with the policy* $\pi \leftarrow \pi_0$*.*

Rather than requiring the nominal system to be stable, we will use controllers based on $\boldsymbol{\eta}_t^{\mathrm{nat}}$ (or estimates thereof). This requires only that the π_0 stabilize $A_\star$. Formally:

Assumption 1b (Stabilized Setting) *We assume that an* LDC-EX *is* stabilizing*; namely that* $A_{\pi_0,\mathrm{cl}}$ *is stable, where* $A_{\pi,\mathrm{cl}}$ *be defined in Lemma G.2.*

In order to define our DRC-EX parameterization, we need to introduce the following relevant transfer operators. We note that the 'A' matrix in each of the following Markov operators is $A_{\pi_0,\mathrm{cl}}$, which is stable by the above assumption, so each of the following operators are stable:

Definition 2.2b (Markov Operators for Strongly Stabilized System) *Fix an* LDC-EX *controller, and let* $A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl}}, C_{\pi_0,\mathrm{cl}}, D_{\pi_0,\mathrm{cl}}$ *be as in Lemma G.2, with* $\pi \leftarrow \pi_0$*. Further, define*

$$\begin{aligned}
C_{\pi_0,\mathrm{cl},\eta} &:= \begin{bmatrix} D_{\pi_0,\eta}C_\star & C_{\pi_0,\eta} \end{bmatrix}, \quad D_{\pi_0,\mathrm{cl},\eta} := \begin{bmatrix} 0 & D_{\pi_0,\eta} \end{bmatrix}, \\
B_{\pi_0,\mathrm{cl,ex}} &= \begin{bmatrix} B_\star \\ B_{\pi_0,\mathrm{ex}} \end{bmatrix}, \quad D_{\pi_0,\mathrm{cl,ex}} := \begin{bmatrix} 0 \\ I \end{bmatrix},
\end{aligned}$$

and the transfer functions $G_{\mathrm{ex}\to(y,u)} \in \mathscr{G}^{(d_y+d_u)\times d_u}$ *and* $G_{\mathrm{ex}\to\eta} \in \mathscr{G}^{d_\eta \times d_u}$ *via*

$$\begin{aligned}
G_{\mathrm{ex}\to(y,u)} &:= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl,ex}}, C_{\pi_0,\mathrm{cl}}, D_{\pi_0,\mathrm{cl,ex}}) \\
G_{\mathrm{ex}\to\eta} &:= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl,ex}}, C_{\pi_0,\mathrm{cl},\eta}, 0) \\
G_{(w,e)\to\eta} &:= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl}}, C_{\pi_0,\mathrm{cl},\eta}, D_{\pi_0,\mathrm{cl},\eta}) \\
G_{(w,e)\to(y,u)} &= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl}}, C_{\pi_0,\mathrm{cl}}, D_{\pi_0,\mathrm{cl}}),
\end{aligned}$$

and will decompose $G_{\mathrm{ex}\to(y,u)} = \begin{bmatrix} G_{\mathrm{ex}\to y} \\ G_{\mathrm{ex}\to u} \end{bmatrix}$ *for appropriate* $G_{\mathrm{ex}\to y} \in \mathscr{G}^{d_y\times d_u}, G_{\mathrm{ex}\to u} \in \mathscr{G}^{d_u\times d_u}$.

We can now write a "Nature's y's" representation of all relevant quantities:

Lemma 3.2b *We have the following identities for* $(\mathbf{y}^{\mathsf{alg}}_t, \mathring{\mathbf{u}}^{\mathsf{alg}}_t, \mathbf{u}^{\mathsf{alg}}_t)$*:*

$$\begin{bmatrix} \mathbf{y}^{\mathsf{alg}}_t \\ \mathring{\mathbf{u}}^{\mathsf{alg}}_t \end{bmatrix} = \begin{bmatrix} \mathbf{y}^{\mathrm{nat}}_t \\ \mathbf{u}^{\mathrm{nat}}_t \end{bmatrix} + \sum_{i=1}^{t-1} G^{[i]}_{\mathrm{ex}\to(y,u)} \mathbf{u}^{\mathrm{ex}}_{t-i}$$

Moreover, we have the following identity for $\boldsymbol{\eta}^{\mathsf{alg}}_t$*:*

$$\boldsymbol{\eta}^{\mathsf{alg}}_t = \boldsymbol{\eta}^{\mathrm{nat}}_t + \sum_{i=1}^{t} G^{[i]}_{\mathrm{ex}\to\eta} \mathbf{u}^{\mathrm{ex}}_{t-i}$$

Finally, we can express Nature's y's, u's and η*'s as functions of the noise via*

$$\mathbf{v}^{\mathrm{nat}}_t := \begin{bmatrix} \mathbf{y}^{\mathrm{nat}}_t \\ \mathbf{u}^{\mathrm{nat}}_t \end{bmatrix} = \sum_{i=1}^{t} G^{[i]}_{(w,e)\to(y,u)} \begin{bmatrix} \mathbf{w}_{t-i} \\ \mathbf{e}_{t-i} \end{bmatrix}, \quad \boldsymbol{\eta}^{\mathrm{nat}}_t = \sum_{i=1}^{t} G^{[i]}_{(w,e)\to\eta} \begin{bmatrix} \mathbf{w}_{t-i} \\ \mathbf{e}_{t-i} \end{bmatrix}$$

The proof of the above lemma is a consequence of computation augmenting that of Lemma G.2 computation, whose proof we omit in the interest of brevity. We now state the relevant generalization of Assumption 3, which by the above lemma and a similar computation for the mapping of $(w, e) \to (y, u)$, holds for any bounded noise sequence:

Assumption 3b (Bounded Nature's y, u, η) *We assume that that* $\mathbf{w}_t$ *and* $\mathbf{e}_t$ *are chosen by an oblivious adversary, and that* $\|\mathbf{v}^{\mathrm{nat}}_t\|_2 := \|(\mathbf{y}^{\mathrm{nat}}_t, \mathbf{u}^{\mathrm{nat}}_t)\|_2 \le R_{\mathrm{nat}}$ *and* $\|\boldsymbol{\eta}^{\mathrm{nat}}_t\|_2 \le R_{\mathrm{nat}}$ *for all* t. [13]

G.3. DRC-EX Parametrization and Algorithm

Let us now describe the DRC-EX parametrization. Throughout, we will supress dependence on π_0.

Definition 3.3b (Disturbance Response Controller with Exogenous Inputs) *A Disturbance Response Controller with Exogenous Inputs (DRC-EX), parameterized by a* m*-length sequence of matrices* $M = (M^{[i]})_{i=0}^{m-1}$*, chooses the control input as* $\mathbf{u}^{\mathrm{ex}}_t = \sum_{s=0}^{m-1} M^{[s]} \boldsymbol{\eta}^{\mathrm{nat}}_{t-s}$*. For a fixed* M*, we denote the resultant inputs, ouputs, and control-outputs* $(\mathbf{y}^M_t, \mathbf{u}^M_t, \boldsymbol{\eta}^M_t)$*, and let* $J_T(M)$ *the loss functional. We also set* $\mathbf{v}^M = (\mathbf{y}^M, \mathbf{u}^M)$.

Parallel to the stable setting, if $G_{\mathrm{ex}\to\eta}$ is known exactly, one can exactly recover $\boldsymbol{\eta}^{\mathrm{nat}}_{t-s}$ via Lemma 3.2b. When unknown, we can approximately recover $\boldsymbol{\eta}^{\mathrm{nat}}_{t-1}$ using an estimate $\widehat{G}_{\mathrm{ex}\to\eta}$, namely (Line 7)

$$\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t := \boldsymbol{\eta}^{\mathsf{alg}}_t - \sum_{i=1}^{t} \widehat{G}^{[i]}_{\mathrm{ex}\to\eta} \mathbf{u}^{\mathrm{ex}}_{t-i}$$

13. Note that, if the system is stable and perturbations bounded, that $\mathbf{y}^{\mathrm{nat}}_t$ will be bounded for all t.

Thus, we propose to use the estimates $\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$ to define our controller. Moreover, to estimate the consequence of a given input, we also need to estimate the $\mathbf{v}_t = (\mathbf{y}_t^{\mathrm{nat}}, \mathbf{u}_t^{\mathrm{nat}})$ so can ascertan the baseline in the absence of exogenous input. Thus we take

$$\widehat{\mathbf{v}}_t^{\mathrm{nat}} := \mathbf{v}_t^{\mathsf{alg}} - \sum_{i=1}^{t} \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]} \mathbf{u}_{t-i}^{\mathrm{ex}}.$$

The above definitions give rise to the following counterfactual dynamics and losses:

Definition 4.1b (Counterfactual Costs and Dynamics, Stabilized Systems) *Let $\mathbf{v}_t^{\mathrm{nat}} = (\mathbf{y}_t^{\mathrm{nat}}, \mathbf{u}_t^{\mathrm{nat}})$, and $\widehat{\mathbf{v}}_t^{\mathrm{nat}}$ denote estimates of $\mathbf{v}_t^{\mathrm{nat}}$. We define the counterfactual costs and dynamics*

$$\begin{aligned}
&\mathbf{u}_t^{\mathrm{ex}}\left(M_t \mid \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}\right) := \sum_{i=0}^{m-1} M_t^{[i]} \cdot \widehat{\boldsymbol{\eta}}_{t-i}^{\mathrm{nat}} \\
&\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right] := \widehat{\mathbf{v}}_t^{\mathrm{nat}} + \sum_{i=1}^{h} \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]} \cdot \mathbf{u}_{t-i}^{\mathrm{ex}}\left(M_{t-i} \mid \widehat{\boldsymbol{\eta}}_{1:t-i}^{\mathrm{nat}}\right), \\
&F_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right] := \ell_t\left(\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right]\right)
\end{aligned}$$

Overloading notation, we let $\mathbf{v}_t(M_t \mid \cdot) := \mathbf{v}_t[M_t, \dots, M_t \mid \cdot]$ denote the unary (single M_t) specialization of $\mathbf{v}_t$, and lower case $f_t\left(M|\cdot\right) = F_t\left[M, \dots, M|\cdot\right]$ the specialization of F_t. Throughout, we use paranthesis for unary functions of M_t, and brackets for functions of $M_{t:t-h}$.

The gradient feedback controller (Algorithm 1) and estimation procedure (Algorithm 2), and DRC-GD algorithm for unknown algorithm (Algorithm 3) are modified in algorithms Algorithms 4 to 6, respectively.

Algorithm 4: Disturbance Response Controller via Gradient Descent, with Exogenous Inputs (DRC-GD-EX)

Input: Stepsize η_t, radius R, memory m, Markov operators $\widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{G}_{\mathrm{ex}\to\eta}$, rollout h.
Define $\mathcal{M} = \mathcal{M}(m, R) = \{M = (M^{[i]})_{i=0}^{m-1} : \|M\|_{\ell_1,\mathrm{op}} \leq R\}$.
Initialize $\mathbf{M}_1 \in \mathcal{M}$ arbitrarily.
for $t = 1, \dots, T$ **do**
  Observe $\mathbf{v}_t^{\mathsf{alg}} = (\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})$
  Update $\mathring{\mathbf{s}}_t, \mathring{\mathbf{u}}_t, \boldsymbol{\eta}_t^{\mathsf{alg}}$ as in Definition G.4
  Estimate $\widehat{\mathbf{v}}_t^{\mathrm{nat}}$ and $\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$ via

$$\widehat{\mathbf{v}}_t^{\mathrm{nat}} := \mathbf{v}_t^{\mathsf{alg}} - \sum_{i=1}^{t-1} \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]} \mathbf{u}_{t-i}^{\mathrm{ex}}, \quad \widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}} := \boldsymbol{\eta}_t^{\mathsf{alg}} - \sum_{i=1}^{t} \widehat{G}_{\mathrm{ex}\to\eta}^{[i]} \mathbf{u}_{t-i}^{\mathrm{ex}}$$

  Choose the exogenous control input as

$$\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}} \leftarrow \mathbf{u}_t^{\mathrm{ex}}(\mathbf{M}_t \mid \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}) = \sum_{i=0}^{m-1} \mathbf{M}_t^{[i]} \, \widehat{\boldsymbol{\eta}}_{t-i}^{\mathrm{nat}}.$$

  Play total input $\mathbf{u}_t^{\mathsf{alg}} = \mathbf{u}_t^{\mathrm{ex},\mathsf{alg}} + \mathring{\mathbf{u}}_t$
  Observe the loss function ℓ_t and suffer a loss of $\ell_t(y_t, u_t)$.
  Recalling $f_t(\cdot|\cdot)$ from Definition 4.1b,update the disturbance feedback controller as

$$\mathbf{M}_{t+1} = \Pi_{\mathcal{M}}\left(\mathbf{M}_t - \eta_t \partial f_t\left(\mathbf{M}_t \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right)\right)$$

G.4. Examples of LDC's with exogenous inputs

Let us now provide examples of possible LDC's with exogenous inputs π_0 which can be used. The first three examples (Examples 3 to 5) are only pertain to a subset of dynamical systems - namely those that are (a) internally stable, (b) stabilizable by static feedback, or (c) stabilized by an internally stable controller.

In general, the are certain pathological which are unstable, and cannot be stabilized by static feedback or internally stable controller (see e.g. Halevi (1994)). For general systems, Appendix G.4.1 describes an LDC-ex formulation based on powerful parametrization known as the "Youla parametrization" Youla et al. (1976), also attributed to Kučera (1975), which uses an *observer-feedback* controller to provide an internally stabilizing, convex controller parametrization for *arbitrary* systems.

Unfortunately, realizing an exact Youla parametrization requires exact system knowledge. To adress this, we consider introduce an LDC-ex parametrization based on *approximate youla parametrization*. Under mild conditions, we shall show that these parametrizations have the same expressive power as the exact Youla parametrization, despite allowing for inexact system knowledge.

Example 3 (Stable System) *The internally stable system case (Assumption 1) corresponds to the setting where $\rho(A_\star) < 1$. Hence, we can $A_{\pi_0}, B_{\pi_0}, C_{\pi_0}, D_{\pi_0}$ to be identically zero, $\boldsymbol{\eta}_t^{\mathsf{alg}} = \mathbf{y}_t^{\mathsf{alg}}$, corresponding to $C_{\pi_0,\eta} = 0$ and $D_{\pi_0,\eta} = I$, . This satisfies Assumption 1b because $A_{\pi_0,\mathrm{cl}} = A_\star$, which is stable by assumption. This identical to the stable system setting in the body of the paper.*

Algorithm 5: Estimation of Unknown System

Input: Number of samples N, system length h.

Initialize

$$\widehat{G}^{[0]}_{\mathrm{ex}\to(y,u)} = \begin{bmatrix} 0_{d_y \times d_u} \\ I_{d_u} \end{bmatrix}, \quad \widehat{G}^{[i]}_{\mathrm{ex}\to(y,u)} = 0_{(d_y+d_u)\times d_u}, \quad \widehat{G}^{[i]}_{\mathrm{ex}\to\eta} = 0_{d_\eta \times d_u}, \quad i > 0$$

for $t = 1, 2, \dots, N$ **do**

  Play $\mathbf{u}^{\mathrm{ex,alg}}_t \sim \mathcal{N}(0, I_{d_u})$, recieve $\mathbf{v}^{\mathsf{alg}}_t = (\mathbf{u}^{\mathsf{alg}}_t, \mathbf{y}^{\mathsf{alg}}_t)$ and $\boldsymbol{\eta}^{\mathsf{alg}}_t$

Estimate $\widehat{G}^{[1:h]}_{u\to y}$ and $\widehat{G}^{[1:h]}_{u\to u}$ via least squares:

$$\widehat{G}^{[1:h]}_{\mathrm{ex}\to(y,u)} \leftarrow \arg\min_{G^{[1:h]}} \sum_{t=h+1}^{N} \|\mathbf{v}^{\mathsf{alg}}_t - \sum_{i=1}^{h} G^{[i]} \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\|_2^2$$

$$\widehat{G}^{[0:h]}_{\mathrm{ex}\to\eta} \leftarrow \arg\min_{G^{[0:h]}} \sum_{t=h+1}^{N} \|\boldsymbol{\eta}^{\mathsf{alg}}_t - \sum_{i=0}^{h} G^{[i]} \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\|_2^2$$

Return $\widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{G}_{\mathrm{ex}\to\eta}$.

Algorithm 6: DRC-GD-EX for Unknown System

Input: Stepsizes $(\eta_t)_{t\geq 1}$, radius $R_{\mathcal{M}}$, memory m, rollout h, Exploration length N,

Run the estimation procedure (Algorithm 2) for N steps with system length h to estimate $\widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{G}_{\mathrm{ex}\to\eta}$

Run the regret minimizing algorithm (Algorithm 1) for $T - N$ remaining steps with estimated Markov operators $\widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{G}_{\mathrm{ex}\to\eta}$, stepsizes $(\eta_{t+N})_{t\geq 1}$, radius $R_{\mathcal{M}}$, memory m, rollout parameter h.

Example 4 (Static Feedback) *Under static feedback, we take $A_{\pi_0}, B_{\pi_0}, C_{\pi_0}$ to be zero, but set $D_{\pi_0} = K$ for a static-feedback matrix $K \in \mathbb{R}^{d_y \times d_x}$. Again, we set $\boldsymbol{\eta}^{\mathsf{alg}}_t = \mathbf{y}^{\mathsf{alg}}_t$, corresponding to $C_{\pi_0,\eta} = 0$ and $D_{\pi_0,\eta} = I$. From Lemma G.2, the closed-loop matrix $A_{\pi_0,\mathrm{cl}}$ is given by $A_\star + B_\star K C_\star$. Thus, we require K such that $\rho(A_\star + B_\star K C_\star) < 1$. For general partially observed systems, it may not be the case that such a K exists, even if the system is stabilizable (i.e. there exists a control policy π_0 which stabilizes it). However, for stabilizable fully observed systems, such a K is always guaranteed to exist, and can be obtained by solving the discrete algebraic Riccati equation, or DARE (Anderson and Moore, 2007). Observe that static feedback reduces to the stable-system setting when $K = 0$.*

Example 5 (Stabilizing Feedback) *More generally, we can select a stabilizing controller π_0 such that $A_{\pi_0}, B_{\pi_0}, C_{\pi_0}, D_{\pi_0}$ need not be zero, but both the internal controller dynamics, and the closed-loop dynamics are stable. That is, $\rho(A_{\pi_0}) < 1$ and $\rho(A_{\pi_0,\mathrm{cl}}) < 1$. Yet again, we set $\boldsymbol{\eta}^{\mathsf{alg}}_t = \mathbf{y}^{\mathsf{alg}}_t$, corresponding to $C_{\pi_0,\eta} = 0$ and $D_{\pi_0,\eta} = I$. Note that this strictly generalizes Examples 3 and 4:*

Static feedback is recovered by setting $A_{\pi_0}, B_{\pi_0}, C_{\pi_0} = 0$ *and* $D_{\pi_0} = K$*, and stable systems by setting* $D_{\pi_0} = 0$ *as well.*

G.4.1. EXACT YOULA LDC-EX

As described above, certain pathological systems may not admit any stabilizing controller π_0 satisfying Example 5, and thus no controllers satisfying either of the special cases Examples 3 and 4. However, all stabilizable system and detectable systems *do* admit stabilizing controllers of the following form:

Example 6 (Exact Observer Feedback) *Consider a stabilizable and detectable system, and fix matrices* L, F *that satisfy* $\rho(A_\star + B_\star F) < 1$ *and* $\rho(A_\star + LC_\star) < 1$*. Exact Observer Feedback with Exogenous inputs denotes the internal state* $\mathring{\mathbf{s}}_t$ *via* $\widetilde{\mathbf{x}}_t \in \mathbb{R}^{d_x}$*, and has the dynamics*

$$\widetilde{\mathbf{x}}_{t+1} = (A_\star + LC_\star)\widetilde{\mathbf{x}}_t - L\mathbf{y}_t + B_\star \mathbf{u}_t^{\mathsf{alg}}$$
$$\boldsymbol{\eta}_t^{\mathrm{nat}} = C_\star \widetilde{\mathbf{x}}_t - \mathbf{y}_t^{\mathsf{alg}}, \quad \mathbf{u}_t^{\mathsf{alg}} = \mathbf{u}_t^{\mathrm{ex}} + F\widetilde{\mathbf{x}}_t,$$

with $\widetilde{\mathbf{x}}_1 = 0$*. This yields an LDC-ex* $d_\eta = d_y$*, with* $A_{\pi_0} = (A_\star + LC_\star + B_\star F)$*,* $B_{\pi_0} = -L$*,* $C_{\pi_0} = F$*,* $D_{\pi_0} = 0$*,* $C_{\pi_0,\eta} = C_\star$*, and* $D_{\pi_0,\eta} = -I$*.*

Note that the optimal LQG controller is an observer-feedback controller. However, for this parametrization, we don't need to know this optimal LQG controller. Rather, *any* observer-feedback controller will suffice.

Lemma G.5 *Under Example 6, following identities hold:*

1. $G^{[i]}_{\mathrm{ex}\to\eta} = 0$ *for all* $i > 0$*. In other words,* $\boldsymbol{\eta}_t^{\mathsf{alg}} = \boldsymbol{\eta}_t^{\mathrm{nat}}$ *for all* t*, regardless of exogenous inputs.*

2. *We have the identity.*

$$G^{[i]}_{(w,e)\to\eta} = \mathbb{I}_{i=0}\begin{bmatrix} 0 & I_{d_y}\end{bmatrix} + \mathbb{I}_{i>0} C_\star (A_\star + LC_\star)^{i-1}\begin{bmatrix} I_{d_x} & F\end{bmatrix}.$$

3. *We have the identity*

$$G^{[i]}_{\mathrm{ex}\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix} 0 \\ I_{d_u}\end{bmatrix} + \mathbb{I}_{i>0}\begin{bmatrix} C_\star \\ F\end{bmatrix}(A_\star + B_\star F)^{i-1} B_\star.$$

4. *We have the identity*

$$G^{[i]}_{(w,e)\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix} 0 & I \\ 0 & 0\end{bmatrix} + \mathbb{I}_{i>0}\begin{bmatrix} I & 0 \\ 0 & -L\end{bmatrix}\begin{bmatrix} A_\star & B_\star F \\ -LC_\star & A_\star + BF + LC_\star\end{bmatrix}^{i-1}\begin{bmatrix} C_\star & 0 \\ 0 & F\end{bmatrix}$$

   Moreover, via a change of basis, we can write

$$G^{[i]}_{(w,e)\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix} 0 & I \\ 0 & 0\end{bmatrix} + \mathbb{I}_{i>0}\begin{bmatrix} I & 0 \\ -L & -L\end{bmatrix}\begin{bmatrix} A_\star + B_\star F & B_\star F \\ 0 & A_\star + LC_\star\end{bmatrix}^{i-1}\begin{bmatrix} C_\star & -F \\ 0 & F\end{bmatrix}$$

Proof The first four computations may be verified directly. Alternatively, Lemma G.6 suffices to establish this while substituting $(A_\star, B_\star, C_\star) = (\widehat{A}, \widehat{B}, \widehat{C})$. For the last claim, a change of basis conjugating the $A_{\pi_0,\mathrm{cl}}$ matrix by $T = \begin{bmatrix} I & 0 \\ I & I \end{bmatrix}$, via $T^{-1} A_{\pi_0,\mathrm{cl}} T$ suffices. ■

In particular, since $\rho(A_\star + B_\star F), \rho(A_\star + LC_\star) < 1$ hold by assumption due to stabilizability and detectability, all of the above systems are guaranteed to be stable.

G.4.2. APPROXIMATE YOULA LDC-EX (LDC-EX)

The previously suggested parameterization requires exact specification of the system parameters $(A_\star, B_\star, C_\star)$. However, for an unknown system, one can only hope to estimate parameters approximately. This section details the effects of executing a Youla controller with approximate estimates of the system parameters.

Example 7 (Approximate Youla LDC-Ex) *An Approximate Observer-Feedback controller when given parameter estimates $\widehat{A}, \widehat{B}, \widehat{C}$ and executed under the influence of exogenous inputs follows:*

$$\begin{aligned} \widehat{\mathbf{x}}_{t+1} &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - L\mathbf{y}_t + \widehat{B}\mathbf{u}_t \\ \widehat{\boldsymbol{\eta}}_t &= \widehat{C}\widehat{\mathbf{x}}_t - \mathbf{y}_t \\ \mathbf{u}_t &= \mathbf{u}_t^{\mathrm{ex}} + F\widehat{\mathbf{x}}_t. \end{aligned}$$

Note that $\widehat{\boldsymbol{\eta}}_t$ depends on the history of exogenous inputs $\mathbf{u}_t^{\mathrm{ex}}$. Still, we can give a closed form representation of the overall system dynamics, and the map from exogenous inputs to outputs/controls:

Lemma G.6 *Set $\boldsymbol{\delta}_t := \widehat{\mathbf{x}}_t - \mathbf{x}_t$ and $\Delta_{\mathrm{youl}} := \widehat{A} - A_\star + L(\widehat{C} - C_\star)$. Then, the dynamics induced by Example 7 satisfy that*

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \boldsymbol{\delta}_{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} A_\star + B_\star F & B_\star F \\ \Delta_{\mathrm{youl}} & \widehat{A} + L\widehat{C} \end{bmatrix}}_{:=A_{:,\mathrm{in}}} \begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\delta}_t \end{bmatrix} + \underbrace{\begin{bmatrix} B_\star \\ \widehat{B} - B_\star \end{bmatrix}}_{B_{:,\mathrm{in}}} \mathbf{u}_t^{\mathrm{ex}} + \begin{bmatrix} I & 0 \\ -I & -L \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{e}_t \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbf{y}_t \\ \widehat{\boldsymbol{\eta}}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} C_\star & 0 \\ \widehat{C} - C_\star & \widehat{C} \\ F & F \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\delta}_t \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \mathbf{u}_t^{\mathrm{ex}} + \begin{bmatrix} I \\ -I \\ 0 \end{bmatrix} \mathbf{e}_t.$$

Denoting by $G_{:,\mathrm{in}}$ the Markov operator describing the map from $\mathbf{u}_t^{\mathrm{ex}} \to (\mathbf{y}_t, \mathbf{u}_t)$, we then have the identity that

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t^{\mathrm{nat}} \\ \mathbf{u}_t^{\mathrm{nat}} \end{bmatrix} + \sum_{i=0}^{t-1} G_{:,\mathrm{in}}^{[i]} \mathbf{u}_{t-i}^{\mathrm{ex}}.$$

Proof Let's change variables.

$$\begin{aligned}
\mathbf{x}_{t+1} &= A_\star \mathbf{x}_t + B_\star F \widehat{\mathbf{x}}_t + B_\star \mathbf{u}_t^{\mathrm{ex}} + \mathbf{w}_t \\
&= (A_\star + B_\star F)\mathbf{x}_t + B_\star F \boldsymbol{\delta}_t + B_\star \mathbf{u}_t^{\mathrm{ex}} + \mathbf{w}_t \\
\widehat{\mathbf{x}}_{t+1} &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - L\mathbf{y}_t + \widehat{B}F\widehat{\mathbf{x}}_t + \widehat{B}\mathbf{u}_t^{\mathrm{ex}} \\
&= (\widehat{A} + \widehat{B}F)\widehat{\mathbf{x}}_t + L(\widehat{C}\widehat{\mathbf{x}} - C_\star \mathbf{x}_t) + \widehat{B}\mathbf{u}_t^{\mathrm{ex}} - L\mathbf{e}_t \\
\boldsymbol{\delta}_{t+1} = \widehat{\mathbf{x}}_{t+1} - \mathbf{x}_{t+1} &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - (A_\star + LC_\star)\mathbf{x}_t - L\mathbf{e}_t - \mathbf{w}_t + (\widehat{B} - B_\star)\mathbf{u}_t^{\mathrm{ex}} \\
&= (\widehat{A} + L\widehat{C})\boldsymbol{\delta}_t + (\widehat{A} - A_\star + L(\widehat{C} - C_\star))\mathbf{x}_t - L\mathbf{e}_t - \mathbf{w}_t + (\widehat{B} - B_\star)\mathbf{u}_t^{\mathrm{ex}}.
\end{aligned}$$

Once again, changing variables, we have

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}_t &= \widehat{C}\boldsymbol{\delta}_t + (\widehat{C} - C_\star)\mathbf{x}_t - \mathbf{e}_t \\
\mathbf{u}_t &= \mathbf{u}_t^{\mathrm{ex}} + F\boldsymbol{\delta}_t + F\mathbf{x}_t.
\end{aligned}$$

■

G.5. Expressivity of DRC-EX

In this section generalize the expressivity guarantee of Theorem 1 to our more general setting. To begin, let us define a notion of an operator which translates the dynamics under the nominal controller π_0 to target dynamics π:

Definition G.7 *Given a dynamical system π, we say that $G_{\pi_0 \to \pi}$ is a $\pi_0 \to \pi$* conversion operator *if the following under dynamics induced by any noise sequence $(\mathbf{w}_t, \mathbf{e}_t)$: If $(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ are the input-output sequence under π (Lemma G.2), then the sequence defined by*

$$\mathbf{u}_t^{\mathrm{ex}, \pi \to \pi_0} := \sum_{i=1}^{t} G_{\pi_0 \to \pi}^{[t-i]} \boldsymbol{\eta}_i^{\mathrm{nat}}$$

satisfies the following for all t:

$$\begin{bmatrix} \mathbf{y}_t^\pi \\ \mathbf{u}_t^\pi \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t^{\mathrm{nat}} \\ \mathbf{u}_t^{\mathrm{nat}} \end{bmatrix} + \sum_{i=1}^{t} G_{\mathrm{ex} \to (y,u)}^{[t-i]} \mathbf{u}_i^{\mathrm{ex}, \pi \to \pi_0}.$$

In other words, if one selects exogenous inputs $\mathbf{u}_t^{\mathrm{ex}, \pi \to \pi_0}$, then one recovers the dynamics of the controller π. Note that, it is enough to show that one recovers the dynamics of $\mathbf{u}_t^\pi$, since the inputs and noise to the system uniquely determine the dynamics of $\mathbf{y}_t$ via Eq. (1.1). With the above definition, we define our comparator class accordingly:

Definition 2.3b (Decay Functions & Policy Class) *Given an* LDC-EX *π_0, we define the comparator class $\Pi(\psi)$ as the set of all π for which there exists a $\pi_0 \to \pi$ conversion operator $G_{\pi_0 \to \pi}$ which decay dominated by ψ: that is, $\psi_{G_{\pi_0 \to \pi}}(n) \le \psi(n)$ for all n. Moreover, we define $R_{G_\star} := 1 \vee \|G_{\mathrm{ex} \to \eta}\|_{\ell_1, \mathrm{op}} \vee \|G_{\mathrm{ex} \to (y,u)}\|_{\ell_1, \mathrm{op}}$, and $\psi_{G_\star}(n) := \max\{\psi_{G_{\mathrm{ex} \to (y,u)}}(n), \psi_{G_{\mathrm{ex} \to \eta}}(n)\}$.*

With this definition in mind, Theorem 1b follows by direct analogy to Theorem 1:

Theorem 1b *Let $\Pi(\psi)$ be as in Definition 2.3b, $R_\psi = \psi(0)$, and let $J_T(M)$ of the strongly stabilized DRC controller (Definition 3.3b). Given a proper decay function ψ and $\pi \in \Pi(\psi)$, there exists an $M \in \mathcal{M}(m, R_\psi)$ such that*

$$J_T(M) - J_T(\pi) \le 2LTR_\psi R_{G_\star}^2 R_{\mathrm{nat}}^2\, \psi(m). \tag{G.3}$$

Proof The proof is analogous to the first part of the proof of Theorem 1, where the control approximation identity (Claim C.7) is built into the definition of $\psi(\cdot)$ by assumption. We omit the proof in the interest of simplicity. ■

While quite general, Theorem 1b guarantees competition with policies whose conversion operators ($\Pi(\psi)$ in Definition 2.3b) have reasonable decay, and unlike Theorem 1, it does not make this explicit. Thus it remains to show that this class $\Pi(\psi)$ is reasonable expressive.

In what follows, we will show that an analogoue holds in all of our examples. Let's make this formal:

Definition G.8 (Convolution of Markov Operator) *Let $G_1 \in \mathscr{G}^{d_1 \times d_0}$, and $G_2 \in \mathscr{G}^{d_0 \times d_2}$. We define $G = G_1 \odot G_2$ as the operator*

$$G^{[i]} = \sum_{j=0}^{i} G_1^{[j]} \cdot G_2^{[i-j]}.$$

Theorem 10 *For any policy π, the matrix $G_{\pi_0 \to \pi}$ can be represented as follows:*

1. *If the system is internally stable (Example 3), $G_{\pi_0 \to \pi} = G_{\pi,\mathrm{cl},e\to u}$, for π_0 which is identically zero.*

2. *If the system is stabilized by static feedback π_0 (Example 4), $G_{\pi_0\to\pi} = \bar{G}_{\pi_0\to\pi} \circ G_{\pi_0, y\to(y,u)}$ is as detailed in Proposition G.10 since a static controller is internally stable too, with $A_{\pi_0} = 0$. Furthermore, since $\rho(A_{\pi_0,\mathrm{cl}}) = \rho(A_\star + B_\star K C_\star) < 1$, both $\bar{G}_{\pi_0\to\pi}$ and $G_{\pi_0,y\to(y,u)}$ exhibit geometric decay.*

3. *If the system is stabilized by internally stable feedback (Example 5), $G_{\pi_0\to\pi} = \bar{G}_{\pi_0\to\pi} \circ G_{\pi_0,y\to(y,u)}$ is as detailed in Proposition G.10. In particular, both $\bar{G}_{\pi_0\to\pi}$ and $G_{\pi_0,y\to(y,u)}$ exhibit geometric decay as long as π is stabilizing, since $\rho(A_{\pi_0}) < 1$ and $\rho(A_{\pi_0,\mathrm{cl}}) < 1$*

4. *If the system is stabilized by exact observer feedback (Example 6), the $G_{\pi_0\to\pi}$ is as detailed in Proposition G.11. The latter exhibits geometric decay as long as π is stabilizing.*

5. *If the system is stabilized by inexact observer feedback (Example 7), then $G_{\pi_0\to\pi}$ is as Proposition G.12 details. In particular, it is a convolution of three Markov operators of stable systems, as long as $\max\{\rho(A_\star + B_\star F), \rho(\widehat{A} + \widehat{B}F), \rho(A_\star + LC_\star), \rho(\widehat{A} + L\widehat{C})\} < 1$.*

In each of the above cases, $G_{\pi_0\to\pi}$ is either the Markov operator of a stable system, or can be expressed by a convolution of two (Example 5) or three (Example 5) Markov operators of stable systems.

Specifically, we show that there we can represent $G_{\pi_0\to\pi}$ as an *convolution* of stable transfer operators. Since a convolution of operators with geometric decay itself has geometric decay, we find that we obtain the same expressive power as in the stable system case.

G.5.1. EXPRESSIVITY OF INTERNALLY STABLE FEEDBACK

Let us begin by defining a closed form expression for the $\pi_0 \to \pi$ operator that arises under internally stable feedback:

Definition G.9 (Internally Stable Dynamical System Conversion) *Given a nominal controller π_0 given by $(A_{\pi_0}, B_{\pi_0}, C_{\pi_0}, D_{\pi_0})$, and a target controller π given by $(A_\pi, B_\pi, C_\pi, D_\pi)$, and recalling the closed loop matrix $A_{\pi,\mathrm{cl}}$ from Lemma G.2, define the matrices $A_{\pi_0\to\pi}, B_{\pi_0\to\pi}, C_{\pi_0\to\pi}$ by*

$$A_{\pi_0\to\pi} := \left[\begin{array}{cc|c} A_{\pi,\mathrm{cl}} & & 0 \\ & & 0 \\ \hline B_{\pi_0}C_\star & 0 & A_{\pi_0} \end{array}\right], \quad B_{\pi_0\to\pi} := \begin{bmatrix} B_\star D_\pi & -B_\star \\ B_\pi & 0 \\ 0 & 0 \end{bmatrix},$$
$$C_{\pi_0\to\pi} := \begin{bmatrix} (D_\pi - D_{\pi_0})C_\star & C_\pi & -C_{\pi_0} \end{bmatrix}$$

and $D_{\pi_0\to\pi} = \begin{bmatrix} D_\pi & 0 \end{bmatrix}$ Define $\bar{G}_{\pi_0\to\pi} := \mathsf{Transfer}(A_{\pi_0\to\pi}, B_{\pi_0\to\pi}, C_{\pi_0\to\pi}, D_{\pi_0\to\pi})$, and define:

$$G^{[i]}_{\pi_0, y\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix} I \\ D_{\pi_0} \end{bmatrix} + \mathbb{I}_{i\geq 1}\begin{bmatrix} 0 \\ C_{\pi_0}A_{\pi_0}^{i-1}B_{\pi_0} \end{bmatrix}.$$

Finally, we define the $\pi_0 \to \pi$ coversion operator

$$G^{[i]}_{\pi_0\to\pi} = \sum_{j=0}^{i} \bar{G}^{[i-j]}_{\pi_0\to\pi} G^{[j]}_{\pi_0, y\to(y,u)}.$$

Proposition G.10 *For any stabilizing π and internally stable π_0, the Markov operator $G_{\pi_0\to\pi}$ defined in Definition G.9 is the convolution of two stable Markov operators, and is a $\pi_0 \to \pi$ conversion operator. That is, for all t, the exogenous inputs*

$$\mathbf{u}_t^{\mathrm{ex},\pi\to\pi_0} = \sum_{i=0}^{t-1} \bar{G}^{[i]}_{\pi_0\to\pi}\mathbf{y}^{\mathrm{nat}}_{t-i}.$$

produce the input-output pairs $(\mathbf{y}_t^\pi, \mathbf{u}_t^\pi)$ via

$$\begin{bmatrix} \mathbf{y}_t^\pi \\ \mathbf{u}_t^\pi \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t^{\mathrm{nat}} \\ \mathbf{u}_t^{\mathrm{nat}} \end{bmatrix} + \sum_{i=0}^{t-1} G^{[i]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex},\pi\to\pi_0}_{t-i}.$$

G.5.2. EXPRESSIVITY OF OBSERVER-FEEDBACK (YOULA PARAMETRIZATION)

Proposition G.11 *Define the matrices*

$$A_{\mathrm{yla},\pi} := A_{\pi,\mathrm{cl}}, \quad B_{\mathrm{yla},\pi} := \begin{bmatrix} B_\star D_\pi - L \\ B_\star \end{bmatrix}, \quad C_{\mathrm{yla},\pi} = \begin{bmatrix} D_\pi C_\star - F \end{bmatrix}, \quad D_{\mathrm{yla},\pi} = D_\pi$$

Then,

$$G_{\mathrm{yla},\pi} = \mathsf{Transfer}(A_{\mathrm{yla},\pi}, B_{\mathrm{yla},\pi}, C_{\mathrm{yla},\pi}, D_{\mathrm{yla},\pi})$$

is a $\pi_0 \to \pi$ conversion operator for the Youla LDC-Ex of Example 6. That is

$$\begin{bmatrix} \mathbf{y}_t^\pi \\ \mathbf{u}_t^\pi \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t^{\mathrm{nat}} \\ \mathbf{u}_t^{\mathrm{nat}} \end{bmatrix} + \sum_{i=0}^{t-1} G^{[i]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex},\pi\to\pi_0}_{t-i} \quad \textit{for} \quad \mathbf{u}_s^{\mathrm{ex},\pi\to\pi_0} = \sum_{j=0}^{s-1} G^{[i]}_{\mathrm{yla},\pi}\boldsymbol{\eta}^{\mathrm{nat}}_{s-i}.$$

The statement of the Youla parametrization is standard, though varies source-to-source. We use the expression in cite Megretski (2004, Theorem 10.1).

G.5.3. EXPRESSIVITY OF APPROXIMATIVE OBSERVER-FEEDBACK (DRC-YOULA PARAMETRIZATION)

Proposition G.12 *Let $G_{\mathrm{yla},\pi}$ be as in Proposition G.11, and define $\bar{G}_{\mathrm{yla},\pi} \in \mathscr{G}^{d_u+d_y \times d_y}$ via*

$$\bar{G}_{\mathrm{yla},\pi}^{[i]} = \begin{bmatrix} G_{\mathrm{yla},\pi}^{[i]} \\ I_{d_y} \cdot \mathbb{I}_{i=0} \end{bmatrix}.$$

Further, define the operators

$$G_{\widehat{:}\to\star} := \mathsf{Transfer}\left(\begin{bmatrix} A_\star + B_\star F & 0 \\ \widehat{B}F - LC_\star & \widehat{A} + L\widehat{C} \end{bmatrix}, \begin{bmatrix} B_\star & \widehat{B} \\ L & L \end{bmatrix}, \begin{bmatrix} F & -F \end{bmatrix}, \begin{bmatrix} I & 0 \end{bmatrix}\right)$$
$$G_{\star\to\widehat{:}} := \mathsf{Transfer}\left(\begin{bmatrix} A_\star + LC_\star & B_\star F - L\widehat{C} \\ 0 & \widehat{A} + \widehat{B}F \end{bmatrix}, \begin{bmatrix} L \\ L \end{bmatrix}, \begin{bmatrix} C_\star & -\widehat{C} \end{bmatrix}, I\right).$$

Then, the transfer operator $G_{\pi_0\to\pi} := G_{\widehat{:}\to\star} \odot \bar{G}_{\mathrm{yla},\pi} \odot G_{\star\to\widehat{:}}$ is a $\pi_0 \to \pi$ conversion operator for the Approximate Youla LCD-Ex of Example 7.

G.6. Proofs of Expressivity Guarantes

G.6.1. PROOF OF PROPOSITION G.10

Define

$$\mathbf{w}_{\pi_0,t} := \begin{bmatrix} I & B_\star D_{\pi_0} \\ 0 & B_{\pi_0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{e}_t \end{bmatrix}, \quad \mathbf{e}_{\pi_0,t} := \begin{bmatrix} I \\ D_\pi \end{bmatrix} \mathbf{e}_t$$

From the closed loop matrices $A_{\pi_0,\mathrm{cl}}, C_{\pi_0,\mathrm{cl}}$ described in Lemma G.2, the nomimal system with exogenous inputs is then described by the equations

$$\bar{\mathbf{x}}_{t+1} := \begin{bmatrix} \mathbf{x}_t \\ \mathring{\mathbf{s}}_t \end{bmatrix} = A_{\pi_0,\mathrm{cl}}\bar{\mathbf{x}}_t + \underbrace{\begin{bmatrix} B_\star \\ 0 \end{bmatrix}}_{B_{\pi_0,\mathrm{cl}}} \mathbf{u}_t^{\mathrm{ex}} + \mathbf{w}_{\pi_0,t}$$
$$\bar{\mathbf{y}}_t := \begin{bmatrix} \mathbf{y}_t \\ \mathring{\mathbf{u}}_t \end{bmatrix} = C_{\pi_0,\mathrm{cl}}\bar{\mathbf{x}}_t + \mathbf{e}_{\pi_0,t}. \tag{G.4}$$

We then put Eq. (G.4) in feedback with the following system via $\mathbf{u}_t^{\mathrm{ex}} = \mathbf{u}_t^{\Delta}$:

$$\mathbf{a}_{t+1}^{\Delta} = A_\pi \mathbf{a}_t^{\Delta} + \underbrace{\begin{bmatrix} B_\pi & 0 \end{bmatrix}}_{:=\bar{B}_\pi} \bar{\mathbf{y}}_t$$
$$\mathbf{u}_t^{\Delta} = C_\pi \mathbf{a}_t^{\Delta} + \underbrace{\begin{bmatrix} D_\pi & -I \end{bmatrix}}_{:=\bar{D}_\pi} \bar{\mathbf{y}}_t. \tag{G.5}$$

Then, the joint dynamics of Eqs. (G.5) and (G.4) are given by

$$\begin{bmatrix}\bar{\mathbf{x}}_{t+1}\\ \mathbf{a}^{\Delta}_{t+1}\end{bmatrix} = \underbrace{\begin{bmatrix} A_{\pi_0,\mathrm{cl}} + B_{\pi_0,\mathrm{cl}}\bar{D}_{\pi}C_{\pi_0,\mathrm{cl}} & B_{\pi_0,\mathrm{cl}}C_{\pi}\\ \bar{B}_{\pi}C_{\pi_0,\mathrm{cl}} & A_{\pi}.\end{bmatrix}}_{:=A_{\Delta\pi,\mathrm{cl}}}\begin{bmatrix}\bar{\mathbf{x}}_{t+1}\\ \mathbf{a}^{\Delta}_{t+1}\end{bmatrix} + \underbrace{\begin{bmatrix} I & B_{\pi_0,\mathrm{cl}}\bar{D}_{\pi}\\ 0 & \bar{B}_{\pi}\end{bmatrix}}_{B_{\Delta\pi,\mathrm{cl}}}\begin{bmatrix}\mathbf{w}_{\pi_0,t}\\ \mathbf{e}_{\pi_0,t}\end{bmatrix}$$

$$\begin{bmatrix}\bar{\mathbf{y}}_{t}\\ \mathbf{u}^{\Delta}_{t}\end{bmatrix} = \underbrace{\begin{bmatrix} C_{\pi_0,\mathrm{cl}} & 0\\ \bar{D}_{\pi}C_{\pi_0,\mathrm{cl}} & C_{\pi}\end{bmatrix}}_{C_{\Delta\pi,\mathrm{cl}}}\begin{bmatrix}\bar{\mathbf{x}}_{t}\\ \mathbf{a}^{\Delta}_{t}\end{bmatrix} + \begin{bmatrix} I\\ \bar{D}_{\pi}\end{bmatrix}\mathbf{e}_{\pi_0,t}. \tag{G.6}$$

First, we claim that, for all t, the system Eq. (G.6) yields inputs an outputs equivalent to $\mathbf{u}^{\pi}_t, \mathbf{u}^{\pi}_t$:

Lemma G.13 *Let $\bar{\mathbf{y}}_t = (\mathbf{y}_t, \mathring{\mathbf{u}}_t)$ and $\mathbf{u}^{\Delta}_t$ be as given by Eq. (G.6). Then, for $\mathbf{u}^{\Delta}_t$ defined above,*

$$\forall t, \mathbf{y}_t = \mathbf{y}^{\pi}_t \text{ and } \mathring{\mathbf{u}}_t + \mathbf{u}^{\Delta}_t = \mathbf{u}^{\pi}_t.$$

In particular, $\begin{bmatrix}\mathbf{y}^{\pi}_t\\ \mathbf{u}^{\pi}_t\end{bmatrix} = \sum_{i=1}^{t} G^{[i]}_{\pi_0,\mathrm{cl},u\to(y,u)}\mathbf{u}^{\Delta}_t$.

Proof Let us consider the update of the state $\mathbf{x}_t$: $\mathbf{x}_{t+1} = A_{\star}\mathbf{x}_{t+1} + B_{\star}(\mathring{\mathbf{u}}_t + \mathbf{u}^{\Delta}_t) + \mathbf{w}_t = A_{\star}\mathbf{x}_{t+1} + B_{\star}C_{\pi}\mathbf{a}^{\Delta}_t + B_{\star}D_{\pi}\mathbf{y}^{\mathrm{en}}_t + \mathbf{w}_t$. First, note that

$$\mathbf{u}_t = \mathring{\mathbf{u}}_t + \mathbf{u}^{\Delta}_t = D_{\pi}\mathbf{y}_t + C_{\pi}\mathbf{a}^{\Delta}_t = D_{\pi}C_{\star}\mathbf{x}_t + D\pi\mathbf{e}_t.$$

Thus,

$$\begin{aligned}\mathbf{x}_{t+1} &= A_{\star}\mathbf{x}_{t+1} + \mathring{\mathbf{u}}_t + \mathbf{u}^{\Delta}_t + \mathbf{w}_t\\ &= (A_{\star} + B_{\star}D_{\pi}C_{\star})\mathbf{x}_t + B_{\star}C_{\pi}\mathbf{a}^{\Delta}_t + B_{\star}D_{\pi}\mathbf{e}_t + \mathbf{w}_t\end{aligned}$$

Moreover, we have that

$$\begin{aligned}\mathbf{a}^{\Delta}_{t+1} = A_{\pi}\mathbf{a}^{\Delta}_t + B_{\pi}\mathbf{y}_t &= A_{\pi}\mathbf{a}^{\Delta}_t + B_{\pi}C_{\star}\mathbf{x}_t + B_{\pi}C_{\star}\mathbf{e}_t\\ &= A_{\pi}\mathbf{a}^{\Delta}_t + (B_{\pi}C_{\star}+)\mathbf{x}_t + B_{\pi}\mathbf{e}_t.\end{aligned}$$

Thus, $(\mathbf{x}_t, \mathbf{a}^{\Delta}_t)$ have the same evolution as $(\mathbf{x}^{\pi}_t, \mathbf{a}^{\pi}_t)$, where $\mathbf{a}^{\pi}_t$ is the internal state of the system when placed in feedback with π. Thus,

$$\begin{aligned}\mathbf{y}_t &= C_{\star}\mathbf{x}_t + \mathbf{e}_t = C_{\star}\mathbf{x}^{\pi}_t + \mathbf{e}_t = \mathbf{y}^{\pi}_t\\ \mathring{\mathbf{u}}_t + \mathbf{u}^{\Delta}_t &= D_{\pi}C_{\star}\mathbf{x}_t + C_{\pi}\mathbf{a}^{\Delta}_t + D_{\pi}C_{\star}\mathbf{e}_t = D_{\pi}C_{\star}\mathbf{x}^{\pi}_t + C_{\pi}\mathbf{a}^{\pi}_t + D_{\pi}\mathbf{e}_t = \mathbf{u}^{\pi}_t.\end{aligned}$$

■

Next, we show that $\mathbf{u}^{\Delta}_t$ can be represented as a linear function of the sequence $\bar{\mathbf{y}}^{\pi_0}_t = (\mathbf{y}^{\pi_0}, \mathbf{u}^{\pi_0})$:

Claim G.14 *Define*

$$C_{\Delta\pi,\mathrm{cl},u} := \begin{bmatrix}\bar{D}_{\pi}C_{\pi_0,\mathrm{cl}} & C_{\pi}\end{bmatrix}, \quad B_{\Delta\pi,\mathrm{cl},e} := \begin{bmatrix} B_{\pi_0,\mathrm{cl,in}}\bar{D}_{\pi}\\ \bar{B}_{\pi}\end{bmatrix}$$

Then, the matrices $N^{[0]} = \bar{D}_{\pi}$, $N^{[i]} = C_{\Delta\pi,\mathrm{cl},u}A^{i-1}_{\Delta\pi,\mathrm{cl}}B_{\Delta\pi,\mathrm{cl},e}$ satisfy

$$\mathbf{u}^{\Delta}_t = \sum_{i=0}^{t-1} N^{[i]}\bar{\mathbf{y}}^{\pi_0}_{t-i}, \quad \text{where } \bar{\mathbf{y}}^{\pi_0}_s = \begin{bmatrix}\mathbf{y}^{\pi_0}_t\\ \mathbf{u}^{\pi_0}_t\end{bmatrix}.$$

Proof Analogous to Claim C.7, and the fact that, in the absence of $\mathbf{u}_t^{\Delta}$, $\bar{\mathbf{y}}_t = \begin{bmatrix}\mathbf{y}_t^{\pi_0}\\ \mathbf{u}_t^{\pi_0}\end{bmatrix}$. ■

Let us now show that $N^{[i]}$ is given by $\bar{G}_{\pi_0\to\pi}$:

Claim G.15 *For all $i \geq 0$, $N^{[i]} = \bar{G}_{\pi_0\to\pi}^{[i]}$. As a consequence,*

$$\mathbf{u}_t^{\Delta} = \sum_{i=0}^{t-1} \bar{G}_{\pi_0\to\pi}^{[i]} \bar{\mathbf{y}}_{t-i}^{\pi_0}, \qquad \textit{where } \bar{\mathbf{y}}_s^{\pi_0} = (\mathbf{y}_t^{\pi_0}, \mathbf{u}_t^{\pi_0}).$$

Proof By definition $\bar{D}_\pi = D_{\pi_0\to\pi}$. To establish the identity, define the block permutation matrix T, where the blocks correspond to the $\mathbf{x}_t, \mathring{\mathbf{s}}_t, \mathbf{a}_t^{\Delta}$ states:

$$T = \begin{bmatrix} I & 0 & 0\\ 0 & 0 & I \\ 0 & I & 0\end{bmatrix}.$$

Since $T^2 = I$, it suffices to show that

$$TA_{\Delta\pi,\mathrm{cl}}T = A_{\pi_0\to\pi}, \quad TB_{\Delta\pi,\mathrm{cl}} = B_{\pi_0\to\pi}, \quad TC_{\Delta\pi,\mathrm{cl}} = C_{\pi_0\to\pi}.$$

Recall that

$$A_{\Delta\pi,\mathrm{cl}} := \begin{bmatrix} A_{\pi_0,\mathrm{cl}} + B_{\pi_0,\mathrm{cl,in}}\bar{D}_\pi C_{\pi_0,\mathrm{cl}} & B_{\pi_0,\mathrm{cl,in}}C_\pi \\ \bar{B}_\pi C_{\pi_0,\mathrm{cl}} & A_\pi. \end{bmatrix}$$

We begin with

$$\begin{aligned} A_{\pi_0,\mathrm{cl}} + B_{\pi_0,\mathrm{cl,in}}\bar{D}_\pi C_{\pi_0,\mathrm{cl}} &= \begin{bmatrix} A_\star + B_\star D_{\pi_0} C_\star & B_\star C_{\pi_0} \\ B_{\pi_0} C_\star & A_{\pi_0}\end{bmatrix} + \begin{bmatrix} B_\star \\ 0 \end{bmatrix}\begin{bmatrix} D_\pi & -I\end{bmatrix}\begin{bmatrix} C_\star & 0 \\ D_{\pi_0}C_\star & C_{\pi_0}\end{bmatrix} \\ &= \begin{bmatrix} A_\star + B_\star D_{\pi_0} C_\star & B_\star C_{\pi_0} \\ B_{\pi_0} C_\star & A_{\pi_0}\end{bmatrix} + \begin{bmatrix} B_\star \\ 0 \end{bmatrix}\begin{bmatrix} (D_\pi - D_{\pi_0})C_\star & -C_{\pi_0}\end{bmatrix} \\ &= \begin{bmatrix} A_\star + B_\star D_\pi C_\star & 0 \\ B_{\pi_0}C_\star & A_{\pi_0}\end{bmatrix} \end{aligned}$$

Moreover, recalling $\bar{B}_\pi = [B_\pi \mid 0]$, we have

$$\bar{B}_\pi C_{\pi_0,\mathrm{cl}} = \begin{bmatrix} B_\pi D_{\pi_0} C_\star & 0\end{bmatrix} + \begin{bmatrix} (D_\pi - D_{\pi_0})C_\star & -C_{\pi_0}\end{bmatrix} = \begin{bmatrix} B_\pi C_\star & 0 \end{bmatrix}$$

Finally, since $B_{\pi_0,\mathrm{cl,in}}C_\pi = \begin{bmatrix} B_\star C_\pi \\ 0\end{bmatrix}$, we have

$$A_{\Delta\pi,\mathrm{cl}} = \begin{bmatrix} A_{\pi_0,\mathrm{cl}} + B_{\pi_0,\mathrm{cl,in}}D_{\Delta\pi}C_{\pi_0,\mathrm{cl}} & B_{\pi_0,\mathrm{cl,in}}C_\pi \\ B_\pi C_{\pi_0,\mathrm{cl}} & A_\pi. \end{bmatrix} = \begin{bmatrix} A_\star + B_\star D_\pi C_\star & 0 & B_\star C_\pi \\ B_{\pi_0}C_\star & A_{\pi_0} & 0 \\ B_\pi C_\star & 0 & A_\pi \end{bmatrix}$$

Thus,

$$TA_{\Delta\pi,\mathrm{cl}}T = \begin{bmatrix} A_\star + B_\star D_\pi C_\star & B_\star C_\pi & 0 \\ B_\pi C_\star & A_\pi & 0 \\ B_{\pi_0}C_\star & 0 & A_{\pi_0}\end{bmatrix} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{A_{\pi,\mathrm{cl}}} & \begin{matrix}0\\0\end{matrix} \\ \hline B_{\pi_0}C_\star & 0 & A_{\pi_0}\end{array}\right] := A_{\pi_0\to\pi}$$

Now, recall $C_{\pi_0,\mathrm{cl}} = \begin{bmatrix} C_\star & 0 \\ D_{\pi_0}C_\star & C_\pi \end{bmatrix}$ and $B_{\pi_0,\mathrm{cl}} = \begin{bmatrix} I & B_\star D_{\pi_0} \\ 0 & B_{\pi_0} \end{bmatrix}$. Then,

$$\begin{aligned} C_{\Delta\pi,\mathrm{cl},u}T &:= \begin{bmatrix} \bar{D}_\pi C_{\pi_0,\mathrm{cl}} & C_\pi \end{bmatrix} T \\ &:= \begin{bmatrix} (D_\pi - D_{\pi_0})C_\star & -C_{\pi_0} & C_\pi \end{bmatrix} T \\ &:= \begin{bmatrix} (D_\pi - D_{\pi_0})C_\star & C_\pi & -C_{\pi_0} \end{bmatrix} = C_{\pi_0 \to \pi} \end{aligned}$$

and

$$\begin{aligned} TB_{\Delta\pi,\mathrm{cl},e} := TB_{\Delta\pi,\mathrm{cl},e} &:= \begin{bmatrix} B_{\pi_0,\mathrm{cl,in}}\bar{D}_\pi \\ \bar{B}_\pi \end{bmatrix} T \\ &:= \begin{bmatrix} B_\star D_\pi & -B_\star \\ 0 & 0 \\ B_\pi & 0 \end{bmatrix} T := \begin{bmatrix} B_\star D_\pi & -B_\star \\ B_\pi & 0 \\ 0 & 0 \end{bmatrix} = B_{\pi_0 \to \pi} \end{aligned}$$

■

We conclude the proof by showing that $\bar{\mathbf{y}}_t^{\pi_0} = (\mathbf{y}_t^{\pi_0}, \mathbf{u}_t^{\pi_0})$ can be represented in terms of $\mathbf{y}_t^{\pi_0}$:

Claim G.16 *Recall* $G_{\pi_0}^{[i]} = \mathbb{I}_{i=0}D_{\pi_0} + \mathbb{I}_{i\geq 1}C_{\pi_0}A_{\pi_0}^{i-1}(B_{\pi_0} + B_{\pi_0,u}D_{\pi_0})$. *Then,* $\mathbf{u}_t^{\pi_0} = \sum_{i=1}^t G_{\pi_0}^{[i]}\mathbf{y}_{t-1}^{\pi_0}$ *As a consequence,*

$$\bar{\mathbf{y}}_t^{\pi_0} = \sum_{i=0}^{t-1} G_{\pi_0,y\to(y,u)}^{[i]}\mathbf{y}_{t-i}^{\pi_0}$$

Proof Directly from the LDC equations. ■

In sum,

$$\mathbf{u}_t^{\Delta} = \sum_{i=0}^{t-1} G_{\pi_0\to\pi}^{[i]}\bar{\mathbf{y}}_{t-i}^{\pi_0} = \sum_{i=0}^{t-1}\sum_{j=0}^{t-i-1} \bar{G}_{\pi_0\to\pi}^{[i]} G_{\pi_0,y\to(y,u)}^{[j]}\mathbf{y}_{t-i-j}^{\pi_0},$$

which concludes the proof. □.

G.6.2. PROOF OF PROPOSITION G.12

Consider the system

$$\begin{aligned} \tilde{\mathbf{x}}_{t+1}^\star &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^\star - L\mathbf{y}_t + B_\star\mathbf{u}_t \\ \boldsymbol{\eta}_t^\star &= C_\star\tilde{\mathbf{x}}_t^\star - \mathbf{y}_t \\ \mathbf{v}_{t+1}^{\Delta\pi} &= A_{\Delta\pi}\mathbf{v}_t^{\Delta\pi} + B_{\Delta\pi}\boldsymbol{\eta}_t^\star \\ \mathbf{u}_t^{\Delta\pi} &= C_{\Delta\pi}\mathbf{v}_t^{\Delta\pi} + D_{\Delta\pi}\boldsymbol{\eta}_t^\star \\ \mathbf{u}_t &= F\tilde{\mathbf{x}}_t^\star + \mathbf{u}_t^{\Delta\pi}. \end{aligned} \tag{G.7}$$

From Proposition G.10, the inputs $\mathbf{u}_t$ coincide with $\mathbf{u}_t^{\pi}$ for all $t \geq 1$. Thus, if we set $\mathbf{u}_t^{\mathrm{ex}} = F(\tilde{\mathbf{x}}_t^{\star} - \tilde{\mathbf{x}}_t) + \mathbf{u}_t^{\Delta\pi}$, the system

$$\begin{aligned}\widehat{\mathbf{x}}_{t+1} &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - L\mathbf{y}_t + \widehat{B}\mathbf{u}_t \\ \mathbf{u}_t &= F\widehat{\mathbf{x}}_t + \mathbf{u}_t^{\mathrm{ex}} \\ \boldsymbol{\eta}_t &= \widehat{C}\widehat{\mathbf{x}}_{t+1} - \mathbf{y}_t\end{aligned} \tag{G.8}$$

also generates $\mathbf{u}_t = \mathbf{u}_t^{\pi}$. Now, let us represent the above as a system with inputs $\boldsymbol{\eta}_t^{\star}, \mathbf{u}_t^{\Delta\pi}$. We shall show that these can all be represented in terms of $\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$, concluding the proof.

First, we write

$$\begin{aligned}\tilde{\mathbf{x}}_{t+1}^{\star} &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\star} - L\mathbf{y}_t + B_\star\mathbf{u}_t \\ &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\star} + L\boldsymbol{\eta}_t^{\star} + L(\mathbf{y}_t - \boldsymbol{\eta}_t^{\star}) + B_\star F\tilde{\mathbf{x}}_t^{\star} + B_\star\mathbf{u}_t^{\Delta\pi} \\ &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\star} + L\boldsymbol{\eta}_t^{\star} - LC_\star\tilde{\mathbf{x}}_t^{\star} + B_\star F\tilde{\mathbf{x}}_t^{\star} + B_\star\mathbf{u}_t^{\Delta\pi} \\ &= (A_\star + B_\star F)\tilde{\mathbf{x}}_t^{\star} + L\boldsymbol{\eta}_t^{\star} + B_\star\mathbf{u}_t^{\Delta\pi},\end{aligned}$$

where we use the fact that $\boldsymbol{\eta}_t^{\star} = C_\star\tilde{\mathbf{x}}_t^{\star} - \mathbf{y}_t$. Next, we write

$$\begin{aligned}\widehat{\mathbf{x}}_{t+1} &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - L\mathbf{y}_t + \widehat{B}\mathbf{u}_t \\ &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t - L\mathbf{y}_t + \widehat{B}F\tilde{\mathbf{x}}_t^{\star} + \widehat{B}\mathbf{u}_t^{\Delta\pi} \\ &= (\widehat{A} + L\widehat{C})\widehat{\mathbf{x}}_t + L\boldsymbol{\eta}_t^{\star} + (\widehat{B}F - LC_\star)\tilde{\mathbf{x}}_t^{\star} + \widehat{B}\mathbf{u}_t^{\Delta\pi},\end{aligned}$$

where in the last line we use $\boldsymbol{\eta}_t^{\star} = C_\star\tilde{\mathbf{x}}_t^{\star} - \mathbf{y}_t$. This gives that

$$\begin{aligned}\begin{bmatrix}\tilde{\mathbf{x}}_{t+1}^{\star} \\ \widetilde{\mathbf{x}}_{t+1}\end{bmatrix} &= \begin{bmatrix}A_\star + B_\star F & 0 \\ \widehat{B}F - LC_\star & \widehat{A} + L\widehat{C}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}_t^{\star} \\ \widetilde{\mathbf{x}}_t\end{bmatrix} + \begin{bmatrix}B_\star & \widehat{B} \\ L & L\end{bmatrix}\begin{bmatrix}\mathbf{u}_t^{\Delta\pi} \\ \boldsymbol{\eta}_t^{\star}\end{bmatrix} \\ \mathbf{u}_t^{\mathrm{ex}} &= \begin{bmatrix}F & -F\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}_t^{\star} \\ \widetilde{\mathbf{x}}_t\end{bmatrix} + \begin{bmatrix}I & 0\end{bmatrix}\begin{bmatrix}\mathbf{u}_t^{\Delta\pi} \\ \boldsymbol{\eta}_t^{\star}\end{bmatrix}.\end{aligned}$$

Thus, letting

$$G_{\hat{\cdot}\to\star} := \mathsf{Transfer}\left(\begin{bmatrix}A_\star + B_\star F & 0 \\ \widehat{B}F - LC_\star & \widehat{A} + L\widehat{C}\end{bmatrix}, \begin{bmatrix}B_\star & \widehat{B} \\ L & L\end{bmatrix}, \begin{bmatrix}F & -F\end{bmatrix}, \begin{bmatrix}I & 0\end{bmatrix}\right)$$

denote the transfer operator mapping $\begin{bmatrix}\mathbf{u}_t^{\Delta\pi} \\ \boldsymbol{\eta}_t^{\star}\end{bmatrix} \to \mathbf{u}_t^{\mathrm{ex}}$, we can render

$$\mathbf{u}_t^{\mathrm{ex}} = \sum_{s=1}^{t} G_{\hat{\cdot}\to\star}^{[t-s]}\begin{bmatrix}\mathbf{u}_t^{\Delta\pi} \\ \boldsymbol{\eta}_t^{\star}\end{bmatrix}.$$

Next, for $G_{\mathrm{yla},\pi} := \mathsf{Transfer}(A_{\mathrm{yla},\pi}, B_{\mathrm{yla},\pi}, C_{\mathrm{yla},\pi}, D_{\mathrm{yla},\pi})$ from Proposition G.11, we have

$$\mathbf{u}_s^{\Delta\pi} = \sum_{j=1}^{s} G_{\mathrm{yla},\pi}^{[s-j]}\boldsymbol{\eta}_j^{\star},$$

giving that, for $\mathbf{u}_t^{\mathrm{ex}}$ defined in Eq. (G.8),

$$\forall t, \quad \mathbf{u}_t^{\mathrm{ex}} = \sum_{s=1}^{t}\sum_{j=1}^{s} G_{\hat{:}\to\star}^{[t-s]} \bar{G}_{\mathrm{yla},\pi}^{[s-j]} \boldsymbol{\eta}_t^{\star} \cdot, \quad \bar{G}_{\mathrm{yla},\pi}^{[i]} = \begin{bmatrix} G_{\mathrm{yla},\pi}^{[i]} \\ I_{d_y} \mathbb{I}_{i=0} \end{bmatrix} \tag{G.9}$$

To conclude, let us represent $\boldsymbol{\eta}_t^{\star}$ in terms of $\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$. Here, we use the crucial fact that the dynamics of $\boldsymbol{\eta}_t^{\star}$ are *non-counterfactual*. Thus, let us instead consider the following "natural" dynamics:

$$\begin{aligned}
\mathbf{x}_{t+1}^{\mathrm{nat}} &= A_\star \mathbf{x}_t^{\mathrm{nat}} + B_\star \mathbf{u}_t^{\mathrm{nat}} + \mathbf{w}_t \\
\mathbf{y}_t^{\mathrm{nat}} &= C_\star \mathbf{x}_t + \mathbf{e}_t \\
\hat{\mathbf{x}}_{t+1}^{\mathrm{nat}} &= (\widehat{A} + L\widehat{C})\hat{\mathbf{x}}_t^{\mathrm{nat}} - L\mathbf{y}_t^{\mathrm{nat}} + \widehat{B}\mathbf{u}_t^{\mathrm{nat}} \\
\mathbf{u}_t^{\mathrm{nat}} &= F\hat{\mathbf{x}}_t^{\mathrm{nat}} \\
\tilde{\mathbf{x}}_{t+1}^{\mathrm{nat}} &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\mathrm{nat}} - L\mathbf{y}_t^{\mathrm{nat}} + B_\star \mathbf{u}_t^{\mathrm{nat}} \\
\boldsymbol{\eta}_t^{\mathrm{nat}} &:= \widehat{C}\hat{\mathbf{x}}_t^{\mathrm{nat}} - \mathbf{y}_t^{\mathrm{nat}} \\
\boldsymbol{\eta}_t^{\star,\mathrm{nat}} &:= C_\star \tilde{\mathbf{x}}_t^{\mathrm{nat}} - \mathbf{y}_t^{\mathrm{nat}}
\end{aligned}$$

From Lemma G.5, the η-dynamics under exact observer feedback do not depend on the exogenous inputs; thus, $\boldsymbol{\eta}_t^{\star,\mathrm{nat}} = \boldsymbol{\eta}_t^{\star}$ for all t, where $\boldsymbol{\eta}^{\star,\mathrm{nat}}$ is defined in Eq. (G.7). Next, we can substitute

$$\begin{aligned}
\tilde{\mathbf{x}}_{t+1}^{\mathrm{nat}} &= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\mathrm{nat}} - L\mathbf{y}_t^{\mathrm{nat}} + B_\star \mathbf{u}_t^{\mathrm{nat}} \\
&= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\mathrm{nat}} + L\boldsymbol{\eta}_t^{\mathrm{nat}} - L\widehat{C}\hat{\mathbf{x}}_t^{\mathrm{nat}} + B_\star F\hat{\mathbf{x}}_t^{\mathrm{nat}} \\
&= (A_\star + LC_\star)\tilde{\mathbf{x}}_t^{\mathrm{nat}} + L\boldsymbol{\eta}_t^{\mathrm{nat}} + (B_\star F - L\widehat{C})\hat{\mathbf{x}}_t^{\mathrm{nat}}
\end{aligned}$$

Furthermore, we can write

$$\begin{aligned}
\hat{\mathbf{x}}_{t+1}^{\mathrm{nat}} &= (\widehat{A} + L\widehat{C})\hat{\mathbf{x}}_t^{\mathrm{nat}} + L\boldsymbol{\eta}_t^{\mathrm{nat}} - L\widehat{C}\hat{\mathbf{x}}_t^{\mathrm{nat}} + \widehat{B}F\hat{\mathbf{x}}_t^{\mathrm{nat}} \\
&= (\widehat{A} + \widehat{B}F)\hat{\mathbf{x}}_t^{\mathrm{nat}} + L\boldsymbol{\eta}_t^{\mathrm{nat}}.
\end{aligned}$$

Thus,

$$\begin{bmatrix} \tilde{\mathbf{x}}_{t+1}^{\mathrm{nat}} \\ \hat{\mathbf{x}}_{t+1}^{\mathrm{nat}} \end{bmatrix} = \begin{bmatrix} A_\star + LC_\star & B_\star F - L\widehat{C} \\ 0 & \widehat{A} + \widehat{B}F \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_t^{\mathrm{nat}} \\ \hat{\mathbf{x}}_t^{\mathrm{nat}} \end{bmatrix} + \begin{bmatrix} L \\ L \end{bmatrix} \boldsymbol{\eta}_t^{\mathrm{nat}}$$

Moreover,

$$\boldsymbol{\eta}_t^{\star,\mathrm{nat}} = C_\star \tilde{\mathbf{x}}_t^{\mathrm{nat}} - \mathbf{y}_t^{\mathrm{nat}} = C_\star \tilde{\mathbf{x}}_t^{\mathrm{nat}} - \widehat{C}\hat{\mathbf{x}}_t^{\mathrm{nat}} + \boldsymbol{\eta}_t^{\mathrm{nat}},$$

or in matrix form

$$\boldsymbol{\eta}_t^{\star,\mathrm{nat}} = \begin{bmatrix} C_\star & -\widehat{C} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_t^{\mathrm{nat}} \\ \hat{\mathbf{x}}_t^{\mathrm{nat}} \end{bmatrix} + I \cdot \boldsymbol{\eta}_t^{\mathrm{nat}}.$$

Hence, defining

$$G_{\star\to :} := \mathsf{Transfer}\left(\begin{bmatrix} A_\star + LC_\star & B_\star F - L\widehat{C} \\ 0 & \widehat{A} + \widehat{B}F \end{bmatrix}, \begin{bmatrix} L \\ L \end{bmatrix}, \begin{bmatrix} C_\star & -\widehat{C} \end{bmatrix}, I \right)$$

as the $\boldsymbol{\eta}_t^{\mathrm{nat}} \to \boldsymbol{\eta}_t^{\mathrm{nat}}$ transfer operator, we see that

$$\boldsymbol{\eta}_j^\star = \boldsymbol{\eta}_j^{\star,\mathrm{nat}} = \sum_{i=1}^{j} G_{\star\to\hat{:}}^{[j-i]} \boldsymbol{\eta}_i^{\mathrm{nat}}.$$

Thus, from Eq. (G.9), the exogenous inputs $\mathbf{u}_t^{\mathrm{ex}}$ from Eq. (G.8) satisfy

$$\begin{aligned} \mathbf{u}_t^{\mathrm{ex}} &= \sum_{s=1}^{t}\sum_{j=1}^{s}\sum_{i=1}^{j} G_{\hat{:}\to\star}^{[t-s]} \bar{G}_{\mathrm{yla},\pi}^{[s-j]} G_{\star\to\hat{:}}^{[j-i]} \boldsymbol{\eta}_i^{\mathrm{nat}}, \\ &= \sum_{s=1}^{t} (G_{\hat{:}\to\star} \odot \bar{G}_{\mathrm{yla},\pi} \odot G_{\star\to\hat{:}})^{[t-s]} \boldsymbol{\eta}_t^{\mathrm{nat}}. \end{aligned}$$

Since $\mathbf{u}_t^{\mathrm{ex}}$ induces the desired inputs $\mathbf{u}_t^{\pi}$, the proposition follows.

□

Appendix H. Regret Analysis: Non-Stochastic

While the theorems in the main paper hold for stable systems, the stated proofs and claims here hold for the more general setting of stabilizable systems, with the following modifications:

Definition H.1 (Modifications for the Stabilized Case) *The following modifications are made for the Stabilized Setting of Appendix G:*

1. *We are given access to a stabilizing controller satisfying Assumption 1b*
2. *We replace Assumption 1 with Assumption 1b.*
3. $\psi_{G_\star}$ *and* $R_{G_\star}$ *are defined as in Definition 2.3b.*
4. *We replace Algorithm 1 with Algorithm 4, and Algorithm 2 with Algorithm 5.*

(where we are granted access to a sub-optimal stabilizing controller).

H.1. Omitted Proofs from Section C

In this section, we present all ommited proofs from Section C, and demonstrate that all bounds either hold verbatim in the more general stabilized system setting, or present generalizations thereof. This ensures that Theorem 2 holds verbatim in the more general setting as well. Before continuing, let us review some of the notation from the stabilized setting, and how the stable system setting can be recovered:

- We use $\mathbf{v} = (\mathbf{y}, \mathbf{u}) \in \mathbb{R}^{d_y+d_u}$ to denote the pair of outputs and inputs on which the loss is measured. In particular, $\mathbf{v}_t^{\mathsf{alg}} = (\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}})$, and $\mathbf{v}^M = (\mathbf{y}_t^M, \mathbf{u}_t^M)$.
- The exogenous inputs $\mathbf{u}_t^{\mathrm{ex}}$ reduce to the inputs $\mathbf{u}_t$ in the stable case.
- The exogenous inputs $\mathbf{u}_t^{\mathrm{ex}}$ are linear in $\boldsymbol{\eta}_t^{\mathrm{nat}}$ or estimates $\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$; in the stable case, these correspond to $\mathbf{y}_t^{\mathrm{nat}}, \widehat{\mathbf{y}}_t^{\mathrm{nat}}$.

Next, we note that the regret decomposition is the same as in the stable case, given by Eq. C.3. We begin with a magnitude bound that generalizes Lemma C.1:

Lemma C.1b (Magnitude Bound) *Recall the notation with variants* $\mathbf{v}_t^{\mathsf{alg}}, \mathbf{v}^M$. *For all* t, *and* $M, M_1, M_2, \dots \in \mathcal{M}$, *we have*

$$\max\left\{\left\|\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}}\right\|_2, \left\|\mathbf{u}_t^{\mathrm{ex},M}\right\|_2, \left\|\mathbf{u}_t^{\mathrm{ex}}(M_t \mid \boldsymbol{\eta}_{1:t}^{\mathrm{nat}})\right\|_2\right\} \le R_{\mathcal{M}} R_{\mathrm{nat}}$$
$$\max\left\{\left\|\mathbf{v}_t\left[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}\right]\right\|, \|\mathbf{v}_t^{\mathsf{alg}}\|, \|\mathbf{v}_t^M\|\right\} \le 2R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}.$$

Proof The proofs of all these bounds are similar; let us focus on the $\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}}$, $\mathbf{v}_t^{\mathsf{alg}}$ sequence. We have $\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}} := \sum_{s=t-h\vee 0}^{T} \mathbf{M}_t^{[t-s]} \mathbf{y}_s^{\mathrm{nat}}$, from which Holder's inequality implies $\|\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}}\|_2 \le R_{\mathcal{M}} R_{\mathrm{nat}}$. Then, $\|\mathbf{v}_t^{\mathsf{alg}}\| = \|\mathbf{v}_t^{\mathrm{nat}} + \sum_{s=1}^t G_{\mathrm{ex}\to u}^{[t-s]} \mathbf{u}_s^{\mathrm{ex},\mathsf{alg}}\|_2 \le R_{\mathrm{nat}} + \|G_{\mathrm{ex}\to u}\|_{\ell_1,\mathrm{op}} \|\mathbf{u}_s^{\mathsf{alg}}\|_2 \le R_{\mathrm{nat}} + R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} \le 2R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}$, since $R_{\mathcal{M}}, R_{G_\star} \ge 1$. ■

We now restate the burn-in bound, which can be checked to hold in the more general present setting:

Lemma C.2 *We have that* (burn-in loss) $\le 4LR_{G_\star}^2 R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 (m+h)$

We now turn to the truncation costs (Lemma C.3):

Lemma C.3 (Bound on Truncation Errors) *We can bound*

$$\text{(algorithm truncation error)} + \text{(comparator truncation error)} \le 4LTR_{G_\star} R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 \psi_{G_\star}(h+1).$$

Proof Let us bound the algorithm truncation cost; the comparator cost is similar. Note that $\mathbf{u}_t^{\mathrm{ex},\mathsf{alg}} = \mathbf{u}_t^{\mathrm{ex}}(\mathbf{M}_t)$. By the magnitude bound (Lemma C.1b) and Lipschitz assumption,

$$\begin{aligned}
&\sum_{t=m+h+1}^{T} \ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \sum_{t=m+h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}] \le \\
&= 2LR_{G_\star} R_\psi R_{\mathrm{nat}} \\
&\qquad \cdot \sum_{t=m+h+1}^{T} \left\| \begin{bmatrix} \mathbf{y}_t[\mathbf{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}] - \mathbf{y}_t^{\mathsf{alg}} \\ \mathbf{u}_t[\mathbf{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}] - \mathbf{u}_t^{\mathsf{alg}} \end{bmatrix} \right\|_2 \\
&= 2LR_{G_\star} R_\psi R_{\mathrm{nat}} \sum_{t=m+h+1}^{T} \left\| \left[\sum_{s=1}^{t-h-1} G_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex},\mathsf{alg}} \right] \right\|_2 \\
&\le 2LTR_{G_\star} R_\psi R_{\mathrm{nat}} \psi_{G_\star}(h+1) \max_{s\in[T]} \|\mathbf{u}_s^{\mathrm{ex},\mathsf{alg}}\|_2.
\end{aligned}$$

By Lemma C.1b, the above is at most $2LTR_{G_\star} R_\psi^2 R_{\mathrm{nat}}^2 \psi_{G_\star}(h+1)$. ■

Next, we turn to bounding the Lipschitz constants. For this, we shall need the following bound:

Lemma H.2 (Norm Relations) *For any $M \in \mathcal{K}$, we have*

$$\|M\|_{\ell_1,\mathrm{op}} \le \sqrt{m}\|M\|_{\mathrm{F}} \quad \textit{and} \quad \max_{i=0}^{h} \|M_{t-i}\|_{\ell_1,\mathrm{op}} \le \sqrt{m}\|M_{t:t-h}\|_{\mathrm{F}}$$

Proof The first inequality follows form Cauchy Schwartz:

$$\|M\|_{\ell_1,\mathrm{op}} = \sum_{i=1}^{m} \|M^{[i-1]}\|_{\mathrm{op}} \le \sum_{i=1}^{m} \|M^{[i-1]}\|_{\mathrm{F}} \le \sqrt{m}\|M\|_{\mathrm{F}}.$$

The second follows from using the first to bound

$$\max_{i=0}^{h} \|M_{t-i}\|_{\ell_1,\mathrm{op}} \le \max_{i} \sqrt{m}\|M_{t-i}\|_{\mathrm{F}} \le \sqrt{m}\|M_{t:t-h}\|_{\mathrm{F}}.$$

■

As a second intermediate step, we show that the maps $M \mapsto \mathbf{y}_t(M \mid G_\star, \mathbf{y}^{\mathrm{nat}}_{1:t})$ and $M \mapsto \mathbf{u}_t(M \mid G_\star, \mathbf{y}^{\mathrm{nat}}_{1:t}, \mathbf{u}^{\mathrm{nat}}_{1:t})$ are Lipschitz:

Lemma H.3 (Lipschitz Bound on Coordinate Mappings) *For any $M, \tilde{M} \in \mathcal{M}(M, R)$,*

$$\begin{aligned}&\left\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}_t(\tilde{M} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t)\right\|_2 \\ &\quad \le R_{\mathrm{nat}}\|M - \tilde{M}\|_{\ell_1,\mathrm{op}} \le \sqrt{m} R_{G_\star} R_{\mathrm{nat}} \|M - \tilde{M}\|_{\mathrm{F}}.\end{aligned}$$

Similarly, for any $M_{t:t-h}, \tilde{M}_{t:t-h} \in \mathcal{M}(M, R)^{h+1}$,

$$\begin{aligned}&\left\|\mathbf{v}_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t] - \mathbf{v}_t[\tilde{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t]\right\|, \\ &\le \sqrt{m} R_{G_\star} R_{\mathrm{nat}} \max_{s=t-h}^{t} \|M_s - \tilde{M}_s\|_{\mathrm{F}}.\end{aligned}$$

Proof Let us prove the bound for $\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}_t(M' \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t)\|_2$, for time varying $M_{t:t-h}$ and $\tilde{M}_{t:t-h}$ are similar. We have

$$\begin{aligned}\|\mathbf{v}_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t] - \mathbf{v}_t[\tilde{M}_{t:t-h}\|_2 &= \left\|\sum_{s=t-h}^{t-1} G_{\mathrm{ex}\to(y,u)} \left(\mathbf{u}^{\mathrm{ex}}(M \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:t} s) - \mathbf{u}^{\mathrm{ex}}(\tilde{M} \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s})\right)\right\| \\ &\le R_{G_\star} \max_{s=t-h}^{t-1} \left\|\mathbf{u}^{\mathrm{ex}}(M \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}(\tilde{M} \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:t} s)\right\| \\ &= R_{G_\star} \max_{s=t-h}^{t-1} \left\|\sum_{j=s-m+1}^{s} (M - \tilde{M})^{[s-j]} \boldsymbol{\eta}^{\mathrm{nat}}_j\right\|_2 \\ &\le R_{\mathrm{nat}} R_{G_\star} \left\|M - \tilde{M}\right\|_{\ell_1,\mathrm{op}} \le \sqrt{m} R_{\mathrm{nat}} R_{G_\star} \left\|M - \tilde{M}\right\|_{\mathrm{F}},\end{aligned}$$

where the last step uses Lemma H.2. ■

We now present and prove the generalization of Lemma C.5 to the stabilized setting:

Lemma C.5b *Define $L_f := L\sqrt{m}R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}$. Then, the functions $f_t(\cdot \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}})$ are L_f-Lipschitz, and $F_t[M_{t:t-h} \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}]$ are L_f-coordinate-wise Lipschitz on $\mathcal{M}$ in the Frobenius norm $\|M\|_{\mathrm{F}} = \|[M^{[0]}, \dots, M^{[m-1]}]\|_{\mathrm{F}}$. Moreover, the Euclidean diameter of $\mathcal{M}$ is at most $D \le 2\sqrt{d}R_{\mathcal{M}}$.*

Proof Let us bound the coordinate-Lipschitz constant of F_t, the bound for f_t is similar.

$$
\begin{aligned}
&\left|F_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}] - F_t[\tilde{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}]\right| \\
&= \left|\ell_t\left(\mathbf{v}_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}]\right) - \ell_t\left(\mathbf{v}_t[\tilde{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}]\right)\right| \\
&\le LR_{\mathrm{nat}}R_{G_\star}R_{\mathcal{M}}\left\|\mathbf{v}_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}] - \mathbf{v}_t[\tilde{M}_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}]\right\| \\
&\le \sqrt{m}LR_{\mathrm{nat}}^2R_{G_\star}^2R_{\mathcal{M}}\left(\max_{s\in[t-h:t]}\|M_s - \tilde{M}_s\|_{\mathrm{F}}\right),
\end{aligned}
$$

where the last inequality is by Lemma H.3. The bound on the diameter of $\mathcal{M}$ follows from Lemma H.2

∎

H.2. Estimation Bounds: Proof of Theorems 7 & 7b

We state a generalization of Proof of Theorems 7 for estimating both respose $G_{\mathrm{ex}\to\eta}$ and $G_{\mathrm{ex}\to(y,u)}$:

Theorem 7b (Guarantee for Algorithm 5, Generalization of Theorem 7) *Let $\delta \in (e^{-T}, T^{-1})$, $N, d_u \le T$, and $\psi_{G_\star}(h+1) \le \frac{1}{\sqrt{N}}$. Define $d_{\max} = \max\{d_y + d_u, d_\eta\}$, and set*

$$
\epsilon_G(N,\delta) = \frac{h^2 R_{\mathrm{nat}}}{\sqrt{N}}C_\delta, \quad \text{where } C_\delta := 14\sqrt{d_{\max} + d_y + \log\tfrac{1}{\delta}}, \quad \text{and } R_{\mathbf{u},\mathrm{est}} := 3\sqrt{d_u + \log(1/\delta)}.
$$

and suppose that $N \ge h^4 C_\delta^2 R_{\mathbf{u},\mathrm{est}}^2 R_{\mathcal{M}}^2 R_{G_\star}^2 + c_0 h^2 d_u^2$ for an appropriately large c_0, which can be satisfied by taking

$$
N \ge 1764(d_{\max} + d_u + \log(1/\delta))^2 h^4 R_{\mathcal{M}}^2 R_{G_\star}^2 + c_0 h^2 d_u^2.
$$

Then with probability $1 - \delta - N^{-\log^2 N}$, Algorithm 5 satisfies the following bounds

1. *$\epsilon_G \le 1/\max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{G_\star}\}$.*
2. *For all $t \in [N]$, $\|\mathbf{u}_t\| \le R_{\mathbf{u},\mathrm{est}} := 3\sqrt{d_u + \log(1/\delta)}$*
3. *For estimation error is bounded as*

$$
\begin{aligned}
\|\widehat{G}_{\mathrm{ex}\to\eta} - G_{\mathrm{ex}\to\eta}\|_{\ell_1,\mathrm{op}} &\le \|\widehat{G}_{\mathrm{ex}\to\eta}^{[0:h]} - G_{\mathrm{ex}\to\eta}^{[0:h]}\|_{\ell_1,\mathrm{op}} + R_{\mathbf{u},\mathrm{est}}\psi_{G_\star}(h+1) \le \epsilon_G \\
\|\widehat{G}_{\mathrm{ex}\to(y,u)} - G_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} &\le \|\widehat{G}_{\mathrm{ex}\to(y,u)}^{[1:h]} - G_{\mathrm{ex}\to(y,u)}^{[1:h]}\|_{\ell_1,\mathrm{op}} + R_{\mathbf{u},\mathrm{est}}\psi_{G_\star}(h+1) \le \epsilon_G.
\end{aligned}
$$

Moreover, Algorithm 2 satisfies the same for $\widehat{G}, G_\star$.

Proof Let us focus on Algorithm 5, the bound for Algorithm 2 is the special case of $\widehat{G}_{\mathrm{ex}\to(y,u)} = \widehat{G}$ and $G_{\mathrm{ex}\to(y,u)} = G_\star$.

The first bound of the lemma is strictly numerical. Lets prove the second part of the lemma. Using standard gaussian concentration (see e.g. Vershynin (2018, Section 4.2)):

Claim H.4 *With probability* $1-\delta/3$ *and* $\delta \le 1/T \le 1/N$ *and* $\delta \le 1/3$, $\|\mathbf{u}_t\| \le \frac{4}{3}\sqrt{2}\sqrt{d_u \log 9 + \log(3N/\delta)} \le R_{\mathbf{u},\mathrm{est}} := 3\sqrt{d_u + \log(1/\delta)}$ *for all* $t \in [N]$. *Denote this event* $\mathcal{E}^{\mathbf{u},\mathrm{bound}}$.

Let us turn to the last part of the lemma. To begin, let us bound the truncation error. We have

$$\begin{aligned}\frac{2}{\epsilon_G} \cdot R_{\mathbf{u},\mathrm{est}} \|G_\star^{[>h]} - \widehat{G}^{[>h]}\|_{\ell_1,\mathrm{op}} &\le \psi_{G_\star}(h+1)\frac{2R_{\mathbf{u},\mathrm{est}}}{h^2 R_{\mathrm{nat}} C_\delta \sqrt{N}} \\ &\le \psi_{G_\star}(h+1)\frac{1}{h^2 R_{\mathrm{nat}}\sqrt{N}} \\ &\le \psi_{G_\star}(h+1)\frac{1}{\sqrt{N}} \le 1\end{aligned}$$

where the second inequality uses $C_\delta \ge 2R_{\mathbf{u},\mathrm{est}}$, the thir uses $h^2 R_{\mathrm{nat}} \ge 1$, and the four holds from our choice of $\psi_{G_\star}(h+1)$. Hence,

$$R_{\mathbf{u},\mathrm{est}} \|G_\star^{[>h]} - \widehat{G}^{[>h]}\|_{\ell_1,\mathrm{op}} \le \epsilon_G/2 \le 1/6 \tag{H.1}$$

, where the last step uses Part 1 of the lemma.

Let us now bound the estimation error. We begin by bounding $\|\widehat{G}^{[0:h]} - G_\star^{[0:h]}\|_{\mathrm{op}}$. To this end, define $\boldsymbol{\delta}_t = \mathbf{v}_t - \sum_{i=1}^h G_\star^{[i]}\mathbf{u}_t$, and define $\boldsymbol{\Delta} = [\boldsymbol{\delta}_{N_1}^\top \mid \cdots \mid \boldsymbol{\delta}_N^\top]$. Simchowitz et al. (2019) develop error bounds in terms of the operator norm of $\boldsymbol{\Delta}$. In the subsubection below, we provide a simplified and self-contaned proof of the estimation guarantees from Simchowitz et al. (2019):

Lemma H.5 (Simplification of Proposition 3.2 in Simchowitz et al. (2019)) *Then, if* N *is sufficiently large that* $N \ge chd_u \log^4(N)$ *for some universal constant* $c > 0$, *and* $\mathcal{E}_\lambda$ *is the event that* $\|\boldsymbol{\Delta}\|_{\mathrm{op}} \le \lambda$, *then, with probability* $1 - N^{-\log^2(N)} - \delta/4$

$$\|\widehat{G}_{\mathrm{LS}} - G_\star^{[1:h]}\|_{\mathrm{op}} \le \frac{5.6}{N}\lambda\sqrt{(h+1)(d_{\max} + d_u + \log(h/\delta))}.$$

In particular, for $h \ge 2$, *we have the simplifid bound*

$$\|\widehat{G}_{\mathrm{LS}} - G_\star^{[1:h]}\|_{\mathrm{op}} \le \frac{5.6}{N}\lambda h\sqrt{d_{\max} + d_u + \log(1/\delta)}$$

Observe that for a sufficiently large constant c_0, taking $N \geq c_0 h^2 d_u^2$ implies our condition in the above lemma $N \geq chd_u \log^4(N)$. Next, let us bound $\|\mathbf{\Delta}\|_{\mathrm{op}}$. We the have on the event $\mathcal{E}^{\mathbf{u},\mathrm{bound}}$:

$$\begin{aligned}
\|\mathbf{\Delta}\|_{\mathrm{op}} &\leq \sqrt{N} \max_{t\in[N]} \|\boldsymbol{\delta}_t\| = \sqrt{N} \max_{t\in[N]} \|\mathbf{y}_t - \sum_{s=1}^{h} G_\star^{[s]} \mathbf{u}_{t-s}\| \\
&= \sqrt{N} \max_{t\in[N]} \|\mathbf{y}_t^{\mathrm{nat}} + \sum_{s>h} G_\star^{[s]} \mathbf{u}_{t-s}\| \leq \sqrt{N} \left(\max_{t\in[N]} \|\mathbf{y}_t^{\mathrm{nat}}\| + \|G_\star^{[>h]}\|_{\ell_1,\mathrm{op}} \max_{t\in[N]} \|\mathbf{u}_{t-s}\| \right) \\
&\leq \sqrt{N} \left(R_{\mathrm{nat}} + \|G_\star^{[>h]}\|_{\ell_1,\mathrm{op}} \cdot R_{\mathbf{u},\mathrm{est}} \right) \\
&\leq \sqrt{N} \left(R_{\mathrm{nat}} + \frac{1}{6}) \right) && \text{(Eq. (H.1))} \\
&\leq \frac{7}{6} \sqrt{N} R_{\mathrm{nat}},
\end{aligned}$$

where we used the assumed upper bound on $\psi_{G_\star}$ from Plugging the above into Lemma H.5 and using $R_{\mathrm{nat}} \geq 1$ by assumption gives gives

$$\begin{aligned}
2\|\widehat{G}_{\mathrm{LS}} - G_\star^{[1:h]}\|_{\mathrm{op}} &\leq \frac{7}{3} R_{\mathrm{nat}} \sqrt{N} \cdot \frac{5.6}{N} h \sqrt{d_{\max} + d_u + \log(1/\delta)} \\
&= \frac{14h\sqrt{d_{\max} + d_u + \log(1/\delta)}}{\sqrt{N}} = hC_\delta/\sqrt{N}.
\end{aligned}$$

Thus, $2\|\widehat{G}_{\mathrm{LS}} - G_\star^{[1:h]}\|_{\ell_1,\mathrm{op}} \leq 14h^2 C_\delta/\sqrt{N} := \epsilon_G(N, \delta)$, as needed.

■

H.2.1. PROOF OF LEMMA H.5

We adopt the argument of Simchowitz et al. (2019), but provide a simpler and self-contained proof. Let us focus on the $G_{\mathrm{ex}\to(y,u)}$ case, which we shall denote $G_\star$ for the present argument. We denote the esimtate of $\widehat{G}_{\mathrm{ex}\to(y,u)}$ and $\widehat{G}$. Further, let $\overline{\mathbf{U}}$ denote the matrix with rows $\mathbf{u}_{t:t+h}^{\mathrm{ex,alg}}$ for $1 \leq t \leq N - h - 1$. Moreover, let $\boldsymbol{\delta}_t = \mathbf{y}_t - \sum_{i=1}^{h} G_\star^{[i]} \mathbf{u}_t$, and let $\mathbf{\Delta}$ denote the matrix with rows $\boldsymbol{\delta}_t$ for $h + 1 \leq t \leq N$. We then have the identity

$$\widehat{G} - G_\star = (\overline{\mathbf{U}}^\top \overline{\mathbf{U}})^{-1} \overline{\mathbf{U}}^\top \mathbf{\Delta}$$

We can crudely bound

$$\|\widehat{G} - G_\star\|_{\mathrm{op}} \leq \|\overline{\mathbf{U}}^\top \overline{\mathbf{U}})^{-1}\|_{\mathrm{op}} \|\overline{\mathbf{U}}^\top \mathbf{\Delta}\|_{\mathrm{op}}.$$

Let us now bound the operator norm of $\|\overline{\mathbf{U}}^\top \mathbf{\Delta}\|_{\mathrm{op}}$. We have that the columns of $\overline{\mathbf{U}}^\top \mathbf{\Delta}$ are of the form

$$[(\mathbf{u}_{t-i}^{\mathrm{ex,alg}} \boldsymbol{\delta}_t)_{h+1 \leq t \leq N}], \quad i \in \{0, 1, \ldots, h\}$$

Thus, by Tsiamis and Pappas (2019, Lemma A.1), and the definition of the operator norm (with $\mathcal{S}^{d_u-1} := \{v :\in \mathbb{R}^{d_u} : \|v\| = 1\}$),

$$\begin{aligned}\|\overline{\mathbf{U}}^\top \mathbf{\Delta}\|_{\mathrm{op}} &\le \sqrt{h+1} \max_{i\in\{0,\dots,h\}} \|[(\mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}]\|_{\mathrm{op}} \\ &\le \max_{i\in\{0,\dots,h\}} \max_{v\in\mathcal{S}^{d_u-1}} \|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2\end{aligned}$$

By the self-normalized martingale bounds (Abbasi-Yadkori et al. (2011, Theorem 3)), and the fact that $\boldsymbol{\delta}_t$ is $\mathcal{F}_{t-h-1}$ measurable, where $(\mathcal{F}_t)$ is the filtration generated by the random inputs, we have that with probability $1-\delta$

$$\|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}(\mathbf{\Delta}^\top\mathbf{\Delta} + \lambda^2 I)\|_2^2 \le 2\log\det(\frac{1}{\lambda^2}\mathbf{\Delta}^\top\mathbf{\Delta} + I) + 2\log(1/\delta)$$

In particular, if λ^2 is any parameter such that the event $\mathcal{E}_\lambda := \|\mathbf{\Delta}\|_{\mathrm{op}} \le \lambda$, then we have that with probability $1-\delta$ that whenever $\mathcal{E}_\lambda$ holds,

$$\begin{aligned}\frac{1}{\lambda^2}\|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2^2 &\le \|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}(\mathbf{\Delta}^\top\mathbf{\Delta} + \lambda^2 I)\|_2 \\ &\le 2\log\det(\frac{1}{\lambda^2}\mathbf{\Delta}^\top\mathbf{\Delta} + I) + 2\log(1/\delta) \\ &\le 2\log\det(2I) + 2\log(1/\delta) = 2(d_{\max}\log 2 + \log(1/\delta)).\end{aligned}$$

So rearranging,

$$\|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2 \le \lambda\sqrt{2(d_{\max}\log 2 + \log(1/\delta))}.$$

Next, by a standard covering argument Vershynin (2018, Section 4.2), we have that if $\mathcal{S}_0$ is an $1/5$-net of $\mathcal{S}^{d_u-1}$, then $\max_{v\in\mathcal{S}^{d_u-1}} \|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2 \le \frac{5}{4}\max_{v\in\mathcal{S}_0} \|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2$, and that we can take $\log|\mathcal{S}_0| \le d_u \log 9$. Thus, by a union bound over $v \in \mathcal{S}_0$ and $i \in \{0,\dots,h\}$, the following holds with probability $1-\delta/4$,

$$\begin{aligned}\|(v^\top \mathbf{u}^{\mathrm{ex,alg}}_{t-i}\boldsymbol{\delta}_t)_{h+1\le t\le N}\|_2 &\le \frac{4\sqrt{2}}{3}\lambda\sqrt{((d_{\max}+d_u)\log 9 + \log(4(h+1)/\delta))} \\ &\le \frac{4\sqrt{2}}{3}\lambda\sqrt{((d_{\max}+d_u)\log 9 + \log(8h/\delta))} \\ &\le 2.8\lambda\sqrt{d_{\max}+d_u+\log(h/\delta)},\end{aligned}$$

Hence, we have that

$$\|\widehat{G} - G_\star\|_{\mathrm{op}} \le 2.8\|(\overline{\mathbf{U}}^\top\overline{\mathbf{U}})^{-1}\|_{\mathrm{op}}\lambda\sqrt{(h+1)(d_{\max}+d_u+\log(h/\delta))}.$$

Finally, by constants in the argument modifing the arguments of Oymak and Ozay (2019, Lemma C.2), we have that for any ϵ, we can ensure that for $T \ge c(\epsilon)(h+1)d_u\log^4(Nd_u)$, we can ensure $\|(\overline{\mathbf{U}}^\top\overline{\mathbf{U}})^{-1}\|_{\mathrm{op}} \le (1-\epsilon)(N-(h+1))^{-1}$ with probability $1-N^{-\log^2 N}$. By enforcing $N \ge (h+1)/4$ and taking $\epsilon = 1/2$, we can obtain $\|(\overline{\mathbf{U}}^\top\overline{\mathbf{U}})^{-1}\|_{\mathrm{op}} \le N/2$, yielding

$$\|\widehat{G} - G_\star\|_{\mathrm{op}} \le 5.6\|(\overline{\mathbf{U}}^\top\overline{\mathbf{U}})^{-1}\|_{\mathrm{op}}\lambda\sqrt{(h+1)(d_{\max}+d_u+\log(h/\delta))},$$

with probability $1-\delta/4-N^{-\log^2 N}$ on $\mathcal{E}_\lambda$.

H.3. Unknown System Regret (Section D.3)

Let us conclude with presenting the omitted proofs from Section D.3, and generalize to the stabilized case. The regret decomposition is identical to Eq. (D.2), modifying the functions if necessary to capture their dependence on $\widehat{\mathbf{u}}^{\mathrm{nat}}_{1:t}$. Throughout, we will assume ϵ_G satisfies a generalization of Condition D.1 to the stabilized setting:

Condition D.1b (Estimation Condition) *We assume that the event of Theorem 7b holds (i.e. accuracy of estimates $\widehat{G}_{\mathrm{ex}\to(y,u)}$ and $\widehat{G}_{\mathrm{ex}\to\eta}$), which entail $\epsilon_G(N,\delta) \le 1/2\max\{R_{\mathcal{M}}R_{G_\star}, R_{\mathbf{u},\mathrm{est}}\}$, where $R_{\mathbf{u},\mathrm{est}} = R_{\mathbf{u},\mathrm{est}}(\delta) := 3\sqrt{d_u + \log(1/\delta)}$. Moreover, these entail that $\|\widehat{G}_{\mathrm{ex}\to\eta}\|_{\ell_1,\mathrm{op}}, \|\widehat{G}_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} \le 2R_{G_\star}$. These also entail that $\max_{s\le N}\|\mathbf{u}^{\mathsf{alg}}_s\| \le R_{\mathbf{u},\mathrm{est}}$.*

We begin with the following generalization of the stability guarantee of Lemma D.1:

Lemma D.1b (Stability of $\widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}$) *Introduce the notation $\overline{R}_{\mathbf{u}} := 2\max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\}$. Then, for $\epsilon_G \le 1/2\max\{R_{\mathcal{M}}R_{G_\star}, R_{\mathbf{u},\mathrm{est}}\}$ (satisfied by Condition D.1b) the following holds $t \in [T]$,*

$$\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_t\|_2 \le \overline{R}_{\mathbf{u}}, \quad \|\mathbf{v}^{\mathsf{alg}}_t\| \le R_{\mathrm{nat}} + R_{G_\star}\overline{R}_{\mathbf{u}} \le 2R_{G_\star}\overline{R}_{\mathbf{u}},$$
$$\|\widehat{\mathbf{v}}^{\mathrm{nat}}_t\|_2, \|\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t\|_2 \le 2R_{\mathrm{nat}},$$

Proof The proof is analogous to that of Lemma D.1, but with the following modifications. Let us sketch the major steps in the proof: we first establish the inequality $\|\boldsymbol{\eta}^{\mathrm{nat}}_t - \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t\|_2 \le \epsilon_G\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_{1:t}\|_{2,\infty}$, where we recall the notation $\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_{1:t}\|_{2,\infty} = \max_{s\in\{1,\dots,t\}}\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_s\|_2$ introduced in the original proof. Next, we can establish that

$$\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_t\|_2 \le \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\} + \epsilon_G R_{\mathcal{M}}\|G_{\mathrm{ex}\to\eta}\|_{\ell_1,\mathrm{op}}\|\mathbf{u}^{\mathsf{alg}}_{1:t-1}\|_{2,\infty}.$$

By assumption, $\|G_{\mathrm{ex}\to\eta}\|_{\ell_1,\mathrm{op}} \le R_{G_\star}$ (Definition 2.3b). Hence, for $\epsilon_G \le 1/2\max\{R_{\mathcal{M}}R_{G_\star}, R_{\mathbf{u},\mathrm{est}}\} \le 1/2R_{\mathcal{M}}R_{G_\star}$, we can recursively verify that $\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_t\|_2 \le 2\max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}}R_{\mathrm{nat}}\}$. Lastly, we can bound $\|\widehat{\mathbf{v}}^{\mathrm{nat}}_t - \mathbf{v}^{\mathrm{nat}}_t\|_2 \le \epsilon_G\|\mathbf{u}^{\mathrm{ex},\mathsf{alg}}_{1:t}\|_{2,\infty} \le R_{\mathrm{nat}}$ under the conditon of the lemma, giving $\|\widehat{\mathbf{v}}^{\mathrm{nat}}_t\| \le 2\|\mathbf{v}^{\mathrm{nat}}_t\|$. Similarly, we can bound $\|\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t\| \le 2\|\boldsymbol{\eta}^{\mathrm{nat}}_t\|$. ■

In the stabilized setting, we shall need to slightly modify our magnitude bounds to account for that norms of the controls:

Lemma H.6 (Magnitude Bounds for Estimated System) *Suppose that Condition D.1b holds. Then, for any $t > N$, and all $M \in \mathcal{M}(m, R_{\mathcal{M}})$ and $M_{t:t-h} \in \mathcal{M}(m, R_{\mathcal{M}})^{h+1}$,*

$$\|\mathbf{u}^{\mathsf{alg}}_t\|, \|\mathbf{u}^{\mathrm{ex}}_t(M \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t})\| \le 2R_{\mathcal{M}}R_{\mathrm{nat}}$$
$$\left\|\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t}, \widehat{\mathbf{v}}^{\mathrm{nat}}_t\right]\right\|_2 \le 6R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}$$

Proof We have that $\|\mathbf{u}^{\mathrm{ex}}_t(M)\|_2 = \|\sum_{i=0}^{m-1} M^{[i]}\widehat{\mathbf{y}}^{\mathrm{nat}}_{t-i}\| \le R_{\mathcal{M}}\max_{s\le t}\|\widehat{\mathbf{y}}^{\mathrm{nat}}_t\| \le 2R_{\mathcal{M}}R_{\mathrm{nat}}$ by Lemma D.1b. The bound on $\mathbf{u}^{\mathsf{alg}}_t$ specializes by setting $M \leftarrow \mathbf{M}_t$.

By the same lemma, and the fact that $\|\widehat{G}_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} \le \epsilon_G + \|G_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} \le 2R_{G_\star}$ by Condition D.1b,

$$\left\|\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t}, \widehat{\mathbf{v}}^{\mathrm{nat}}_t\right]\right\|_2 = \left\|\widehat{\mathbf{v}}^{\mathrm{nat}}_t + \sum_{s=t-h}^{t} \widehat{G}^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex}}_s(M_s \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t})\right\|$$
$$\le 2R_{\mathrm{nat}} + \|\widehat{G}_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} \max_{M\in\mathcal{M}(m,R_{\mathcal{M}})} \max_{s\le t} \|\mathbf{u}^{\mathrm{ex}}_t(M \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t})\|$$
$$\le 2R_{\mathrm{nat}} + 4R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}} \le 6R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}.$$

The bound on $\|\mathbf{y}_t\left[M_{t:t-h} \mid \widehat{G}, \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}\right]\|_2$ is similar. ■

Next, we check that the proof of Lemma D.2 goes through in the general case

Proof [Proof of Lemma D.2 for Stabilized Setting] The proof is analogous to the general case, where we replace the dependence no R_{nat} and $R_{G_\star}$ with an upper bound on $\widehat{\mathbf{y}}^{\mathrm{nat}}_t$, $\widehat{\mathbf{u}}^{\mathrm{nat}}_t$ and $\max\{\|\widehat{G}_{u\to y}\|_{\ell_1,\mathrm{op}}, \|\widehat{G}_{u\to u}\|_{\ell_1}$ In light of the above bounds, these quantities are also $\lesssim R_{\mathrm{nat}}$ and $R_{G_\star}$, up to additional constant factors, yielding the same regret bound up to constants. ■

Lastly, we establish Lemma D.5, encompassing both the stabel and stabilized case. Given that the proof is somewhat involved, we organize it in the following subsection.

H.3.1. PROOF OF LEMMA D.4/D.4B

We bound the error in estimating natures y's and natures u's:

Lemma D.4b (Accuracy of Estimated Nature's y and u and η) *Assume Condition D.1b. Then for $t \ge N+h+1$, we have that*

$$\|\mathbf{v}^{\mathrm{nat}}_t - \widehat{\mathbf{v}}^{\mathrm{nat}}_t\|_2 \le 3R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$$
$$\|\boldsymbol{\eta}^{\mathrm{nat}}_t - \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t\|_2 \le 3R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$$

Proof Let us focus on $\|\mathbf{v}^{\mathrm{nat}}_t - \widehat{\mathbf{v}}^{\mathrm{nat}}_t\|_2$, the error bound on $\|\boldsymbol{\eta}^{\mathrm{nat}}_t - \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_t\|_2$ is similar. Let us use the notation $G^{[0:\ell]}$ to denote the restriction of a Markov operator to $(G^{[i]})_{i=0}^{\ell}$, and $G^{[>\ell]}$ to restrict to $(G^{[i]})_{i>\ell}$. We can then bound:

$$\|\mathbf{v}^{\mathrm{nat}}_t - \widehat{\mathbf{v}}^{\mathrm{nat}}_t\|_2 = \left\|\sum_{s=1}^{t} G^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex,alg}}_t - \widehat{G}^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex,alg}}_t\right\|_2$$
$$\le \left\|G^{[0:t-(N+1)]}_{\mathrm{ex}\to(y,u)} - \widehat{G}^{[0:t-(N+1)]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \cdot \max_{s\in[N+1:t]} \|\mathbf{u}^{\mathrm{ex,alg}}_s\|$$
$$+ \left\|G^{[>t-(N+1)]}_{\mathrm{ex}\to(y,u)} - \widehat{G}^{[>t-N+1]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \cdot \max_{s\le N} \|\mathbf{u}^{\mathrm{ex,alg}}_s\|.$$

For $t \ge N+h-1$, we have $\|\mathbf{u}^{\mathrm{alg}}_t\| \le 2R_{\mathcal{M}}R_{\mathrm{nat}}$, and under Condition D.1b, we have $\max_{s\le N}\|\mathbf{u}^{\mathrm{ex,alg}}_s\| \le \overline{R}_{\mathbf{u}}$. Moreover, we can bound $\left\|G^{[0:t-N+1]}_{\mathrm{ex}\to(y,u)} - \widehat{G}^{[0:t-N+1]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \le \left\|G_{\mathrm{ex}\to(y,u)} - \widehat{G}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \le \epsilon_G$ under Condition D.1b. In addition, since $t \ge N+h+1$, $t-N+1 \ge h$, so

$$\left\|G^{[>t-(N+1)]}_{\mathrm{ex}\to(y,u)} - \widehat{G}^{[>t-N+1]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \le \left\|G^{[>h]}_{\mathrm{ex}\to(y,u)} - \widehat{G}^{[>h]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}}$$
$$= \left\|G^{[>h]}_{\mathrm{ex}\to(y,u)}\right\|_{\ell_1,\mathrm{op}} \le \psi_{G_\star}(h+1),$$

where we use the fact that $\widehat{G}^{[i]}_{\mathrm{ex}\to(y,u)} = 0$ for $i > 0$, and the fact that we define $\psi_{G_\star}(\cdot) := \max\{\psi_{G_{\mathrm{ex}\to(y,u)}}(\cdot), \psi_{G_{\mathrm{ex}\to\eta}}(\cdot)\}$ (Definition 2.3b). Thus, all in all,

$$\|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 \le 2R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G + \psi_{G_\star}(h+1)R_{\mathbf{u},\mathrm{est}} \le 3R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G,$$

where the last step holds because $R_{\mathcal{M}}R_{\mathrm{nat}} \ge 1$ by assumption, and that under Condition D.1b, $\psi_{G_\star}(h+1)R_{\mathbf{u},\mathrm{est}} \le \epsilon_G$. ■

H.3.2. PROOF OF LEMMA D.5

We prove the lemma in the more general stabilized setting, where we require the stronger Condition D.1b instead of Condition D.1. For completeness, we state this general bound here

Lemma D.5b (Approximation Error Bounds: Stabilized) *Under Condition D.1b,*

$$\text{(loss approximation error)} + \text{(comparator approximation error)} \lesssim LTR_{G_\star}R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\epsilon_G$$

Proof Let us start with the loss approximation error. For $t \ge N+h+1$, and using $\mathbf{u}_t^{\mathrm{ex}}(\mathbf{M}_t) = \mathbf{u}_t^{\mathrm{ex,alg}}$, we have

$$\begin{aligned}
&\left\|\mathbf{v}_t^{\mathsf{alg}} - \mathbf{v}_t[\mathbf{M}_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}]\right\|_2 \\
&\quad= \|\mathbf{v}_t^{\mathrm{nat}} + \sum_{s=1}^{t} G^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_s^{\mathrm{ex,alg}} - (\widehat{\mathbf{v}}_t^{\mathrm{nat}} + \sum_{s=t-h}^{t} \widehat{G}^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_s^{\mathrm{ex,alg}})\|_2 \\
&\quad= \|\mathbf{v}_t^{\mathrm{nat}} + \sum_{s=1}^{t} G^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_s^{\mathrm{ex,alg}} - (\widehat{\mathbf{v}}_t^{\mathrm{nat}} + \sum_{s=1}^{t} \widehat{G}^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_s^{\mathrm{ex,alg}})\|_2 \quad \widehat{G}^{[i]}_{\mathrm{ex}\to(y,u)} = 0 \text{ for } i > h \\
&\quad\le \|\mathbf{y}_t^{\mathrm{nat}} - \widehat{\mathbf{y}}_t^{\mathrm{nat}}\|_2 + \left\|\sum_{s=1}^{t} G^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_t^{\mathrm{ex,alg}} - \widehat{G}^{[t-s]}_{\mathrm{ex}\to(y,u)}\mathbf{u}_t^{\mathrm{ex,alg}}\right\|_2 \\
&\quad= \|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 + \|\mathbf{v}_t^{\mathrm{nat}} = \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 \le 6R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G
\end{aligned}$$

where we use Lemma D.1b in the last inequality. Moreover, recalling the following bound from Lemma H.6,

$$\|\mathbf{v}_t\left[\mathbf{M}_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right]\|_2 \le 6R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}$$

we have

$$\begin{aligned}
&\max\left\{\|\mathbf{v}_t^{\mathsf{alg}}\|_2, \left\|\mathbf{v}_t\left[M_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}\right]\right\|_2\right\} \\
&\qquad\qquad\qquad\le 6R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}} + 6R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G \le 9R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}},
\end{aligned}$$

where we use the bound $\epsilon_G \le 1/2R_{\mathcal{M}}R_{G_\star} \le R_{G_\star}/2$ under Condition D.1b and the bounds $R_{\mathcal{M}}, R_{G_\star} \ge 1$. Hence, we have

$$
\begin{aligned}
&\text{(loss approximation error)}\\
&:= \sum_{t=N+m+h+1}^{T} \ell_t(\mathbf{v}_t^{\mathsf{alg}}) - \sum_{t=N+m+h+1}^{T} F_t[\mathbf{M}_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}]\\
&\le \sum_{t=N+m+h+1}^{T} |\ell_t(\mathbf{y}_t^{\mathsf{alg}}, \mathbf{u}_t^{\mathsf{alg}}) - \ell_t(\mathbf{v}_t[\mathbf{M}_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}])]|\\
&\le 9LR_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}\\
&\quad \cdot \sum_{t=N+m+h+1}^{T} \left\| \mathbf{v}_t^{\mathsf{alg}} - \mathbf{v}_t[\mathbf{M}_{t:t-h} \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}] \right\|_2\\
&\le 54LTR_{G_\star}R_{\mathcal{M}}^2R_{\mathrm{nat}}^2\epsilon_G.
\end{aligned}
$$

where we used Assumption 2 and the bounds computed above.

Let us now turn to the comparator approximation error

$$
\begin{aligned}
&\text{(comparator approximation error)}\\
&:= \inf_{M\in\mathcal{M}} \sum_{t=N+m+h+1}^{T} f_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}) - \inf_{M\in\mathcal{M}} \sum_{t=N+m+h+1}^{T} \ell_t(\mathbf{y}_t^M, \mathbf{u}_t^M)\\
&\le \max_{M\in\mathcal{M}} \sum_{t=N+m+h+1}^{T} |\ell_t(\mathbf{v}_t^M) - \ell_t(\mathbf{v}_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}))|\\
&\le 6LTR_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}\\
&\quad \cdot \max_{M\in\mathcal{M}} \sum_{t=N+m+h+1}^{T} \left\| \mathbf{v}_t^M - \mathbf{v}_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}) \right\|_2
\end{aligned}
$$

where again we use the magnitude bounds in Lemmas C.1b and H.6, and the Lipschitz Assumption (Assumption 2). Let us bound the differences between the $\mathbf{v}_t$ terms, taking caree that errors is introduced by both the approximation of the transfer function $G_{\mathrm{ex}\to(y,u)}$ and the Nature's η sequence

$\boldsymbol{\eta}_t^{\mathrm{nat}}$. For $t \geq N + h + 1 + m$, we obtain

$$
\begin{aligned}
&\left\| \mathbf{v}_t^M - \mathbf{v}_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}) \right\|_2 \\
&= \left\| \mathbf{v}_t^{\mathrm{nat}} + \sum_{s=1}^{t} G_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex}}(M \mid \boldsymbol{\eta}_{1:s}^{\mathrm{nat}}) - \left(\widehat{\mathbf{v}}_t^{\mathrm{nat}} + \sum_{s=t-h}^{t} \widehat{G}_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}}) \right) \right\|_2 \\
&= \underbrace{\|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2}_{\leq 3R_{\mathcal{M}} R_{\mathrm{nat}} \epsilon_G} + \underbrace{\|G_{\mathrm{ex}\to(y,u)} - \widehat{G}_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}}}_{\leq \epsilon_G} \underbrace{\max_{t-h \leq s \leq t} \|\mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}})\|}_{\leq 2R_{\mathcal{M}} R_{\mathrm{nat}}} \\
&\quad + \left\| \sum_{s=1}^{t-h+1} G_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}}) \right\|_2 \\
&= 5R_{\mathcal{M}} R_{\mathrm{nat}} \epsilon_G + \left\| \sum_{s=1}^{t} G_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}}) \right\|_2 .
\end{aligned}
$$

where we have used the magnitude bounds in Lemma D.1b and H.6. We can further bound

$$
\begin{aligned}
\left\| \sum_{s=1}^{t} G_{\mathrm{ex}\to(y,u)}^{[t-s]} \mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}}) \right\|_2 &\leq \psi_{G_\star}(h+1) \max_{s \leq t} \|\mathbf{u}_s^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:s}^{\mathrm{nat}})\| \\
&\leq \psi_{G_\star}(h+1) 2 \max\{R_{\mathbf{u},\mathrm{est}}, R_{\mathcal{M}} R_{\mathrm{nat}}\},
\end{aligned}
$$

where we use Lemma D.1b abolve. From Conditions D.1/D.1b, we can bound $\psi_{G_\star}(h+1) R_{\mathbf{u},\mathrm{est}} \leq \epsilon_G$. And since $R_{\mathbf{u},\mathrm{est}} \geq 1$, this implies that the above is at most $2\psi_{G_\star} R_{\mathbf{u},\mathrm{est}} R_{\mathcal{M}} R_{\mathrm{nat}} \leq 2\epsilon_G R_{\mathcal{M}} R_{\mathrm{nat}}$. Thus, from the above previous two displays,

$$
\left\| \mathbf{v}_t^M - \mathbf{v}_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}) \right\|_2 \leq 7\epsilon_G R_{\mathcal{M}} R_{\mathrm{nat}},
$$

giving

$$
(\text{comparator approximation error}) \leq 6LR_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} \cdot 7\epsilon_G R_{\mathcal{M}} R_{\mathrm{nat}} = 42LTR_{G_\star} R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 \epsilon_G
$$

Combining the two bounds, we and using $R_{G_\star}, R_{\mathcal{M}} \geq 1$,

$$
(\text{comparator approximation error}) + (\text{loss approximation error}) \lesssim TLR_{G_\star} R_{\mathcal{M}}^2 R_{\mathrm{nat}}^2 \epsilon_G
$$

■

Appendix I. Strongly Convex, Semi-Adversarial Regret

We begin by stating a slight generalization of the semi-adversarial model described by Assumption 6. Recall the assumption that our noises decompose as follows:

$$\mathbf{w}_t = \mathbf{w}_t^{\mathrm{adv}} + \mathbf{w}_t^{\mathrm{stoch}}$$
$$\mathbf{e}_t = \mathbf{e}_t^{\mathrm{adv}} + \mathbf{e}_t^{\mathrm{stoch}}$$

We make the following assumption on the noise and losses:

Assumption 6b (Semi-Adversarial Noise: Martingale Structure) *We assume that there is a filtration $(\mathcal{F}_t)_{t\ge 0}$ and a matrix $\Sigma_{\mathrm{noise}} \in \mathbb{R}^{(d_x+d_y)^2} \succeq 0$ (possibly degenerate), and $\sigma_{\mathbf{w}}^2, \sigma_{\mathbf{e}}^2 \ge 0$ (possibly zero) such that the following hold:*

1. *The adversarial disturbance sequences $(\mathbf{w}_t^{\mathrm{adv}})$ and $(\mathbf{e}_t^{\mathrm{adv}})$ and the loss sequence $\ell_t(\cdot)$ are* oblivious*, in the sense that they are $\mathcal{F}_0$-adapted.*
2. *The sequences $(\mathbf{w}_t^{\mathrm{stoch}})$ and $(\mathbf{e}_t^{\mathrm{stoch}})$ and $(\mathcal{F}_t)$-adapted*
3. *$\mathbb{E}[\mathbf{e}_t^{\mathrm{stoch}} \mid \mathcal{F}_{t-1}] = 0$, $\mathbb{E}[\mathbf{w}_t^{\mathrm{stoch}} \mid \mathcal{F}_{t-1}] = 0$.*
4. *The noises satisfy*

$$\mathbb{E}\left[\begin{bmatrix}\mathbf{w}_t^{\mathrm{stoch}} \\ \mathbf{e}_t^{\mathrm{stoch}}\end{bmatrix}\begin{bmatrix}\mathbf{w}_t^{\mathrm{stoch}} \\ \mathbf{e}_t^{\mathrm{stoch}}\end{bmatrix}^\top \mid \mathcal{F}_{t-1}\right] \succeq \Sigma_{\mathrm{noise}} \succeq \begin{bmatrix}\sigma_{\mathbf{w}}^2 I_{d_x} & 0 \\ 0 & \sigma_{\mathbf{e}}^2 I_{d_y}\end{bmatrix}.$$

Moreover, at least one of the following hold:

(a) *The system is internally stable has no stabilizing controller π_0, and $\sigma_{\mathbf{w}}^2 + \sigma_{\mathbf{e}}^2 > 0$*

(b) *The system is stabilized by a static feedback controller π_0 (that is, $A_{\pi_0} = 0$ and $\dim(\mathbf{z}^\pi) = 0$), and $\sigma_{\mathbf{w}}^2 > 0$*

(c) *The system is stabilized by a general stabilizing controller, and $\sigma_{\mathbf{e}}^2 > 0$.* [14]

As in thes stable setting, the strong convexity parameter governs the functions

$$f_{t;k}\left(M \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}\right) := \mathbb{E}\left[f_t(M \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}) \mid \mathcal{F}_{t-k}\right]. \tag{I.1}$$

For stabilized settings, Proposition E.2 admits the following generalization:

Proposition E.2b (Strong Convexity for known system) *Suppose that we interact with an internally-controlled system (Definition 2.1). Then, under assumptions Assumptions 2, 3, 5, 1b and 6b, there exists system dependent constants $m_{\mathrm{sys}}, p_{\mathrm{sys}} \ge 0$ and $\alpha_{f;0} > 0$ such that, for $h = \lfloor m/3 \rfloor$, $k = m + 2h$, and $m \ge m_{\mathrm{sys}}$, the functions $f_{t;k}\left(M \mid G_\star,\, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}\right)$ are $\alpha_{f;m}$-strongly convex, where*

$$\alpha_{f;m} := \alpha_{\mathrm{loss}} \cdot \alpha_{\mathrm{sys}} \cdot m^{p_{\mathrm{sys}}}$$

In other words, the strong convexity parameter decays at most polynomially *in m.*

14. This condition can be generalized somewhat to a form of "output controllability" of the noise transfer function, which can potentially accomodate $\sigma_{\mathbf{e}} = 0$. We omit this generalization in the interest of brevity

The above proposition is given in Appendix J.2.1. For general LDC-Ex controllers, we do not have transparent expressions for α_{sys} and $m^{p_{\mathrm{sys}}}$. Nevertheless, we ensure that the above bound is strong enough to ensures rates of $\log^{\mathcal{O}(1)+p_{\mathrm{sys}}} T$ and $\sqrt{T}\log^{\mathcal{O}(1)+p_{\mathrm{sys}}}$, where the exponent hidden by $\mathcal{O}(1)$ does not depend on system parameters (so that the exponents are determined solely by p_{sys}). We make a couple remarks, which in particular describe how p_{sys} is often 0 in many settings:

1. In general, the strong convexity parameters of the system are determined by the properties of the Z-transforms for relevant operators. A general expression is given in Theorem 11 , and the preliminaries and definitions relevant for the theorem are given in Appendix J.1. Proposition E.2b is proven in Appendix J.2 as a consequence of this more general result, and Appendix J contains all details related to establishing strong convexity.

2. In Appendix J.2.2, we show that for systems stabilized via static feedback, we can take $p_{\mathrm{sys}} = 0$, and give explict and transparent bounds on α_{sys}. This recovers the special case of internally stable systems as a special case, where we can take $\alpha_{\mathrm{sys}} = \sigma_{\mathbf{e}}^2 + \frac{\sigma_{\mathbf{w}}^2\sigma_{\min}(C_\star)}{(1+\|A_\star\|_{\mathrm{op}})^2}$.

3. For the special case of internally systems (Proposition E.2,), we present a smaller self-contained proof that does not appeal to Z-transform machinery (Proposition E.2. Note that this resut does not require that $m \geq m_{\mathrm{sys}}$ restriction required by Proposition E.2b.

4. The parameter m_{sys} is related to the decay of the system, and can be deduced from the conditions of Theorem 11.

Theorems 4 and 5 generalize to the stabilized-system setting:

Theorem 4b (Fast Rate for Known System: Stabilized Case) *Suppose assumptions Assumptions 2, 3, 5, 1b and 6b holds. Thenw with the additional condition $h = \lfloor m/3 \rfloor$ and and appropriate modifications as in Definition H.1, Theorem 4 holds verbatim when α_f is replaced with the stabilized analgoue $\alpha_{f;m}$ from Proposition E.2b. In particular, taking $\alpha = \Omega(\alpha_{f;m})$, we obtain regret bounded by*

$$\mathrm{Regret}_T(\psi) \lesssim \frac{L^2 m^{3+p_{\mathrm{sys}}} d_{\min} R_{\mathrm{nat}}^4 R_{G_\star}^4 R_{\mathcal{M}}^2}{\min\left\{\alpha_{\mathrm{sys}}, L R_{\mathrm{nat}}^2 R_{G_\star}^2\right\}} \left(1 + \frac{\beta_{\mathrm{loss}}}{L R_{\mathcal{M}}}\right) \cdot \log \frac{T}{\delta}. \tag{I.2}$$

In particular, under Assumption 4, we obtain

$$\mathrm{Regret}_T(\psi) \lesssim (\tfrac{1}{1-\rho}\log\tfrac{T}{\delta})^{p_{\mathrm{sys}}} \cdot \mathrm{poly}(C, L, \beta_{\mathrm{loss}}, \frac{1}{\alpha}, \sigma_{\mathrm{noise}}^2, \frac{1}{1-\rho}, \log\tfrac{T}{\delta}) \cdot (d_{\max}^2\sqrt{T} + d_{\max}^3),$$

where the exponents in the $\mathrm{poly}(\cdot)$ *term do not depend on system parameters, although p_{sys} does.*

Again, for general stabilized system, we may suffer exponents which depend on this system-dependent p_{sys}. But, as discussed above p_{sys} may be equal to 0 in many cases of interest.

For unknown systems, we have the following:

Theorem 5b (Fast Rate for Unknown System: Stabilized Case) *Suppose assumptions Assumptions 2, 3, 5, 1b and 6b holds. Thenw with the additional condition $h = \lfloor m/3 \rfloor$ and and appropriate* *modifications as in Definition H.1, Theorem 5 holds verbatim when α_f is replaced with the stabilized analgoue $\alpha_{f;m}$ from Proposition E.2b. In particular, taking $\alpha = \Omega(\alpha_{f;m})$ and Assumption 4, we obtain*

$$\mathrm{Regret}_T(\psi) \lesssim \mathrm{poly}(C, L, \beta_{\mathrm{loss}}, \frac{1}{\alpha}, \sigma^2_{\mathrm{noise}}, \frac{1}{1-\rho}, \log\tfrac{T}{\delta}) \qquad \cdot (d^2_{\max}\sqrt{T(\tfrac{1}{1-\rho}\log\tfrac{T}{\delta})^{p_{\mathrm{sys}}}} + (\tfrac{1}{1-\rho}\log\tfrac{T}{\delta})^{p_{\mathrm{sys}}} d^3_{\max}),$$

where the exponent in p_{sys} where the exponents in the $\mathrm{poly}(\cdot)$ *term do not depend on system parameters, although p_{sys} does.*

I.1. Proof Details for Theorems 4 and 4b

The proof of Theorems Theorems 4 and 4b are identical, except for the difference in strong convexity parameters in view of Proposition E.2b and Proposition E.2. Thus, the proof of Proposition E.2b follows from the proof of Proposition E.2b given in Appendix E, ammending α_f to $\alpha_{f;m}$ where it arises.

It remains to supply a the ommited proof of the lemma that establishes smoothness of the objectives, Lemma E.3. We restate the lemma here to include the $\mid G_\star$, $\mathbf{y}^{\mathrm{nat}}_{1:t}, \mathbf{u}^{\mathrm{nat}}_{1:t}$ encountered in the stabilized case:

Lemma E.3b (Smoothness) *The functions $f_t(M \mid G_\star, \mathbf{y}^{\mathrm{nat}}_{1:t}, \mathbf{u}^{\mathrm{nat}}_{1:t})$ are β_f-smooth, where we define $\beta_f := mR^2_{\mathrm{nat}}R^2_{G_\star}\beta_{\mathrm{loss}}$.*

Proof For brevity, we omit $\mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t$. Let $\mathsf{D}\mathbf{v}_t[\cdot]$ the differential of the function as maps from $\mathbb{R}^{(md_yd_u)} \to \mathbb{R}^{d_y+d_u}$, these are elements of $\mathbb{R}^{(md_yd_u)\times(d_y+d_u)}$. These are affine functions, and thus do not depend on the M argument.From the chain rule (with appropriate transpose conventions), and the fact that affine functions have vanishing second derivative

$$\begin{aligned}\nabla f_t(M) &= \mathsf{D}\mathbf{v}_t(\nabla\ell)(\mathbf{v}_t)\\ \nabla^2 f_t(M) &= \mathsf{D}\mathbf{v}_t \cdot (\nabla^2\ell)(\mathbf{v}_t)\cdot \mathsf{D}\mathbf{v}_t^\top \preceq \beta_{\mathrm{loss}}\,\|\mathsf{D}\mathbf{v}_t\|^2_{\mathrm{op}}\, I \end{aligned} \tag{I.3}$$

Let us now bound the norm of the differentials. Observe that $\|\mathsf{D}(\mathbf{u}_t(M)\|_{\mathrm{op}}$ and $\|\mathsf{D}(\mathbf{y}_t(M)\|_{\mathrm{op}}$ are just the Frobenius norm to ℓ_2 Lipschitz constant of $M \mapsto \mathbf{v}_t(M)$ is bounded by $\sqrt{m}R_{G_\star}R_{\mathrm{nat}}$ via Lemma H.3. Thus $\nabla^2 f_t(M) \preceq \beta_{\mathrm{loss}} m R^2_{G_\star} R^2_{\mathrm{nat}} I$, as needed. ∎

I.2. Supporting Proofs for Theorems Theorems 5 and 5b

We now generalize to the stabilized, unknown setting. Throughout, we shall use the various magnitude bounds on $\widehat{\mathbf{y}}^{\mathrm{nat}}, \widehat{\mathbf{u}}^{\mathrm{nat}}, \dots$ developed in Appendix H.3 for unknown system / Lipschitz loss setting.

For this strongly convex, stabilized, unknown setting, we generalize the true prediction losses of Definition F.1 as follows:

Definition F.1b (True Prediction Losses) *We define the true prediction losses as*

$$\begin{aligned}\mathbf{v}^{\mathrm{pred}}_t[M_{t:t-h}] &:= \mathbf{v}_t[M_{t:t-h} \mid G_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t]\\ &= \mathbf{v}^{\mathrm{nat}}_t + \sum_{i=0}^{h} G^{[i]}_{\mathrm{ex}\to(y,u)}\mathbf{u}^{\mathrm{ex}}_t(M \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:t-i})\\ F^{\mathrm{pred}}_t[M_{t:t-h}] &:= \ell_t\left(\mathbf{v}^{\mathrm{pred}}_t[M_{t:t-h}]\right),\end{aligned}$$

and let $f_t^{\mathrm{pred}}(M) = F_t^{\mathrm{pred}}(M, \dots, M)$ *denote the unary specialization. The corresponding conditional functions of interest are*

$$f_{t;k}^{\mathrm{pred}}(M) := \mathbb{E}\left[f_t^{\mathrm{pred}}(M) \mid \mathcal{F}_{t-k}\right].$$

Throughout the proof, it will be useful to adopt the shorthand $\mathbf{v}_t^\star(M) := \mathbf{u}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}})$, and $\mathbf{y}_t^\star(M) := \mathbf{y}_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}})$ to denote the counterfactuals for the true nature's y, and $\widehat{\mathbf{v}}_t(M) := (M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}})$ denote the counterfactuals for the estimates $\widehat{G}$ and $\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}$ and $\widehat{\mathbf{v}}^{\mathrm{nat}}$. Note that $\mathbf{u}^{\mathrm{pred}}$ and $\mathbf{y}^{\mathrm{pred}}$ can be though as interpolating between these two sequences.

We shall also let $\mathsf{D}\mathbf{v}_t^{\mathrm{pred}}$ denote differentials as elements of are elements of $\mathbb{R}^{(md_yd_u)\times(d_y+d_u)}$, and similarly for $\mathsf{D}\mathbf{v}^\star$ and $\mathsf{D}\widehat{\mathbf{v}}$. As these functions are affine, the differential is independent of M-argument

I.2.1. PRELIMINARY NOTATION AND PERTURBATION BOUNDS

Before continuing, we shall state and prove two useful lemmas that will help bound the gradients / Lipschitz constants of various quantities of interest.

Lemma I.1 (Norm and Perturbation Bounds) *The following bounds hold for* $t > N$*:*

(a) $\|\mathsf{D}\mathbf{v}_t^{\mathrm{pred}}\|_{\mathrm{op}} \le 2\sqrt{m}R_{\mathrm{nat}}R_{G_\star}$

(b) $\|\mathbf{v}_t^{\mathrm{pred}}(M) - \mathbf{v}_t^\star(M)\|_2 \le 3\sqrt{m}R_{\mathcal{M}}R_{\mathrm{nat}}R_{G_\star}\|M\|_{\mathrm{F}}$

(c) $\|\mathsf{D}(\mathbf{v}_t^{\mathrm{pred}} - \widehat{\mathbf{v}}_t)\|_{\mathrm{op}} \le 2\sqrt{m}\epsilon_G R_{\mathrm{nat}}$

(d) For all $M \in \mathcal{M}$*,* $\|\mathbf{v}_t^{\mathrm{pred}}(M) - \widehat{\mathbf{v}}_t(M)\|_2 \le 5R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$.

Proof Note that operator norm bounds on the differential are equivalent to the Frobenius-to-ℓ_2 Lipschitz constants of the associated mappings. The proofs are then analogous to the proof of Lemma H.3, where the role of $\mathbf{y}_t^{\mathrm{nat}}$ and $G_\star$ are replaced with the appropraite quantities. For clarity, we provide a relevant generalization of that lemma, without proof.

Lemma I.2 (Lipschitz Bound on Generalized Coordinate Mappings) *Let* $\widetilde{G}, \widetilde{\mathbf{y}}_{1:t}^{\mathrm{nat}}, \widetilde{\mathbf{u}}_{1:t}^{\mathrm{nat}}$ *be arbitrary, let* $R_{\widetilde{G}} = \|\widetilde{G}\|_{\ell_1,\mathrm{op}}$ *and* $\widetilde{R}_{\mathrm{nat},t} := \max\{\|\widetilde{\mathbf{y}}_s^{\mathrm{nat}}\| : s \in [t-h-m+1 : t]\}$*. Then,*

$$\begin{aligned}&\|\mathbf{v}_t(M \mid \widetilde{G}, \widetilde{\mathbf{v}}_t^{\mathrm{nat}}, \widetilde{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}) - \mathbf{v}_t(\tilde{M} \mid \widetilde{G}, \widetilde{\mathbf{v}}_t^{\mathrm{nat}}, \widetilde{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}})\|_2 \\ &\quad \le \widetilde{R}_{\mathrm{nat},t}R_{\widetilde{G}}\|M - \tilde{M}\|_{\ell_1,\mathrm{op}} \le \sqrt{m}R_{\widetilde{G}}\widetilde{R}_{\mathrm{nat},t}\|M - \tilde{M}\|_{\mathrm{F}}.\end{aligned}$$

The generalized to non-unary functions of $M_{t-h:t}$ *is analogous Lemma H.3. Notice the above bound* does not *depend on* $\widetilde{\mathbf{v}}_t^{\mathrm{nat}}$*, which consitutes an affine term.*

. For part (a), the bound follows by bounding $\|\widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}\| \le 2R_{\mathrm{nat}}$ by Lemma D.1b, and applying Lemma I.2 with $\widetilde{\boldsymbol{\eta}}_t^{\mathrm{nat}} \leftarrow \widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$, and $\widetilde{\mathbf{v}}_t^{\mathrm{nat}} \leftarrow \widehat{\mathbf{v}}_t^{\mathrm{nat}}$, and $\widetilde{G} \leftarrow G_{\mathrm{ex}\to(y,u)}$. {**MS:** *from here*}

In part (b), we can compute

$$\begin{aligned}\|\mathbf{v}_t^{\mathrm{pred}}(M) - \mathbf{v}_t^\star(M)\|_2 &\le \|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \mathbf{v}_t^{\mathrm{nat}}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}) - \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \mathbf{v}_t^{\mathrm{nat}}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}})\|_2 \\ &= \|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \mathbf{v}_t^{\mathrm{nat}}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}} - \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}})\|_2.\end{aligned}$$

Let us apply Lemma I.2 with $\widetilde{\boldsymbol{\eta}}_t^{\mathrm{nat}} \leftarrow \boldsymbol{\eta}_t^{\mathrm{nat}} - \widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$ and $\widetilde{G} \leftarrow G_{\mathrm{ex}\to(y,u)}$. The associated value of $\widetilde{R}_{\mathrm{nat},t}$ can be replaced by an upper bound on $\max\{\boldsymbol{\eta}_t^{\mathrm{nat}} - \widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}} : s \in [N+1:t]\}$, which we can take to be $3R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$ by Lemma D.4b. This gives a bound of $3\sqrt{m}R_{\mathcal{M}}R_{\mathrm{nat}}R_{G_\star}\|M\|_{\mathrm{F}}$., as needed.

For part (c), take $\widetilde{\boldsymbol{\eta}}_t^{\mathrm{nat}} \leftarrow \widehat{\boldsymbol{\eta}}_t^{\mathrm{nat}}$ $\widetilde{G} \leftarrow \widehat{G}_{\mathrm{ex}\to(y,u)} - G_{\mathrm{ex}\to(y,u)}$ playing the role of $G_\star$, yielding a $\widetilde{R}_{\mathrm{nat},t} \le 2R_{\mathrm{nat}}$ by Lemma D.1b and $R_{\widetilde{G}} \le \epsilon_G$ from Lemma H.6.

Finally, let us establish part (d). We have

$$\begin{aligned}
\|\mathbf{v}_t^{\mathrm{pred}}(M) - \widehat{\mathbf{v}}_t(M)\|_2 &\le \|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 + \|\sum_{i=0}^{h}(G_{\mathrm{ex}\to(y,u)}^{[i]} - \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]})\mathbf{u}_{t-i}^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:t-i}^{\mathrm{nat}})\| \\
&\le \|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 + \|\sum_{i=0}^{h}(G_{\mathrm{ex}\to(y,u)}^{[i]} - \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]})\mathbf{u}_{t-i}^{\mathrm{ex}}(M \mid \widehat{\boldsymbol{\eta}}_{1:t-i}^{\mathrm{nat}})\| \\
&\le \|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 + 2R_{\mathcal{M}}R_{\mathrm{nat}}\sum_{i=0}^{h}\|(G_{\mathrm{ex}\to(y,u)}^{[i]} - \widehat{G}_{\mathrm{ex}\to(y,u)}^{[i]})\|_{\mathrm{op}} \\
&\le \|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 + 2R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G,
\end{aligned}$$

where the second to last step uses Lemma H.6. Finally, we can bound $\|\mathbf{v}_t^{\mathrm{nat}} - \widehat{\mathbf{v}}_t^{\mathrm{nat}}\|_2 \le 3R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G$ by Lemma D.4b. Combining the bounds yields the proof. ■

I.2.2. GRADIENT ERROR (LEMMAS F.2 AND F.2B)

Lemma F.2b *For any $M \in \mathcal{M}$, we have that*

$$\left\|\nabla f_t(M \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{u}}_{1:t}^{\mathrm{nat}}) - \nabla f_t^{\mathrm{pred}}(M)\right\|_{\mathrm{F}} \le C_{\mathrm{approx}}\,\epsilon_G,$$

where is $C_{\mathrm{approx}} := \sqrt{m}R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}^2(8\beta_{\mathrm{loss}} + 12L)$.

Proof Let denote the differential of the functions as maps from $\mathbb{R}^{(md_yd_u)}$,, respectively. Define differentials analogously for $\mathbf{u}^{\mathrm{pred}}, \mathbf{y}^{\mathrm{pred}}$. Then,

$$\begin{aligned}
&\nabla f_t^{\mathrm{pred}}(M) - \nabla f_t(M \mid \widehat{G}_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{v}}_t^{\mathrm{nat}}) \\
&= \underbrace{\mathsf{D}\mathbf{v}_t^{\mathrm{pred}} \cdot \Big((\nabla \ell_t)(\mathbf{v}_t^{\mathrm{pred}}(M)) - (\nabla \ell_t)(\widehat{\mathbf{v}}_t(M))\Big)}_{(a)} \\
&+ \underbrace{\mathsf{D}(\mathbf{v}_t^{\mathrm{pred}} - \widehat{\mathbf{v}}_t)(\nabla \ell_t)(\widehat{\mathbf{v}}_t(M))}_{(b)}
\end{aligned}$$

We can bound the first term via

$$\begin{aligned}
\|(a)\|_{\mathrm{op}} &\overset{(i)}{\le} \left\|\mathsf{D}\mathbf{v}_t^{\mathrm{pred}}(M)\right\|_{\mathrm{op}} \cdot \beta_{\mathrm{loss}} \cdot \left\|\mathbf{v}_t^{\mathrm{pred}}(M) - \widehat{\mathbf{v}}_t(M)\right\|_{\mathrm{op}} \\
&\overset{(ii)}{\le} \big(2\sqrt{m}R_{G_\star}R_{\mathrm{nat}}\big) \cdot \beta_{\mathrm{loss}} \cdot (4R_{\mathcal{M}}R_{\mathrm{nat}}\epsilon_G) \\
&= 8\beta_{\mathrm{loss}}\sqrt{m}R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}^2\epsilon_G
\end{aligned}$$

where (i) uses smoothness of the loss, (ii) uses Lemma I.1. To bound term (b), we use the Lipschitzness from Assumption 2 to bound the norm of the gradient:

$$\|(\nabla\ell)(\widehat{\mathbf{v}}_t(M))\|_2 \le L\max\{1, \|\widehat{\mathbf{v}}_t(M)\|\} \le 6LR_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}} \quad \text{(by Lemma H.6)}$$

Hence, from Lemma I.1,

$$\begin{aligned}\|(b)\|_{\mathrm{op}} &\le 6LR_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}} \cdot \left\|\mathsf{D}(\mathbf{v}_t^{\mathrm{pred}} - \widehat{\mathbf{v}}_t)\right\|_{\mathrm{op}} \\ &\le 6LR_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}} \cdot 2\sqrt{m}\epsilon_G R_{\mathrm{nat}} = 12\sqrt{m}LR_{G_\star}R_{\mathrm{nat}}^2 R_{\mathcal{M}}\epsilon_G\end{aligned}$$

Hence, we conclude that

$$\|\nabla f_t^{\mathrm{pred}}(M) - \nabla f_t(M \mid \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}, \widehat{\mathbf{u}}_{1:t}^{\mathrm{nat}})\| \le \underbrace{\sqrt{m}R_{G_\star}R_{\mathcal{M}}R_{\mathrm{nat}}^2(8\beta_{\mathrm{loss}} + 12)}_{:=C_{\mathrm{approx}}} \cdot \epsilon_G$$

■

I.2.3. SMOOTHNESS, STRONG CONVEXITY, LIPSCHITZ (LEMMAS F.4, F.5/F.5B, AND F.6)

We begin by checking verifying the smoothness bound, which we recall from Section F.1:

Lemma F.4 *Under Condition D.1, $f_t^{\mathrm{pred}}(M)$ are $4\beta_f$-smooth, for β_f as in Lemma E.3.*

Proof The proof follows by modifying Lemma E.3b, replaced $\mathsf{D}\mathbf{v}_t^\star$ with $\mathsf{D}\mathbf{v}_t^{\mathrm{pred}}$. By Lemma I.1 part (a), we bound the operator norm of these differentials by twice the corresponding bound in Lemma E.3b, incurring an additional factor of four in the final result. ■

Next, we check Lipschitznes:

Lemma F.6 (Lipschitzness: Unknown & Strongly Convex) *Recall the Lipschitz constant L_f from Lemma C.5. Then under Condition D.1, $f_t^{\mathrm{pred}}(M)$ is $4L_f$-Lipschitz, $f_t^{\mathrm{pred}}[M_{t:t-h}]$ is $4L_f$ coordinate Lipschitz. Moreover, $\max_{M\in\mathcal{M}} \|\nabla \widehat{f}_t(M; \widehat{G}, \widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}})\|_2 \le 4L_f$.*

Proof [Proof of Lemma F.6] We prove the general stabilized case. Recall that for the known-system setting, the losses $f_t(| \; G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}})$ and $F_t[|| \; G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}$ are $L_f := L\sqrt{m}R_{\mathrm{nat}}^2 R_{G_\star}^2 R_{\mathcal{M}}$-Lipschitz and L_f-coordinate Lipschitz, respectively. Under Condition D.1b, we have that $\|\widehat{G}_{\mathrm{ex}\to(y,u)}\|_{\ell_1,\mathrm{op}} \le 2R_{G_\star}$, and moreover, by Lemma D.1b, we have that $\|\widehat{\mathbf{v}}_s^{\mathrm{nat}}\|_2, \|\widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}}\|_2 \le 2R_{\mathrm{nat}}$ for all $s \in [t]$, Hence, repeating the computation of the known-system Lipchitz constant in Lemma C.5, but with inflated norms of $\widehat{\mathbf{u}}_{1:t}^{\mathrm{nat}}$ adn $\widehat{\mathbf{y}}_{1:t}^{\mathrm{nat}}$, we find that f_t^{pred} (resp. F_t^{pred}) are $\overline{L}_f := 4L_f$-Lipschitz (resp. -coordinate Lipschitz). ■

Finally, we verify strong convexity in this setting. The following subsumes Lemma F.5:

Lemma F.5b (Strong Convexity: Unknown Stabilized System) *Consider the stabilized setting, with $\alpha_{f;m}$ as in Proposition E.2b. Suppose further that the conditions of that proposition hold, and in addition,*

$$\epsilon_G \le \frac{1}{9R_{\mathcal{M}}R_{\mathrm{nat}}R_{G_\star}}\sqrt{\frac{\alpha_{f;m}}{m\alpha_{\mathrm{loss}}}}$$

Then, the functions are $f^{\mathrm{pred}}_{t;k}$ are $\alpha_{f;m}/4$-strongly convex. Analogously, replacing $\alpha_{f;m}$ by α_f in the stable setting, the functions $f^{\mathrm{pred}}_{t;k}$ are $\alpha_f/4$ strong convex for α_f as in Proposition E.2.

Proof Let us consider the stabilized case; the stable case is identical. Proposition E.2b (proved in Appendix J.2.1) follows from Theorem 11, and an can be used to prove the following intermediate bound:

$$\mathbb{E}\left[\left\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \ge \frac{\alpha_{f;m}}{\alpha_{\mathrm{loss}}}\|M\|_{\mathrm{F}}^2, \quad \forall M = (M^{[i]})_{i=0}^{m-1}$$

To deduce our desired strong convexity bound, it suffices to show that $M = (M^{[i]})_{i=0}^{m-1}$,

$$\mathbb{E}[\|\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}^{\mathrm{nat}}_t\|_2^2 \mid \mathcal{F}_{t-k}] \ge \frac{\alpha_{f;m}}{4\alpha_\ell}\|M\|_{\mathrm{F}}^2,$$

with an additional slack factor of $1/2$. To begin, note the elementary inequality

$$\begin{aligned}\|v+w\|_2^2 = \|v\|_2^2 + \|w\|_2^2 - 2\|v\|\|w\| &\ge \|v\|_2^2 + \|w\|_2^2 - 2(\frac{1}{2}\cdot\frac{1}{2}\|v\|_2^2 + \frac{1}{2}\cdot 2\|w\|_2^2) \\ &= \|v\|_2^2/2 - \|w\|_2^2\end{aligned}$$

This yields

$$\begin{aligned}&\mathbb{E}[\|\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}^{\mathrm{nat}}_t\|_2^2 \mid \mathcal{F}_{t-k}] \\ &= \mathbb{E}[\|(\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) + \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t\|_2^2 \mid \mathcal{F}_{t-k}] \\ &\ge \frac{1}{2}\mathbb{E}[\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t\|_2^2 \mid \mathcal{F}_{t-k}] \\ &\qquad - \mathbb{E}[\|\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t)\|_2^2 \mid \mathcal{F}_{t-k}] \\ &\ge \frac{\alpha_{f;m}\|M\|_{\mathrm{F}}^2}{2\alpha_\ell} - \underbrace{\mathbb{E}[\|\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t)\|_2^2 \mid \mathcal{F}_{t-k}]}_{(i)}.\end{aligned}$$

Moreover, by Lemma I.1 part (b), we have

$$\|\mathbf{v}^{\mathrm{pred}}_t(M) - \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t)\|_2^2 \le (3R_{\mathcal{M}}R_{\mathrm{nat}}R_{G_\star}\sqrt{m}\epsilon_G)^2.$$

Hence, if $\epsilon_G \le \frac{\sqrt{\alpha_{f;m}/\alpha_\ell}}{9R_{\mathcal{M}}R_{\mathrm{nat}}R_{G_\star}\sqrt{m}}$, then the term (i) is bounded by $(i) \le \frac{\alpha_{f;m}\|M\|_{\mathrm{F}}^2}{4\alpha_\ell}$, which concludes the proof. ■

I.3. Proof of Proposition F.8 (Approximation Error)

We prove the proposition in the general stabilized setting, where assume the corresponding Condition D.1b holds. Recall the set $\mathcal{M}_0 := \mathcal{M}(m_0, R_{\mathcal{M}}/2)$, and consider a comparator

$$M_\star \in \arg\inf_{M\in\mathcal{M}_0}\sum_{t=N+m+2h+1}^{T}\ell_t(\mathbf{y}^M_t, \mathbf{u}^M_t)$$

We summarize the conditions of the Proposition F.8 as follows:

Condition I.1 (Conditions for Proposition F.8) *We assume that (a)* $\epsilon_G \le 1/R_{\mathcal{M}}$*, (b)* $m \ge 2m_0 + h$*, and (c)* $\psi_{G_\star}(h+1) \le R_{G_\star}/T$.

Note that the first condition holds from from Condition D.1/D.1b., and the secnd two from the definition of the algorithm paramaters. The proof has two major steps. We begin with the following claim, which reduces the proof to controllng the differences $\|\mathbf{u}^{\mathrm{ex}}_t(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t}) - \mathbf{u}^{\mathrm{ex}}_t(M_\star \mid \mathbf{y}^{\mathrm{nat}}_{1:t})\|_2$ between algorithmic inputs on the $\widehat{\mathbf{y}}^{\mathrm{nat}}$ sequence using M_{apprx}, and on the $\mathbf{y}^{\mathrm{nat}}$ sequence using $M_\star$:

Lemma I.3 *We have the bound:*

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 3LR_{G_\star}^2 R_{\mathcal{M}} R_{\mathrm{nat}} \sum_{t=N+m+h+1}^{T} \max_{s=t-h}^{t} \|\mathbf{u}^{\mathrm{ex}}_s(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}_s(M_\star \mid \mathbf{y}^{\mathrm{nat}}_{1:s})\|_2$$

The above lemma is proven in Appendix I.3.1.

We will neglect the first $m_0 + 2h$ terms in the above sum. Specifically, defining $N_1 = N + m + 3h + 1 + m_0$, we have

$$\begin{aligned}(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) &\le 3LR_{G_\star}^2 R_{\mathcal{M}} R_{\mathrm{nat}} \sum_{t=N_1}^{T} \max_{s=t-h}^{t} \|\mathbf{u}^{\mathrm{ex}}_s(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}_s(M_\star \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s})\|_2 \\ &\quad + 3LR_{G_\star}^2 R_{\mathcal{M}}(m_0+2h) \max_{s=N+m+h+1}^{N_1} \|\mathbf{u}^{\mathrm{ex}}_s(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}_s(M_\star \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s})\|_2.\end{aligned}$$

Moreover, by the triangle inequality and Lemmas H.6 and C.1b

$$\|\mathbf{u}^{\mathrm{ex}}_s(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}_s(M_\star \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s})\|_2 \le 3R_{\mathcal{M}}R_{\mathrm{nat}},$$

giving

$$\begin{aligned}(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) &\le 3LR_{G_\star}^2 R_{\mathcal{M}} R_{\mathrm{nat}} \sum_{t=N_1}^{T} \max_{s=t-h}^{t} \|\mathbf{u}^{\mathrm{ex}}_s(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:s}) - \mathbf{u}^{\mathrm{ex}}_s(M_\star \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:s})\|_2 \\ &\quad + 9L(m_0+2h)R_{G_\star}^2 R_{\mathcal{M}}^2 R_{\mathrm{nat}}. \end{aligned} \tag{I.4}$$

We now turn to bounding these $\mathbf{u}^{\mathrm{ex}}$ differences, which is the main source of difficulty in the proof of Proposition F.8. The next lemma is proven in Appendix I.3.2:

Lemma I.4 *Under Condition I.1, there exists an* $M_{\mathrm{apprx}} \in \mathcal{M}(m, R_{\mathcal{M}})$*, depending only on* ϵ_G *and* $M_\star$*, such that for all* $t \ge m+1$ *and* $\tau > 0$,

$$\begin{aligned}&\|\mathbf{u}^{\mathrm{ex}}_t(M_\star \mid \mathbf{y}^{\mathrm{nat}}_{1:t}) - \mathbf{u}^{\mathrm{ex}}_t(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t})\|_2 \le \\ &\underbrace{\overline{R}_{\mathbf{u}} R_{\mathcal{M}} \psi_{G_\star}(h+1)}_{\text{(truncation term)}} + \underbrace{R_{\mathcal{M}}^2 \epsilon_G^2 \left(\frac{R_{\mathrm{nat}} R_{\mathcal{M}} + \tau^{-1}}{2}\right)}_{\text{(estimation term)}} + \underbrace{\frac{\tau}{2} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}^{\mathrm{ex}}_j(\mathbf{M}_j - M_{\mathrm{apprx}}) \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2^2}_{\text{(movement term)}}\end{aligned} \tag{I.5}$$

From the above lemma and Eq. (I.4), and reparametrizing $\tau \leftarrow 2\tau/3LR_{G_\star}^2 R_{\mathcal{M}} R_{\mathrm{nat}}$, and bounding $m_0 + 2h \le m$, have

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 3L \cdot R_{\mathcal{M}}^2 R_{G_\star}^2 R_{\mathrm{nat}} \left(T\overline{R}_{\mathbf{u}}\psi_{G_\star}(h+1) + 3m + R_{\mathrm{nat}} R_{\mathcal{M}} T \epsilon_G^2 \cdot \left(1 + LR_{G_\star}^2 \tau^{-1}\right)\right) + \tau \sum_{t=N_1}^{T} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2$$

For $\psi_{G_\star}(h+1) \le R_{G_\star}/T$, the above simplifies to

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 3LR_{\mathcal{M}}^3 R_{G_\star}^2 R_{\mathrm{nat}}^2 T \epsilon_G^2 \left(1 + \frac{LR_{G_\star}^2}{\tau}\right) + 3LR_{\mathcal{M}}^2 R_{G_\star}^2 R_{\mathrm{nat}} (\overline{R}_{\mathbf{u}} R_{G_\star} + 3m) + \tau \sum_{t=N_1}^{T} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2$$

Moreover, we can crudely bound

$$\begin{aligned}\sum_{t=N_1}^{T} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2 &\le (m_0+h) \sum_{t=N_1-m_0-h}^{T} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2 \\ &= (m_0+h) \sum_{t=N+m+2h+1}^{T} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2\end{aligned}$$

Thus, again reparametrizing $\tau \leftarrow \tau/(m_0+h)$, and bounding $m_0 + h \le m$,

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 3LR_{\mathcal{M}}^3 R_{G_\star}^2 R_{\mathrm{nat}}^2 T \epsilon_G^2 \left(1 + \frac{LmR_{G_\star}^2}{\tau}\right) + 3LR_{\mathcal{M}}^2 R_{G_\star}^2 R_{\mathrm{nat}} (\overline{R}_{\mathbf{u}} R_{G_\star} + 3m) + \tau \sum_{t=N+m+h+1}^{T} \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2. \tag{I.6}$$

Finally, let us upper bound

$$\begin{aligned}\left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2 &\le \max_{j=0}^{m-1} \|\widehat{\boldsymbol{\eta}}_j^{\mathrm{nat}}\|_2 \cdot \|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\ell_1,\mathrm{op}} \\ &\le 2R_{\mathrm{nat}} \cdot \|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\ell_1,\mathrm{op}} && \text{(Lemma D.1b)} \\ &\le 2R_{\mathrm{nat}} \cdot \sqrt{m}\,\|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\mathrm{F}}. && \text{(Lemma H.2)}\end{aligned}$$

Thus, $\tau \leftarrow \tau/4R_{\mathrm{nat}}^2 m$, we obtain

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 3LR_{\mathcal{M}}^3 R_{G_\star}^2 R_{\mathrm{nat}}^2 T \epsilon_G^2 \left(1 + \frac{Lm^2 R_{\mathrm{nat}}^2 R_{G_\star}^2}{\tau}\right) + 3LR_{\mathcal{M}}^2 R_{G_\star}^2 R_{\mathrm{nat}} (\overline{R}_{\mathbf{u}} R_{G_\star} + 3m) \tag{I.7}$$

$$+ \tau \sum_{t=N+m+h+1}^{T} \|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2.$$

Finally, let us crudely bound the abouve by

$$(\widehat{\mathbf{y}}^{\mathrm{nat}}\text{-approx error}) \le 36m^2 R_{\mathcal{M}}^3 R_{\mathrm{nat}}^4 R_{G_\star}^4 T\epsilon_G \max\{L, L^2/\tau\} + 9LR_{\mathcal{M}}^2 R_{G_\star}^2 R_{\mathrm{nat}}(\overline{R}_{\mathbf{u}} R_{G_\star} + m) + \tau \sum_{t=N+m+h+1}^{T} \|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2 .$$

as needed.

□

I.3.1. PROOF OF LEMMA I.3

Let $M_{\mathrm{apprx}}, M_\star \in \mathcal{M}(m, R_{\mathcal{M}})$, and recall the shorthand

$$\mathbf{v}_t^\star(M) := \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}), \quad \mathbf{v}_t^{\mathrm{pred}}(M) := \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}})$$

Then,

$$\begin{aligned}\left| f_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - f_t(M_\star \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}) \right| &= \left| \ell_t(\mathbf{v}_t^{\mathrm{pred}}(M_{\mathrm{apprx}})) - \ell_t(\mathbf{v}_t^\star(M_\star)) \right| \\ &\le L \max\{\underbrace{\|\mathbf{v}_t^{\mathrm{pred}}(M_{\mathrm{apprx}})\|_2}_{(a)}, \underbrace{\|\mathbf{v}_t^\star(M_\star)\|_2, 1\}}_{(b)} \|\mathbf{v}_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - \mathbf{v}_t^\star(M_\star)\|_2.\end{aligned}$$

From Lemma C.1b, we have $(b) \le 2R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}$. Moreover, combining with Lemma H.6, a similar argument lets us bound $(a) \le R_{\mathrm{nat}} + 2R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} = 3R_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}}$. Since these upper bounds are all assumed to be greater than one,

$$\left| f_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - f_t(M_\star \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}) \right| \le 3LR_{G_\star} R_{\mathcal{M}} R_{\mathrm{nat}} \cdot \|\mathbf{v}_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - \mathbf{v}_t^\star(M_\star)\|_2. \tag{I.8}$$

Unfolding

$$\begin{aligned}\|\mathbf{v}_t^{\mathrm{pred}}(M_{\mathrm{apprx}}) - \mathbf{v}_t^\star(M_\star)\|_2 &= \left\| \sum_{i=0}^{h} G_{\mathrm{ex}\to(y,u)}^{[i]} (\mathbf{u}_{t-i}^{\mathrm{ex}}(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:t-i}^{\mathrm{nat}}) - \mathbf{u}_{t-i}^{\mathrm{ex}}(M_\star \mid \boldsymbol{\eta}_{1:t-i}^{\mathrm{nat}})) \right\|_2 \\ &\le R_{G_\star} \max_{i=0}^{h} \left\| \mathbf{u}_{t-i}^{\mathrm{ex}}(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:t-i}^{\mathrm{nat}}) - \mathbf{u}_{t-i}^{\mathrm{ex}}(M_\star \mid \boldsymbol{\eta}_{1:t-i}^{\mathrm{nat}})) \right\|_2 .\end{aligned}$$

Combining with Eq. (I.8) gives the bound. □

I.3.2. PROOF OF LEMMA I.4

For simplicity, let us use $G_\star, \widehat{G}$ for $G_{\mathrm{ex}\to\eta}, \widehat{G}_{\mathrm{ex}\to\eta}$. Since $M_\star \in \mathcal{M}(m_0, R_{\mathcal{M}}/2)$, we have $M_\star^{[i]} = 0$ for all $i \ge m_0$. Therefore, we can write

$$
\begin{aligned}
&\mathbf{u}_t^{\mathrm{ex}}(M_\star \mid \mathbf{y}_{1:t}^{\mathrm{nat}}) && \text{(I.9)}\\
&= \sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \boldsymbol{\eta}_s^{\mathrm{nat}} \\
&= \sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \left(\widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}} + (\boldsymbol{\eta}_s^{\mathrm{nat}} - \widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}})\right) \\
&= \sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \left(\widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}} + \left(\left(\boldsymbol{\eta}_s^{\mathsf{alg}} - \sum_{j=1}^{s} G_\star^{[s-j]} \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}}\right) - \left(\boldsymbol{\eta}_s^{\mathsf{alg}} - \sum_{j=s-h}^{s} \widehat{G}^{[j]} \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}}\right)\right)\right) \\
&= \underbrace{\left(\sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \left(\widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}} + \left(\sum_{j=s-h}^{s} \left(\widehat{G}^{[j]} - G_\star^{[j]}\right) \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}}\right)\right)\right)}_{:=\mathbf{u}_t^{\mathrm{main}}} \\
&\qquad - \underbrace{\sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \sum_{1 \le j < s-h} G_\star^{[s-j]} \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}}}_{:=\mathbf{u}_t^{\mathrm{trunc}}} && \text{(I.10)}
\end{aligned}
$$

Here, $\mathbf{u}_t^{\mathrm{trunc}}$ is a lower order truncation term:

Claim I.5 *For $t \ge N + m + 1$, we have that $\|\mathbf{u}_t^{\mathrm{trunc}}\|_2 \le \overline{R}_{\mathbf{u}} R_{\mathcal{M}} \psi_{G_\star}(h+1)$.*

Proof We have that $\|\mathbf{u}_j^{\mathrm{ex},\mathsf{alg}}\|_2 \le \overline{R}_{\mathbf{u}}$ by Lemma D.1b. This gives

$$
\begin{aligned}
\left\|\mathbf{u}_t^{\mathrm{trunc}}\right\| &= \left\| \sum_{s=(t-m_0+1)_+}^{t} M_\star^{[t-s]} \sum_{1 \le j < s-h-1} G_\star^{[s-j]} \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}} \right\| \\
&\le R_{\mathcal{M}} \overline{R}_{\mathbf{u}} \sum_{1 \le j < s-h} \|G_\star^{[s-j]}\| \;\le\; R_{\mathcal{M}} \overline{R}_{\mathbf{u}} \psi_{G_\star}(h+1).
\end{aligned}
$$

■

To bound the dominant term $\mathbf{u}_t^{\mathrm{main}}$, we express $\mathbf{u}_j^{\mathsf{alg}}$ in terms of $\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}$ and the controller $M_\star$:

$$
\begin{aligned}
\mathbf{u}_j^{\mathsf{alg}} = \mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) &= \sum_{q=j-m_0+1}^{j} \mathbf{M}_j^{[j-q]} \widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}} \\
&= \underbrace{\sum_{q=j-m_0+1}^{j} M_\star^{[j-q]} \widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}}}_{(b)} + \sum_{q=j-m+1}^{j} (\mathbf{M}_j^{[j-q]} - M_\star^{[j-q]}),
\end{aligned}
$$

where sum only over $q \in \{(j-m_0)_+, \dots, j\}$ in the bracketed term (b) because $M_\star^{[n]} = 0$ for $n > m_0$ since $M_\star \in \mathcal{M}(m_0, R_\mathcal{M}/2)$. Next, the equalities

$$\begin{aligned}
\mathbf{u}_j^{\mathrm{ex},\mathsf{alg}} &= \mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) \\
&= \mathbf{u}_j^{\mathrm{ex}}(M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) + \mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) - \mathbf{u}_j^{\mathrm{ex}}(M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) \\
&= \mathbf{u}_j^{\mathrm{ex}}(M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) + (\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}) \\
&= \left(\sum_{q=j-m_0+1}^{j} M_\star^{[j-q]} \widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}} \right) + \mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}}),
\end{aligned}$$

and introducing the shorthand $\Delta_G^{[j]} := \widehat{G}^{[j]} - G_\star^{[j]}$, we can further develop

$$\begin{aligned}
\mathbf{u}_t^{\mathrm{main}} &:= \sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \left(\widehat{\mathbf{y}}_s^{\mathrm{nat}} + \sum_{j=s-h}^{s} \Delta_G^{[j]} \mathbf{u}_j^{\mathrm{ex},\mathsf{alg}} \right) \\
&:= \underbrace{\sum_{s=t-m_0+1}^{t} M_\star^{[t-s]} \left(\widehat{\mathbf{y}}_s^{\mathrm{nat}} + \sum_{j=s-h}^{s} \sum_{q=j-m_0+1}^{j} \Delta_G^{[j]} M_\star^{[j-q]} \widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}} \right)}_{:= \mathbf{u}_t^{\mathrm{apprx}}} \\
&\quad + \underbrace{\sum_{s=t-m_0+1}^{t} \sum_{j=s-h}^{s} \sum_{q=j-m+1}^{j} M_\star^{[t-s]} \Delta_G^{[j]} \mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})}_{:= \mathbf{u}_t^{\mathrm{err}}} . \qquad \text{(I.11)}
\end{aligned}$$

Here, the input $\mathbf{u}_t^{\mathrm{apprx}}$ is respresents the part of the input which can be represented as $\mathbf{u}_t^{\mathrm{apprx}} = \mathbf{u}_t^{\mathrm{ex}}(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}_t^{\mathrm{nat}})$ for some $M_{\mathrm{apprx}} \in \mathcal{M}(m, R_\mathcal{M})$; the remaining error term, $\mathbf{u}_t^{\mathrm{err}}$, will be bounded shortly thereafter.

Claim I.6 (Existence of a good comparator) *Define the controller*

$$M_{\mathrm{apprx}}^{[i]} = M_\star^{[i]} I_{i \le m_0 - 1} + \sum_{a=0}^{m_0-1} \sum_{b=0}^{h} \sum_{c=0}^{m_0-1} M_\star^{[a]} \Delta_G^{[b]} M_\star^{[c]} \mathbb{I}_{a+b+c=i},$$

which depends only of $M_\star$ and ϵ_G. Then,

1. *We have the identity*

$$\mathbf{u}_t^{\mathrm{apprx}} = \mathbf{u}_t^{\mathrm{ex}}(M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:t}^{\mathrm{nat}}), \quad \forall t \ge N, \qquad \text{(I.12)}$$

2. $\|M_{\mathrm{apprx}} - M_\star\|_{\ell_1,\mathrm{op}} \le \|M_\star\|_{\ell_1,\mathrm{op}}^2 \epsilon_G \le R_\mathcal{M}^2 \epsilon_G / 4$.

3. *If $m \ge 2m_0 - 1 + h$ and $\epsilon_G \le \frac{1}{R_\mathcal{M}}$ (as ensured by Condition I.1), then $M_{\mathrm{apprx}} \in \mathcal{M}(m, R_\mathcal{M})$*

Proof To verify Eq. (I.12),

$$\begin{aligned}
\mathbf{u}_t^{\mathrm{apprx}} &= \sum_{s=t-m_0+1}^{t} M_\star^{[t-s]}\widehat{\boldsymbol{\eta}}_s^{\mathrm{nat}} + \sum_{s=t-m_0+1}^{t}\sum_{s=j-h}^{s}\sum_{q=j-m_0+1}^{j} M_\star^{[t-s]}\Delta_G^{[s-j]}M_\star^{[j-q]}\widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}} \\
&= \sum_{i=0}^{m_0-1} M_\star^{[i]}\widehat{\boldsymbol{\eta}}_{t-q}^{\mathrm{nat}} + \sum_{a=0}^{m_0-1}\sum_{b=0}^{h}\sum_{c=0}^{m_0-1} M_\star^{[a]}\Delta_G^{[b]}M_\star^{[c]}\widehat{\boldsymbol{\eta}}_{t-(a+b+c)}^{\mathrm{nat}} \\
&= \sum_{i=0}^{2(m_0-1)+h}\left(M_\star^{[i]}I_{i\le m_0-1} + \sum_{a=0}^{m_0-1}\sum_{b=0}^{h}\sum_{c=0}^{m_0-1} M_\star^{[a]}\Delta_G^{[b]}M_\star^{[c]}\mathbb{I}_{a+b+c=i}\right)\widehat{\boldsymbol{\eta}}_{t-i}^{\mathrm{nat}} \\
&= \sum_{i=0}^{2(m_0-1)+h} M_{\mathrm{apprx}}^{[i]}\widehat{\boldsymbol{\eta}}_{t-i}^{\mathrm{nat}},
\end{aligned}$$

Next, since $\|M_\star\| \le R_{\mathcal{M}}/2$,

$$\|M_{\mathrm{apprx}} - M_\star\|_{\ell_1,\mathrm{op}} \le \sum_{a=0}^{m_0-1}\sum_{b=0}^{h}\sum_{c=0}^{m_0-1}\|M_\star^{[a]}\|_{\mathrm{op}}\|\Delta_G^{[b]}\|_{\mathrm{op}}\|M_\star^{[c]}\|_{\mathrm{op}} \le \|M_\star\|_{\ell_1,\mathrm{op}}^2\epsilon_G \le \frac{R_{\mathcal{M}}^2\epsilon_G}{4},$$

which verifies point 2. Therefore, for $\epsilon_G \le 1/R_{\mathcal{M}}$,

$$\|M_{\mathrm{apprx}}\| \le \|M_\star\|_{\ell_1,\mathrm{op}} + \|M_{\mathrm{apprx}} - M_\star\|_{\ell_1,\mathrm{op}} \le \frac{R_{\mathcal{M}}}{2} + \frac{R_{\mathcal{M}}^2\varepsilon}{4} \le R_{\mathcal{M}}.$$

Moreover, by assumption on $m \ge 2m_0+h-1$, we have $M_{\mathrm{apprx}}^{[i]} \ge 0$ for $i > m \ge m_0 2(m_0-1)+h$.
■

Lastly, we control the error term. We shall do this incrementally via two successive claims. First, we "re-center" $\mathbf{u}_t^{\mathrm{err}}$ arround the comparator M_0, rather than $M_\star$, and uses AM-GM to isolate terms $\|\mathbf{M}_j - M_{\mathrm{apprx}}\|_{\mathrm{F}}^2$:

Claim I.7 *For $m \ge 2m_0 - 1 + h$, the following bound holds for all $\tau > 0$*

$$\|\mathbf{u}_t^{\mathrm{err}}\|_2 \le \frac{R_{\mathrm{nat}}R_{\mathcal{M}}^3\epsilon_G^2}{2} + \frac{R_{\mathcal{M}}^2\epsilon_G^2}{2\tau} + \frac{\tau}{2}\max_{j=t-m_0+1-h}^{t}\left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2^2.$$

Proof Using $\|\widehat{\boldsymbol{\eta}}_q^{\mathrm{nat}}\|_2 \le 2R_{\mathrm{nat}}$ (Lemma D.1b) and $\|M_\star\|_{\ell_1,\mathrm{op}} \le R_{\mathcal{M}}/2$ by assumption,

$$\begin{aligned}
\|\mathbf{u}_t^{\mathrm{err}}\|_2 &= \left\|\sum_{s=t-m_0+1}^{t}\sum_{j=s-h}^{s} M_\star^{[t-s]}\Delta_G^{[j]}\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2 \\
&\le R_{\mathcal{M}}\epsilon_G \max_{j=t-m_0+1-h}^{t}\left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_\star \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2 \\
&\le R_{\mathcal{M}}\epsilon_G \max_{j=t-m_0+1-h}^{t}\left(\left\|\mathbf{u}_j^{\mathrm{ex}}(M_\star - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2 + \left\|\mathbf{u}_j^{\mathrm{ex}}(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}_{1:j}^{\mathrm{nat}})\right\|_2\right).
\end{aligned}$$

Next, from (Lemma D.1b) and Claim I.6, we can bound

$$\begin{aligned}\left\|\mathbf{u}^{\mathrm{ex}}_j(M_\star - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2 &\le \|M_\star - M_{\mathrm{apprx}}\|_{\ell_1,\mathrm{op}} \max_{s \le j} \|\widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_s\|_2 \\ &\le 2R_{\mathrm{nat}} \|M_\star - M_{\mathrm{apprx}}\|_{\ell_1,\mathrm{op}} \\ &\le \frac{R_{\mathrm{nat}} R_{\mathcal{M}}^2 \epsilon_G}{2}.\end{aligned}$$

This yield

$$\begin{aligned}\|\mathbf{u}^{\mathrm{err}}_t\|_2 &\le \frac{R_{\mathrm{nat}} R_{\mathcal{M}}^3 \epsilon_G^2}{2} + R_{\mathcal{M}} \epsilon_G \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}^{\mathrm{ex}}_j(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2 \\ &\le \frac{R_{\mathrm{nat}} R_{\mathcal{M}}^3 \epsilon_G^2}{2} + \frac{R_{\mathcal{M}}^2 \epsilon_G^2}{2\tau} + \frac{\tau}{2} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}^{\mathrm{ex}}_j(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2^2 \\ &\le R_{\mathcal{M}}^2 \epsilon_G^2 \left(\frac{R_{\mathrm{nat}} R_{\mathcal{M}} + \tau^{-1}}{2}\right) + \frac{\tau}{2} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}^{\mathrm{ex}}_j(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2^2.\end{aligned}$$

■

Putting things together, we have that

$$\begin{aligned}&\|\mathbf{u}^{\mathrm{ex}}_t(M_\star \mid \mathbf{y}^{\mathrm{nat}}_{1:t}) - \mathbf{u}^{\mathrm{ex}}_t(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t})\|_2 \\ &\quad = \|\mathbf{u}^{\mathrm{trunc}}_t + \mathbf{u}^{\mathrm{main}}_t - \mathbf{u}^{\mathrm{ex}}_t(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t})\|_2 && \text{(Eq. (I.10))} \\ &\quad = \|\mathbf{u}^{\mathrm{trunc}}_t + \mathbf{u}^{\mathrm{err}}_t + \mathbf{u}^{\mathrm{apprx}}_t - \mathbf{u}^{\mathrm{ex}}_t(M_{\mathrm{apprx}} \mid \widehat{\mathbf{y}}^{\mathrm{nat}}_{1:t})\|_2 && \text{(Eq. (I.11))} \\ &\quad = \|\mathbf{u}^{\mathrm{trunc}}_t + \mathbf{u}^{\mathrm{err}}_t\|_2 && \text{(by Claim I.6)} \\ &\quad \le \|\mathbf{u}^{\mathrm{trunc}}_t\|_2 + \|\mathbf{u}^{\mathrm{err}}_t\|_2 \\ &\le \overline{R}_{\mathbf{u}} R_{\mathcal{M}} \psi_{G_\star}(h+1) + R_{\mathcal{M}}^2 \epsilon_G^2 \left(\frac{R_{\mathrm{nat}} R_{\mathcal{M}} + \tau^{-1}}{2}\right) \\ &\qquad + \frac{\tau}{2} \max_{j=t-m_0+1-h}^{t} \left\|\mathbf{u}^{\mathrm{ex}}_j(\mathbf{M}_j - M_{\mathrm{apprx}} \mid \widehat{\boldsymbol{\eta}}^{\mathrm{nat}}_{1:j})\right\|_2^2. \\ &= \text{(RHS of Eq. (I.5))},\end{aligned}$$

as needed. □

Appendix J. Establishing Strong Convexity

This appendix is devoted to establishing strong convexity of the DRC and DRC-EX parameterizations under semi-adversarial noise, described by Assumption 6b in the previous appendix. The organization is as follows:

1. Appendix J.1 introduces the necessary preliminaries to state our bound, including the Markov operators of the dynamics that arise from an internal stabilizing controller, and the notion of the Z-transform.

2. Appendix J.2 presents Theorem 11, which describes the strong convexity of internally stabilized systems in terms of certain functionals of the Z-transforms of relevant Markov operators. Combining with Proposition J.6 which characterzes the behavior of these functionals,

we this section concludes with the proof of Proposition E.2b. This section then specializes this bounds for systems with internal controllers which are given by static feedback (Appendix J.2.2), and exact observer-feedback (Appendix J.2.3).

3. Appendix J.3 adresses the proof of Theorem 11.
4. Appendix J.4 establishes the proof of Proposition E.2. It borrows one lemam from the proof of Theorem 11, but bypasses the Z-transform to establish bounds via elementary principles.
5. Appendix J.5 proves Proposition J.6 via complex-analytic arguments. The focus is to obtain polynomial dependence in the horizon parameters, and no care is paid to specifying system-dependent constants.

J.1. Strong Convexity Preliminaries

Transfer Functions and Z-Transforms The strong convexity modulus is most succintly described in the Fourier domain, where we work with Markov operators and their Z-tranfsorms. We recall the definition of an abstract Markov operator as follows:

Definition G.1 (Markov Operator) *Let $\mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the set of Markov operators $G = (G^{[i]})_{i \geq 0}$ with $G^{[i]} \in \mathbb{R}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, such that $\|G\|_{\ell_1,\mathrm{op}} < \infty$. Given a system (A, B, C, D) with input dimension d_{in} and output dimension d_{o}, we let $G = \mathsf{Transfer}(A, B, C, D) \in \mathscr{G}_{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the system $G^{[0]} = D$ and $G^{[i]} = CA^{i-1}B$.*

We shall also use the notation

$$G^\top = \mathsf{Transfer}(A, B, C, D)^\top = \mathsf{Transfer}(A^\top, C^\top, B^\top, D^\top),$$

where $(A^\top, C^\top, B^\top, D^\top)$ is commonly referred to as the adjoint system. For an abstract Markov operator G, its Z-transform is the following power series:

Definition J.1 (Z-Transform) *For $G \in \mathscr{G}_{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, the* Z-transform *is the mapping from $\mathbb{C} \to \mathbb{C}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$*

$$\check{G}(z) : z \mapsto \sum_{i=0}^{n} G^{[i]} z^{-i}$$

For finite-order linear dynamical systems, the Z-transform can be expressed in closed form via:

Lemma J.2 *If $G = \mathsf{Transfer}(A, B, C, D)$, then $\check{G}(z) = D + C(zI - A)^{-1}B$.*

Closed Loop Dynamics: For stabilized systems, the relevant Markov operators that arise correspond to the closed-loop dyanics of the nominal system placed in feedback with the stabilizing controller π_0. From Lemma 3.2b, we recall the operators $G_{\mathrm{ex}\to(y,u)}$ and $G_{(w,e)\to\eta}$ which satisfy:

$$\begin{bmatrix} \mathbf{y}_t^{\mathsf{alg}} \\ \mathbf{u}_t^{\mathsf{alg}} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_t^{\mathrm{nat}} \\ \mathbf{u}_t^{\mathrm{nat}} \end{bmatrix} + \sum_{i=0}^{t-1} G_{\mathrm{ex}\to(y,u)}^{[i]} \mathbf{u}_{t-i}^{\mathrm{ex}}$$

and

$$\boldsymbol{\eta}_t^{\mathrm{nat}} = \sum_{i=1}^{t} G_{(w,e)\to\eta}^{[i]} \begin{bmatrix} \mathbf{w}_{t-i} \\ \mathbf{e}_{t-i} \end{bmatrix}.$$

The Markov operators in terms of which we bound the strong convexity modulus are as follows:

Definition J.3 (Markov Operators for Strong Convexity) *Recall the Markov operators*

$$\begin{aligned} G_{\mathrm{ex}\to(y,u)} &:= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl,ex}}, C_{\pi_0,\mathrm{cl}}, D_{\pi_0,\mathrm{cl,ex}}) \\ G_{(w,e)\to\eta} &:= \mathsf{Transfer}(A_{\pi_0,\mathrm{cl}}, B_{\pi_0,\mathrm{cl}}, C_{\pi_0,\mathrm{cl},\eta}, D_{\pi_0,\mathrm{cl},\eta}) \end{aligned}$$

from Definition 2.2b. We define the noise transfer function as

$$G_{\mathrm{noise}} = G_{(w,e)\to\eta}\Sigma_{\mathrm{noise}}^{\frac{1}{2}} \in \mathscr{G}^{d_\eta \times (d_x+d_u)},$$

where the above notation is short hand for $G_{\mathrm{noise}}^{[i]} = G_{(w,e)\to\eta}^{[i]}\Sigma_{\mathrm{noise}}^{\frac{1}{2}}$ for all i. Note that $G_{\mathrm{ex}\to(y,u)}$ has $d_{\mathsf{in}} = d_u$ and $d_{\mathsf{o}} = d_y + d_u$, whereas $G_{\mathrm{noise}}^\top$ has $d_{\mathsf{in}} = d_y$ and $d_{\mathsf{o}} = d_x + d_u$. We further define the function $\psi_{G_{\mathrm{ex}\to(y,u)}}$ and $\psi_{G_{\mathrm{noise}}}$ denote the corresponding decay functions, which are proper by Assumption 1b.

Here, $G_{\mathrm{ex}\to(y,u)}$ describes the dependence of $(\mathbf{y},\mathbf{u})$ on exogenous inputs $\mathbf{u}^{\mathrm{ex}}$, and G_{noise} is the transpose of the system which describes the effect that the noise in the system has on natures $\mathbf{y}_t^{\mathrm{nat}}$. Since $\mathbf{u}_t^{\mathrm{ex}}(M)$ is linear in natures y, G_{noise} needs to be sufficiently well conditioned (in a sense we will describe) to ensure strong convexity. Note that Σ_{noise} above need not be full-covariance, provided that it satisfies Assumption 6b. Moreover, since $f_t(M)$ depends on $\mathbf{u}_s^{\mathrm{ex}}(M)$ via the Markov operator $G_{\mathrm{ex}\to(y,u)}$, this operator also needs to be sufficiently well conditioned. [15]

J.2. Internally Stabilized Strong Convexity and Proof of Proposition E.2b

The relevant strong convexity parameter is bounded most precisely in terms of what we call "$\mathcal{H}$" functions, which describe the behavior of the Z-transform $\check{G}(z)$ of a Markov operator along the torus: $\mathbb{T} := \{e^{\iota\theta} \mid \theta \in [0, 2\pi]\}$:

Definition J.4 ($\mathcal{H}_{\min}$-Functional) *Let $d_{\mathsf{in}} \le d_{\mathsf{o}}$*[16]*, $G \in \mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$. We define the $\mathcal{H}_{\min}$ and $\mathcal{H}_\infty$ functionals as*

$$\mathcal{H}_{\min}[G] := \min_{\theta\in[0,2\pi]} \sigma_{d_{\mathsf{in}}}(\check{G}(e^{\iota\theta})) \quad \textit{and} \quad \|G\|_{\mathcal{H}_\infty} := \max_{\theta\in[0,2\pi]} \|\check{G}(e^{\iota\theta})\|_{\mathrm{op}}.$$

We will show that for h, k sufficiently large, the strong convexity parameter is lower bounded by $\gtrsim \mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}] \cdot \mathcal{H}_{\min}[G_{\mathrm{noise}}]$. Unfortunately, for certain pathological systems, one or both of these terms may vanish. To ensure fast rates for *all* systems, we will need a more refined notion:

Definition J.5 *Let $d_{\mathsf{in}} \le d_{\mathsf{o}}$, $G \in \mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, and let $\omega = (\omega^{[i]})_{i\ge 0}$ denote elements of $\mathscr{G}^{d_{\mathsf{in}}} := \mathscr{G}^{1\times d_{\mathsf{in}}}$, with $\|\omega\|_{\ell_2}^2 := \sum_{i\ge 0}\|\omega^{[i]}\|_2^2$ and Z-transform $\check{\omega}$. Further, define $\mathscr{W}_h := \{\omega \in \mathscr{G}^{d_{\mathsf{in}}} : \|\omega\|_{\ell_2} = 1,\ \omega^{[i]} = 0,\ \forall i > h\}$. We define the $\mathcal{H}_{[h]}$-functional as*

$$\mathcal{H}_{[h]}[G]^2 := \min_{\omega\in\mathscr{W}_h} \frac{1}{2\pi}\int_0^{2\pi} \|\check{G}(e^{\iota\theta})\check{\omega}(e^{\iota\theta})\|_2^2 \mathrm{d}\theta.$$

15. In the full observation setting, with controllers depending directly on noise $\mathbf{w}_t$, Agarwal et al. (2019b) only needs to verify that (the appropriate equivalent of) $G_{\mathrm{ex}\to(y,u)}$ is well conditioned, since the noise terms $\mathbf{w}_t$ are independent by assumption.
16. The restriction $d_{\mathsf{in}} \le d_{\mathsf{o}}$ is to remind the reader that, if $d_{\mathsf{in}} > d_{\mathsf{o}}$, then $\mathcal{H}_{\min}[G]$ is identically zero.

Abusing notation, we also will write $\mathcal{H}_{\min}[\check{G}]$ and other relevant functionals as a function of the Z-transform, where convenient. We also note that, just as $\sigma_{\min}(\cdot)$ is not a norm, $\mathcal{H}_{\min}[\cdot]$ and $\mathcal{H}_{[h]}[\cdot]$ are not norms as well.

Having defined the relevant functions, the following bound gives us a precise bound on the relevant strong convexity parameter. We consider the functions

$$f_{t;k}(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}) = \mathbb{E}[f_t(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}}) \mid \mathcal{F}_{t-k}],$$

where $\mathcal{F}_t$ is the filtration from Assumption 6b. In what follows, we will adopt the shorthand $f_{t;k}(M \mid G_\star, \mathbf{y}_{1:t}^{\mathrm{nat}}, \mathbf{u}_{1:t}^{\mathrm{nat}})$. Our main theorem is as follows:

Theorem 11 *Fix, m, h and let $k = m + 2h$. Further, define*

$$\begin{aligned}\alpha_{m,h} &= \frac{1}{2} \cdot \mathcal{H}_{[m]}[G_{\mathrm{ex}\to(y,u)}]^2 \cdot \mathcal{H}_{[m+h]}[G_{\mathrm{noise}}^\top]^2 \\ &\geq \frac{1}{2} \cdot \mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}]^2 \cdot \mathcal{H}_{\min}[G_{\mathrm{noise}}^\top]^2 := \alpha_\infty,\end{aligned}$$

Then

$$\mathbb{E}\left[\left\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}_{1:t}^{\mathrm{nat}}, \mathbf{v}_t^{\mathrm{nat}}) - \mathbf{v}_t^{\mathrm{nat}}\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \geq \alpha_{m,h,k}\|M\|_{\mathrm{F}}^2 \geq \alpha_\infty \|M\|_{\mathrm{F}}^2,$$

provided that

$$\psi_{G_{\mathrm{ex}\to(y,u)}}(h) \leq \frac{(\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}])}{8(m+h)}, \quad \textit{and} \quad \Psi_{G_{\mathrm{noise}}}(h) \leq \frac{\mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]}{2(m+h)}. \tag{J.1}$$

Thus, if each ℓ_t is chosen by an oblivious adversary and is α-strongly convex, the functions $f_{t;k}(M)$ are $\alpha_{\mathrm{loss}} \cdot \alpha_{m,h,k}$ and $\alpha_{\mathrm{loss}} \cdot \alpha_\infty$strongly-convex, provided that Eq. (J.1) holds.

The above theorem is proved in Appendix J.3. Some remarks are in order:

1. While m, h are algorithm parameters, k appears only in the analysis. The constraints on h reflects how the m-history long inputs $\mathbf{u}_t^{\mathrm{ex}}(M)$ must be given time to propogate through the system, and the constrain $k \geq m + h$ reflects the sufficient excitation required from past noise to ensure the last $m + h$ natures y's are well conditioned.

2. As we shall show in Proposition J.6, the functional $\mathcal{H}_{[h]}[G]$ decays polynomially as a function of h. On the other hand, decay functions decay geometrically, so these constrains on m, h, k can always be satisfied for h and k sufficiently large.

3. We consider $G_{\mathrm{noise}}^\top$ to insure that the input dimension is greater than output dimension, as per the restriction in J.4.

Appendix J.2.2 provides a transparent lower bound on α_∞ when the system is stabilized by an static feedback controller. For general controllers, however, $\mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}]$ may be equal to zero. We introduce the following condition. However, we can show that $\mathcal{H}_{[h]}[G]$ degrades at most polynomially in h:

Proposition J.6 *Let $d_{\mathsf{in}} \leq d_{\mathsf{o}}$ and $G = \mathsf{Transfer}(A, B, C, D) \in \mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, with $\sigma_{d_{\mathsf{in}}}(D) > 0$, or more generally, that G is Then, there exists constants c, n depending only on G such that, for all $h \geq 0$, $\mathcal{H}_{[h]}[G] \geq c/(h+1)^n$.*

We are now in a place to prove our intended proposition:

J.2.1. PROOF OF PROPOSITION E.2B

Recall the settings $h = \lfloor m/3 \rfloor$, and $k = m + 2h$. First, we lower bound $\alpha_{m,h}$. From Proposition J.6, there exists constances $c_1, c_2 > 0$ and n_1, n_2 such that $\mathcal{H}_{[m]}[G_{\mathrm{ex}\to(y,u)}] \geq c_1 m^{-n_1}$ $\mathcal{H}_{[m+h]}[G_{\mathrm{noise}}^\top] \geq c_2(m+h)^{-n_2} \geq \frac{c_2}{2^{n_2}} m^{n_2}$, since $h \leq m$. Thus, there exists somes $\alpha_{\mathrm{sys}} > 0$ and p_{sys} such that $\alpha_{m,h} \geq \alpha_{\mathrm{sys}} m^{p_{\mathrm{sys}}}$.

Now, let us show that there exist an $m \geq m_{\mathrm{sys}}$ for which conditions of Theorem 11 hold. From the stability assumption of the stabilized system (Assumption 1b), there exists constants $C > 0$ and $\rho \in (0,1)$ for which $\psi_{G_{\mathrm{ex}\to(y,u)}}(n) \vee \psi_{G_{\mathrm{noise}}}(n) \leq C\rho^n$. Thus, for $m \geq 4$ and $h = \lfloor m/3 \rfloor \geq m/4$, we have

$$(\frac{(\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}])}{8(m+h)}) \cdot \psi_{G_{\mathrm{ex}\to(y,u)}}(h) \leq \frac{C\rho^{\lfloor m/3 \rfloor}}{8(m + \lfloor m/3 \rfloor)c_1 \lfloor m/3 \rfloor^{-n_1}},$$

which is at most 1 for all m sufficiently large. A similar argume,tm applies to checking the bound $\Psi_{G_{\mathrm{noise}}}(h) \leq \frac{\mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]}{2(m+h)}$. □

J.2.2. EXAMPLE: STATIC FEEDBACK CONTROLLERS

Consider the static feedback setting (Example 4), where we have a stabilizing controller with $A_{\pi_0} = 0$, and $\boldsymbol{\eta}_t = \mathbf{y}_t$. For consistency with conventiona notational, we set $F = D_{\pi_0} \in \mathbb{R}^{d_u d_y}$. This includes the full observation setting via the laws $(A_\star + B_\star F)$, but may also include settings with partial observation which admit a matrix F such that $(A_\star + B_\star K C_\star)$ is stable: Note that taking $K = 0$ subsumes full-feedback as well. The proposition shows that α_∞ from Theorem 11 admits a transparent lower bound:

Proposition J.7 *Consider a static feedback controller with $K = D_{\pi_0}$, and recall $\alpha_\infty := \frac{1}{2} \cdot \mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}]^2 \cdot \mathcal{H}_{\min}[G_{\mathrm{noise}}^\top]^2$. Then,*

1. *$\Sigma_{\mathrm{noise}} \succeq \sigma^2 I$, then $\alpha_\infty \geq \frac{\sigma^2}{32} \min\left\{1, \|K\|_{\mathrm{op}}^{-2}\right\} \cdot \min\left\{1, \|B_\star K\|_{\mathrm{op}}^{-2}\right\}$.*
2. *If only $\sigma_{\mathbf{w}}^2 > 0$ (but Σ_{noise} may not be positive definite), then*

$$\alpha_\infty \geq \frac{\sigma_{\mathbf{w}}^2}{16} \min\left\{1, \|K\|_{\mathrm{op}}^{-2}\right\} \cdot \frac{\sigma_{\min}(C_\star)^2}{(1 + \|A_K\|_{\mathrm{op}})^2}$$

3. *Finally, if $K = 0$, then*

$$\alpha_\infty \geq \frac{\sigma_{\mathbf{e}}^2}{2} + \frac{\sigma_{\mathbf{w}}^2}{2} \frac{\sigma_{\min}(C_\star)}{(1 + \|A_\star\|_{\mathrm{op}})^2}.$$

Note that if $\sigma_{\min}(C_\star) > 0$, then we only need $\sigma_{\mathbf{w}}^2 > 0$ to ensure $\alpha_\infty > 0$. In particular, with with $C_\star = I$, we recover the bounds from Agarwal et al. (2019b), even with stabilizing feedback. Note that, unlike Agarwal et al. (2019b), these bounds don't require any assumptions on the system, or any approximate diagonalizability.[17] It order to illustrate how useful it is to the represent strong convexity in terms of Z-transform and $\mathcal{H}$-functionals, we provide a proof of the above proposition

17. However, to conclude strong convexity via Theorem 11, we require $h - m > 0$. Still, we note that these bounds apply to more general settings where one has (a) observation noise and (b) partial observation.

Proof [Proof of Proposition J.7] In static feedback, we have a stabilizing controller with $A_{\pi_0} = 0$, and set $D_{\pi_0} = K$ and $A_K = A_\star + B_\star K C_\star$, and $\check{A}_K(z) = (zI - A_K)^{-1}$. Then, we can verify

$$G^{[i]}_{\mathrm{ex}\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix}0\\ I\end{bmatrix} + \mathbb{I}_{i>0}\begin{bmatrix}C_\star\\ KC_\star\end{bmatrix} A_K^{i-1}B_\star, \quad G^{[i]}_{(w,e)\to\eta} = \mathbb{I}_{i=0}\begin{bmatrix}0 & I\end{bmatrix} + \mathbb{I}_{i>0}C_\star A_K^{i-1}\begin{bmatrix}0 & K\end{bmatrix}$$

Thus,

$$\check{G}_{\mathrm{ex}\to(y,u)}(z) = \begin{bmatrix}C_\star \check{A}_K(z)B_\star\\ I + KC_\star\check{A}_K(z)B_\star\end{bmatrix}, \quad \check{G}_{\mathrm{noise}}(z)^\top = \Sigma^{1/2}_{\mathrm{noise}}\begin{bmatrix}\check{A}_K(z)^\top C_\star^\top\\ I + B_\star^\top K^\top \check{A}_K(z)^\top C_\star^\top\end{bmatrix}.$$

We now invoke a simple lemma:

Lemma J.8 *Consider a matrix of the form* $W = \begin{bmatrix}YZ\\ I+XZ\end{bmatrix} \in \mathbb{R}^{(d_1+d)\times d}$.*, with* $Y \in \mathbb{R}^{d_1\times d_1}$*,* $X, Z^\top \in \mathbb{R}^{d\times d_1}$*. Then,* $\sigma_{\min}(W) \ge \frac{1}{2}\min\{1, \frac{\sigma_{\min}(Y)}{\|X\|_{\mathrm{op}}}\}$.

Proof [Proof of Lemma J.8] Consider $\|Wv\|_2$ for $v \in \mathbb{R}^d$ with $\|v\| = 1$. If $\|Zv\|_2 \le 1/2\|X\|_{\mathrm{op}}$, then

$$\|Wv\|_2 \ge \|I + XZv\|_2 \ge 1 - \|X\|_{\mathrm{op}}\|Zv\|_2 \ge 1 - \frac{1}{2} = \frac{1}{2}.$$

Otherwise, $\|Wv\|_2 \ge \|YZv\| \ge \sigma_{\min}(Y)\cdot\|Zv\| \ge \frac{\sigma_{\min}(Y)}{2\|X\|_{\mathrm{op}}}$ ■

By Lemma J.8, we see that

$$\mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}] \ge \frac{1}{2}\min\{1, \|K\|^{-1}_{\mathrm{op}}\},$$

and if $\Sigma_{\mathrm{noise}} \succeq \sigma^2 I$, then

$$\mathcal{H}_{\min}[G^\top_{\mathrm{noise}}] \ge \frac{\sigma}{2}\min\{1, \|B_\star K\|^{-1}_{\mathrm{op}}\}.$$

This establishes the first result of the Proposition. Moreover, if we just have state noise σ_w^2 but possibly no observation noise, then $\mathcal{H}_{\min}[G_{\mathrm{noise}}] \ge \sigma_{\min}(C_\star)\mathcal{H}_{\min}[\check{A}_K] \ge \frac{\sigma_{\min}(C_\star)}{1+\|A_J\|_{\mathrm{op}}}$, where we note that $\sigma_{\min}(\check{A}_K(z)) = \frac{1}{\|\check{A}_K(z)^{-1}\|_{\mathrm{op}}} = \frac{1}{1+\|zI-A_K\|_{\mathrm{op}}}$, which is at least $\frac{1}{1+\|A_K\|_{\mathrm{op}}}$ for $z \in \mathbb{T}$.

Lastly, when $K = 0$, we can direclty lower bound $\mathcal{H}_{\min}[G_{\mathrm{ex}\to(y,u)}] \ge 1$, and lower bound $\mathcal{H}_{\min}[G_{\mathrm{noise}}]^2 \ge \sigma^2_{\mathbf{w}}\sigma_{\min}(C_\star)^2\mathcal{H}_{\min}[\check{A}_K]^2 + \mathcal{H}_{\min}[I + B_\star^\top K^\top \check{A}_K(z)^\top C_\star^\top]^2\sigma^2_{\mathbf{e}}$ By specializing $F = 0$ in the argument adopted for the previous part of the proposition, $\mathcal{H}_{\min}[\mathbb{I} + B_\star^\top K^\top \check{A}_K(z)^\top C_\star^\top] \ge \frac{\sigma_{\min}(C_\star)}{1+\|A_\star\|_{\mathrm{op}}}$, and by setting $K = 0$, $\mathcal{H}_{\min}[I + B_\star^\top K^\top \check{A}_K(z)^\top C_\star^\top]^2 = I$. ■

J.2.3. EXAMPLE: YOULA LDC-EX WITH EXACT OBSERVER FEEDBACK

In general, static feedback is not sufficient to stabilize a partially observed linear dynamic system. Let us consider what arises from the the Youla LDC-Ex parametrization. From Lemma G.5, we have

$$G^{[i]}_{(w,e)\to\eta} = \mathbb{I}_{i=0}\begin{bmatrix}0 & I_{d_y}\end{bmatrix} + \mathbb{I}_{i>0}C_\star(A_\star + LC_\star)^{i-1}\begin{bmatrix}I_{d_x} & F\end{bmatrix}.$$

and

$$G^{[i]}_{\mathrm{ex}\to(y,u)} = \mathbb{I}_{i=0}\begin{bmatrix}0\\ I_{d_u}\end{bmatrix} + \mathbb{I}_{i>0}\begin{bmatrix}C_\star\\ F\end{bmatrix}(A_\star + B_\star F)^{i-1}B_\star.$$

Thus, introducing $\check{A}_{BF}(z) = (zI-(A+BK))^{-1}$, and $\check{A}_{LC}(z) = (zI-(A-LC))^{-1}$, we have

$$\check{G}_{\mathrm{ex}\to(y,u)}(z) = \begin{bmatrix}0\\ I_{d_u}\end{bmatrix} + \begin{bmatrix}C_\star\\ F\end{bmatrix}\check{A}_{BF}(z)B_\star = \begin{bmatrix}C_\star\check{A}_{BF}(z)B_\star\\ I_{d_u}+F\check{A}_{BF}(z)B_\star\end{bmatrix}$$

Moreover,

$$\check{G}_{(w,e)\to\eta}(z) = \begin{bmatrix}0 & I_{d_y}\end{bmatrix} + C_\star\check{A}_{LC}(z)\begin{bmatrix}I_{d_x} & F\end{bmatrix},$$

giving

$$\check{G}_{\mathrm{noise}}(z) = \Sigma_{\mathrm{noise}}^{\frac{1}{2}}\begin{bmatrix}\check{A}_{LC}(z)^\top C_\star^\top\\ I_{d_y}+F^\top\check{A}_{LC}(z)^\top C_\star^\top\end{bmatrix}.$$

From Lemma J.8, we have

$$\mathcal{H}_{\min}[\check{G}_{\mathrm{noise}}(z)] \ge \min\left\{1, \frac{1}{2\|F\|_{\mathrm{op}}}\right\}.$$

Lower bounding $\check{G}_{\mathrm{ex}\to(y,u)}(z)$ is a little trickier. Define $X(z) = zI - A_\star$. We have

$$\begin{aligned}
I_{d_u} + F\check{A}_{BF}(z)B_\star &= I + F(zI - A_\star - B_\star F)^{-1}B_\star = I + F(X(z) - B_\star F)^{-1}B_\star\\
&= I + F(X(z)-B_\star F)^{-1}B_\star\\
&= -((-I) - F(X(z)+B_\star(-I)F)^{-1}B_\star)\\
&= -(-I + FX(z)^{-1}B_\star)^{-1}.
\end{aligned}$$

Then,

$$\sigma_{\min}(I_{d_u} + F\check{A}_{BF}(z)B_\star) = \frac{1}{\|-I+FX(z)^{-1}B_\star\|_{\mathrm{op}}} \le \frac{1}{1+\|F\|_{\mathrm{op}}\|B_\star\|_{\mathrm{op}}\|X(z)^{-1}\|_{\mathrm{op}}}.$$

Substituting $X(z) = (zI - A_\star)$, we have

$$\mathcal{H}_{\min}[\check{G}_{\mathrm{ex}\to(y,u)}] \ge \frac{1}{1+\|F\|_{\mathrm{op}}\|B_\star\|_{\mathrm{op}}\max_{z\in\mathbb{T}}\|(zI-A_\star)^{-1}\|_{\mathrm{op}}}$$

In otherwise, if the eigenvalues of $A_\star$ are bounded away from 1 in magnitude, then $\mathcal{H}_{\min}[\check{G}_{\mathrm{ex}\to(y,u)}] > 0$, yielding a bound of $\alpha_\infty > 0$.

J.3. Proof of Theorem 11

The proof of Theorem 11 proceeds by first representing the strong convexity in terms of the Toeplitz operator defined below:

Definition J.9 *Let $k, \ell, h \in \mathbb{N}$ with $k \geq h \geq 0$, and $\ell > 0$. Given a Markov operator $G \in \mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ with $G \in \mathbb{R}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, let $G_\ell \in \mathbb{R}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the Markov operator with $G_\ell^{[i]} = \mathbb{I}_{i \leq \ell} G^{[i]}$. The* Toeplitz *operator is defined by*

$$\mathsf{Toep}_{h;k,\ell}(G) := \begin{bmatrix} G_\ell^{[0]} & 0 & 0 & \dots & 0 \\ G_\ell^{[1]} & G_\ell^{[0]} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ G_\ell^{[h]} & G_\ell^{[h-1]} & \dots & G_\ell^{[1]} & G_\ell^{[0]} \\ G_\ell^{[h+1]} & G_\ell^{[h]} & \dots & G_\ell^{[1]} & G_\ell^{[1]} \\ \dots & \dots & \dots & \dots & \dots \\ G_\ell^{[k]} & G_\ell^{[k-1]} & \dots & \dots & G_\ell^{[k-h]} \end{bmatrix} \in \mathbb{R}^{(k+1)d_{\mathsf{o}} \times (h+1)d_{\mathsf{in}}}$$

We use the shorthand $\mathsf{Toep}_{h;k}(G) = \mathsf{Toep}_{h;k,k}(G)$.

Our first lemma establishes strong convexity in terms of the above operator:

Lemma J.10 *For any $M = (M^{[i]})_{i=0}^{m-1}$, we have the bound*

$$\mathbb{E}\left[\left\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \geq \underline{\alpha}_{m,h,k}\|M\|_{\mathrm{F}}^2$$

where $\|M\|_{\mathrm{F}}^2 := \sum_{i=0}^{m-1} \|M^{[i]}\|_{\mathrm{F}}^2$, and

$$\underline{\alpha}_{m,h,k} = \sigma_{d_u m}(\mathsf{Toep}_{m-1;m+h-1,h}(G_{\mathrm{ex}\to(y,u)}))^2 \cdot \sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}^\top))^2$$

Next, we show that the smallest singular value of a Toeplitz operators is lower bounded by the $\mathcal{H}_{[h]}[G]$

Lemma J.11 *Let $G \in \mathscr{G}^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ be a Markov operator, which in particular means $\|G\|_{\ell_1,\mathrm{op}} < \infty$. Further, let $\ell, k, h \in \mathbb{N}$, with $\ell \geq 1$, and $k \geq h \geq 1$. Finally, set $c_1 = c_2$ if $\ell \geq k$, and otherwise, let $c_1 = (1-\tau)$, $c_2 = \frac{1}{\tau} - 1$ for some $\tau > 0$. Then,*

$$\sigma_{d_{\mathsf{in}}(h+1)}(\mathsf{Toep}_{h;k,\ell}(G))^2 \geq c_1 \mathcal{H}_{[h]}[G]^2 - c_2(h+1)^2 \psi_G(k - h \vee \ell)^2,$$

where $\mathcal{H}_{[h]}[G]$ is as in Definition J.5.

The above lemma is proved in Appendix J.3.2. Theorem 11 noq follows readily:
Proof [Proof of Theorem 11] From Lemma J.10, the functions $f_{t;k}(M) = \mathbb{E}[f_t(M) \mid \mathcal{F}_{t-k}]$ are $\alpha_{\mathrm{loss}} \cdot \underline{\alpha}_{m,h,k}$ strongly-convex, where

$$\underline{\alpha}_{m,h,k} = \sigma_{d_u m}(\mathsf{Toep}_{m-1;m+h-1,h}(G_{\mathrm{ex}\to(y,u)}))^2 \cdot \sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}^\top))^2$$

Applying Lemma J.11 with $\tau = \frac{1}{8}$,

$$\sigma_{d_u m}(\mathsf{Toep}_{m-1;m+h-1,h}(G_{\mathrm{ex}\to(y,u)}))^2 \geq \frac{7}{8}\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}]^2 - 7(h+1)^2 \psi_{G_{\mathrm{ex}\to(y,u)}}(h)^2$$

. Taking

$$\psi_{G_{\mathrm{ex}\to(y,u)}}(h) \leq \frac{h + 1(\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}])^2}{8},$$

we obtain

$$\sigma_{d_u m}(\mathsf{Toep}_{m-1;m+h-1,h}(G_{\mathrm{ex}\to(y,u)}))^2 \ge \frac{3}{4}\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}]^2$$

. Further, applying Lemma J.11 with $\ell = k$, we obtain

$$\sigma_{d_u(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}))^2 \ge \mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]^2 - (m+h-1)^2\Psi_{G_{\mathrm{noise}}}(k-(m+h-1))^2.$$

Taking $k = m+2h$, it suffices that $\Psi_{G_{\mathrm{noise}}}(h)^2 \le \frac{\mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]}{2(m+h)}$, we obtain $\sigma_{d_u(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}))^2 \ge \frac{3}{4}\mathcal{H}_{[m+h]}[G_{\mathrm{noise}}]^2$. Thus,

$$\underline{\alpha}_{m,h,k} \ge \frac{3}{4}\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}]^2 \cdot \frac{3}{4}\mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]^2 \ge \frac{1}{2}\mathcal{H}_{[h]}[G_{\mathrm{ex}\to(y,u)}]^2\mathcal{H}_{[m+h-1]}[G_{\mathrm{noise}}]^2 := \alpha_{m,h,k}.$$

Since $\mathcal{H}_{[h]}[G] \ge \mathcal{H}_{\min}[G]$, we conclude that $\alpha_{m,h} \ge \alpha_\infty$. Therefore,

$$\mathbb{E}\left[\left\|\mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \ge \alpha_{m,h,k}\|M\|_{\mathrm{F}}^2 \ge \alpha_\infty\|M\|_{\mathrm{F}}^2 \tag{J.2}$$

Finally, we observe that if $f(z) = \ell(Xz+v)$ is a function with a random variable X and α-strongly convex loss ℓ, then $\mathbb{E}[f(z)]$ is $\alpha \cdot \alpha'$ strongly convex as long as $\mathbb{E}[\|Xz\|_2^2] \ge \alpha'\|z\|_2^2$ for all z. This means that Eq. (J.2) entails that $f_{t;k}$ is both $\alpha_{\mathrm{loss}} \cdot \alpha_{m,h,k}$ and $\alpha_{\mathrm{loss}} \cdot \alpha_\infty$-strongly convex.

■

J.3.1. PROOF OF LEMMA J.10

For simplicity, we assume that $M \in \mathcal{M}(m+1, R_{\mathcal{M}})$; this simplifies the indexing. Further, introduce the row-toeptliz operator

$$\mathsf{ToepRow}_h(G) := \begin{bmatrix} G^{[0]} & G^{[1]} & \dots & G^{[h]} \end{bmatrix}.$$

Further, lets us introduce the shorthand

$$\delta\mathbf{v}_t(M) = \mathbf{v}_t(M \mid G_{\mathrm{ex}\to(y,u)}, \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}, \mathbf{v}^{\mathrm{nat}}_t) - \mathbf{v}^{\mathrm{nat}}_t, \quad \mathbf{u}^{\mathrm{ex}}_t(M) := \mathbf{u}^{\mathrm{ex}}_t(M \mid \boldsymbol{\eta}^{\mathrm{nat}}_{1:t}).$$

We can directly check that

$$\delta\mathbf{v}_t(M) = \mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)}) \begin{bmatrix} \mathbf{u}^{\mathrm{ex}}_t(M) \\ \mathbf{u}^{\mathrm{ex}}_{t-1}(M) \\ \dots \\ \mathbf{u}^{\mathrm{ex}}_{t-h}(M) \end{bmatrix}.$$

Moreover,

$$\begin{aligned}
\begin{bmatrix} \mathbf{u}^{\mathrm{ex}}_t(M) \\ \mathbf{u}^{\mathrm{ex}}_{t-1}(M) \\ \dots \\ \mathbf{u}^{\mathrm{ex}}_{t-h}(M) \end{bmatrix} &= \begin{bmatrix} M^{[0]} & M^{[1]} & \dots & M^{[m]} & 0 & \dots \\ 0 & M^{[0]} & \dots & M^{[m-1]} & M^{[m]} & \dots \\ \dots & \dots & \dots & \dots & & \\ \dots & \dots & M^{[0]} & M^{[1]} & \dots & M^{[m]} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}^{\mathrm{nat}}_t \\ \boldsymbol{\eta}^{\mathrm{nat}}_{t-1} \\ \dots \\ \boldsymbol{\eta}^{\mathrm{nat}}_{t-(m+h)} \end{bmatrix} \\
&= \mathsf{Toep}_{h;m+h}(M) \begin{bmatrix} \boldsymbol{\eta}^{\mathrm{nat}}_t \\ \boldsymbol{\eta}^{\mathrm{nat}}_{t-1} \\ \dots \\ \boldsymbol{\eta}^{\mathrm{nat}}_{t-(m+h)} \end{bmatrix}.
\end{aligned}$$

Letting $\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h)}$ denote the vector above, this us gives the compact representation:

$$\delta \mathbf{v}_t(M) = \mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h)}.$$

Recall that, to establish the lemma, we wish to lower bound

$$\mathbb{E}[\|\delta \mathbf{v}_t(M)\|_2^2 \mid \mathcal{F}_{t-k}] \geq \underline{\alpha}_{h,m,k}\|M\|_{\mathrm{F}}^2,$$

where $\|M\|_{\mathrm{F}}^2 := \sum_{i=0}^{m-1} \|M^{[i]}\|_{\mathrm{F}}^2$. To this end, define the random variable

$$\mathbf{N}^{\mathrm{nat}}_{t;t-(m+h);k} := \mathbf{N}^{\mathrm{nat}}_{t;t-(m+h)} - \mathbb{E}[\mathbf{N}^{\mathrm{nat}}_{t;t-(m+h)} \mid \mathcal{F}_{t-k}]. \tag{J.3}$$

Since $\mathbf{N}^{\mathrm{nat}}_{t;(t-(m+h);k}$ is uncorrelated with $\mathbb{E}[\mathbf{N}^{\mathrm{nat}}_{t;(t-(m+h)} \mid \mathcal{F}_{t-k}]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\|\delta \mathbf{v}_t(M)\|_2^2 \mid \mathcal{F}_{t-k}\right] \\
&= \mathbb{E}\left[\left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h)}\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \\
&= \mathbb{E}\left[\left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);l}\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \\
&\qquad + \left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\cdot\mathbb{E}\left[\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k} \mid \mathcal{F}_{t-k}\right]\right\|_2^2 \quad \text{(By uncorrelation)} \\
&\geq \mathbb{E}\left[\left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k}\right\|_2^2 \mid \mathcal{F}_{t-k}\right] \\
&= \left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\right\|_2^2 \\
&\qquad \times \sigma_{d_y(1+m+h)}\left(\mathbb{E}\left[\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k}(\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k})^\top \mid \mathcal{F}_{t-k}\right]\right).
\end{aligned}$$

Thus, to conclude the proof, it suffices to establish that, for $\mathbf{N}^{\mathrm{nat}}_{t;t-(m+h);k}$ defined in Eq. (J.3), we have

$$\sigma_{d_y(1+m+h)}\left(\mathbb{E}\left[\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k}(\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k})^\top \mid \mathcal{F}_{t-k}\right]\right) \geq \sigma_{d_y(m+1+h)}(\mathsf{Toep}_{m+h;k}(G_{\mathrm{noise}}))^2 \tag{J.4}$$

and

$$\left\|\mathsf{ToepRow}_h(G_{\mathrm{ex}\to(y,u)})\mathsf{Toep}_{h;m+h}(M)\right\|_{\mathrm{F}}^2 \geq \sigma_{d_u(m+1)}(\mathsf{Toep}_{m;m+h,h}(G_{\mathrm{ex}\to(y,u)}))^2 \tag{J.5}$$

Note that m in the above display in fact corresponds to $m-1$ in the statement of the lemma, since for the proof we assume $M \in \mathcal{M}(m+1, R)$ to simplify indices.

Let us now establish both equations in the above display.

Proving Eq. (J.4) Recall that $G_{(w,e)\to\eta}$ denotes the Markov operator mapping disturbances to outputs, which satisfies by Lemma 3.2b the following

$$\boldsymbol{\eta}^{\mathrm{nat}}_t = \sum_{i=1}^{t} G^{[i]}_{(w,e)\to\eta}\begin{bmatrix}\mathbf{w}_{t-i}\\ \mathbf{e}_{t-i}\end{bmatrix}.$$

Then, since $\mathbf{N}^{\mathrm{nat}}_{t;t-(m+h)}$ is the component of nature's y's depending only on noises $(\mathbf{w}_s, \mathbf{e}_s)$ for $s \in \{t-k, t-k+1, \dots, t\}$, we deduce:

$$\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k} = \mathsf{ToepTrans}_{m+h;k}(G_{(w,e)\to\eta}) \begin{bmatrix} \begin{bmatrix} \mathbf{e}^{\mathrm{stoch}}_t \\ \mathbf{w}^{\mathrm{stoch}}_t \end{bmatrix} \\ \begin{bmatrix} \mathbf{e}^{\mathrm{stoch}}_{t-1} \\ \mathbf{w}^{\mathrm{stoch}}_{t-1} \end{bmatrix} \\ \dots \\ \begin{bmatrix} \mathbf{e}^{\mathrm{stoch}}_{t-k} \\ \mathbf{w}^{\mathrm{stoch}}_{t-k} \end{bmatrix} \end{bmatrix},$$

where we have defined the Toeplitz Transpose operator

$$\mathsf{ToepTrans}_{h;k}(G) := \begin{bmatrix} G^{[0]} & G^{[1]} & G^{[2]} & \dots & G^{[k-h]} & \dots & G^{[k]} \\ 0 & G^{[0]} & G^{[1]} & \dots & G^{[k-h-1]} & \dots & G^{[k-1]} \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ \dots & \dots & \dots & 0 & G^{[0]} & \dots & G^{[k-h]}. \end{bmatrix}$$

Thus, letting $\mathrm{Diag}_{k+1}(\Sigma_{\mathrm{noise}})$ denote a block diagonal matrix with Σ_{noise} along the diagonal, we have from Assumption 6b

$$\begin{aligned} &\mathbb{E}\left[\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k}(\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k})^\top \mid \mathcal{F}_{t-k}\right] \\ &= \mathsf{ToepTrans}_{m+h;k}\left(G_{(w,e)\to\eta}\right)\mathrm{Diag}_{k+1}(\Sigma_{\mathrm{noise}})\mathsf{ToepTrans}_{m+h;k}\left(G_{(w,e)\to\eta}\right)^\top \\ &= \mathsf{ToepTrans}_{m+h;k}\left(G_{(w,e)\to\eta}\cdot\Sigma^{\frac{1}{2}}_{\mathrm{noise}}\right)\mathsf{ToepTrans}_{m+h;k}\left(G_{(w,e)\to\eta}\cdot\Sigma^{\frac{1}{2}}_{\mathrm{noise}}\right)^\top \\ &= \mathsf{ToepTrans}_{m+h;k}\left(G_{\mathrm{noise}}\right)\mathsf{ToepTrans}_{m+h;k}(G_{\mathrm{noise}})^\top, \end{aligned}$$

where we use the convention $G_{(w,e)\to\eta}\cdot\Sigma^{\frac{1}{2}}_{\mathrm{noise}}$ denotes the Markov operator whose i-th component is $G^{[i]}_{(w,e)\to\eta}\Sigma^{\frac{1}{2}}_{\mathrm{noise}}$, and recall the definition $G_{\mathrm{noise}} := G_{(w,e)\to\eta}\cdot\Sigma^{\frac{1}{2}}_{\mathrm{noise}}$ (Definition J.3).

The following fact is straightforward:

Claim J.12 *For all $\sigma_d(\mathsf{ToepTrans}_{h;k}(G)) = \sigma_d(\mathsf{Toep}_{h;k}(G^\top))$ for all d.*

Thus, combining the above with Claim J.12,

$$\begin{aligned} \sigma_{d_y(1+m+h)}\left(\mathbb{E}[\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k}(\mathbf{N}^{\mathrm{nat}}_{t:t-(m+h);k})^\top \mid \mathcal{F}_{t-k}]\right) &\geq \sigma_{d_y(1+m+h)}\left(\mathsf{ToepTrans}_{m+h;k}(G^\top_{\mathrm{noise}})\right)^2 \\ &= \sigma_{d_y(1+m+h)}\left(\mathsf{Toep}_{m+h;k}(G^\top_{\mathrm{noise}})\right)^2, \end{aligned}$$

concluding the proof of Eq. (J.4).

Proof of Eq. (J.5) This bound is a direct consequence of the following claim, which thereby concludes the proof of Lemma J.10.

Claim J.13 *Suppose that G and M are of conformable shapes. Then,*

$$\|\mathsf{ToepRow}_h(G)\mathsf{Toep}_{h;m+h}(M)\|_{\mathrm{F}} \geq \sigma_{d_u(m+1)}(\mathsf{Toep}_{m;m+h,h}(G^\top)) \left\| \begin{bmatrix} M^{[0]} \\ M^{[1]} \\ \dots \\ M^{[m]} \end{bmatrix} \right\|_{\mathrm{F}}$$

Proof Keeping the convention $M^{[i]} = 0$ for $i > m$, we can write

$$\begin{aligned} &\mathsf{ToepRow}_h(G)\mathsf{Toep}_{h;m+h}(M) \\ &= \left[G^{[0]}M^{[0]} \mid G^{[0]}M^{[1]} + G^{[1]}M^{[0]} \mid \cdots \mid \textstyle\sum_{i=0}^h G^{[i]}M^{[h-i]} \mid \sum_{i=0}^h G^{[i]}M^{[1+h-i]} \mid \cdots \mid \sum_{i=0}^h G^{[i]}M^{[m+h-i]} \right]. \\ &= \left[\boldsymbol{N}^{[0]} \mid \boldsymbol{N}^{[1]} \mid \dots \quad \boldsymbol{N}^{[m+h]} \right], \quad \text{where } \boldsymbol{N}^{[j]} := \sum_{i=0}^{m+h} \mathbb{I}_{i\leq h} G^{[i]}M^{[j-i]}. \end{aligned}$$

The above block-row matrix has Frobenius norm equal to that of the following block-column matrix,

$$\overline{\boldsymbol{N}} := \begin{bmatrix} \boldsymbol{N}^{[0]} \\ \boldsymbol{N}^{[1]} \\ \dots \\ \boldsymbol{N}^{[m+h]} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^0 \mathbb{I}_{i\leq h} G^{[i]}M^{[-i]} \\ \sum_{i=0}^1 \mathbb{I}_{i\leq h} G^{[i]}M^{[1-i]} \\ \dots \\ \sum_{i=0}^{m+h} \mathbb{I}_{i\leq h} G^{[i]}M^{[m+h-i]}. \end{bmatrix}$$

which can be expressed as the product

$$\mathscr{N} = \begin{bmatrix} G^{[0]} & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \\ G^{[1]} & G^{[0]} & 0 & \dots & 0 & 0 & \dots & \dots & 0 & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ G^{[h]} & G^{[h-1]} & \dots & \dots & \dots & G^{[0]} & 0 & \dots & \dots & 0 \\ 0 & G^{[h]} & G^{[h-1]} & \dots & \dots & \dots & G^{[0]} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & & & \\ 0 & \dots & 0 & G^{[h]} & G^{[h-1]} & \dots & \dots & \dots & \dots & G^{[0]} \end{bmatrix} \cdot \begin{bmatrix} M^{[0]} \\ M^{[1]} \\ \dots \\ M^{[m]} \\ 0 \\ \dots \\ 0 \end{bmatrix},$$

where we have use $M^{[i]} = 0$ for $i > m$. Let us denote the first m column blocks of the above matrix as X_m. Letting $G_h^{[i]} = \mathbb{I}_{i\leq h} G_h^{[i]}$, we have

$$X_m := \begin{bmatrix} G_h^{[0]} & 0 & \dots & 0 & 0 & 0 \\ G_h^{[1]} & G_h^{[0]} & 0 & \dots & 0 & \\ \dots & \dots & \dots & \dots & \dots & \\ G_h^{[m]} & G_h^{[m-1]} & \dots & \dots & \dots & G_h^{[0]} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ G_h^{[h+m]} & G_h^{[m+h-1]} & \dots & \dots & \dots & G_h^{[h]} \end{bmatrix}$$

Then, we have that $\overline{\boldsymbol{N}} = X_m \begin{bmatrix} M^{[0]} \\ M^{[1]} \\ \dots \\ M^{[m]} \end{bmatrix}$ and in particular,

$$\|\overline{\boldsymbol{N}}\|_{\mathrm{F}} \geq \sigma_{d_u(m+1)}(X_m) \left\| \begin{bmatrix} M^{[0]} \\ M^{[1]} \\ \dots \\ M^{[m]} \end{bmatrix} \right\|_{\mathrm{F}}$$

To conclude, we recognize X_m are the matrix $\mathsf{Toep}_{m;m+h,h}(G)$, so that

$$\|\overline{\boldsymbol{N}}\|_{\mathrm{F}} \geq \sigma_{d_u(m+1)}(\mathsf{Toep}_{m;m+h,h}(G)) \left\| \begin{bmatrix} M^{[0]} \\ M^{[1]} \\ \dots \\ M^{[m]} \end{bmatrix} \right\|_{\mathrm{F}}$$

■

J.3.2. PROOF OF LEMMA J.11

Let $G = (G^{[i]})_{i \geq 0}$ be a Markov operator. We define its z-series as the series

$$\check{G}(z) := \sum_{i \geq 0} z^{-i} \cdot G^{[i]}.$$

Our goal is to prove a lower bound on $\sigma_{\min}(\mathsf{Toep}_{h;k}(G))$. Introduce a "signal" $\{\omega^{[i]}\}_{i \geq 0}$, and annotate the ℓ_2-ball $\mathscr{W}_h := \{\omega : \sum_{i=0}^{h} \|\omega^{[i]}\|_2^2 \leq 1, \omega^{[i]} = 0, \forall i > h\}$. Let us introduce the convention $G^{[i]} = 0$ for $i < 0$. Then, we can express

$$\begin{aligned} \sigma_{\min}(\mathsf{Toep}_{h;k,\ell}(G))^2 &= \min_{\omega \in \mathscr{W}_h} \left\| \mathsf{Toep}_{h;k,\ell}(G)(\omega^{[0]}, \dots, \omega^{[h]}) \right\|_2^2 \\ &= \min_{\omega \in \mathscr{W}_h} \sum_{i=0}^{k} \left\| \sum_{j=0}^{h} \mathbb{I}_{(i-j) \leq \ell} G^{[i-j]} \omega^{[j]} \right\|_2^2 \end{aligned} \tag{J.6}$$

Let us first pass to the $k, \ell \to \infty$ limit.

Lemma J.14 *Let $c_1 = c_2$ if $\ell \geq k$, and otherwise, let $c_1 = (1-\tau)$, $c_2 = \frac{1}{\tau} - 1$ for some $\tau > 0$. Then, for $\|G\|_{\ell_1,\mathrm{op}} < \infty$, we have*

$$\sigma_{\min}(\mathsf{Toep}_{h;k,\ell}(G)) \geq c_1 \min_{\omega \in \mathscr{W}_h} \sum_{i \in \mathbb{Z}} \left\| \sum_{j \in \mathbb{Z}} \mathbb{I}_{(i-j) \leq \ell} G^{[i-j]} \omega^{[j]} \right\|_2^2 - c_2 (h+1)^2 \psi_G(k-h)^2,$$

where if $\ell \geq k$, c_1, c_2 sa $c_1 = c_2 = 1$.
c_1, c_2 satisfy either , or,

Proof For $\omega \in \mathscr{W}_h$, we have

$$\left\|\mathsf{Toep}_{h;k,\ell}(G)(\omega^{[0]},\dots,\omega^{[h]})\right\|_2^2 = \sum_{i=0}^{k}\left\|\sum_{j=0}^{h}\mathbb{I}_{(i-j)\le\ell}G^{[i-j]}\omega^{[j]}\right\|_2^2$$
$$= \sum_{i=0}^{k}\left\|\sum_{j=0}^{h}(G^{[i-j]} - \mathbb{I}_{(i-j)>\ell}G^{[i-j]})\omega^{[j]}\right\|_2^2$$
$$= \sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}(G^{[i-j]} - \mathbb{I}_{(i-j)>\ell \text{ or } i>k}G^{[i-j]})\omega^{[j]}\right\|_2^2.$$

where in the last line we use that $\ell \le k$. Let us introduce the shorthand $\mathbb{I}_{i,j} := \mathbb{I}_{(i-j)>\ell \text{or} i>k}$. Using the elementary vector inequality $\|v+w\|_2^2 \ge (1-\tau)\|v\|_2^2 + (1-\frac{1}{\tau})\|w\|_2^2$. This gives

$$\left\|\mathsf{Toep}_{h;k,\ell}(G)(\omega^{[0]},\dots,\omega^{[h]})\right\|_2^2 \ge c_1\sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}G^{[i-j]}\omega^{[j]}\right\|_2^2 - c_2\sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}\mathbb{I}_{i,j}G^{[i-j]}\omega^{[j]}\right\|_2^2,$$

where $c_1 = (1-\tau)$ and $c_2 = (\frac{1}{\tau}-1)$, and where all sums converge due to $\|G\|_{\ell_1,\mathrm{op}} < 1$. Moreover, one can see that if $\ell \ge k$, then we can simplify the above argument and take $c_1 = c_2 = 1$, as in this case

$$\sum_{i=0}^{k}\left\|\sum_{j=0}^{h}(G^{[i-j]} - \mathbb{I}_{(i-j)>\ell}G^{[i-j]})\omega^{[j]}\right\|_2^2 = \sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}G^{[i-j]}\omega^{[j]}\right\|_2^2 - \sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}\mathbb{I}_{i\ge k}G^{[i-j]}\omega^{[j]}\right\|_2^2$$

Observe that since ω has ℓ_2-norm bounded by 1, we have

$$\|\sum_{j=0}^{h}\mathbb{I}_{i,j}G^{[i-j]}\omega^{[j]}\|_2^2 \le \|\left[\mathbb{I}_{i,j}G^{[i]} \mid \cdots \mid \mathbb{I}_{(i-j)>\ell}G^{[i-h]}\right]\|_{\mathrm{op}}^2 \le \sum_{j=0}^{h}\|\mathbb{I}_{i,j}G^{[i-j]}\|_{\mathrm{op}}^2$$

Thus,

$$\sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}\mathbb{I}_{i,j}G^{[i-j]}\omega^{[j]}\right\|_2^2 \le \sum_{i=0}^{\infty}\sum_{j=0}^{h}\|\mathbb{I}_{i,j}G^{[i-j]}\|_{\mathrm{op}}^2 \le \left(\sum_{i=0}^{\infty}\sum_{j=0}^{h}\|\mathbb{I}_{i,j}G^{[i-j]}\|_{\mathrm{op}}\right)^2$$
$$\le (h+1)\left(\sum_{i\ge 0}\|\mathbb{I}_{i\le\min\{\ell,k-h\}}G^{[i-j]}\|_{\mathrm{op}}\right)^2 = (h+1)^2\psi_G(\ell\wedge k-h)^2.$$

Thus, we conclude that, for any $\tau > 0$,

$$\left\|\mathsf{Toep}_{h;k,\ell}(G)(\omega^{[0]},\dots,\omega^{[h]})\right\|_2^2 \ge (1-\tau)\sum_{i=0}^{\infty}\left\|\sum_{j=0}^{h}G^{[i-j]}\omega^{[j]}\right\|_2^2 + (1-\frac{1}{\tau})(h+1)^2\psi_G(\min\{\ell,k-h\})^2.$$

Finally, since $\omega^{[j]} = 0$ for $j \notin \{0, \dots, h\}$, and $G^{[i-j]} = 0$ for $j \geq 0$ and $i < 0$, we can pass to a double-sum over all indices $i, j \in \mathbb{Z}$. ■

Next, for each $\omega \in \mathscr{W}_h$, we introduce

$$G_{*\omega}^{[i]} := (G * \omega)^{[i]} = \sum_{j \in \mathbb{Z}} G^{[i-j]} \omega^{[j]},$$

and let $\check{G}_{*\omega}[z]$ denote its Z-transform. By the convolution theorem,

$$\check{G}_{*\omega}(z) = \check{G}(z)\check{\omega}(z),$$

where $\check{\omega}(z) = \sum_{i \in \mathbb{Z}} \omega^{[i]} z^i$ is the Z-transform induced by ω. Moreover, by parseval's indentity,

$$\sum_{i \in \mathbb{Z}} \|G_{*\omega}^{[i]}\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} \|\check{G}_{*\omega}(e^{\iota\theta})\|_2^2 \mathrm{d}\theta = \frac{1}{2\pi} \int_{\mathrm{i}}^{2\pi} \|\check{\mathrm{G}}(\mathrm{e}^{\iota\theta})\check{\omega}(\mathrm{e}^{\iota\theta}\|_2^2 \mathrm{d}\theta = \mathcal{H}_{[\mathrm{h}]}[\mathrm{G}]^2$$

Therefore, for c_1, c_2 as in Lemma J.14,

$$\begin{aligned}
\sigma_{\min}(\mathsf{Toep}_{h;k}(G)) &\geq c_1 \min_{\omega \in \mathscr{W}_h} \sum_{i \in \mathbb{Z}} \left\| \sum_{j \in \mathbb{Z}} G^{[i-j]} \omega^{[j]} \right\|_2^2 - c_2(h+1)^2 \psi_G(\ell \wedge k - h)^2 \\
&= c_1 \min_{\omega \in \mathscr{W}_h} \sum_{i \in \mathbb{Z}} \|G_{*\omega}^{[i]}\|_2^2 - c_2(h+1)^2 \psi_G(\ell \wedge k - h)^2 \\
&= c_1 \mathcal{H}_{[h]}[G]^2 - c_2(h+1)^2 \psi_G(\ell \wedge k - h)^2.
\end{aligned}$$

J.4. Proof of Proposition E.2 (Strong Convexity for the Stable Case)

For *stable systems* – that is, systems without a stabilizing controller – we can directly lower bound bound the strong convexity without passing to the Z-transform. This has the advantage of not requiring the conditions on $\Psi_{G_{\mathrm{ex}\to(y,u)}}$ and $\Psi_{G_{\mathrm{noise}}}$ stipulated by Theorem 11. As our starting bound, we recall from Lemma J.10 the bound that $f_{t;k}(M)$ are $\alpha_{\mathrm{loss}} \cdot \underline{\alpha}_{m,h,k}$ strongly-convex, where

$$\underline{\alpha}_{m,h,k} = \sigma_{d_u m}(\mathsf{Toep}_{m-1;m+h-1,h}(G_{\mathrm{ex}\to(y,u)}))^2 \cdot \sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}^\top))^2$$

The followign lemma bounds the quantities in the above display, directly implying Proposition E.2:

Lemma J.15 *For any $m \geq 1$ and any $k \geq m+h$, we have that $\sigma_{d_u(h+1)}(\mathsf{Toep}_{h;m+h-1}(G_{\mathrm{ex}\to(y,u)})) \geq 1$, and $\sigma_{d_u(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}}))^2 \geq \sigma_{\mathbf{e}}^2 + \frac{\sigma_{\mathbf{w}}^2 \sigma_{\min}(C_\star)^2}{(1+\|A_\star\|_{\mathrm{op}})^2}$*

Proof

In the stable case, we have that

$$G_{\mathrm{ex}\to(y,u)}^{[i]} = \begin{bmatrix} G_\star^{[i]} \\ I_{d_u} \mathbb{I}_{i=0}. \end{bmatrix}$$

Thus, for $m \geq 1$, $\mathsf{Toep}_{m-1;m+h-1,h-1}(G_{\mathrm{ex}\to(y,u)}))$ can be repartioned so as to contain a submatrix $I_{d_u m \times d_u m}$. Thus, $\sigma_{d_u m}(\mathsf{Toep}_{h;m+h-1}(G_{\mathrm{ex}\to(y,u)})) \geq 1$.

To lower bound $\sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}})^\top)^2$, we use the diagonal covariance lower bound $\Sigma_{\mathrm{noise}} = \begin{bmatrix} \sigma_{\mathbf{w}}^2 I_{d_x} & 0 \\ 0 & \sigma_{\mathbf{e}}^2 I_{d_y} \end{bmatrix}$. For this covariance, G_{noise} takes the bform

$$G_{\mathrm{noise}}^{[i]} = \begin{bmatrix} \mathbb{I}_{i\geq 1}\cdot (C_\star A^{[i-1]})^\top \sigma_{\mathbf{w}} \\ \mathbb{I}_{i=0}\cdot I_{d_u}\sigma_{\mathbf{e}}. \end{bmatrix}$$

Thus, $\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}})$ can be partitioned as a two-block row matrix, where one block is an

$$X = \begin{bmatrix} \sigma_{\mathbf{e}}\cdot I_{(h+m)d_u} \\ \mathbf{0}_{(k-(m+h-1))d_u\times(m+h)d_u} \\ \sigma_{\mathbf{w}}\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top), \end{bmatrix}$$

where $G_{\star,\mathbf{w}} = \mathbb{I}_{i\geq 1} C_\star A_\star^{i-1}$. From this structure (and use the short hand)

$$\begin{aligned}
&\sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}})^\top)^2 \\
&= \sigma_{(m+h)d_y}^2(X) \\
&= \sigma_{(m+h)d_y}(\sigma_{\mathbf{e}}^2 I_{(h+m)d_y} + \sigma_{\mathbf{w}}^2 \mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top)^\top \mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top)) \\
&\geq \sigma_{\mathbf{e}}^2 + \sigma_{\mathbf{w}}^2 \cdot \sigma_{(m+h)d_y}(\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top))^2
\end{aligned}$$

It remains to lower bound $\sigma_{(m+h)d_u}(\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top))^2$. We can recognize that, for $k \geq m+h$, $\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top)$ has an $(m+h)d_y \times (m+h)d_y$ submatrix which takes the form

$$\left(\mathrm{Diag}(\underbrace{C_\star,\ldots,C_\star}_{m+h\text{ times}})\cdot \mathsf{PowToep}_{m+h}(A_\star)\right),$$

where we have defined

$$\mathsf{PowToep}_p(A) := \begin{bmatrix} I & A & A^2 & \ldots & A^{p-1} \\ 0 & I & A & \ldots & A^{p-2} \\ \ldots \\ 0 & 0 & 0 & \ldots & I \end{bmatrix}$$

Thus,

$$\begin{aligned}
\sigma_{(m+h)d_y}(\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top)) &\geq \sigma_{\min}(C_\star)\cdot\sigma_{\min}(\mathsf{PowToep}_{m+h}(A)) \\
&= \sigma_{\min}(C_\star)\cdot \|\mathsf{PowToep}_{m+h}(A)^{-1}\|_{\mathrm{op}}
\end{aligned}$$

We can verify by direct computation that

$$\mathsf{PowToep}(A)^{-1} = \begin{bmatrix} I & -A & 0 & 0 & \ldots & 0 \\ 0 & I & -A & 0 & \ldots & 0 \\ \ldots \\ 0 & 0 & \ldots & 0 & 0 & I \end{bmatrix},$$

giving $\|\mathsf{PowToep}_{m+h}(A)^{-1}\|_{\mathrm{op}} \leq 1 + \|A\|_{\mathrm{op}}$. Thus, $\sigma_{(m+h)d_y}(\mathsf{Toep}_{m+h-1;k}(G_{\star,\mathbf{w}}^\top)) \geq \frac{\sigma_{\min}(C_\star)}{1+\|A\|_{\mathrm{op}}}$, yielding $\sigma_{d_y(m+h)}(\mathsf{Toep}_{m+h-1;k}(G_{\mathrm{noise}})^\top)^2 \geq \sigma_{\mathbf{e}}^2 + \sigma_{\mathbf{w}}^2\left(\frac{\sigma_{\min}(C_\star)}{1+\|A\|_{\mathrm{op}}}\right)^{-1}$, as needed. ■

J.5. Proof of Proposition J.6

The proof of all subsequence lemmas are provided in sequence at the end of the section. As we continue, set $\mathbb{T} := \{e^{\iota\theta} : \theta \in [2\pi]\} \subset \mathbb{C}$. For $\omega \in \mathscr{G}^{d_{\mathsf{in}}}$, we the following signal norms:

Definition J.16 (Signal norms) *For a complex function $p(z) : z \to \mathbb{C}^d$, we define*

$$\|p\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi}\int_0^{2\pi} \|p(e^{\iota\theta})\|_2^2 \mathrm{d}\theta, \qquad \|p\|_{\mathcal{H}_\infty} := \max_{\theta \in [0,2\pi]} \|p(e^{\iota\theta})\|_2$$

By Parseval's theorem, $\|\omega\|_{\ell_2} = \|\check{\omega}\|_{\mathcal{H}_2}^2$ whenever $\|\omega\|_{\ell_2} < \infty$. Importantly, whenever $\omega \in \mathscr{W}_h$, the signals $\check{\omega}$ are not too "peaked", in the sense that they haved bounded $\mathcal{H}_\infty$-norm:

Lemma J.17 *Suppose that $\omega \in \mathscr{W}_h$. Then, $\check{\omega}$ is a rational function, $\|\check{\omega}\|_{\mathcal{H}_2} = 1$, and $\|\check{\omega}\|_{\mathcal{H}_\infty}^2 \leq h + 1$.*

Using this property, we show that integrating against $\check{\omega}$, for $\omega \in \mathscr{W}_h$, is lower bounded by integrating against the indicator function of a set with mass proportional to $1/h$:

Lemma J.18 (Holder Converse) *Let Let $\mathscr{C}(\epsilon)$ denote the set of Lebesgue measure subsets $\mathcal{C} \subseteq [0, 2\pi]$ with Lebesgue measure $|\mathcal{C}| \geq \epsilon$. Then,*

$$\mathcal{H}_{[h]}[G] = \min_{\omega \in \mathscr{W}_h} \|\check{G}(z)\check{\omega}(z)\|_{\mathcal{H}_2} \geq \frac{1}{8\pi} \cdot \min_{\mathcal{C} \in \mathscr{C}(\frac{\pi}{h+1})} \int_{\theta \in \mathcal{C}} \sigma_{d_{\mathsf{in}}}\left(\check{G}(e^{\iota\theta})\right)^2 \mathrm{d}\theta$$

The next steps of the proof argue that the function $\sigma_{d_{\mathsf{in}}}\left(\check{G}(e^{\iota\theta})\right)^2$ can be lower bounded by a function which, roughly speaking, cannot spend "too much time" close to zero. First, we verify that $\sigma_{d_{\mathsf{in}}}(G(z))$ can only reach zero finitely many times:

Lemma J.19 *Let $G = (A, B, C, D)$ be a Markov operator from $\mathbb{R}^{d_{\mathsf{in}}} \to \mathbb{R}^{d_{\mathsf{o}}}$. Suppose that $d_{\mathsf{o}} \geq d_{\mathsf{in}}$, and $\sigma_{d_{\mathsf{in}}}(D) > 0$. Then, $\sigma_{d_{\mathsf{in}}}(G(z)) = 0$ for at most finitely many $z \in \mathbb{C}$.*

Using this property, we lower bound $\sigma_{\min}(\check{G}(e^{\iota\theta}))^2$ by an *analytic* function. Recall that $f : \mathbb{R} \to \mathbb{R}$ is analytic if it is infinitely differentiable, and for each $x \in \mathbb{R}$, there exists a radius r such that, for all $\delta \in (0, r)$, the Taylor series of f converges on $(x - \delta, x + \delta)$, as is equal to f.

Lemma J.20 *There exists a non-negative, analytic, function $f : \mathbb{R} \to \mathbb{R}$, which is is not identically zero such that, for any $\theta \in \mathbb{R}$, $\sigma_{\min}(\check{G}(e^{\iota\theta}))^2 \geq f(\theta)$, for all $\theta \in \mathbb{R}$.*

Finally, we use the fact that analytic functions cannot spend "too much time" close to zero (unless of course they vanish identically):

Lemma J.21 *Let $f : \mathbb{R} \to \mathbb{R}$ be a real analytic, nonegative, periodic function with period 2π, which is not identically zero. Then, there exists a constants $c > 0$ and $n \in \mathbb{N}$ depending on f such that the following holds: for all $\epsilon \in (0, 2\pi]$, and any set $\mathcal{C} \subseteq [0, 2\pi]$ of Lebesgue measure $|\mathcal{C}| \geq \epsilon$, then,*

$$\int_{\mathcal{C}} f(\theta)\mathrm{d}\theta \geq c\epsilon^n.$$

The proof of Proposition J.6 follows from applying the above lemmas in sequence. Recalling that $\mathscr{C}(\frac{\pi}{h+1})$ denotes the set of $\mathcal{C} \subset [0, 2\pi]$ with $|\mathcal{C}| \geq \frac{\pi}{h+1}$, for some non-vanishing, periodic, analytic f, we obtain

$$\begin{aligned}
\mathcal{H}_{[h]}[G(z)] &\geq \frac{1}{8\pi} \cdot \min_{\mathcal{C} \in \mathscr{C}(\frac{\pi}{h+1})} \int_{\theta \in \mathcal{C}} \sigma_{d_{\mathsf{in}}} \left(\check{G}(e^{\iota\theta})\right)^2 \mathrm{d}\theta && \text{(Lemma J.18)} \\
&\geq \frac{1}{8\pi} \cdot \min_{\mathcal{C} \in \mathscr{C}(\frac{\pi}{h+1})} \int_{\theta \in \mathcal{C}} f(\theta) \mathrm{d}\theta && \text{(Lemma J.20)} \\
&\geq \frac{c}{8\pi} \min_{\mathcal{C} \in \mathscr{C}(\frac{\pi}{h+1})} |\mathcal{C}|^n && \text{(Lemma J.21)} \\
&\geq \frac{c'}{8\pi}(h+1)^n,
\end{aligned}$$

for some $c' > 0$, and $n \in \mathbb{N}$.

J.5.1. PROOF OF LEMMA J.17

Proof Since $\check{\omega}(z) = \sum_{i=0}^{h} \omega^{[i]} z^{-i}$, $\check{\omega}(z)$ is rational. The bound $\|\check{\omega}\|_{\mathcal{H}_2} = 1$ follows from Parsevals identity with $\sum_{i \geq 0} \|\omega^{[i]}\|_2 = 1$ for $\omega \in \mathscr{W}_h$. The third point explicitly uses that $\omega \in \mathscr{W}_h$ is an $h+1$-length signal. Namely, by Cauchy-Schwartz,

$$\|\check{\omega}(z)\|_2 = \| \sum_{i=0}^{h} z^i \omega^{[i]} \|_2^2 \leq \sqrt{\sum_{i=0}^{h} |z^i|^2} \sqrt{\sum_{i=0}^{h} \|\omega^{[i]}\|_2^2} \overset{(i)}{=} \sqrt{\sum_{i=0}^{h} |z^i|^2},$$

where we use that the ℓ_2 norm of ω is bounded by 1. If $z = e^{\iota\theta}$, then $|z^i|^2 = 1$, so $\|\check{\omega}(z)\|_2 \leq \sqrt{h+1}$. ■

J.5.2. PROOF OF LEMMA J.18

We argue that rational functions p with unit $\mathcal{H}_2$-norm and bounded $\mathcal{H}_\infty$ norm must be large on a set of sufficiently large measure:

Lemma J.22 *Let $|\cdot|$ denote Lebesgue measure. Let p be a rational function on $\mathbb{C}$ with $\|p\|_{\mathcal{H}_2} = 1$ and $\|p\|_{\mathcal{H}_\infty}^2 \leq B$. Then, there exists a Lebesgue measurable $\mathcal{C}$ which is a finite union of intervals with Lebesgue measure $|\mathcal{C}| \geq \frac{\pi}{B}$ for which*

$$\forall z \in \mathcal{C}, \|p(z)\|_2 \geq \frac{1}{2}.$$

Proof Let $\mathcal{C}_t := \{\theta \in [0, 2\pi] : \|p(e^{\iota\theta})\|_2 \geq t\}$, which is Lebesgue measurable by rationality of p. Then, by a Chebyschev-like arugment,

$$\begin{aligned}
1 = \|p\|_{\mathcal{H}_2}^2 &\leq \frac{1}{2\pi} \int_{\theta \in \mathcal{C}_t} \|p(e^{\iota\theta})\| + \frac{1}{2\pi} \int_{\theta \in [0, 2\pi] - \mathcal{C}_t} \|p(e^{\iota\theta})\|^2 \\
&\leq \frac{|\mathcal{C}_t| \|p\|_{\mathcal{H}_\infty}^2}{2\pi} + (1 - \frac{|\mathcal{C}_t|}{2\pi})t \\
&\leq \frac{B|\mathcal{C}_t|}{2\pi} + t
\end{aligned}$$

Hence, $\frac{|\mathcal{C}_t|}{2\pi} \geq \frac{1-t}{B}$. In particular, if we $\mathcal{C} = \mathcal{C}_{1/2}$, then $\frac{|\mathcal{C}_t|}{2\pi} \geq \frac{1}{2B}$, as needed. ■

We can now prove Lemma J.18 as follows. For each $\omega \in \mathscr{W}_h$, let $\mathcal{C}_\omega$) denote the corresponding subset of $[0, 2\pi]$ guaranteed by Lemma J.22, that is $\|\check{\omega}(e^{\iota\theta})\|_2 \geq \frac{1}{2}\mathbb{I}(\theta \in \mathcal{C}_\omega$. Then, for $\omega \in \mathscr{W}_h$,

$$\begin{aligned}
\|\check{G}(z)\omega(z)\|_{\mathcal{H}_\infty} &= \frac{1}{2\pi} \min_{\omega \in \mathscr{W}_h} \int_0^{2\pi} \|G(e^{\iota\theta})\check{\omega}(e^{\iota\theta})\|_2^2 \mathrm{d}\theta \\
&\geq \frac{1}{2\pi} \int_0^{2\pi} \sigma_{d_{\mathsf{in}}}(G(e^{\iota\theta}))^2 \|\check{\omega}(e^{\iota\theta})\|_2^2 \mathrm{d}\theta \\
&\geq \frac{1}{2\pi} \int_0^{2\pi} \sigma_{d_{\mathsf{in}}}(G(e^{\iota\theta}))^2 \cdot (\frac{1}{2}\mathbb{I}(\theta \in \mathcal{C}_\omega))^2 \mathrm{d}\theta \\
&= \frac{1}{8\pi} \int_{\theta \in \mathcal{C}_\omega}^{2\pi} \sigma_{d_{\mathsf{in}}}(G(e^{\iota\theta}))^2 \mathrm{d}\theta.
\end{aligned}$$

Since each set $\mathcal{C}_\omega \subset [0, 2\pi]$ has Lebesgue measure at least $\pi/(h+1)$, and since $\mathscr{C}(\pi/(h+1))$ denotes the collection of all subsets with this property, $\min_{\omega \in \mathscr{W}_h} \|\check{G}(z)\omega(z)\|_{\mathcal{H}_\infty}$ is lower bounded by

$$\min_{\mathcal{C} \in \mathscr{C}(\pi/(h+1))} \frac{1}{8\pi} \int_{\theta \in \mathcal{C}_\omega}^{2\pi} \sigma_{d_{\mathsf{in}}}(G(e^{\iota\theta}))^2 \mathrm{d}\theta,$$

as needed.

J.5.3. PROOF OF LEMMA J.19

Proof Next, Note that if $G = (A, B, C, D)$, then $\check{G}(z) = D + C(zI - A)^{-1}B$. Since $\sigma_{d_{\mathsf{in}}}(D) > 0$, there exists a projection matrix $P \in \mathbb{R}^{d_{\mathsf{in}} d_{\mathsf{o}}}$ such that PD is rank d_{in}. Moreover, if $\sigma_{d_{\mathsf{in}}}(P\check{G}(z)) = 0$, then $\sigma_{d_{\mathsf{in}}}(\check{G}(z)) = 0$, so it suffices to show that $\sigma_{d_{\mathsf{in}}}(P\check{G}(z)) = 0$ for only finitely many z. Since $P\check{G}(z) \in \mathbb{R}^{d_{\mathsf{in}} \times d_{\mathsf{in}}}$ is square-matrix valued, it suffices to show that determinant $\det(P\check{G}(z)) = 0$ for at most finitely z. Since $\det$ is a polynomial function, and $P\check{G}(z)$ has rational-function entries, (this can be verified by using Cramers rule), there exists polynomials f, g such that $\det(P\check{G}(z)) = \frac{f(z)}{g(z)}$ for $z \in \mathbb{C}$. This means that either $\det(P\check{G}(z)) = 0$ for all $z \in \mathbb{C}$, or is identically zero on $\mathbb{C}$. Let us show that the second option is not possible. Consider taking $z \to \infty$ (on the real axis). Then $\lim_{z \to \infty} \check{G}(z) = \lim_{z \to \infty} D + C(zI - A)^{-1}B = D$. Hence, $\lim_{z \to \infty} \det(P\check{G}(z)) = \det(PD) > 0$, since $\sigma_{d_{\mathsf{in}}}(PD) > 0$. ■

J.5.4. PROOF OF LEMMA J.20

Proof We have the following lower bound

$$\begin{aligned}
\lambda_{\min}(\check{G}(z)^{\mathsf{H}}\check{G}(z)) &= \frac{\prod_{i=1}^{d_{\mathsf{in}}} \lambda_i \check{G}(z)^{\mathsf{H}}\check{G}(z)}{\prod_{i=1}^{d_{\mathsf{in}}-1} \lambda_i \check{G}(z)^{\mathsf{H}}\check{G}(z)} \\
&\geq \frac{\prod_{i=1}^{d_{\mathsf{in}}} \lambda_i \check{G}(z)^{\mathsf{H}}\check{G}(z)}{(\max_{z \in \mathbb{T}} \|\check{G}(z)\|_{\mathrm{op}}^2)^{d_{\mathsf{in}}-1}} \\
&\geq \frac{\det(\check{G}(z)^{\mathsf{H}}\check{G}(z))}{\|\check{G}\|_{\mathcal{H}_\infty}^{2(d_{\mathsf{in}}-1)}} := \phi(z)
\end{aligned}$$

By assumption, $\|\check{G}(z)\|_{\mathcal{H}_\infty} < \infty$, and so $\phi(z)$ only vanishes when $\det(\check{G}(z)^{\mathsf{H}}\check{G}(z))$ does, which itself only vanishes when $\lambda_{\min}(\check{G}(z)^{\mathsf{H}}\check{G}(z)) = 0$. By Lemma J.19, this means that $\phi(z)$ is not indentically zero on $\mathbb{T}$.

Now, let $\phi(\theta) = \det(\check{G}(e^{\iota\theta})^{\mathsf{H}}\check{G}(e^{\iota\theta}))$. It suffices to show that this function is real analytic. We argue by expressing $\phi(\theta) = \Phi_2(\Phi_1(\theta))$, where Φ_1 is a real analytic map from $\mathbb{R} \to (\mathbb{R}^2)^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, and Φ_2 an analytic map from $(\mathbb{R}^2)^{d_{\mathsf{o}} \times d_{\mathsf{in}}} \to \mathbb{R}$.

Let $\mathrm{embed} : \mathbb{C}^{d_{\mathsf{o}} \times d_{\mathsf{in}}} \to (\mathbb{R}^2)^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$ denote the cannonical complex to real embedding. We define $\Phi_1(\theta) = \mathrm{embed}(\check{G}(e^{\iota\theta})$. To see that $\Phi_1(\theta)$ is real analytic, we observe that the map $z \mapsto e^{\iota\theta}$ is complex analytic, and since $\check{G}$ is a rational function, $u \mapsto \check{G}(u)$ is analytic away from the poles of $\check{G}$. Since $\check{G}$ has no poles $u \in \mathbb{T}$ (by assumption of stability/bounded $\mathcal{H}_\infty$ norm), we conclude that $z \mapsto \check{G}(e^{\iota z})$ is complex analytic at any $z \in \mathbb{R}$. Thus, $\theta \mapsto \mathrm{embed}(\check{G}(e^{\iota z}))$ is real analytic for $\theta \in \mathbb{R}$.

Second, given $X \in (\mathbb{R}^2)^{d_{\mathsf{o}} \times d_{\mathsf{in}}}$, let $\Phi_2(X) = \det((\mathrm{embed}^{-1}(X))^{\mathsf{H}}(\mathrm{embed}^{-1}(X)))$. It is easy to see that $\Phi_2(X)$ is a polynomial in the entries of X, and thus also real analytic. Immediately, we verify that $\phi(\theta) = \Phi_2(\Phi_1(\theta))$, demonstrating that ϕ is given by the composition of two real analytic maps, and therefore real analytic. ■

J.5.5. PROOF OF LEMMA J.21

We begin with a simple claim:

Claim J.23 *f has finitely many zeros on* $[0, 2\pi]$.

Proof Since f is real analytic on $\mathbb{R}$, it can be extended to a complex analytic function $\bar{f}$ on a open subset $U \subset \mathbb{C}$ containing the real line $\mathbb{R}$. Since f is not identically zero on $\mathbb{R}$ assumption, $\bar{f}$ is not identically zero on $\mathbb{R}$, and thus by "Principle of Permanence", $\overline{f}$ can have no accumulation points of zeros on U. In particular, its restriction f can have no accumulation points of zeros on $\mathbb{R}$. As $[0, 2\pi]$ is compact, f has finitely many zeros on $[0, 2\pi]$.[18] ■

We now turn to the proof of our intended lemma:
Proof [Proof of Lemma J.21]

Let $\theta_1, \dots, \theta_m$ denote the zeros on $f(\theta)$ which lie on $[0, 2\pi)$, of which there are finitely many by the above argument. The Taylor coefficients of f cannot be all zero at any of these θ_i, for otherwise analyticity would imply that f would locally vanish. Thus, by Taylor's thoerem, at each zero θ_i, we have that for some constants $c_i > 0$, $r_i > 0$, $n_i \in \mathbb{N}$,

$$f(\theta) \geq c_i|\theta - \theta_i|^{n_i}, \ \forall \theta \in \mathbb{R} : |\theta - \theta_i| \leq r_i$$

Letting $c = \min_i c_i$, $n = \max_i n_i$, and $r_i = \min\{1, \min_i r_i\}$, we have that for all

$$f(\theta) \geq c|\theta - \theta_i|^{n}, \ \forall \theta \in \mathbb{R} : |\theta - \theta_i| \leq r$$

18. As a proof of this fact, note that if f has no accumulation points, then for each $x \in [0, 2\pi]$, there exists an open set set $U_x \subset \mathbb{R}$ containing x which has at most 1 zero. The sets U_x form an open cover of $[0, 2\pi]$. By compactness, there exists a finite number of these sets $U_{x_1}, \dots, U_{x_m}$ which cover $[0, 2\pi]$. Since each U_{x_i} has at most one zero, there are at most m zeros of f on $[0, 2\pi]$.

By shrinking r if necessary, we may assume that the intervals $I_i = [\theta_i - r, \theta_i + r]$ are disjoint, and that there exists a number θ_0 such that $[\theta_0, 2\pi + \theta_0] \supset \bigcup_{i=1}^m I_i$. By periodicity, one can check then $f(\theta)$ only vanishes on $[\theta_0, 2\pi + \theta_0]$ at $\{\theta_1, \dots, \theta_m\} \subset \bigcup_{i=1}^m \mathrm{Interior}(I_i)$. Compactness of the set $S := [\theta_0, 2\pi + \theta_0] - \bigcup_{i=1}^m \mathrm{Interior}(I_i)$ and the fact that $f(\theta) \neq 0$ for all $\theta \in S$ implies that $\inf_{\theta \in S} f(\theta) > 0$. By shrinking c if necessary, we may assume $\inf_{\theta \in S} f(\theta) \geq c$. Therefore, we have shown that

$$\forall \theta \in [\theta_0, 2\pi + \theta_0], \quad f(\theta) \geq \underline{\mathrm{f}}(\theta) := c\left(\mathbb{I}\left(\min_{i \in [m]} |\theta - \theta_i| > r\right) + \sum_{i=1}^m \mathbb{I}((\theta - \theta_i) > r)|\theta - \theta_i|^n\right).$$

Now, let $\mathscr{C}(\epsilon)$ denote the set of subsets $\mathcal{C} \subset [\theta_0, 2\pi + \theta_0]$ with Lebesgue measure ϵ. By translation invariance of the Lebesgue measure, and periodicity of $f(\theta)$, it suffices to show that, for constants c', n', the following holds for all $\epsilon \in (0,1)$, the following holds

$$\min_{\mathcal{C} \in \mathscr{C}(\epsilon)} \int_{\theta \in \mathcal{C}} \underline{\mathrm{f}}(\theta) \mathrm{d}\theta \geq c' \epsilon^{n'}. \tag{J.7}$$

In fact, by shrinking c' if necessary, it suffices to show that the above holds only for $\epsilon \in (0, 2mr)$. Examining the above display, we that any set of the form $\mathcal{C} := \{\theta : \underline{\mathrm{f}}(\theta) \leq t\}$ with $|\mathcal{C}| = \epsilon$ is a minimizer. Assuming the restriction $\epsilon < 2mr$, this implies that the minimum (J.7) is attained by the set $\mathcal{C}_\epsilon := \bigcup_i I_i(\epsilon)$, where we define the intervals $I_i(\epsilon) = [\theta_i - \epsilon/2m, \theta_i + \epsilon/2m]$. We can compute then that,

$$\begin{aligned}
\int_{\theta \in \mathcal{C}_\epsilon} \underline{\mathrm{f}}(\theta) \mathrm{d}\theta &= \sum_{i=1}^m \int_{\theta_i - \epsilon/2m}^{\theta_i - \epsilon/2m} \underline{\mathrm{f}}(\theta) \mathrm{d}\theta \\
&= \sum_{i=1}^m \int_{\theta_i - \epsilon/2m}^{\theta_i - \epsilon/2m} c|\theta - \theta_i|^n \mathrm{d}\theta \\
&= \sum_{i=1}^m \int_{-\epsilon/2m}^{\epsilon/2m} c|\theta|^n \mathrm{d}\theta \\
&= 2cm \int_0^{\epsilon/2m} |r|^n \mathrm{d}\theta \\
&= \frac{2cm}{n+1} (\epsilon/2m)^{n+1},
\end{aligned}$$

which has the desired form.

■

Appendix K. Gradient Descent with Conditional Strong Convexity

We begin by recalling Condtions E.1, E.2and F.1 under which we argue the subsequent bounds. First:

Condition E.1 (Unary Regularity Condition (uRC) for Conditionally-Strongly Convex Losses) *Suppose that $\mathcal{K} \subset \mathbb{R}^d$. Let $f_t := \mathcal{K} \to \mathbb{R}$ denote a sequence of functions and $(\mathcal{F}_t)_{t \geq 1}$ a filtration. We suppose f_t is L_{f}-Lipschitz, and $\max_{x \in \mathcal{K}} \|\nabla^2 f_t(x)\|_{\mathrm{op}} \leq \beta$, and that $f_{t;k}(x) := \mathbb{E}[f_t(x) \mid \mathcal{F}_{t-k}]$ is α-strongly convex on $\mathcal{K}$.*

Note that, by Jensen's inequality, $f_{t;k}$ are β-smooth and L_{f}-Lipschitz on $\mathcal{K}$. Second, we recall the with-memory analogue:

Condition E.2 (With-Memory Regularity Condition (wmRC)) *Suppose that $\mathcal{K} \subset \mathbb{R}^d$ and $h \geq 1$. We let $F_t := \mathcal{K}^{h+1} \to \mathbb{R}$ be a sequence of L_{c} coordinatewise-Lipschitz functions with the induced unary functions $f_t(x) := F_t(x, \ldots, x)$ satisfying Condition E.1.*

Lastly, we formalize the fashion in which the iterates are generated:

Condition F.1 *We suppose that $z_{t+1} = \Pi_{\mathcal{K}}(z_t - \eta \boldsymbol{g}_t)$, where $\boldsymbol{g}_t = \nabla f_t(z_t) + \boldsymbol{\epsilon}_t$. We further assume that the gradient descent iterates applied for $t \geq t_0$ for some $t_0 \leq k$, with $z_0 = z_1 = \cdots = z_{t_0} \in \mathcal{K}$. We assume that $\|\boldsymbol{g}_t\|_2 \leq L_{\mathbf{g}}$, and $\mathrm{Diam}(\mathcal{K}) \leq D$.*

The remainder of the section is as follows. Section K.1 proves Lemma K.1, which relates the regret on the non-conditioned unary sequence f_t to standard strongly convex $\frac{\log T}{\alpha}$, plus additional correction for the errors $\boldsymbol{\epsilon}_t$, the negative regret, and a correction $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ for the mismatch between f_t and $f_{t;k}$. For $k = 0$, $f_{t;k} = t_t$ and $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ is zero, recovering Proposition F.3. Next, Section K.2 proves Lemma K.2, which bounds theterms $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ in terms of a mean-zero sequence $Z_t(z_\star)$ depending on the comparator $z_\star$.

Next, Section K.3.1 states and proves our main high-probability regret bound for unary functions, Theorem. Lastly, Section K.3.3 extends

K.1. Basic Regret Lemma and Proposition F.3

We begin by proving the following "basic" inequality for the unary setting, which provides a key intermediate regret bound adressing both conditional strong convexity and error in the gradients, as well as incorporating negative regret:

Lemma K.1 (Basic Inequality for Conditional-Expectation Regret) *Consider the setting of Conditions E.1 and F.1. For step size $\eta_t = \frac{3}{\alpha t}$,*

$$\forall z_\star \in \mathcal{K}, \quad \sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) \leq -\frac{\alpha}{6} \sum_{t=k+1}^{T} \|z_t - z_\star\|_2^2 + \frac{6L_{\mathrm{f}}^2}{\alpha} \log(T+1)$$
$$+ \frac{6}{\alpha} \sum_{t=k+1}^{T} \|\boldsymbol{\epsilon}_t\|_2^2 - \sum_{t=1}^{T} \left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle + \frac{\alpha D^2 (k+1)}{2},$$

where we define $\boldsymbol{\epsilon}_t^{\mathrm{stoch}} := \nabla f_t(z_t) - \nabla f_{t;k}(z_t)$

Note that Proposition F.3 in the body arises in the special case where $k = 0$. For $k > 1$, the error $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ is required to relate the updates based on $\nabla f_t(z_t)$ to to based on $\nabla f_{t;k}(z_t)$, the latter of which corresponding to functions which are strongly convex.

Proof Let $\boldsymbol{g}_t := \nabla f_t(z_t) + \boldsymbol{\epsilon}_t$. From (Hazan et al., 2016, Eq. 3.4), strong convexity of $f_{t;k}$ implies that

$$2(f_{t;k}(z_t) - f_{t;k}(z_\star)) \le 2\nabla_{t;k}^\top(z_t - z_\star) - \alpha\|z_\star - z_t\|_2^2 \tag{K.1}$$

Now, we let gradient descent correspond to the update $\mathbf{y}_{t+1} = z_t - \eta_{t+1}\boldsymbol{g}_t$, where $\boldsymbol{g}_t = \nabla_{t;k} + \boldsymbol{\epsilon}_t^{\mathrm{stoch}} + \boldsymbol{\epsilon}_t$, for stochastic error $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ and deterministic noise $\boldsymbol{\epsilon}_t$. The Pythagorean Theorem implies

$$\|z_{t+1} - z_\star\|_2^2 \le \|z_t - z_\star - \eta_{t+1}\boldsymbol{g}_t\|_2^2 = \|z_t - z_\star\|_2^2 + \eta_{t+1}^2\|\boldsymbol{g}_t\|^2 - 2\eta_{t+1}\boldsymbol{g}_t^\top(z_t - z_\star), \tag{K.2}$$

which can be re-expressed as

$$-2\boldsymbol{g}_t^\top(z_t - z_\star) \ge \frac{\|z_{t+1} - z_\star\|_2^2 - \|z_t - z_\star\|_2^2}{\eta_{t+1}} - \eta_{t+1}\|\boldsymbol{g}_t\|^2 \tag{K.3}$$

Furthermore, using the elementary inequality $ab \le \frac{a^2}{2\tau} + \frac{\tau}{2}b^2$ for any a, b and $\tau > 0$, we have that for any $\tau > 0$

$$\begin{aligned}
-\langle \boldsymbol{g}_t, z_t - z_\star\rangle &= -\left\langle \nabla_{t;k} + \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle - \langle \boldsymbol{\epsilon}_t, z_t - z_\star\rangle \\
&\le -\left\langle \nabla_{t;k} + \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle + \frac{\alpha\tau}{2}\|z_t - z_\star\|_2^2 + \frac{1}{2\alpha\tau}\|\boldsymbol{\epsilon}_t\|_2^2
\end{aligned} \tag{K.4}$$

Combining Equations (K.3) and (K.4), and rearranging,

$$\begin{aligned}
2\nabla_{t;k}^\top(z_t - z_\star) &\le \frac{\|z_t - z_\star\|^2 - \|z_{t+1} - z_\star\|^2}{\eta_{t+1}} + \eta_{t+1}\|\boldsymbol{g}_t\|^2 + \frac{1}{\tau\alpha}\|\boldsymbol{\epsilon}_t\|_2^2 \\
&\quad + \tau\alpha\|z_t - z_\star\|^2 - 2\left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle \\
&\le \frac{\|z_t - z_\star\|^2 - \|z_{t+1} - z_\star\|^2}{\eta_{t+1}} + 2\eta_{t+1}L^2 + \left(2\eta_{t+1} + \frac{1}{\tau\alpha}\right)\|\boldsymbol{\epsilon}_t\|_2^2 \\
&\quad + \tau\alpha\|z_t - z_\star\|^2 - 2\left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle,
\end{aligned}$$

where we used $\|\boldsymbol{g}_t\|_2^2 \le 2(\|\nabla f(z_t)\|_2^2 + \|\boldsymbol{\epsilon}_t\|_2^2) \le 2(L^2 + \|\boldsymbol{\epsilon}_t\|_2^2)$. Combining with (K.1), we have

$$\begin{aligned}
&\sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) \\
&\le \frac{1}{2}\sum_{t=k+1}^{T}\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} - (1-\tau)\alpha\right)\|z_t - z_\star\|_2^2 \\
&\quad + \sum_{t=k+1}^{T} 2\eta_{t+1}L^2 + \left(\frac{1}{\tau\alpha} + 2\eta_{t+1}\right)\|\boldsymbol{\epsilon}_t\|_2^2 - \sum_{t=1}^{T}\left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle + \frac{\|z_{k+1}\|_2^2}{\eta_{1+k}}
\end{aligned}$$

Finally, let us set $\eta_t = \frac{3}{\alpha t}$, $\tau = \frac{1}{3}$, and recall $D = \mathrm{Diam}(\mathcal{K})$ and $\|\nabla_t\|_2^2 \le L_{\mathrm{f}}^2$. Then, we have that

1. $\frac{1}{2}\sum_{t=1}^{T}\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} - (1-\tau)\alpha\right)\|z_t - z_\star\|_2^2 = \frac{1}{2}\sum_{t=1}^{T}(\alpha/3 - 2\alpha/3)\|z_t - z_\star\|_2^2$, which is equal to $\frac{-\alpha}{6}\sum_{t=1}^{T}\|z_t - z_\star\|^2$
2. $\left(\frac{1}{\tau\alpha} + 2\eta_{t+1}\right) \le \frac{6}{\alpha}$, and $2\sum_{t=k+1}^{T}\eta_{t+1}L_{\mathrm{f}}^2 \le 2\cdot 3L_{\mathrm{f}}^2\log(T+1)/\alpha$
3. $\frac{\|z_{k+1}\|_2^2}{\eta_{1+k}} \le \alpha(k+1)D^2/3$.

Putting things together,

$$\sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) \le -\frac{\alpha}{6}\sum_{t=k+1}^{T}\|z_t - z_\star\|_2^2 + 6L_{\mathrm{f}}^2\log(T+1) + \frac{3}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2 - \sum_{t=1}^{T}\left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star\right\rangle + \frac{\alpha D^2(k+1)}{3}$$

Finally, to conclude, we bound

$$\begin{aligned}\frac{\alpha D^2(k+1)}{3} - \frac{\alpha}{6}\sum_{t=k+1}^{T}\|z_t - z_\star\|_2^2 &\le \frac{(k+1)\alpha D^2}{3} + \frac{(k+1)\alpha D^2}{6} - \frac{\alpha}{6}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2 \\ &= \frac{(k+1)\alpha D^2}{2} - \frac{\alpha}{6}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2.\end{aligned}$$

■

K.2. De-biasing the Stochastic Error

The next step in the proof is to unpack the stochastic error term from Lemma K.1, yielding a bound in terms of a mean-zero sequence $Z_t(z_\star)$:

Lemma K.2 (De-biased Regret Inequality) *Under Conditions E.1 and F.1 step size $\eta_t = \frac{3}{\alpha t}$, the following bound holds deterministically for any $z_\star \in \mathcal{K}$*

$$\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) \le \frac{\alpha D^2(k+1)}{2} + \frac{6L_{\mathrm{f}}^2 + kL_{\mathbf{g}}(6\beta + 12L_{\mathrm{f}})}{\alpha}\log(T+1) + \frac{6}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2 + \sum_{t=k+1}^{T} Z_t(z_\star) - \frac{\alpha}{6}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2$$

where we define $Z_t(z_\star) := (f_{t;k} - f)(z_{t-k}) - (f_{t;k} - f)(z_\star) + \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_{t-k} - z_\star\rangle$.

One can readily check that $\mathbb{E}[Z_t(z_\star) \mid \mathcal{F}_{t-k}] = 0$.

Proof

Let $z_\star \in \mathcal{K}$ denote an arbitrary competitor point. We recall that $f_{t;k} := \mathbb{E}[f_t \mid \mathcal{F}_{t-k}]$, and set $\boldsymbol{\epsilon}_t^{\mathrm{stoch}} := \nabla f_t(z_t) - \nabla f_{t;k}(z_t)$. Proceeding from Lemma K.1, there are two challenges: (a) first, we wish to convert a regret bound on the conditional expectations $f_{t;k}$ of the functions to the actual functions f_t and (b) the errors $\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \rangle$ do not form a martingale sequence, because the errors $\boldsymbol{\epsilon}_t^{\mathrm{stoch}}$ are correlated with z_t. We adress both points with a decoupling argument. Begin by writing

$$\boldsymbol{\epsilon}_t^{\mathrm{stoch}} = \nabla f_t(z_t) - \nabla f_{t;k}(z_t) = \nabla(f_t - f_{t;k})(z_{t-k}) + \nabla(f_t - f_{t;k})(z_t) - \nabla(f_t - f_{t;k})(z_{t-k}).$$

We then have that

$$\begin{aligned}
&\sum_{t=k+1}^{T} \left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle \\
&= \sum_{t=k+1}^{T} \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_{t-k} - z_\star \rangle \\
&\quad + \sum_{t=k+1}^{T} \langle \nabla(f_t - f_{t;k})(z_t) - \nabla(f_t - f_{t;k})(z_{t-k}), z_t - z_\star \rangle + \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_t - z_{t-k} \rangle
\end{aligned}$$

Since $f_{t;k}$ and f_t are L-Lipschitz and β-smooth, we have

$$\begin{aligned}
&\left| \sum_{t=k+1}^{T} \langle \nabla(f_t - f_{t;k})(z_t) - \nabla(f_t - f_{t;k})(z_{t-k}), z_t - z_\star \rangle + \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_t - z_{t-k} \rangle \right| \\
&\quad \le \sum_{t=k+1}^{T} 2\beta \|z_t - z_{t-k}\|_2 \|z_t - z_\star\|_2 + 2L_{\mathrm{f}} \|z_t - z_{t-k}\| \quad \le \sum_{t=k+1}^{T} 2(\beta D + L_{\mathrm{f}}) \|z_t - z_{t-k}\|.
\end{aligned}$$

Similarly, we decouple,

$$\begin{aligned}
\sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) &= \sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) + \sum_{t=k+1}^{T} (f_{t;k} - f)(z_{t-k}) - (f_{t;k} - f)(z_\star) \\
&\quad + \sum_{t=k+1}^{T} f_t(z_{t-k}) - f_t(z_t) + f_{t;k}(z_t) - f_{t;k}(z_{t-k}).
\end{aligned}$$

Similarly, we can bound

$$\left| \sum_{t=k+1}^{T} f_t(z_{t-k}) - f_t(z_t) + f_{t;k}(z_t) - f_{t;k}(z_{t-k}) \right| \le \sum_{t=k+1}^{2} 2L_{\mathrm{f}} \|z_t - z_{t-k}\|.$$

Putting the above together, we find that

$$\begin{aligned}
&\sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) + \sum_{t=k+1}^{T} \left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle \le \underbrace{\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star)}_{(i)} \\
&+ \sum_{t=k+1}^{T} \underbrace{(f_{t;k} - f)(z_{t-k}) - (f_{t;k} - f)(z_\star) + \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_{t-k} - z_\star \rangle}_{:= Z_t(z_\star)} + \underbrace{\sum_{t=k+1}^{T} (2\beta D + 4L_{\mathrm{f}}) \|z_t - z_{t-k}\|}_{(iii.a)},
\end{aligned}$$

To conclude, let us bound the term $(iii.a)$:

$$
\begin{aligned}
(iii.a) &\le \sum_{t=k+1}^{T} (2\beta D + 4L_{\mathrm{f}}) \sum_{i=1}^{k} \|z_{t-i} - z_{t-k-j}\| \\
&\le \sum_{t=k+1}^{T} (2\beta D + 4L_{\mathrm{f}}) \sum_{i=1}^{k} \eta_{t+1} \|\nabla f_t(z_t) + \boldsymbol{\epsilon}_t\| \\
&\le kL_{\mathbf{g}}(2\beta D + 4L_{\mathrm{f}}) \sum_{t=k+1}^{T} \eta_{t+1} \le \frac{3}{\alpha} \cdot kL_{\mathbf{g}}(2\beta D + 4L_{\mathrm{f}}) \log(T+1).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sum_{t=k+1}^{T} f_{t;k}(z_t) - f_{t;k}(z_\star) + \sum_{t=k+1}^{T} \left\langle \boldsymbol{\epsilon}_t^{\mathrm{stoch}}, z_t - z_\star \right\rangle \\
&\qquad \le \sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) + \sum_{t=k+1}^{T} Z_t(z_\star) + \frac{kL_{\mathbf{g}}(6\beta D + 12L_{\mathrm{f}})}{\alpha} \log(T+1),
\end{aligned}
$$

Lemma K.2 follows directly from combining the above with Lemma K.1.

■

K.3. High Probability Regret

K.3.1. HIGH PROBABILITY FOR UNARY FUNCTIONS

Our main high-probability guarantee for unary functions is as follows:

Theorem 12 *Consider a sequence of functions $f_1, f_2, \dots$ satisfying Condtions E.1 and F.1. Then, with step size $\eta_t = \frac{3}{\alpha t}$, the following bound holds with probability $1-\delta$ for all $z_\star \in \mathcal{K}$ simultaenously:*

$$
\begin{aligned}
&\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) - \left(\frac{6}{\alpha} \sum_{t=k+1}^{T} \|\boldsymbol{\epsilon}_t\|_2^2 - \frac{\alpha}{12} \sum_{t=1}^{T} \|z_t - z_\star\|_2^2 \right) \\
&\qquad \lesssim \alpha k D^2 + \frac{kdL_{\mathrm{f}}^2 + kL_{\mathrm{f}}L_{\mathbf{g}} + k\beta L_{\mathbf{g}}}{\alpha} \log(T) + \frac{kL_{\mathrm{f}}^2}{\alpha} \log\left(\frac{1 + \log_+(\alpha D^2)}{\delta} \right).
\end{aligned}
$$

Proof Starting from Lemma K.2, we have

$$
\begin{aligned}
\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) &\le \frac{\alpha D^2(k+1)}{2} + \mathcal{O}\left(\frac{L_{\mathrm{f}}^2 + k(\beta + L_{\mathrm{f}})L_{\mathbf{g}}}{\alpha} \right) \log(T+1) + \frac{6}{\alpha} \sum_{t=k+1}^{T} \|\boldsymbol{\epsilon}_t\|_2^2 \\
&\quad + \underbrace{\sum_{t=k+1}^{T} Z_t(z_\star) - \frac{\alpha}{12} \sum_{t=1}^{T} \|z_t - z_\star\|_2^2}_{(i)} - \frac{\alpha}{12} \sum_{t=k+1}^{T} \|z_t - z_\star\|_2^2
\end{aligned}
$$

where we we recall $Z_t(z_\star) := (f_{t;k} - f)(z_{t-k}) - (f_{t;k} - f)(z_\star) + \langle \nabla(f_t - f_{t;k})(z_{t-k}), z_{t-k} - z_\star \rangle$. We now state a high-probability upper bound on term (i), proved in Section K.3.2 below:

Lemma K.3 (Point-wise concentration) *Fix a $z_\star \in \mathcal{K}$. Then, with probability $1-\delta$, the following bound holds*

$$\sum_{t=k+1}^{T} Z_t(z_\star) - \frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2 \le \mathcal{O}\left(\frac{kL_{\mathrm{f}}^2}{\alpha}\right)\log\left(\frac{k(1+\log_+(\alpha T D^2)}{\delta}\right),$$

where $\log_+(x) = \log(x \vee 1)$.

Together with $k \le T$ and some algebra, the following holds probabilty $1-\delta$ for any fixed $z_\star \in \mathcal{K}$,

$$\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) \le \frac{\alpha D^2(k+1)}{2} + \mathcal{O}\left(\frac{L_{\mathrm{f}}^2 + k(\beta + L_{\mathrm{f}})L_{\mathbf{g}}}{\alpha}\right)\log(T+1) + \frac{6}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2$$
$$+\, \mathcal{O}\left(\frac{kL_{\mathrm{f}}^2}{\alpha}\right)\log\left(\frac{T(1+\log_+(\alpha D^2)}{\delta}\right) - \frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2$$

To extend from a fixed $z_\star$ to a uniform bound, we adopt a covering argument. Note that the only terms that depend explicitly on the comparators $z_\star$ are $-f_t(z_\star)$ and $\|z_t - z_\star\|_2^2$. We then establish the following bound:

Claim K.4 *Let $\mathcal{N}$ denote a D/T-cover of $\mathcal{K}$. Then, for ay $z_\star \in \mathcal{K}$, there exists a $z \in \mathcal{N}$ with*

$$\left|-\sum_{t=k+1}^{T}(f_t(z_\star) - f_t(z)) + \frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2 - \|z_t - z\|_2^2\right| \lesssim \frac{L_{\mathrm{f}}^2}{\alpha} + \alpha D^2.$$

Proof $\left|\|z_t - z_\star\|_2^2 - \|z_t - z\|_2^2\right| = |\langle z_t - z_\star, (z_t - z_\star) - (z_t - z)\rangle + \langle z_t - z_\star, (z_t - z_\star) - (z_t - z)\rangle| \le 2D\|z_\star - z\|$. Moreover, $|f_t(z) - f_t(z_\star)| \le L_{\mathrm{f}}\|z - z_\star\|$. From the triangle inequality, we have that the sum in the claim is bounded by $(L_{\mathrm{f}} + \alpha D)T\|z - z_\star\| \le L_{\mathrm{f}}D + \alpha D^2 \lesssim L_{\mathrm{f}}^2/\alpha + \alpha D^2$. ■

Next, we bound the size of our covering

Claim K.5 *There exists an D/T covering of $\mathcal{N}$ with cardinality at most $(1+2T)^d$.*

Proof Observe that $\mathcal{K}$ is contained in ball of radius D. Set $\epsilon = D/T$. By a standard volumetric covering argument, it follows that we can select $|\mathcal{N}| \le ((D+\epsilon/2)/(\epsilon/2))^d = (1+\frac{2D}{\epsilon})^d = (1+2T)^d$. ■

Absorbing the approximation error of $L^2/\alpha + \alpha D^2$ from Claim K.4, and applying a union bound over the cover from Claim K.4, we have with probability $1-\delta$ that

$$\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) - \left(\frac{\alpha}{12}\sum_{t=1}^{T}\|z_t - z_\star\|_2^2 + \frac{6}{\alpha}\sum_{t=k+1}^{T}\|\boldsymbol{\epsilon}_t\|_2^2\right)$$
$$\lesssim \alpha D^2 k + \frac{L_{\mathrm{f}}^2 + k(\beta + L_{\mathrm{f}})L_{\mathbf{g}}}{\alpha}\log(T) + \frac{kL^2}{\alpha}\log\left(\frac{T(1+2T)^d(1+\log_+(\alpha D^2)}{\delta}\right)$$
$$\lesssim \alpha k D^2 + \frac{kdL^2 + k(\beta + L_{\mathrm{f}})L_{\mathbf{g}}}{\alpha}\log(T) + \frac{kL_{\mathrm{f}}^2}{\alpha}\log\left(\frac{1+\log_+(\alpha D^2)}{\delta}\right).$$

■

K.3.2. PROOF OF LEMMA K.3

For simplicity, drop the dependence on $z_\star$, and observe that, since $f_t, f_{t;k}$ are L_f-Lipschitz, we can bound $|Z_t| \le 4L_\mathrm{f}\|z_t - z_\star\|_2$. Moreover, $\mathbb{E}[Z_t \mid \mathcal{F}_{t-k}] = 0$. We can therefore write

$$\sum_{t=k+1}^{T} Z_t = \sum_{t=k+1}^{T} U_t \cdot \overline{Z}_t,$$

where we set $U_t := 4L_\mathrm{f}\|z_t - z_\star\|_2$ and $\overline{Z}_t := Z_t/U_t$. We can check that

$$|\overline{Z}_t| \mid \mathcal{F}_{t-k} \le 1 \text{ a.s.} \quad \text{and} \quad \mathbb{E}[\overline{Z}_t \mid \mathcal{F}_{t-k}] = 0.$$

Hence, $\overline{Z}_t$ is a bounded, random variable with $\mathbb{E}[\overline{Z}_t \mid \mathcal{F}_{t-k}] = 0$ multiplied by a $\mathcal{F}_{t-k}$-measurable non-negative term. Note that this does not quite form a martingale sequence, since $\overline{Z}_t$ has mean zero conditional on $\mathcal{F}_{t-k}$, not $\mathcal{F}_{t-1}$.

This can be adressed by a blocking argument: let $t_{i,j} = k + i + jk - 1$ for $i \in [k]$, and $j \in \{1, \dots, T_i\}$, where $T_i := \max\{j : t_{i,j} \le T\}$. Then, we can write

$$\sum_{t=k+1}^{T} U_t \cdot \overline{Z}_t = \sum_{i=1}^{k} \left(\sum_{j=1}^{T_i} U_t \cdot \overline{Z}_{t_{i,j}} \right)$$

Now, each term in the inner sum is a martingale sequence with respect to the filtration $\{\mathcal{F}_{t_{i,j}}\}_{j \ge 1}$. Moreover, $\overline{Z}_{t_{i,j}} \mid \mathcal{F}_{t_{i,j-1}}$ is $\frac{1}{4}$-sub-Gaussian. We now invoke the a modification of Simchowitz et al. (2018, Lemma 4.2 (b)), which follows straightforwardly from adjusting the last step of its proof

Lemma K.6 *Let X_j, Y_j be two random processes. Suppose $(\mathcal{G}_j)_{j \ge 0}$ is a filtration such that (X_j) is $(\mathcal{G}_j)$-adapted, Y_j is $(\mathcal{G}_{j-1})$ adapted, and $X_j \mid \mathcal{G}_{j-1}$ is σ^2 subGaussian. Then, for any $0 < \beta_- \le \beta_+$,*

$$\mathbb{P}\left[\mathbb{I}(\sum_{j=1}^{T} Y_j^2 \le \beta_+) \sum_{j=1}^{T} X_j Y_j \ge u \max\left\{ \sqrt{\sum_{j=1}^{T} Y_j^2}, \beta_- \right\} \cdot \right] \le \log\lceil \frac{\beta_+}{\beta_-} \rceil \exp(-u^2/6\sigma^2)$$

For each i, apply the above lemma with $\beta_- = L_\mathrm{f}^2/\alpha$ and $\beta_+ = \max\{T L_\mathrm{f}^2 D^2, L_\mathrm{f}^2/\alpha\}$, $X_j = \overline{Z}_{t_{i,j}}$ and $Y_j = U_{t_{i,j}}$, $\sigma^2 = 1/4$, and $u = \sqrt{3\log(1/k\delta)/2}$. Then, we have $\sum_{j=1}^{T_i} Y_j^2 \le \beta_+$ almost surely, so we conclude that, with probability $(1 + \log(1 \vee \alpha T D^2))\delta$, the following holds for any $\tau > 0$

$$\begin{aligned}
\forall i : \sum_{j=1}^{T_i} U_t \cdot \overline{Z}_{t_{i,j}} &\le \beta_- \sqrt{3\log(k/\delta)/2} + \sqrt{\frac{3}{2}\log(k/\delta)\sum_{j=1}^{T_i} U_{t_{i,j}}^2} \\
&= \beta_- \sqrt{3\log(k/\delta)/2} + \sqrt{\frac{3 \cdot 16 L_{\mathrm{f}}^2}{2}\log(k/\delta)\sum_{j=1}^{T_i} \|z_{t_{i,j-1}} - z_\star\|_2^2} \\
&= \beta_- \sqrt{3\log(k/\delta)/2} + \sqrt{24 L_{\mathrm{f}}^2 \log(k/\delta)\sum_{j=1}^{T_i} \|z_{t_{i,j-1}} - z_\star\|_2^2} \\
&\le \frac{L_{\mathrm{f}}^2}{\alpha}(\sqrt{3\log(k/\delta)/2} + \frac{12}{\tau}\log(k/\delta)) + \frac{\tau}{2}\sum_{j=1}^{T_i} \|z_{t_{i,j-1}} - z_\star\|_2^2 \\
&\le L_{\mathrm{f}}^2 \log(k/\delta)(\frac{3}{2\alpha} + \frac{12}{\tau}) + \frac{\tau}{2}\sum_{j=1}^{T_i} \|z_{t_{i,j-1}} - z_\star\|_2^2
\end{aligned}$$

Therefore, with probability with probability $1 - (1 + \log(\alpha T D^2))\delta$, for any $\tau, \tau_1 > 0$,

$$\begin{aligned}
\sum_{t=k+1}^{T} Z_t &\le \sum_{i=1}^{k}\sum_{j=1}^{T_i}\sum_{j=1}^{T_i} U_t \cdot \overline{Z}_{t_{i,j}} \\
&\le k L_{\mathrm{f}}^2 \log(k/\delta)\left(\frac{3}{2\alpha} + \frac{27}{\tau}\right) + \frac{\tau}{2}\sum_{t=k+1}^{T} \|z_{t-k} - z_\star\|_2^2
\end{aligned}$$

$\tau = \alpha/6$, $\delta \leftarrow \delta/(1 + \log(1 \vee \alpha T D^2))$, we have that with probability $1 - \delta$

$$\begin{aligned}
\sum_{t=k+1}^{T} Z_t - \frac{\alpha}{12}\sum_{t=1}^{T} \|z_t - z_\star\|_2^2 &\le \sum_{t=k+1}^{T} Z_t - \frac{\alpha}{12}\sum_{t=k+1}^{T} \|z_{t-k} - z_\star\|_2^2 \\
&\lesssim \frac{k L_{\mathrm{f}}^2}{\alpha}\log\left(\frac{k(1 + \log_+(\alpha T D^2))}{\delta}\right).
\end{aligned}$$

□.

K.3.3. HIGH PROBABILITY REGRET WITH MEMORY: PROOF OF THEOREM 9

Proof We reitarate the argument of Anava et al. (2015). Decompose

$$\begin{aligned}
\sum_{t=k+1}^{T} \tilde{f}_t(z_t, z_{t-1}, z_{t-2}, \ldots, z_{t-h}) - f_t(z_\star) = &\sum_{t=k+1}^{T} f_t(z_t) - f_t(z_\star) \\
&+ \sum_{t=k+1}^{T} \tilde{f}_t(z_t, z_{t-1}, z_{t-2}, \ldots, z_{t-h}) - f_t(z_t).
\end{aligned}$$

We can bound the first sum directly from Theorem 12. The second term can be bounded as follows:

$$\begin{aligned}
\sum_{t=k+1}^{T} \tilde{f}_t(z_t, z_{t-1}, z_{t-2}, \ldots, z_{t-h}) - f_t(z_t) &\leq L_{\mathrm{c}} \sum_{t=k+1}^{T} \|(0, z_{t-1} - z_t, \ldots, z_{t-h} - z_t)\|_2 \\
&\leq \sum_{t=k+1}^{T} \sum_{i=1}^{h} \|z_t - z_{t-i}\|_2 \\
&\leq L_{\mathrm{c}} \sum_{t=k+1}^{T} \sum_{i=1}^{h} \sum_{j=1}^{i} \|z_{t-j+1} - z_{t-j}\|_2 \\
&\leq L_{\mathrm{c}} \sum_{t=k+1}^{T} \sum_{i=1}^{h} \sum_{j=1}^{i} \eta_{t-j+1} \|\boldsymbol{g}_{t-j}\| \\
&\leq h L_{\mathrm{c}} \sum_{t=k+1}^{T} \sum_{i=1}^{h} \eta_{t-j+1} \|\boldsymbol{g}_{t-j}\| \\
&\leq h L_{\mathrm{c}} L_{\mathbf{g}} \sum_{t=k+1}^{T} \sum_{i=1}^{h} \eta_{t-j+1} \\
&\leq h^2 L_{\mathrm{c}} L_{\mathbf{g}} \sum_{t=1}^{T} \eta_{t+1} \lesssim \frac{h^2 L_{\mathrm{c}} L_{\mathbf{g}} \log T}{\alpha}
\end{aligned}$$

This establishes the desired bound. ■