Logarithmic Regret for Online Control

Naman Agarwal[1] Elad Hazan[1 2] Karan Singh[1 2]

[1] Google AI Princeton
[2] Department of Computer Science, Princeton University
namanagarwal@google.com, {ehazan,karans}@princeton.edu

September 11, 2019

Abstract

We study optimal regret bounds for control in linear dynamical systems under adversarially changing strongly convex cost functions, given the knowledge of transition dynamics. This includes several well studied and fundamental frameworks such as the Kalman filter and the linear quadratic regulator. State of the art methods achieve regret which scales as $O(\sqrt{T})$, where T is the time horizon.

We show that the optimal regret in this setting can be significantly smaller, scaling as $O(\text{poly}(\log T))$. This regret bound is achieved by two different efficient iterative methods, online gradient descent and online natural gradient.

1 Introduction

Algorithms for regret minimization typically attain one of two performance guarantees. For general convex losses, regret scales as square root of the number of iterations, and this is tight. However, if the loss function exhibit more curvature, such as quadratic loss functions, there exist algorithms that attain poly-logarithmic regret. This distinction is also known as "fast rates" in statistical estimation.

Despite their ubiquitous use in online learning and statistical estimation, logarithmic regret algorithms are almost non-existent in control of dynamical systems. This can be attributed to fundamental challenges in computing the optimal controller in the presence of noise.

Time-varying cost functions in dynamical systems can be used to model unpredictable dynamic resource constraints, and the tracking of a desired sequence of exogenous states. At a pinch, if we have changing (even, strongly) convex loss functions, the optimal controller for a linear dynamical system is not immediately computable via a convex program. For the special case of quadratic loss, some previous works [9] remedy the situation by taking a semi-definite relaxation, and thereby obtain a controller which has provable guarantees on regret and computational requirements. However, this semi-definite relaxation reduces the problem to regret minimization over linear costs, and removes the curvature which is necessary to obtain logarithmic regret.

In this paper we give the first efficient poly-logarithmic regret algorithms for controlling a linear dynamical system with noise in the dynamics (i.e. the standard model). Our results apply to general convex loss functions that are strongly convex, and not only to quadratics.

Reference	Noise	Regret	loss functions
[1]	**none**	$O(\log^2 T)$	quadratic (fixed hessian)
[4]	adversarial	$O(\sqrt{T})$	convex
[9]	stochastic	$O(\sqrt{T})$	quadratic
here	stochastic	$O(\log^7 T)$	strongly convex

1.1 Our Results

The setting we consider is a linear dynamical system, a continuous state Markov decision process with linear transitions, described by the following equation:

$$x_{t+1} = Ax_t + Bu_t + w_t. \tag{1.1}$$

Here x_t is the state of the system, u_t is the action (or control) taken by the controller, and w_t is the noise. In each round t, the learner outputs an action u_t upon observing the state x_t and incurs a cost of $c_t(x_t, u_t)$, where c_t is convex. The objective here is to choose a sequence of adaptive controls u_t so that a minimum total cost may be incurred.

The approach taken by [9] and other previous works is to use a semi-definite relaxation for the controller. However, this removes the properties associated with the curvature of the loss functions, by reducing the problem to an instance of online linear optimization. It is known that without curvature, $O(\sqrt{T})$ regret bounds are tight (see [13]).

Therefore we take a different approach, initiated by [4]. We consider controllers that depend on the previous noise terms, and take the form $u_t = \sum_{i=1}^{H} M_i w_{t-i}$. While this resulting convex relaxation does not remove the curvature of the loss functions altogether, it results in an overparametrized representation of the controller, and it is not a priori clear that the loss functions are strongly convex with respect to the parameterization. We demonstrate the appropriate conditions on the linear dynamical system under which the strong convexity is retained.

Henceforth we present two methods that attain poly-logarithmic regret. They differ in terms of the regret bounds they afford and the computational cost of their execution. The online gradient descent update (OGD) requires only gradient computation and update, whereas the online natural gradient (ONG) update, in addition, requires the computation of the preconditioner, which is the expected Gram matrix of the Jacobian, denoted J, and its inverse. However, the natural gradient update admits an instance-dependent upper bound on the regret, which while being at least as good as the regret bound on OGD, offers better performance guarantees on benign instances (See Corollary 4.5, for example).

Algorithm	Update rule (simplified)	Applicability
OGD	$M_{t+1} \leftarrow M_t - \eta_t \nabla f_t(M_t)$	$\exists K$, diag L s.t. $A - BK = QLQ^{-1}$
ONG	$M_{t+1} \leftarrow M_t - \eta_t (\mathbb{E}[J^\top J])^{-1} \nabla f_t(M_t)$	$\|L\| \leq 1 - \delta$, $\|Q\|, \|Q\|^{-1} \leq \kappa$

1.2 Related Work

For a survey of linear dynamical systems (LDS), as well as learning, prediction and control problems, see [17]. Recently, there has been a renewed interest in learning dynamical systems in the machine learning literature. For fully-observable systems, sample complexity and regret bounds for control (under Gaussian noise) were obtained in [3, 10, 2]. The technique of spectral filtering for learning and open-loop control of partially observable systems was introduced and studied in [15, 7, 14]. Provable control in the Gaussian noise setting via the policy gradient method was also studied in [11].

The closest work to ours is that of [1] and [9], aimed at controlling LDS with adversarial loss functions. The authors in [3] obtain a $O(\log^2 T)$ regret algorithm for changing quadratic costs (with a fixed hessian), but for dynamical systems that are noise-free. In contrast, our results apply to the full (noisy) LDS setting, which presents the main challenges as discussed before. Cohen et al. [9] consider changing quadratic costs with stochastic noise to achieve a $O(\sqrt{T})$ regret bound.

We make extensive use of techniques from online learning [8, 16, 13]. Of particular interest to our study is the setting of online learning with memory [5]. We also build upon the recent control work of [4], who use online learning techniques and convex relaxation to obtain provable bounds for LDS with adversarial perturbations.

2 Problem Setting

We consider a linear dynamical system as defined in (1.1) with costs $c_t(x_t, u_t)$, where c_t is strongly convex. In this paper we assume that the noise w_t is a random variable generated independently at every time step. For any algorithm $\mathcal{A}$, we attribute a cost defined as

$$J_T(\mathcal{A}) = \mathbb{E}_{\{w_t\}}\left[\sum_{t=1}^{T} c_t(x_t, u_t)\right],$$

where $x_{t+1} = Ax_t + Bu_t + w_t$, $u_t = \mathcal{A}(x_1, \ldots x_t)$ and $\mathbb{E}_{\{w_t\}}$ represents the expectation over the entire noise sequence. For the rest of the paper we will drop the subscript $\{w_t\}$ from the expectation as it will be the only source of randomness. Overloading notation, we shall use $J_T(K)$ to denote the cost of a linear controller K which chooses the action as $u_t = -Kx_t$.

Assumptions. In the paper we assume that $x_1 = 0$ [1], as well as the following conditions.

Assumption 2.1. *We assume that $\|B\| \leq \kappa_B$. Furthermore, the perturbation introduced per time step is bounded, i.i.d, and zero-mean with a lower bounded covariance i.e.*

$$\forall t \; w_t \sim \mathcal{D}_w, \mathbb{E}[w_t] = 0, \mathbb{E}[w_t w_t^\top] \succeq \sigma^2 I \text{ and } \|w_t\| \leq W$$

While we make the assumption that the noise vectors are bounded with probability 1, we can generalize to the case of sub-gaussian noise by conditioning on the event that none of the noise vectors are ever large. This can be done at an expense of another multiplicative $\log(T)$ factor in the regret. Furthermore we assume the following,

Assumption 2.2. *The costs $c_t(x, u)$ are α-strongly convex. Further, as long as it is guaranteed that $\|x\|, \|u\| \leq D$, it holds that*

$$\|\nabla_x c_t(x, u)\|, \|\nabla_u c_t(x, u)\| \leq GD.$$

The class of linear controllers we work with are defined as follows.

Definition 2.3 (Diagonal Strong Stability)**.** *Given a dynamics (A, B), a linear policy/matrix K is (κ, γ)-diagonal strongly stable for real numbers $\kappa \geq 1, \gamma < 1$, if there exists a complex diagonal matrix L and a non-singular complex matrix Q, such that $A - BK = QLQ^{-1}$ and the following conditions are met:*

1. *The spectral norm of L is strictly smaller than one, i.e., $\|L\| \leq 1 - \gamma$.*
2. *The controller and the transforming matrices are bounded, i.e., $\|K\| \leq \kappa$ and $\|Q\|, \|Q^{-1}\| \leq \kappa$.*

The notion of strong stability was introduced by [9]. Both strong stability and diagonal strong stability are quantitative measures of the classical notion of stabilizing controllers [2] that permit a discussion on non-asymptotic regret bounds. We note that an analogous notion for quantification of open-loop stability appears in the work of [14].

On the generality of the diagonal strong stability notion, the following comment may be made: while not all matrices are complex diagonalizable, an exhaustive characterization of $m \times m$ complex diagonal matrices is the existence of m linearly independent eigenvectors; for the later, it suffices, but is not necessary, that a matrix has m distinct eigenvalues (See [18]). It may be observed that almost all matrices admit distinct eigenvalues, and hence, are complex diagonalizable insofar the complement set admits a zero-measure. By this discussion, *almost all* stabilizing controllers are diagonal strongly stable for some κ, γ. The astute reader may note the departure here from the more general notion – strongly stability – in that *all* stabilizing controllers are strongly stable for some choice of parameters.

[1] This is only for convenience of presentation. The case with a bounded x_1 can be handled similarly.

[2] A controller K is stabilizing if the spectral radius of $A - BK \leq 1 - \delta$

Regret Formulation. Let $\mathcal{K} = \{K : K \text{ is } (\kappa, \gamma)\text{-diagonal strongly stable}\}$. For an algorithm $\mathcal{A}$, the notion of regret we consider is *pseudo-regret*, i.e. the sub-optimality of its cost with respect to the cost for the best linear controller i.e.,

$$\texttt{Regret} = J_T(\mathcal{A}) - \min_{K \in \mathcal{K}} J_T(K).$$

3 Preliminaries

Notation. We reserve the letters x, y for states and u, v for actions. We denote by d_x, d_u to be the dimensionality of the state and the control space respectively. Let $d = \max(d_x, d_u)$. We reserve capital letters A, B, K, M for matrices associated with the system and the policy. Other capital letters are reserved for universal constants in the paper. We use the shorthand $M_{i:j}$ to denote a subsequence $\{M_i, \ldots, M_j\}$. For any matrix U, define U_{vec} to be a flattening of the matrix where we stack the columns upon each other. Further for a collection of matrices $M = \{M^{[i]}\}$, let M_{vec} be the flattening defined by stacking the flattenings of $M^{[i]}$ upon each other. We use $\|x\|_U^2 = x^\top U x$ to denote the matrix induced norm. The rest of this section provides a recap of the relevant definitions and concepts introduced in [4].

3.1 Reference Policy Class

For the rest of the paper, we fix a (κ, γ)-diagonally strongly stable matrix $\mathbb{K}$ (The bold notation is to stress that we treat this matrix as fixed and not a parameter). Note that this can be any such matrix and it can be computed via a semi-definite feasibility program [9] given the knowledge of the dynamics, before the start of the game. We work with following the class of policies.

Definition 3.1 (Disturbance-Action Policy)**.** *A disturbance-action policy $M = (M^{[0]}, \ldots, M^{[H-1]})$, for horizon $H \geq 1$ is defined as the policy which at every time t, chooses the recommended action u_t at a state x_t, defined [3] as*

$$u_t(M) \triangleq -\mathbb{K}x_t + \sum_{i=1}^{H} M^{[i-1]} w_{t-i}.$$

For notational convenience, here it may be considered that $w_i = 0$ for all $i < 0$.

The policy applies a linear transformation to the disturbances observed in the past H steps. Since (x, u) is a linear function of the disturbances in the past under a linear controller K, formulating the policy this way can be seen as a relaxation of the class of linear policies. Note that $\mathbb{K}$ is a fixed matrix and is not part of the parameterization of the policy. As was established in [4] (and we include the proof for completeness), with the appropriate choice of parameters, superimposing such a $\mathbb{K}$, to the policy class allows it to approximate any linear policy in terms of the total cost suffered with a finite horizon parameter H.

We refer to the policy played at time t as $M_t = \{M_t^{[i]}\}$ where the subscript t refers to the time index and the superscript $[i-1]$ refers to the action of M_t on w_{t-i}. Note that such a policy can be executed because w_{t-1} is perfectly determined on the specification of x_t as $w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}$.

3.2 Evolution of State

This section describes the evolution of the state of the linear dynamical system under a non-stationary policy composed of a sequence of T policies, where at each time the policy is specified by $M_t = (M_t^{[0]}, \ldots, M_t^{[H-1]})$. We will use $M_{0:T-1}$ to denote such a non-stationary policy. The following definitions ease the burden of notation.

1. Define $\tilde{A} = A - B\mathbb{K}$. $\tilde{A}$ shall be helpful in describing the evolution of state starting from a non-zero state in the absence of disturbances.

[3] x_t is completely determined given $w_0 \ldots w_{t-1}$. Hence, the use of x_t only serves to ease the burden of presentation.

2. For any sequence of matrices $M_{0:H}$, define Ψ_i as a linear function that describes the effect of w_{t-i} on the state x_t, formally defined below.

Definition 3.2. *For any sequence of matrices $M_{0:H}$, define the disturbance-state transfer matrix Ψ_i for $i \in \{0, 1, \ldots H\}$, to be a function with $h+1$ inputs defined as*

$$\Psi_i(M_{0:H}) \triangleq \tilde{A}^i \mathbf{1}_{i \leq H} + \sum_{j=0}^{H} \tilde{A}^j B M_{H-j}^{[i-j-1]} \mathbf{1}_{i-j \in [1,H]}.$$

It will be important to note that ψ_i is a **linear** function of its argument.

3.3 Surrogate State and Surrogate Cost

This section introduces a couple of definitions required to describe our main algorithm. In essence they describe a notion of state, its derivative and the expected cost if the system evolved solely under the past H steps of a non-stationary policy.

Definition 3.3 (Surrogate State & Surrogate Action)**.** *Given a sequence of matrices $M_{0:H+1}$ and $2H$ independent invocations of the random variable w given by $\{w_j \sim \mathcal{D}_w\}_{j=0}^{2H-1}$, define the following random variables denoting the surrogate state and the surrogate action:*

$$y(M_{0:H}) = \sum_{i=0}^{2H} \Psi_i(M_{0:H}) w_{2H-i-i},$$

$$v(M_{0:H+1}) = -\mathbb{K} y(M_{0:H}) + \sum_{i=1}^{H} M_{H+1}^{[i-1]} w_{2H-i}.$$

When M is the same across all arguments we compress the notation to $y(M)$ and $v(M)$ respectively.

Definition 3.4 (Surrogate Cost)**.** *Define the surrogate cost function f_t to be the cost associated with the surrogate state and the surrogate action defined above, i.e.,*

$$f_t(M_{0:H+1}) = \mathbb{E}\left[c_t(y(M_{0:H}), v(M_{0:H+1}))\right].$$

When M is the same across all arguments we compress the notation to $f_t(M)$.

Definition 3.5 (Jacobian)**.** *Let $z(M) = \begin{bmatrix} y(M) \\ v(M) \end{bmatrix}$. Since $y(M), v(M)$ are random linear functions of M, $z(M)$ can be reparameterized as $z(M) = JM_{vec} = \begin{bmatrix} J_y \\ J_v \end{bmatrix} M_{vec}$, where J is a random matrix, which derives its randomness from the random perturbations w_i.*

3.4 OCO with Memory

We now describe the setting of online convex optimization with memory introduced in [5]. In this setting, at every step t, an online player chooses some point $x_t \in \mathcal{K} \subset \mathbb{R}^d$, a loss function $f_t : \mathcal{K}^{H+1} \mapsto \mathbb{R}$ is then revealed, and the learner suffers a loss of $f_t(x_{t-H:t})$. We assume a certain coordinate-wise Lipschitz regularity on f_t of the form such that, for any $j \in \{0, \ldots, H\}$, for any $x_{0:H}, \tilde{x}_j \in \mathcal{K}$,

$$|f_t(x_{0:j-1}, x_j, x_{j+1:H}) - f_t(x_{0:j-1}, \tilde{x}_j, x_{j+1:H})| \leq L\|x_j - \tilde{x}_j\|. \tag{3.1}$$

In addition, we define $f_t(x) = f_t(x, \ldots, x)$, and we let

$$G_f = \sup_{t \in \{0,\ldots,T\}, x \in \mathcal{K}} \|\nabla f_t(x)\|, \quad D = \sup_{x,y \in \mathcal{K}} \|x - y\|. \tag{3.2}$$

The resulting goal is to minimize the *policy regret* [6], which is defined as

$$\texttt{PolicyRegret} = \sum_{t=H}^{T} f_t(x_{t-H:t}) - \min_{x \in \mathcal{K}} \sum_{t=H}^{T} f_t(x).$$

Algorithm 1 Online Control Algorithm

1: **Input:** Step size schedule η_t, Parameters $\kappa_B, \kappa, \gamma, T$.
2: Define $H = \gamma^{-1} \log(T\kappa^2)$
3: Define $\mathcal{M} = \{M = \{M^{[0]} \dots M^{[H-1]}\} : \|M^{[i-1]}\| \leq \kappa^3 \kappa_B (1-\gamma)^i\}$.
4: Initialize $M_0 \in \mathcal{M}$ arbitrarily.
5: **for** $t = 0, \dots, T-1$ **do**
6: Choose the action:

$$u_t = -\mathbb{K}x_t + \sum_{i=1}^{H} M_t^{[i-1]} w_{t-i}.$$

7: Observe the new state x_{t+1} and record $w_t = x_{t+1} - Ax_t - Bu_t$.
8: **Online Gradient Update:**

$$M_{t+1} = \Pi_{\mathcal{M}}(M_t - \eta_t \nabla f_t(M_t))$$

9: **Online Natural Gradient Update:**

$$M_{vec,t+1} = \Pi_{\mathcal{M}}(M_{vec,t} - \eta_t (\mathbb{E}[J^T J])^{-1} \nabla_{M_{vec,t}} f_t(M_t))$$

10: **end for**

4 Algorithms & Statement of Results

The two variants of our method are spelled out in Algorithm 1. Theorems 4.1 and 4.3 provide the main guarantees for the two algorithms.

Online Gradient Update

Theorem 4.1 (Online Gradient Update)**.** *Suppose Algorithm 1 (Online Gradient Update) is executed with* $\mathbb{K}$ *being any* (κ, γ)*-diagonal strongly stable matrix and* $\eta_t = \Theta\left(\alpha\sigma^2 t\right)^{-1}$*, on an LDS satisfying Assumption 2.1 with control costs satisfying Assumption 2.2. Then, it holds true that*

$$J_T(\mathcal{A}) - \min_{K \in \mathcal{K}} J_T(K) \leq \tilde{O}\left(\frac{G^2 W^4}{\alpha\sigma^2} \log^7(T)\right).$$

The above result leverages the following lemma which shows that the function $f_t(\cdot)$ is strongly convex with respect to its argument M. Note that strong convexity of the cost functions c_t over the state-action space does not by itself imply the strong convexity of the surrogate cost f_t over the space of controllers M. This is because, in the surrogate cost f_t, c_t is applied to $y(M), v(M)$ which themselves are linear functions of M; the linear map M is necessarily column-rank-deficient. To observe this, note that M maps from a space of dimensionality $H \times \dim(x) \times \dim(u)$ to that of $\dim(x) + \dim(u)$. The next theorem, which forms the core of our analysis, shows that this is not the case using the inherent stochastic nature of the dynamical system.

Lemma 4.2. *If the cost functions* $c_t(\cdot, \cdot)$ *are* α*-strongly convex,* $\mathbb{K}$ *is a* (κ, γ) *diagonal strongly stable matrix and Assumption 2.1 is met then the idealized functions* $f_t(M)$ *are* λ*-strongly convex with respect to* M *where*

$$\lambda = \frac{\alpha\sigma^2\gamma^2}{36\kappa^{10}}$$

We present the proof of simpler instances, including a one dimensional version of the theorem, in Section 8, as they present the core ideas without the tedious notation necessitated by the general setting. We provide the general proof in Section D of the Appendix.

Online Natural Gradient Update

Theorem 4.3 (Online Natural Gradient Update)**.** *Suppose Algorithm 1 (Online Natural Gradient Update) is executed with $\eta_t = \Theta(\alpha t)^{-1}$, on an LDS satisfying Assumptions 2.1 and with control costs satisfying Assumption 2.2. Then, it holds true that*

$$J_T(\mathcal{A}) - \min_{K \in \mathcal{K}} J_T(K) \leq \tilde{O}\left(\frac{GW^2}{\alpha\mu}\log^7(T)\right) \quad \textit{where} \quad \mu^{-1} \triangleq \max_{M \in \mathcal{M}} \|(\mathbb{E}[J^T J])^{-1}\nabla_{M_{vec}} f_t(M)\|.$$

In Theorem 4.3, the regret guarantee depends on an instance-dependent parameter μ, which is a measure of hardness of the problem. First, we note that the proof of Lemma 4.2 establishes that the Gram matrix of the Jacobian (Defintion 3.5) is strictly positive definite and hence we recover the logarithmic regret guarantee achieved by the Online Gradient Descent Update, with the constants preserved.

Corollary 4.4. *In addition to the assumptions in Theorem 4.3, if $\mathbb{K}$ is a (κ, γ)-diagonal strongly stable matrix, then for the natural gradient update*

$$J_T(\mathcal{A}) - \min_{K \in \mathcal{K}} J_T(K) \leq \tilde{O}\left(\frac{G^2W^4}{\alpha\sigma^2}\log^7(T)\right),$$

Proof. The conclusion follows from Lemma 5.2 and Lemma 8.1 which is the core component in the proof of Lemma 4.2 showing that $\mathbb{E}[J^T J] \geq \frac{\gamma^2\sigma^2}{36\kappa^{10}} \cdot \mathbb{I}$. □

Secondly, we note that, being instance-dependent, the guarantee the Natural Gradient update offers can potentially be stronger than that of the Online Gradient method. A case in point is the following corollary involving spherically symmetric quadratic costs, in which case the Natural Gradient update yields a regret guarantee under demonstrably more general conditions, in that the bound does not depend on the minimum eigenvalue of the covariance of the disturbances σ^2, unlike the one OGD affords [4].

Corollary 4.5. *Under the assumptions on Theorem 4.3, if the cost functions are of the form $c_t(x, u) = r_t(\|x\|^2 + \|u\|^2)$, where $r_t \in [\alpha, \beta]$ is an adversarially chosen sequence of numbers and $\mathbb{K}$ is chosen to be a (κ, γ)-diagonal strongly stable matrix, then the natural gradient update guarantees*

$$J_T(\mathcal{A}) - \min_{K \in \mathcal{K}} J_T(K) \leq \tilde{O}\left(\frac{\beta^2W^2}{\alpha}\log^7(T)\right),$$

Proof. It suffices to note $\|\nabla_{M_{\text{vec}}} f_t(M)\|_{(\mathbb{E}[J^T J])^{-2}} = \|\mathbb{E}[J^T(r_t \cdot I)JM_{\text{vec}}]\|_{(\mathbb{E}[J^T J])^{-2}} \leq \beta\|M_{\text{vec}}\|$. □

5 Reduction to Low Regret with Memory

The next lemma is a condensation of the results from [4] which we present in this form to highlight the reduction to OCO with memory. It shows that achieving low policy regret on the memory based function f_t is sufficient to ensure low regret on the overall dynamical system. Since the proof is essentially provided by [4], we provide it in the Appendix for completeness. Define,

$$\mathcal{M} \triangleq \{M = \{M^{[0]} \dots M^{[H-1]}\} : \|M^{[i-1]}\| \leq \kappa^3\kappa_B(1-\gamma)^i\}.$$

Lemma 5.1. *Let the dynamical system satisfy Assumption 2.1 and let $\mathbb{K}$ be any (κ, γ)-diagonal strongly stable matrix. Consider a sequence of loss functions $c_t(x, u)$ satisfying Assumption 2.2 and a sequence of policies $M_0 \dots M_T$ satisfying*

$$\texttt{PolicyRegret} = \sum_{t=0}^{T} f_t(M_{t-H-1:t}) - \min_{M \in \mathcal{M}} \sum_{t=0}^{T} f_t(M) \leq R(T)$$

[4] A more thorough analysis of the improvement in this case shows a multiplicative gain of $\frac{WDH\sqrt{d}\kappa^{10}}{\sigma^2\gamma^2}$. Furthermore, Theorem 4.3 and Corollary 4.4 hold more generally under strong stability of the comparator class and $\mathbb{K}$, as opposed to diagonal strong stability.

for some function $R(T)$ and f_t as defined in Definition 3.4. Let A be an online algorithm that plays the non-stationary controller sequence $\{M_0, \dots M_T\}$. Then as long as H is chosen to be larger than $\gamma^{-1}\log(T\kappa^2)$ we have that

$$J(A) - \min_{K^* \in \mathcal{K}} J(K^*) \le R(T) + O(GW^2 \log(T)),$$

Here $O(\cdot)$, $\Theta(\cdot)$ contain polynomial factors in $\gamma^{-1}, \kappa_B, \kappa, d$.

Lemma 5.2. *The function f_t as defined in Definition 3.4 is coordinate-wise L-lipschitz and the norm of the gradient is bounded by G_f, where*

$$L = \frac{2DGW\kappa_B\kappa^3}{\gamma}, \quad G_f \le GDWHd\left(H + \frac{2\kappa_B\kappa^3}{\gamma}\right)$$

$$\text{where } D \triangleq \frac{W\kappa^2(1 + H\kappa_B^2\kappa^3)}{\gamma(1-\kappa^2(1-\gamma)^{H+1})} + \frac{\kappa_B\kappa^3 W}{\gamma}.$$

The proof of this lemma is identical to the analogous lemma in [4] and hence is omitted.

6 Analysis for Online Gradient Descent

In the setting of Online Convex Optimization with Memory, as shown by [5], by running a memory-based OGD, we can bound the policy regret by the following theorem.

Theorem 6.1. *Consider the OCO with memory setting defined in Section 3.4. Let $\{f_t\}_{t=H}^T$ be Lipschitz loss functions with memory such that $f_t(x)$ are λ-strongly convex, and let L and G_f be as defined in (3.1) and (3.2). Then, there exists an algorithm which generates a sequence $\{x_t\}_{t=0}^T$ such that*

$$\sum_{t=H}^{T} f_t(x_{t-H:t}) - \min_{x\in\mathcal{K}} \sum_{t=H}^{T} \tilde{f}_t(x) \le \frac{G_f^2 + LH^2 G_f}{\lambda}(1 + \log(T)).$$

We provide the requisite algorithm and the proof of the above theorem in the Appendix.

Specialization to the Control Setting: We combine bound the above with the listed reduction.

Proof of Theorem 4.1. Setting $H = \gamma^{-1}\log(T\kappa^2)$, Theorem 6.1, in conjunction with Lemma 5.2, implies that policy regret is bounded by $\tilde{O}\left(\frac{G^2W^4H^6}{\alpha\sigma^2}\log T\right)$. An invocation of Lemma 5.1 now suffices to conclude the proof of the claim. □

7 Analysis for Online Natural Gradient Descent

In this section, we consider structured loss functions of the form $f_t(M_{0:H+1}) = \mathbb{E}[c_t(z)]$, where $z = \sum_{i=0}^{H+1} J_i[M_i]_{\text{vec}}$. J_i is a random matrix, and c_t's are adversarially chosen strongly convex loss functions. In a similar vein, define $f_t(M)$ to be the specialization of f_t when input the same argument, i.e. M, $H+1$ times. Define $J = \sum_{i=0}^{H+1} J_i$.

The following lemma provides upper bounds on the regret bound as well as the norm of the movement of iterate at every round for the Online Natural Gradient Update (Algorithm 1).

Lemma 7.1. *For α-strongly convex c_t, if the iterates M_t are chosen as per the update rule:*

$$[M_{t+1}]_{vec} = \Pi_{\mathcal{M}}\left([M_t]_{vec} - \eta_t(\mathbb{E}[J^T J])^{-1}\nabla_{[M_t]_{vec}} f_t(M_t)\right)$$

with a decreasing step size of $n_t = \frac{1}{\alpha t}$, *it holds that*

$$\sum_{t=1}^{T} f_t(M_t) - \min_{M^* \in \mathcal{M}} \sum_{t=1}^{T} f_t(M^*) \leq (2\alpha)^{-1} \max_{M \in \mathcal{M}} \|\nabla_{M_{vec}} f_t(M)\|^2_{(\mathbb{E}[J^T J])^{-1}} \log T.$$

Moreover, the norm of the movement of consecutive iterates is bounded for all t *as*

$$\|[M_{t+1}]_{vec} - [M_t]_{vec}\| \leq (\alpha t)^{-1} \max_{M \in \mathcal{M}} \|(\mathbb{E}[J^T J])^{-1} \nabla_{M_{vec}} f_t(M)\|.$$

The following theorem now bounds the total for the online game with memory.

Theorem 7.2. *In the setting desribed in this subsection, let* c_t *be* α*-strongly convex, and* f_T *be such that it satisfies equation (3.1) with constant* L*, and* $G_f = \max_{M \in \mathcal{M}} \|(\mathbb{E}[J^T J])^{-1} \nabla_{M_{vec}} f_t(M)\|$. *Then, the online natural gradient update generates a sequence* $\{M_t\}_{t=0}^{T}$ *such that*

$$\sum_{t=H}^{T} f_t(M_{t-H:t}) - \min_{M \in \mathcal{M}} \sum_{t=H}^{T} \tilde{f}_t(M) \leq \frac{\max_{M \in \mathcal{M}} \|\nabla_{M_{vec}} f_t(M)\|^2_{(\mathbb{E}[J^T J])^{-1}} + LH^2 G_f}{\alpha} (1 + \log(T)).$$

Proof of Theorem 7.2. We know by (3.1) that, for any $t \geq H$,

$$\begin{aligned}
|f_t(M_{t-H:t}) - f_t(M)| &\leq L \sum_{j=1}^{H} \|[M_t]_{\text{vec}} - [M_{t-j}]_{\text{vec}}\| \leq L \sum_{j=1}^{H} \sum_{l=1}^{j} \|[M_{t-l+1}]_{\text{vec}} - [M_{t-l}]_{\text{vec}}\| \\
&\leq L \sum_{j=1}^{H} \sum_{l=1}^{j} \eta_{t-l} \max_{M \in \mathcal{M}} \|(\mathbb{E}[J^T J])^{-1} \nabla_{M_{\text{vec}}} f_t(M)\| \\
&\leq LH^2 \eta_{t-H} \max_{M \in \mathcal{M}} \|(\mathbb{E}[J^T J])^{-1} \nabla_{M_{\text{vec}}} f_t(M)\|,
\end{aligned}$$

and so we have that

$$\left| \sum_{t=H}^{T} f_t(M_{t-H:t}) - \sum_{t=H}^{T} f_t(M_t) \right| \leq \frac{LH^2 G_f}{\alpha} (1 + \log(T)).$$

The result follows by invoking Lemma 7.1. ☐

Specialization to the Control Setting: We combine bound the above with the listed reduction.

Proof of Theorem 4.3. First observe that $\|\nabla_{M_{\text{vec}}} f_t(M)\|^2_{(\mathbb{E}[J^T J])^{-1}} \leq \mu^{-1} \|\nabla_{M_{\text{vec}}} f_t(M)\|$. Setting $H = \gamma^{-1} \log(T\kappa^2)$, Theorem 7.2, in conjunction with Lemma 5.2, imply the stated bound on policy regret. An invocation of Lemma 5.1 suffices to conclude the proof of the claim. ☐

8 Proof of Strong Convexity in simpler cases

In this section we illustrate the proof of strong convexity of the function $f_t(M)$ with respect to M, i.e. Lemma 4.2, in two settings.

1. The case when $\mathbb{K} = 0$ is a diagonal strongly stable policy.

2. A specialization of Lemma 4.2 to one-dimensional state and one-dimensional control.

This latter case highlights the difficulty caused in the proof due to a choosing a non-zero $\mathbb{K}$ and presents the main ideas of the proof without the tedious tensor notations necessary for the general case.

We will need some definitions and preliminaries that are outlined below. By definition we have that $f_t(M) = \mathbb{E}[c_t(y_t(M), v_t(M))]$. Since we know that c_t is strongly convex we have that

$$\nabla^2 f_t(M) = \mathbb{E}_{\{w_k\}_{k=0}^{2H-1}}[\nabla^2 c_t(y(M), v(M))] \succeq \alpha \mathbb{E}_{\{w_k\}_{k=0}^{2H-1}}[J_y^\top J_y + J_v^\top J_v].$$

We remind the reader that J_y, J_v are random matrices dependent on the noise vectors $\{w_k\}_{k=0}^{2H-1}$. In each of the above cases, we will demonstrate the truth of the following lemma implying Lemma 4.2.

Lemma 8.1. *If Assumption 2.1 is satisfied and $\mathbb{K}$ is chosen to be a (κ, γ)-diagonal strongly stable matrix, then the following holds,*

$$\mathbb{E}_{\{w_k\}_{k=0}^{2H-1}}[J_y^\top J_y + J_v^\top J_v] \succeq \frac{\gamma^2 \sigma^2}{36\kappa^{10}} \cdot \mathbb{I}.$$

To analyze J_y, J_v, we will need to rearrange the definition of $y(M)$ to make the dependence on each individual $M^{[i]}$ explicit. To this end consider the following definition for all $k \in [H+1]$.

$$\tilde{v}_k(M) \triangleq \sum_{i=1}^{H} M^{[i-1]} w_{2H-i-k}$$

Under this definition it follows that

$$y(M) = \sum_{k=1}^{H} (A - B\mathbb{K})^{k-1} B \tilde{v}_k(M) + \sum_{k=1}^{H} (A - B\mathbb{K})^{k-1} w_{2H-k}$$

$$v(M) = -\mathbb{K} y(M) + \tilde{v}_0(M)$$

From the above definitions, (J_y, J_v) may be characterized in terms of the Jacobian of $\tilde{v}_k$ with respect to M, which we define for the rest of the section as $J_{\tilde{v}_k}$. Defining M_{vec} as the stacking of rows of each $M^{[i]}$ vertically, i.e. stacking the columns of $(M^{[i]})^\top$, it can be observed that for all k,

$$J_{\tilde{v}_k} = \frac{\partial \tilde{v}_k(M)}{\partial M} = \begin{bmatrix} I_{d_u} \otimes w_{2H-k-1}^\top & I_{d_u} \otimes w_{2H-k-2}^\top & \dots & I_{d_u} \otimes w_{H-k}^\top \end{bmatrix}$$

where d_u is the dimension of the controls. We are now ready to analyze the two simpler cases. Further on in the section we drop the subscripts $\{w_k\}_{k=0}^{2H-1}$ from the expectations for brevity.

8.1 Proof of Lemma 8.1: $\mathbb{K} = 0$

In this section we assume that $\mathbb{K} = 0$ is a (κ, γ)-diagonal strongly stable policy for (A, B). Be definition, we have $v(M) = \tilde{v}_0(M)$. One may conclude the proof with the following observation.

$$\mathbb{E}[J_y^\top J_y + J_v^\top J_v] \succeq \mathbb{E}[J_v^\top J_v] = \mathbb{E}[J_{\tilde{v}_0}^\top J_{\tilde{v}_0}] = I_{d_u} \otimes \Sigma \succeq \sigma^2 \mathbb{I}.$$

8.2 Proof of Lemma 8.1: 1-dimensional case

Note that in the one dimensional case, the policy given by $M = \{M^{[i]}\}_{i=0}^{H-1}$ is an H dimensional vector with $M^{[i]}$ being a scalar. Furthermore $y(M), v(M), \tilde{v}_k(M)$ are scalars and hence their Jacobians $J_y, J_v, J_{\tilde{v}_k}$ with respect to M are $1 \times H$ vectors. In particular we have that,

$$J_{\tilde{v}_k} = \frac{\partial \tilde{v}_k(M)}{\partial M} = \begin{bmatrix} w_{2H-k-1} & w_{2H-k-2} & \dots & w_{H-k} \end{bmatrix}$$

Therefore using the fact that $E[w_iw_j] = 0$ for $i \neq j$ and $\mathbb{E}[w_i^2] = \sigma^2$, it can be observed that for any k_1, k_2, we have that

$$\mathbb{E}[J_{v_{\tilde{k}_1}}^\top J_{v_{\tilde{k}_2}}] = \mathcal{T}_{k_1-k_2} \cdot \sigma^2 \tag{8.1}$$

where $\mathcal{T}_m$ is defined as an $H \times H$ matrix with $[\mathcal{T}_m]_{ij} = 1$ if and only if $i - j = m$ and 0 otherwise. This in particular immediately gives us that,

$$\mathbb{E}[J_y^\top J_y] = \underbrace{\left(\sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \cdot (A - B\mathbb{K})^{k_1-1+k_2-1}\right)}_{\triangleq \mathbb{G}} \cdot B^2 \cdot \sigma^2 \tag{8.2}$$

$$\mathbb{E}[J_{\tilde{v}_0}^\top J_y] = \underbrace{\left(\sum_{k=1}^{H} \mathcal{T}_{-k}(A - B\mathbb{K})^{k-1}\right)}_{\triangleq \mathbb{Y}} \cdot B \cdot \sigma^2 \tag{8.3}$$

First, we prove a few spectral properties of the matrices $\mathbb{G}$ and $\mathbb{Y}$ defined above. From Gershgorin's circle theorem, and the fact that $\mathbb{K}$ is (κ, γ)-diagonal strongly stable, we have

$$\|\mathbb{Y} + \mathbb{Y}^\top\| \leq \|\sum_{k=1}^{H}(\mathcal{T}_{-k} + \mathcal{T}_k)(A - B\mathbb{K})^{k-1}\| \leq 2\gamma^{-1} \tag{8.4}$$

The spectral properties of $\mathbb{G}$ summarized in the lemma below form the core of our analysis.

Lemma 8.2. *$\mathbb{G}$ is a symmetric positive definite matrix. In particular*

$$\mathbb{G} \succeq \frac{1}{4} \cdot I.$$

Now consider the statements which follow by the respective definitions.

$$\begin{aligned}\mathbb{E}[J_v^\top J_v] &= \mathbb{K}^2 \cdot \mathbb{E}[J_y^\top J_y] - \mathbb{K} \cdot \mathbb{E}[J_y^\top J_{\tilde{v}_0}] - \mathbb{K} \cdot \mathbb{E}[J_{\tilde{v}_0}^\top J_y] + \mathbb{E}[J_{\tilde{v}_0}^\top J_{\tilde{v}_0}] \\ &= \sigma^2 \cdot \underbrace{\left(B^2\mathbb{K}^2 \cdot \mathbb{G} - B\mathbb{K} \cdot (\mathbb{Y} + \mathbb{Y}^\top) + I\right)}_{\triangleq \mathbb{F}}.\end{aligned}$$

Now $\mathbb{F} \succeq 0$. To prove Lemma 8.1, it suffices that for every vector m of appropriate dimensions, we have that

$$m^\top \left(\mathbb{F} + B^2 \cdot \mathbb{G}\right) m \geq \frac{\gamma^2\|m\|^2}{36\kappa^{10}}.$$

To prove the above we will consider two cases. The first case is when $3|B|\gamma^{-1}\kappa \geq 1$. Noting $\kappa \geq 1$, in this case Lemma 8.2 immediately implies that

$$m^\top \left(\mathbb{F} + B^2 \cdot \mathbb{G}\right) m \geq m^\top \left(B^2 \cdot \mathbb{G}\right) m \geq \frac{\frac{1}{4}\|m\|^2}{9\gamma^{-2}\kappa^2} \geq \frac{\gamma^2\|m\|^2}{36\kappa^{10}},$$

In the second case (when $3|B|\gamma^{-1}\kappa \leq 1$), (8.4) implies that

$$m^\top \left(\mathbb{F} + B^2 \cdot \mathbb{G}\right) m \geq m^\top \left(I - B\mathbb{K} \cdot (\mathbb{Y} + \mathbb{Y}^\top)\right) m \geq (1/3)\|m\|^2 \geq \frac{\gamma^2\|m\|^2}{36\kappa^{10}}.$$

8.2.1 Proof of Lemma 8.2

Define the following matrix for any complex number $|\psi| < 1$.

$$\mathbb{G}(\psi) = \sum_{k_1=1}^{H} \sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \left(\psi^\dagger\right)^{k_1-1} \psi^{k_2-1}$$

Note that $\mathbb{G}$ in Lemma 8.2 is equal to $\mathbb{G}(A - B\mathbb{K})$. The following lemma provides a lower bound on the spectral properties of the matrix $\mathbb{G}(\psi)$. The lemma presents the proof of a more general case (ϕ is complex) that while unnecessary in the one dimensional case, aids the multi-dimensional case. A special case when $\phi = 1$ was proven in [12], and we follow a similar approach relying on the inverse of such matrices.

Lemma 8.3. *Let ψ be a complex number such that $|\psi| \leq 1$. Furthermore let $\mathcal{T}_m$ is defined as an $H \times H$ matrix with $[\mathcal{T}_m]_{ij} = 1$ if and only if $i - j = m$ and 0 otherwise. Define the matrix $\mathbb{G}(\psi)$ as*

$$\mathbb{G}(\psi) = \sum_{k_1=1}^{H} \sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \left(\psi^\dagger\right)^{k_1-1} \psi^{k_2-1}.$$

We have that

$$\mathbb{G}(\psi) \succeq (1/4) \cdot I_H$$

8.2.2 Proof of Lemma 8.3

Proof of Lemma 8.3. The following definitions help us express the matrix $\mathbb{G}$ in a more convenient form. For any number $\psi \in \mathbb{C}$, such that $|\psi| < 1$ and any h define,

$$S_\psi(h) = \sum_{i=1}^{h} |\psi|^{2(i-1)} = \frac{1 - |\psi|^{2h}}{1 - |\psi|^2}.$$

With the above definition it can be seen that the entries $\mathbb{G}(\psi)$ can be expressed in the following manner,

$$[\mathbb{G}(\psi)]_{ij} = S_\psi(H - |i-j|) \cdot \psi^{i-j} \qquad \text{if } j \geq i$$
$$[\mathbb{G}(\psi)]_{ij} = (\psi^\dagger)^{j-i} \cdot S_\psi(H - |i-j|) \qquad \text{if } i \geq j$$

Schematically the matrix $\mathbb{G}(\psi)$ looks like

$$\mathbb{G}(\psi) = \begin{bmatrix} S_\psi(H) & S_\psi(H-1)\psi & S_\psi(H-2)\psi^2 & . & . & S(2)\psi^{H-2} & S(1)\psi^{H-1} \\ \psi^\dagger S_\psi(H-1) & S_\psi(H) & S_\psi(H-1)\psi & . & . & S(3)\psi^{H-3} & S(2)\psi^{H-2} \\ (\psi^\dagger)^2 S_\psi(H-2) & \psi^\dagger S_\psi(H-1) & S_\psi(H) & . & . & S(4)\psi^{H-4} & S(3)\psi^{H-3} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ (\psi^\dagger)^{H-1} S_\psi(1) & (\psi^\dagger)^{H-2} S_\psi(2) & (\psi^\dagger)^{H-3} S_\psi(3) & . & . & \psi^\dagger S_\psi(H-1) & S_\psi(H) \end{bmatrix}.$$

We analytically compute the inverse of the matrix $\mathbb{G}(\psi)$ below and bound its spectral norm.

Claim 8.4. *The inverse of $\mathbb{G}(\psi)$ has the following form.*

$$[\mathbb{G}(\psi)]^{-1} = \begin{bmatrix} \alpha & b & 0 & . & . & 0 & 0 & \beta^\dagger \\ b^\dagger & a & b & . & . & 0 & 0 & 0 \\ 0 & b^\dagger & a & . & . & 0 & 0 & . \\ . & 0 & b^\dagger & . & . & b & 0 & . \\ . & 0 & 0 & . & . & a & b & 0 \\ 0 & 0 & 0 & . & . & b^\dagger & a & b \\ \beta & 0 & 0 & . & . & 0 & b^\dagger & \alpha \end{bmatrix},$$

where the relevant quantities above are given by the following formula

$$b = \frac{-\psi}{1+|\psi|^{2H}} \qquad a = -b(\psi^\dagger + \psi^{-1}) = \frac{1+|\psi|^2}{1+|\psi|^{2H}}$$

$$\beta = \frac{(1-|\psi|^2)}{(1-(|\psi|^2)^{H+1})} \frac{(\psi^\dagger)^H \psi}{(1+|\psi|^{2H})} \qquad \alpha = \frac{1-(|\psi|^2)^{H+2}}{(1-(|\psi|^2)^{H+1})(1+(|\psi|^{2H}))}.$$

Since $|\psi| < 1$, it is easy to see that $|\alpha|, |a| \leq 2$ and $|\beta|, |b| \leq 1$. This immediately implies that $\|(\mathbb{G}(\psi))^{-1}\| \leq 4$ and therefore the lemma follows.

To prove the remnant claim, the following may be verified, implying $\mathbb{G}(\psi)[\mathbb{G}(\psi)]^{-1} = I$.

- Lets first consider the diagonal entries and in particular $i = j \in [1, H-2]$. We have that

$$\left[\mathbb{G}(\psi)[\mathbb{G}(\psi)]^{-1}\right]_{i,i} = b \cdot \psi^\dagger S_\psi(H-1) + b^\dagger \cdot \psi S_\psi(H-1) + a S_\psi(H) = \frac{-2|\psi|^2 S_\psi(H-1) + (1+|\psi|^2) S_\psi(H)}{1+|\psi|^{2H}} = 1$$

- Lets consider the diagonal entry $(0,0)$. (The (H,H) entry is the complement and hence equal to 1).

$$\begin{aligned}\left[\mathbb{G}(\psi)[\mathbb{G}(\psi)]^{-1}\right]_{0,0} &= \alpha \cdot S_\psi(H) + b^\dagger \psi S_\psi(H-1) + \beta^\dagger (\psi^\dagger)^{H-1} S_\psi(1) \\ &= \frac{(1-(|\psi|^2)^{H+2}) S_\psi(H) - (1-(|\psi|^2)^{H+1})|\psi|^2 S_\psi(H-1) + (1-|\psi|^2)(|\psi|^{2H})}{(1-(|\psi|^2)^{H+1})(1+(|\psi|^{2H}))} \\ &= 1\end{aligned}$$

- Now lets consider non diagonal entries, in particular for $j \in [1, H-2]$ and $i \in [0, H-1]$ and $i > j$. (The case with the same conditions and $j > i$ follows by replacing ψ with $\psi^\dagger$ in the computation below)

$$\begin{aligned}\left[\mathbb{G}(\psi)[\mathbb{G}(\psi)]^{-1}\right]_{i,j} &= (\psi^\dagger)^{i-j-1}\left(b(\psi^\dagger)^2 S_{H-i+j-1} + b^\dagger S_{H-i+j+1} + a(\psi^\dagger) S_{H-i+j}\right) \\ &= (\psi^\dagger)^{i-j}\left(-|\psi|^2 S_{H-i+j-1} - S_{H-i+j+1} + (|\psi|^2+1) S_{H-i+j}\right) \\ &= 0\end{aligned}$$

- Lastly lets consider the first column, i.e. $j = 0$ and $i > 0$. (The case of the last column follows as it is the complement and hence equal to 0.)

$$\left[\mathbb{G}(\psi)[\mathbb{G}(\psi)]^{-1}\right]_{i,j} = \alpha \cdot (\psi^\dagger)^i S_\psi(H-i) + b \cdot (\psi^\dagger)^{i-1} S_\psi(H-i+1) + \beta \psi^{H-i-1} S_\psi(i+1) = 0.$$

□

9 Conclusion

We presented two algorithms for controlling linear dynamical systems with strongly convex costs, under certain stability assumptions, with regret that scales poly-logarithmically with time. This improves state-of-the-art known regret bounds that scale as $O(\sqrt{T})$. It remains open to extend the poly-log regret guarantees to more general systems and loss functions, such as exp-concave losses, or alternatively, show that this is impossible.

Acknowledgements

The authors thank Sham Kakade and Cyril Zhang for various thoughtful discussions. Elad Hazan acknowledges funding from NSF grant # CCF-1704860.

References

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

[2] Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117, 2019.

[3] 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.

[4] Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *Proceedings of the 36th International Conference on Machine Learning*, pages 111–119, 2019.

[5] 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.

[6] Raman Arora, Ofer Dekel, and Ambuj Tewari. Online bandit learning against an adaptive adversary: from regret to policy regret. In *Proceedings of the 29th International Conference on Machine Learning*, pages 1503–1510, 2012.

[7] Sanjeev Arora, Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. Towards provable control for unknown linear dynamical systems. 2018.

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

[9] Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1028–1037, 2018.

[10] 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.

[11] Maryam Fazel, Rong Ge, Sham M Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1466–1475, 2018.

[12] Surbhi Goel, Adam Klivans, and Raghu Meka. Learning one convolutional layer with overlapping patches. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1783–1791, 2018.

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

[14] 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.

[15] 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.

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

[17] Robert F Stengel. *Optimal control and estimation*. Courier Corporation, 1994.

[18] Gilbert Strang. *Introduction to linear algebra*, volume 3.

Algorithm 2 OGD with Memory (OGD-M).

1: **Input:** Step size η, functions $\{f_t\}_{t=m}^T$
2: Initialize $x_0, \dots, x_{H-1} \in \mathcal{K}$ arbitrarily.
3: **for** $t = H, \dots, T$ **do**
4: Play x_t, suffer loss $f_t(x_{t-H}, \dots, x_t)$
5: Set $x_{t+1} = \Pi_{\mathcal{K}} \left(x_t - \eta \nabla \tilde{f}_t(x) \right)$
6: **end for**

Appendix

A Proof of Theorem 6.1

Proof. By the standard OGD strong convexity analysis, if $\eta_t = (\lambda \cdot (t - H))^{-1}$, we have that

$$\sum_{t=H}^{T} \tilde{f}_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=H}^{T} \tilde{f}_t(x) \leq \frac{G^2}{2\lambda}(1 + \log(T)).$$

In addition, we know by (3.1) that, for any $t \geq H$,

$$\begin{aligned} |f_t(x_{t-H}, \dots, x_t) - f_t(x_t, \dots, x_t)| &\leq L \sum_{j=1}^{H} \|x_t - x_{t-j}\| \leq L \sum_{j=1}^{H} \sum_{l=1}^{j} \|x_{t-l+1} - x_{t-l}\| \\ &\leq L \sum_{j=1}^{H} \sum_{l=1}^{j} \eta_{t-l} \|\nabla \tilde{f}_{t-l}(x_{t-l})\| \leq LH^2 \eta_{t-H} G, \end{aligned}$$

and so we have that

$$\left| \sum_{t=H}^{T} f_t(x_{t-H}, \dots, x_t) - \sum_{t=H}^{T} f_t(x_t, \dots, x_t) \right| \leq \frac{LH^2G}{\lambda}(1 + \log(T)).$$

It follows that

$$\sum_{t=H}^{T} f_t(x_{t-H}, \dots, x_t) - \min_{x \in \mathcal{K}} \sum_{t=H}^{T} f_t(x, \dots, x) \leq \frac{G^2 + LH^2G}{\lambda}(1 + \log(T)).$$

□

B Proof of Lemma 7.1

Proof of Theorem 7.1. Let $M^* = \arg\min_{M \in \mathcal{M}} \sum_{t=1}^{T} f_t(M)$, $z_t = JM_{vec,t}$ and $z^* = JM^*_{\text{vec}}$. Now, we have, as consequence of strong convexity of c_t, that

$$\sum_{t=1}^{T} f_t(M_t) - \sum_{t=1}^{T} f_t(M^*) \leq \mathbb{E}\left[\langle \nabla_z c_t(z_t), z_t - z^* \rangle - \frac{\alpha}{2} \|z_t - z^*\|^2 \right].$$

With $P = \mathbb{E}[J^T J]$, the choice of the update rule ensures that

$$\|[M_{t+1}]_{\text{vec}} - M^*_{vec}\|_P^2 = \|[M_t]_{\text{vec}} - M^*_{\text{vec}}\|_P^2 - 2\eta_t \langle \nabla_{[M_t]_{\text{vec}}} f_t(M_t), [M_t]_{\text{vec}} - M^*_{\text{vec}} \rangle + \eta_t^2 \|\nabla_{[M_t]_{\text{vec}}} f_t(M_t)\|_{P^{-1}}.$$

Observe by the application of chain rule and linearity of expectation that

$$\begin{aligned}\mathbb{E}[\langle \nabla_z c_t(z_t), z_t - z^*\rangle] &= \mathbb{E}[\langle \nabla_z c_t(z_t), J([M_t]_{\text{vec}} - M^*_{\text{vec}})\rangle] \\ &= \langle \nabla_{[M_t]_{\text{vec}}} f_t(M_t), [M_t]_{\text{vec}} - M^*_{\text{vec}}\rangle, \\ \mathbb{E}[\|z_t - z^*\|^2] &= \|[M_t]_{\text{vec}} - M^*_{\text{vec}}\|_P^2.\end{aligned}$$

Combining these (in)equalities, we have

$$\begin{aligned}&\sum_{t=1}^T f_t(M_t) - \sum_{t=1}^T f_t(M^*) \\ \leq &\sum_{t=1}^T \left(\frac{\|[M_t]_{\text{vec}} - M^*_{vec}\|_P^2 - \|[M_{t+1}]_{\text{vec}} - M^*_{vec}\|_P^2}{2\eta_t} + \frac{\eta_t}{2}\|\nabla_{[M_t]_{\text{vec}}} f_t(M_t)\|^2_{P^{-1}} \right) \\ &- \frac{\alpha}{2}\|[M_t]_{vec} - M^*_{vec}\|_P^2 \\ \leq &(2\alpha)^{-1} \max_{M\in\mathcal{M}} \|\nabla_{M_{\text{vec}}} f_t(M)\|^2_{P^{-1}} \log T\end{aligned}$$

□

C Proof of Lemma 5.1

Since the proof of Lemma will borrow heavily from the definitions introduced by [4], we restate those definitions here for convenience. Please note that some of these definitions overload our previous definitions but it will be clear from the context.

C.1 Definitions

1. Let $x_t^K(M_{0:t-1})$ is the state attained by the system upon execution of a non-stationary policy $\pi(M_{0:t-1}, K)$. We similarly define $u_t^K(M_{0:t-1})$ to be the action executed at time t. If the same policy M is used across all time steps, we compress the notation to $x_t^K(M), u_t^K(M)$. Note that $x_t^K(0), u_t^K(0)$ refers to running the linear policy K.

2. $\Psi_{t,i}^{K,h}(M_{t-h:t})$ is a transfer matrix that describes the effect of w_{t-i} with respect to the past $h+1$ policies on the state x_{t+1}, formally defined below. When M is the same across all arguments we compress the notation to $\Psi_{t,i}^{K,h}(M)$.

Definition C.1. *For any $t, h \leq t, i \leq H + h$, define the disturbance-state transfer matrix $\Psi_{t,i}^{K,h}$ to be a function with $h+1$ inputs defined as*

$$\Psi_{t,i}^{K,h}(M_{t-h:t}) = \tilde{A}_K^i \mathbf{1}_{i\leq h} + \sum_{j=0}^{h} \tilde{A}_K^j B M_{t-j}^{[i-j-1]} \mathbf{1}_{i-j\in[1,H]}.$$

Definition C.2 (Surrogate State & Surrogate Action)**.** *Define,*

$$y_{t+1}^K(M_{t-H:t}) = \sum_{i=0}^{2H} \Psi_{t,i}^{K,H}(M_{t-H:t}) w_{t-i},$$

$$v_{t+1}^K(M_{t-H:t+1}) = -K y_{t+1}^K(M_{t-H:t}) + \sum_{i=1}^{H} M_{t+1}^{[i-1]} w_{t+1-i}.$$

When M is the same across all arguments we compress the notation to $y_{t+1}^K(M), v_{t+1}^K(M)$.

Definition C.3 (Surrogate Cost)**.** *Define the surrogate cost function f_t to be the cost associated with the surrogate state and surrogate action, i.e.,*

$$f_t(M_{t-H-1:t}) = \mathbb{E}\left[c_t(y_t^K(M_{t-H-1:t-1}), v_t^K(M_{t-H-1:t}))\right].$$

When M is the same across all arguments we compress the notation to $f_t(M)$.

Note that this definition coincides exactly with Definition 3.4 in the main text.

C.2 Prerequisites

In this section we state some lemmas and theorems which were proved in [4]. Due to consistency of definitions the proofs of these are omitted and can be found in [4].

Lemma C.4 (Sufficiency)**.** *For any two (κ, γ)-diagonal strongly stable matrices K^*, K, there exists $M_* = (M_*^{[0]}, \ldots, M_*^{[H-1]}) \in \mathcal{M}$ defined as*

$$M_*^{[i]} = (K - K^*)(A - BK^*)^i$$

such that

$$\sum_{t=0}^{T} \left(c_t(x_t^K(M_*), u_t^K(M_*)) - c_t(x_t^{K^*}(0), u_t^{K^*}(0))\right) \leq T \cdot \frac{2GDWH\kappa_B^2\kappa^5(1-\gamma)^H}{\gamma}.$$

Theorem C.5. *For any (κ, γ)-diagonal strongly stable K, any $\tau > 0$, and any sequence of policies $M_1 \ldots M_T$ satisfying $\|M_t^{[i]}\| \leq \tau(1-\gamma)^i$, if the perturbations are bounded by W, we have that*

$$\sum_{t=1}^{T} f_t(M_{t-H-1:t}) - \sum_{t=1}^{T} c_t(x_t^K(M_{0:t-1}), u_t^K(M_{0:t})) \leq 2TGD^2\kappa^3(1-\gamma)^{H+1},$$

where

$$D \triangleq \frac{W\kappa^3(1 + H\kappa_B\tau)}{\gamma(1 - \kappa^2(1-\gamma)^{H+1})} + \frac{\tau W}{\gamma}.$$

C.3 Proof of Lemma 5.1

Proof of Lemma 5.1. Let D be defined as

$$D \triangleq \frac{W\kappa^3(1 + H\kappa_B\tau)}{\gamma(1 - \kappa^2(1-\gamma)^{H+1})} + \frac{\kappa_B\kappa^3 W}{\gamma}.$$

Let K^* be the optimal linear policy in hindsight. By definition K^* is a (κ, γ)-diagonal strongly stable matrix. Using Lemma C.4 and Theorem C.5, we have that

$$\begin{aligned}
&\min_{M_* \in \mathcal{M}} \left(\sum_{t=0}^{T} f_t(M_*)\right) - \sum_{t=0}^{T} c_t(x_t^{K^*}(0), u_t^{K^*}(0)) \\
&\leq \min_{M_* \in \mathcal{M}} \left(\sum_{t=0}^{T} c_t(x_t^K(M_*), u_t^K(M_*))\right) - \sum_{t=0}^{T} c_t(x_t^{K^*}(0), u_t^{K^*}(0)) + 2TGD^2\kappa^3(1-\gamma)^{H+1} \\
&\leq 2TGD(1-\gamma)^{H+1}\left(\frac{WH\kappa_B^2\kappa^5}{\gamma} + D\kappa^3\right).
\end{aligned} \tag{C.1}$$

Note that by definition of $\mathcal{M}$, we have that

$$\forall t \in [T], \forall i \in [H] \quad \|M_t^{[i]}\| \leq \kappa_B\kappa^3(1-\gamma)^i.$$

Using Theorem C.5 we have that

$$\sum_{t=0}^{T} c_t(x_t^K(M_{0:t-1}), u_t^K(M_{0:t-1})) - \sum_{t=0}^{T} f_t(M_{t-H-1:t}) \leq 2TGD^2\kappa^3(1-\gamma)^{H+1}. \tag{C.2}$$

Summing up (C.1) and (C.2) and using the condition that $H \geq \frac{1}{\gamma}\log(T\kappa^2)$, we get the result. ☐

D Proof of Strong Convexity(Lemma 4.2): Multi-dimensional

Proof of Lemma 8.1. Building on Section 8, we prove Lemma 8.1 for multi-dimensional systems. Using the fact that $E[w_i w_j^\top] = 0$ for different i, j and $\mathbb{E}[w_i w_i^\top] = \Sigma$, it can be observed that for any k_1, k_2 and any $d_u \times d_u$ matrix P, we have that

$$\mathbb{E}[J_{v_{\tilde{k}_1}}^\top P J_{v_{\tilde{k}_2}}] = \mathcal{T}_{k_1-k_2} \otimes P \otimes \Sigma \tag{D.1}$$

where $\mathcal{T}_m$ is defined as an $H \times H$ matrix with $[\mathcal{T}_m]_{ij} = 1$ if and only if $i - j = m$ and 0 otherwise. This in particular immediately gives us that for any matrix P,

$$\begin{aligned}\mathbb{E}[J_y^\top P J_y] &= \left(\sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \otimes \left(\left(B^\top(A-B\mathbb{K})^\top\right)^{k_1-1} P(A-B\mathbb{K})^{k_2-1}B\right)\right) \otimes \Sigma \\ &= \left(\left(I_H \otimes B^\top\right) \underbrace{\left(\sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \otimes \left(\left((A-B\mathbb{K})^\top\right)^{k_1-1} P(A-B\mathbb{K})^{k_2-1}\right)\right)}_{\triangleq \mathbb{G}_P} (I_H \otimes B)\right) \otimes \Sigma \end{aligned} \tag{D.2}$$

Furthermore consider the following calculation

$$\mathbb{E}[J_{\tilde{v}_0}^\top \mathbb{K} J_y] = \left(\sum_{k=1}^{H} \mathcal{T}_{-k} \otimes \mathbb{K}(A-B\mathbb{K})^{k-1}B\right) \otimes \Sigma \tag{D.3}$$

$$= \left((I_H \otimes \mathbb{K}) \underbrace{\left(\sum_{k=1}^{H} \mathcal{T}_{-k} \otimes (A-B\mathbb{K})^{k-1}\right)}_{\triangleq \mathbb{Y}} (I_H \otimes B)\right) \otimes \Sigma \tag{D.4}$$

As before, we state the following bounds on the spectral properties of the matrices $\mathbb{G}$ and $\mathbb{Y}$ defined above.

Lemma D.1.

$$\|\mathbb{Y}\| \leq \|\sum_{k=1}^{H} \mathcal{T}_{-k}(A-B\mathbb{K})^{k-1}\| \leq \gamma^{-1}\kappa^2 \tag{D.5}$$

Lemma D.2. $\mathbb{G}_I$ *(where I represents the Identity matrix) is a symmetric positive definite matrix with*

$$\mathbb{G}_I \succeq \frac{1}{4\kappa^4} \cdot I_{Hd_x}$$

Consider the following calculations which follows by definitions.

$$\begin{aligned}\mathbb{E}[J_v^\top J_v] &= \mathbb{E}[J_y^\top \mathbb{K}^\top \mathbb{K} J_y] - \mathbb{E}[J_y^\top \mathbb{K}^\top J_{\tilde{v}_0}] - \mathbb{E}[J_{\tilde{v}_0}^\top \mathbb{K} J_y] + \mathbb{E}[J_{\tilde{v}_0}^\top J_{\tilde{v}_0}] \\ &= \underbrace{\left((I_H \otimes B^\top)\mathbb{G}_{\mathbb{K}^\top\mathbb{K}}(I_H \otimes B) - \mathbb{Y}(I_H \otimes B) - (I_H \otimes B^\top)\mathbb{Y}^\top + I_{Hd_u}\right)}_{\triangleq \mathbb{F}} \otimes \Sigma\end{aligned}$$

Since we know that $\Sigma \succeq 0$ we immediately get that $\mathbb{F} \succeq 0$. Using the above calculations it is enough to show that the following matrix has lower bounded eigenvalues, i.e. for every vector m of appropriate dimensions, we have that

$$m^\top \left(\mathbb{F} + (I_H \otimes B^\top)\mathbb{G}_I(I_H \otimes B)\right) m \geq \frac{\gamma^2\|m\|^2}{36\kappa^{10}}$$

To prove the above we will consider two cases. The first case is when $\|(I_H \otimes B)m\| \geq \frac{\gamma\|m\|}{3\kappa^3}$. In this case note that

$$m^\top \left(\mathbb{F} + (I_H \otimes B^\top)\mathbb{G}_I(I_H \otimes B)\right) m \geq m^\top \left((I_H \otimes B^\top)\mathbb{G}_I(I_H \otimes B)\right) m \geq \frac{\frac{1}{4\kappa^4}\gamma^2\|m\|^2}{9\kappa^6}$$

In the second case (when $\|(I_H \otimes B)m\| \leq \frac{\gamma\|m\|}{3\kappa^3}$), we have that

$$\begin{aligned} m^\top \left(\mathbb{F} + (I_H \otimes B^\top)\mathbb{G}_I(I_H \otimes B)\right) m &\geq m^\top \left(I_{Hd_u} - (I_H \otimes \mathbb{K})\mathbb{Y}(I_H \otimes B) - (I_H \otimes B^\top)\mathbb{Y}^\top(I_H \otimes \mathbb{K}^\top)\right) m \\ &\geq (1/3)\|m\|^2 \geq \frac{\gamma^2\|m\|^2}{36\kappa^{10}}. \end{aligned}$$

□

We now finish the proof with the proof of Lemmas D.1 and D.2.

Proof of Lemma D.1. Since $\mathbb{K}$ is (κ,γ)-diagonal strongly stable, we can diagonalize the matrix $A - B\mathbb{K}$ as $A - B\mathbb{K} = QLQ^{-1}$ with $\|Q\|, \|Q\|^{-1} \leq \kappa$. Therefore,

$$\mathbb{Y} = \left(\sum_{k=1}^{H} \mathcal{T}_{-k} \otimes QL^{k-1}Q^{-1}\right) = (I_H \otimes Q)\left(\sum_{k=1}^{H} \mathcal{T}_{-k} \otimes L^{k-1}\right)(I_H \otimes Q^{-1}).$$

Now consider the matrix P for any complex number ϕ with $|\phi| < 1$.

$$P = \sum_{k=1}^{H} \mathcal{T}_{-k}\phi^{k-1}$$

We wish to bound $\|P\|$. To this end consider $PP^\top$ and consider the ℓ_1 norm of any row. It can easily be seen that the ℓ_1 norm of any row of $PP^\top$ is bounded by $\frac{1}{1-|\phi|} \cdot \frac{1}{1-|\phi|^2}$, and therefore

$$\|P\| = \sqrt{\|PP^\top\|} \leq \sqrt{\frac{1}{(1-|\phi|)(1-|\phi|^2)}}.$$

Using that L is diagonal with entries bounded in magnitude by $1-\gamma$, we get that $\|\mathbb{Y}\| \leq \gamma^{-1}\kappa^2$. □

Proof of Lemma D.2. We need to consider the following matrix

$$\mathbb{G}_I = \sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \otimes \left(\left((A - B\mathbb{K})^\top\right)^{k_1-1}(A - B\mathbb{K})^{k_2-1}\right)$$

Since $\mathbb{K}$ is (κ,γ)-diagonal strongly stable, we can diagonalize the matrix $A - B\mathbb{K}$ as $A - B\mathbb{K} = QLQ^{-1}$ with $\|Q\|, \|Q\|^{-1} \leq \kappa$. Further since $A - B\mathbb{K}$ is a real valued matrix we have that $(A - B\mathbb{K})^\top = (Q^{-1})^\dagger L^\dagger Q^\dagger$. Therefore we have that

$$\mathbb{G}_I = \sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2} \otimes \left((Q^{-1})^\dagger \left(L^\dagger\right)^{k_1-1} Q^\dagger Q L^{k_2-1} Q^{-1}\right)$$

Further consider the following matrix $\hat{\mathbb{G}}$.

$$\hat{\mathbb{G}} = \begin{bmatrix} 0 & 0 & . & . & I \\ 0 & . & . & . & L \\ . & . & . & . & L^2 \\ . & 0 & . & . & . \\ 0 & I & . & . & . \\ I & L & . & . & L^{H-1} \\ L & L^2 & . & . & 0 \\ L^2 & . & . & . & 0 \\ . & . & . & . & . \\ . & L^{H-1} & . & . & . \\ L^{H-1} & 0 & . & . & 0 \end{bmatrix}$$

It can be seen that,

$$\left((I_{2H-1} \otimes Q)\hat{\mathbb{G}}(I_{2H-1} \otimes Q^{-1})\right)^{\dagger}\left((I_{2H-1} \otimes Q)\hat{\mathbb{G}}(I_{2H-1} \otimes Q^{-1})\right) = \mathbb{G}_I. \tag{D.6}$$

Furthermore note that since $\|Q\|, \|Q^{-1}\| \leq \kappa$, therefore all singular values of Q lie in the range $[\kappa^{-1}, \kappa]$. Therefore it follows that

$$Q^{\dagger}Q \succeq \kappa^{-2}I \qquad (Q^{-1})^{\dagger}Q^{-1} \succeq \kappa^{-2}I \tag{D.7}$$

Using (D.6),(D.7) it follows that

$$\mathbb{G}_I \succeq \kappa^{-4} \cdot \left(\hat{\mathbb{G}}\right)^{\dagger}\left(\hat{\mathbb{G}}\right) \tag{D.8}$$

Therefore we only need to show that $\left(\hat{\mathbb{G}}\right)^{\dagger}\left(\hat{\mathbb{G}}\right)$ has a lower bounded eigenvalue. To that end notice that since L is a diagonal matrix with diagonal values whose magnitude is upper bounded by 1. Therefore, it sufficient to consider the case when L is a scalar complex number with magnitude upper bounded by 1. To this end we can consider the following simplification of $\mathbb{G}_I$ defined for a complex number ψ with $|\psi| < 1$ as defined earlier.

$$\mathbb{G}(\psi) = \sum_{k_1=1}^{H}\sum_{k_2=1}^{H} \mathcal{T}_{k_1-k_2}\left(\psi^{\dagger}\right)^{k_1-1}\psi^{k_2-1}$$

Invoking Lemma 8.3 we immediately get that

$$\mathbb{G}_I \succeq \kappa^{-4} \cdot \left(\hat{\mathbb{G}}\right)^{\dagger}\left(\hat{\mathbb{G}}\right) \succeq \frac{1}{4\kappa^4} \cdot I_{Hd_x}.$$

□