2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Robust Online Convex Optimization for Disturbance Rejection

Joyce Lai and Peter Seiler

2024 IEEE 63rd Conference on Decision and Control (CDC) | 979-8-3503-1633-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/CDC56724.2024.10886354

***Abstract*— Online convex optimization (OCO) is a powerful tool for learning sequential data, making it ideal for high precision control applications where the disturbances are arbitrary and unknown in advance. However, the ability of OCO-based controllers to accurately learn the disturbance while maintaining closed-loop stability relies on having an accurate model of the plant. This paper studies the performance of OCO-based controllers for linear time-invariant (LTI) systems subject to disturbance and model uncertainty. The model uncertainty can cause the closed-loop to become unstable. We provide a sufficient condition for robust stability based on the small gain theorem. This condition is easily incorporated as an on-line constraint in the OCO controller. Finally, we verify via numerical simulations that imposing the robust stability condition on the OCO controller ensures closed-loop stability.**

I. Introduction

This paper considers a class of controllers recently developed using online convex optimization (OCO). Online machine learning and convex optimization methods are powerful tools for learning sequential data. This makes these techniques ideal for high precision control applications like satellite pointing and photolithography. These systems have reliable physics-based models with small error (within the control bandwidth) but are subject to unknown arbitrary disturbances.

This has motivated a large body of recent work using online learning and convex optimization for control [1]–[9]. The most closely related work is the class of OCO controllers defined in [10]. Here, OCO with memory is introduced to the discrete-time control setting as an ideal cost minimization problem (which we describe in detail in Section IV-B) to handle arbitrary disturbances and general time-varying convex cost functions. The OCO controller has promising regret guarantees and makes less restrictive assumptions about the disturbance characteristics (e.g., white noise or worst-case) than that of H_2 and H_∞ optimal control techniques [11], [12]. This makes OCO methods well suited for high precision control applications with unknown, arbitrary disturbances that degrade the system performance.

The OCO framework in [10] aims to learn the disturbance characteristics in real time. However, small model errors can cause instability and thus must be explicitly considered in the design. There are additional works that attempt to learn the model from data [13]–[19]. However, dynamic uncertainties in many high precision applications arise due to high frequency, time-varying, and/or nonlinear effects. It is difficult to learn such unmodeled effects from real-time data. In these cases, it is useful to design a robust OCO-based controller that can learn the disturbance features and tolerate model uncertainty, thus motivating our work.

There are three main contributions of our work. First, we provide a robust stability condition for OCO control of a discrete linear time-invariant (LTI) plant (Theorem 1 in Section III-B) and show how it can be used to compute the stability bound (Section V). The robust stability condition is a scaled version of the small gain condition which holds for an arbitrary induced system norm. Second, we show how the robust stability condition can be imposed as a pointwise in time constraint on the OCO controller to ensure robustness to nonparametric uncertainties. This implementation of the robust stability condition is specific to using the induced ℓ_∞-norm (Section III-C and Section IV-C), resulting in an easy extension of [10]. Lastly, we present numerical results that illustrate the effect of the robust stability constraint on the OCO controller (Section V).

II. Problem Formulation

This section formulates the OCO control problem for discrete-time LTI plants subject to both model uncertainty and unknown disturbances.

A. Notation

Let $v \in \mathbb{R}^n$ be a vector. The p-norm of this vector is defined as $\|v\|_p := \left[\sum_{i=1}^n |v_i|^p\right]^{\frac{1}{p}}$. Next, $\mathbb{N}$ denotes the set of non-negative integers. Let $d : \mathbb{N} \to \mathbb{R}^n$ denote a vector-valued sequence $\{d_0, d_1, \ldots\}$. The ℓ_p-norm of d is defined as:

$$\|d\|_p = \left[\sum_{t=0}^{\infty} \|d_t\|_p^p\right]^{\frac{1}{p}}. \tag{1}$$

Note that $\|d_t\|_p$ is the p-norm of the vector $d_t \in \mathbb{R}^n$ at time t while $\|d\|_p$ is the ℓ_p-norm of the sequence. The set ℓ_p consists of sequences that have finite ℓ_p-norm. The subset $\ell_{pe} \subset \ell_p$ is the extended space of sequences that have finite ℓ_p-norm on all finite intervals, i.e. $\sum_{t=0}^T \|d_t\|_p^p < \infty$ for all $T \geq 0$. Finally, let $G : \ell_{pe} \to \ell_{pe}$ denote systems that map an input signal $u \in \ell_{pe}$ to an output signal $y \in \ell_{pe}$. The induced ℓ_p-norm for this system is defined as:

$$\|G\|_{p\to p} = \sup_{0\neq u\in\ell_p} \frac{\|y\|_p}{\|u\|_p}. \tag{2}$$

To simplify notation, we'll often use $\|d\|$ and $\|G\|$ for the signal norm and system induced norm when the specific p-norm is not important.

J. Lai and P. Seiler are with the Department of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor, MI 48109, USA. {joycelai,pseiler}@umich.edu

B. Model Uncertainty

In this section, we consider the feedback system in Figure 1 and discuss the model uncertainty $\Delta(z)$ in more detail.

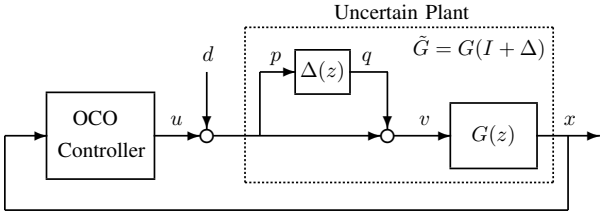

Fig. 1. Discrete-time feedback system with unknown disturbance d and uncertainty $\Delta(z)$. OCO control is used to reject the disturbance d without knowledge of the uncertainty $\Delta(z)$.

Consider the nominal discrete-time, LTI plant $G(z)$ with dynamics:

$$x_{t+1} = A\,x_t + B\,v_t, \tag{3}$$

where $x_t \in \mathbb{R}^{n_x}$ and $v_t \in \mathbb{R}^{n_u}$ are the nominal plant state and input at time t, respectively. We assume $x_0 = 0$ for simplicity.

Model uncertainty for systems with physics-based models often shows up as unmodeled actuator dynamics affecting the plant input [11], [12], [20]. We can account for these unmodeled dynamics by defining an input-multiplicative uncertainty set $\mathcal{G}_\delta$ as:

$$\mathcal{G}_\delta = \left\{ \tilde{G}(z) = G(z)\big(I + \Delta(z)\big) : \|\Delta\| \leq \delta \right\}, \tag{4}$$

where $\delta \in [0, \infty)$. Note that the induced 2-norm is common choice to bound the uncertainty. However, our main result in Section III holds for any induced p-norm.

Let $\tilde{G}_0(z)$ denote the true plant dynamics. We assume that the true plant is within the uncertainty set, i.e. $\tilde{G}_0(z) \in \mathcal{G}_\delta$. In other words, there exists a specific $\Delta_0(z)$ such that $\|\Delta_0\| \leq \delta$ and $\tilde{G}_0(z) = G(z)(I + \Delta_0(z)) \in \mathcal{G}_\delta$. More generally, we refer to $\tilde{G}(z) = G(z)(I+\Delta(z))$ as the uncertain plant. Here, we assume the uncertainty $\Delta(z)$ is LTI. However, our main result in Section III can be extended to the case where Δ is a possibly nonlinear time-varying (NLTV) system.

C. OCO Control

Unknown disturbances are often caused by environmental factors and moving physical components which degrade system performance. However, these disturbances often also have learnable characteristics. It is typical to model such disturbances as entering at the plant input as shown in Figures 1 and 2.

OCO control can be used to learn and reject the disturbance without a priori knowledge of the disturbance [1]–[9]. Here, we describe a class of OCO controllers closely related to [10] which considers the case without uncertainty $\Delta(z) = 0$. The OCO controller has the block diagram representation shown in Figure 2. This corresponds to the class of disturbance action controllers defined as:

$$u_t = -Kx_t + \sum_{i=0}^{H-1} M_t^{[i]} \hat{w}_{t-i}, \tag{5}$$

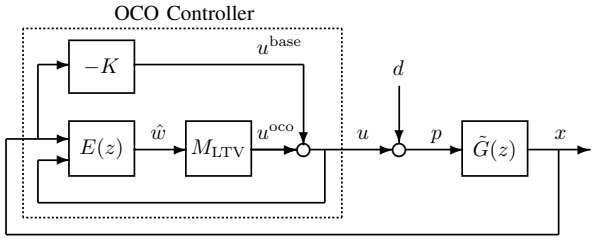

Fig. 2. Block diagram representation of the OCO controller in a discrete-time feedback system with unknown disturbance d_t and uncertain plant $\tilde{G}(z)$. The OCO controller is composed of a state-feedback gain K, an estimator $E(z)$, and an LTV system M_{LTV}.

where $K \in \mathbb{R}^{n_u \times n_x}$, $M_t^{[i]} \in \mathbb{R}^{n_u \times n_x}$, and $\hat{w}_t \in \mathbb{R}^{n_x}$ are the state-feedback gain, learned coefficients, and disturbance estimate, at time t, respectively. The state-feedback gain K is user-selected while the learned coefficients $\{M_t\}_{i=0}^{H-1}$ are typically updated via some online optimization method. For example, [10] uses online projected gradient descent (OPGD) with memory (see Section IV-B).

The online optimization often uses an estimate of the disturbance to learn the coefficients. When there is no uncertainty $\Delta(z) = 0$, the disturbance estimate $\hat{w}_t$ can be perfectly estimated from x_t and u_t using the nominal plant dynamics [10]. We discuss this in more detail in Section IV-A. Thus, we assume the disturbance estimate $\hat{w}_t$ to be the output of an LTI estimator $E(z)$ of the following form:

$$\begin{aligned} x_{t+1}^e &= A_e x_t^e + B_{e1} x_t + B_{e2} u_t \\ \hat{w}_t &= C_e x_t^e + D_{e1} x_t + D_{e2} u_t, \end{aligned} \tag{6}$$

where $x_t^e \in \mathbb{R}^{n_e}$ and $\hat{w}_t \in \mathbb{R}^{n_x}$ are the estimator state and output at time t, respectively. Note that x_t and u_t are inputs to the estimator.

The first term in (5) is considered the baseline controller which we denote by:

$$u_t^{\text{base}} = -Kx_t. \tag{7}$$

The main results in Section III can be generalized to the case when the baseline control u_t^{base} is the output of an LTI controller $K(z)$ with input x_t. We assume the baseline controller is a static, state-feedback gain for simplicity.

The second term in (5) is the output of an finite impulse response (FIR) filter with time-varying coefficients. We denote the FIR filter as a linear time-varying (LTV) system M_{LTV} with input-output dynamics defined as:

$$u_t^{\text{oco}} = \sum_{i=0}^{H-1} M_t^{[i]} \hat{w}_{t-i}. \tag{8}$$

where $\hat{w}_t \in \mathbb{R}^{n_x}$ and $u_t^{\text{oco}} \in \mathbb{R}^{n_u}$ are the input and output at time t, respectively. The FIR filter order H is also referred to as the learning horizon since the coefficients are often updated via OCO using the past H disturbance estimates. We provide an example of online optimization in Sections IV and V, but the main results in Section III assume only that the coefficients are time-varying.

The OCO controller (5) can be interpreted as a baseline controller u_t^{base} plus an adapting term u_t^{oco} which corrects for the unknown disturbance d_t based on disturbance estimates.

D. Model Uncertainty Effects on OCO Control

The uncertainty $\Delta(z)$ and disturbance d_t have different effects on closed-loop stability. Suppose the state-feedback gain K is stabilizing, i.e., all eigenvalues of $(A - BK)$ are strictly inside the unit disk. Without uncertainty $\Delta(z) = 0$, OCO control can be designed to achieve disturbance rejection with provable guarantees [10]. In this case, a bounded disturbance d cannot cause signals $x, u, \hat{w}$, etc. to grow unbounded. However, small amounts of model uncertainty can cause the system to become unstable.

As shown in Figures 1 and 2, the uncertain plant input is the control input perturbed by an unknown disturbance:

$$p_t = u_t + d_t, \tag{9}$$

where $u_t, d_t, p_t \in \mathbb{R}^{n_u}$ are the control input, disturbance, and perturbed uncertain plant input at time t, respectively. The perturbed input p_t is further distorted by the uncertainty $\Delta(z)$. The resulting input to the nominal plant $G(z)$ is:

$$v_t = (I + \Delta)\, p_t = u_t + d_t + q_t, \tag{10}$$

where $q_t = \Delta p_t \in \mathbb{R}^{n_u}$. Again, v_t is the nominal plant input at time t. Not only is there an unknown disturbance d_t, but also a distorted signal q_t due to the uncertainty $\Delta(z)$.

The additional perturbation q_t can lead to unexpected behaviors that affect the disturbance estimate and FIR filter coefficient update when left unaccounted for in the OCO design. This can occur even when the state-feedback gain K is stabilizing for the uncertain plant $\tilde{G}(z)$. Thus, the OCO controller is required to: i) learn and compensate for the disturbance, and ii) stabilize the system in the presence of uncertainty. The OCO controller must achieve these objectives without a priori knowledge of the disturbance or uncertainty.

III. Main Result

This section provides a condition on M_{LTV} that ensures the feedback system with OCO control remains stable even in the presence of the model uncertainty.

A. Linear Fractional Transformation

As a first step, we transform the feedback system of the uncertain plant and OCO controller (Figures 1 and 2) to a standard form as shown in Figure 3. This diagram separates the LTI dynamics P from the uncertainty Δ and time-varying OCO dynamics M_{LTV}. Here P includes the dynamics due to the plant, estimator, and state-feedback gain. This diagram is called a linear fractional transformation (LFT) in the robust control literature [11], [12]. We use the notation $F_U(P, \Gamma)$ for this interconnection with $\Gamma = \left[\begin{smallmatrix} \Delta & 0 \\ 0 & M_{\text{LTV}} \end{smallmatrix}\right]$ closed around the upper channels of P.

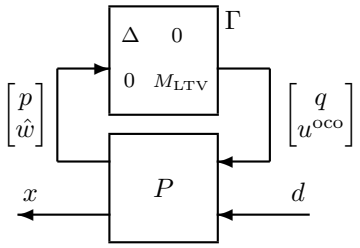

Fig. 3. Equivalent LFT $F_U(P, \Gamma)$ of original system separating LTI dynamics P from uncertainty Δ and time-varying learning dynamics M_{LTV}.

An explicit state-space model for P can be determined from the various components of the feedback system described in Section II. The dynamics of P are given by:

$$\begin{bmatrix} x_{t+1} \\ x_{t+1}^e \end{bmatrix} = \begin{bmatrix} A - BK & 0 \\ B_{e1} - B_{e2}K & A_e \end{bmatrix} \begin{bmatrix} x_t \\ x_t^e \end{bmatrix} + \begin{bmatrix} B & B & B \\ 0 & B_{e2} & 0 \end{bmatrix} \begin{bmatrix} q_t \\ u_t^{\text{oco}} \\ d_t \end{bmatrix}$$

$$\begin{bmatrix} p_t \\ \hat{w}_t \\ x_t \end{bmatrix} = \begin{bmatrix} -K & 0 \\ D_{e1} - D_{e2}K & C_e \\ I & 0 \end{bmatrix} \begin{bmatrix} x_t \\ x_t^e \end{bmatrix} + \begin{bmatrix} 0 & I & I \\ 0 & D_{e2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} q_t \\ u_t^{\text{oco}} \\ d_t \end{bmatrix}.$$

Next, we use the LFT representation $F_U(P, \Gamma)$ to formulate and state our robust stability condition.

B. Scaled Small Gain Theorem

Our first stability result is a variation of the standard small gain theorem (see Section 5.4 of [21]). This provides a sufficient condition for the dynamics $F_U(P, \Gamma)$ to have a bounded gain from disturbance d to state x. Note stability here is in the sense of bounded gain in some induced norm.

Lemma 1. *Consider the interconnection $F_U(P, \Gamma)$ where $P : \ell_{pe} \to \ell_{pe}$ and $\Gamma : \ell_{pe} \to \ell_{pe}$ are linear systems with finite induced ℓ_p-norm. Partition P as:*

$$\begin{bmatrix} \bar{p} \\ x \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{21} \end{bmatrix} \begin{bmatrix} \bar{q} \\ d \end{bmatrix}, \tag{11}$$

where $\bar{p} := \left[\begin{smallmatrix} p \\ \hat{w} \end{smallmatrix}\right]$ and $\bar{q} := \left[\begin{smallmatrix} q \\ u^{\text{oco}} \end{smallmatrix}\right]$ are the inputs and outputs of Γ. The interconnection has finite induced ℓ_p-norm, i.e. $\|F_U(P, \Gamma)\| < \infty$, if $\|P_{11}\|\,\|\Gamma\| < 1$.

Proof: The system P is LTI so by the principle of superposition (assuming zero initial conditions):

$$\bar{p} = P_{11}\bar{q} + P_{12}d. \tag{12}$$

We can bound $\bar{p}$ using the triangle inequality and the definition of the induced norm:

$$\|\bar{p}\| \le \|P_{11}\|\,\|\bar{q}\| + \|P_{12}\|\,\|d\|. \tag{13}$$

Next, $\bar{q} = \Gamma\bar{p}$ so that $\|\bar{q}\| \le \|\Gamma\|\,\|\bar{p}\|$. Substitute this bound into (13) and re-arrange to obtain:

$$\|\bar{p}\| \le \frac{\|P_{12}\|}{1 - \|P_{11}\|\|\Gamma\|}\,\|d\|. \tag{14}$$

This last step requires the small gain condition $\|P_{11}\|\,\|\Gamma\| < 1$ to obtain the bound on $\|\bar{p}\|$.

Finally, the state is $x = P_{21}\bar{q} + P_{22}d$. We can use similar steps and the bound on $\bar{p}$ to obtain:

$$\|x\| \leq \left[\|P_{22}\| + \frac{\|P_{21}\|\,\|P_{12}\|\,\|\Gamma\|}{1 - \|P_{11}\|\,\|\Gamma\|}\right]\|d\|. \tag{15}$$

Hence, $F_U(P,\Gamma)$ has finite induced ℓ_p-norm. □

The small gain condition in the previous lemma can be conservative as it does not exploit the block structure $\Gamma = \left[\begin{smallmatrix}\Delta & 0\\ 0 & M_{\mathrm{LTV}}\end{smallmatrix}\right]$. We can reduce the conservatism by normalizing the blocks and introducing scalings. Specifically, assume $\|\Delta\| \leq \delta$ and $\|M_{\mathrm{LTV}}\| \leq \beta$. Define the normalized uncertainty and learning dynamics as: $\tilde{\Delta} = \frac{1}{\delta}\Delta$ and $\tilde{M}_{\mathrm{LTV}} = \frac{1}{\beta}M_{\mathrm{LTV}}$. Stacking these together yields

$$\tilde{\Gamma} := \begin{bmatrix}\frac{1}{\delta}I & 0\\ 0 & \frac{1}{\beta}I\end{bmatrix}\Gamma = \begin{bmatrix}\frac{1}{\delta}\Delta & 0\\ 0 & \frac{1}{\beta}M_{\mathrm{LTV}}\end{bmatrix}. \tag{16}$$

The scaling normalizes each block so that $\|\tilde{\Gamma}\| \leq 1$.

Next, the uncertainty is LTI and hence $d_1\Delta = \Delta d_1$ for any scalar $d_1 > 0$. (In fact, this relation holds even if d_1 is also an LTI system but we do not pursue this generalization here.) Similarly, the learning dynamics are also linear and hence $d_2 M_{\mathrm{LTV}} = M_{\mathrm{LTV}} d_2$ for any scalar $d_2 > 0$. It follows that the normalized systems can be equivalently written, for any $d_1, d_2 > 0$, as:

$$\tilde{\Gamma} := \begin{bmatrix}\frac{1}{d_1\delta}I & 0\\ 0 & \frac{1}{d_2\beta}I\end{bmatrix}\Gamma\begin{bmatrix}d_1 & 0\\ 0 & d_2\end{bmatrix}. \tag{17}$$

This leads to the following scaled small gain result.

Theorem 1. *Consider the interconnection $F_U(P,\Gamma)$ where $P : \ell_{pe} \to \ell_{pe}$ and $\Gamma : \ell_{pe} \to \ell_{pe}$ are linear systems with finite induced ℓ_p-norm. Assume $\Gamma := \left[\begin{smallmatrix}\Delta & 0\\ 0 & M_{\mathrm{LTV}}\end{smallmatrix}\right]$ where $\|\Delta\| \leq \delta$ and $\|M_{\mathrm{LTV}}\| \leq \beta$. Partition P as:*

$$\begin{bmatrix}\bar{p}\\ x\end{bmatrix} = \begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{21}\end{bmatrix}\begin{bmatrix}\bar{q}\\ d\end{bmatrix}, \tag{18}$$

where $\bar{p} := \left[\begin{smallmatrix}p\\ \hat{w}\end{smallmatrix}\right]$ and $\bar{q} := \left[\begin{smallmatrix}q\\ u^{\mathrm{oco}}\end{smallmatrix}\right]$ are the inputs and outputs of Γ. The interconnection has finite induced ℓ_p-norm, i.e., $\|F_U(P,\Gamma)\| < \infty$, if there exists scalars $d_1, d_2 > 0$ such that

$$\tilde{P}_{11} := \begin{bmatrix}\frac{1}{d_1}I & 0\\ 0 & \frac{1}{d_2}I\end{bmatrix}P_{11}\begin{bmatrix}d_1\delta\, I & 0\\ 0 & d_2\beta\, I\end{bmatrix} \tag{19}$$

satisfies $\|\tilde{P}_{11}\| < 1$.

Proof: Define a scaled version of the nominal dynamics P as:

$$\tilde{P} = \left[\begin{array}{cc|c}\frac{1}{d_1}I & 0 & 0\\ 0 & \frac{1}{d_2}I & 0\\ \hline 0 & 0 & I\end{array}\right]\begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{bmatrix}\left[\begin{array}{cc|c}d_1\delta\, I & 0 & 0\\ 0 & d_2\beta\, I & 0\\ \hline 0 & 0 & I\end{array}\right].$$

The constants introduced in the scaled plant $\tilde{P}$ cancel those introduced for $\tilde{\Gamma}$ in (16). In other words, $F_U(P,\Gamma)$ and $F_U(\tilde{P},\tilde{\Gamma})$ define the same dynamics from d to x. Moreover, $\|\tilde{P}_{11}\| < 1$ and $\|\tilde{\Gamma}\| \leq 1$ by assumption. It follows from the small gain result (Lemma 1) that $F_U(\tilde{P},\tilde{\Gamma}) = F_U(P,\Gamma)$ has finite induced ℓ_p-norm. □

The scalings d_1 and d_2 in the robust stability condition (Theorem 1) can be used to reduce the conservatism of the small gain condition (Lemma 1). They are known as D-scales in the robust control literature (see [22] and Chapter 11 in [11]) and are used in structured singular value robust stability tests. Note that without loss of generality, we can set $d_2 = 1$ and express (19) in terms of only d_1. This will be useful in Section V when we use Theorem 1 to compute the stability bound.

C. Bounding the LTV Dynamics

In this section, we provide a result specific to the induced ℓ_∞-norm for the OCO control implementation. The induced ℓ_∞-norm is useful as it allows us to relate $\|M_{\mathrm{LTV}}\|_{\infty\to\infty}$ to $\|M_t\|_{\infty\to\infty}$. The robust stability constraint can then be imposed as a point-wise in time constraint β on the coefficients $\|M_t\|_{\infty\to\infty} \leq \beta$ in the projection step of OPGD. We discuss this further in Section IV-B and IV-C.

The dynamics M_{LTV} in (8) can be expressed as:

$$u_t^{\mathrm{oco}} = M_t\hat{W}_t, \tag{20}$$

where

$$M_t := \begin{bmatrix}M_t^{[0]} & \cdots & M_t^{[H-1]}\end{bmatrix} \in \mathbb{R}^{n_u \times n_x H}, \text{ and} \tag{21}$$

$$\hat{W}_t := \begin{bmatrix}\hat{w}_t\\ \vdots\\ \hat{w}_{t-H+1}\end{bmatrix} \in \mathbb{R}^{n_x H} \tag{22}$$

are the stacked FIR coefficients and estimated disturbance history. The following lemma relates the induced ℓ_∞-norm of the system M_{LTV} to the matrix induced ∞-norm of M_t. The proof is based on standard norm properties but is included for completeness.

Lemma 2. *Let M_{LTV} be the LTV system defined in* (20) *and M_t be the stacked gains defined in* (21)*. Then*

$$\|M_{\mathrm{LTV}}\|_{\infty\to\infty} = \sup_t \|M_t\|_{\infty\to\infty}. \tag{23}$$

Proof: The equality in (23) is shown in two steps: (A) $\|M_{\mathrm{LTV}}\|_{\infty\to\infty} \leq \sup_t \|M_t\|_{\infty\to\infty}$ and (B) $\|M_{\mathrm{LTV}}\|_{\infty\to\infty} \geq \sup_t \|M_t\|_{\infty\to\infty}$.

First, we show direction (A). Let $\hat{w}$ and u^{oco} be any input-output pair of M_{LTV}. By definition of the induced matrix norm and equation (20), we can show that

$$\begin{aligned}\|u^{\mathrm{oco}}\|_\infty &= \sup_t \|M_t\hat{W}_t\|_\infty\\ &\leq \sup_t \|M_t\|_{\infty\to\infty}\cdot\|\hat{w}\|_\infty.\end{aligned} \tag{24}$$

It follows that $\frac{\|u^{\mathrm{oco}}\|_\infty}{\|\hat{w}\|_\infty} \leq \sup_t \|M_t\|_{\infty\to\infty}$, and thus we have that $\|M_{\mathrm{LTV}}\|_{\infty\to\infty} \leq \sup_t \|M_t\|_{\infty\to\infty}$ by definition of the induced ℓ_∞-norm. Hence, claim (A) holds.

Next, we show direction (B). Suppose $\sup_t \|M_t\|_{\infty\to\infty}$ achieves its maximum at some finite time t_0. (The proof can be modified if the supremum occurs as $t \to \infty$.) Then there exists a vector $\hat{W}^*$ such that $\|\hat{W}^*\| = 1$ and

$$\|M_{t_0}\hat{W}^*\|_\infty = \sup_t \|M_t\|_{\infty\to\infty}.$$

We can use the vector $\hat{W}^*$ to construct a signal $\hat{w}^*$ for which: (a) $\|\hat{w}^*\|_\infty = 1$, and (b) the corresponding $\hat{W}^*$ constructed from $\hat{w}^*$ satisfies $\hat{W}^*_{t_0} = \hat{W}^*$. Since $\|u^{\text{oco}}\|_\infty \geq \|u^{\text{oco}}_{t_0}\|_\infty = \|M_{t_0}\hat{W}^*_{t_0}\|_\infty$, we have that

$$\|u^{\text{oco}}\|_\infty \geq \sup_t \|M_t\|_{\infty\to\infty} \cdot \|\hat{w}^*\|_\infty.$$

It follows that $\frac{\|u^{\text{oco}}\|_\infty}{\|\hat{w}^*\|_\infty} \geq \sup_t \|M_t\|_{\infty\to\infty}$, and thus we have that $\|M_{\text{LTV}}\|_{\infty\to\infty} \geq \sup_t \|M_t\|_{\infty\to\infty}$ by definition of the induced ℓ_∞-norm. Hence, claim (B) holds. □

IV. Application to OCO

In this section, we demonstrate how the main results can be applied to ensure robust stability of existing OCO controllers. We focus on the OCO controllers in [7], [10] where the coefficients of M_{LTV} are updated via OPGD.

A. Estimator Design

The class of OCO controllers defined by [10] considers the feedback system with OCO control (Figure 2) and no uncertainty (Figure 1) when $\Delta(z) = 0$. In this case, a perfect plant model is assumed $\tilde{G}(z) = G(z)$. Thus, the nominal plant dynamics can be used to design an estimator $E(z)$ and OPGD to update the coefficients in M_{LTV}. Later, we will show how the OPGD projection step can be modified to ensure robust stability for the case that there is uncertainty $\Delta(z) \neq 0$.

Without uncertainty, the plant dynamics with unknown disturbance reduce to:

$$x_{t+1} = Ax_t + Bu_t + Bd_t.$$

Note that Bd_t is the effective disturbance on the state at time t. Assuming the state x_t is measurable, we can perfectly reconstruct this effective disturbance at the previous time step. Use the measured state and rearranging the plant dynamics:

$$\hat{w}_t = x_t - Ax_{t-1} - Bu_{t-1}. \tag{25}$$

With no uncertainty, this estimator perfectly reconstructs the effective disturbance with a one-step delay: $\hat{w}_t = Bd_{t-1}$. However, perfect reconstruction is no longer guaranteed with uncertainty, i.e. if $\Delta(z) \neq 0$ then $\hat{w}_t \neq Bd_{t-1}$. In this case, $\hat{w}_t$ is considered an estimate of Bd_{t-1}.

The disturbance reconstruction (25) can be expressed in state-space form as:

$$\begin{aligned} x^e_{t+1} &= 0\, x^e_t - A\, x_t - Bu_t \\ \hat{w}_t &= x^e_t + x_t, \end{aligned}$$

where $x^e_t = -Ax_{t-1} - Bu_{t-1}$ is the estimator state. This has the form of the general LTI estimator $E(z)$ in (6). The estimates $\hat{w}_t$ of past disturbances are used to update the FIR coefficients M_t defined in (21) by minimizing an "ideal" cost which we describe next.

B. OPGD on an Ideal Cost

The coefficients M_t are updated at each time step via OPGD in the direction of an "ideal" (per-step) cost. This cost is associated with the nominal plant dynamics (3) and a per-step cost function. Here, we consider quadratic per-step costs:

$$c(x_t, u_t) = x_t^\top Q\, x_t + u_t^\top R\, u_t, \tag{26}$$

where $Q = Q^\top \succeq 0 \in \mathbb{R}^{n_x \times n_x}$ and $R = R^\top \succ 0 \in \mathbb{R}^{n_u \times n_u}$. Note that the finite-horizon cost is defined as:

$$J_T(x, u) = \sum_{t=0}^{T} c(x_t, u_t), \tag{27}$$

where T is the total time horizon. The ideal cost $g(M)$ is defined for any static gain $M \subset \mathbb{R}^{n_u \times n_x H}$ based on this per-step cost (26) which is computed and defined as follows.

Let $\tilde{x}_\tau \in \mathbb{R}^{n_x}$ and $\tilde{u}_\tau \in \mathbb{R}^{n_u}$ denote the ideal state and control input at time τ, respectively. The ideal state and input are initialized at $\tau = t - H$ by:

$$\tilde{x}_{t-H} = 0 \text{ and } \tilde{u}_{t-H} = \sum_{i=0}^{H-1} M^{[i-1]}\, w_{t-H-i}. \tag{28}$$

where t is the current time. The ideal state and control input are then computed for $\tau = t-H+1, \ldots, t$ by iterating over the plant dynamics with the static gains M:

$$\tilde{x}_\tau = A\, \tilde{x}_{\tau-1} + B\, \tilde{u}_{\tau-1} + \hat{w}_{\tau-1} \tag{29}$$

$$\tilde{u}_\tau = -K\, \tilde{x}_\tau + \sum_{i=0}^{H-1} M^{[i]}\, \hat{w}_{\tau-i}. \tag{30}$$

The ideal cost is then defined as $g(M) := c(\tilde{x}_t, \tilde{u}_t)$. In other words, the ideal cost $g(M)$ is the cost of the plant dynamics evolving with static gain M over the learning horizon H, neglecting dynamics beyond time $t - H$. The coefficients M_t are updated via OPGD on this ideal cost:

$$M_{t+1} = \Pi_{\mathcal{M}}\left(M_t - \eta \nabla_M g(M_t)\right), \tag{31}$$

where η is the learning rate, and $\Pi_{\mathcal{M}}$ is the projection of the gradient step of M_t onto a constraint set $\mathcal{M}$. Additional details are given in [7], [10]. Next, we show how the constraint set $\mathcal{M}$ can be modified to ensure the robust stability of the OCO feedback system (Figures 1 and 2) when $\Delta(z) \neq 0$.

C. Robust OCO Control

Theorem 1 states that the closed-loop system (Figures 1 and 2) will be stable for bounded uncertainties $\|\Delta\|_{\infty\to\infty} \leq \delta$ and bounded LTV learning dynamics $\|M_{\text{LTV}}\|_{\infty\to\infty} \leq \beta$ if the robust stability condition $\|\tilde{P}_{11}\|_{\infty\to\infty} < 1$ holds. Larger values of β risk stability, yet can improve disturbance rejection as they allow the OCO more freedom to adapt the coefficients M_t. Thus, it is important to determine the largest possible value of β such that the robust stability condition

holds. We refer to this value of β as the robust stability bound β^* which is the solution to:

$$\begin{aligned} \sup_{\beta} \quad & \beta \\ \text{subject to} \quad & \|\tilde{P}_{11}\|_{\infty\to\infty} < 1, \end{aligned} \tag{32}$$

where $\tilde{P}_{11}$ is defined in (19). Once the stability bound β^* is known, we can select the constraint β to be within $[0, \beta^*)$ to ensure stability. Selecting β as close as possible to β^* gives the full benefit of OCO.

Next, it follows from Lemma 2 that if $\|M_t\|_{\infty\to\infty} \leq \beta$, then $\|M_{\text{LTV}}\|_{\infty\to\infty} \leq \beta$ (and vice versa). This means we can simply impose a constraint on the coefficients M_t as they are updated via OPGD by defining the constraint set $\mathcal{M}$ as:

$$\mathcal{M} := \Big\{ M \in \mathbb{R}^{n_u \times n_x H} : \|M\|_{\infty\to\infty} \leq \beta \Big\}. \tag{33}$$

However, instead of computing the exact projection $\Pi_{\mathcal{M}}$ in (31), we can scale down the coefficients until they are within the constraint set $\mathcal{M}$ for simpler online implementation. The projection step (see Algorithm 1 from [10]) can easily be modified to update the coefficients as:

$$M_{t+1} = \begin{cases} M_{\text{step}}, & \|M_{\text{step}}\|_{\infty\to\infty} \leq \beta \\ \beta \left(\frac{M_{\text{step}}}{\|M_{\text{step}}\|_{\infty\to\infty}} \right), & \|M_{\text{step}}\|_{\infty\to\infty} > \beta, \end{cases} \tag{34}$$

where $M_{\text{step}} := M_t - \eta \nabla_M g(M_t)$ is the gradient step of the coefficients M_t at time t. The results in the following section are based on the coefficient update described in (34).

V. Numerical Results

In this section, we perform numerical studies to illustrate the effect of β and explicitly use the robust stability condition (Theorem 1) to compute the stability bound β^*. Note that we recover state-feedback when $\beta = 0$ and unconstrained OCO (U-OCO) when $\beta = \infty$. We refer to the case when $0 < \beta < \infty$ as constrained OCO (C-OCO).

Our example considers the following nominal plant model $G(z)$ and uncertainty $\Delta(z)$:

$$G(z) = \frac{0.1}{z - 0.9} \text{ and } \Delta(z) = \frac{-z^2 + 1.79z - 0.7903}{z^2 - 1.672z + 0.9048},$$

where $\Delta(z)$ is small in magnitude at low frequencies and larger at higher frequencies. For the baseline controller, we use a state-feedback gain of $K = 0.15$. Note that $K = 0.15$ is stabilizing for both the nominal and uncertain plant. For the OCO, we use the quadratic per-step cost $c(x_t, u_t)$ in (26) with $Q = 1$ and $R = 10^{-1}$, learning rate $\eta = 5 \times 10^{-4}$, and learning horizon $H = 1$. Lastly, we use a step disturbance: $d_t = 100$ for $0 \leq t \leq 500$ and $d_t = -100$ for $500 < t \leq T$. All simulations were run with a time horizon of $T = 1000$.

Figure 4 shows the per-step cost $c(x_t, d_t)$ and estimated disturbance $\hat{w}_t$ of U-OCO at each time t. We compare the performance of the nominal (red dashed) and uncertain (blue solid) plant. The disturbance is perfectly reconstructed $\hat{w}_t = Bd_{t-1}$(see Section IV-A) with the nominal plant. However, with the uncertain plant, this is not the case $\hat{w}_t \neq Bd_{t-1}$. The ideal cost $g(M)$ computation assumes the nominal plant

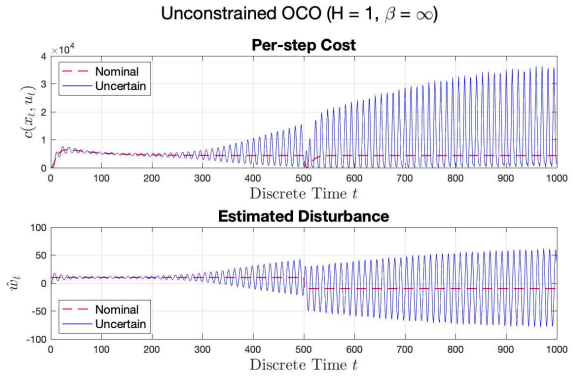

Fig. 4. Per-step cost (top) and disturbance estimate (bottom) of running U-OCO on the nominal (red dashed) and uncertain (blue solid) plant. U-OCO is stable for the nominal plant and unstable for the uncertain plant.

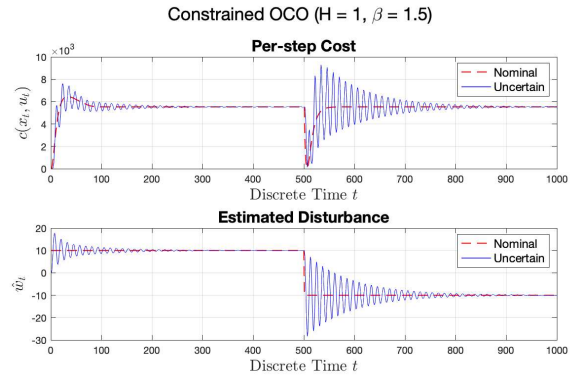

Fig. 5. Per-step cost (top) and disturbance estimate (bottom) of running C-OCO with $\beta = 1.5$ on the nominal (red dashed) and uncertain (blue solid) plant. C-OCO is stable for the nominal and uncertain plants.

model and perfect disturbance estimation. This mismatch introduces an error in the coefficient update M_{t+1} which causes instability. The instability is illustrated by the per-step cost and estimated disturbance growing unbounded. On the other hand, U-OCO performance is stable without uncertainty because the disturbance is estimated perfectly. Thus, the constraint β is needed for the coefficient update to ensure stability for the uncertain plant.

Figure 5 shows the per-step cost $c(x_t, d_t)$ and estimated disturbance $\hat{w}_t$ of C-OCO for $\beta = 1.5$ at each time t. Again, we compare the performance of the nominal (red dashed) and uncertain (blue solid) plant. As mentioned before, an error in the disturbance estimate introduces an error in the ideal cost gradient. The ideal cost gradient error can cause the gradient step $M_{step} = M_t - \nabla_M g(M_t)$ to grow too large in the wrong direction. When the constraint β is chosen such that the robust stability condition (Theorem 1) is satisfied, the effect of uncertainty induced error on the gradient step of the coefficient update is limited. This is illustrated in Figure 5 as the performance of C-OCO on the uncertain plant eventually recovers the performance on the nominal model with $\beta = 1.5$. Thus, imposing the constraint β can ensure that OCO is robust to uncertainty.

As mentioned in Section IV-C, the choice of β is critical. In order to compute the stability bound β^*, we first make an additional assumption about the uncertainty. We assume the uncertainty $\Delta(z)$ can be decomposed into a normalized uncertainty $\tilde{\Delta}(z)$ and a frequency-dependent uncertainty weight $W_u(z)$ in the form:

$$\Delta(z) = \tilde{\Delta}(z) W_u(z), \tag{35}$$

where $\|\tilde{\Delta}\| \leq 1$ and $W_u(z)$ is stable. Roughly, $W_u(z)$ represents the amount of uncertainty present at each frequency. The set of uncertain plants $\mathcal{G}_\delta$ can then be expressed as:

$$\mathcal{G}_\delta = \left\{ \tilde{G}(z) = G(z)\big(I + \tilde{\Delta}(z) W_u(z)\big) : \|\tilde{\Delta}\| \leq 1 \right\}. \tag{36}$$

We use the following uncertainty weight $W_u(z)$:

$$W_u(z) = \frac{-z^2 + 1.79z - 0.7903}{z^2 - 1.672z + 0.9048}.$$

$W_u(z)$ is small in magnitude at low frequencies and larger at higher frequencies. Note that we now consider the uncertainty $\Delta(z)$ from the previous example as our uncertainty weight $W_u(z)$. The previous example is thus a special case where the normalized uncertainty is a static gain $\tilde{\Delta}(z) = 1$.

Given the uncertainty weight W_u, we compute the stability bound β^* by first modifying the block diagram (Figure 1) and LFT (Figure 3) to reflect the uncertainty decomposition described here. Likewise, the LFT equations can be derived by separating the normalized uncertainty $\tilde{\Delta}$ and LTV learning dynamics M_{LTV} apart from the remaining LTI dynamics P. The key point is that P now includes the uncertainty weight W_u which will reduce the conservativeness of the stability bound β^*.

As mentioned in Section III-B, we can set $d_2 = 1$ without loss of generality. Additionally, we know that $\delta = 1$ since we assume $\|\tilde{\Delta}\|_{\infty\to\infty} \leq 1$. Note again that we have chosen to use the induced ℓ_∞-norm for reasons discussed in Section III-C. Next, we select any $\beta > 0$ as our initial guess for the robust stability bound β^*. Finally, we partition the dynamics P according (18), construct the $\tilde{P}_{11}$ dynamics according to (19), and compute $\|\tilde{P}_{11}\|_{\infty\to\infty}$ for a range of d_1 values. If the minimum $\|\tilde{P}_{11}\|_{\infty\to\infty}$ over the range of d_1 values is greater than 1, we decrease β until we obtain the largest β such that $\|\tilde{P}_{11}\|_{\infty\to\infty} < 1$. Note that the induced ℓ_∞-norm of an LTI system is equal to the ℓ_1-norm of its impulse response. We compute an upper bound on the ℓ_1-norm based on bounding the tail end of the impulse response (see Section 4.3 of [23]). For our example, the robust stability bound is $\beta^* = 1.063$.

Figure 6 shows the averaged per-step cost $J_T(x,u)/T$ of C-OCO with $H = 1$ as a function of β. Again, we compare the performance with the nominal (red dashed) and uncertain (blue solid) plants. The uncertainty was constructed by randomly generating normalized uncertainties $\tilde{\Delta}(z)$ with $\|\tilde{\Delta}\|_{\infty\to\infty} = 1$ and multiplying by $W_u(z)$ according to (35). When $\beta = 0$, the OCO has no freedom to learn the disturbance, and pure state-feedback (SF) is recovered for both the nominal (red square) and uncertain plants. As β is increased, the OCO is allowed more freedom to learn the disturbance, and we see similar improved performance in both the nominal and uncertain plants. However, when β is "too large" such that the robust stability condition (Theorem 1) no longer holds, C-OCO on the uncertain plant may become unstable. Figure 6 shows the stability bound at $\beta^* = 1.063$ which appears somewhat conservative. Once the constraint β becomes inactive, C-OCO recovers U-OCO performance for the nominal and uncertain plants. For the nominal plant, this indicates a limit as to how much the OCO can improve upon the baseline controller. For the uncertain plants, this indicates a limit as to how much the OCO performance can be degraded by uncertainty. Hence, there is a trade off between OCO performance and robustness to uncertainty.

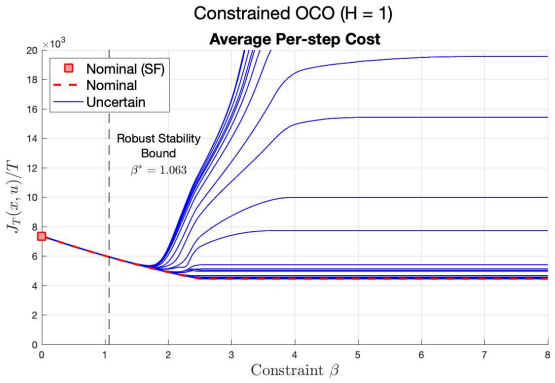

Fig. 6. Average per-step cost of running C-OCO for $H = 1$ and varying β on the nominal (red dashed) and 100 uncertain (blue solid) plants. C-OCO improves performance for the nominal and uncertain plants until β is too large, causing some uncertain plants to become unstable.

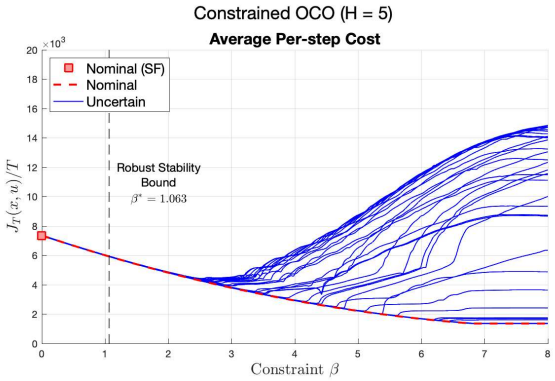

Fig. 7. Average per-step cost of running C-OCO for $H = 5$ and varying β on the nominal (red dashed) and 100 uncertain (blue solid) plants. C-OCO improves performance for the nominal and uncertain plants until β is too large, causing some uncertain plants to become unstable.

Again, Figure 7 shows the averaged per-step cost $J_T(x,u)/T$ of C-OCO as a function of β, this time with $H = 5$. The results are similar to the case when $H = 1$, but the nominal plants appear to go unstable at larger values of β. This illustrates that increasing the learning horizon H may increase robustness, however the stability bound $\beta^* = 1.063$ is more conservative. This suggests that the true stability

bound also depends on the learning horizon H which is not taken advantage of in the robust stability condition in Theorem 1.

VI. Conclusion

In this paper, we establish a robust stability condition using the small gain theorem for a class of OCO controllers with memory and use the result to compute the robust stability bound. In particular, we impose the stability bound as a constraint on the controller point-wise in time. We provide numerical results to illustrate that imposing the robust stability constraint will ensure stability in the presence of bounded uncertainties. Future work will focus on reducing the conservatism of the robust stability bound β^*, computing the exact projection, and developing an OCO controller for the output-feedback case.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 2347026. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] O. Anava, E. Hazan, and S. Mannor, "Online convex optimization against adversaries with memory and application to statistical arbitrage," 2014.

[2] E. Hazan, "The Convex Optimization Approach to Regret Minimization," in *Optimization for Machine Learning*. The MIT Press, 09 2011. [Online]. Available: https://doi.org/10.7551/mitpress/8996.003.0012

[3] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proceedings of the 20th international conference on machine learning (icml-03)*, 2003, pp. 928–936.

[4] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, no. 2-3, pp. 169–192, 2007.

[5] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[6] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends® in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.

[7] N. Agarwal, E. Hazan, and K. Singh, "Logarithmic regret for online control," in *Advances in Neural Information Processing Systems*, 2019, pp. 10 175–10 184.

[8] D. Foster and M. Simchowitz, "Logarithmic regret for adversarial online control," in *International Conference on Machine Learning*, 2020, pp. 3211–3221.

[9] G. Goel, N. Agarwal, K. Singh, and E. Hazan, "Best of both worlds in online control: Competitive ratio and policy regret," *arXiv preprint arXiv:2211.11219*, 2022.

[10] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh, "Online control with adversarial disturbances," in *International Conference on Machine Learning*. PMLR, 2019, pp. 111–119.

[11] K. Zhou, J. Doyle, and K. Glover, *Robust and optimal control*. Pearson, 1995.

[12] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. John Wiley and Sons, 2005.

[13] Y. Rahman, A. Xie, J. B. Hoagg, and D. S. Bernstein, "A tutorial and overview of retrospective cost adaptive control," in *2016 American Control Conference*, 2016, pp. 3386–3409.

[14] R. Venugopal and D. S. Bernstein, "Adaptive disturbance rejection using ARMARKOV/Toeplitz models," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 2, pp. 257–269, 2000.

[15] M. A. Santillo and D. S. Bernstein, "Adaptive control based on retrospective cost optimization," *Journal of guidance, control, and dynamics*, vol. 33, no. 2, pp. 289–304, 2010.

[16] G. Goel and B. Hassibi, "Measurement-feedback control with optimal data-dependent regret," *arXiv preprint arXiv:2209.06425*, 2022.

[17] ——, "Regret-optimal estimation and control," *arXiv preprint arXiv:2106.12097*, 2021.

[18] ——, "Regret-optimal measurement-feedback control," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 1270–1280.

[19] ——, "Regret-optimal control in dynamic environments," *arXiv preprint arXiv:2010.10473*, 2020.

[20] J. Doyle, "Guaranteed margins for LQG regulators," *IEEE Transactions on Automatic Control*, vol. 23, no. 4, pp. 756–757, 1978.

[21] H. K. Khalil, *Nonlinear Systems*, third ed. ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[22] A. Packard and J. Doyle, "The complex structured singular value," *Automatica*, vol. 29, no. 1, pp. 71–109, 1993.

[23] M. Dahleh and I. Diaz-Bobillo, *Control of uncertain systems: a linear programming approach*. Prentice-Hall, 1995.