Online Convex Programming and Regularization in Adaptive Control

Maxim Raginsky, Alexander Rakhlin, and Serdar Yüksel

Abstract— **Online Convex Programming (OCP) is a recently developed model of sequential decision-making in the presence of time-varying uncertainty. In this framework, a decision-maker selects points in a convex feasible set to respond to a dynamically changing sequence of convex cost functions. A generic algorithm for OCP, often with provably optimal performance guarantees, is inspired by the Method of Mirror Descent (MD) developed by Nemirovski and Yudin in the 1970's. This paper highlights OCP as a common theme in adaptive control, both in its classical variant based on parameter tuning and in a more modern supervisory approach. Specifically, we show that: (1) MD leads to a generalization of classical adaptive control schemes based on recursive parameter tuning; (2) A supervisory controller switching policy that uses OCP to estimate system parameters from a sequence of appropriately regularized output prediction errors can flexibly adapt to presence or absence of output disturbances in the system.**

I. Introduction

A. *Certainty equivalence and adaptive control*

Suppose we have an unknown discrete-time deterministic linear SISO system Σ with input (control) sequence $\{u_t\}$ and output sequence $\{y_t\}$. We wish to adaptively control Σ so as to achieve *output regulation*, i.e., drive the output to zero, whenever the noise and disturbance signals are zero. Moreover, all system signals must remain bounded in response to arbitrary bounded noise and disturbance signals. Similar considerations apply to stochastic systems.

Suppose we have an indexed class of models $\mathcal{M} = \{\Sigma_\theta : \theta \in \Theta\}$, where Θ is a compact convex subset of a finite-dimensional linear space, and we have reason to believe that there is some $\theta^* \in \Theta$, such that Σ is reasonably well modeled by Σ_{θ^*}. Suppose we also have an indexed class of controllers $\mathcal{K} = \{K_\gamma : \gamma \in \Gamma\}$, where Γ is either a finite set or a compact subset of a finite-dimensional linear space, and a model-to-controller mapping $\chi : \Theta \to \Gamma$, such that, for each $\theta \in \Theta$, $K_{\chi(\theta)}$ achieves "satisfactory" performance on Σ_θ. A wide variety of adaptive control schemes is based on the "certainty equivalence" heuristic: at each time step, the choice of the controller is based on the current estimate of the system model given all currently available information. We can specify a generic certainty-equivalent adaptive control (CEAC) scheme by a tuple $(\mathcal{M}, \mathcal{K}, \chi, \pi, \delta)$, where $\mathcal{M}$, $\mathcal{K}$ and χ have already been defined, π is a sequence of mappings $\pi_t : \mathbb{R}^t \times \mathbb{R}^{t-1} \to \Theta$ that serve as model parameter estimators, and δ is a sequence of mappings $\delta_t : \mathbb{R}^t \times \mathbb{R}^{t-1} \to \{0,1\}$ that are used to decide whether to switch to a new controller or to keep the currently used one. The operation of the scheme is displayed in Algorithm 1.

Algorithm 1 A Generic CEAC Scheme

```
for $t = 0, 1, \dots$ do
    Compute the parameter estimate $\theta_t = \pi_t(y^t, u^{t-1})$
    Compute $\beta_t = \delta_t(y^t, u^{t-1})$
    Let $\gamma_t = \beta_t \chi(\theta_t) + (1 - \beta_t)\gamma_{t-1}$
    Switch $K_{\gamma_t}$ into the loop
end for
```

Example 1 (Classical adaptive control based on recursive parameter updates)**.** Let $\mathcal{M}$ consist of models of the form

$$\Sigma_\theta : \qquad A_\theta(z^{-1}) y_{t+1} = B_\theta(z^{-1}) u_t + w_{t+1} \tag{1}$$

where $A_\theta(z^{-1}) = 1 - \sum_{i=1}^p a_i z^{-i}, B_\theta(z^{-1}) = \sum_{i=0}^q b_i z^{-i}$ are polynomials in the unit-delay operator z^{-1}, parametrized by $\theta = (a_1, \dots, a_p, b_0, \dots, b_q)^\intercal \in \mathbb{R}^{p+q+1}$, and $\{w_t\}$ is a scalar disturbance process. Let $\Gamma \equiv \Theta$ and to each $\theta \in \Theta$ associate the controller

$$K_\theta : \qquad G_\theta(z^{-1}) u_t = H_\theta(z^{-1}) y_t, \tag{2}$$

where $G_\theta(z^{-1}) = B_\theta(z^{-1}), H_\theta(z^{-1}) = -\sum_{i=1}^p a_i z^{-(i-1)}$. We assume that for each $\theta \in \Theta$ the feedback interconnection of Σ_θ and K_θ is minimum phase, and so let $\chi(\theta) \equiv \theta$.

Let $\theta_t = \pi_t(y^t, u^{t-1})$ be a sequence of recursively computed parameter estimates, and let $\delta_t(y^t, u^{t-1}) \equiv 1$ for all t and all (y^t, u^{t-1}). This is the classical one-step-ahead direct adaptive control (see, e.g., Goodwin and Sin [1]).

Example 2 (Supervisory control)**.** Let $\{u_t\}$ and $\{y_t\}$ denote the input and output sequences of Σ. To construct the class of models $\mathcal{M}$, we first define a family of observers $\mathcal{O} = \{O_\theta : \theta \in \Theta\}$ with a common state realization

$$O_\theta : \qquad v_{t+1} = A v_t + B y_t + D u_t \tag{3a}$$

$$y_t^\theta = C_\theta v_t \tag{3b}$$

The corresponding model class $\mathcal{M} = \{\Sigma_\theta : \theta \in \Theta\}$ is obtained by injecting y_t^θ back into O_θ, leading to

$$\Sigma_\theta : \qquad x_{t+1}^\theta = (A + BC_\theta) x_t^\theta + D u_t \tag{4a}$$

$$y_t^\theta = C_\theta x_t^\theta \tag{4b}$$

M. Raginsky is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA; m.raginsky@duke.edu. Research supported by U.S. National Science Foundation under grant CCF-1017564.

A. Rakhlin is with the Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104, USA; rakhlin@wharton.upenn.edu.

S. Yüksel is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6; yuksel@mast.queensu.ca. Research supported by the Natural Sciences and Engineering Research Council of Canada.

To define the controllers, we assume that Γ is a finite set and let $\mathcal{K} = \{K_\gamma : \gamma \in \Gamma\}$ consist of models with a common state realization

$$K_\gamma : \qquad z_{t+1} = F_\gamma z_t + G_\gamma y_t \tag{5a}$$

$$u_t = H_\gamma z_t + S_\gamma y_t \tag{5b}$$

The feedback interconnection of Σ_θ and K_γ has the state-space representation

$$\begin{pmatrix} x_{t+1} \\ z_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} A + BC_\theta + DS_\gamma C_\theta & DH_\gamma \\ G_\gamma C_\theta & F_\gamma \end{pmatrix}}_{A_{\theta\gamma}} \begin{pmatrix} x_t \\ z_t \end{pmatrix} \tag{6a}$$

$$y_t = \begin{pmatrix} C_\theta & 0 \end{pmatrix} \begin{pmatrix} x_t \\ z_t \end{pmatrix} \tag{6b}$$

Assume that there exists a partition $\Theta = \bigcup_{\gamma \in \Gamma} \Theta_\gamma$, such that every Θ_γ is compact and K_γ is a good controller for all $\{\Sigma_\theta : \theta \in \Theta_\gamma\}$, and let $\chi(\theta) = \gamma$ if $\theta \in \Theta_\gamma$. We assume also that the system Σ_θ is detectable.

To construct π, define for each $\theta \in \Theta$ the output prediction error $e_t(\theta) \triangleq y_t^\theta - y_t$. Now, choose a sequence $\{a_t\}$ of nonnegative reals, not all of which are zero, define

$$J_t(\theta) \triangleq \sum_{\tau=0}^{t} a_{t-\tau} |e_\tau(\theta)|^2, \qquad \forall t \geq 0, \theta \in \Theta$$

and let $\theta_t = \pi_t(y^t, u^{t-1}) \triangleq \arg\min_{\theta \in \Theta} J_t(\theta)$. To construct δ, choose positive sequences $\{h_t\}, \{\varepsilon_t\}$ and let

$$\begin{aligned} \beta_t &= \delta_t(y^t, u^{t-1}) \\ &\triangleq \begin{cases} 1, & \text{if } (1 + h_t) J_t(\theta_t) \leq \min\limits_{\theta \in \Theta_{\gamma_{t-1}}} J_t(\theta) - h_t \varepsilon_t \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

The (discrete-time variant of) scale-independent hysteresis-based switching [2] is a special case of this construction (which was analyzed in [3]).

B. Common themes and our contribution

Speaking conceptually, the purposes of π and δ in a CEAC scheme are distinct: the former estimates the best model given the current data, while the latter determines whether to the controller currently in the loop is to be replaced with one matched to the latest model estimate. In general, therefore, the tasks of gathering information about the system and deciding how to control it are separated. This is similar to the distinction between the averaging step in stochastic approximation (SA), which generates approximate solutions, and the SA update itself, which gathers information about the unknown objective function [4], [5]. The hysteresis switching policy of Example 2 illustrates this separation principle quite nicely: the new model estimate affects the controller selection only if it does significantly better than the current controller. There is no such separation in Example 1, although it is possible to implement a parameter update scheme based on SA trajectory averaging.

Another common theme underlying both examples is that, in order to obtain the parameter estimates, one has to solve a time-varying sequence of optimization problems, which are often convex in the underlying parameter θ. Convexity plays an important role in control, both as an analytic technique [6] and as a structural characteristic permitting efficient solution of a wide variety of problems in controller design, verification, and implementation [7]. The contribution of the present paper is a unifying perspective on parameter estimation in adaptive control (both in its classical variant based on parameter tuning and in a more recent supervisory approach based on controller switching) through the concept of *online convex programming* and the *Mirror Descent* scheme of Nemirovski and Yudin [4], [5], [8]. At this point, we are not aiming for a definitive treatment; our goal in this paper is to look at a fundamental topic in control theory from a fresh perspective and hopefully to stimulate further research, both in the optimization and in the control communities.

II. Online convex programming and the method of mirror descent

The term "Online Convex Programming" (OCP) [9] refers broadly to a class of problems where a decision-maker sequentially chooses points in a convex set in response to a time-varying sequence of convex cost functions. Formally, OCP is a *game* between two players, the Decision-Maker (DM) and the Environment (E), specified by a pair $(\Theta, \mathcal{L})$, where $\Theta \subset \mathbb{R}^n$ is a compact and convex and $\mathcal{L}$ is a class of convex functions $\ell : \Theta \to \mathbb{R}$, which unfolds as follows:

Algorithm 2 Online Convex Programming

```
for $t = 0, 1, 2, \dots$ do
    DM chooses $\theta_t \in \Theta$
    E chooses $\ell_t \in \mathcal{L}$ and reveals it to DM
    DM suffers loss $\ell_t(\theta_t)$
end for
```

In the learning theory community, OCP has emerged as one of the basic models of sequential (online) learning processes. In that context, the objective of the DM, for each horizon T, is to minimize the *regret*, i.e., the difference between the total cost $\sum_{t=0}^{T} \ell_t(\theta_t)$, and the smallest total cost that could have been achieved in hindsight by a single $\theta \in \Theta$. Then learning is synonymous with *Hannan consistency*[1], i.e., the existence of a strategy that achieves regret sublinear in T.

Now, regret does not play such a major role in the context of control. Nevertheless, as we shall show, the same technique used in learning theory to achieve Hannan consistency will lead to good adaptive control strategies. This technique is based on the method of Mirror Descent (MD) developed by Nemirovski and Yudin [4] as a robust alternative to standard projected subgradient methods (cf. [5], [8]). The structure of MD hinges on the concept of a *distance-generating function* and its induced *Bregman divergence* [11]:

Definition 1 (Distance-generating function)**.** *Fix an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^n$. A function $\omega : \Theta \to \mathbb{R}$ is a* distance-

[1] This term reflects J. Hannan's seminal work on learning Bayes optimal policies in repeated games [10].

generating function *with modulus $\alpha > 0$ with respect to $\|\cdot\|$, provided it has the following properties:*

1) *ω is convex and continuous on Θ*
2) *The set $\Theta^\circ \triangleq \{\theta \in \Theta : \partial\omega(\theta) \neq \varnothing\}$, where $\partial\omega(\theta)$ denotes the subdifferential of ω at θ, is convex*
3) *Restricted to Θ°, ω is C^1 and* strongly convex *with parameter $\alpha > 0$, i.e., for all $\theta, \theta' \in \Theta^\circ$*

$$\omega(\theta') \geq \omega(\theta) + \nabla\omega(\theta)^\intercal(\theta' - \theta) + \frac{\alpha}{2}\|\theta - \theta'\|^2.$$

Definition 2 (Bregman divergence)**.** *Let ω be a distance generating function on Θ. Then the* Bregman divergence *induced by ω is the function $D_\omega : \Theta \times \Theta^\circ \to \mathbb{R}^+$ defined by*

$$D_\omega(\theta, \theta') \triangleq \omega(\theta) - \omega(\theta') - \nabla\omega(\theta')^\intercal(\theta - \theta').$$

As an example, take $\omega(\theta) = \frac{1}{2}\|\theta\|_2^2$. Then $D_\omega(\theta, \theta') = \frac{1}{2}\|\theta - \theta'\|_2^2$; see [5], [8] for other examples.

Lemma 1 (Properties of Bregman divergences)**.** *$D_\omega(\cdot,\cdot)$ is nonnegative and strongly convex: $D_\omega(\theta, \theta') \geq \frac{\alpha}{2}\|\theta - \theta'\|^2$ for all $\theta, \theta' \in \Theta^\circ$. Moreover for any $\theta \in \Theta$ and $\theta', \theta'' \in \Theta^\circ$,*

$$\begin{aligned} D_\omega(\theta, \theta') &+ D_\omega(\theta', \theta'') - D_\omega(\theta, \theta'') \\ &= (\nabla\omega(\theta'') - \nabla\omega(\theta'))^\intercal(\theta - \theta') \end{aligned} \qquad (7)$$

Now, following Nemirovski et al. [5], define for every $\theta \in \Theta^\circ$ the *prox-mapping* $\Pi_\theta : \mathbb{R}^n \to \Theta^\circ$ as follows:

$$\Pi_\theta(\xi) \triangleq \operatorname*{arg\,min}_{\theta' \in \Theta} \left[\xi^\intercal(\theta' - \theta) + D_\omega(\theta', \theta)\right] \qquad (8)$$

The method of Mirror Descent is shown in Algorithm 3, where $\{\eta_t\}$ is a decreasing sequence of nonnegative step sizes and $\bar\nabla\ell_t(\theta_t)$ is an arbitrary subgradient of ℓ_t at θ_t.

Algorithm 3 Mirror Descent

```
Choose an arbitrary $\theta_0 \in \Theta$
for $t = 0, 1, 2, \dots$ do
    Receive $\ell_t \in \mathcal{L}$ and incur loss $\ell_t(\theta_t)$
    Output $\theta_{t+1} = \Pi_{\theta_t}\left(\eta_t \bar\nabla\ell_t(\theta_t)\right)$
end for
```

The following lemma (cf. [5] for the proof) is basic:

Lemma 2. *For every $\theta \in \Theta$, the MD updates satisfy*

$$V_{t+1}(\theta) \leq V_t(\theta) + \eta_t\bar\nabla\ell_t(\theta_t)^\intercal(\theta - \theta_t) + \frac{\eta_t^2\|\bar\nabla\ell_t(\theta_t)\|_*^2}{2\alpha},$$

where $V_t(\theta) \triangleq D_\omega(\theta, \theta_t)$, and $\|\cdot\|_*$ denotes the norm dual to $\|\cdot\|$ defined via $\|u\|_* = \sup_{\|v\| \leq 1} v^\intercal u$.

We close this section with a discussion of computational issues. First of all, it can be shown that the MD updates can be computed recursively as follows [8]: Suppose that we can efficiently compute the Legendre–Fenchel dual of ω, defined as $\Omega(\xi) \triangleq \sup_\theta [\xi^\intercal\theta - \omega(\theta)]$. Suppose also that ω is *steep*, i.e., for any sequence $\{\theta_k\}$ converging to a boundary point of Θ, we have $\|\nabla\omega(\theta_k)\|_* \to +\infty$. Then at time t the MD update takes the following form:

$$\xi_{t+1} = \nabla\omega(\theta_t) - \eta_t\bar\nabla\ell_t(\theta_t), \; \theta_{t+1} = \nabla\Omega(\xi_{t+1}) \qquad (9)$$

This structure is what gives the MD method its name: the current point θ_t is mapped to its "mirror image" in the dual space $\nabla\omega(\Theta)$, updated via gradient descent, then mapped back to Θ by means of the inverse mapping $\nabla\Omega$. When all ℓ_t are equal to the same C^1 function ℓ, we can view (9) as a discretization of the following continuous-time evolution:

$$\dot\xi(t) = -\nabla\ell(\nabla\Omega(t)), \qquad \theta(t) = \nabla\Omega(t). \qquad (10)$$

Denoting by θ^* any minimizer of ℓ on Θ, we can show that $V(\xi) \triangleq D_\Omega(\xi, \nabla\omega(\theta^*))$ is a Lyapunov function for (10), i.e., it decreases along any trajectory of (10); in fact, this was the motivation originally given by Nemirovski and Yudin [4].

III. Adaptive control based on parameter tuning

We now consider the setting of Example 1. For future convenience, let us cast the model (1) in the regressive form

$$\Sigma_\theta : \qquad y_{t+1} = \phi_t^\intercal\theta + w_{t+1}, \qquad (11)$$

where $\phi_t \triangleq (y_t, \dots, y_{t-p-1}, u_t, \dots, u_{t-q})^\intercal$. We assume that the true system Σ has the form Σ_{θ^*} for an unknown $\theta^* \in \Theta$.

Let $\{u_t\}$ be an arbitrary sequence of inputs, and let $\{y_t\}$ be the resulting sequence of outputs, $y_{t+1} = \phi_t^\intercal\theta^* + w_{t+1}$. For each $t = 0, 1, 2, \dots$ define the loss functions

$$\ell_t(\theta) = (y_{t+1} - \phi_t^\intercal\theta)^2. \qquad (12)$$

Let $\mathcal{L}$ denote the class of all functions of the form (12) as θ^* ranges over Θ and $\{u_t\}$ ranges over all arbitrary control sequences. Hence, we can cast our adaptive control problem as an instance of OCP over $\mathcal{L}$, but with an additional twist: at time t, we set u_t ourselves and therefore have full knowledge of ϕ_t. The only unknown is y_{t+1}, which is revealed to us once we apply the control u_t. What is essential is that, by modifying u_t, we adjust one coordinate of ϕ_t to ensure that $\phi_t^\intercal\theta_t = 0$. Hence, we have $\ell_t(\theta_t) = y_{t+1}^2$. Furthermore, in the absence of disturbances ($w_t \equiv 0$ for all t), the loss of θ^* is zero: $\ell_t(\theta^*) \equiv 0$ for all t. We now show the following:

Theorem 1. *Assume that there are no disturbances, $w_t \equiv 0$ for all t. Choose a norm $\|\cdot\|$ on $\mathbb{R}^n$, a constant $\alpha > 0$, and a $(\alpha, \|\cdot\|)$ distance-generating function ω on Θ. For each $t = 0, 1, 2, \dots$ let $\eta_t \triangleq \frac{a}{c + \|\phi_t\|_*^2}$ for some $c > 0, 0 < a < \frac{\alpha}{2}$. Then the MD scheme for updating θ_t, coupled with the certainty-equivalent rule $\phi_t^\intercal\theta_t = 0$, achieves output regulation, i.e.,*

$$\lim_{t\to\infty} y_t = 0 \quad \textit{and} \quad \sup_{t \geq 0} |u_t| < \infty.$$

Proof. The updates $\{\theta_t\}$ satisfy the relations

$$\nabla\ell_t(\theta_t) = -2(y_{t+1} - \phi_t^\intercal\theta_t)\phi_t = -2y_{t+1}\phi_t. \qquad (13)$$

For each t, define $V_t \triangleq V_t(\theta^*, \theta_t) \equiv D_\omega(\theta^*, \theta_t)$. Using (13) and Lemma 2, we get the following *Lyapunov recursion*:

$$V_{t+1} \leq V_t + 2a\left(\frac{a\|\phi_t\|_*^2}{\alpha(c + \|\phi_t\|_*^2)} - 1\right)\frac{y_{t+1}^2}{c + \|\phi_t\|_*^2},$$

where we used $\phi_t^\intercal\theta^* = y_{t+1}$ and $\phi_t^\intercal\theta_t = 0$. Since $c > 0$ and $0 < a < \frac{\alpha}{2}$, we have

$$\frac{a\|\phi_t\|_*^2}{\alpha(c + \|\phi_t\|_*^2)} < \frac{1}{2} \;\Rightarrow\; \frac{y_{t+1}^2}{c + \|\phi_t\|_*^2} \leq \frac{V_t - V_{t+1}}{a}, \forall t.$$

Summing from $t = 0$ to $t = T$, we obtain

$$\sum_{t=0}^{T} \frac{y_{t+1}^2}{c + \|\phi_t\|_*^2} \le \frac{V_0 - V_{T+1}}{a} \le \frac{V_0}{a},$$

which implies that

$$\lim_{T\to\infty} \sum_{t=0}^{T} \frac{y_{t+1}^2}{c + \|\phi_t\|_*^2} \le \frac{V_0}{a} < \infty \Rightarrow \lim_{T\to\infty} \frac{y_{t+1}^2}{c + \|\phi_t\|_*^2} = 0.$$

Since the system is minimum phase, there exist constants $K_1, K_2 \ge 0$ that depend on p and q, such that

$$\|\phi_t\|_* \le K_1 + K_2 \max_{0\le\tau\le t} |y_\tau|, \qquad \forall t \ge 0$$

(cf. [1], [12]). Hence, we can apply the Key Technical Lemma of Goodwin and Sin [1, Lm. 6.2.1] to conclude that $y_t \to 0$ and that u_t remain bounded. □

Theorem 2. *Consider the stochastic model $y_{t+1} = \phi_t^\intercal \theta^* + w_{t+1}$ with i.i.d zero-mean noise, $\mathbb{E}w_t^2 = \sigma^2$. Choose a norm $\|\cdot\|$ on $\mathbb{R}^n$, a constant $\alpha > 0$, and a $(\alpha, \|\cdot\|)$ distance-generating function ω on Θ. For each $t = 0, 1, 2, \ldots$ let $\eta_t \triangleq \frac{\alpha}{\alpha + \sum_{\tau=0}^{t} \|\phi_\tau\|_*^2}$. Then the MD update for θ_t, coupled with the certainty-equivalent rule $\phi_t^\intercal \theta_t = 0$, is self-optimizing:*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_t^2 = \sigma^2 \qquad a.s.$$

Proof. The proof follows Goodwin et al. [13] (cf. also [12, Sec. 13.4]), except we use the Lyapunov function induced by ω. From Lemma 2, convexity of ℓ_t, and (13),

$$\eta_t(\ell_t(\theta_t) - \ell_t(\theta^*)) \le V_t(\theta^*) - V_{t+1}(\theta^*) + \frac{\eta_t^2 y_{t+1}^2 \|\phi_t\|_*^2}{2\alpha}.$$

Denote the conditional expectation $\mathbb{E}_t[A] \triangleq \mathbb{E}[A|w_1^{t-1}]$. Since $\mathbb{E}_{t+1}[y_{t+1}] = \phi_t^\intercal \theta^*$ and $\phi_t^\intercal \theta_t = 0$, it is easy to verify that $\mathbb{E}_{t+1} y_{t+1}^2 = (\mathbb{E}_{t+1} y_{t+1})^2 + \sigma^2$ and $\mathbb{E}_{t+1}\left[\ell_t(\theta_t) - \ell_t(\theta^*)\right] = (\mathbb{E}_{t+1} y_{t+1})^2$, yielding

$$\begin{aligned}\eta_t(\mathbb{E}_{t+1} y_{t+1})^2 \le V_t(\theta^*) &- \mathbb{E}_{t+1} V_{t+1}(\theta^*) \\ &+ \frac{\eta_t^2((\mathbb{E}_{t+1} y_{t+1})^2 + \sigma^2)\|\phi_t\|_*^2}{2\alpha}.\end{aligned}$$

Rearranging and using the fact that $\eta_t \le \alpha \|\phi_t\|_*^{-2}$,

$$\frac{1}{2}\eta_t(\mathbb{E}_{t+1} y_{t+1})^2 \le V_t(\theta^*) - \mathbb{E}_{t+1} V_{t+1}(\theta^*) + \frac{\eta_t^2 \sigma^2 \|\phi_t\|_*^2}{2\alpha}.$$

It is easy to show that

$$\sum_{t=0}^{T} \eta_t^2 \|\phi_t\|_*^2 = \sum_{t=0}^{T} \frac{\alpha^2 \|\phi_t\|_*^2}{(\alpha + \sum_{s=0}^{t} \|\phi_s\|_*^2)^2} \le \alpha.$$

The Robbins–Siegmund supermartingale convergence theorem [14, Thm. 1.3.12] implies that $\sum_{t=1}^{\infty} \eta_t (\mathbb{E}_{t+1} y_{t+1})^2 < \infty$ a.s. The rest of the proof follows exactly as in [12]. □

Remark 1. With the choice $\omega(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, we recover the one-step-ahead adaptive controller based on the projection algorithm in the deterministic case (cf. [1, Sec. 6.3.1]) and the gradient approach of [13] in the stochastic case. However, we can choose other distance-generating functions depending on the geometry of Θ. In particular, judicious choice of ω can vastly improve the dimension dependence of MD convergence rates [5]. This degree of freedom deserves a closer look in the context of robust control, where one may have structured uncertainty sets Θ (e.g., polyhedra or simplices). Put another way, by choosing ω we also choose the Lyapunov function $V(\cdot) \equiv D_\omega(\theta^*, \cdot)$. In view of this, it would be of interest to connect the continuous-time version of MD, Eq. (10), to adaptive control. This would lead to an intriguing generalization of the ODE method of Ljung [15].

Remark 2. For a reader familiar with the online learning literature, the following interpretation of the results of this section might be useful. As we have mentioned earlier, the control problem can be cast as sequential minimization of time-varying cost functions. If there is a θ^* with zero loss, we expect to have *constant* regret (see, e.g., Theorem 11.2 in [16]). Theorem 1 then follows immediately since individual costs on each round must decay to zero. When there is noise in the system, the cumulative loss of the best comparator θ^* is non-zero. However, generic Hannan-consistent strategies yielding $o(T)$ regret are not strong enough to recover the results of Theorem 2, and the specific structure of the loss function becomes essential. It is also interesting to note that in the stochastic case the stepsize η_t is kept small, on the order of $O(1/t)$, similar to the case of regret minimization over strongly convex functions [17].

IV. Supervisory control

We now consider the supervisory control framework of Example 2, where we also assume the following:

- **Exact matching with output disturbances:** the unknown system Σ is of the form

$$\Sigma: \quad \begin{aligned} x_{t+1} &= (A + BC_{\theta^*})x_t + Du_t + Rw_t \\ y_t &= C_{\theta^*} x_t \end{aligned}$$

  for some $\theta^* \in \Theta$, where $\{w_t\}$ is a bounded disturbance signal (we do not require that this bound be known, only that it exists). The matrix R is assumed to be such that the mapping from w_t to y_t is stable. Hence, we can model the effect of the disturbance as a bounded additive sequence $\widehat{w}_t$ in the *output* of Σ [3].
- **Stability margin:** there exists $\lambda_0 \in (0, 1)$, such that

$$\sup_{\theta\in\Theta} \|A_{\theta\chi(\theta)}\| < \lambda_0. \tag{14}$$

- **Smooth parametrization:** the mappings $\theta \mapsto A_\theta$ and $\theta \mapsto C_\theta$ are continuous.
- **Convexity:** the square of the output prediction error $e_t(\theta) = y_t^\theta - y_t$ is convex in θ for all t (with $\{u_t\}$ and $\{y_t\}$ fixed).

We will consider estimation strategies of the following type. We choose a distance-generating function $\omega : \Theta \to \mathbb{R}$ satisfying the bounds $0 < \mu_1 \le \omega(\theta) \le \mu_2 < \infty$ for all $\theta \in \Theta$, a monotone decreasing sequence $\{\eta_t\}$ with $\eta_0 = 1$, and a constant $\lambda \in (\lambda_0, 1)$, and define

$$J_t(\theta) \triangleq \sum_{\tau=0}^{t} \lambda^{t-\tau}|e_\tau(\theta)|^2 + \eta_t \omega(\theta). \qquad (15)$$

Our parameter estimators will be of the form

$$\theta_t = \pi_t(y^t, u^{t-1}) \triangleq \arg\min_{\theta \in \Theta} J_t(\theta).$$

Note that the problem of solving for $\{\theta_t\}$ is an instance of OCP with an additive structure. In particular, defining $\alpha_t = \lambda^{-t}\eta_t$, we can rewrite (15) as a regularized objective

$$J_t(\theta) = \frac{1}{\alpha_t}\sum_{\tau=0}^{t} \ell_\tau(\theta) + \omega(\theta),$$

where, for each τ, $\ell_\tau(\theta) \triangleq \lambda^{-\tau}|e_\tau(\theta)|^2$ is a convex function of θ. We also fix a hysteresis constant $h > 0$ and consider the hierarchical hysteresis switching rule

$$\delta_t(y^t, u^{t-1}) \triangleq \begin{cases} 1, & \text{if } (1+h)J_t(\theta_t) \le \min\limits_{\theta \in \Theta_{\gamma_{t-1}}} J_t(\theta) \\ 0, & \text{otherwise} \end{cases}$$

For each t, also define

$$\widehat{\theta}_t \triangleq \begin{cases} \theta_t, & \text{if } \beta_t = \delta_t(y^t, u^{t-1}) = 1 \\ \widehat{\theta}_{t-1}, & \text{otherwise} \end{cases}$$

to be the model index that actually determines the controller choice at time t, since $\gamma_t = \chi(\widehat{\theta}_t)$. The evolution of the switched system can be described by the state-space model

$$x_{t+1} = \tilde{A}_{\widehat{\theta}_t} x_t - L_{\gamma_t} e_t(\widehat{\theta}_t) \qquad (16a)$$

$$y_t = \begin{pmatrix} C_{\theta^*} & 0 \end{pmatrix} x_t - e_t(\theta^*) \qquad (16b)$$

$$u_t = \begin{pmatrix} 0 & H_{\gamma_t} \end{pmatrix} x_t + S_{\gamma_t} y_t \qquad (16c)$$

where $x_t \triangleq \begin{pmatrix} v_t \\ z_t \end{pmatrix}$, $\tilde{A}_\theta \triangleq A_{\theta\chi(\theta)}$, and $L_{\gamma_t} \triangleq \begin{pmatrix} B + DS_{\gamma_t} \\ G_{\gamma_t} \end{pmatrix}$.

Before presenting the proof of stability, we list several useful facts (cf. [18, Sec. 5.7], for example): For each $\theta \in \Theta$, let P_θ denote the solution of the Lyapunov equation

$$(\lambda_0^{-1}\tilde{A}_\theta)^\intercal P(\lambda_0^{-1}\tilde{A}_\theta) - P = -I.$$

Owing to the stability margin condition (14), P_θ is positive definite. Let $\bar{K} \triangleq \sup_{\theta \in \Theta} \operatorname{cond}(P_\theta)$, where $\operatorname{cond}(\cdot)$ denotes the condition number with respect to the 2-norm. Owing to the smoothness condition and to the fact that each Θ_γ is compact, we conclude that, for every $\gamma \in \Gamma$, $\theta \mapsto \tilde{A}_\theta$ is continuous on Θ_γ, hence $\theta \mapsto P_\theta$ and $\theta \mapsto \operatorname{cond}(P_\theta)$ are also continuous on Θ_γ [19]. Hence $\sup_{\theta \in \Theta_\gamma} \operatorname{cond}(P_\theta)$ exists and is finite for each γ, and therefore so is $\bar{K}$. For every $\theta \in \Theta$, the autonomous system $x_{t+1} = \tilde{A}_\theta x_t$ is uniformly exponentially stable, i.e.,

$$\|x_t\| \le \rho \lambda_0^{t-t_0}\|x_{t_0}\|, \qquad \forall t \ge t_0 \ge 0 \qquad (17)$$

where $\rho = \sqrt{\bar{K}}$. Now we can state and prove our main result:

Theorem 3. *Suppose that the regularization parameters $\{\eta_t\}$ are chosen so that the sequence $\alpha_t = \eta_t\lambda^{-t}$ is monotone nondecreasing. Then one can choose h and λ in such a way that all the signals in the supervisory control system remain bounded for every set of initial conditions. Moreover, $y_t^2 = O(\lambda^t + \eta_t) + O(1)$, where the $O(1)$ term converges to zero whenever $w_t \to 0$. In particular, if there are no output disturbances, then choosing $\eta_t = \lambda^t$ for all t we will get $y_t^2 = O(\lambda^t)$, i.e., exponential stability.*

Proof. Given $t > 0$, let $N(t)$ denote the number of switchings over $0 < \tau < t$, i.e., $N(t) = |\{0 < \tau < t : \beta_\tau = 1\}|$. Let $\bar{J}_t(\theta) \triangleq \lambda^{-t}J_t(\theta)$. From (15) we see that (1) $\bar{J}_0(\theta) \ge \omega(\theta) \ge \mu_1 > 0$ and (2) $\bar{J}_{t+1}(\theta) \ge \bar{J}_t(\theta), \forall t \ge 0$. The Hierarchical Hysteresis Switching Lemma [2] then gives

$$N(t) \le 1 + m + \frac{m}{\log(1+h)} \log \frac{\bar{J}_t(\theta)}{\bar{J}_0(\theta_0)}, \qquad \forall \theta \in \Theta \quad (18)$$

where $m = |\Gamma|$. Exact matching with bounded output disturbance implies that

$$\sum_{\tau=0}^{t} \lambda^{-\tau}|e_\tau(\theta^*)|^2 \le M_1 + M_2\lambda^{-t}, \qquad \forall t$$

for some $0 < M_1, M_2 < \infty$, where M_2 converges to zero whenever the bound on the disturbance signal does. This implies that $\bar{J}_t(\theta^*) \le M_1 + (M_2 + \mu_2\eta_t)\lambda^{-t}$. Using this in (18) with $\theta = \theta^*$, we obtain

$$\begin{aligned} N(t) &\le 1 + m + \frac{m}{\log(1+h)} \log \frac{M_1 + (M_2 + \mu_2\eta_t)\lambda^{-t}}{\mu_1} \\ &\le N_0 + \frac{t}{\tau_{\mathrm{AD}}} \end{aligned} \qquad (19)$$

with $N_0 = 1 + m + \frac{m}{\log(1+h)} \log \frac{M_1 + M_2 + \mu_2}{\mu_1}$ and $\tau_{\mathrm{AD}} = \frac{\log(1+h)}{m \log(1/\lambda)}$. In the terminology of Hespanha and Morse [20], Eq. (19) states that the switching policy δ defined above has *chatter bound* N_0 and *average dwell time* τ_{AD}.

Now let $0 < t_1 < t_2 < \ldots < t_{N(t)} < t$ be the switching times between 0 and t. Consider the autonomous system $x_{t+1} = \tilde{A}_{\widehat{\theta}_t} x_t$. During each interval $t_i \le \tau < t_{i+1}$, this system is time-invariant and uniformly exponentially stable. Applying (17) repeatedly, we obtain the bound

$$\|x_t\| \le \rho^{N(t)+1}\lambda_0^t\|x_0\| \le \rho^{N_0+1+t/\tau_{\mathrm{AD}}}\lambda_0^t\|x_0\|.$$

By choosing h and λ appropriately, we can guarantee that $\rho^{1/\tau_{\mathrm{AD}}}\lambda_0 \le \lambda$. Assuming this holds, we get $\|x_t\| \le \bar{\rho}\lambda^t\|x_0\|$ with $\bar{\rho} \equiv \rho^{N_0+1}$. Now consider the switched system (16). Defining $\bar{L} \triangleq \max_\gamma \|L_\gamma\|$, we have

$$\|x_t\| \le \bar{\rho}\lambda^t\|x_0\| + \bar{L}\bar{\rho}\sum_{\tau=0}^{t} \lambda^{t-\tau}|e_\tau(\widehat{\theta}_\tau)|.$$

Applying Cauchy–Schwarz, we get

$$\|x_t\| \le \bar{\rho}\lambda^t\|x_0\| + \frac{\bar{L}\bar{\rho}\lambda^{t/2}}{\sqrt{1-\lambda}}\sqrt{\sum_{\tau=0}^{t} \lambda^{-\tau}|e_\tau(\widehat{\theta}_\tau)|^2}$$

Using the monotonicity of $\{\alpha_t\}$ and the positivity of ω,

$$\begin{aligned}\sum_{\tau=0}^{t} \lambda^{-\tau} |e_\tau(\widehat{\theta}_\tau)|^2 &= \sum_{k=0}^{N(t)} \left(\bar{J}_{t_{k+1}}(\widehat{\theta}_{t_k}) - \bar{J}_{t_k}(\widehat{\theta}_{t_k}) \right) \\ &\qquad - \sum_{k=0}^{N(t)} (\alpha_{t_{k+1}} - \alpha_{t_k}) \omega(\widehat{\theta}_{t_k}) \\ &\le \sum_{k=0}^{N(t)} \left(\bar{J}_{t_{k+1}}(\widehat{\theta}_{t_k}) - \bar{J}_{t_k}(\widehat{\theta}_{t_k}) \right).\end{aligned}$$

We can now apply the Hierarchical Hysteresis Switching Lemma again to bound the above sum by $m(1+h)\bar{J}_t(\theta^*) \le m(1+h)(M_1 + (M_2 + \mu_2\eta_t)\lambda^{-t})$. Hence,

$$\|x_t\| \le \bar{\rho}\lambda^t \|x_0\| + \frac{\bar{L}\bar{\rho}\sqrt{m(1+h)(M_1\lambda^t + M_2 + \mu_2\eta_t)}}{\sqrt{1-\lambda}},$$

which means that $\|x_t\|^2 = O\left(\lambda^t + \eta_t + M_2\right)$. If there are no output disturbances, we will have $M_2 = 0$ and $\|x_t\|^2 = O(\lambda^t + \eta_t)$, i.e., $x_t \to 0$. Moreover, $|e_t(\theta^*)|^2 \le M_1\lambda^t + M_2$, which together with detectability of $\Sigma = \Sigma_{\theta^*}$ in turn implies that $y_t^2 = O(\lambda^t + \eta_t + M_2)$. The control sequence $\{u_t\}$ remains bounded, again by detectability. □

Remark 3. From the above proof it is evident that the role of the decaying regularization parameter η_t is to guarantee that output regulation is achieved when there are no disturbances. Setting $\eta_t \equiv 1$ for all t will not achieve this effect.

V. Conclusion

Our goal in this paper was to highlight the role of online convex programming (OCP) in adaptive control. We have shown that the well-known Mirror Descent scheme of Nemirovski and Yudin can be viewed as a generalization of gradient-based parameter tuning in classical adaptive control. The main import of MD in that setting is the degree of freedom the control designer has in choosing the underlying Lyapunov function (by varying the distance-generating function appropriately). In addition, we have highlighted the importance of regularization in switching control: adding a slowly decaying regularization term to the cumulative sum of output prediction errors ensures that the output and the control signals remain bounded in the presence of bounded disturbances, yet output regulation is achieved when disturbances are absent. This is an improvement over the hysteresis-based scheme of [2], which uses a small positive constant instead of a decaying regularizer. As part of future work, we will investigate adaptive control schemes that combine parameter tuning with controller switching [21].

Over the past decade, OCP has been extensively used in the online learning literature to model sequential decision-making in an uncertain dynamic environment. This led to a number of new insights into fairly advanced concepts from convex optimization, including not only MD, but also self-concordance and interior point methods [22]. Moreover, some connections to control have emerged as well. For instance, recent work by De Farias and Megiddo [23] addresses the problem of combining expert advice in a reactive environment, which can be thought of as OCP over the unit simplex where the extreme points correspond to the different strategies. That work highlights the importance of "dwelling" on a single strategy for a number of rounds, which is quite reminiscent of a similar idea in supervisory control (cf. [2] and references therein). Further exploration of these concepts will be extremely beneficial in control, learning, and optimization.

References

[1] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[2] J. P. Hespanha, D. Liberzon, and A. S. Morse, "Hysteresis-based switching algorithms for supervisory control of uncertain systems," *Automatica*, vol. 39, pp. 263–272, 2003.

[3] S. R. Kulkarni and P. J. Ramadge, "Model and controller selection policies based on output prediction errors," *IEEE Trans. Automat. Control*, vol. 41, no. 11, pp. 1594–1604, 1996.

[4] A. S. Nemirovski and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[5] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1574–1609, 2009.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia: SIAM, 1994.

[7] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Trans. Control Sys. Technol.*, vol. 18, no. 2, pp. 267–278, 2010.

[8] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Res. Lett.*, vol. 31, pp. 167–175, 2003.

[9] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient descent," in *Proc. Int. Conf. on Machine Learning*, 2003, pp. 928–936.

[10] J. Hannan, "Approximation to Bayes risk in repeated play," *Contributions to the Theory of Games*, vol. 3, pp. 97–139, 1957.

[11] L. M. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. and Math. Phys.*, vol. 7, pp. 200–217, 1967.

[12] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice-Hall, 1986.

[13] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, "Discrete time stochastic adaptive control," *SIAM J. Control Optim.*, vol. 19, no. 6, pp. 829–853, 1981.

[14] M. Duflo, *Random Iterative Models*, ser. Applications of Mathematics. Springer, 1997, vol. 34.

[15] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Control*, vol. AC-22, no. 4, pp. 551–575, 1977.

[16] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge, 2006.

[17] P. L. Bartlett, E. Hazan, and A. Rakhlin, "Adaptive online gradient descent," in *Advances in Neural Information Processing Systems 20*, 2007.

[18] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd ed. New York: Springer, 1998.

[19] D. F. Delchamps, "Analytic feedback control and the algebraic Riccati equation," *IEEE Trans. Automat. Control*, vol. AC-29, pp. 1031–1033, 1984.

[20] J. P. Hespanha and A. S. Morse, "Stability of switched systems with average dwell-time," in *Proc. 38th IEEE Conf. on Decision and Control*, Phoenix, AZ, 1999, pp. 2655–2660.

[21] K. S. Narendra and J. Balakrishnan, "Adaptive control using multiple models," *IEEE Trans. Automat. Control*, vol. 42, no. 2, pp. 171–187, 1997.

[22] J. Abernethy, E. Hazan, and A. Rakhlin, "Competing in the dark: An fficient algorithm for bandit linear optimization," in *Proc. 21st Annual Conf. on Learning Theory (COLT)*, 2008.

[23] D. P. De Farias and N. Megiddo, "Combining expert advice in reactive environments," *J. Assoc. Comput. Mach.*, vol. 53, no. 5, pp. 762–799, 2006.