

Switching Supervisory Control Using Calibrated Forecasts

Ibrahim Al-Shyoukh, *Member, IEEE*, and Jeff S. Shamma, *Fellow, IEEE*

Abstract—In this paper, we approach supervisory control as an online decision problem. In particular, we introduce “calibrated forecasts” as a mechanism for controller selection in supervisory control. The forecasted quantity is a candidate controller’s performance level, or reward, over finite implementation horizon. Controller selection is based on using the controller with the maximum calibrated forecast of the reward. The proposed supervisor does not perform a pre-routed search of candidate controllers and does not require the presence of exogenous inputs for excitation or identification. Assuming the existence of a stabilizing controller within the set of candidate controllers, we show that under the proposed supervisory controller, the output of the system remains bounded for any bounded disturbance, even if the disturbance is chosen in an adversarial manner. The use of calibrated forecasts enables one to establish overall performance guarantees for the supervisory scheme even though non-stabilizing controllers may be persistently selected by the supervisor because of the effects of initial conditions, exogenous disturbances, or random selection. The main results are obtained for a general class of system dynamics and specialized to linear systems.

Index Terms—Adaptive control, calibrated forecast, machine learning, supervisory control.

I. INTRODUCTION

ONLINE decision problems [4], [10] arise in several fields of study including, statistical analysis, game theory, computer science, and systems and control. In these problems, a decision maker must take a decision in each of a sequence of stages based on the information available at each stage. Online algorithms, and the related topic of learning algorithms, seek to improve the quality of these decisions as more data is gathered about some underlying process. In the systems and control field, the utilization of learning is extensive. Adaptive control [1], adaptive filtering [22], and supervisory control [12] are all examples.

Supervisory control [11], [15], [19], [21] is an approach to adaptive control that uses information obtained online to decide on an appropriate control action. The “decision maker” is composed of a set of candidate controllers and a supervisor.

Manuscript received May 02, 2007; December 07, 2007. First published March 27, 2009; current version published April 08, 2009. This work was supported by National Science Foundation (NSF) Grant ECS-0501394, ARO Grant W911NF-04-1-0316, and AFOSR Grant FA9550-05-1-0239. Recommended by Associate Editor J. P. Hespanha.

I. Al-Shyoukh is with the Department of Molecular and Medical Pharmacology, University of California Los Angeles, Los Angeles, CA 90095 USA (e-mail: shyoukh@ucla.edu).

J. S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: shamma@gatech.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2009.2014923

At least one of the candidate controllers is assumed to provide the desirable closed-loop characteristics. The supervisor uses simple logic rules to switch to a controller whose performance is “better” than the other controllers in the candidate set of controllers based on the performance of each controller using the online measurements available.

There is a large body of work on supervisory control. See reference [12] and references therein for an extensive overview. Early work on supervisory control used a strategy of sequentially stepping through controllers until a controller is found that stabilizes the system [11]. Switching is based on monitoring the output of the system over a moving window of finite time. The next controller is switched into the loop if the value of the output monitoring function for the second half of the monitoring window is higher than the value of that for the first half, provided that the length of the half window is sufficiently large. References [19], [20] present a model-based approach to supervisory control. The supervisor evaluates a set of performance signals that are estimates of the output error with respect to a set of candidate reference models. The supervisor then switches into the loop the controller with the best performance signal. Extension of this work for a certain class of nonlinear systems can be found in [13]. Other model-based approaches can be found in [14], [16].

An alternate approach to supervisory control is the cost based unfalsified control approach [21]. This approach utilizes the measured data to assess the performance of controllers in and out of the loop in real time. Based on these performance assessments, controllers that do not meet a prespecified desired performance condition are said to be falsified by the measured data. These controllers are rejected from the set of candidate controllers. For a cost based supervisory scheme to be effective, it is imperative that the cost should be representative of the objectives of the problem. With regard to stabilizing an unknown plant, a cost function should encode stability properties. Such reasoning motivated the notion of cost detectability [23].

In this paper, we approach supervisory control as an online decision problem. In particular, we introduce “calibrated forecasts” as a mechanism for controller selection in supervisory control.

To give some background on calibrated forecasts, suppose we sequentially measure outcomes taken from a finite set. For any stage, a forecast is a probability measure of the next outcome given the data of prior outcomes. A *calibrated* forecast [5], [9] guarantees that forecasts are “consistent” in hindsight. The following excerpt from [5] illustrates the main idea:

Suppose that, in a long (conceptually infinite) sequence of weather forecasts, we look at all those days for which the forecast probability of precipitation was, say, close to

some given value ω and (assuming these form an infinite sequence) determine the long run proportion p of such days on which the forecast event (rain) in fact occurred. The plot of p against ω is termed the forecaster's empirical calibration curve. If the curve is diagonal, $p = \omega$, the forecaster may be termed (empirically) well calibrated.

In other words, for a calibrated forecast, rain occurs on $x\%$ of all of the days (in the limit) in which the forecast was approximately $x\%$.

Interestingly, there are algorithms that are guaranteed to provide calibrated forecasts *regardless* of the underlying process generating the sequence of outcomes. This process can even be adversarial in the sense that it would like to thwart the calibration condition. Unfortunately, the computational requirements to implement these algorithms is prohibitive. An important exception is the case of binary outcomes, recently analyzed in [18].

In this paper, we will use calibrated forecasts as part of a supervisory controller. We first extend the concept of calibrated forecasting to accommodate scalar valued non-binary sequences. In this framework, the forecast is no longer interpreted as a probability, but a forecast of the sequence value. We show that the forecast for scalar sequences has similar statistical properties as conventional calibrated forecasts for binary sequences. We apply these forecasts to select a controller with the highest forecasted level of performance. In the present approach, the supervisor does not perform a pre-routed search of candidate controllers. Rather, the selection is based on online performance, with an element of randomness for "exploration". Moreover, the algorithm does not require the presence of exogenous inputs for excitation or identification. The main model-based assumption is a feasibility assumption amounting to the existence of a stabilizing controller within the set of candidate controllers. The use of calibrated forecasts will enable us to establish overall performance guarantees for the supervisory scheme even though non-stabilizing controllers may be persistently selected by the supervisor because of the effects of initial conditions, exogenous disturbances, or random selection. The main results are obtained for a general class of system dynamics and specialized to linear systems. Our approach resembles machine learning methods for so called adversarial multi-armed bandit problems [2], [7].

The remainder of this paper is organized as follows. The following Section II provides some background on calibrated forecasts and presents an extension of the calibration concept from binary signals to scalar signals. Section III presents a description of the setup, architecture of the supervisory controller. Section IV begins with a preliminary switching algorithm that employs a binary measure of controller performance and continues with the main results on calibrated forecast based supervisory switching with quantitative but imperfect measures of controller performance. Section V presents a numerical example. Finally, Section VI offers some concluding remarks and possible extensions of the proposed supervisory control algorithm.

The following notation will be used throughout.

- $|x|$ denotes the Euclidian norm of $x \in \mathbb{R}^n$.
- $|S|$ denotes the cardinality of the finite set S .

- For $y : \mathbb{R}^+ \rightarrow \mathbb{R}^n$, $\|y|_{[0,t]}\|$ denotes the usual \mathcal{L}_∞ norm

$$\|y|_{[0,t]}\| = \sup_{\tau \in [0,t]} |y(\tau)|.$$

- Similarly, $\|y|_{[t_1,t_2]}^\sigma\|$ denotes the backwards exponentially weighted \mathcal{L}_∞ norm of y over the interval $[t_1, t_2]$, i.e.,

$$\|y|_{[t_1,t_2]}^\sigma\| = \sup_{\tau \in [t_1,t_2]} e^{-\sigma(t_2-\tau)} |y(\tau)|$$

- $\|A\|$ denotes the induced norm of the matrix, A , i.e., $\|A\| = \sup_{x \neq 0} |Ax| / |x|$.
- $\mathcal{I}(z)$ denotes the indicator function

$$\mathcal{I}(z) = \begin{cases} 1, & z \text{ is TRUE;} \\ 0, & \text{otherwise.} \end{cases}$$

- $\text{rand}([-a, a])$ denotes a random number generated using a uniform probability distribution function on the interval $[-a, a]$.
- $\text{rand}(\{1, 2, \dots, N\})$ denotes a random integer generated using a uniform probability mass function on the set $\{1, 2, \dots, N\}$.
- Boldface $\mathbf{1}$ and $\mathbf{0}$ denote appropriately dimensioned vectors or matrices of ones or zeros, respectively.

II. CALIBRATED FORECASTS

A. Background: Binary Sequences

In this section we review the concept of calibrated forecasts specialized to binary sequences. The discussion follows that of [17], [18].

At every stage, $k = 0, 1, 2, \dots$, there is an outcome, $\xi(k) \in \{0, 1\}$. A forecaster observes outcomes sequentially, and at stage k , makes a forecast, $f(k) \in [0, 1]$, of the current outcome based on previously observed outcomes, $\{\xi(0), \xi(1), \dots, \xi(k-1)\}$. Note that the forecast, $f(k)$, may belong to the entire interval $[0, 1]$. Accordingly, we interpret $f(k)$ as the forecasted *probability* that $\xi(k) = 1$. In general, we allow for the possibility of *randomized* forecasts, where $f(k)$ is a non-deterministic function of the observed outcomes.

We now define criteria under which a forecasting scheme is considered to be "calibrated". For any $p \in [0, 1]$ and $\delta > 0$, the indicator function

$$\mathcal{I}(|f(k) - p| < \delta)$$

reflects when the forecast, $f(k)$, is within a specified tolerance of a specified value (probability) p . Now, define the calibration error with respect to the pair (p, δ) as

$$e_{p,\delta}(K) = \frac{1}{K+1} \sum_{k=0}^K \mathcal{I}(|f(k) - p| < \delta) (\xi(k) - f(k)). \quad (1)$$

The calibration error, $e_{p,\delta}$, compares the predicted frequency with the actual realized frequency when the prediction is δ close to p .

Definition 1: A forecasting scheme is ε -calibrated if for all outcome sequences, $\{\xi(0), \xi(1), \xi(2), \dots\}$, and all $p \in [0, 1]$ and $\delta > 0$, the calibration error satisfies

$$\limsup_{K \rightarrow \infty} |e_{p,\delta}(K)| \leq \varepsilon \quad (2)$$

almost surely.

The statement “almost surely” in the definition refers to the set of realizations of randomization during forecasting. Note that a probabilistic structure has *not* been imposed on the space of outcome sequences. A sequence is called *calibrated* if it is ε -calibrated for every $\varepsilon > 0$. Prior work [6] has shown that there does not exist a deterministic forecasting scheme that satisfies the calibration criterion for *all* outcome sequences, and so randomized forecasting is necessary.

The standard intuition behind the calibration criterion is as follows. Define

$$N(K, p, \delta) = \{k \in [0, K] : \mathcal{I}(|f(k) - p| < \delta) = 1\}.$$

In words, $N(K, p, \delta)$ denotes the set of stages where the forecast, $f(k)$, approximately equaled the specified value, p . The calibration error can be rewritten as

$$e_{p,\delta}(K) \approx \frac{|N(K, p, \delta)|}{K+1} \left(\left(\frac{1}{|N(K, p, \delta)|} \sum_{k \in N(K, p, \delta)} \xi(k) \right) - p \right).$$

(The \approx sign is because the forecast f may be slightly different than p on $N(K, p, \delta)$.) We see that there are two ways for the calibration error to vanish. First, the forecasted value of p may be asymptotically unused in the sense that

$$\limsup_{K \rightarrow \infty} \frac{|N(K, p, \delta)|}{K+1} = 0.$$

If this is not the case, then we require that for large K ,

$$\frac{1}{|N(K, p, \delta)|} \sum_{k \in N(K, p, \delta)} \xi(k) \approx p,$$

implying that the empirical frequency of the outcomes over the stages where the forecast was (approximately) p is consistent with the forecast of p .

There are forecasting algorithms (see [9], [17] and references therein) that are “universally calibrated”, i.e., the calibration condition of Definition 1 is satisfied for all sequences, even “adversarial” sequences that seek to violate the calibration condition. In the adversarial sequence setting, the universally calibrated forecast is necessarily random. That is, the algorithm computes a probability density function of the next forecast, and then randomly selects a forecast based on this density function. It is assumed that the adversarial sequence generator has knowledge of the forecasting algorithm and the data available to the algorithm. Accordingly, an adversarial sequence can compute the probability density function of the forecast but does *not* have access to the forecast itself.

For more general non-binary sequences, the computational requirements of constructing calibrated forecasts is prohibitive. In general, these algorithms require a discretization of the probability simplex and have an “internal state” associated with each

discretized element. In the binary case, this amounts to a scalar state associated with each p_i in the range $0 \leq p_1 \leq p_2 \leq \dots \leq p_n \leq 1$. The number of requisite discretization values increases as the desired accuracy ε decreases.

Recent work [18] has analyzed a forecasting algorithm (suggested by [8]) that circumvents the requirement of discretization. The algorithm, called “tracking forecasts” in [18], produces calibrated forecasts for all binary sequences and special classes of non-binary sequences. While not being “universal”, its computational and memory requirements are very modest, being on par with taking a running average of a sequence of values.

The tracking forecast is defined as follows. First, define the projection $\Pi_{[0,1]} : \mathbb{R} \rightarrow [0, 1]$ as

$$\Pi_{[0,1]}[s] = \arg \min_{p \in [0,1]} |s - p|.$$

Let η be a positive constant, and let $0 < \rho < 1$. Define

$$\tilde{f}(k+1) = \tilde{f}(k) + \left(\frac{1}{k+1} \right)^\rho (\xi(k) - \tilde{f}(k)) \quad (3a)$$

$$f(k+1) = \Pi_{[0,1]}[\tilde{f}(k+1) + \text{rand}([- \eta, \eta])]. \quad (3b)$$

The quantity $\tilde{f}(k)$ can be interpreted as a weighted average of $\xi(0), \dots, \xi(k-1)$ but with an increased weighting on recent outcomes. Indeed, if $\rho = 1$, then $\tilde{f}(k)$ is simply the average of $\xi(0), \dots, \xi(k-1)$. The forecast, $f(k)$, consists of a randomized perturbation of $\tilde{f}(k)$.

Proposition 1 ([18]): For any $\varepsilon > 0$ there exists an $\eta > 0$ such that the tracking forecast, $f(k)$, defined by (3) is ε -calibrated for all binary sequences.

B. Non-Binary Sequences

In this paper, we will need to broaden the set of allowable outcome sequences to non-binary sequences. In particular, we are interested in scalar sequences of the form

$$\hat{\xi}(k) = \xi(k) + \nu(k), \quad k = 0, 1, 2, \dots, \quad (4)$$

where $\nu(k)$ is a random independent zero-mean finite-variance sequence, and $\xi(\cdot)$ is a bounded sequence with $\xi(k) \in [\xi_{\min}, \xi_{\max}]$. We shall refer to sequences (4) as *admissible noisy bounded scalar non-binary sequences*. We are also interested in the special case when $\nu(k) = 0, \forall k = 0, 1, 2, \dots$, we shall refer to such sequences as *admissible bounded scalar non-binary sequences*.

The calibration error for any $f^* \in [\xi_{\min}, \xi_{\max}]$ and δ is defined as

$$e_{f^*,\delta}(K) = \frac{1}{K+1} \sum_{k=0}^K \mathcal{I}(|f(k) - f^*| < \delta) (\xi(k) - f(k)). \quad (5)$$

The following definition parallels Definition 1.

Definition 2: A forecasting scheme is ε -calibrated if for all sequences (4), all $f^* \in [\xi_{\min}, \xi_{\max}]$, and $\delta > 0$, the calibration error (5) satisfies

$$\limsup_{K \rightarrow \infty} |e_{f^*,\delta}(K)| \leq \varepsilon \quad (6)$$

almost surely.

Even though the reward is not binary, the tracking forecast assures a property similar to calibration, but with a different interpretation. As before, we will define the tracking forecast, but now without any projection to the unit interval.

$$\tilde{f}(k+1) = \tilde{f}(k) + \left(\frac{1}{k+1}\right)^\rho (\hat{\xi}(k) - \tilde{f}(k)) \quad (7a)$$

$$f(k+1) = \tilde{f}(k+1) + \text{rand}([- \eta, \eta]). \quad (7b)$$

The forecast, $f(k)$, is no longer interpreted as a probability. Rather, the calibration condition now implies the following. For any specified forecast value, e.g., f^* , the average of the outcome sequence on the stages that the forecast was (approximately) f^* is approximately equal to f^* .¹

Proposition 2: For any $\varepsilon > 0$ there exists an $\eta > 0$ such that the tracking forecast, $f(k)$, defined by (7) is ε -calibrated for the following classes of sequences

- 1) All admissible bounded scalar non-binary sequences for any $0 < \rho < 1$;
- 2) All admissible noisy bounded scalar non-binary sequences for $1/2 < \rho < 1$.

Proof: The proof for the first class of sequences parallels the binary sequences case in [18]. Alternatively, the proof can be considered as a special case of the proof of the second class of sequences. The proof for the second class of sequences is provided in the Appendix. ■

III. PLANT ASSUMPTIONS AND SWITCHING CONTROLLER STRUCTURE

A. Setup

Our objective is to control an unknown plant which we represent by the nonlinear system

$$\dot{x} = g(x, u, w), \quad x(0) = x_o \quad (8a)$$

$$y = h(x, u), \quad (8b)$$

where $x_o \in \mathbb{R}^n$ is a fixed (but unknown) initial condition, $w(\cdot) \in \mathcal{L}_\infty$ is a fixed (but unknown) exogenous disturbance, and $u(\cdot)$ is the control input.

It will be evident that the structure of a finite-dimensional plant will not be essential. For example, an input-output plant model along with suitably modified versions of the forthcoming assumptions will also result in the desired stability properties of the supervisory controller. We avoid such generality here in favor of clarity of exposition.

The switching controller has the form

$$u(t) = F_{\alpha(t)}(y(t)), \quad \alpha(t) \in \{1, 2, \dots, N_c\}, \quad (9)$$

i.e., the controller switches among a finite set of static output control laws, $F_i : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_u}$. This setup also encompasses switching among dynamic output control laws (cf., the forthcoming subsection on specialization to linear systems). We will call a switching signal $\alpha : [0, \infty) \rightarrow \{1, 2, \dots, N_c\}$ *admissible* if it is piecewise constant with a minimum dwell time, $\tau_{\min} > 0$. That is for consecutive switching times $t_a < t_b$,

$$t_b - t_a \geq \tau_{\min}.$$

¹Or the forecasted value f^* is asymptotically unused.

We assume that the closed-loop equations

$$\dot{x} = g(x, u, w), \quad x(0) = x_o \quad u(t) = F_{\alpha(t)}(y(t))$$

are well posed, i.e., for any admissible switching signal, there exists a unique solution over $[0, \infty)$.

The following are the main assumptions on the plant and candidate control laws. Define a *finitely switching* control input as

$$u_{\text{fs}}(t) = \begin{cases} F_{\alpha(t)}(y(t)), & 0 \leq t < t_o; \\ F_i(y(t)), & t \geq t_o, \end{cases} \quad (10)$$

for some $t_o > 0$.

Assumption 1: There exist continuous strictly increasing functions $d_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and $d_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ and constant $\sigma > 0$, such that for any finitely switching input (10) and any $\Delta T > 0$,

$$\left\| |y|_{[0, t_o + \Delta T]}^\sigma \right\| \leq d_1(\Delta T) \left\| |y|_{[0, t_o]}^\sigma \right\| + d_2(\Delta T).$$

Assumption 2: There exist a control law, F_i^* , and positive constants σ and ℓ^* , such that for any $0 < \gamma < 1$, there exists a $\Delta T^* \geq 0$ that satisfies the following condition. Let $\sigma > 0$ be as in Assumption 1. For any finitely switching control input (10) with $F_i = F_i^*$,

$$\left\| |y|_{[0, t_o + \Delta T]}^\sigma \right\| \leq \gamma \left\| |y|_{[0, t_o]}^\sigma \right\| + \ell^*,$$

for all $\Delta T \geq \Delta T^*$.

Assumption 1 provides an upper bound on the growth rate of solutions over all finite histories and candidate controllers. Assumption 2 states that there is at least one controller that satisfies a desired stabilization condition.

B. Specialization to Linear Systems

This section shows explicitly how Assumptions 1–2 can be satisfied in the special case of linear systems. To this end, consider the linear setup

$$\dot{x} = Ax + Bu + w, \quad x(0) = x_o \quad (11a)$$

$$y = \begin{pmatrix} Cx \\ u \end{pmatrix} \quad (11b)$$

$$u = (-F_{\alpha(t)} \quad \mathbf{0})y \quad (11c)$$

$$F_{\alpha(t)} \in \{F_1, F_2, \dots, F_{N_c}\}. \quad (11d)$$

The control input switches between N_c linear static-output feedback controllers. The plant output has been augmented to include the control input. However, we assume that there is no direct algebraic loop. That is, the control input can be rewritten as

$$u = -F_{\alpha(t)}Cx.$$

In the case of switching between dynamic output feedback controllers, simple manipulations can convert the system into the static output feedback form considered above. Consider a dynamically controlled system

$$\dot{x}_p = A_p x_p + B_p u + w$$

$$\dot{x}_c = A_c x_c + B_c y_p$$

$$y_p = C_p x_p$$

$$u = D_c y_p + C_c x_c,$$

where x_p is the plant state, x_c is the controller state, y_p is an output measurement, and (A_c, B_c, D_c, C_c) are matrices associated with a fixed dynamic controller. By augmenting the state, output, and control, we can rewrite this system as

$$\begin{aligned} \begin{pmatrix} \dot{x}_p \\ \dot{x}_c \end{pmatrix} &= \begin{pmatrix} A_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} x_p \\ x_c \end{pmatrix} + \begin{pmatrix} B_p & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} u_{\text{aug}} + \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix} w \\ y_{\text{aug}} &= \begin{pmatrix} C_p & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} x_p \\ x_c \end{pmatrix} \\ u_{\text{aug}} &= \begin{pmatrix} D_c & C_c \\ B_c & A_c \end{pmatrix} y_{\text{aug}}. \end{aligned}$$

In this form, switching between dynamic controllers means switching the “static output feedback” matrix $\begin{pmatrix} D_c & C_c \\ B_c & A_c \end{pmatrix}$ which only depends on controller matrices. The only restriction is that all dynamic controllers in this setup must be of the same order.

We now show that under natural assumptions, the linear system (11) will satisfy Assumptions 1–2.

Assumption 3: The switched linear system (11) satisfies:

- $A - BF_i C$ is a stability matrix for some $i \in \{1, 2, \dots, N_c\}$.
- The pair $[A, C]$ is observable.

Proposition 3: Under Assumption 3, the linear system (11) satisfies Assumptions 1–2.

The proof of Proposition 3 relies on standard bounding arguments and is presented in the appendix.

IV. SUPERVISORY ALGORITHM AND MAIN RESULTS

A. Falsification Based Supervisory Switching

For the sake of contrast, we begin with a switching algorithm that is based on the notion controller falsification (as in [21]). This algorithm does not employ calibrated forecasts. Rather, the algorithm exploits Assumption 2 to sequentially test whether a candidate controller satisfies a specified stability condition. In case all controllers fail, the stability condition is relaxed and the testing process is repeated. Assumption 2 assures that a controller will emerge that does not fail the (sufficiently relaxed) stability condition. Assumption 1 assures that all signals remain bounded in the process.

In order to precisely define this and forthcoming supervisory algorithms, let

- $t_k, k = 0, 1, 2, \dots$, denote the switching times, with $t_0 = 0$;
- $I_k = [t_{k-1}, t_k)$ denote the k th interval; and
- $i(k)$ denote the control law implemented over the k th interval, i.e.,

$$u(t) = F_{i(k)}(y(t)), \quad t_{k-1} \leq t < t_k.$$

In terms of the preceding discussion, at time t_k the supervisory algorithm will determine 1) the duration of the next interval $I_{k+1} = [t_k, t_{k+1})$ and 2) which control law $i(k+1) \in \{1, 2, \dots, N_c\}$ to implement over I_{k+1} .

After each interval, I_k , the most recently implemented control law, $i(k)$, is evaluated according to a binary reward:

$$r(k) = \begin{cases} 1, & \|y|_{[0, t_k]}\|^\sigma \leq \gamma \|y|_{[0, t_{k-1}]}\|^\sigma + \ell(k); \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

where $\sigma > 0$ is as in Assumptions 1–2, γ is a fixed positive scalar with $\gamma < 1$, and $\ell(k)$ is yet to be specified. This reward reflects a stability condition that bounds the growth of the size of the output over the interval I_k relative to the size of the output prior to the interval I_k .

The following are fixed parameters of the supervisory algorithms:

- T_{inc} := the increment by which $\Delta T(k)$ is increased;
- ℓ_{inc} := the increment by which $\ell(k)$ is increased;
- γ := a positive scalar with $\gamma < 1$;
- σ := a positive scalar satisfying Assumptions 1–2;
- τ_{min} := the length of the initial interval I_1 ; and

The following scalars will be updated with each interval, I_k :

- $\Delta T(k)$:= the length of interval I_k ; and
- $\ell(k)$:= the scalar offset used in the reward function (12).

Likewise, the following vector of dimension N_c will be updated with each interval, I_k : $\phi(k)$, which is defined as follows:

- $\phi_i(k)$:= the lowest value of the realized reward (12) when the i th control law was implemented *since the last increase* in interval lengths.

In the following description, we adopt the convention that any k -dependent variable is held constant, e.g., $X(k+1) = X(k)$ for the “generic” variable $X(\cdot)$, unless specified otherwise.

Algorithm 1 (Falsification Based Switching)

• Initialization:

- Select $\gamma < 1$, τ_{min} , T_{inc} , ℓ_0 , ℓ_{inc} , and (small) σ .
- Set $k = 1$, $N(0) = \mathbf{0}$, $\phi(0) = \mathbf{1}$, $\Delta T(1) = \tau_{\text{min}}$, $t_1 = \tau_{\text{min}}$, and $\ell(1) = \ell_0$.
- Select $i(1) = \text{rand}(\{1, 2, \dots, N_c\})$.

• Evaluation of Control Law $i(k)$: At time t_k , set

$$r(k) = \begin{cases} 1, & \|y|_{[0, t_k]}\|^\sigma \leq \gamma \|y|_{[0, t_{k-1}]}\|^\sigma + \ell(k); \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

• Parameter Update: At time t_k , set

- $\phi_{i(k)}(k) = \min(\phi_{i(k)}(k), r(k))$
- If $\max_j(\phi_j(k)) = 0$,
 - * $\Delta T(k+1) = \Delta T(k) + T_{\text{inc}}$
 - * $\ell(k+1) = \ell(k) + \ell_{\text{inc}}$
 - * $\phi(k) = \mathbf{1}$
- $t_{k+1} = t_k + \Delta T(k+1)$.

• Control Law Selection: Select any $i(k+1)$ satisfying

$$\phi_{i(k+1)}(k) = 1.$$

• Loop: Update $k \leftarrow k+1$ and repeat.

In terms of the falsification concept, whenever the reward $r(k) = 0$, the control law $i(k)$ did not pass the assumed stability criterion test and hence has been falsified. If all control laws fail the stability criterion test, then the test is made less

stringent and the process is repeated. Eventually, a control law will emerge that never fails the stability criterion test.

Proposition 4: Under Assumptions 1–2 on the plant model (8), Algorithm 1 results in y bounded, i.e., $y \in \mathcal{L}_\infty$.

The proof of Proposition 4 is based on the following claim.

1) *Claim 1:* The parameters $\Delta T(k)$ and $\ell(k)$ are uniformly bounded, i.e.,

$$\begin{aligned}\Delta T_{\max} &= \lim_{k \rightarrow \infty} \Delta T(k) < \infty \\ \ell_{\max} &= \lim_{k \rightarrow \infty} \ell(k) < \infty.\end{aligned}$$

Proof: *Claim 1* The parameters $\Delta T(k)$ and $\ell(k)$ are increased whenever the condition

$$\max_j(\phi_j(k)) = 0 \quad (14)$$

holds. In other words, every control law in its most recent utilization resulted in a reward of zero. However, by Assumption 2, there exists a $\Delta T^* \geq 0$, a positive constant ℓ^* , and at least one control law, i^* , which satisfies the condition

$$\|y|_{[0,t_k]}^\sigma\| \leq \gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell^*,$$

provided that $\Delta T(k) \geq \Delta T^*$. This implies that the condition (14) cannot be satisfied infinitely often with $\Delta T(k)$ and $\ell(k)$ increasing without bound. ■

Proof: (*Proposition 4*) *Claim 1* implies that eventually (i.e., for all $k > k^*$ for some k^*), the bound on $\|y|_{[0,t_k]}^\sigma\|$ will satisfy the condition

$$\|y|_{[0,t_k]}^\sigma\| \leq \gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell_{\max}. \quad (15)$$

This condition implies the desired result. ■

B. Calibrated Forecast Based Supervisory Switching

We now introduce a supervisory switching algorithm based on calibrated forecasts. This algorithm differs from the falsification based algorithm in two important aspects:

- *Quantitative assessment:* Algorithm 1 uses a binary indicator of whether or not a controller meets the desired stability criterion test (15). The algorithm does not distinguish between gross failure or slight failure. The forthcoming framework takes into account a quantitative assessment of performance in the controller selection process.
- *Falsification uncertainty:* The setup for Algorithm 1 allowed a control law to be falsified with certainty with a single implementation. The forthcoming framework introduces imperfect performance assessments, thereby eliminating certain falsification.

Both of these issues will be addressed by evaluating an implemented control law according to the non-binary reward

$$r(k) = -\frac{\|y|_{[0,t_k]}^\sigma\|}{\gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell(k)} + \nu(k), \quad (16)$$

where the $\nu(k)$ are a zero-mean finite-variance sequence. Assume for now that $\nu(k) = 0$. In terms of the preceding discussion, the binary reward in (12) is replaced by the *ratio*

$$-\frac{\|y|_{[0,t_k]}^\sigma\|}{\gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell(k)}.$$

This ratio is greater than -1 if and only if

$$\|y|_{[0,t_k]}^\sigma\| \leq \gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell(k).$$

Accordingly, a controller passes the stability criterion test whenever the ratio is sufficiently large. Like the binary reward, a controller can be falsified on the basis of this ratio. Unlike the binary reward, the ratio takes into account a quantitative measure of performance.

Now suppose that the $\nu(k)$ are non-zero. In this case, one can no longer falsify a controller based on a single implementation. The inclusion of $\nu(k)$ is motivated by eliminating the possibility of controller falsification. However, it does not necessarily correspond to “measurement noise” on the measured output variable, y .

For clarity of exposition, we will first present a supervisory switching algorithm for the special case of $\nu(k) = 0$, followed by the general case of $\nu(k) \neq 0$.

1) *Non-Binary Rewards With Perfect Assessment:* The following are new variables to be employed in the calibration based switching algorithm:

- $N_i(k)$:= the number of times the i th control law has been implemented up to and including interval I_k ;
- $\tilde{R}_i(k)$:= a tracking forecast of the reward of the i th control law; and
- ρ & η := calibration parameters, with $0 < \rho < 1$ and η satisfying Proposition 1.

Algorithm 2 (Calibrated Forecast Based Switching)

• Initialization:

- Select $0 < \rho < 1$, $\gamma < 1$, τ_{\min} , T_{inc} , ℓ_0 , ℓ_{inc} , (small) σ , and (sufficiently small) $\tilde{\eta}$.
- Set $k = 1$, $N(0) = \mathbf{0}$, $\tilde{R}(1) = \mathbf{0}$, $\phi(0) = \mathbf{1}$, $\Delta T(1) = \tau_{\min}$, $t_1 = \tau_{\min}$, and $\ell(1) = \ell_0$.
- Select $i(1) = \text{rand}(\{1, 2, \dots, N_c\})$.

• Evaluation of Control Law $i(k)$: At time t_k , set

$$r(k) = -\frac{\|y|_{[0,t_k]}^\sigma\|}{\gamma \|y|_{[0,t_{k-1}]}^\sigma\| + \ell(k)} \quad (17)$$

• Parameter Update: At time t_k , set

- $N_{i(k)}(k) = N_{i(k)}(k-1) + 1$
- $\tilde{R}_{i(k)}(k+1) = \tilde{R}_{i(k)}(k) + (1/N_{i(k)}(k))^\rho (r(k) - \tilde{R}_{i(k)}(k))$
- $\tilde{R}_j(k+1) = \tilde{R}_j(k+1) + \text{rand}([- \eta, \eta])$, for all $j = 1, 2, \dots, N_c$
- If $r(k) < -1$, set $\phi_{i(k)}(k) = 0$.
- If $\max_j(\phi_j(k)) = 0$,
 - * $\Delta T(k+1) = \Delta T(k) + T_{\text{inc}}$
 - * $\ell(k+1) = \ell(k) + \ell_{\text{inc}}$

- * $\phi(k) = 1$
- $t_{k+1} = t_k + \Delta T(k+1)$.
- **Control Law Selection:**
 - With probability $1/k + 1$, select $i(k+1) = \text{rand}(\{1, 2, \dots, N_c\})$
 - With probability $1 - 1/k + 1$, select

$$i(k+1) = \arg \max_{j \in \{1, 2, \dots, N_c\}} R_j(k+1)$$
- **Loop:** Update $k \leftarrow k+1$ and repeat.

The following distinctions between Algorithms 1 and 2 are worth noting:

- Control law selection is based on a calibrated forecast of the performance (reward) of a control law.
- There is probabilistic exploration through random selection of control laws, but with diminishing probability.
- Implemented control laws are *not* limited to controllers that have yet to fail for current values of $\Delta T(\cdot)$ and $\ell(\cdot)$. As opposed to sequential implementation, a control law may be implemented that fails the stability criterion but still delivers the best performance.

Theorem 1: Under Assumptions 1 and 2 on the plant model (8), Algorithm 2 results in y bounded, i.e., $y \in \mathcal{L}_\infty$, with probability 1, for sufficiently small calibration parameters ε and η .

The remainder of this subsection is devoted to the proof of Theorem 1.

— *Bounded rewards:*

The reward is bounded from above by zero. Because of the update rule for $\phi(k)$, Claim 1 still holds. Therefore, the reward sequence is bounded from below as well via Assumption 1.

— *Calibrated forecasts:*

Since the rewards are bounded, $R_j(k)$ is a calibrated forecast for the performance of the j th feedback law for all $j \in \{1, 2, \dots, N_c\}$ (via Proposition 2).

— *Feasible controller:*

Claim 2 For some $i_o \in \{1, 2, \dots, N_c\}$,

$$\lim_{k \rightarrow \infty} \tilde{R}_{i_o}(k) \geq -1.$$

Proof: Following Claim 1, condition (14) will be satisfied only a finite number of times. According to the update rule of $\phi(k)$ in Algorithm 2, this means that for some i_o , the component $\phi_{i_o}(k) = 1$ for all but a finite number of $k \geq 1$. Stated differently, for sufficiently large k , if $\alpha(k) = i_o$ then $r(k) \geq -1$. Finally, the randomization in the control law selection step assures that every control law is implemented infinitely often. In particular, $r(k) \geq -1$ for all but a finite number of $k \geq 1$ when implementing control law $F_{i_o}(\cdot)$, which implies the desired result. ■

Note that the control law index i_o in Claim 2 need *not* be the stabilizing control law index i^* in Assumption 2. Because of the effects of initial conditions and exogenous disturbances, a stabilizing control law need not “reveal itself”.

— *Overall average reward:*

Claim 3: For any $\bar{\varepsilon} > 0$,

$$\liminf_{k \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K R(k) \geq -1 - \bar{\varepsilon}$$

for sufficiently small calibration parameters ε and η .

Proof: It will be useful to distinguish the intervals in which the controller selection was based on a random selection or on its forecasted performance. We will refer to the former as “exploration” and the latter as “exploitation”.

We will present the proof for the simpler case of $N_c = 2$ in detail. The general case follows from similar arguments.

For $N_c = 2$, assume that control law $F_1(\cdot)$ satisfies the condition of Claim 2, i.e.,

$$\lim_{k \rightarrow \infty} \tilde{R}_{i_o}(k) \geq -1. \quad (18)$$

Let $\{I_1, I_2, \dots, I_K\}$ be a succession of switching intervals. We will divide these intervals into different categories:

$$\begin{aligned} \mathcal{G}_K &= \{k : i(k) = 1\} \\ \mathcal{B}_K &= \{k : i(k) = 2 \text{ under exploitation}\} \\ \mathcal{E}_K &= \{k : i(k) = 2 \text{ under exploration}\} \end{aligned}$$

These sets represent using the “good” control law, the “bad” control law under exploitation, or the bad control law under exploration.

We can average the rewards over these intervals using

$$\begin{aligned} \frac{1}{K} \sum_{k=1}^K r(k) &= \frac{1}{\mathcal{G}_K + \mathcal{B}_K + \mathcal{E}_K} \\ &\left(\sum_{k \in \mathcal{G}_K} r(k) + \sum_{k \in \mathcal{B}_K} r(k) + \sum_{k \in \mathcal{E}_K} r(k) \right). \quad (19) \end{aligned}$$

By definition of \mathcal{G}_K ,

$$\begin{aligned} \liminf_{K \rightarrow \infty} \frac{1}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} \sum_{k \in \mathcal{G}_K} r(k) \\ \geq \liminf_{K \rightarrow \infty} \frac{\mathcal{G}_K}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} (-1 - \eta). \end{aligned}$$

Likewise, with probability one,

$$\liminf_{K \rightarrow \infty} \frac{1}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} \sum_{k \in \mathcal{E}_K} r(k) = 0,$$

because exploration occurs with vanishing probability.

It remains to analyze

$$\liminf_{K \rightarrow \infty} \frac{1}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} \sum_{k \in \mathcal{B}_K} r(k).$$

Whenever $k \in \mathcal{B}_K$, the reward forecasted for control law $i = 2$ exceeded the reward forecasted for control law $i = 1$. From (18),

$$\liminf_{k \rightarrow \infty} R_1(k) \geq -1 - \eta.$$

Accordingly, for sufficiently large k ,

$$k \in \mathcal{B}_K \Rightarrow R_2(k) \geq -1 - 2\eta.$$

In terms of an indicator function, there exists an f^* and δ such that for sufficiently large k ,

$$k \in \mathcal{B}_K \Rightarrow \mathcal{I}(|R_2(k) - f^*| \leq \delta) = 1,$$

where $f^* \geq -1 - 2\eta$. This means we can rewrite

$$\begin{aligned} & \liminf_{K \rightarrow \infty} \frac{1}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} \sum_{k \in \mathcal{B}_K} r(k) \\ &= \frac{|\mathcal{B}_K|}{|\mathcal{G}_K| + |\mathcal{B}_K| + |\mathcal{E}_K|} \\ & \left(\frac{1}{|\mathcal{B}_K|} \sum_{k \in \mathcal{B}_K} \mathcal{I}(|R_2(k) - f^*| \leq \delta) r(k) \right). \end{aligned}$$

We are now in a position to take advantage of the calibration property. The calibration (1)–(2) together imply that

$$\frac{1}{|\mathcal{B}_K|} \sum_{k \in \mathcal{B}_K} \mathcal{I}(|R_2(k) - f^*| \leq \delta) r(k) \geq -1 - 2\eta - \varepsilon.$$

Combining the lower bounds for the categories of \mathcal{G}_K , \mathcal{B}_k , and \mathcal{E}_K , we can lower bound the above overall average reward (19) by

$$\liminf_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K r(k) \geq -1 - 2\eta - \varepsilon.$$

Both η and ε are calibration parameters that can be chosen to be sufficiently small, which proves the desired result.

The proof is similar in the general case of $N_c > 2$. Again, let us suppose that the control law $F_1(\cdot)$ satisfies the conditions of Claim 2. The only modification is to subdivide the sets \mathcal{B}_K and \mathcal{E}_K according to control laws $i = 2, \dots, N_c$. ■

— *Bounded output:*

The overall average reward property can be rewritten as

$$\limsup_{k \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \frac{\|y|_{[0, t_k]}^\sigma\|}{\gamma \|y|_{[0, t_{k-1}]}^\sigma\| + \ell(k)} \leq 1 + \bar{\varepsilon}.$$

Introduce the shorthand notation

$$\begin{aligned} z(k) &= \|y|_{[0, t_k]}^\sigma\| \\ a(k) &= -r(k)\gamma \end{aligned}$$

Then using the definition of the reward,

$$z(k) = a(k)z(k-1) - r(k)\ell(k).$$

The input/output stability of this system, and accordingly the boundedness of $z(k)$, is determined by the products $\prod_{k=1}^K a(k)$ for large K . In particular if these terms satisfy

$$\limsup_{K \rightarrow \infty} \sum_{j=1}^K \left(\prod_{k=j}^K a(k) \right) < \infty, \quad (20)$$

then the $z(k)$ will be bounded. Since $a(k) \geq 0$, we can use the arithmetic/geometric mean inequality,

$$\prod_{k=1}^K a(k) \leq \left(\frac{1}{K} \sum_{k=1}^K a(k) \right)^K$$

to bound the product terms. Therefore, for sufficiently large K ,

$$\prod_{k=1}^K -r(k) \leq (1 + \bar{\varepsilon})^K$$

which implies that

$$\prod_{k=1}^K a(k) \leq ((1 + \bar{\varepsilon})\gamma)^K.$$

As long as $(1 + \bar{\varepsilon})\gamma < 1$, the dynamics describing the evolution of the $z(k)$ will be input/output stable (cf., (20)). These dynamics are driven by the sequence $r(k)\ell(k)$, which is uniformly bounded. Therefore, $z(k) = \|y|_{[0, t_k]}^\sigma\|$ is uniformly bounded for all $k = 0, 1, 2, \dots$

2) *Non-Binary Rewards With Imperfect Assessment:* In this part, we address the case when performance assessment is corrupted by independent zero-mean finite variance (possibly unbounded) sequence $\nu(k)$.

As previously discussed, in this setup a single measurement does not provide sufficient information on whether a given controller satisfies the desired stabilization condition. Therefore it is not possible to rule out or falsify a controller based on any single measurement. However, a calibrated forecast provides a measure of the consistency of satisfying the stabilization condition.

We argue that simple changes to the supervisor's algorithm can provide the same performance guarantees on the output as these in the perfect measurements case. To that end, consider the following changes to Algorithm 2:

- **Evaluation of Control Law $i(k)$:** At time t_k , set

$$r(k) = -\frac{\|y|_{[0, t_k]}^\sigma\|}{\gamma \|y|_{[0, t_{k-1}]}^\sigma\| + \ell(k)} + \nu(k)$$

- **Parameter Update:** At time t_k , set
 - $R_j(k+1) = \hat{R}_j(k+1) + \text{rand}([- \eta, \eta])$, for all $j = 1, 2, \dots, N_c$
 - If $r(k) < -1$, set $\phi_{i(k)}(k) = 0$.
 - If $\max_j(\phi_j(k)) = 0$, set $\phi(k) = \mathbf{1}$. **Moreover**, if $\max_j(\phi_j(k)) = 0$ AND $\max_j \hat{R}_j(k) < -1 - \varepsilon$, set
 - * $\Delta T(k+1) = \Delta T(k) + T_{\text{inc}}$
 - * $\ell(k+1) = \ell(k) + \ell_{\text{inc}}$.

The remaining parts of the algorithm are unchanged. For this setup, the tracking forecast defined in (7) will be used to track each controller's performance.

Theorem 2: Under Assumptions 1 and 2 on the plant model (8), the above modifications to Algorithm 2 results in y bounded, i.e., $y \in \mathcal{L}_\infty$, with probability 1, for sufficiently small calibration parameters ε and η and with $1/2 < \rho < 1$.

The proof of Theorem 2 is analogous to the proof of Theorem 1 in the preceding subsection. The changes in the algorithm, along with the ensured calibration, guarantee that Claim 1 holds. The remaining parts of the proof parallel the ones in the preceding subsection. Though, it is important to note that in order to include the effects of imperfect rewards, we make use of the following:

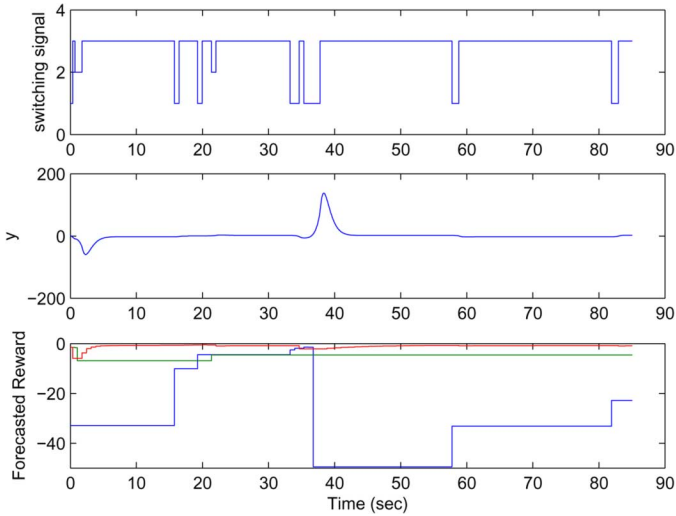


Fig. 1. Non-Binary Rewards: Perfect Assessment—Switching Signal, output, and average reward for the example considered with adversarial disturbance and model mismatch.

$$\begin{aligned} & \liminf_{K \rightarrow \infty} \sum_{k=1}^K r(k) \\ &= \liminf_{K \rightarrow \infty} \left(\frac{1}{K} \sum_{k=1}^K - \frac{\|y|_{[0,k]}^\sigma\|}{\gamma \|y|_{[0,k-1]}^\sigma\| + \ell(k)} + \nu(k) \right) \\ &\approx \liminf_{K \rightarrow \infty} \left(\frac{1}{K} \sum_{k=1}^K - \frac{\|y|_{[0,k]}^\sigma\|}{\gamma \|y|_{[0,k-1]}^\sigma\| + \ell(k)} \right) \pm \epsilon_\nu, \end{aligned}$$

where $\epsilon_\nu > 0$ is arbitrarily small. Moreover, arguments similar to those for Claim 3 assure that for any $\bar{\epsilon}$,

$$\liminf_{K \rightarrow \infty} \sum_{k=1}^K r(k) \geq -1 - \bar{\epsilon}.$$

Therefore, it follows that

$$\limsup_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K \frac{\|y|_{[0,k]}^\sigma\|}{\gamma \|y|_{[0,k-1]}^\sigma\| + \ell(k)} \leq 1 + \bar{\epsilon} + \epsilon_\nu.$$

The remaining steps in the proof are identical to those in the preceding subsection.

V. ILLUSTRATIVE SIMULATIONS

In this section, we provide an illustrative numerical example. We shall consider the unstable nonminimum phase system

$$P(s) = 0.3 \frac{s-10}{(s-1)(s+3)},$$

which has the state space representation

$$\begin{aligned} \dot{x} &= \begin{pmatrix} -2 & 3 \\ 1 & 0 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u + w; \\ y &= (0.3 \quad -3) x. \end{aligned}$$

The set of candidate controllers is composed of three dynamic output feedback controllers. The dynamics of the controllers are given by

$$\begin{aligned} \dot{x}_c &= \underbrace{\begin{pmatrix} 19.1026 & -208.0256 \\ 6.4103 & -54.1026 \end{pmatrix}}_{A_c} x_c + \underbrace{\begin{pmatrix} -70.3419 \\ -18.0342 \end{pmatrix}}_{B_c} y \\ u &= C_c^i x_c, \end{aligned}$$

where $C_c^1 = (-4 \quad 4)$, $C_c^2 = (1 \quad -5)$, and $C_c^3 = (1 \quad 5)$. The controller with C_c^3 is the only stabilizing controller. The simulations incorporate model mismatch in the form of two high frequency poles at $s = -35$. The transfer function of the resulting system is

$$P(s) = 0.3 \frac{s-10}{(s-1)(s+3)} \left(\frac{35}{s+35} \right)^2.$$

Moreover, an *adversarial* disturbance is constructed in the following manner:

$$w(t) = \begin{cases} -\text{sat}(Bu(t)), & i = 3; \\ 0, & \text{otherwise,} \end{cases}$$

where $\text{sat}(\cdot)$ is the saturation function. In words, the disturbance cancels the control action of the stabilizing controller whenever it is used. However, this cancelation is saturated so that the disturbance is bounded.

Case 1: Non-Binary Rewards with Perfect Assessment

This case uses the non-binary reward case with perfect assessment presented in Section IV-B1. The simulations were carried out for the system with disturbance and model mismatch as presented above. The simulation results are shown in Fig. 1.

Case 2: Non-Binary Rewards with Imperfect Assessment

This case uses the non-binary reward case with noisy measurements presented in Section IV-B2. The simulations were carried out for the system with disturbance and model mismatch as presented above. The assessment error was generated using a zero-mean Gaussian distribution with a variance equal to 4. The simulation results are shown in Fig. 2.

The algorithm parameters that were used are as follows. $\rho = 0.5$, $\sigma = 0.5$, $\gamma = 0.999$, $\tau_{\min} = 0.35$, $T_{\text{inc}} = 0.35$, $\ell_0 = 0.4$, $\ell_{\text{inc}} = 0.3$, $\eta = 1e^{-10}$. For Case 4, $\rho = 0.51$ was used. The (unknown) initial condition of the system is $x(0) = (0.2 \quad 0.1 \quad 0 \quad 0)^T$. The exploration probability in the algorithm was set to $(1/k + 1)^{0.75}$, where k is the stage index.

VI. CONCLUSION

We have introduced a calibrated forecasts approach to switching supervisory control. We showed that using a calibrated forecast strategy along with the proposed supervisor has minimal informational and structural assumptions on the unknown plant while guaranteeing that the output of the system will remain bounded in the presence of bounded disturbances — even if the disturbances are adversarial. The results clearly indicate the potential for use of online algorithms in switching control and opens the possibility of applying other online learning algorithms to adaptive control.

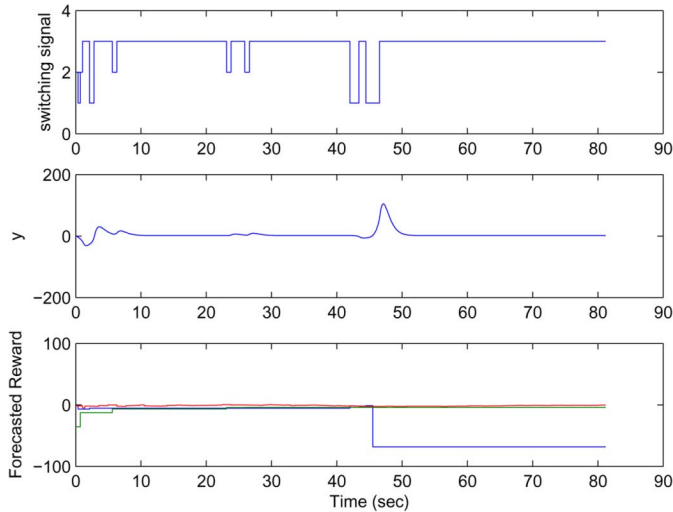


Fig. 2. Non-Binary Rewards: Imperfect Assessment—Switching Signal, output, and average reward for the example considered with adversarial disturbance and model mismatch.

It is worthwhile noting that the stability criterion employed in the paper does reflect performance in terms of disturbance rejection. For example, the values of γ and ℓ^* could distinguish two stabilizing controllers in that a superior controller will satisfy the stability condition in Assumption 2 “sooner”, i.e., for smaller values of $\Delta T(k)$ and $\ell(k)$ than an inferior controller.

The main assumption on the system is the existence of a stabilizing controller within the set of candidate controllers. The stability results do not require knowledge of certain system parameters or signal bounds or the use of external inputs or excitation for identification purposes. Moreover, the setup avoids going through a prerouted search of the action space. Rather, the algorithm reinforces positive learned behavior by selecting the control laws with the highest forecasted reward. For contrast, we also presented a simpler falsification based algorithm that searches through controllers in an attempt to discover a feasible controller. However, this approach ignores the accumulated information on control law performance. Furthermore, this approach would not work in the presence of imperfect reward measurements because a controller can never be “falsified” for a fixed $\Delta T(k)$ and $\ell(k)$ by a single measurement. However, calibrated forecasts-based switching still can be used in the imperfect reward scenario.

The present algorithm does not use estimation of the behavior of control laws are not in the loop. Hence, our supervisor updates its forecast of control law performance one at a time. We conjecture that incorporating assessments of non-implemented control laws, as in unfalsified control [21], could help to further improve transient responses in the present approach. One caveat is that the estimate of non-implemented control law performance may be conservative or pessimistic, and this could adversely affect transient behavior. Finally, another interesting extension would be to examine the case of slowly drifting parameters in the plant.

APPENDIX

PROOF OF PROPOSITION 2 (STATEMENT 2)

Recall the tracking forecast defined in (7)

$$\begin{aligned}\tilde{f}(k+1) &= \tilde{f}(k) + \left(\frac{1}{k+1}\right)^\rho (\xi(k) + \nu(k) - \tilde{f}(k)) \\ f(k+1) &= \tilde{f}(k+1) + h(k+1),\end{aligned}$$

where $h(k+1) = \text{rand}([- \eta, \eta])$ for some $\eta > 0$. Note that the tracking forecast uses the noisy outcome sequence $\xi(k) + \nu(k)$. Furthermore, standard results from stochastic approximation (e.g., [3]) assure that the sequence $\{\tilde{f}(0), \tilde{f}(1), \tilde{f}(2), \dots\}$ is *uniformly bounded*, almost surely for $\rho > 1/2$. Specifically, $\sum_k (1/k+1)^\rho = \infty$, $\sum_k (1/k+1)^{2\rho} < \infty$, and $\nu(k)$ have finite variance.²

Now recall the calibration error (5) with respect to the pair (f^*, δ) as

$$e_{f^*, \delta}(K) = \frac{1}{K+1} \sum_{k=0}^K \mathcal{I}(|f(k) - f^*| < \delta) (\xi(k) - f(k)),$$

where $f^* \in [\xi_{\min}, \xi_{\max}]$.

We will show that for any $\varepsilon > 0$, the calibration error for the above tracking forecast with $\rho > 1/2$ and sufficiently small η satisfies

$$\limsup_{K \rightarrow \infty} |e_{f^*, \delta}(K)| \leq \varepsilon$$

for all noisy outcome sequences, $\{\hat{\xi}(0), \hat{\xi}(1), \hat{\xi}(2), \dots\}$, $f^* \in [\xi_{\min}, \xi_{\max}]$ and $\delta > 0$, almost surely.

The proof is a modification of the arguments in [18] for calibration of binary sequence. We will need some standard results from stochastic approximation [3] for the proof of Proposition 2. In particular, we will say that a sequence $M(k)$, $k = 0, 1, 2, \dots$, satisfies the *Kushner-Clark condition* if for all real $T > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\ell \geq n: \sum_{k=n}^{\ell} (1/k+1) \leq T} \left| \sum_{k=n}^{\ell} \frac{1}{k+1} M(k) \right| = 0.$$

For any f^* and δ , define

$$w(\tilde{f}) = \mathbb{E} \left[\mathcal{I} \left(\left| \tilde{f} + h - f^* \right| < \delta \right) \right],$$

where the expectation is taken over $h = \text{rand}([- \eta, \eta])$. The dependence of $w(\cdot)$ on f^* and δ is dropped for ease of notation. It is straightforward to verify that $w(\cdot)$ is Lipschitz continuous.

We can rewrite the calibration error in recursive form as

$$\begin{aligned}e_{f^*, \delta}(k+1) &= e_{f^*, \delta}(k) \\ &+ \left(\frac{1}{k+1}\right) \mathcal{I} \left(\left| \tilde{f}(k) + h(k) - f^* \right| < \delta \right) \\ &\left((\xi(k) - \tilde{f}(k) - h(k)) - e_{f^*, \delta}(k) \right).\end{aligned}$$

²See [3] for background on these conditions and the connection to the Kushner-Clarke conditions to be defined in the appendix.

Alternatively

$$\begin{aligned} e_{f^*,\delta}(k+1) &= e_{f^*,\delta}(k) \\ &+ \left(\frac{1}{k+1}\right)(M_1(k) - M_2(k)) \\ &+ M_3(k) - M_4(k) - e_{f^*,\delta}(k) \end{aligned}$$

where

$$\begin{aligned} M_1(k) &= w(\tilde{f}(k)) \left(\frac{\tilde{f}(k+1) - \tilde{f}(k)}{\left(\frac{1}{k+1}\right)^\rho} \right) \\ M_2(k) &= w(\tilde{f}(k))\nu(k) \\ M_3(k) &= \left(\mathcal{I} \left(\left| \tilde{f}(k) + h(k) - f^* \right| < \delta \right) - w(\tilde{f}(k)) \right) \\ &\quad (\xi(k) - \tilde{f}(k)) \\ M_4(k) &= \mathcal{I} \left(\left| \tilde{f}(k) + h(k) - f^* \right| < \delta \right) h(k). \end{aligned}$$

It can be verified using arguments from [18, Section 3.4.3], that the $M_1(k)$ terms satisfy the Kushner-Clark conditions. Since the $\nu(k)$ terms are zero-mean finite-variance and independent from $\tilde{f}(k)$, the $M_2(k)$ satisfy the Kushner-Clark conditions. Furthermore, the $M_3(k)$ are zero-mean and uniformly bounded (almost surely), and hence also satisfy the Kushner-Clark conditions. The implication is that the M_1 , M_2 , and M_3 terms have zero long term effects on the value of $e_{f^*,\delta}(k)$, almost surely. This is because the resulting ODE of stochastic approximation is

$$\dot{e}_{f^*,\delta} = -e_{f^*,\delta},$$

which is globally Lipschitz and globally asymptotically stable. Finally, we see that the long term effect of M_4 can be made arbitrarily small through η which bounds the magnitude of $h(k)$.

Proof of Proposition 3: Since the pair $[A, C]$ is observable, it is possible to build a state observer of the form

$$\dot{\hat{x}} = (A - LC)\hat{x} + (L \ B)y, \quad \hat{x}(0) = 0.$$

The initial conditions are deliberately set to zero. Note that this construction is *not required* for the switching algorithm. Rather, it is only a device for the proof of Proposition 3. The observer gain L is such that $(A - LC)$ is a stability matrix. In particular, for some $m_o, \lambda_o > 0$

$$\left\| e^{(A-LC)t} \right\| \leq m_o e^{-\lambda_o t}.$$

For any $\sigma < \lambda_o$, we can bound the size of the observer state, $|\hat{x}(t)|$, by

$$\begin{aligned} |\hat{x}(t)| &\leq \int_0^t m_o e^{-\lambda_o(t-\tau)} \|(L \ B)\| |y(\tau)| d\tau \\ &= \int_0^t m_o e^{-(\lambda_o - \sigma)(t-\tau)} \|(L \ B)\| e^{-\sigma(t-\tau)} |y(\tau)| d\tau \\ &\leq \frac{m_o \|(L \ B)\|}{\lambda_o - \sigma} \left\| y|_{[0,t]}^\sigma \right\|. \end{aligned}$$

Likewise, we can bound the size of the state estimation error, $x(t) - \hat{x}(t)$, by

$$|x(t) - \hat{x}(t)| \leq m_o e^{-\lambda_o t} |x(0)| + \frac{m_o}{\lambda_o} \left\| w|_{[0,t]} \right\|.$$

Combining these two inequalities results in

$$|x(t)| \leq \frac{m_o \|(L \ B)\|}{\lambda_o - \sigma} \left\| y|_{[0,t]}^\sigma \right\| + m_o e^{-\lambda_o t} |x(0)| + \frac{m_o}{\lambda_o} \left\| w|_{[0,t]} \right\|. \quad (21)$$

Let $u_{fs}(t)$ be a finitely switching control input as in (10). Let t_o denote the time of the last switch, and let F_i be the final gain matrix. There exist positive constants m_{fb} and λ_{fb} such that for any F_i, i i) $\left\| \begin{pmatrix} C \\ -F_i C \end{pmatrix} \right\| \leq m_{fb}$ and ii) $\left\| e^{(A-BF_i C)t} \right\| \leq m_{fb} e^{\lambda_{fb} t}$. From these bounds, for any $t \geq t_o$,

$$|y(t)| \leq m_{fb}^2 e^{\lambda_{fb}(t-t_o)} |x(t_o)| + \int_{t_o}^t m_{fb}^2 e^{\lambda_{fb}(t-\tau)} |w(\tau)| d\tau. \quad (22)$$

Substituting (21) evaluated at $t = t_o$ into (22) and applying standard bounding arguments establishes Assumption 1.

The restriction on σ is that it satisfies $\sigma < \lambda_o$. The observability assumptions implies that the analysis can be carried out with any λ_o by appropriately constructing L . Accordingly, σ can be arbitrarily chosen.

To show Assumption 2, let us assume that $A - BF_{i^*} C$ is a stability matrix satisfying

$$\left\| e^{(A-BF_{i^*} C)t} \right\| \leq m_s e^{-\lambda_s t} \quad (23)$$

for some strictly positive constants m_s and λ_s . As in (22), we can bound the output magnitude (but this time using (23)) by

$$|y(t)| \leq m_{fb} m_s e^{-\lambda_s(t-t_o)} |x(t_o)| + \frac{m_{fb} m_s}{\lambda_s} \left\| w|_{[t_o,t]} \right\|.$$

Likewise,

$$\begin{aligned} \left\| y|_{[t_o, t_o + \Delta T]}^\sigma \right\| &\leq \sup_{\tau \in [0, \Delta T]} e^{-\sigma(\Delta T - \tau)} \\ &\quad \times \left(m_{fb} m_s e^{-\lambda_s \tau} |x(t_o)| + \frac{m_{fb} m_s}{\lambda_s} \left\| w|_{[t_o, t_o + \tau]} \right\| \right). \end{aligned}$$

Assuming that $\sigma < \lambda_s$, the above inequality results in

$$\begin{aligned} \left\| y|_{[t_o, t_o + \Delta T]}^\sigma \right\| &\leq e^{-\sigma \Delta T} m_{fb} m_s |x(t_o)| \\ &\quad + \frac{m_{fb} m_s}{\lambda_s} \left\| w|_{[t_o, t_o + \Delta T]} \right\|. \quad (24) \end{aligned}$$

Moreover,

$$\begin{aligned} \left\| y|_{[0, t_o + \Delta T]}^\sigma \right\| &= \max \left(e^{-\sigma \Delta T} \left\| y|_{[0, t_o]}^\sigma \right\|, \left\| y|_{[t_o, t_o + \Delta T]}^\sigma \right\| \right) \\ &\leq e^{-\sigma \Delta T} \left\| y|_{[0, t_o]}^\sigma \right\| + \left\| y|_{[t_o, t_o + \Delta T]}^\sigma \right\|. \end{aligned}$$

Combining the above inequality with (24), substituting (21) evaluated at $t = t_o$, and applying standard bounding arguments leads to the desired result for sufficiently large ΔT .

REFERENCES

- [1] K. J. Astrom and B. Wittenmark, *Adaptive Control*. Boston, MA: Addison-Wesley, 1994.

- [2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The non-stochastic multiarmed bandit problem," *SIAM J. Comput.*, vol. 32, no. 1, pp. 48–77, 2003.
- [3] M. Benaim, "Dynamics of stochastic approximation algorithms," in *Seminaire de Probabilites XXXIII*, J. Azema, Ed. New York: Springer-Verlag, 1999, vol. 1709, pp. 1–68.
- [4] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. New York, NY: Cambridge University Press, 2006.
- [5] A. P. Dawid, "The well-calibrated bayesian," *J. Amer. Stat. Assoc.*, vol. 77, no. 379, pp. 605–610, 1982.
- [6] A. P. Dawid, "The impossibility of inductive inference," *J. Amer. Stat. Assoc.*, vol. 80, no. 390, pp. 340–341, 1985.
- [7] D. P. de Farias and N. Megiddo, "Combining expert advice in reactive environments," *J. Assoc. Comput. Mach. (J. ACM)*, vol. 53, no. 5, pp. 762–799, 2006.
- [8] D. P. Foster, Personal Communication, 2005.
- [9] D. P. Foster and R. Vohra, "Asymptotic Calibration," *Biometrika*, vol. 85, no. 2, pp. 379–390, 1998.
- [10] D. P. Foster and R. Vohra, "Regret in the on-line decision problem," *Games Econ. Behav.*, vol. 29, no. 1-2, pp. 7–35, Oct. 1999.
- [11] M. Fu and B. R. Barmish, "Adaptive stabilization of linear systems via switching control," *IEEE Trans. Automat. Control*, vol. AC-31, no. 12, pp. 1097–1103, Dec. 1986.
- [12] J. P. Hespanha, "Tutorial on supervisory control," in *Lecture Notes Workshop Control Logic Switching 40th IEEE Conf. Decision Control*, Dec. 2001, pp. 1–46.
- [13] J. P. Hespanha, D. Liberzon, and A. S. Morse, "Logic-based switching control of a non-holonomic system with parametric modeling uncertainty," *Syst. Control Lett.*, vol. 38, no. 3, pp. 167–177, Nov. 1999.
- [14] 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.
- [15] J. P. Hespanha, D. Liberzon, and A. S. Morse, "Overcoming the limitations of adaptive control by means of logic-based switching," *Syst. Control Lett.*, vol. 49, no. 1, pp. 49–65, May 2003.
- [16] J. Hocherman-Frommer, S. R. Kulkarni, and P. J. Ramadge, "Controller switching based on output prediction errors," *IEEE Trans. Automat. Control*, vol. AC-43, no. 5, pp. 596–607, May 1998.
- [17] S. M. Kakade, D. P. Foster, and D. P. Foster, J. Shawe-Taylor and Y. Singer, Eds., "Deterministic calibration and nash equilibrium," in *Proc. 17th Annu. Conf. Learning Theory*, 2004, pp. 33–48.
- [18] S. Mannor, J. S. Shamma, and G. Arslan, "Online calibrated forecasts: Memory efficiency versus universality for learning in games," *Machine Learning*, vol. 67, no. 1-2, pp. 77–115, May 2007.
- [19] A. S. Morse, "Supervisory control of families of linear set-point controllers—part 1: Exact matching," *IEEE Trans. Automat. Control*, vol. 41, no. 10, pp. 1413–1431, Oct. 1996.
- [20] A. S. Morse, "Supervisory control of families of linear set-point controllers—part 2: Robustness," *IEEE Trans. Automat. Control*, vol. 42, no. 11, pp. 1500–1515, Nov. 1997.
- [21] M. G. Safonov and T.-C. Tsao, "The unfalsified control concept and learning," *IEEE Trans. Automat. Control*, vol. 42, no. 6, pp. 843–847, Jun. 1997.
- [22] A. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.
- [23] R. Wang, A. Paul, M. Stefanovic, and M. G. Safonov, "Cost detectability and stability of adaptive control systems," *Int. J. Robust Nonlin. Control*, vol. 17, no. 5-6, pp. 549–561, Apr. 2007.



Ibrahim Al-Shyouch (M'07) received the B.Sc. degree from Jordan University of Science and Technology, Irbid, Jordan, in 1997, the M.Sc. degree from the Illinois Institute of Technology, Chicago, in 2000, and the Ph.D. degree from the University of California Los Angeles (UCLA) in 2007, all in mechanical engineering.

He is currently a Postdoctoral Scholar with the Department of Molecular and Medical Pharmacology, UCLA. His current research interests include learning and decision making in complex systems with applications to multi-agent and biological systems, combination therapy, and cell control.



Jeff S. Shamma (S'85–M'88–SM'98–F'06) received the B.S. degree from the Georgia Institute of Technology (Georgia Tech), Atlanta, in 1983 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1988, both in mechanical engineering.

He has held faculty positions at the University of Minnesota, Minneapolis, the University of Texas at Austin, and the University of California, Los Angeles. He returned to Georgia Tech in 2007, where he is a Professor of Electrical and Computer Engineering and the Julian T. Hightower Chair of Systems and Controls.