Mathematics of Control, Signals, and Systems (2018) 30:19
https://doi.org/10.1007/s00498-018-0225-1

ORIGINAL ARTICLE

Classical pole placement adaptive control revisited: linear-like convolution bounds and exponential stability

Daniel E. Miller[1] · **Mohamad T. Shahab**[1]

Received: 11 September 2018 / Accepted: 23 October 2018 / Published online: 7 November 2018

Abstract

While the original classical parameter adaptive controllers do not handle noise or unmodelled dynamics well, redesigned versions have been proven to have some tolerance; however, exponential stabilization and a bounded gain on the noise are rarely proven. Here we consider a classical pole placement adaptive controller using the original projection algorithm rather than the commonly modified version; we impose the assumption that the plant parameters lie in a convex, compact set, although some progress has been made at weakening the convexity requirement. We demonstrate that the closed-loop system exhibits a very desirable property: there are linear-like convolution bounds on the closed-loop behaviour, which confers exponential stability and a bounded noise gain, and which can be leveraged to prove tolerance to unmodelled dynamics and plant parameter variation. We emphasize that there is no persistent excitation requirement of any sort; the improved performance arises from the vigilant nature of the parameter estimator.

Keywords Adaptive control · Projection algorithm · Exponential stability · Bounded gain

1 Introduction

Adaptive control is an approach used to deal with systems with uncertain or time-varying parameters. The classical adaptive controller consists of a linear time-invariant (LTI) compensator together with a tuning mechanism to adjust the compensator parameters to match the plant. The first general proofs that adaptive controllers could work came around 1980, e.g. see [1–5]. However, such controllers were typically not robust

This research was supported by a grant from the Natural Sciences Research Council of Canada.

✉ Daniel E. Miller
miller@uwaterloo.ca

[1] Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

to unmodelled dynamics, did not tolerate time variations well, and did not handle noise or disturbances well, e.g. see [6]. During the following two decades, a great deal of effort was made to address these shortcomings. The most common approach was to make small controller design changes, such as the use of signal normalization, deadzones, and σ-modification, to ameliorate these issues, e.g. see [7–11]. Quite surprisingly, it turns out that simply using projection (onto a convex set of admissible parameters) has proved quite powerful, and the resulting controllers typically provide a bounded-noise bounded-state property, as well as tolerance of some degree of unmodelled dynamics and/or time variations, e.g. see [12–17]. Of course, it is clearly desirable that the closed-loop system exhibits LTI-like system properties, such as a bounded gain on the noise[1] and exponential stability. As far as the author is aware, in the classical approach to adaptive control a bounded gain on the noise is proven only in [12,13]; however, a crisp exponential bound on the effect of the initial condition is not provided, and a minimum phase assumption is imposed. While it is possible to prove a form of exponential stability if the reference input is sufficiently persistently exciting, e.g. see [18], this places a stringent requirement on an exogenous input.

There are several non-classical approaches to adaptive control which provide LTI-like system properties. First of all, in [19,20] a logic-based switching approach is used to sequence through a predefined list of candidate controllers; while exponential stability is proven, the transient behaviour can be quite poor and a bounded gain on the noise is not proven. A more sophisticated logic-based approach, labelled supervisory control, was proposed by Morse; here a supervisor switches in an efficient way between candidate controllers—see [21–25]. In certain circumstances a bounded gain on the noise can be proven—see [26,27], and the Concluding Remarks section of [22]. A related approach, called localization-based switching adaptive control, uses a falsification approach to prove exponential stability as well as a degree of tolerance of disturbances, e.g. see [28].

Another non-classical approach, proposed by the first author, is based on periodic probing, estimation, and control: rather than estimate the plant or controller parameters, the goal is to estimate what the control signal would be if the plant parameters and plant state were known and the 'optimal controller' were applied. Exponential stability and a bounded gain on the noise are achieved, as well as near optimal performance, e.g. see [29–31]; a degree of unmodelled dynamics and time variations can be allowed. Roughly speaking, the idea is to estimate the 'optimal control signal' at every step; this differs from the classical approach to adaptive control wherein the goal is to (at best) obtain an asymptotic estimate of the 'optimal control signal'. In order to carry out this estimation, a sampled data approximation of a differentiator is used, with optimality being achieved as the sampling period tends to zero. The drawback is that while a bounded gain on the noise is always achieved, it tends to increase dramatically the closer that one gets to optimality. Because of the nature of the approach, it only works in the continuous-time domain.

In this paper we consider the discrete-time setting and we propose an alternative approach to obtaining LTI-like system properties. We return to a common approach in

[1] Since the closed-loop system is nonlinear, a bounded-noise bounded-state property does not automatically imply a bounded gain on the noise.

classical adaptive control—the use of the projection algorithm to carry out parameter estimation together with the certainty equivalence principle. In the literature it is the norm to use a modified version of the ideal projection algorithm in order to avoid division by zero;[2] in this paper we prove that an unexpected consequence of this minor adjustment is that some inherent properties of the scheme have been destroyed. Here we use the original version of the projection algorithm coupled with a pole placement certainty equivalence-based controller. In the general case we impose compactness and convexity assumptions on the set of admissible parameters; however, some progress has been made at weakening the convexity requirement. We obtain linear-like convolution bounds on the closed-loop behaviour, which immediately confers exponential stability and a bounded gain on the noise; **such convolution bounds are, as far as the authors are aware, a first in adaptive control, and it allows us to use a modular approach to analyse robustness and tolerance to time-varying parameters, which is a highly desirable feature.** Indeed, this allows us to utilize all of the intuition that we have developed for LTI systems in the adaptive setting, something which has not been present using other techniques. To this end, the results will be presented in a very pedagogically desirable fashion: we first deal with the ideal plant (with disturbances); we then leverage that result to prove that a large degree of time variations is tolerated; we then demonstrate that the approach tolerates a degree of unmodelled dynamics, in a way familiar to those versed in the analysis of LTI systems.

In a recent short paper we consider the first-order case [32]. Here we consider the general case, which requires much more sophisticated analysis and proofs. Furthermore, in comparison to [32], here we (i) present a more general estimation algorithm, which alleviates the classical concern about dividing by zero, (ii) prove that the controller achieves the objective in the presence of a more general class of time variations, and (iii) prove robustness to unmodelled dynamics. An early version of this paper has appeared in a conference [33].

Before proceeding, we present some mathematical preliminaries. Let $\mathbf{Z}$ denote the set of integers, $\mathbf{Z}^+$ the set of non-negative integers, $\mathbf{N}$ the set of natural numbers, $\mathbf{R}$ the set of real numbers, and $\mathbf{R}^+$ the set of non-negative real numbers. We let $\mathbf{D}^0$ denote the open unit disc of the complex plane. We use the Euclidean 2-norm for vectors and the corresponding induced norm for matrices, and denote the norm of a vector or matrix by $\|\cdot\|$. We let $l_\infty(\mathbf{R}^n)$ denote the set of $\mathbf{R}^n$-valued bounded sequences; we define the norm of $u \in l_\infty(\mathbf{R}^n)$ by $\|u\|_\infty := \sup_{k\in\mathbf{Z}} \|u(k)\|$. Occasionally, we will deal with a map $F : l_\infty(\mathbf{R}^n) \to l_\infty(\mathbf{R}^n)$; the gain is given by $\sup_{u\neq 0} \frac{\|Fu\|_\infty}{\|u\|_\infty}$ and denoted by $\|F\|$. With $T \in \mathbf{Z}$, the truncation operator $P_T : l_\infty(\mathbf{R}^n) \to l_\infty(\mathbf{R}^n)$ is defined by

$$(P_T x)(t) = \begin{cases} x(t) & t \leq T \\ 0 & t > T. \end{cases}$$

We say that the map $F : l_\infty(\mathbf{R}^n) \to l_\infty(\mathbf{R}^n)$ is causal if $P_T F P_T = P_T F$ for every $T \in \mathbf{Z}$.

[2] An exception is the work of Ydstie [12,13], who considers the ideal Projection Algorithm as a special case; however, a crisp bound on the effect of the initial condition is not proven and a minimum phase assumption is imposed.

If $\mathcal{S} \subset \mathbf{R}^p$ is a convex and compact set, we define $\|\mathcal{S}\| := \max_{x \in \mathcal{S}} \|x\|$ and the function $\pi_{\mathcal{S}} : \mathbf{R}^p \to \mathcal{S}$ denotes the projection onto $\mathcal{S}$; it is well known that $\pi_{\mathcal{S}}$ is well defined.

2 The setup

In this paper we start with an nth-order linear time-invariant discrete-time plant given by

$$
\begin{aligned}
y(t+1) &= -\sum_{i=0}^{n-1} a_{i+1} y(t-i) + \sum_{i=0}^{n-1} b_{i+1} u(t-i) + d(t) \\
&= \underbrace{\left[\, y(t) \cdots y(t-n+1)\ u(t) \cdots u(t-n+1) \,\right]}_{=:\phi(t)^T} \underbrace{\begin{bmatrix} -a_1 \\ \vdots \\ -a_n \\ b_1 \\ \vdots \\ b_n \end{bmatrix}}_{=:\theta^*} + d(t), \\
\phi(t_0) &= \phi_0, \quad t \geq t_0,
\end{aligned} \tag{1}
$$

with $y(t) \in \mathbf{R}$ being the measured output, $u(t) \in \mathbf{R}$ the control input, $\phi(t)$ a vector of input–output data, and $d(t) \in \mathbf{R}$ the disturbance (or noise) input. We assume that θ^* is unknown but belongs to a known set $\mathcal{S} \subset \mathbf{R}^{2n}$. Associated with this plant model are the polynomials

$$A(z^{-1}) := 1 + a_1 z^{-1} + \cdots + a_n z^{-n}, \quad B(z^{-1}) := b_1 z^{-1} + \cdots + b_n z^{-n}$$

and the transfer function $\frac{B(z^{-1})}{A(z^{-1})}$.

Remark 1 It is straightforward to verify that if the system has a disturbance at both the input and output, then it can be converted to a system of the above form.

We impose an assumption on the set of admissible plant parameters.

Assumption 1: $\mathcal{S}$ is convex and compact, and for each $\theta^* \in \mathcal{S}$, the corresponding pair of polynomials $A(z^{-1})$ and $B(z^{-1})$ are coprime.

The convexity part of the above assumption is common in a branch of the adaptive control literature—it is used to facilitate constrained parameter estimation, e.g. see [34], and it is a key assumption in arguably the simplest technique to ensuring that the associated pole placement adaptive controller has tolerance to a degree of unmodelled dynamics and to noise, e.g. see [12–17,35]. That being said, in Sect. 8 we will show that it is possible to weaken this to assuming that $\theta^* \in \mathcal{S}_1 \cup \mathcal{S}_2$ with each S_i convex and compact, and for each $\theta^* \in \mathcal{S}_1 \cup \mathcal{S}_2$, the corresponding pair of polynomials $A(z^{-1})$

and $B(z^{-1})$ are coprime; however, we will require that all closed-loop poles be placed at zero, and a more complicated controller will be required. The boundedness part of the assumption is less common, but it is quite reasonable in practical situations; it is used here to ensure that we can prove uniform bounds and decay rates on the closed-loop behaviour.

The main goal here is to prove a form of stability, with a secondary goal that of asymptotic tracking of an **exogenous reference signal** $y^*(t)$; since the plant may be non-minimum phase, there are limits on how well the plant can be made to track $y^*(t)$. To proceed, we use a parameter estimator together with an adaptive pole placement control law. At this point, we discuss the most critical aspect—the parameter estimator.

2.1 Parameter estimation

We can write the plant as

$$y(t+1) = \phi(t)^T \theta^* + d(t).$$

Given an estimate $\hat{\theta}(t)$ of θ^* at time t, we define the **prediction error** by

$$e(t+1) := y(t+1) - \phi(t)^T \hat{\theta}(t);$$

this is a measure of the error in $\hat{\theta}(t)$. The common way to obtain a new estimate is from the solution of the optimization problem

$$\text{argmin}_\theta \{\|\theta - \hat{\theta}(t)\| : y(t+1) = \phi(t)^T \theta\},$$

yielding the **ideal (projection) algorithm**

$$\hat{\theta}(t+1) = \begin{cases} \hat{\theta}(t) & \text{if } \phi(t) = 0 \\ \hat{\theta}(t) + \frac{\phi(t)}{\phi(t)^T \phi(t)} e(t+1) & \text{otherwise.} \end{cases} \tag{2}$$

Of course, if $\phi(t)$ is close to zero, numerical problems can occur, so it is the norm in the literature (e.g. [3,34]) to replace this by the following **classical algorithm**: with $0 < \alpha < 2$ and $\beta > 0$, define[3]

$$\hat{\theta}(t+1) = \hat{\theta}(t) + \frac{\alpha \phi(t)}{\beta + \phi(t)^T \phi(t)} e(t+1). \tag{3}$$

This latter algorithm is widely used and plays a role in many discrete-time adaptive control algorithms; however, when this algorithm is used, all of the results are asymptotic, and exponential stability and a bounded gain on the noise are never proven. It is not hard to guess why—a careful look at the estimator shows that the gain on the update law is small if $\phi(t)$ is small. A more mathematically detailed argument is given in the following example.

[3] It is common to make this more general by letting α be time-varying.

Remark 2 Consider the simple first-order plant

$$y(t+1) = -a_1 y(t) + b_1 u(t) + d(t)$$

with $a_1 \in [-2, -1]$ and $b_1 \in [1, 2]$. We use the estimator (3) with $\alpha \in (0, 2)$ and $\beta > 0$, and, as in [12–17], we use projection to keep the parameters estimates inside $\mathcal{S}$ so as to guarantee a bounded-input bounded-state property. Further suppose that $y^* = d = 0$, and that a classical pole placement adaptive controller places the closed-loop pole at zero: $u(t) = \frac{\hat{a}_1(t)}{\hat{b}_1(t)} y(t) =: \hat{f}(t) y(t)$. Suppose that

$$y(0) = y_0 = \varepsilon \in (0, 1), \quad \hat{\theta}(0) = \begin{bmatrix} -\hat{a}_1(0) \\ \hat{b}_1(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \theta^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

so that $\hat{f}(0) = -0.5$ and $-a_1 + b_1 \hat{f}(0) = 1.5$, i.e. the system is initially unstable. An easy calculation verifies that $\hat{f}(t) \in [-2, -0.5]$ and $-a_1 + b_1 \hat{f}(t) \in [0, 1.5]$ for $t \geq 0$, which leads to a crude bound on the closed-loop behaviour: $|y(t)| \leq (1.5)^t \varepsilon$ for $t \geq 0$. With $N(\varepsilon) := \text{int}[\frac{1}{2\ln(1.5)} \ln(\frac{1}{\varepsilon})]$, it follows that

$$|y(t)| \leq \varepsilon^{1/2}, \quad t \in [0, N(\varepsilon)].$$

A careful examination of the parameter estimator shows that

$$\|\hat{\theta}(t) - \theta_0\| \leq 20(2)^{1/2} \frac{\varepsilon}{\beta}, \quad t \in [0, N(\varepsilon)].$$

From the form of $\hat{f}(t)$, it is easy to prove that if $\frac{\varepsilon}{\beta}$ is small enough,[4] then we have $|-a + b_1 \hat{f}(t)| \geq 1.25$ for $t \in [0, N(\varepsilon)]$, in which case

$$|y(N(\varepsilon))| \geq (1.25)^{N(\varepsilon)} \varepsilon \quad \Rightarrow \quad \left| \frac{y(N(\varepsilon))}{\varepsilon} \right| = \left| \frac{y(N(\varepsilon))}{y(0)} \right| \geq (1.25)^{N(\varepsilon)};$$

since $N(\varepsilon) \to \infty$ as $\varepsilon \to 0$, we see that exponential stability is not achieved. A similar kind of analysis can be used to prove that a bounded gain on the noise is not achieved either.

Now we return to the problem at hand—analysing the ideal algorithm (2). We will be using the ideal algorithm with projection to ensure that the estimate remains in $\mathcal{S}$ for all time. With an initial condition of $\hat{\theta}(t_0) = \theta_0 \in \mathcal{S}$, for $t \geq t_0$ we set

$$\check{\theta}(t+1) = \begin{cases} \hat{\theta}(t) & \text{if } \phi(t) = 0 \\ \hat{\theta}(t) + \frac{\phi(t)}{\phi(t)^T \phi(t)} e(t+1) & \text{otherwise,} \end{cases} \tag{4}$$

[4] Here $\beta > 0$ is fixed, so this is equivalent to ε being small enough.

which we then project onto $\mathcal{S}$:

$$\hat{\theta}(t+1) := \pi_{\mathcal{S}}(\check{\theta}(t+1)). \tag{5}$$

Because of the closed and convex property of $\mathcal{S}$, the projection function is well defined; furthermore, it has the nice property that, for every $\theta \in \mathbf{R}^{2n}$ and every $\theta^* \in \mathcal{S}$, we have

$$\|\pi_{\mathcal{S}}(\theta) - \theta^*\| \leq \|\theta - \theta^*\|,$$

i.e. projecting θ onto $\mathcal{S}$ never makes it further away from the quantity θ^*.

2.2 Revised parameter estimation

Some readers may be concerned that the original problem of dividing by a number close to zero, which motivates the use of classical algorithm, remains. Of course, this is balanced against the soon-to-be-proved benefit of using (4)–(5). We propose a middle ground as follows. A straightforward analysis of $e(t+1)$ reveals that

$$e(t+1) = -\phi(t)^T[\hat{\theta}(t) - \theta^*] + d(t),$$

which means that

$$|e(t+1)| \leq 2\|\mathcal{S}\| \times \|\phi(t)\| + |d(t)|.$$

Therefore, if

$$|e(t+1)| > 2\|\mathcal{S}\| \times \|\phi(t)\|,$$

then the update to $\hat{\theta}(t)$ will be greater than $2\|\mathcal{S}\|$, which means that there is little information content in $e(t+1)$—it is dominated by the disturbance. With this as motivation, and with $\delta \in (0, \infty]$, let us replace (4) with

$$\check{\theta}(t+1) = \begin{cases} \hat{\theta}(t) + \frac{\phi(t)}{\phi(t)^T\phi(t)}\, e(t+1) & \text{if } |e(t+1)| < (2\|\mathcal{S}\| + \delta)\|\phi(t)\| \\ \hat{\theta}(t) & \text{otherwise;} \end{cases} \tag{6}$$

in the case of $\delta = \infty$, we will adopt the understanding that $\infty \times 0 = 0$, in which case the above formula collapses into the original one (4). In the case that $\delta < \infty$, we can be assured that the update term is bounded above by $2\|\mathcal{S}\| + \delta$, which should alleviate concerns about having infinite gain. We would now like to rewrite the update to make it more concise. To this end, we now define $\rho_\delta : \mathbf{R}^{2n} \times \mathbf{R} \to \{0, 1\}$ by

$$\rho_\delta(\phi(t), e(t+1)) := \begin{cases} 1 & \text{if } |e(t+1)| < (2\|\mathcal{S}\| + \delta)\|\phi(t)\| \\ 0 & \text{otherwise,} \end{cases}$$

yielding a more concise way to write the estimation algorithm update:

$$\check{\theta}(t+1) = \hat{\theta}(t) + \rho_\delta(\phi(t), e(t+1)) \frac{\phi(t)}{\phi(t)^T\phi(t)} e(t+1); \tag{7}$$

once again, we project this onto $\mathcal{S}$:

$$\hat{\theta}(t+1) := \pi_{\mathcal{S}}(\check{\theta}(t+1)). \tag{8}$$

Remark 3 If the disturbance $d(t) = 0$, then the estimation algorithm (7)–(8) has a nice scaling property. In this case, if $\phi(t) \neq 0$ then $\rho_\delta(\phi(t), e(t+1)) = 1$, so (7) becomes

$$\check{\theta}(t+1) = \hat{\theta}(t) + \frac{\phi(t)\phi(t)^T}{\phi(t)^T\phi(t)}[\theta^* - \hat{\theta}(t)];$$

so if $\phi(t)$ is replaced by $\gamma\phi(t)$ with $\gamma \neq 0$, then $\check{\theta}(t+1)$ (and $\hat{\theta}(t+1)$) remains the same. Hence, scaling the pair $(y(t), u(t))$ makes no difference to the estimator, which is clearly a desirable feature; notice that the classical algorithm (3) does not enjoy that property. This is the first clue that this algorithm may provide closed-loop properties not provided by the classical algorithm (3).

2.3 Properties of the estimation algorithm

Analysing the closed-loop system behaviour will require a careful examination of the estimation algorithm. We define the parameter estimation error by

$$\tilde{\theta}(t) := \hat{\theta}(t) - \theta^*,$$

and the corresponding Lyapunov function associated with $\tilde{\theta}(t)$, namely $V(t) := \tilde{\theta}(t)^T\tilde{\theta}(t)$. In the following result we list a property of $V(t)$; it is a generalization of what is well known for the classical algorithm (3).

Proposition 1 *For every $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\theta_0 \in \mathcal{S}$, $\theta^* \in \mathcal{S}$, $d \in l_\infty$, and $\delta \in (0, \infty]$, when the estimator (7) and (8) is applied to the plant (1), the following holds:*

$$\|\hat{\theta}(t+1) - \hat{\theta}(t)\| \leq \rho_\delta(\phi(t), e(t+1)) \frac{|e(t+1)|}{\|\phi(t)\|}, \quad t \geq t_0, \tag{9}$$

$$V(t) \leq \quad V(t_0) + \sum_{j=t_0}^{t-1} \rho_\delta(\phi(j), e(j+1))\Big[-\frac{1}{2}\frac{[e(j+1)]^2}{\|\phi(j)\|^2} + 2\frac{[d(j)]^2}{\|\phi(j)\|^2}\Big], \quad t \geq t_0+1.$$

Proof See Appendix. □

2.4 The control law

The elements of $\hat{\theta}(t)$ are partitioned in a natural way as

$$\hat{\theta}(t) =: \left[-\hat{a}_1(t) \cdots -\hat{a}_n(t)\ \hat{b}_1(t) \cdots \hat{b}_n(t) \right]^T .$$

Associated with $\hat{\theta}(t)$ are the polynomials

$$\hat{A}(t, z^{-1}) := 1 + \hat{a}_1(t)z^{-1} + \cdots + \hat{a}_n(t)z^{-n},$$
$$\hat{B}(t, z^{-1}) := \hat{b}_1(t)z^{-1} + \cdots + \hat{b}_n(t)z^{-n}.$$

While we can use an $(n-1)$th-order **proper** controller to carry out pole placement, it will be convenient to use an nth-order **strictly proper** controller, such as in [15,16,36–38]. In particular, we first choose a $2n$th-order monic polynomial

$$A^*(z^{-1}) = 1 + a_1^* z^{-1} + \cdots + a_{2n}^* z^{-2n}$$

so that $z^{2n}A^*(z^{-1})$ has all of its zeros in $\mathbf{D}^o$. Next, we choose two polynomial

$$\hat{L}(t, z^{-1}) = 1 + \hat{l}_1(t)z^{-1} + \cdots + \hat{l}_n(t)z^{-n}$$
$$\text{and} \quad \hat{P}(t, z^{-1}) = \hat{p}_1(t)z^{-1} + \cdots + \hat{p}_n(t)z^{-n}$$

which satisfy the equation

$$\hat{A}(t, z^{-1})\hat{L}(t, z^{-1}) + \hat{B}(t, z^{-1})\hat{P}(t, z^{-1}) = A^*(z^{-1}); \tag{10}$$

given the assumption that the $\hat{A}(t, z^{-1})$ and $\hat{B}(t, z^{-1})$ are coprime, it is well known that there exist **unique** $\hat{L}(t, z^{-1})$ and $\hat{P}(t, z^{-1})$ which satisfy this equation. Indeed, it is easy to prove that the coefficients of $\hat{L}(t, z^{-1})$ and $\hat{P}(t, z^{-1})$ are analytic functions of $\hat{\theta}(t) \in \mathcal{S}$.

In our setup we have an exogenous signal $y^*(t)$. At time t we choose $u(t)$ so that

$$\begin{aligned} u(t) = &-\hat{l}_1(t-1)u(t-1) - \cdots - \hat{l}_n(t-1)u(t-n) \\ &-\hat{p}_1(t-1)[y(t-1) - y^*(t-1)] \\ &- \cdots - \hat{p}_n(t-1)[y(t-n) - y^*(t-n)]. \end{aligned} \tag{11}$$

So the overall controller consists of the estimator (7)–(8) together with (11).[5]

It turns out that we can write down a state-space model of our closed-loop system with $\phi(t) \in \mathbf{R}^{2n}$ as the state. Proceeding as in Kreisselmeier [37], only two elements of ϕ have a complicated description:

[5] We also implicitly use a pole placement procedure to obtain the controller parameters from the plant parameter estimates; this entails solving a linear equation.

$$\phi_1(t+1) = y(t+1) = e(t+1) + \hat{\theta}(t)^T\phi(t),$$
$$\phi_{n+1}(t+1) = u(t+1) = -\sum_{i=1}^{n}\left\{\hat{l}_i(t)u(t+1-i) + \hat{p}_i(t)[y(t+1-i) - y^*(t+1-i)]\right\}$$
$$= \left[\, -\hat{l}_1(t) \ \cdots \ -\hat{l}_n(t) \ -\hat{p}_1(t) \ \cdots \ -\hat{p}_n(t) \,\right]\phi(t) + \sum_{i=1}^{n}\hat{p}_i(t)y^*(t+1-i).$$

With $e_i \in \mathbf{R}^{2n}$ the ith normal vector, if we now define

$$\bar{A}(t) := \begin{bmatrix} -\hat{a}_1(t) & -\hat{a}_2(t) & \cdots & -\hat{a}_n(t) & \hat{b}_1(t) & \cdots & \cdots & \hat{b}_n(t) \\ 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ & \ddots & & \vdots & \vdots & \cdots & \cdots & \vdots \\ & & 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\hat{p}_1(t) & -\hat{p}_2(t) & \cdots & -\hat{p}_n(t) & -\hat{l}_1(t) & -\hat{l}_2(t) & \cdots & -\hat{l}_n(t) \\ 0 & \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & & & 1 & 0 \end{bmatrix},$$

$$B_1 := e_1, \quad B_2 := e_{n+1}, \quad r(t) := \sum_{i=1}^{n}\hat{p}_i(t)y^*(t+1-i), \tag{12}$$

then the following key equation holds:

$$\phi(t+1) = \bar{A}(t)\phi(t) + B_1 e(t+1) + B_2 r(t); \tag{13}$$

notice that the characteristic equation of $\bar{A}(t)$ always equals $z^{2n}A^*(z^{-1})$. Before proceeding, define

$$\bar{a} := \max\{\|\bar{A}(\hat{\theta})\| : \hat{\theta} \in \mathcal{S}\}.$$

Remark 4 While the proposed adaptive controller (7)–(8) together with (11) is nonlinear, when it is applied to the plant the closed-loop system enjoys the homogeneity property. More precisely, fix the initial parameter estimate θ_0 and starting time $t_0 \in \mathbf{Z}$; suppose that an initial condition, reference signal, and disturbance signal combination (ϕ_0, r, d) yields a system response of ϕ, and with $\gamma \in \mathbf{R}$ suppose that an initial condition, reference signal, and disturbance signal combination of $(\gamma\phi_0, \gamma r, \gamma d)$ yields a system response of ϕ^{γ}. Using induction it is easy to prove that[6]

$$\phi^{\gamma}(t) = \gamma\phi(t), \ t \geq t_0.$$

Hence, with minimal effort we see that the closed-loop behaviour enjoys one of the two required properties of a linear system, namely that of homogeneity. While it does not enjoy the other property needed for linearity, we will soon see that we are still able to prove linear-like convolution bounds on the closed-loop behaviour.

3 Preliminary analysis

The closed-loop system given in (13) arises in classical adaptive control approaches in slightly modified fashion, so we will borrow some tools from there. More specifically, the following result was proven by Kreisselmeier [37], in the context of proving that a slowly time-varying adaptive control system is stable (in a weak sense); we are providing a special case of his technical lemma to minimize complexity.[7]

Proposition 2 *[37] Consider the discrete-time system*

$$x(t+1) = [A_{nom}(t) + \Delta(t)]x(t)$$

with $\Phi(t,\tau)$ denoting the corresponding state transition matrix. Suppose that there exist constants $\sigma \in (0,1)$, $\gamma_1 > 1$, $\alpha_i \geq 0$, and $\beta_i \geq 0$ so that

(i) for all $t \geq t_0$, we have $\|A_{nom}(t)^i\| \leq \gamma_1\sigma^i, \ i \geq 0$;
(ii) for all $t > \tau$ we have

$$\sum_{i=\tau}^{t-1} \|A_{nom}(i+1) - A_{nom}(i)\| \leq \alpha_0 + \alpha_1(t-\tau)^{1/2} + \alpha_2(t-\tau)$$

and $\sum_{i=\tau}^{t-1} \|\Delta(i)\| \leq \beta_0 + \beta_1(t-\tau)^{1/2} + \beta_2(t-\tau)$;
(iii) there exists a $\mu \in (\sigma, 1)$ and $N \in \mathbf{N}$ satisfying $\alpha_2 + \frac{\beta_2}{N} < \frac{1}{N\gamma_1}(\frac{\mu}{\gamma_1^{1/N}} - \sigma)$.

Then there exists a constant γ_2 so that the transition matrix satisfies

$$\|\Phi(t,\tau)\| \leq \gamma_2\mu^{t-\tau}, \ t \geq \tau.$$

6 In addition, if we define $\hat{\theta}(t)$ and $\hat{\theta}^{\gamma}(t)$ in the natural way, then it is easy to prove that for $\gamma \neq 0$ we have

$$\hat{\theta}^{\gamma}(t) = \hat{\theta}(t), \ t \geq t_0.$$

7 Furthermore, in [37] it is assumed that α_i and β_i are strictly greater than zero, but it is trivial to extend this to allow for zero as well.

Remark 5 We apply the above proposition in the following way. Suppose that $\sigma \in (0,1)$, $\gamma_1 > 1$, $\alpha_i \geq 0$, $\beta_i \geq 0$ are such conditions (i) and (ii) hold. If $\mu \in (\sigma, 1)$, then it follows that $\frac{\mu}{\gamma_1^{1/N}} - \sigma > 0$ for large enough $N \in \mathbf{N}$, so condition (iii) will hold as well as long as α_2 and β_2 are small enough.

In applying Proposition 2, the matrix $\bar{A}(t)$ will play the role of $A_{nom}(t)$. A key requirement is that Condition (i) holds: the following provides relevant bounds. Before proceeding, let

$$\underline{\lambda} := \max\{|\lambda| : \lambda \text{ is a root of } z^{2n}A^*(z^{-1})\}.$$

Lemma 1 *For every $\delta \in (0,\infty]$ and $\sigma \in (\underline{\lambda}, 1)$ there exists a constant $\gamma \geq 1$ so that for every $t_0 \in \mathbf{Z}$, $\theta_0 \in \mathcal{S}$, $\theta^* \in \mathcal{S}$, and $y^*, d \in l_\infty$, when the controller (7), (8) and (11) is applied to the plant (1), the matrix $\bar{A}(t)$ satisfies, for every $t \geq t_0$:*

$$\|\bar{A}(t)^k\| \leq \gamma\sigma^k, \;\; k \geq 0,$$

and for every $t > k \geq t_0$:

$$\sum_{j=k}^{t-1} \|\bar{A}(j+1) - \bar{A}(j)\| \leq \gamma[\sum_{j=k}^{t-1} \rho_\delta(\phi(j), e(j+1))\frac{e(j+1)^2}{\|\phi(j)\|^2}]^{1/2}(t-k)^{1/2}.$$

Proof See Appendix. □

4 The main result

Theorem 1 *For every $\delta \in (0,\infty]$ and $\lambda \in (\underline{\lambda}, 1)$ there exists a $c > 0$ so that for every $t_0 \in \mathbf{Z}$, $\theta_0 \in \mathcal{S}$, $\theta^* \in \mathcal{S}$, $\phi_0 \in \mathbf{R}^{2n}$, and $y^*, d \in \ell_\infty$, when the adaptive controller (7), (8) and (11) is applied to the plant (1), the following bound holds:*

$$\|\phi(k)\| \leq c\lambda^{k-t_0}\|\phi_0\| + \sum_{j=t_0}^{k-1} c\lambda^{k-1-j}(|r(j)| + |d(j)|), \;\; k \geq t_0. \tag{14}$$

Remark 6 We see from (12) that $r(t)$ is a weighted sum of $\{y^*(t), ..., y^*(t-n+1)\}$. Hence, there exists a constant $\bar{c}$ so that the bound (14) can be rewritten as

$$\|\phi(k)\| \leq c\lambda^{k-t_0}\|\phi_0\| + \sum_{j=t_0-n+1}^{k-1} \bar{c}\lambda^{k-1-j}|y^*(j)| + \sum_{j=t_0}^{k-1} c\lambda^{k-1-j}|d(j)|, \;\; k \geq t_0.$$

Remark 7 Theorem 1 implies that the system has a bounded gain (from d and r to y) in every $p-$norm. More specifically, for $p = \infty$ we see immediately from (14) that

$$\|\phi(k)\| \leq c\|\phi_0\| + \frac{c}{1-\lambda} \sup_{\tau \in [t_0,k]} [|r(\tau)| + |d(\tau)|)], \quad k \geq t_0.$$

Furthermore, for $1 \leq p < \infty$ it follows from Young's Inequality applied to (14) that

$$\left[\sum_{j=t_0}^{k} \|\phi(j)\|^p\right]^{1/p} \leq \frac{c}{(1-\lambda^p)^{1/p}} \|\phi_0\|$$
$$+ \frac{c}{1-\lambda} \left\{ \left[\sum_{j=t_0}^{k} \|r(j)\|^p\right]^{1/p} + \left[\sum_{j=t_0}^{k} \|d(j)\|^p\right]^{1/p} \right\}, \; k \geq t_0.$$

Remark 8 Most pole placement adaptive controllers are proven to yield a weak form of stability, such as boundedness (in the presence of a nonzero disturbance) or asymptotic stability (in the case of a zero disturbance), which means that details surrounding initial conditions can be ignored. Here the goal is to prove a stronger, linear-like, convolution bound, so it requires a more detailed and nuanced analysis. A key tool is Proposition 2, which was introduced by Kreisselmeier [37] to analyse slowly time-varying adaptive pole placement problems. It has been used in a number of places in the adaptive control literature, including the work of [15,16,35], all of whom utilize the classical projection algorithm (3). However, as pointed out in Remark 2, an adaptive controller based on the classical projection algorithm (3) does not, in general, provide exponential stability or a bounded gain on the noise, **regardless of how small the parameter** $\beta > 0$ **is**; indeed, what is proven in [15,16] is that for every set of initial conditions and every pair of exogenous disturbance and reference signal inputs, the state $\phi(t)$ is bounded, i.e. the system enjoys the bounded-input bounded-state property. What is surprising and unexpected is that for $\beta = 0$, the closed-loop system enjoys much nicer properties, and clearly this does not follow in any obvious way by taking the limit as $\beta \to 0$ of what is proven in the classical setup of [15,16].

Remark 9 The approach taken in our proof is motivated by our earlier work on the first-order one-step-ahead adaptive controller [32]; here we use Kreisselmeier's result on time-varying systems given in Proposition 2 in place of a lemma used in [32] for time-varying first-order systems. While the layout of our proof has a superficial similarity to that to [15,16], in that we both partition the timescale in terms of the size of state (in the case of [15,16]) or the size of the disturbance scaled by the state (here), on closer inspection it is clear that they differ significantly.

Remark 10 With $\hat{G}(t, z^{-1}) = \sum_{i=1}^{2n} \hat{g}_i(t) z^{-i} := \hat{B}(t, z^{-1})\hat{P}(t, z^{-1})$ it is possible to use arguments like those in [34] to prove, when the disturbance d is identically zero, a weak tracking result of the form

$$\lim_{t\to\infty}\left[\sum_{i=0}^{2n} a_i^* y(t-i) - \sum_{i=1}^{2n} \hat{g}_i(t) y^*(t-i)\right] = 0.$$

Since the main goal of the paper is on stability issues, we omit the proof. However, we do discuss step tracking in a later section.

Proof Fix $\delta \in (0, \infty]$ and $\lambda \in (\underline{\lambda}, 1)$. Let $t_0 \in \mathbf{Z}$, $\theta_0 \in \mathcal{S}$, $\theta^* \in \mathcal{S}$, $\phi_0 \in \mathbf{R}^{2n}$, and $y^*, d \in l_\infty$ be arbitrary. Define r via (12). Now choose $\lambda_1 \in (\underline{\lambda}, \lambda)$.

We have to be careful in how to apply Proposition 2 to (13)—we need the $\Delta(t)$ term to be something which we can bound using Proposition 1. So define

$$\Delta(t) := \rho_\delta(\phi(t), e(t+1)) \frac{e(t+1)}{\|\phi(t)\|^2} B_1 \phi(t)^T; \tag{15}$$

it is easy to check that

$$\Delta(t)\phi(t) = \rho_\delta(\phi(t), e(t+1)) B_1 e(t+1)$$

and that

$$\|\Delta(t)\| = \rho_\delta(\phi(t), e(t+1)) \frac{|e(t+1)|}{\|\phi(t)\|},$$

which is a term which plays a key role in Proposition 1. We can now rewrite (13) as

$$\phi(t+1) = [\bar{A}(t) + \Delta(t)]\phi(t) + B_1 \underbrace{[1-\rho_\delta(\phi(t), e(t+1))]e(t+1)}_{=:\eta(t)} + B_2 r(t). \tag{16}$$

If $\rho_\delta(\phi(t), e(t+1)) = 1$ then $\eta(t) = 0$, but if $\rho_\delta(\phi(t), e(t+1)) = 0$ then

$$|e(t+1)| \geq (2\|\mathcal{S}\| + \delta)\|\phi(t)\|;$$

but we also know that

$$e(t+1) = -\tilde{\theta}(t)\phi(t) + d(t) \;\Rightarrow\; |e(t+1)| \leq 2\|\mathcal{S}\| \times \|\phi(t)\| + |d(t)|; \tag{17}$$

combining these equations we have

$$(2\|\mathcal{S}\| + \delta)\|\phi(t)\| \leq 2\|\mathcal{S}\| \times \|\phi(t)\| + |d(t)|,$$

which implies that $\|\phi(t)\| \leq \frac{1}{\delta}|d(t)|$; it is easy to check that this holds even when $\delta = \infty$. Using (17) we conclude that

$$|\eta(t)| \leq \left(\frac{2\|\mathcal{S}\|}{\delta} + 1\right)|d(t)|, \quad t \geq t_0. \tag{18}$$

We now analyse (16). We let $\Phi(t, \tau)$ denote the transition matrix associated with $\bar{A}(t) + \Delta(t)$; this matrix clearly implicitly depends on θ_0, θ^*, d and r. From Lemma 1 there exists a constant γ_1 so that

$$\|\bar{A}(t)^i\| \leq \gamma_1 \lambda_1^i, \quad i \geq 0, \; t \geq t_0, \tag{19}$$

and for every $t > k \geq t_0$, we have

$$\begin{aligned} &\sum_{j=k}^{t-1} \|\bar{A}(j+1) - \bar{A}(j)\| \\ &\leq \gamma_1 \left[\sum_{j=k}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{|e(j+1)|^2}{\|\phi(j)\|^2} \right]^{1/2} (t-k)^{1/2}. \end{aligned} \tag{20}$$

Using the definition of Δ given in (15) and the Cauchy–Schwarz inequality, we also have

$$\begin{aligned} &\sum_{j=k}^{t-1} \|\Delta(j)\| \\ &\leq \left[\sum_{j=k}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{|e(j+1)|^2}{\|\phi(j)\|^2} \right]^{1/2} (t-k)^{1/2}, \; t > k \geq t_0. \end{aligned} \tag{21}$$

At this point we consider two cases: the easier case in which there is no noise, and the harder case in which there is noise.
Case 1: $d(t) = 0, t \geq t_0$.
Using the bound on $\eta(t)$ given in (18), in this case (16) becomes

$$\phi(t+1) = [\bar{A}(t) + \Delta(t)]\phi(t) + B_2 r(t), \quad t \geq t_0. \tag{22}$$

The bound on $V(t)$ given by Proposition 1 simplifies to

$$V(t) \leq V(t_0) - \frac{1}{2} \sum_{j=t_0}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{[e(j+1)]^2}{\|\phi(j)\|^2}, \quad t \geq t_0 + 1.$$

Since $V(\cdot) \geq 0$ and $V(t_0) = \|\theta_0 - \theta^*\|^2 \leq 4\|\mathcal{S}\|^2$, this means that

$$\sum_{j=t_0}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{[e(j+1)]^2}{\|\phi(j)\|^2} \leq 2V(t_0) \leq 8\|\mathcal{S}\|^2.$$

Hence, from (20) and (21) we conclude that

$$\sum_{j=k}^{t-1} \|\bar{A}(j+1) - \bar{A}(j)\| \leq 8^{1/2}\gamma_1\|\mathcal{S}\|(t-k)^{1/2},$$
$$\sum_{j=k}^{t-1} \|\Delta(j)\| \leq 8^{1/2}\|\mathcal{S}\|(t-k)^{1/2}, \ t > k \geq t_0.$$

Now we apply Proposition 2: we set

$$\alpha_0 = \beta_0 = \alpha_2 = \beta_2 = 0, \quad \alpha_1 = 8^{1/2}\gamma_1\|\mathcal{S}\|, \quad \beta_1 = 8^{1/2}\|\mathcal{S}\|, \quad \mu = \lambda.$$

Following Remark 3 it is now trivial to choose $N \in \mathbf{N}$ so that $\frac{\lambda}{\gamma_1^{1/N}} - \lambda_1 > 0$, namely

$$N = \text{int}\left[\frac{\ln(\gamma_1)}{\ln(\lambda) - \ln(\lambda_1)}\right] + 1, \tag{23}$$

which means that

$$0 = \alpha_2 + \frac{\beta_2}{N} < \frac{1}{N\gamma_1}\left(\frac{\lambda}{\gamma_1^{1/N}} - \lambda_1\right).$$

From Proposition 2 we see that there exists a constant γ_2 so that the state transition matrix $\Phi(t,\tau)$ corresponding to $\bar{A}(t) + \Delta(t)$ satisfies

$$\|\Phi(t,\tau)\| \leq \gamma_2\lambda^{t-\tau}, \quad t \geq \tau \geq t_0.$$

If we now apply this to (22), we end up with the desired bound:

$$\|\phi(k)\| \leq \gamma_2\lambda^{k-t_0}\|\phi(t_0)\| + \sum_{j=t_0}^{k-1} \gamma_2\lambda^{k-1-j}|r(j)|, \quad k \geq t_0.$$

Case 2: $d(t) \neq 0$ for some $t \geq t_0$.

This case is much more involved since noise can radically affect parameter estimation. Indeed, even if the parameter estimate is quite accurate at a point in time, the introduction of a large noise signal (large relative to the size of $\phi(t)$) can create a highly inaccurate parameter estimate. To proceed, we partition the timeline into two parts: one in which the noise is small versus ϕ and one where it is not; the actual choice of the line of division will become clear as the proof progresses. To this end, with $\varepsilon > 0$ to be chosen shortly, partition $\{j \in \mathbf{Z} : j \geq t_0\}$ into two sets:

$$S_{\text{good}} := \left\{ j \geq t_0 : \phi(j) \neq 0 \text{ and } \frac{[d(j)]^2}{\|\phi(j)\|^2} < \varepsilon \right\},$$

$$S_{\text{bad}} := \left\{ j \geq t_0 : \phi(j) = 0 \text{ or } \frac{[d(j)]^2}{\|\phi(j)\|^2} \geq \varepsilon \right\};$$

clearly $\{j \in \mathbf{Z} : \quad j \geq t_0\} = S_{\text{good}} \cup S_{\text{bad}}$. Observe that this partition clearly depends on $\theta_0, \theta^*, \phi_0, d$ and r/y^*. We will apply Proposition 2 to analyse the closed-loop system behaviour on S_{good}; on the other hand, we will easily obtain bounds on the system behaviour on S_{bad}. Before doing so, we partition the time index $\{j \in \mathbf{Z} : j \geq t_0\}$ into intervals which oscillate between S_{good} and S_{bad}. To this end, it is easy to see that we can define a (possibly infinite) sequence of intervals of the form $[k_i, k_{i+1})$ satisfying:

(i) $k_1 = t_0$, and
(ii) $[k_i, k_{i+1})$ either belongs to S_{good} or S_{bad}, and
(iii) if $k_{i+1} \neq \infty$ and $[k_i, k_{i+1})$ belongs to S_{good} (respectively, S_{bad}), then the interval $[k_{i+1}, k_{i+2})$ must belong to S_{bad} (respectively, S_{good}).

Now we turn to analysing the behaviour during each interval.
Sub-Case 2.1: $[k_i, k_{i+1})$ lies in S_{bad}.

Let $j \in [k_i, k_{i+1})$ be arbitrary. In this case either $\phi(j) = 0$ or $\frac{[d(j)]^2}{\|\phi(j)\|^2} \geq \varepsilon$ holds. In either case we have

$$\|\phi(j)\| \leq \frac{1}{\varepsilon^{1/2}} |d(j)|, \ \ j \in [k_i, k_{i+1}). \tag{24}$$

From (13) and (17) we see that

$$\begin{aligned} \|\phi(j+1)\| &\leq \bar{a}\|\phi(j)\| + (2\|\mathcal{S}\| \times \|\phi(j)\| + |d(j)| + |r(j)|) \\ &\leq [1 + \underbrace{(\bar{a} + 2\|\mathcal{S}\|)}_{=:\gamma_3} \frac{1}{\varepsilon^{1/2}}]|d(j)| + |r(j)|, \ \ j \in [k_i, k_{i+1}). \end{aligned} \tag{25}$$

If we combine this with (24), we conclude that

$$\|\phi(j)\| \leq \begin{cases} \frac{1}{\varepsilon^{1/2}}|d(j)| & j = k_i \\ (1 + \frac{\gamma_3}{\varepsilon^{1/2}})|d(j-1)| + |r(j-1)| & j = k_i + 1, ..., k_{i+1}. \end{cases} \tag{26}$$

Sub-Case 2.2: $[k_i, k_{i+1})$ lies in S_{good}.

Let $j \in [k_i, k_{i+1})$ be arbitrary. In this case $\phi(j) \neq 0$ and

$$\frac{[d(j)]^2}{\|\phi(j)\|^2} < \varepsilon, \quad j \in [k_i, k_{i+1}),$$

which implies that

$$\rho_\delta(\phi(j), e(j+1)) \frac{d(j)^2}{\|\phi(j)\|^2} < \varepsilon, \quad j \in [k_i, k_{i+1}). \tag{27}$$

From Proposition 1 we have that

$$V(\bar{k}) \leq V(\underline{k}) + \sum_{j=\underline{k}}^{\bar{k}-1} \rho_\delta(\phi(j), e(j+1)) \frac{-\frac{1}{2}e(j+1)^2 + 2d(j)^2}{\|\phi(j)\|^2}, \quad k_i \leq \underline{k} < \bar{k} \leq k_{i+1};$$

using (27) and the fact that $0 \leq V(\cdot) \leq 4\|\mathcal{S}\|^2$, we obtain

$$\begin{aligned}\sum_{j=\underline{k}}^{\bar{k}-1} \rho_\delta(\phi(j), e(j+1)) \frac{e(j+1)^2}{\|\phi(j)\|^2} &\leq 2V(\underline{k}) + 2\sum_{j=\underline{k}}^{\bar{k}-1} \rho_\delta(\phi(j), e(j+1)) \frac{2d(j)^2}{\|\phi(j)\|^2} \\ &\leq 8\|\mathcal{S}\|^2 + 4\varepsilon(\bar{k} - \underline{k}), \quad k_i \leq \underline{k} < \bar{k} \leq k_{i+1}.\end{aligned}$$

Hence, using this in (20) and (21) yields

$$\begin{aligned}\sum_{j=\underline{k}}^{\bar{k}-1} \|\bar{A}(j+1) - \bar{A}(j)\| &\leq \gamma_1[8\|\mathcal{S}\|^2 + 4\varepsilon(\bar{k} - \underline{k})]^{1/2}(\bar{k} - \underline{k})^{1/2} \\ &\leq \gamma_1 8^{1/2}\|\mathcal{S}\|(\bar{k} - \underline{k})^{1/2} + 2\gamma_1\varepsilon^{1/2}(\bar{k} - \underline{k}), \quad k_i \leq \underline{k} < \bar{k} \leq k_{i+1},\end{aligned}$$

as well as

$$\begin{aligned}\sum_{j=\underline{k}}^{\bar{k}-1} \|\Delta(j)\| &\leq [8\|\mathcal{S}\|^2 + 4\varepsilon(\bar{k} - \underline{k})]^{1/2}(\bar{k} - \underline{k})^{1/2} \\ &\leq 8^{1/2}\|\mathcal{S}\|(\bar{k} - \underline{k})^{1/2} + 2\varepsilon^{1/2}(\bar{k} - \underline{k}), \quad k_i \leq \underline{k} < \bar{k} \leq k_{i+1}.\end{aligned}$$

Now we will apply Proposition 2: we set

$$\begin{aligned}&\alpha_0 = \beta_0 = 0, \quad \alpha_1 = \gamma_1 8^{1/2}\|\mathcal{S}\|, \quad \beta_1 = 8^{1/2}\|\mathcal{S}\|, \\ &\alpha_2 = 2\gamma_1\varepsilon^{1/2}, \quad \beta_2 = 2\varepsilon^{1/2}, \quad \mu = \lambda.\end{aligned}$$

With N chosen as in Case 1 via (23), we have that $\underline{\delta} := \frac{\lambda}{\gamma_1^{1/N}} - \lambda_1 > 0$; we need

$$\alpha_2 + \frac{\beta_2}{N} < \frac{1}{N\gamma_1}\underline{\delta},$$

which will certainly be the case if we set $\varepsilon := \frac{\underline{\delta}^2}{8\gamma_1^2(\gamma_1 N+1)^2}$. From Proposition 2 we see that there exists a constant γ_4 so that the state transition matrix $\Phi(t, \tau)$ corresponding to $\bar{A}(t) + \Delta(t)$ satisfies

$$\|\Phi(t, \tau)\| \leq \gamma_4 \lambda^{t-\tau}, \ k_i \leq \tau \leq t \leq k_{i+1}.$$

If we now apply this to (16) and use (18) to provide a bound on $\eta(t)$, we end up with

$$\|\phi(k)\| \leq \gamma_4 \lambda^{k-k_i} \|\phi(k_i)\| + \left(2\frac{\|\mathcal{S}\|}{\delta} + 1\right) \sum_{j=k_i}^{k-1} \gamma_4 \lambda^{k-1-j} (|r(j)| + |d(j)|), \quad k_i \leq k \leq k_{i+1}. \tag{28}$$

This completes Sub-Case 2.2.

Now we combine Sub-Case 2.1 and Sub-Case 2.2 into a general bound on $\phi(t)$. Define

$$\gamma_5 := \max\left\{1, 1 + \frac{\gamma_3}{\varepsilon^{1/2}}, \gamma_4, \gamma_4\left(2\frac{\|\mathcal{S}\|}{\delta} + 2 + \frac{\gamma_3}{\varepsilon^{1/2}}\right)\right\}.$$

It remains to prove

Claim The following bound holds:

$$\|\phi(k)\| \leq \gamma_5 \lambda^{k-t_0} \|\phi_0\| + \sum_{j=t_0}^{k-1} \gamma_5 \lambda^{k-1-j} (|r(j)| + |d(j)|), \quad k \geq t_0. \tag{29}$$

Proof of the Claim If $[k_1, k_2) = [t_0, k_2) \subset S_{\text{good}}$, then (29) holds for $k \in [t_0, k_2]$ by (28). If $[t_0, k_2) \subset S_{\text{bad}}$, then from (26) we obtain

$$\|\phi(j)\| \leq \begin{cases} \|\phi(k_1)\| = \|\phi_0\| & j = k_1 = t_0 \\ (1 + \frac{\gamma_3}{\varepsilon^{1/2}})|d(j-1)| + |r(j-1)| & j = k_1 + 1, ..., k_2, \end{cases}$$

which means that (29) holds for $k \in [t_0, k_2]$ for this case as well.

We now use induction—suppose that (29) holds for $k \in [k_1, k_i]$; we need to prove that it holds for $k \in (k_i, k_{i+1}]$ as well. If $[k_i, k_{i+1}) \subset S_{\text{bad}}$ then from (26) we have

$$\|\phi(j)\| \leq \left(1 + \frac{\gamma_3}{\varepsilon^{1/2}}\right) |d(j-1)| + |r(j-1)|, \ \ j = k_i + 1, ..., k_{i+1},$$

which means that (29) holds for $k \in (k_i, k_{i+1}]$. On the other hand, if $[k_i, k_{i+1}) \subset S_{\text{good}}$, then $k_i - 1 \in S_{\text{bad}}$; from (26) we have that

$$\|\phi(k_i)\| \leq \left(1 + \frac{\gamma_3}{\varepsilon^{1/2}}\right) |d(k_i - 1)| + |r(k_i - 1)|.$$

Using (28) to analyse the behaviour on $[k_i, k_{i+1}]$, we have

$$\begin{aligned} \|\phi(k)\| &\leq \gamma_4 \lambda^{k-k_i} \|\phi(k_i)\| + \left(2\frac{\|\mathcal{S}\|}{\delta} + 1\right) \gamma_4 \sum_{j=k_i}^{k-1} \lambda^{k-1-j} (|r(j)| + |d(j)|) \\ &\leq \gamma_4 \lambda^{k-k_i} [(1 + \frac{\gamma_3}{\varepsilon^{1/2}})|d(k_i - 1)| + |r(k_i - 1)|] + \end{aligned}$$

$$+\gamma_4\left(2\frac{\|\mathcal{S}\|}{\delta}+1\right)\sum_{j=k_i}^{k-1}\lambda^{k-1-j}(|r(j)|+|d(j)|)$$

$$\leq\left[\gamma_4\left(1+\frac{\gamma_3}{\varepsilon^{1/2}}\right)+\gamma_4(2\frac{\|\mathcal{S}\|}{\delta}+1)\right]\sum_{j=k_i-1}^{k-1}\lambda^{k-1-j}(|r(j)|+|d(j)|)$$

$$\leq\gamma_5\sum_{j=t_0}^{k-1}\lambda^{k-1-j}(|d(j)|+|r(j)|),\quad k=k_i+1,\ldots,k_{i+1},$$

as desired. □

This completes the proof. □

5 Tolerance to time variations

The linear-like bound proven in Theorem 1 can be leveraged to prove that the same behaviour will result even in the presence of slow time variations with occasional jumps. So suppose that the actual plant model is

$$y(t+1)=\phi(t)^T\theta^*(t)+d(t),\ \phi(t_0)=\phi_0, \tag{30}$$

with $\theta^*(t)\in\mathcal{S}$ for all $t\in\mathbf{R}$. We adopt a common model of acceptable time variations used in adaptive control: with $c_0\geq 0$ and $\varepsilon>0$, we let $s(\mathcal{S},c_0,\varepsilon)$ denote the subset of $l_\infty(\mathbf{R}^{2n})$ whose elements θ^* satisfy $\theta^*(t)\in\mathcal{S}$ for every $t\in\mathbf{Z}$ as well as

$$\sum_{t=t_1}^{t_2-1}\|\theta^*(t+1)-\theta^*(t)\|\leq c_0+\varepsilon(t_2-t_1),\quad t_2>t_1 \tag{31}$$

for every $t_1\in\mathbf{Z}$. We will now show that, for every $c_0\geq 0$, the approach tolerates time-varying parameters in $s(\mathcal{S},c_0,\varepsilon)$ if ε is small enough.

Theorem 2 *For every $\delta\in(0,\infty]$, $\lambda_1\in(\underline{\lambda},1)$ and $c_0\geq 0$, there exists a $c_1>0$ and $\varepsilon>0$ so that for every $t_0\in\mathbf{Z}$, $\theta_0\in\mathcal{S}$, $\theta^*\in s(\mathcal{S},c_0,\varepsilon)$, $\phi_0\in\mathbf{R}^{2n}$, and $y^*,d\in\ell_\infty$, when the adaptive controller (7), (8) and (11) is applied to the time-varying plant (30), the following holds:*

$$\|\phi(k)\|\leq c_1\lambda_1^{k-t_0}\|\phi_0\|+\sum_{j=t_0}^{k-1}c_1\lambda_1^{k-1-j}(|r(j)|+|d(j)|),\quad k\geq t_0.$$

Proof Fix $\delta\in(0,\infty]$, $\lambda_1\in(\underline{\lambda},1)$, $\lambda\in(\underline{\lambda},\lambda_1)$ and $c_0>0$. Let $t_0\in\mathbf{Z}$, $\theta_0\in\mathcal{S}$, $\phi_0\in\mathbf{R}^{2n}$, and $y^*,d\in\ell_\infty$ be arbitrary. With $m\in\mathbf{N}$, we will consider $\phi(t)$ on intervals of the form $[t_0+im,t_0+(i+1)m]$; we will be analysing these intervals in groups of m (to be chosen shortly); we set $\varepsilon = \frac{c_0}{m^2}$, and let $\theta^* \in s(\mathcal{S}, c_0, \varepsilon)$ be arbitrary.

First of all, for $i \in \mathbf{Z}^+$ we can rewrite the plant equation as

$$y(t+1) = \phi(t)^T \theta^*(t_0+im) + d(t) + \underbrace{\phi(t)^T[\theta^*(t) - \theta^*(t_0+im)]}_{=:\tilde{n}(t)},$$
$$t \in [t_0+im, t_0+(i+1)m]. \tag{32}$$

Theorem 1 applied to (32) says that there exists a constant $c > 0$ so that

$$\|\phi(t)\| \le c\lambda^{t-t_0-im}\|\phi(t_0+im)\| + \sum_{j=t_0+im}^{t-1} c\lambda^{t-1-j}(|r(j)| + |d(j)| + |\tilde{n}(j)|),$$
$$t \in [t_0+im, t_0+(i+1)m].$$

The above is a difference inequality associated with a first-order system; using this observation together with the fact that $c \ge 1$, we see that if we define

$$\psi(t+1) = \lambda\psi(t) + |r(t)| + |d(t)| + |\tilde{n}(t)|, \quad t \in [t_0+im, t_0+(i+1)m-1],$$

with $\psi(t_0+im) = \|\phi(t_0+im)\|$, then

$$\|\phi(t)\| \le c\psi(t), \quad t \in [t_0+im, t_0+(i+1)m].$$

Now we analyse this equation for $i = 0, 1, ..., m-1$.

Case 1: $|\tilde{n}(t)| \le \frac{1}{2c}(\lambda_1 - \lambda)\|\phi(t)\|$ for all $t \in [t_0+im, t_0+(i+1)m]$.

In this case

$$\begin{aligned}
\psi(t+1) &\le \lambda\psi(t) + |r(t)| + |d(t)| + |\tilde{n}(t)| \\
&\le \lambda\psi(t) + |r(t)| + |d(t)| + \frac{1}{2c}(\lambda_1 - \lambda)c\psi(t) \\
&\le \left(\frac{\lambda+\lambda_1}{2}\right)\psi(t) + |r(t)| + |d(t)|, \ t \in [t_0+im, t_0+(i+1)m],
\end{aligned}$$

which means that

$$|\psi(t)| \le \left(\frac{\lambda+\lambda_1}{2}\right)^{t-t_0-im}|\psi(t_0+im)| + \sum_{j=t_0+im}^{t-1}\left(\frac{\lambda+\lambda_1}{2}\right)^{t-1-j}(|r(j)| + |d(j)|),$$
$$t = t_0+im, ..., t_0+(i+1)m.$$

This, in turn, implies that

$$\|\phi(t_0+(i+1)m)\| \le c\left(\frac{\lambda+\lambda_1}{2}\right)^m \|\phi(t_0+im)\| + \sum_{j=t_0+im}^{t_0+(i+1)m-1} c\left(\frac{\lambda+\lambda_1}{2}\right)^{t_0+(i+1)m-1-j} (|r(j)|+|d(j)|). \tag{33}$$

Case 2: $|\tilde{n}(t)| > \frac{1}{2c}(\lambda_1-\lambda)\|\phi(t)\|$ for some $t \in [t_0+im, t_0+(i+1)m]$.
Since $\theta^*(t) \in \mathcal{S}$ for $t \ge t_0$, we see

$$|\tilde{n}(t)| \le 2\|\mathcal{S}\| \times \|\phi(t)\|,\ t \in [t_0+im, t_0+(i+1)m].$$

This means that

$$\begin{aligned}\psi(t+1) &\le \lambda\psi(t)+|r(t)|+|d(t)|+|\tilde{n}(t)| \\ &\le \lambda\psi(t)+|r(t)|+|d(t)|+2\|\mathcal{S}\|c\psi(t) \\ &\le \underbrace{(1+2c\|\mathcal{S}\|)}_{=:\gamma_1}\psi(t)+|r(t)|+|d(t)|,\ t \in [t_0+im, t_0+(i+1)m],\end{aligned}$$

which means that

$$|\psi(t)| \le \gamma_1^{t-t_0-im}\|\psi(t_0+im)\| + \sum_{j=t_0+im}^{t-1} \gamma_1^{t-j-1}(|r(j)|+|d(j)|),\ t=t_0+im, ..., t_0+(i+1)m.$$

This, in turn, implies that

$$\begin{aligned}\|\phi(t_0+(i+1)m)\| &\le c\gamma_1^m\|\phi(t_0+im)\| \\ &\quad +c\sum_{j=t_0+im}^{t_0+(i+1)m-1}(\gamma_1)^{t_0+(i+1)m-j-1}(|r(j)|+|d(j)|) \\ &\le c\gamma_1^m\|\phi(t_0+im)\| \\ &\quad +c\left(\frac{2\gamma_1}{\lambda+\lambda_1}\right)^m \sum_{j=t_0+im}^{t_0+(i+1)m-1} \\ &\quad \left(\frac{\lambda+\lambda_1}{2}\right)^{t_0+(i+1)m-j-1}(|r(j)|+|d(j)|).\end{aligned} \tag{34}$$

On the interval $[t_0, t_0+m^2]$ there are m subintervals of length m; furthermore, because of the choice of ε we have that

$$\sum_{j=t_0}^{t_0+m^2-1} \|\theta^*(j+1)-\theta^*(j)\| \le c_0+\varepsilon m^2 \le 2c_0.$$

A simple calculation reveals that there are at most $N_1 := \frac{4c_0 c}{\lambda_1 - \lambda}$ subintervals which fall into the category of Case 2, with the remaining number falling into the category of Case 1. Henceforth, we assume that $m > N_1$. If we use (33) and (34) to analyse the behaviour of the closed-loop system on the interval $[t_0, t_0 + m^2]$, we end up with a crude bound of

$$\|\phi(t_0 + m^2)\| \leq c^m \gamma_1^{N_1 m} \left(\frac{\lambda_1 + \lambda}{2}\right)^{m(m-N_1)} \|\phi(t_0)\|$$
$$+ (\frac{2\gamma_1}{\lambda+\lambda_1})^m (c\gamma_1^m)^m \left(\frac{2}{\lambda+\lambda_1}\right)^{m^2} \sum_{j=t_0}^{t_0+m^2-1} \left(\frac{\lambda_1+\lambda}{2}\right)^{t_0+m^2-j-1} (|r(j)| + |d(j)|). \tag{35}$$

At this point we would like to choose m so that

$$c^m \gamma_1^{N_1 m} \left(\frac{\lambda_1 + \lambda}{2}\right)^{m^2 - mN_1} \leq \lambda_1^{m^2} \Leftrightarrow c^m \gamma_1^{N_1 m} \left(\frac{2}{\lambda + \lambda_1}\right)^{mN_1} \leq \left(\frac{2\lambda_1}{\lambda_1 + \lambda}\right)^{m^2};$$

notice that $\frac{2\lambda_1}{\lambda_1+\lambda} > 1$, so if we take the log of both sides, we see that we need

$$m \ln(c) + N_1 m \ln(\gamma_1) + N_1 m \ln\left(\frac{2}{\lambda + \lambda_1}\right) \leq m^2 \ln\left(\frac{2\lambda_1}{\lambda_1 + \lambda}\right),$$

which will clearly be the case for large enough m, so at this point we choose such an m. It follows from (35) that there exists a constant γ_2 so that

$$\|\phi(t_0 + m^2)\| \leq \lambda_1^{m^2} \|\phi(t_0)\| + \gamma_2 \sum_{j=t_0}^{t_0+m^2-1} \lambda_1^{t_0+m^2-j-1} (|r(j)| + |d(j)|).$$

Indeed, the same bound holds regardless of the interval of analysis:

$$\|\phi(\bar{t} + m^2)\| \leq \lambda_1^{m^2} \|\phi(\bar{t})\| + \gamma_2 \sum_{j=\bar{t}}^{\bar{t}+m^2-1} \lambda_1^{\bar{t}+m^2-j-1} (|r(j)| + |d(j)|), \quad \bar{t} \geq t_0.$$

Solving iteratively yields

$$\|\phi(t_0 + im^2)\|$$
$$\leq \lambda_1^{im^2} \|\phi(t_0)\| + \gamma_2 \sum_{j=t_0}^{t_0+im^2-1} \lambda_1^{t_0+im^2-j-1} (|r(j)| + |d(j)|), \quad i \geq 0. \tag{36}$$

We now combine this bound with the bounds which hold on the good intervals (33) and the bad intervals (34) and conclude that there exists a constant γ_3 so that

$$\|\phi(t)\| \leq \gamma_3 \lambda_1^{t-t_0} \|\phi(t_0)\| + \gamma_3 \sum_{j=t_0}^{t-1} \lambda_1^{t-j-1} (|r(j)| + |d(j)|), \quad t \geq t_0,$$

as desired. □

6 Tolerance to unmodelled dynamics

Due to the linear-like bounds proven in Theorems 1 and 2, we can use the Small Gain Theorem to good effect to prove the tolerance of the closed-loop system to unmodelled dynamics. However, since the controller, and therefore the closed-loop system, is nonlinear, handling initial conditions is more subtle: in the linear time-invariant case we can separate out the effect of initial conditions from that of the forcing functions (r/y^* and d), but in our situation they are intertwined. We proceed by looking at two cases—with and without initial conditions. In all of the cases we consider the time-varying plant (30) with $d_\Delta(t)$ added to represent the effect of unmodelled dynamics:

$$y(t+1) = \phi(t)^T \theta^*(t) + d(t) + d_\Delta(t), \quad \phi(t_0) = \phi_0. \tag{37}$$

To proceed, fix $\delta \in (0, \infty]$, $\lambda_1 \in (\underline{\lambda}, 1)$ and $c_0 \geq 0$; from Theorem 2 there exists a $c_1 > 0$ and $\varepsilon > 0$ so that for every $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\theta_0 \in \mathcal{S}$, $y^*, d \in \ell_\infty$, and $\theta^* \in s(\mathcal{S}, c_0, \varepsilon)$, when the adaptive controller (7), (8) and (11) is applied to the time-varying plant (37), the following bound holds:

$$\|\phi(k)\| \leq c_1 \lambda_1^{k-t_0} \|\phi_0\| + \sum_{j=t_0}^{k-1} c_1 \lambda_1^{k-1-j} (|r(j)| + |d(j)| + |d_\Delta(j)|), \quad k \geq t_0. \tag{38}$$

6.1 Zero initial conditions

In this case we assume that $\phi(t) = 0$ for $t \leq t_0$; we derive a bound on the closed-loop system behaviour in the presence of unmodelled dynamics. Suppose that the unmodelled dynamics is of the form $d_\Delta(t) = (\Delta\phi)(t)$ with $\Delta : l_\infty(\mathbf{R}^{2n}) \to l_\infty(\mathbf{R}^{2n})$ a (possibly nonlinear time-varying) causal map with a finite gain of $\|\Delta\|$. It is easy to prove that if $\|\Delta\| < \frac{1-\lambda_1}{c_1}$, then

$$\|\phi(k)\| \leq \frac{c_1}{1 - \lambda_1 - c_1\|\Delta\|} (\sup_{t \geq t_0} \|r(t)\| + \sup_{t \geq t_0} \|d(t)\|), \quad k \geq t_0,$$

i.e. a form of closed-loop stability is attained. Following the approach of Remark 7, we could also analyse the closed-loop system using l_p-norms with $1 \leq p < \infty$.

6.2 Nonzero initial conditions

Now we consider the case of unmodelled LTI dynamics when the plant has nonzero initial conditions, and we develop convolution-like bounds on the closed-loop system. To this end suppose that the unmodelled dynamics are of the form

$$d_{\Delta}(t) := \sum_{j=0}^{\infty} \Delta_j \phi(t-j), \tag{39}$$

with $\Delta_j \in \mathbf{R}^{1\times 2n}$; the corresponding transfer function is $\Delta(z^{-1}) := \sum_{j=0}^{\infty} \Delta_j z^{-j}$. It is easy to see that this model subsumes the classical additive uncertainty, multiplicative uncertainty, and uncertainty in a coprime factorization, which is common in the robust control literature, e.g. see [39], with the only constraint being that the perturbations correspond to strictly causal terms. In order to obtain linear-like bounds on the closed-loop behaviour, we need to impose more constraints on $\Delta(z)$ than in the previous subsection: after all, if $\Delta(z^{-1}) = \Delta_p z^{-p}$, it is clear that $\|\Delta\| = \|\Delta_p\|$ for all p, but the effect on the closed-loop system varies greatly—a large value of p allows the behaviour in the far past to affect the present. To this end, with $\mu > 0$ and $\beta \in (0, 1)$, we shall restrict $\Delta(z^{-1})$ to a set of the form

$$\mathcal{B}(\mu, \beta) := \left\{ \sum_{j=0}^{\infty} \Delta_j z^{-j} : \Delta_j \in \mathbf{R}^{1\times 2n} \text{ and } \|\Delta_j\| \le \mu\beta^j,\ j \ge 0 \right\}.$$

It is easy to see that every transfer function in $\mathcal{B}(\mu, \beta)$ is analytic in $\{z \in \mathbf{C} : |z| > \beta\}$, so it has no poles in that region.

Now we fix $\mu > 0$ and $\beta \in (0, 1)$ and let $\Delta(z^{-1})$ belong to $\mathcal{B}(\mu, \beta)$; the goal is to analyse the closed-loop behaviour of (37) for $t \ge t_0$ when d_Δ is given by (39). We first partition $d_\Delta(t)$ into two parts—that which depends on $\phi(t)$ for $t \ge t_0$ and that which depends on $\phi(t)$ for $t < t_0$:

$$d_{\Delta}(t) = \sum_{j=0}^{\infty} \Delta_j \phi(t-j) = \sum_{j=-\infty}^{t} \Delta_{t-j}\phi(j) = \underbrace{\sum_{j=-\infty}^{t_0-1} \Delta_{t-j}\phi(j)}_{=:d_{\Delta}^{-}(t)} + \underbrace{\sum_{j=t_0}^{t} \Delta_{t-j}\phi(j)}_{=:d_{\Delta}^{+}(t)}.$$

It is clear that

$$\|d_{\Delta}^{+}(t)\| \le \sum_{j=t_0}^{t} \mu\beta^{t-j}\|\phi(j)\|,$$

$$\|d_{\Delta}^{-}(t)\| \le \sum_{j=-\infty}^{t_0-1} \mu\beta^{t-j}\|\phi(j)\| = \mu\beta^{t-t_0}\sum_{j=1}^{\infty}\beta^j\|\phi(t_0-j)\|, \quad t \ge t_0.$$

If $\phi(t)$ is bounded on $\{t \in \mathbf{Z} : t < t_0\}$ then $\sum_{j=1}^{\infty} \beta^j \|\phi(t_0 - j)\|$ is finite, in which case we see that $d_\Delta^-(t)$ goes to zero exponentially fast; henceforth, we make the reasonable assumption that this is the case. It turns out that we can easily bound $d_\Delta(t)$ with a difference equation. To this end, consider

$$m(t+1) = \beta m(t) + \beta\|\phi(t)\|, \quad t \geq t_0, \tag{40}$$

with $m(t_0) = m_0 := \sum_{j=1}^{\infty} \beta^j \|\phi(t_0 - j)\|$; it is straightforward to prove that

$$|d_\Delta(t)| \leq |d_\Delta^+(t)| + |d_\Delta^-(t)| \leq \mu m(t) + \mu\|\phi(t)\|, \quad t \geq t_0. \tag{41}$$

This model of unmodelled dynamics is similar to that used in the adaptive control literature, e.g. see [10].

Theorem 3 *For every $\beta \in (0, 1)$ and $\lambda_2 \in (\max\{\lambda_1, \beta\}, 1)$, there exist $\bar{\mu} > 0$ and $c_2 > 0$ so that for every $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $m_0 \in \mathbf{R}$, $\theta_0 \in \mathcal{S}$, $y^*, d \in l_\infty$, $\theta^* \in s(\mathcal{S}, c_0, \varepsilon)$ and $\mu \in (0, \bar{\mu})$, when the adaptive controller (7), (8) and (11) is applied to the time-varying plant (37) with d_Δ satisfying (40) and (41), the following bound holds:*

$$\|\phi(k)\| \leq c_2 \lambda_2^{k-t_0}(\|\phi_0\| + |m_0|) + \sum_{j=t_0}^{k-1} c_2 \lambda_2^{k-1-j}(|d(j)| + |r(j)|), \quad k \geq t_0.$$

Proof Fix $\beta \in (0, 1)$ and $\lambda_2 \in (\max\{\lambda_1, \beta\}, 1)$. The first step is to convert difference inequalities to difference equations. To this end, consider the difference equation

$$\begin{aligned}\tilde{\phi}(t+1) = \lambda_1 \tilde{\phi}(t) + c_1|r(t)| + c_1|d(t)| \\ + c_1\mu\tilde{m}(t) + c_1\mu\tilde{\phi}(t), \ \tilde{\phi}(t_0) = c_1\|\phi(t_0)\|,\end{aligned} \tag{42}$$

together with the difference equation based on (40):

$$\tilde{m}(t+1) = \beta\tilde{m}(t) + \beta\tilde{\phi}(t), \quad \tilde{m}(t_0) = |m_0|. \tag{43}$$

It is easy to use induction together with (38), (40), and (41) to prove that

$$\|\phi(t)\| \leq \tilde{\phi}(t), \quad |m(t)| \leq \tilde{m}(t), \quad t \geq t_0. \tag{44}$$

If we combine the difference equations (42) with (43), we end up with

$$\begin{bmatrix} \tilde{\phi}(t+1) \\ \tilde{m}(t+1) \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_1 + c_1\mu & c_1\mu \\ \beta & \beta \end{bmatrix}}_{A_{cl}(\mu)} \begin{bmatrix} \tilde{\phi}(t) \\ \tilde{m}(t) \end{bmatrix} + \begin{bmatrix} c_1 \\ 0 \end{bmatrix} (|d(t)| + |r(t)|), \quad t \geq t_0. \tag{45}$$

Now we see that $A_{cl}(\mu) \to \begin{bmatrix} \lambda_1 & 0 \\ \beta & \beta \end{bmatrix}$ as $\mu \to 0$, and this matrix has eigenvalues of $\{\lambda_1, \beta\}$. Now choose $\bar{\mu} > 0$ so that all eigenvalues are less than $(\frac{\lambda_2}{2} + \frac{1}{2}\max\{\lambda_1, \beta\})$ in magnitude for $\mu \in (0, \bar{\mu}]$, and define $\varepsilon := \frac{\lambda_2}{2} - \frac{1}{2}\max\{\lambda_1, \beta\}$. Using the proof technique of Desoer in [40], we can conclude that for $\mu \in (0, \bar{\mu}]$, we have

$$\|A_{cl}(\mu)^k\| \le \underbrace{\left(\frac{3 + 2\beta + 2c_1\bar{\mu}}{\varepsilon^2}\right)}_{=:\gamma_1} \lambda_2^k, \quad k \ge 0;$$

if we use this in (45) and then apply the bounds in (44), it follows that

$$\|\phi(k)\| \le c_1\gamma_1\lambda_2^{k-t_0}(\|\phi_0\| + |m_0|) + \sum_{j=t_0}^{k-1} c_1\gamma_1\lambda_2^{k-1-j}(|d(j)| + |r(j)|), \quad k \ge t_0,$$

as desired. □

7 Step tracking

If the plant is non-minimum phase, it is not possible to track an arbitrary bounded reference signal using a bounded control signal. However, as long as the plant does not have a zero at $z = 1$, it is possible to modify the controller design procedure to achieve asymptotic step tracking if there is no noise/disturbance. So at this point assume that the corresponding plant polynomial $B(z^{-1})$ has no zero at $z = 1$ for any plant model $\theta^* \in \mathcal{S}$. To proceed, we use the standard trick from the literature, e.g. see [34]: we still estimate $A(z^{-1})$ and $B(z^{-1})$ as before, but we now design the control law slightly differently. To this end, we first define

$$\tilde{A}(t, z^{-1}) := (1 - z^{-1})\hat{A}(t, z^{-1}),$$

and then let $A^*(z^{-1})$ be a $2(n+1)$th monic polynomial (rather than a $2n$th one) of the form

$$A^*(z^{-1}) = 1 + a_1^* z^{-1} + \cdots + a_{2n+2}^* z^{-2n-2}$$

so that $z^{2(n+1)}A^*(z^{-1})$ has all of its zeros in $\mathbf{D}^o$. Next, we choose two polynomial

$$\tilde{L}(t, z^{-1}) = 1 + \tilde{l}_1(t)z^{-1} + \cdots + \tilde{l}_{n+1}(t)z^{-n-1}$$
$$\text{and } \hat{P}(t, z^{-1}) = \hat{p}_1(t)z^{-1} + \cdots + \hat{p}_{n+1}(t)z^{-n-1}$$

which satisfy the equation

$$\tilde{A}(t, z^{-1})\tilde{L}(t, z^{-1}) + \hat{B}(t, z^{-1})\hat{P}(t, z^{-1}) = A^*(z^{-1}); \tag{46}$$

since $\tilde{A}(t, z^{-1})$ and $\hat{B}(t, z^{-1})$ are coprime, there exist **unique** $\tilde{L}(t, z^{-1})$ and $\hat{P}(t, z^{-1})$ which satisfy this equation. We now define

$$\hat{L}(t, z^{-1}) = 1 + \hat{l}_1(t)z^{-1} + \cdots + \hat{l}_{n+2}(t)z^{-n-2} := (1 - z^{-1})\tilde{L}(t, z^{-1});$$

at time t we choose $u(t)$ so that

$$\begin{aligned} u(t) = &-\hat{l}_1(t-1)u(t-1) - \cdots - \hat{l}_{n+2}(t-1)u(t-n-2) \\ &-\hat{p}_1(t-1)[y(t-1) - y^*(t-1)] \\ &- \cdots - \hat{p}_{n+1}(t-1)[y(t-n-1) - y^*(t-n-1)]. \end{aligned}$$

We can use a modified version of the argument used in the proof of Theorem 1 to conclude that a similar type of result holds here; we can also prove that asymptotic step tracking will be attained if the noise is zero and the reference signal y^* is constant. The details are omitted.

8 Relaxing the convexity requirement

The convexity and coprimeness assumptions on the set of admissible plant parameters play a crucial role in obtaining the nice closed-loop properties provided in Theorems 1–3. Here we will show that it is possible to weaken the convexity requirement if the goal is to place all closed-loop poles at zero, although it is at the expense of using a more complicated controller. Our proposed approach is modelled on the first-order one-step-ahead control setup (see [32,41]) which is deadbeat in nature; of course, here the plant may not be first order, which increases the complexity. While we would like to remove the convexity requirement completely, at present we are only able to weaken it. So in this section we replace **Assumption 1** with

Assumption 2: $\mathcal{S} \subset \mathcal{S}_1 \cup \mathcal{S}_2$ with $\mathcal{S}_1$ and $\mathcal{S}_2$ convex and compact, and for each $\theta \in \mathcal{S}_1 \cup \mathcal{S}_2$, the corresponding pair of polynomials $A(z^{-1})$ and $B(z^{-1})$ are coprime.

The idea is to use an estimator for each of $\mathcal{S}_1$ and $\mathcal{S}_2$, and at each point in time we choose which one to use in the control law. Before proceeding, define

$$\bar{\mathbf{s}} := \max\{\|\mathcal{S}_1\|, \|\mathcal{S}_2\|\}.$$

8.1 Parameter estimation

For each $\mathcal{S}_i$ and $\hat{\theta}_i(t_0) \in \mathcal{S}_i$, we construct an estimator which generates an estimate $\hat{\theta}_i(t) \in \mathcal{S}_i$ at each $t > t_0$. Motivated by (4) and (5), the associated prediction error is defined as

$$e_i(t+1) = y(t+1) - \phi(t)^T \hat{\theta}_i(t), \tag{47}$$

and the parameter update law is given by

$$\check{\theta}_i(t+1) = \begin{cases} \hat{\theta}_i(t) + \frac{\phi(t)}{\|\phi(t)\|^2} e_i(t+1) & \text{if } \phi(t) \neq 0 \\ \hat{\theta}_i(t) & \text{if } \phi(t) = 0, \end{cases} \tag{48}$$

$$\hat{\theta}_i(t+1) = \pi_{\mathcal{S}_i}(\check{\theta}_i(t+1)). \tag{49}$$

(For ease of exposition, we do not use the more general version of the estimator equation in (7) and (8).) Associated with this estimator is the parameter estimation error $\tilde{\theta}_i(t) := \hat{\theta}_i(t) - \theta^*$ as well as the corresponding Lyapunov function $V_i(t) := \tilde{\theta}_i(t)^T \tilde{\theta}_i(t)$.

8.2 The switching control law

The elements of $\hat{\theta}_i(t)$ are partitioned as

$$\hat{\theta}_i(t) =: [-\hat{a}_{i,1}(t) \ \ \cdots \ \ -\hat{a}_{i,n}(t) \ \ \hat{b}_{i,1}(t) \ \ \cdots \ \ \hat{b}_{i,n}(t)]^T;$$

associated with these estimates are the polynomials

$$\hat{A}_i(t, z^{-1}) = 1 + \hat{a}_{i,1}(t)z^{-1} + \hat{a}_{i,2}(t)z^{-2} \cdots + \hat{a}_{i,n}(t)z^{-n},$$
$$\hat{B}_i(t, z^{-1}) = \hat{b}_{i,1}(t)z^{-1} + \hat{b}_{i,2}(t)z^{-2} \cdots + \hat{b}_{i,n}(t)z^{-n}.$$

Next, we choose the following polynomials

$$\hat{L}_i(t, z^{-1}) = 1 + \hat{l}_{i,1}(t)z^{-1} + \hat{l}_{i,2}(t)z^{-2} + \cdots + \hat{l}_{i,n}(t)z^{-n},$$
$$\hat{P}_i(t, z^{-1}) = \hat{p}_{i,1}(t)z^{-1} + \hat{p}_{i,2}(t)z^{-2} + \cdots + \hat{p}_{i,n}(t)z^{-n}$$

to place all closed-loop poles at zero, so we need

$$\hat{A}_i(t, z^{-1})\hat{L}_i(t, z^{-1}) + \hat{B}_i(t, z^{-1})\hat{P}_i(t, z^{-1}) = 1.$$

Given the assumption that the $\hat{A}_i(t, z^{-1})$ and $\hat{B}_i(t, z^{-1})$ are coprime, we know that there exist **unique** $\hat{L}_i(t, z^{-1})$ and $\hat{P}_i(t, z^{-1})$ which satisfy this equation; it is also easy to prove that the coefficients of $\hat{L}_i(t, z^{-1})$ and $\hat{P}_i(t, z^{-1})$ are analytic functions of $\hat{\theta}_i(t) \in \mathcal{S}_i$.

We can now discuss the candidate control law to be used. Define the controller gain by

$$\hat{K}_i(t) := [-\hat{p}_{i,1}(t) \ \ \cdots \ \ -\hat{p}_{i,n}(t) \ \ -\hat{l}_{i,1}(t) \ \ \cdots \ \ -\hat{l}_{i,n}(t)]^T$$

and a switching signal $\sigma : \mathbf{Z} \mapsto \{1, 2\}$ that decides which gain to use at any given point in time. A natural choice for a control law is

$$u(t) = \hat{K}_{\sigma(t-1)}(t-1)^T \phi(t-1) + \sum_{j=1}^{n} \hat{p}_{\sigma(t-1),j}(t-1) y^*(t-j); \tag{50}$$

the above control law is similar to that given in (11) but with the controller gains chosen between the two choices at each point in time.

The most obvious choice of $\sigma(t)$ is to define it by

$$\operatorname{argmin}\{|e_1(t)|, |e_2(t)|\}.$$

While this works in every simulation that we have tried, the proof remains elusive.

Hence, we try another approach, which is based on our earlier work in the first-order one-step-ahead setting [41], and exploits the deadbeat nature of the problem. Using the natural notation of $\bar{A}_{\sigma(t)}(t)$ to represent the $\bar{A}(t)$ matrix of (12) as well as $\hat{r}(t) := \sum_{j=1}^{n} \hat{p}_{\sigma(t),j}(t)y^*(t-j+1)$, the closed-loop behaviour is captured by

$$\phi(t+1) = \bar{A}_{\sigma(t)}(t)\phi(t) + B_1 e_{\sigma(t)}(t+1) + B_2\hat{r}(t). \tag{51}$$

While $\bar{A}_{\sigma(t)}(t)$ is a deadbeat matrix for every t, the product

$$\bar{A}_{\sigma(t)}(t) \times \bar{A}_{\sigma(t-1)}(t-1) \times \cdots \times \bar{A}_{\sigma(t_0)}(t_0), \qquad t \geq t_0$$

will not usually have all eigenvalues at zero. Hence, the method of analysis used in [41] will not work for the proposed control law (50). A natural attempt to alleviate the problem is to hold $\sigma(t)$ constant in (50) for $2n$ steps at a time; the difficulty now is that $\bar{A}_{\sigma(t)}(t)$ is still changing since $\hat{\theta}_i(t)$ still changes. A natural solution to this problem is to update the estimators every $2n$ steps as well; the difficulty here is that we end up with no information about $e_i(t+1)$ between the updates, so the closed-loop system is not amenable to analysis. So our proposed solution procedure will need to be different: we are going to change $\sigma(t)$ every $N \geq 2n$ steps; we keep the estimators running, but adjust the control parameters every $N \geq 2n$ steps as well. The effect of this will become clear in the proof of the main result of this section. To this end, we define a sequence of switching times as follows: we initialize $\hat{t}_0 := t_0$ and then define

$$\hat{t}_\ell := t_0 + \ell N, \qquad \ell \in \mathbf{N}.$$

So now define the associated controller parameters by

$$K_i(t) := [-\hat{p}_{i,1}(\hat{t}_\ell) \ \cdots \ -\hat{p}_{i,n}(\hat{t}_\ell) \ \ -\hat{l}_{i,1}(\hat{t}_\ell) \ \cdots \ -\hat{l}_{i,n}(\hat{t}_\ell)]^T, \ \ t \in [\hat{t}_\ell, \hat{t}_{\ell+1}), \ \ell \in \mathbf{Z}^+, \tag{52}$$

the switching signal by

$$\sigma(t) = \sigma(\hat{t}_\ell), \ \ t \in [\hat{t}_\ell, \hat{t}_{\ell+1}), \ \ell \in \mathbf{Z}^+, \tag{53}$$

and the suitably revised definition of $r(\cdot)$:

$$r(t) := \sum_{j=1}^{n} \hat{p}_{\sigma(\hat{t}_\ell),j}(\hat{t}_\ell)y^*(t-j+1), \ \ t \in [\hat{t}_\ell, \hat{t}_{\ell+1}), \ \ell \in \mathbf{Z}^+.$$

We now define the control law as

$$u(t) = K_{\sigma(t-1)}(t-1)^T \phi(t-1) + r(t-1), \qquad t \geq t_0. \tag{54}$$

What remains to be defined is the choice of switching signal $\sigma(\hat{t}_\ell)$. To this end, we define a performance signal $J_i : \{\hat{t}_0, \hat{t}_1, \ldots\} \mapsto \mathbf{R}$ for estimator i, which produces a measure of "accuracy" of estimation; for $\ell \in \mathbf{Z}^+$, we define

$$J_i(\hat{t}_\ell) := \begin{cases} 0 & \text{if } \phi(j) = 0 \text{ for all } j \in [\hat{t}_\ell, \hat{t}_{\ell+1}), \\ \max\limits_{j \in [\hat{t}_\ell, \hat{t}_{\ell+1}), \phi(\cdot) \neq 0} \frac{|e_i(j+1)|}{\|\phi(j)\|} & \text{otherwise.} \end{cases} \tag{55}$$

Before proceeding, assume that $\theta^* \in \mathcal{S}$, so there exists one or more $j \in \{1, 2\}$ so that $\theta^* \in \mathcal{S}_j$; **throughout the remainder of this section, let i^* denote the smallest such j**. With $\sigma(\hat{t}_0) = \sigma_0$, we use the following switching rule:

$$\sigma(\hat{t}_{\ell+1}) = \operatorname*{argmin}_{i \in \{1,2\}} J_i(\hat{t}_\ell), \qquad \ell \in \mathbf{Z}^+; \tag{56}$$

for the case when $J_1(\hat{t}_\ell) = J_2(\hat{t}_\ell)$, we (somewhat arbitrarily) select $\sigma(\hat{t}_{\ell+1})$ to be 1. Before presenting the main result of this section, we first show that the logic in (56) yields a desirable closed-loop property.

Lemma 2 *Consider the plant (1) subject to Assumption 2 and suppose that the controller consisting of the estimator* (48) *and* (49)*, the control law* (54)*, the performance signal* (55) *and the switching rule* (56) *is applied. Then for every $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\sigma_0 \in \{1, 2\}$, $N \geq 1$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ $(i = 1, 2)$ and $y^*, d \in \ell_\infty$, we have that, for any $\ell \geq 0$, either*

(a) $J_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) \leq J_{i^*}(\hat{t}_\ell)$, *or*
(b) $J_{\sigma(\hat{t}_{\ell+1})}(\hat{t}_{\ell+1}) \leq J_{i^*}(\hat{t}_{\ell+1})$.

Proof Fix $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\sigma_0 \in \{1, 2\}$, $N \geq 1$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ $(i = 1, 2)$, and $y^*, d \in \ell_\infty$; let $\ell \geq 0$ be arbitrary. We know that $\theta^* \in \mathcal{S}_{i^*}$. Assume that (a) does not hold, i.e. $J_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) > J_{i^*}(\hat{t}_\ell)$; then according to (56), this means that $\sigma(\hat{t}_{\ell+1}) = i^*$, i.e. (b) will hold. □

In the above we do not make any claim that $\theta^* \in \mathcal{S}_{\sigma(t)}$ at any time; it only makes a statement about the size of the prediction error. It turns out that this is enough to ensure that closed-loop stability is attained. Next, we present the main result of this section.

8.3 The result

Theorem 4 *Consider the plant* (1) *subject to Assumption 2 and suppose that the controller consisting of the estimator* (48), (49), *the control law* (54), *the performance signal* (55) *and the switching rule* (56) *is applied to the plant. For every* $\lambda \in (0, 1)$ *and* $N \geq 2n$, *there exists a constant* $\gamma > 0$ *such that for every* $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\sigma_0 \in \{1, 2\}$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ $(i = 1, 2)$, *and* $y^*, d \in \ell_\infty$, *the following bound holds:*

$$\|\phi(t)\| \leq \gamma\lambda^{t-t_0}\|\phi_0\| + \gamma \sum_{j=t_0}^{t-1} \lambda^{t-1-j}(|d(j)| + |r(j)|), \quad t \geq t_0. \tag{57}$$

Proof See Appendix. □

Remark 11 The approach taken in this proof differs a fair bit from that of Theorem 1 and does not make use of Kreisselmeier's result given in Proposition 2. Indeed, we spend the first half of the proof converting the closed-loop system into a first-order difference inequality which describes the closed-loop system every $2N$ steps, and the remainder of the proof consists of a modification of the arguments from our earlier work on the first-order one-step-ahead controller [32] to fit this new setting.

Remark 12 It turns out that the controller presented in this section enjoys the same tolerance to slowly time-varying parameters and to unmodelled dynamics as the one designed for the case of $\mathcal{S}$ convex. First of all, the proof for the case of unmodelled dynamics is identical to that of Theorem 3. Second of all, the proof for the case of time-varying parameters given in Theorem 2 just requires a small adjustment: simply choose the free parameter $m > N_1$ to be an integer multiple of N. Here we have made use of one of the most desirable features of the approach, namely that of modularity.

Remark 13 It is natural to ask if the proposed approach would work if $\mathcal{S} \subset \bigcup_{i=1}^{p} \mathcal{S}_i$ with each $\mathcal{S}_i$ compact and convex sets and for which the corresponding pair of polynomials $A(z^{-1})$ and $B(z^{-1})$ are coprime. (After all, if $\mathcal{S}$ is simply compact we can use the Heine–Borel Theorem to prove the existence of such $\mathcal{S}_i$'s.) While the proposed controller (48), (49), (54), (55) and (56) is well defined in this case, we have been unable to prove that it will work; a potential problem is that the switching algorithm could oscillate between two bad choices and never (or rarely) choose the correct one. We are presently working on a more complicated switching algorithm which does not have that problem.

9 Some simulation examples

Here we start with an example which satisfies Assumption 1; in Sect. 9.1 we focus on stability and in Sect. 9.2 we expand this to step tracking. We then move to an example which satisfies Assumption 2 and illustrate the switching controller of Sect. 8.

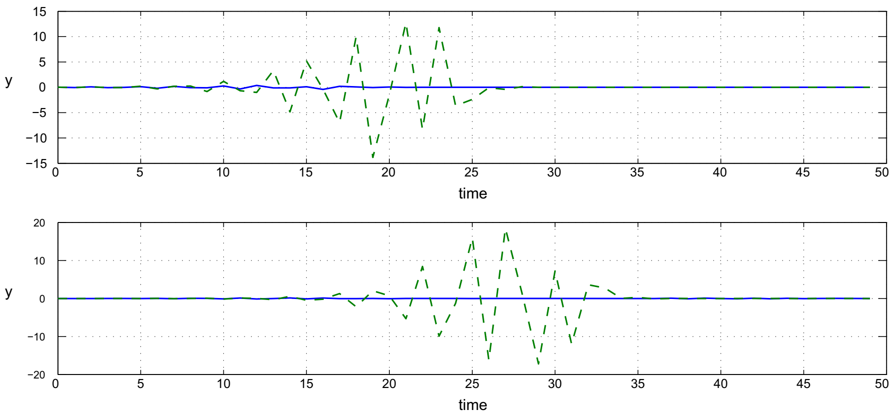

Fig. 1 A comparison of the ideal algorithm (solid) and the classical algorithm (dashed) with a nonzero initial condition and no noise (top plot) and a zero initial condition and noise (bottom plot)

9.1 Stability

Here we provide an example to illustrate the benefit of the proposed adaptive controller. To this end, consider the second-order plant

$$y(t+1) = -a_1(t)y(t) - a_2(t)y(t-1) + b_1(t)u(t) + b_2(t)u(t-1) + d(t)$$

with $a_1(t) \in [0, 2]$, $a_2(t) \in [1, 3]$, $b_1(t) \in [0, 1]$, and $b_2(t) \in [-5, -2]$. If the parameters were fixed, then every admissible model is unstable and non-minimum phase, which makes this a challenging plant to control; indeed, it has two complex unstable poles together with a zero that can lie anywhere in $[2, \infty)$. Last of all, for simplicity set $\delta = \infty$.

In this subsection we consider the problem of stability only—we set $y^* = 0$. First, we compare the ideal algorithm (4)–(5) (with projection onto $\mathcal{S}$) with the classical one (3) (suitably modified to have projection onto $\mathcal{S}$); in both cases we couple the estimator with the adaptive pole placement controller (11) where we place all closed-loop poles at zero. In the case of the classical estimator (3), we arbitrarily set $\alpha = \beta = 1$. Suppose that the plant parameters are constant: $(a_1, a_2, b_1, b_2) = (2, 3, 1, -2)$, but the initial estimate is set to the midpoint of the interval. In the first simulation we set $y(0) = y(-1) = 0.01$ and $u(-1) = 0$ and set the noise $d(t)$ to zero—see the top plot of Fig. 1. In the second simulation we set $y(0) = y(-1) = u(-1) = 0$ and the noise to $d(t) = 0.01 * \sin(5t)$—see the bottom plot of Fig. 1. In both cases the controller based on the ideal algorithm (4)–(5) is clearly superior to the one based on the classical algorithm (3).

Now we compare the two adaptive controllers when applied to a time-varying version of the plant with unmodelled dynamics, a zero initial condition, and a nonzero noise. More specifically, we set

$$a_1(t) = 1 + \sin(.001t),\ a_2(t) = 2 + \cos(.001t),\ b_1(t) = 0.5 + 0.5\sin(.005t),$$

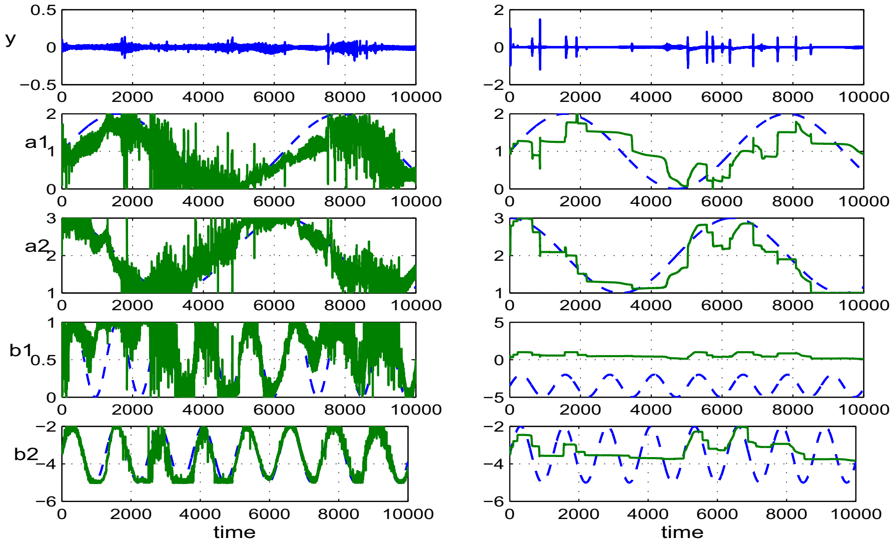

Fig. 2 The system behaviour with time-varying parameters and unmodelled dynamics; the parameters are dashed and the estimates are solid. The ideal algorithm is used in the left-most plots, while the classical algorithm is used in the right-most plots

$$b_2(t) = -3.5 + 1.5\sin(.005t), \quad d(t) = 0.01\sin(5t).$$

For the unmodelled part of the plant, we use a term of the form discussed in Sect. 6.2:

$$m(t+1) = 0.75m(t) + 0.75\|\phi(t)\|, \quad m(0) = 0,$$
$$d_\Delta(t) = \begin{cases} 0 & t = 0, 1, ..., 4999 \\ 0.025m(t) + 0.025\|\phi(t)\| & t \geq 5000. \end{cases}$$

We plot the result in Fig. 2. We see that the controller based on the ideal algorithm is clearly superior to the one based on the classical algorithm, in the sense that the average size of the output $y(t)$ is smaller (by a factor of about three), and the parameter estimates are more accurate; the latter property stems from the fact that the classical estimator tends to have a low gain when the signals are small, unlike the ideal estimator.

9.2 Step tracking

The plant in the previous subsection has a large amount of uncertainty, as well as a wide range of unstable poles and non-minimum phase zeros, which means that there are limits on the quality of the transient behaviour even if the parameters were fixed and known. Hence, to illustrate the tracking ability we look at a subclass of systems: one with a_1 and b_1 as before, namely $a_1(t) \in [0, 2]$ and $b_1(t) \in [0, 1]$, but now with $a_2 = 1$ and $b_2 = -3.5$. With fixed parameters the corresponding system is still unstable and non-minimum phase.

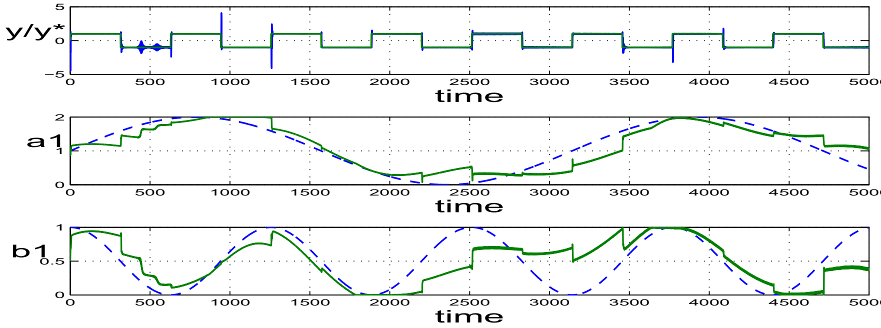

Fig. 3 The pole placement tracking controller with time-varying parameters and small noise; the parameters are dashed and the estimates are solid

We simulate the closed-loop pole placement step tracking controller of Sect. 7 with a zero initial condition, initial parameter estimates at the midpoints of the admissible intervals, and with time-varying parameters:

$$a_1(t) = 1 + \sin(.002t), b_1(t) = 0.5 + 0.5\cos(.005t),$$

with a nonzero disturbance:

$$d(t) = \begin{cases} 0.01\sin(5t) & t = 0, 1, ..., 2499 \\ 0.05\sin(5t) & t = 2500, ..., 4999, \end{cases}$$

and a square wave reference signal of $y^*(t) = \text{sgn}[sin(0.01t)]$. We plot the result in Fig. 3; we see that the parameter estimates crudely follows the system parameters, with less accuracy than in the previous subsection, partly due to the fact that the constant setpoint dominates the estimation process and leads to higher inaccuracy. As a result, $y(t)$ does a good job of following y^* on average, but with the occasional flurry of activity when the parameter estimates are highly inaccurate. When the noise is increased fivefold at $k = 2500$, the behaviour degrades only slightly.

9.3 Removing convexity assumption

Next we illustrate the case when the set of admissible plant parameters satisfies Assumption 2 but not Assumption 1; we will consider a second-order plant with θ^* belonging to $\mathcal{S} \subset \mathcal{S}_1 \cup \mathcal{S}_2$, with $\mathcal{S}_1$ equal to the set $\mathcal{S}$ of the previous example and $\mathcal{S}_2$ equal to $\mathcal{S}_1$ but with a sign added on the b_i parameters:

$$\mathcal{S}_1 := \left\{ \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} \in \mathbf{R}^4 : a_1 \in [0, 2], a_2 \in [1, 3], b_1 \in [0, 1], b_2 \in [-5, -2] \right\},$$

$$\mathcal{S}_2 := \left\{ \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} \in \mathbf{R}^4 : a_1 \in [0, 2], a_2 \in [1, 3], b_1 \in [-1, 0], b_2 \in [2, 5] \right\};$$

notice that convex hull of $\mathcal{S}_1 \cup \mathcal{S}_2$ includes the case of having $b_1 = b_2 = 0$, which means that the corresponding system is not stabilizable. We will apply the proposed controller of Sect. 8 to the time-varying plant with a zero initial condition and reference signal, and a nonzero noise. Specifically, we set $a_1(t)$ and $a_2(t)$ as in Sect. 9.1:

$$a_1(t) = 1 + \sin(.001t), \quad a_2(t) = 2 + \cos(.001t),$$

but now we set $b_1(t)$ and $b_2(t)$ to be

$$b_1(t) = \begin{cases} -0.5 - 0.5\sin(.005t) & 1500 \le t < 8000 \\ 0.5 + 0.5\sin(.005t) & \text{otherwise,} \end{cases}$$
$$b_2(t) = \begin{cases} 3.5 - 1.5\sin(.005t) & 1500 \le t < 8000 \\ -3.5 + 1.5\sin(.005t) & \text{otherwise.} \end{cases}$$

We apply the proposed switching controller consisting of the estimator (48), (49), the control law (54), the performance signal (55) and the switching rule (56); we choose $N = 2n = 4$. Here we also set $y^* = 0$, $y(0) = y(-1) = u(-1) = 0$ and the noise to $d(t) = 0.01\sin(5t)$. Initial parameter estimates $\hat{\theta}_i(0)$ are set to the midpoints of each respective interval, and we set $\sigma_0 = 2$.

The result for this case is plotted in Fig. 4; we see that the controller does a reasonable job, even though the switching algorithm often chooses the wrong model. Larger transients (than in the simulation of Sect. 9.1) occasionally ensue, but on average the adaptive controller provides good performance. Furthermore, the estimator does a fairly good job of tracking the time-varying parameter.

10 Summary and conclusions

Here we show that if the original, ideal, projection algorithm is used in the estimation process (subject to the assumption that the plant parameters lie in a convex, compact set), then the corresponding pole placement adaptive controller guarantees linear-like convolution bounds on the closed-loop behaviour, which confers exponential stability and a bounded noise gain (in every p-norm with $1 \le p \le \infty$), unlike almost all other parameter adaptive controllers. This can be leveraged to prove tolerance to unmodelled dynamics and plant parameter variation. We emphasize that there is no persistent excitation requirement of any sort; the improved performance arises from the vigilant nature of the ideal parameter estimation algorithm.

As far as the author is aware, **the linear-like convolution bound proven here is a first in parameter adaptive control**. It allows a modular approach to be used in analysing time-varying parameters and unmodelled dynamics. This approach avoids all of the fixes invented in the 1980s, such as signal normalization and deadzones, used to deal with the lack of robustness to unmodelled dynamics and time-varying parameters.

In the present paper the standard assumption is that the set of plant parameters lies in a compact and convex set. In Sect. 8 we relaxed the convexity requirement a bit: there the goal is to place all poles at the origin, and the convexity requirement is

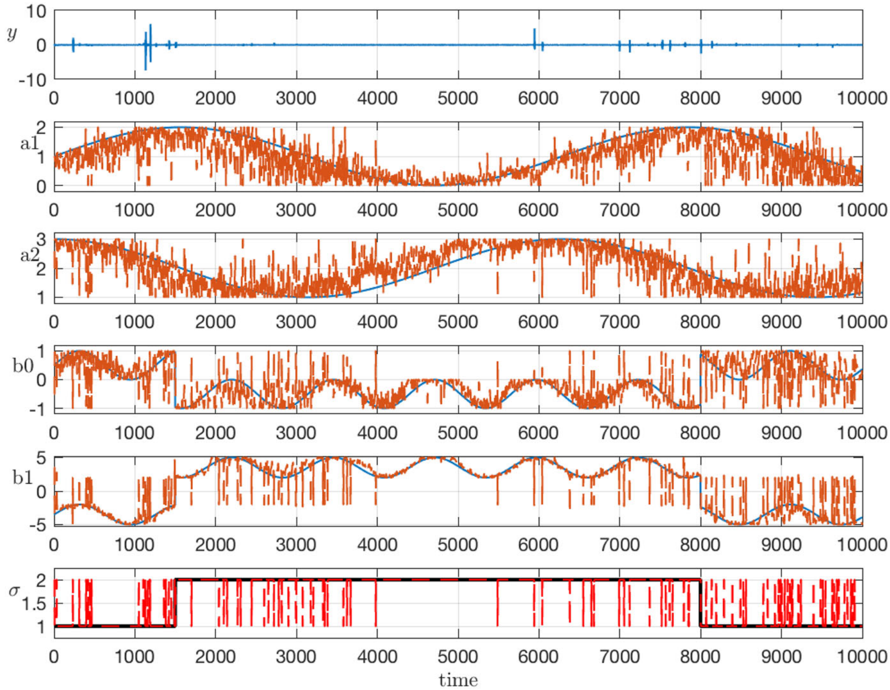

Fig. 4 The upper plot shows the system output. The next four plots show the parameter estimates $\hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)$ (dashed) and actual plant parameters (solid). The bottom plot shows the switching signal $\sigma(\hat{t}_\ell)$ (dashed) and the correct index (solid)

weakened to requiring that the set of admissible parameters lie in the union of two convex sets; two parameter estimators are used together with a switching algorithm to choose which estimator to use at each point in time. We are working at removing the convexity requirement altogether: the idea is to utilize the Heine–Borel theorem to prove that the compact set of admissible parameters lies in a finite union of convex set (for which the corresponding numerator and denominator polynomials are coprime), use a parameter estimator for each, and then design a switching algorithm to switch between them. While the (natural extension of) the switching algorithm for the case of two convex sets does not appear to work in the general case, we are presently analysing a more complicated algorithm.

We are presently working on extending the approach to the model reference adaptive control setting; the analysis is turning out to be more complicated than here, in large part due to the facts that the controller is not strictly causal (as it is here) and that a system delay (or the relative degree of the plant transfer function) creates additional complexity. Extending the approach to the continuous-time setting may prove challenging, since a direct application would yield a non-Lipschitz continuous estimator, which brings with it mathematical solvability issues.

11 Appendix

Proof of Proposition 1: Since projection does not make the parameter estimate worse, it follows from (7) that

$$\|\hat{\theta}(t+1)-\hat{\theta}(t)\| \leq \|\check{\theta}(t+1)-\hat{\theta}(t)\| \leq \|\rho_\delta(\phi(t),e(t+1))\frac{\phi(t)}{\phi(t)^T\phi(t)}e(t+1)\| \\ \leq \rho_\delta(\phi(t),e(t+1))\frac{|e(t+1)|}{\|\phi(t)\|}, \quad t \geq t_0.$$

so the first inequality holds.

We now turn to energy analysis. We first define $\tilde{\check{\theta}}(t) := \check{\theta}(t) - \theta^*$ and $\check{V}(t) := \tilde{\check{\theta}}(T)^T\tilde{\check{\theta}}(t)$. Next, we subtract θ^* from each side of (7), yielding

$$\tilde{\check{\theta}}(t+1) = \tilde{\theta}(t) + \rho_\delta(\phi(t),e(t+1))\frac{\phi(t)}{\phi(t)^T\phi(t)}[-\phi(t)^T\tilde{\theta}(t)+d(t)] \\ = \left[I - \underbrace{\rho_\delta(\phi(t),e(t+1))\frac{\phi(t)\phi(t)^T}{\phi(t)^T\phi(t)}}_{=:W_1(t)}\right]\tilde{\theta}(t) + \underbrace{\rho_\delta(\phi(t),e(t+1))\frac{\phi(t)}{\phi(t)^T\phi(t)}}_{=:W_2(t)}d(t).$$

Then

$$\check{V}(t+1) = [(I-W_1(t))\tilde{\theta}(t)+W_2(t)d(t)]^T \times [(I-W_1(t))\tilde{\theta}(t)+W_2(t)d(t)] \\ = \tilde{\theta}(t)^T[I-W_1(t)][I-W_1(t)]\tilde{\theta}(t) \\ +2\tilde{\theta}(t)^T[I-W_1(t)]W_2(t)d(t) + W_2(t)^TW_2(t)d(t)^2.$$

Now let us analyse the three terms on the RHS: the fact that $W_1(t)^2 = W_1(t)$ allows us to simplify the first term; the fact that $W_1(t)W_2(t) = W_2(t)$ means that the second term is zero; $W_2(t)^TW_2(t) = \rho_\delta(\phi(t),e(t+1))\frac{1}{\phi(t)^T\phi(t)}$, which simplifies the third term. We end up with

$$\check{V}(t+1) = \tilde{\theta}(t)^T[I-W_1(t)]\tilde{\theta}(t) + \rho_\delta(\phi(t),e(t+1))\frac{d(t)^2}{\phi(t)^T\phi(t)} \\ = V(t) - \rho_\delta(\phi(t),e(t+1))\frac{[\tilde{\theta}(t)^T\phi(t)]^2}{\phi(t)^T\phi(t)} + \rho_\delta(\phi(t),e(t+1))\frac{d(t)^2}{\phi(t)^T\phi(t)} \\ = V(t) + \rho_\delta(\phi(t),e(t+1))\frac{d(t)^2-[d(t)-e(t+1)]^2}{\phi(t)^T\phi(t)} \\ \leq V(t) + \rho_\delta(\phi(t),e(t+1))\frac{-\frac{1}{2}e(t+1)^2+2d(t)^2}{\phi(t)^T\phi(t)}.$$

Since projection never makes the estimate worse, it follows that

$$V(t+1) \leq V(t) + \rho_\delta(\phi(t), e(t+1)) \frac{-\frac{1}{2}e(t+1)^2 + 2d(t)^2}{\phi(t)^T\phi(t)}.$$

□

Proof of Lemma 1: Fix $\delta \in (0, \infty]$ and $\sigma \in (\underline{\lambda}, 1)$. First of all, it is well known that the characteristic polynomial of $\bar{A}(t)$ is exactly $z^{2n}A^*(z^{-1})$ for every $t \geq t_0$. Furthermore, it is well known that the coefficients of $\hat{L}(t, z^{-1})$ and $\hat{P}(t, z^{-1})$ are the solution of a linear equation, and are analytic functions of $\hat{\theta}(t) \in \mathcal{S}$. Hence, there exists a constant γ_1 so that, for every set of initial conditions, $y^* \in l_\infty$ and $d \in l_\infty$, we have $\sup_{t \geq t_0} \|\bar{A}(t)\| \leq \gamma_1$.

To prove the first bound, we now invoke the argument used in [40], who considered a more general time-varying situation but with more restrictions on σ. By making a slight adjustment to the first part of the proof given there, we can prove that with $\gamma_2 := \sigma \frac{(\sigma+\gamma_1)^{2n-1}}{(\sigma-\underline{\lambda})^{2n}}$, then for every $t \geq t_0$ we have $\|\bar{A}(t)^k\| \leq \gamma_2\sigma^k, \quad k \geq 0$, as desired.

Now we turn to the second bound. From Proposition 1 and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\sum_{j=k}^{t-1} \|\hat{\theta}(j+1) - \hat{\theta}(j)\| &\leq \sum_{j=k}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{|e(j+1)|}{\|\phi(j)\|} \\ &\leq \left[\sum_{j=k}^{t-1} \rho_\delta(\phi(j), e(j+1)) \frac{e(j+1)^2}{\|\phi(j)\|^2}\right]^{1/2} (t-k)^{1/2}.\end{aligned}$$

Now notice that

$$\begin{aligned}\|\bar{A}(t+1) - \bar{A}(t)\| &\leq \|\hat{\theta}(t+1) - \hat{\theta}(t)\| \\ &\quad + \sum_{i=1}^{n} (|\hat{l}_i(t+1) - \hat{l}_i(t)| + |\hat{p}_i(t+1) - \hat{p}_i(t)|).\end{aligned}$$

The fact that the coefficients of $\hat{L}(t, z^{-1})$ and $\hat{P}(t, z^{-1})$ are analytic functions of $\hat{\theta}(t) \in \mathcal{S}$ means that there exists a constant $\gamma_3 \geq 1$ so that

$$\sum_{j=k}^{t-1} \|\bar{A}(j+1) - \bar{A}(j)\| \leq \gamma_3 \sum_{j=k}^{t-1} \|\hat{\theta}(j+1) - \hat{\theta}(j)\|,$$

so we conclude that the second bound holds as well. □

In order to prove Theorem 4, we need some preliminary results. The first step is to extend Proposition 1 to the case when θ^* may not lie in $\mathcal{S}_i$.

Proposition 3 *For every $t_0 \in \mathbf{Z}$, $t_2 \geq t_1 \geq t_0$, $\phi_0 \in \mathbf{R}^{2n}$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$, and $d \in \ell_\infty$, when the estimator (48) and (49) is applied to the plant (1),*

$$\|\hat{\theta}_i(t_2) - \hat{\theta}_i(t_1)\| \leq \sum_{j=t_1,\phi(j)\neq 0}^{t_2-1} \frac{|e_i(j+1)|}{\|\phi(j)\|}, \quad i = 1, 2. \tag{58}$$

Proof Since projection does not make the parameter estimate worse, it follows from (48) and (49) that when $\phi(t) \neq 0$,

$$\|\hat{\theta}_i(t+1) - \hat{\theta}_i(t)\| \leq \|\check{\theta}_i(t+1) - \hat{\theta}_i(t)\| \leq \left\| \frac{\phi(t)e_i(t+1)}{\|\phi(t)\|^2} \right\| \leq \frac{|e_i(t+1)|}{\|\phi(t)\|}, \tag{59}$$

and when $\phi(t) = 0$,

$$\|\hat{\theta}_i(t+1) - \hat{\theta}_i(t)\| = 0.$$

The result follows by iteration. □

The next result produces a crude bound on the closed-loop behaviour.

Proposition 4 *Consider the plant (1) and suppose that the controller consisting of the estimator (48), (49) and the control law (54) is applied.*[8] *Then for every $p \geq 0$, there exists a constant $\bar{c} \geq 1$ such that for every $t_0 \in \mathbf{Z}$, $t \geq t_0$, $\phi_0 \in \mathbf{R}^{2n}$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ ($i = 1, 2$) and $y^*, d \in \boldsymbol{\ell}_\infty$:*

$$\|\phi(t+p)\| \leq \bar{c}\|\phi(t)\| + \bar{c} \sum_{j=0}^{p-1} (|d(t+j)| + |r(t+j)|). \tag{60}$$

Proof Fix $p \geq 0$. Let $t_0 \in \mathbf{Z}$, $t \geq t_0$, $\phi_0 \in \mathbf{R}^{2n}$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ ($i = 1, 2$) and $y^*, d \in \boldsymbol{\ell}_\infty$ be arbitrary. From (1) we see that

$$|y(t+1)| \leq \|\mathcal{S}\| \|\phi(t)\| + |d(t)|.$$

From (54) and Assumption 2, we have that there exists a constant γ so that

$$|u(t+1)| \leq \gamma \|\phi(t)\| + |r(t)|.$$

From the definition of $\|\phi(t+1)\|$, we have that

$$\|\phi(t+1)\| \leq \|\phi(t)\| + |y(t+1)| + |u(t+1)|.$$

[8] The choice of N and the value of the switching signal $\sigma(t)$ play no role.

Combining these three bounds, we end up with

$$\|\phi(t+1)\| \leq \underbrace{(1 + \|\mathcal{S}\| + \gamma)}_{=:\bar{a}} \|\phi(t)\| + |r(t)| + |d(t)|.$$

Solving iteratively, we have

$$\begin{aligned}\|\phi(t+p)\| &\leq \bar{a}^p \|\phi(t)\| + \sum_{j=0}^{p-1} \bar{a}^{p-j-1}(|d(t+j)| + |r(t+j)|) \\ &\leq \bar{a}^p \|\phi(t)\| + \bar{a}^{p-1} \sum_{j=0}^{p-1}(|d(t+j)| + |r(t+j)|).\end{aligned}$$

Put $\bar{c} := \bar{a}^p$ to conclude the proof. □

We now state a technical result which we used in [32] to analyse the first-order one-step-ahead adaptive control problem.

Lemma 3 *[32] (i) With $n \in \mathbf{N} \cup \{\infty\}$, suppose that $a_j \in \mathbf{R}$ and $c > 0$ satisfy*

$$\sum_{j=0}^{n} a_j^2 \leq c.$$

Then for every $\lambda \in (0, 1)$, if we define $\gamma := c^{\frac{c+1}{2}} (\frac{1}{\lambda})^{\frac{c}{\lambda^2}+1}$, then the following holds:

$$\left|\prod_{l=0}^{j-1} a_l\right| \leq \gamma\lambda^j \qquad j = 0, 1, \ldots, n.$$

(ii) With $n < p \leq \infty$, suppose that $a_j \in \mathbf{R}$ and $c_1 > 0$ satisfy

$$\sum_{l=j}^{j+n} a_l^2 \leq c_1 \qquad j = 0, 1, \ldots, p-n.$$

Then for every $\lambda \in (0, 1)$, if c_1, λ and n satisfy

$$n \geq \frac{\frac{c_1+1}{2}\ln(c_1) + (4\frac{c_1}{\lambda^2} + 1)(\ln(2) - \ln(\lambda))}{\ln(2)}$$

and $\gamma_1 := c_1^{\frac{c_1+1}{2}} (\frac{2}{\lambda})^{\frac{4c_1}{\lambda^2}+1}$, then

$$\left|\prod_{l=0}^{j-1} a_l\right| \leq \gamma_1\lambda^j \qquad j = 0, 1, \ldots, p.$$

Proof of Theorem 4: Fix $\lambda \in (0, 1)$ and $N \geq 2n$. Let $t_0 \in \mathbf{Z}$, $\phi_0 \in \mathbf{R}^{2n}$, $\sigma_0 \in \{1, 2\}$, $\theta^* \in \mathcal{S}$, $\hat{\theta}_i(t_0) \in \mathcal{S}_i$ $(i = 1, 2)$, and $y^*, d \in \ell_\infty$ be arbitrary; as usual, we let i^* denote the smallest $j \in \{1, 2\}$ which satisfies $\theta^* \in \mathcal{S}_j$.

As mentioned at the beginning of Sect. 8, the proposed controller is based on the first-order one-step-ahead control setup [41], although it is more complicated. The proof also uses similar ideas as those used in [41], but as our system is more complicated, it should not be surprising that the proof is significantly different and much more complicated. Hence, before proceeding we provide a proof outline: using the definition of $\hat{t}_\ell$ given in Sect. 8.2:

1. first, we define a state-space equation describing $\phi(t)$ which holds on intervals of the form $[\hat{t}_\ell, \hat{t}_{\ell+1})$;
2. second, we analyse this equation, getting a bound on $\|\phi(\hat{t}_{\ell+1})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$ and the exogenous inputs;
3. we apply Lemma 2 and Proposition 4 to obtain a bound on $\|\phi(\hat{t}_{\ell+2})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$; i.e. we analyse two intervals at a time;
4. fourth, we then analyse the associated difference inequality (relating $\|\phi(\hat{t}_{\ell+2})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$) in a way similar (though not identical) to that used in [41].

Step 1: Obtain a state-space model describing $\phi(t)$ for $t \in [\hat{t}_\ell, \hat{t}_{\ell+1})$.

By definition of the prediction error (47) and by the property of the switching signal (53) being constant on $[\hat{t}_\ell, \hat{t}_{\ell+1})$, we have

$$\begin{aligned} y(t+1) &= \phi(t)^T \hat{\theta}_{\sigma(t)}(t) + e_{\sigma(t)}(t+1) \\ &= \phi(t)^T \hat{\theta}_{\sigma(\hat{t}_\ell)}(t) + e_{\sigma(\hat{t}_\ell)}(t+1) + \phi(t)^T \hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) - \phi(t)^T \hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) \\ &= \hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)^T \phi(t) + \left[\hat{\theta}_{\sigma(\hat{t}_\ell)}(t) - \hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^T \phi(t) \\ &\quad + e_{\sigma(\hat{t}_\ell)}(t+1), \quad t \in [\hat{t}_\ell, \hat{t}_{\ell+1}). \end{aligned} \tag{61}$$

From the control law (54) and the control gains (52), we have

$$\begin{aligned} u(t+1) &= K_{\sigma(t)}(t)^T \phi(t) + r(t) \\ &= K_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)^T \phi(t) + r(t), \qquad t \in [\hat{t}_\ell, \hat{t}_{\ell+1}). \end{aligned} \tag{62}$$

We now derive a state-space equation for $\phi(t)$ in much the same way as (13) was derived; we first define

$$\bar{A}_{\sigma(j)}(j) := \begin{bmatrix} -\hat{a}_{\sigma(j),1}(j) & -\hat{a}_{\sigma(j),2}(j) & \cdots & -\hat{a}_{\sigma(j),n}(j) & \hat{b}_{\sigma(j),1}(j) & \cdots & \cdots & \hat{b}_{\sigma(j),n}(j) \\ 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ & \ddots & & \vdots & \vdots & \cdots & \cdots & \vdots \\ & & 1 & 0 & 0 & \cdots & \cdots & 0 \\ -\hat{p}_{\sigma(j),1}(j) & -\hat{p}_{\sigma(j),2}(j) & \cdots & -\hat{p}_{\sigma(j),n}(j) & -\hat{l}_{\sigma(j),1}(j) & -\hat{l}_{\sigma(j),2}(j) & \cdots & -\hat{l}_{\sigma(j),n}(j) \\ 0 & \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & & & 1 & 0 \end{bmatrix};$$

then, in light of (61) and (62), the following holds:

$$\begin{aligned}\phi(t+1) &= \bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\phi(t) \\ &+B_1\left(\left[\hat{\theta}_{\sigma(\hat{t}_\ell)}(t)-\hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^T\phi(t)+e_{\sigma(\hat{t}_\ell)}(t+1)\right)+B_2r(t), \\ t &\in [\hat{t}_\ell,\hat{t}_{\ell+1}),\ \ell\in\mathbf{Z}^+; \end{aligned} \tag{63}$$

notice the additional term $\left[\hat{\theta}_{\sigma(\hat{t}_\ell)}(t)-\hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]$ on the right-hand side which (13) does not have.

Step 2: Obtain a bound on $\|\phi(\hat{t}_{\ell+1})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$.

In (63) we have $\bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\in\mathbf{R}^{2n\times 2n}$ to be a constant matrix with all eigenvalues equal to zero; since $N\geq 2n$, clearly

$$\left[\bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^{\hat{t}_{\ell+1}-\hat{t}_\ell}=\left[\bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^N=0.$$

So, solving (63) for $\phi(\hat{t}_{\ell+1})$ yields

$$\begin{aligned}&\phi(\hat{t}_{\ell+1}) \\ &=\sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1}\left[\bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^{\hat{t}_{\ell+1}-j-1}\Bigg(B_1\left(\left[\hat{\theta}_{\sigma(\hat{t}_\ell)}(j)-\hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^T\phi(j)+e_{\sigma(\hat{t}_\ell)}(j+1)\right) \\ &\quad +B_2r(j)\Bigg). \end{aligned} \tag{64}$$

It follows from the compactness of $\mathcal{S}$ and the $\mathcal{S}_i$'s that $\left\|\left[\bar{A}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right]^j\right\|,\ j=0,1\ldots,N-1$, is bounded above by a constant which we label c_1. Using this fact together with Proposition 3 which provides a bound on the difference between parameter estimates at two different points in time, we obtain

$$\begin{aligned}\|\phi(\hat{t}_{\ell+1})\| &\leq c_1\sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1}\left(\left\|\hat{\theta}_{\sigma(\hat{t}_\ell)}(j)-\hat{\theta}_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right\|\|\phi(j)\|+|e_{\sigma(\hat{t}_\ell)}(j+1)|+|r(j)|\right) \\ &\leq c_1\sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1}\left(\left[\sum_{q=\hat{t}_\ell,\phi(q)\neq 0}^{j-1}\frac{|e_{\sigma(\hat{t}_\ell)}(q+1)|}{\|\phi(q)\|}\right]\|\phi(j)\|+|e_{\sigma(\hat{t}_\ell)}(j+1)|+|r(j)|\right). \end{aligned}$$

By definition of the prediction error, if $\phi(j)=0$ then

$$|e_i(j+1)|=|d(j)|,$$

and if $\phi(j)\neq 0$, then

$$|e_i(j+1)|=\frac{|e_i(j+1)|}{\|\phi(j)\|}\|\phi(j)\|.$$

Incorporating this into the above inequality yields

$$\begin{aligned}\|\phi(\hat{t}_{\ell+1})\| &\leq c_1 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} \left(\left[\sum_{q=\hat{t}_\ell,\phi(q)\neq 0}^{j} \frac{|e_{\sigma(\hat{t}_\ell)}(q+1)|}{\|\phi(q)\|} \right] \|\phi(j)\| + |d(j)| + |r(j)| \right) \\ &\leq c_1 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} \left(\left[\sum_{q=\hat{t}_\ell,\phi(q)\neq 0}^{\hat{t}_{\ell+1}-1} \frac{|e_{\sigma(\hat{t}_\ell)}(q+1)|}{\|\phi(q)\|} \right] \|\phi(j)\| + |d(j)| + |r(j)| \right) \\ &= c_1 \left[\sum_{q=\hat{t}_\ell,\phi(q)\neq 0}^{\hat{t}_{\ell+1}-1} \frac{|e_{\sigma(\hat{t}_\ell)}(q+1)|}{\|\phi(q)\|} \right] \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} \|\phi(j)\| + c_1 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|) \\ &\leq c_1 (\hat{t}_{\ell+1} - \hat{t}_\ell) \left[\max_{j\in[\hat{t}_\ell,\hat{t}_{\ell+1}),\phi(j)\neq 0} \frac{|e_{\sigma(\hat{t}_\ell)}(j+1)|}{\|\phi(j)\|} \right] \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} \|\phi(j)\| \\ &\quad + c_1 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|). \end{aligned} \tag{65}$$

Since $\hat{t}_{\ell+1} - \hat{t}_\ell = N$, it follows from Proposition 4 that there exists a constant c_2 so that the following holds:

$$\sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} \|\phi(j)\| \leq c_2 \|\phi(\hat{t}_\ell)\| + c_2 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-2} (|d(j)| + |r(j)|); \tag{66}$$

so, substituting (66) into (65) and using the definition of the performance signal $J_{\sigma(\hat{t}_\ell)}(\cdot)$ given in (55) it follows that there exists a constant c_3 so that

$$\begin{aligned}\|\phi(\hat{t}_{\ell+1})\| &\leq c_1 N J_{\sigma(\hat{t}_\ell)} \left(c_2 \|\phi(\hat{t}_\ell)\| + c_2 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-2} (|d(j)| + |r(j)|) \right) \\ &\quad + c_1 \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|) \\ &\leq c_3 J_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) \|\phi(\hat{t}_\ell)\| + c_3 \left(1 + J_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell)\right) \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|). \end{aligned} \tag{67}$$

Step 3: Apply Lemma 2 and Proposition 4 to obtain a bound on $\|\phi(\hat{t}_{\ell+2})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$.

From Lemma 2 either

$$J_{\sigma(\hat{t}_\ell)}(\hat{t}_\ell) \leq J_{i^*}(\hat{t}_\ell) \tag{68}$$

or

$$J_{\sigma(\hat{t}_{\ell+1})}(\hat{t}_{\ell+1}) \leq J_{i^*}(\hat{t}_{\ell+1}). \tag{69}$$

If (68) is true, then we can substitute this into (67) to obtain a bound on $\|\phi(\hat{t}_{\ell+1})\|$ in terms of $J_{i^*}(\hat{t}_\ell)$ and then apply Proposition 4 to get a bound on $\|\phi(\hat{t}_{\ell+2})\|$ in terms of $\|\phi(\hat{t}_{\ell+1})\|$ and the exogenous inputs; it follows that there exists a constant c_4 so that

$$\begin{aligned}\|\phi(\hat{t}_{\ell+2})\| &\leq c_3 c_4 J_{i^*}(\hat{t}_\ell)\|\phi(\hat{t}_\ell)\| \\ &\quad + c_3 c_4 \left(1 + J_{i^*}(\hat{t}_\ell)\right) \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|) + c_4 \sum_{j=\hat{t}_{\ell+1}}^{\hat{t}_{\ell+2}-1} (|d(j)| + |r(j)|). \end{aligned} \tag{70}$$

On the other hand, if (69) is true, we can use (67) to get a bound on $\|\phi(\hat{t}_{\ell+2})\|$ in terms of $J_{i^*}(\hat{t}_{\ell+1})$, and then apply Proposition 4 to get a bound on $\|\phi(\hat{t}_{\ell+1})\|$ in terms of $\|\phi(\hat{t}_\ell)\|$; it follows that there exists a constant c_5 so that

$$\begin{aligned}\|\phi(\hat{t}_{\ell+2})\| &\leq c_3 c_5 J_{i^*}(\hat{t}_{\ell+1})\|\phi(\hat{t}_\ell)\| + c_3 c_5 J_{i^*}(\hat{t}_{\ell+1}) \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+1}-1} (|d(j)| + |r(j)|) \\ &\quad + c_3 \left(1 + J_{i^*}(\hat{t}_{\ell+1})\right) \sum_{j=\hat{t}_{\ell+1}}^{\hat{t}_{\ell+2}-1} (|d(j)| + |r(j)|). \end{aligned} \tag{71}$$

If we define $\alpha(\hat{t}_\ell) := \max\{J_{i^*}(\hat{t}_\ell), J_{i^*}(\hat{t}_{\ell+1})\}$, then there exists a constant c_6 so that (70) and (71) can be combined to yield

$$\|\phi(\hat{t}_{\ell+2})\| \leq c_6 \alpha(\hat{t}_\ell)\|\phi(\hat{t}_\ell)\| + c_6 \left(1 + \alpha(\hat{t}_\ell)\right) \sum_{j=\hat{t}_\ell}^{\hat{t}_{\ell+2}-1} (|d(j)| + |r(j)|), \qquad \ell \in \mathbf{Z}^+. \tag{72}$$

Step 4: Analyse the first-order difference inequality (72).

First, we change notation in (72) to facilitate analysis:

$$\|\phi(\hat{t}_{2j+2})\| \leq c_6 \alpha(\hat{t}_{2j})\|\phi(\hat{t}_{2j})\| + c_6 \left(1 + \alpha(\hat{t}_{2j})\right) \sum_{q=\hat{t}_{2j}}^{\hat{t}_{2j+2}-1} (|d(q)| + |r(q)|), \qquad j \in \mathbf{Z}^+. \tag{73}$$

Next, we will analyse (73) to obtain a bound on the closed-loop behaviour; we consider two cases—one with noise and one without.

Case 1: $d(t) = 0$ for all $t \geq t_0$.

From Proposition 1 and the definition of $\alpha(\cdot)$, we have

$$\sum_{q=0}^{j-1} \alpha(\hat{t}_{2q})^2 \leq \sum_{p=t_0,\phi(p)\neq 0}^{t_0+jN-1} \frac{|e_{i^*}(p+1)|^2}{\|\phi(p)\|^2}$$
$$\leq 2[V(t_0) - V(t_0+jN)] \leq 2V(t_0) \leq 8\|\mathcal{S}_{i^*}\|^2$$
$$\leq 8\bar{\mathbf{s}}^2 =: c_7, \; j \geq 1. \tag{74}$$

If we use this bound in the second occurrence of $\alpha(\hat{t}_{2j})$ in (73), we obtain

$$\|\phi(\hat{t}_{2j+2})\| \leq c_6\alpha(\hat{t}_{2j})\|\phi(\hat{t}_{2j})\| + \underbrace{c_6(1+\sqrt{c_7})}_{=:c_8} \underbrace{\sum_{q=\hat{t}_{2j}}^{\hat{t}_{2j+2}-1} |r(q)|}_{=:\bar{r}(j)}, \qquad j \in \mathbf{Z}^+. \tag{75}$$

Since $\lambda \in (0,1)$ and $c_6 \geq 1$, then it follows that $\lambda_1 := \frac{\lambda^{2N}}{c_6} \in (0,1)$. By Lemma 3(i) if we define $c_9 := c_7^{\frac{c_7+1}{2}} (\frac{1}{\lambda_1})^{\frac{c_7}{\lambda_1^2}+1}$ and use the fact that $\alpha(\hat{t}_{2j}) \geq 0$, we see that

$$\prod_{q=0}^{j-1} \alpha(\hat{t}_{2q}) \leq c_9 \lambda_1^j \;\; j \in \mathbf{Z}^+, \tag{76}$$

which, in turn, implies that

$$\prod_{q=0}^{j-1} [c_6\alpha(\hat{t}_{2q})] \leq c_9 \lambda^{2jN}, \;\; j \in \mathbf{Z}^+. \tag{77}$$

Solving (75) iteratively and using this bound, we obtain

$$\|\phi(\hat{t}_{2j})\| \leq c_9\lambda^{2jN}\|\phi(\hat{t}_0)\| + \sum_{q=0}^{j-1} c_9 c_8 \left(\lambda^{2N}\right)^{j-1-q} \bar{r}(q), \;\; j \in \mathbf{Z}^+. \tag{78}$$

Using Proposition 4 to obtain a bound on $\phi(t)$ between $\hat{t}_{2j}$ and $\hat{t}_{2j+2}$, we conclude that there exists a constant c_{10} so that

$$\|\phi(t)\| \leq c_{10}\lambda^{t-t_0}\|\phi(t_0)\| + \sum_{j=t_0}^{t-1} c_{10}\lambda^{t-j-1}|r(j)|, \qquad t \geq t_0. \tag{79}$$

Case 2: $d(t) \neq 0$ for some $t \geq t_0$.

We now analyse the case when there is noise entering the system. Here the analysis will use a similar (but not identical) approach to that of Case 2 in the proof of Theorem 1. Motivated by Case 1, in the following we will be applying Lemma 3(ii) with a larger bound than in (74); define $c_{11} := 8(1+N)\bar{\mathbf{s}}^2$. We also define $\lambda_1 = \frac{\lambda^{2N}}{c_6}$ and $\nu := \left\lceil \frac{\frac{c_{11}+1}{2}\ln(c_{11}) + (4\frac{c_{11}}{\lambda_1^2}+1)(\ln(2)-\ln(\lambda_1))}{\ln(2)} \right\rceil$.

We now partition the timeline into two parts: one in which the noise is small versus ϕ and one where it is not. With ν defined above, we define

$$S_{\text{good}} = \left\{ j \geq t_0 : \phi(j) \neq 0 \text{ and } \frac{|d(j)|^2}{\|\phi(j)\|^2} < \frac{\bar{\mathbf{s}}^2}{\nu} \right\},$$
$$S_{\text{bad}} = \left\{ j \geq t_0 : \phi(j) = 0 \text{ or } \frac{|d(j)|^2}{\|\phi(j)\|^2} \geq \frac{\bar{\mathbf{s}}^2}{\nu} \right\};$$

clearly $\{j \in \mathbf{Z} : \quad j \geq t_0\} = S_{\text{good}} \cup S_{\text{bad}}$. We can clearly define a (possibly infinite) sequence of intervals of the form $[k_l, k_{l+1})$ which satisfy:

(i) $k_0 = t_0$ serves as the initial instant of the first interval;
(ii) $[k_l, k_{l+1})$ either belongs to S_{good} or S_{bad}; and
(iii) if $k_{l+1} \neq \infty$ and $[k_l, k_{l+1})$ belongs to S_{good} then $[k_{l+1}, k_{l+2})$ belongs to S_{bad} and vice versa.

Now we analyse the behaviour during each interval.

Sub-Case 2.1: $[k_l, k_{l+1})$ lies in S_{bad}.

Let $j \in [k_l, k_{l+1})$ be arbitrary. In this case

$$\frac{|d(j)|^2}{\|\phi(j)\|^2} \geq \frac{\bar{\mathbf{s}}^2}{\nu} \qquad \text{or} \qquad \|\phi(j)\| = 0;$$

in either case

$$\|\phi(j)\| \leq \underbrace{\frac{\nu^{\frac{1}{2}}}{\bar{\mathbf{s}}}}_{=:c_{12}} |d(j)|.$$

Also, applying Proposition 4 for one step, there exists constant c_{13} so that

$$\|\phi(j)\| \leq c_{13}(|d(j-1)| + |r(j-1)|).$$

Then for $j \in [k_l, k_{l+1})$, we have

$$\|\phi(j)\| \leq \begin{cases} c_{12}|d(j)| & j = k_l \\ c_{13}(|d(j-1)| + |r(j-1)|) & j = k_l + 1, k_l + 2, \ldots, k_{l+1}. \end{cases} \tag{80}$$

Sub-Case 2.2: $[k_l, k_{l+1})$ lies in S_{good}.

Let $j \in [k_l, k_{l+1})$ be arbitrary. First, suppose that $k_{l+1} - k_l \leq 4N$. From Proposition 4 it can be easily proven that there exists a constant c_{14} so that

$$\|\phi(t)\| \leq c_{14}\lambda^{t-k_l}\|\phi(k_l)\| + c_{14}\sum_{j=k_l}^{t-1} \lambda^{t-j-1}(|d(j)| + |r(j)|), \qquad t \in [k_l, k_{l+1}]. \tag{81}$$

Now suppose that $k_{l+1} - k_l > 4N$. This means, in particular, that there exist $j_1 < j_2$ so that

$$k_l \le \hat{t}_{2j_1} \le \hat{t}_{2j_2} \le k_{l+1}.$$

To proceed, observe that $\|\phi(j)\| \neq 0$ and

$$\frac{|d(j)|^2}{\|\phi(j)\|^2} < \frac{\bar{\mathbf{s}}^2}{\nu},\ \ j \in [k_l, k_{l+1}). \tag{82}$$

With $0 \le j_1 < j_2$, it follows from Proposition 1 and the definition of $\alpha(\cdot)$ that

$$\begin{aligned}\sum_{q=j_1}^{j_2-1} \alpha(\hat{t}_{2q})^2 &\le \sum_{p=t_0+2j_1N,\phi(p)\neq 0}^{t_0+2j_2N-1} \frac{|e_{i^*}(p+1)|^2}{\|\phi(p)\|^2} \\ &\le 2V(\hat{t}_{2j_1}) + 4 \sum_{p=\hat{t}_{2j_1},\phi(p)\neq 0}^{\hat{t}_{2j_2}-1} \frac{|d(p)|^2}{\|\phi(p)\|^2};\end{aligned} \tag{83}$$

using the bound given in (82) which holds on $[k_l, k_{l+1})$, this becomes

$$\begin{aligned}&\sum_{q=j_1}^{j_2-1} \alpha(\hat{t}_{2q})^2 \\ &\le 8\bar{\mathbf{s}}^2 + 8N(j_2 - j_1)\frac{\bar{\mathbf{s}}^2}{\nu}, \quad \text{for all } j_1, j_2 \in \mathbf{Z}^+ \text{ s.t. } k_l \le \hat{t}_{2j_1} < \hat{t}_{2j_2} \le k_{l+1}.\end{aligned} \tag{84}$$

If $j_2 - j_1 \le \nu$ then

$$\sum_{q=j_1}^{j_2-1} \alpha(\hat{t}_{2q})^2 \le 8\bar{\mathbf{s}}^2 + 8N\nu\frac{\bar{\mathbf{s}}^2}{\nu} = (8+8N)\bar{\mathbf{s}}^2 = c_{11}; \tag{85}$$

so by Lemma 3(i), with λ_1 defined above, if we define $c_{15} := c_{11}^{\frac{c_{11}+1}{2}} (\frac{2}{\lambda_1})^{\frac{4c_{11}}{\lambda_1^2}+1}$, then

$$\prod_{q=j_1}^{j_2-1} [c_6\alpha(\hat{t}_{2q})] \le c_{15}\lambda^{2N(j_2-j_1)}, \quad \text{for all } j_1, j_2 \in \mathbf{Z}^+ \text{ s.t. } k_l \le \hat{t}_{2j_1} < \hat{t}_{2j_2} \le k_{l+1}. \tag{86}$$

If $j_2 - j_1 > \nu$, then by Lemma 3(ii) and our choice of ν we have that

$$\prod_{q=j_1}^{j_2-1} [c_6\alpha(\hat{t}_{2q})] \le c_{15}\lambda^{2N(j_2-j_1)}, \quad \text{for all } j_1, j_2 \in \mathbf{Z}^+ \text{ s.t. } k_l \le \hat{t}_{2j_1} < \hat{t}_{2j_2} \le k_{l+1}, \tag{87}$$

as well.

Now we can proceed to solve (73). The first step is to use (85) to bound the second occurrence of $\alpha(\hat{t}_{2j})$ in (73), yielding

$$\|\phi(\hat{t}_{2j+2})\| \leq c_6\alpha(\hat{t}_{2j})\|\phi(\hat{t}_{2j})\| + \underbrace{c_6(1+\sqrt{c_{11}})}_{=:c_{16}} \underbrace{\sum_{q=\hat{t}_{2j}}^{\hat{t}_{2j+2}-1} (|r(q)| + |d(q)|)}_{=:\bar{w}(j)}. \tag{88}$$

If we solve this iteratively and use the bounds in (86) and (87), we see that

$$\|\phi(\hat{t}_{2j_2})\| \leq c_{15}\lambda^{2N(j_2-j_1)}\|\phi(\hat{t}_{2j_1})\| + \sum_{q=j_1}^{j_2-1} c_{16}c_{15}\left(\lambda^{2N}\right)^{j_2-1-q} \bar{w}(q),$$
$$\text{for all } j_1, j_2 \in \mathbf{Z}^+ \text{ s.t. } k_l \leq \hat{t}_{2j_1} < \hat{t}_{2j_2} \leq k_{l+1}. \tag{89}$$

We can now use Proposition 4:

- to provide a bound on $\|\phi(t)\|$ between consecutive $\hat{t}_{2j}$'s;
- to provide a bound on $\|\phi(t)\|$ on the beginning part of the interval $[k_l, k_{l+1})$ (until we get to the first admissible $\hat{t}_{2j}$);
- to provide a bound on $\|\phi(t)\|$ on the last part of the interval $[k_l, k_{l+1})$ (after the last admissible $\hat{t}_{2j}$).

We conclude that there exist a constant $c_{17} \geq c_{14}$ so that

$$\|\phi(t)\| \leq c_{17}\lambda^{t-k_l}\|\phi(k_l)\| + c_{17}\sum_{j=k_l}^{t-1} \lambda^{t-j-1}(|d(j)| + |r(j)|), \qquad t \in [k_l, k_{l+1}]. \tag{90}$$

Now we combine Sub-Case 2.1 and Sub-Case 2.2 into a general bound on ϕ. The following analysis is almost identical to the one at the end of the proof of Theorem 1. Define $c_{18} := \max\{c_{17}, c_{13}, c_{13}c_{17}\}$.

Claim The following bound holds:

$$\|\phi(t)\| \leq c_{18}\lambda^{t-t_0}\|\phi(t_0)\| + \sum_{j=t_0}^{t-1} c_{18}\lambda^{t-j-1}(|d(j)| + |r(j)|), \qquad t \geq t_0. \tag{91}$$

Proof of the Claim If $[k_0, k_1) = [t_0, k_1) \subset S_{\text{good}}$, then (91) is true for $t \in [k_0, k_1]$ by (90). If $[k_0, k_1) \subset S_{\text{bad}}$, then from (80) we obtain

$$\|\phi(j)\| \leq \begin{cases} \|\phi(k_0)\| = \|\phi_0\| & j = k_0 = t_0 \\ c_{13}(|d(j-1)| + |r(j-1)|) & j = k_0+1, k_0+2, \ldots, k_1, \end{cases}$$

which means that (91) holds on $[k_0, k_1]$ for this case as well.

We now use induction: suppose that (91) is true for $t \in [k_0, k_l]$; we need to prove it holds for $t \in (k_l, k_{l+1}]$ as well. If $k \in [k_l, k_{l+1}) \subset S_{\text{bad}}$, then from (80) we see that

$$\|\phi(j)\| \leq c_{13}(|d(j-1)| + |r(j-1)|), \;\; j = k_l + 1, k_l + 2, \ldots, k_{l+1},$$

which means (91) holds on $(k_l, k_{l+1}]$. On the other hand, if $[k_l, k_{l+1}) \subset S_{\text{good}}$, then $k_l - 1 \in S_{\text{bad}}$; from (80) we have that

$$|\phi(k_l)| \leq c_{13}(|d(k_l - 1)| + |r(k_l - 1)|).$$

Using (90) to analyse the behaviour on $[k_l, k_{l+1}]$, we have

$$\begin{aligned} \|\phi(k)\| &\leq c_{15}\lambda^{k-k_l}[c_{13}(|d(k_l-1)|+|r(k_l-1)|)]+\sum_{j=k_l}^{k-1} c_{17}\lambda^{k-j-1}(|d(j)|+|r(j)|), \\ &\leq c_{18} \sum_{j=k_l-1}^{k-1} \lambda^{k-j-1}(|d(j)| + |r(j)|), \qquad k \in [k_l, k_{l+1}], \end{aligned} \tag{92}$$

which implies that (91) holds. □

This concludes the proof. □

References

1. Feuer A, Morse AS (1978) Adaptive control of single-input, single-output linear systems. IEEE Trans Autom Control 23(4):557–569
2. Morse AS (1980) Global stability of parameter-adaptive control systems. IEEE Trans Autom Control 25:433–439
3. Goodwin GC, Ramadge PJ, Caines PE (1980) Discrete time multivariable control. IEEE Trans Autom Control 25:449–456
4. Narendra KS, Lin YH (1980) Stable discrete adaptive control. IEEE Trans Autom Control 25(3):456–461
5. Narendra KS, Lin YH, Valavani LS (1980) Stable adaptive controller design, Part II: proof of stability. IEEE Trans Autom Control 25:440–448
6. Rohrs CE et al (1985) Robustness of continuous-time adaptive control algorithms in the presence of unmodelled dynamics. IEEE Trans Autom Control 30:881–889
7. Middleton RH, Goodwin GC (1988) Adaptive control of time-varying linear systems. IEEE Trans Autom Control 33(2):150–155
8. Middleton RH et al (1988) Design issues in adaptive control. IEEE Trans Autom Control 33(1):50–58
9. Tsakalis KS, Ioannou PA (1989) Adaptive control of linear time-varying plants: a new model reference controller structure. IEEE Trans Autom Control 34(10):1038–1046
10. Kreisselmeier G, Anderson BDO (1986) Robust model reference adaptive control. IEEE Trans Autom Control 31:127–133
11. Ioannou PA, Tsakalis KS (1986) A robust direct adaptive controller. IEEE Trans Autom Control 31(11):1033–1043
12. Ydstie BE (1989) Stability of discrete-time MRAC revisited. Syst Control Lett 13:429–439
13. Ydstie BE (1992) Transient performance and robustness of direct adaptive control. IEEE Trans Autom Control 37(8):1091–1105
14. Naik SM, Kumar PR, Ydstie BE (1992) Robust continuous-time adaptive control by parameter projection. IEEE Trans Autom Control 37:182–297

15. Wen C, Hill DJ (1992) Global boundedness of discrete-time adaptive control using parameter projection. Automatica 28(2):1143–1158
16. Wen C (1994) A robust adaptive controller with minimal modifications for discrete time-varying systems. IEEE Trans Autom Control 39(5):987–991
17. Li Y, Chen H-F (1996) Robust adaptive pole placement for linear time-varying systems. IEEE Trans Autom Control 41:714–719
18. Narendra KS, Annawswamy AM (1987) A new adaptive law for robust adaptation without persistent excitation. IEEE Trans Autom Control 32:134–145
19. Fu M, Barmish BR (1986) Adaptive stabilization of linear systems via switching control. IEEE Trans Autom Control AC–31:1097–1103
20. Miller DE, Davison EJ (1989) An adaptive controller which provides Lyapunov stability. IEEE Trans Autom Control 34:599–609
21. Morse AS (1996) Supervisory control of families of linear set-point controllers–Part 1: exact matching. IEEE Trans Autom Control 41:1413–1431
22. Morse AS (1997) Supervisory control of families of linear set-point controllers–Part 2: robustness. IEEE Trans Autom Control 42:1500–1515
23. Hespanha JP, Liberzon D, Morse AS (2003) Hysteresis-based switching algorithms for supervisory control of uncertain systems. Automatica 39:263–272
24. Vu L, Chatterjee D, Liberzon D (2007) Input-to-state stability of switched systems and switching adaptive control. Automatica 43:639–646
25. Hespanha JP, Liberzon D, Morse AS (2003) Overcoming the limitations of adaptive control by means of logic-based switching. Syst Control Lett 49(1):49–65
26. Morse AS (1998) A bound for the disturbance to tracking error gain of a supervised set-point control system. In: Normand-Cyrot D (ed) Perspectives in control. Springer, London
27. Vu L, Liberzon D (2011) Supervisory control of uncertain linear time-varying systems. IEEE Trans Automat Control 1:27–42
28. Zhivoglyadov PV, Middleton RH, Fu M (2001) Further results on localization-based switching adaptive control. Automatica 37:257–263
29. Miller DE (2003) A new approach to model reference adaptive control. IEEE Trans Autom Control 48:743–757
30. Miller DE (2006) Near optimal LQR performance for a compact set of plants. IEEE Trans Autom Control 51:1423–1439
31. Vale JR, Miller DE (2011) Step tracking in the presence of persistent plant changes. IEEE Trans Autom Control 56:43–58
32. Miller DE (2017) A parameter adaptive controller which provides exponential stability: the first order case. Syst Control Lett 103:23–31
33. Miller DE (2017) Classical discrete-time adaptive control revisited: exponential stabilization. In: 1st IEEE conference on control technology and applications, pp 1975–1980. The submitted version is posted at arXiv:1705.01494
34. Goodwin GC, Sin KS (1984) Adaptive filtering prediction and control. Prentice Hall, Englewood Cliffs
35. Jerbi A, Kamen EW, Dorsey J (1993) Construction of a robust adaptive regulator for time-varying discrete-time systems. Int J Adapt Control 7:1–12
36. Praly L (1984) Towards a globally stable direct adaptive control scheme for not necessarily minimum phase system. IEEE Trans Autom Control 29:946–949
37. Kreisselmeier G (1986) Adaptive control of a class of slowly time-varying plants. Syst Control Lett 8:97–103
38. Kreisselmeier G, Smith MC (1986) Stable adaptive regulation of arbitrary n^{th}-order plants. IEEE Trans Autom Control 31:299–305
39. Zhou K, Doyle JC, Glover K (1995) Robust and optimal control. Prentice Hall, New Jersey
40. Desoer CA (1970) Slowly varying discrete time system $x_{t+1} = A_t x_t$. Electron Lett 6(11):339–340
41. Shahab MT, Miller DE (2018) Multi-estimator based adaptive control which provides exponential stability: the first-order case. In: 2018 IEEE 57th conference on decision and control (to appear)