Kybernetika

Kostas S. Tsakalis

Bursting scenaria in adaptive algorithms: performance limitations and some remedies

Kybernetika, Vol. 33 (1997), No. 1, 17--40

Persistent URL: `http://dml.cz/dmlcz/125667`

KYBERNETIKA — VOLUME 33 (1997), NUMBER 1, PAGES 17–40

BURSTING SCENARIA IN ADAPTIVE ALGORITHMS: PERFORMANCE LIMITATIONS AND SOME REMEDIES[1]

KOSTAS S. TSAKALIS

A simple, yet general, bursting scenario is presented for a wide class of parameter estimation and system identification algorithms in the absence of sufficient excitation. This allows for an analytical derivation of a lower bound on the worst-case performance of such algorithms in the presence of perturbations. A simple example is constructed to illustrate the implications of these results in adaptive control and interpret the design some burst suppression mechanisms.

1. INTRODUCTION

Adaptive identification algorithms are a fundamental component of most adaptive control schemes where the basic idea is to use input-output (I/O) data to identify on-line an appropriate I/O operator (either of the plant or the desired controller). This is typically performed by deriving (or assuming) a parametric model of the plant and then employing an algorithm to estimate the unknown parameters. The parameter estimation algorithm is designed by using fairly standard optimization tools, e.g. gradient or Newton search, least squares etc. The properties of such algorithms in the context of system identification have been extensively studied, establishing their applicability to a variety of practical problems [15, 30].

In the context of adaptive control the same ideas have also proven successful in achieving the control objective despite the presence of pure parametric uncertainty in the plant model [2, 10, 12, 20, 27]. However, a fundamental and serious problem arises when non-parametric forms of uncertainty appear in the plant description, e.g., unmodeled dynamics and bounded disturbances. In such cases, analytical examples and simulation studies have shown that the original adaptive control algorithms may fail to guarantee boundedness of the parameter estimates and the other closed-loop signals [7, 11, 25, 26] These phenomena are caused by the lack of "sufficient" excitation which allows the perturbations to dominate the error signal and cause the failure of the identification algorithm to obtain a "good" model of the plant. The fundamental obstacle and difference from open-loop system identification in

[1] This work was supported in part by ARPA under grant F 49620-93-1-0062 and in part by NSF under RIA grant ECS-9111346.

overcoming such problems is that the designer has limited or no control over the external inputs and, consequently, the level of excitation. Nevertheless, a variety of recent studies has established that with some modifications, the basic identification algorithms can yield "robust" adaptive controllers without requiring any excitation conditions (e.g., see [7, 12, 21, 27] and references therein). A similar result has also been established in the practically interesting case where the plant is slowly time-varying [14, 19, 34].

However, the performance of these adaptive control schemes is typically characterized by fairly weak measures such as root-mean-square (RMS) criteria. The implication of this observation is that the closed-loop performance may be poor in terms of stronger but practically important measures such as peak steady-state error. (Such a performance measure can be conveniently characterized by the lim sup absolute value of the error and, hence, is referred to as "lim sup performance;" note that, like RMS, this is an asymptotic performance measure and does not account for transient behavior.) This was found to be the case in situations where a disturbance together with the lack of sufficiently high levels of Persistent Excitation (PE) causes the identification process to fail, at least temporarily. Although signal boundedness is maintained with the modified algorithms the identification failure is now manifested by short but persistent time intervals where the various error signals attain large values. The term "large" is used here to signify a magnitude that does not vanish as the magnitude of the perturbation approaches zero. Such a behavior is typically referred to as burst phenomena [1, 8, 17, 24, 29, 36]. Partial remedies include the use of dead-zones with linear time-invariant (LTI) plants e.g., [7, 10, 12, 16, 20, 23]. On the other hand, employing a form of high-gain feedback, an improvement of the tracking error lim sup performance has recently been obtained in model reference adaptive control, but at the expense of the closed-loop robustness properties [5, 31]. Despite the (partial) success of these remedies, however, bursting still remains as one of the major obstacles in designing practically useful and reliable adaptive algorithms.

In view of these results, one may pose the natural question of whether an adaptive controller can be found to provide practical lim sup performance guarantees in the absence of any excitation conditions. At this point, the possibility that an affirmative answer to this question exists is, at best, remote. For example, [8, 29] studied bursting phenomena involving adaptation in two different environments, albeit with the same conclusion. That is, in the presence of disturbances, basic gradient laws with small adaptation gains can result in a bursting behavior.

Motivated by these studies, in this paper we adopt a different point of view, namely, that bursting is a consequence of the optimization objective in the parameter estimation process rather than the form of the estimator itself. More specifically, we address the problem of fundamental performance limitations of the parameter estimation/identification process occurring in environments where perturbations are present but there is lack of "sufficient" excitation. We begin by considering the standard linear-model parameter estimation problem where we provide an analytical method to construct bursting scenaria for a wide class of adaptive algorithms. Based on this simple bursting mechanism we obtain lower bounds on the worst-case lim sup performance of adaptive algorithms in various situations arising in parameter estimation and system identification problems. In all cases, our results show that in the absence of any input constraints, arbitrarily small perturbations, such as bounded disturbances, unmodeled dynamics, and slow time variations of the system parameters, impose a fundamental performance limitation. This limitation is rather severe in the sense that the worst-case lim sup performance deteriorates proportionally with the size of the parametric uncertainty set. Finally, guided by the results of our analysis, we construct a simple example to illustrate the appearance of bursting phenomena in adaptive control. Viewing the violation of the developed sufficient conditions for bursting as necessary conditions for burst suppression, we also discuss some of the possible mechanisms that can lead to the design of adaptive laws that offer "reasonable" lim sup performance guarantees.

2. BURSTING IN PARAMETER ESTIMATION

2.1. LTI models

Let us consider the linear process model with output disturbance

$$y = w^\top \theta_* + d \tag{1}$$

where $y : \mathbb{R}_+ \mapsto \mathbb{R}$ is the output of the process, $w : \mathbb{R}_+ \mapsto \mathbb{R}^n$ is a vector of signals (regressor) available for measurement, $\theta_* \in \mathbb{R}^n$ is an unknown constant parameter vector and d is an unknown disturbance. For such a process the standard parameter estimation problem is to design an algorithm to estimate θ_*, given the measurements y and w.

Denoting by θ the current estimate of θ_*, the "quality" of this estimate is simply its distance from θ_*. However, since the latter is unknown, a typical measure of the quality of the estimate θ is given in terms of the estimation error

$$\epsilon_1 = \hat{y} - y = w^\top \theta - y = w^\top \phi - d \tag{2}$$

where ϕ is the parameter error $\theta - \theta_*$.

This problem is fairly standard and is encountered in several applications ranging from modeling and system identification problems to echo cancellation/noise attenuation and adaptive control problems. In the case of system identification, the input vector w is largely at the disposal of the designer and several studies can be found addressing the problem of selecting the input in order to minimize (in some sense) the effect of the disturbance in the identification/estimation process [15, 30].

On the other hand, there are several important applications where the designer has little or no access to the process inputs, e. g., when the parameter estimator is part of a closed-loop control system. In such cases, estimation algorithms may produce periodic bursting of the estimation error or, even, an unbounded parameter drift [1, 29]. This phenomenon has a simple interpretation from an optimization point of view: the argument of the minimization of, say, ϵ_1^2 over θ is not continuous at $d = 0$, $w = 0$. This simple observation can be used to motivate a constructive proof of estimation error bursting. For this purpose, we consider a class of parameter estimation algorithms $\mathcal{A}$ that, loosely speaking, minimize a "fading memory" functional of the estimation error with "finite" speed of adaptation. More precisely,

2.1. Assumption. $\mathcal{A}$ is a parameter estimation algorithm for the linear model $y = w^\top \theta_*$, generating parameter estimates θ such that:

1. $\dot{\theta}(t) = \mathcal{A}[(y)_t, (w)_t, (\theta)_t]$, where $(\cdot)_t$ denotes truncation at t.

2. Given any bounded, piecewise continuous y, w for which there exist positive constants t_0, T, δ_w and a constant vector θ_* such that for all $t \geq t_0$,

$$\int_t^{t+T} w(\tau)\, w^\top(\tau)\, \mathrm{d}\tau \geq \delta_w I \tag{3}$$

$$y(t) = w^\top(t)\, \theta_* \tag{4}$$

   the parameter estimates θ converge to θ_*, for any $\theta(0)$.

3. Whenever $\|(y)_t\|_\infty \leq c_y$, $\|(w)_t\|_\infty \leq c_w$, $\|(\theta)_t\|_\infty \leq c_\theta$ where c_y, c_w, c_θ are (finite) constants, there exists a (finite) constant Γ such that $|\dot{\theta}(t)| \leq \Gamma$

4. Suppose that a convex, closed and bounded set $\mathcal{M} \subset \mathbb{R}^n$ such that $\theta_* \in \mathcal{M}$ is known a priori. Then, in addition to the above properties, the parameter estimates θ generated by $\mathcal{A}$ remain in $\mathcal{M}$ for all $t \geq 0$.

Adaptive algorithms that satisfy the seemingly reasonable and perhaps desirable Assumption 2.1, are susceptible to "bursting" as quantified by the following result.

2.2. Proposition. Consider the case where an algorithm $\mathcal{A}$ is used to estimate the parameter vector θ_* of the perturbed linear model (1).

1. Suppose that $\mathcal{A}$ satisfies Assumption 2.1, 1–2. Then for any $\delta > 0$ and any $\theta_0 \in \mathbb{R}^n$, there exist bounded, piecewise continuous w, d with $\|d\|_\infty \leq \delta$ and such that $\theta \to \theta_0$ as $t \to \infty$.

2. Suppose that $\mathcal{A}$ satisfies Assumption 2.1 and $\theta_* \in \mathcal{M}$. Then for any $\delta > 0$, there exist bounded, piecewise continuous w, d with $\|d\|_\infty \leq \delta$ and such that $\limsup_{t\to\infty} |\epsilon_1| = \|w\|_\infty \max_{\theta\in\mathcal{M}} |\theta - \theta_*| + \delta$.

Proof. (1) Let w be such that (3) is satisfied and $|w^\top(\theta_* - \theta_0)| \leq \delta$. Note that such an w can always be found, e. g., $w = w_e\delta/|\theta_* - \theta_0|$ where w_e is PE (satisfies (3)) and $|w_e| \leq 1$. Further, define $d = w^\top(\theta_0 - \theta_*)$; clearly $\|d\|_\infty \leq \delta$. With this choice, (1) becomes $y = w^\top\theta_0$ which, by the properties of $\mathcal{A}$, implies that $\theta \to \theta_0$ as $t \to \infty$.

(2) Let $\theta_0 = \arg\max_{\theta\in\mathcal{M}} |\theta - \theta_*|$ and define $c_w = \|w\|_\infty$. In view of Part 1 of the proposition, given $\epsilon > 0$ there exist w, d : $\|d\|_\infty \leq \delta$ and a time T_1 such that $|\theta(T_1) - \theta_0| \leq \epsilon/c_w$. Next, let $w(T_1) = c_w(\theta_0 - \theta_*)/|\theta_0 - \theta_*|$ and $d(T_1) = -\delta$. Since $\dot{\theta}$ is bounded, it follows that $\epsilon_1(T_1) \geq c_w|\theta_0 - \theta_*| + \delta - \epsilon$. The same principle can be invoked to establish, using an induction argument, the existence of a sequence T_i where ϵ_1 satisfies the above inequality. Further, by letting $\epsilon = 1/2^i$, we obtain the

desired right hand-side while equality follows from the fact that the latter is also an upper bound of ϵ_1. □

This result provides a constructive proof that arbitrarily small bounded disturbances can cause a class of adaptive algorithms to exhibit burst phenomena in the absence of any excitation (or other) conditions on the input. Moreover, as the magnitude of the disturbance approaches zero, the worst-case lim sup performance of the estimation error approaches a constant which depends only on the parametric uncertainty set and the magnitude of the input signal but is independent of the disturbance bound.

2.2. LTV models

One of the most important justifications behind the study of adaptive algorithms has traditionally relied on their intuitive applicability in slowly time-varying environments. In this case, the analysis of several adaptive laws has produced results analogous to those for LTI models with disturbances, with the notable exception that dead-zone-like remedies are now unable to provide the corresponding lim sup performance guarantees for the estimation error. This is, in fact, a fundamental problem in the absence of any excitation conditions, regardless of the existence of any other perturbation terms. One simple explanation of this problem is as follows: During a period of insufficient excitation, there is a nontrivial manifold where the contribution of the parameter error to the estimation error is zero. Since all standard algorithms rely on such an error signal to assess the quality of the parameter estimates, the actual parameters may drift in a way that does not contribute any information to the estimation process. Thus, the estimator is "blind" to such parameter drift and an error burst will occur as soon as the excitation changes direction/magnitude revealing the current value of the actual parameters.

To quantify this simple argument, let us consider the linear time-varying (LTV) process model

$$y = w^{\top}\theta_* \tag{5}$$

where $y : \mathbb{R}_+ \mapsto \mathbb{R}$ is the output of the process, $w : \mathbb{R}_+ \mapsto \mathbb{R}^n$ is the regressor vector and $\theta_* : \mathbb{R}_+ \mapsto \mathbb{R}^n$ is the unknown time-varying parameter vector. The speed of variation of the unknown parameters can be characterized in a simple way by the magnitude of their derivative. For example, assuming that μ is a positive constant such that $\|\dot{\theta}_*\|_\infty \leq \mu$, smaller values of μ indicate slower varying parameters. Next, we consider algorithms that satisfy the following:

2.3. Assumption. $\mathcal{A}$ is a parameter estimation algorithm satisfying Assumption 2.1 with the following modifications:

(i) In 2.1,2, and whenever θ_* is constant for $t \geq t_0$, θ is only required to converge to a residual set

$$\mathcal{B} = \left\{\theta : |\theta - \theta_*| \leq d_z\sqrt{T/\delta_w}\right\}$$

for any $\theta(0)$, where $d_z \geq 0$ is a constant.

(ii) In 2.1,3, there also exists a (finite) constant Γ' such that $|\dot{\theta}(t)| \leq \Gamma'|y(t) - w^{\top}(t)\theta(t)|$.

This assumption is weaker than the one used in the previous section in that, under persistent excitation, asymptotic convergence of the parameter error to zero is not required. Instead the parameters are allowed to converge to a residual set of nonzero radius, thus including dead-zone-like algorithms. (Note that the expression used above is inspired by the typical dead-zone estimator with threshold d_z where the radius of the residual set is $d_z\sqrt{T/\delta_w}$.) On the other hand, a stronger condition is used in part 3 of the assumption which essentially reflects the fact that the quality of the parameter estimates is inferred by the instantaneous estimation error. This part of the assumption can be relaxed to include algorithms minimizing an exponentially weighted L_2 norm of the error [13] or an error functional over a finite moving window.

2.4. Proposition. Consider the case where an algorithm $\mathcal{A}$ is used to estimate the TV parameter vector θ_* of the linear model (5) and suppose that $\mathcal{A}$ satisfies Assumption 2.3. Then for any $\mu > 0$, there exist bounded, piecewise continuous w and $\theta_* : \mathbb{R}_+ \mapsto \mathcal{M}$ with $\|\dot{\theta}_*\|_\infty \leq \mu$, such that $\limsup_{t\to\infty} |\epsilon_1| \geq \|w\|_\infty \operatorname{diam}\mathcal{M} - d_z\sqrt{n}$.

Proof. Let $\theta_1, \theta_2 \in \mathcal{M}$ such that $|\theta_1 - \theta_2| = \operatorname{diam}\mathcal{M}$. Also let w_1 be such that (3) is satisfied and $|w_1(t)| \leq c_w$, $\forall t$. Then, for any $\epsilon > 0$, there exists T_1 such that when $\mathcal{A}$ is applied to the model (5) with $\theta_* = \theta_1$ and $w = w_1$, $|\theta(T_1) - \theta_1| < d_z\sqrt{T/\delta_w} + \epsilon$. Next, define $w = w_1$, $\theta_*(t) = \theta_1$ in the interval $[0, T_1)$ and $w = 0$, $\theta_*(t) = \theta_1 + \mu(\theta_2 - \theta_1)(t - T_1)/\operatorname{diam}\mathcal{M}$ in the interval $[T_1, T_2)$, where $T_2 = T_1 + \operatorname{diam}\mathcal{M}/\mu$. Then at time T_2, $\theta_*(T_2) = \theta_2$ and $\theta(T_2) = \theta(T_1)$. Hence, choosing $w(T_2) = c_w(\theta_1 - \theta_2)/|\theta_1 - \theta_2|$, we have that $\epsilon_1(T_2) \geq c_w \operatorname{diam}\mathcal{M} - c_w(d_z\sqrt{T/\delta_w} + \epsilon)$. Clearly, the sequence can be repeated ad infinitum with θ_* oscillating between θ_1 and θ_2. Hence, $\limsup_{t\to\infty} |\epsilon_1| \geq c_w(\operatorname{diam}\mathcal{M} - d_z\sqrt{T/\delta_w})$. Finally, for the special case where during the excitation intervals w attains its maximum magnitude in the direction of each unit vector for a subinterval of length T/n, it follows that $\delta_w = c_w^2 T/n$ yielding the desired expression. □

For adaptive algorithms satisfying Assumption 2.1 ($d_z = 0$) this lower bound on the worst case performance is sharp. On the other hand, for dead-zone-like algorithms ($d_z > 0$) the bound given in the proposition is conservative and makes sense only when d_z is small relative to $\operatorname{diam}\mathcal{M}$. Less conservative bounds or bounds independent of d_z can be derived for specific cases, e. g., for the standard dead-zone algorithm $\limsup_{t\to\infty} |\epsilon_1| \geq \frac{1}{2}\|w\|_\infty \operatorname{diam}\mathcal{M}$. Nevertheless, Proposition 2.4 conveys an important qualitative message, that is, the size of the parametric uncertainty set imposes a fundamental worst-case lim sup performance limitation for a general class of adaptive algorithms operating in TV environments.[1]

[1]Notice that Proposition 2.4 also provides a rigorous proof of the conjecture that dead-zone techniques do not provide any lim sup performance guarantees in the TV case.

3. BURSTING IN SYSTEM IDENTIFICATION

In this section we briefly discuss the implications of the above observations and results in the identification of linear systems. In particular, we consider the case where the I/O map of a linear system is identified via adaptive linear-model parameter estimation. This approach amounts to expressing the I/O relationship in terms of a standard linear model and using a parameter estimator to estimate the partially unknown parameters. For example, given a uniformly observable linear system $[A, b, c]$ (possibly time-varying) we can always express the I/O relationship $u \mapsto y$ in the form

$$\dot{x} = Fx + \theta_{1*}u + \theta_{2*}y \ ; \quad y = q^\top x \tag{6}$$

where F is a Hurwitz matrix and (q, F) is a completely observable pair at the disposal of the designer (e. g. see [34]). Further, using the definitions

$$\begin{aligned} w &= [G(s)\,[Iu],\ G(s)\,[Iy]]^\top \\ \eta &= G(s)\left\{G'(s)\,[Iu]\,\dot{\theta}_{1*} + G'(s)\,[Iy]\,\dot{\theta}_{2*}\right\} \end{aligned}$$

where $G(s) = q^\top(sI - F)^{-1}$, $G'(s) = (sI - F)^{-1}$ the above relationship assumes the convenient linear-model form

$$y = w^\top \theta_* - \eta + \varepsilon_t. \tag{7}$$

Here, ε_t denotes exponentially decaying terms due to initial conditions and η is a perturbation due to the swapping of the possibly time-varying parameters.

From (7) it becomes apparent that linear-model parameter estimation algorithms can be employed in performing a (parametric) identification of the system (6). In such a case we distinguish two types of error signals measuring the quality of the parameter estimation and identification processes. One is the usual *estimation error* $\epsilon_1 = w^\top\theta - y$ driving the parameter estimator. The other is the *identification error* $e_1 = G(s)\,[u\theta_1 + y\theta_2] - y$ arising when the estimates θ are interpreted as an I/O operator and serves as an approximation error of (6). The relationship between these two errors is given by $e_1 = \epsilon_1 - \hat{\eta}$ where $\hat{\eta} = G(s)\{G'(s)\,[uI]\,\dot{\theta}_1 + G'(s)\,[yI]\,\dot{\theta}_2\}$.

In this framework, we are interested in assessing the performance limitations of system identification algorithms applied to a perturbed version of (6). For simplicity, throughout the rest of our discussion we assume that the system (6) is exponentially stable. For the same reason, we need to further restrict the class of algorithms under consideration by introducing the following technical condition.

3.1. Assumption. In Assumptions 2.1 and 2.3 equation (4) is replaced by $y(t) = w^\top(t)\theta_* + \varepsilon_t$ where ε_t is any exponentially decaying term. Furthermore, there exists a constant $\gamma > 0$ such that the quantity Γ in Assumptions 2.1 and 2.3 satisfies $\Gamma \leq \gamma|\theta(t) - \theta_*(t)|$, uniformly in $||u_t||_\infty$.

Under this condition, it is possible to extend the bursting scenaria of the previous section to the system identification process. Note, however, that some technical modifications are required to account for the specific way that the perturbations enter the system as well as the fact that the regressor vector w can only be manipulated through the input u. In particular, the latter constraint takes the form of a minimum time required for w to be steered from the origin to any point on a ball in $\mathbb{R}^n$ whose radius depends on the bound of $\|u\|_\infty$. (Notice that w is controllable from u [21, 27].) With this observation, conservative but intuitively appealing statements on the performance limitations of a class of adaptive identification algorithms are given below.

3.2. Proposition. [33] Consider the case where the system (6), perturbed by setting $y = q^\top x + d$, is identified by means of a parameter estimation algorithm $\mathcal{A}$, which is designed based on the linear model $y = w^\top \theta_*$. Further, suppose that $\theta_* \in \mathcal{M}$ and $\mathcal{A}$ satisfies Assumptions 2.1 and 3.1. Then for any $\delta > 0$, there exist bounded, piecewise continuous u, d with $\|d\|_\infty \leq \delta$ and such that

$$\begin{aligned}
\limsup_{t\to\infty} |\epsilon_1| &\geq \|u\|_\infty \max_{\theta\in\mathcal{M}} |\theta - \theta_*| C_u \left[2 - e^{\gamma\tau_u}\right] \\
\limsup_{t\to\infty} |e_1| &\geq \|u\|_\infty \max_{\theta\in\mathcal{M}} |\theta - \theta_*| [C_u(2 - e^{\gamma\tau_u}) - O(\gamma/(\alpha+\gamma))(e^{\gamma\tau_u} - e^{-\alpha\tau_u})] \\
\limsup_{t\to\infty} |e_1| &\geq O\left(M_r/(1+\gamma)\right) \|u\|_\infty
\end{aligned}$$

where C_u, τ_u, α are positive constants depending on the bound of u, the system (6) and the regressor filters; $M_r = \max r$ s.t. $\{\theta : |\theta - \theta_*| \leq r\} \subseteq \mathcal{M}$.

The proof of the proposition follows the same basic idea outlined in Proposition 1, except that during the bursting phase the regressor vector must be driven to the desired value by u; since this process consumes time τ_u, the maximum possible adjustment of the estimated parameters must also be taken into account, e. g., using the Bellman–Gronwall Lemma.

Similar results are obtained for the adaptive identification problem of an LTV plant of the form (6) where (q, F) is an observable pair and $\theta_* : \mathbb{R}_+ \mapsto \mathcal{M}$. Here, to ensure that this identification problem makes sense, we need to impose some restrictions on the set of admissible parameter vectors θ_*. For example, in a typical identification problem such a condition may be expressed as

$$\theta_* : \mathbb{R}_+ \mapsto \mathcal{M}' \subseteq \mathcal{M} \ ; \quad \|\dot{\theta}_*\|_\infty \leq \mu_0$$

for some $\mu_0 > 0$, where $\mathcal{M}'$ denotes the largest (in diameter) connected part of $\mathcal{M}$ such that any $\theta \in \mathcal{M}'$ corresponds to a system that is pointwise strongly controllable and observable and exponentially stable, uniformly in $\theta \in \mathcal{M}'$.

3.3. Proposition. [33] Consider the case where the system (6) is identified by means of a parameter estimation algorithm $\mathcal{A}$, which is designed based on the linear model $y = w^\top \theta_*$. Suppose that $\mathcal{A}$ satisfies Assumptions 2.3 and 3.1 and $\theta_* \in \mathcal{M}'$. Then there exists $\mu_0 > 0$ such that for any $\mu \in (0, \mu_0)$, there exist $\theta_* : \mathbb{R}_+ \mapsto \mathcal{M}'$

with $\|\dot{\theta}_*\|_\infty \leq \mu$, and a bounded, piecewise continuous u such that

$$\begin{aligned}
\limsup_{t\to\infty} |\epsilon_1| &\geq \|u\|_\infty M' C_u [2 - e^{\gamma\tau_u}] \\
\limsup_{t\to\infty} |e_1| &\geq \|u\|_\infty M' \{C_u(2 - e^{\gamma\tau_u}) - O[\gamma/(\alpha+\gamma)](e^{\gamma\tau_u} - e^{-\alpha\tau_u})\} \\
\limsup_{t\to\infty} |e_1| &\geq \|u\|_\infty O\{[M_r' - d_z\sqrt{T/\delta_w} - O(\Gamma')]/(1+\gamma)\}
\end{aligned}$$

where C_u, τ_u, α are as in Proposition 3.2, $M' = [\operatorname{diam}\mathcal{M}' - d_z\sqrt{T/\delta_w} - O(\Gamma')]$ and $M_r' = 2\max_{\theta_o\in\mathcal{M}'}[r]$ s.t. $\{\theta : |\theta - \theta_o| \leq r\} \subseteq \mathcal{M}'$.[2]

The proof follows along the lines of the previous results, with the addition of a regulating input during the parameter drift phase; this input ensures that when the system parameters drift, the regressor and estimation error maintain small magnitudes which, in turn, limits the maximum possible adjustment of the estimated parameters to an arbitrarily small value.

Thus, as in the case of linear-model parameter estimation, the presence of arbitrarily small disturbances or parameter time-variations combined with lack of sufficient excitation, can induce persistent estimation and identification error bursts whose magnitude is proportional to the size of the parametric uncertainty set. For the estimation error, this is immediately apparent from the respective lower bound, gi en in the propositions, by letting τ_u become sufficiently small. On the other hand, the first lower bound for the identification error is meaningful only when the "adaptation gain" γ is sufficiently small but becomes to conservative for large adaptation gains. (Note that $C_u \to 0$ as $\tau_u \to 0$.) In the latter case, the second lower bound offers a qualitatively similar conclusion at the expense of a reduction in the size of the parametric uncertainty.

3.4. Remark. The qualitative characteristics of this behavior are not limited to a specific model (or structure) of perturbations. Indeed, the same effect can be obtained by output disturbances or unmodeled dynamics. For the latter, in particular, the perturbation d takes the form $d = \Delta_1[u] + \Delta_2[y]$ where Δ_1, Δ_2 are stable operators. In order for this problem to be practically meaningful, the class of admissible perturbations should be restricted to those for which the perturbed system is "close" to the original one e. g., by specifying an upper bound for the induced gains of Δ_i. Under these conditions, the results of Proposition 3.2 remain valid when δ is such that Δ_1, Δ_2 have induced L_2 (or L_∞) gains less than δ and $\delta \in (0, \delta_0)$, for some $\delta_0 > 0$.

It is not surprising that the construction of a bursting scenario for this case involves high-frequency inputs u and perturbations that are "large" at high frequencies. For example, the effect of the previous burst-inducing disturbance can be emulated by choosing

$$\Delta_i = g_t q^\top (sI - F)^{-1}(\theta_{i0} - \theta_{i*})G_H(s)$$

[2] If, in addition, $\mathcal{A}$ satisfies $|\dot{\theta}(t)| \leq \gamma'\operatorname{dist}(\theta(t), \mathcal{B})$ for some constant γ' and for all $\theta \in \{\theta : 0 < \beta \leq \operatorname{dist}(\theta(t), \mathcal{B})\}$, then the terms $O(\Gamma')$ drop out of the performance lower bounds.

where $G_H(s)$ is a stable, high-pass transfer function with "cut-off" frequency v_o, while during the PE intervals the input is a sum of high frequency ($\gg v_o$) sinusoids.[3] On the other hand, some modification of the results is necessary when the admissible perturbations are further restricted to enter the system in a multiplicative or additive form. For such a case, it can be shown that the error lower bounds in Proposition 3.2 remain valid if the term $\max_{\theta \in \mathcal{M}} |\theta - \theta_*|$ is replaced by M_r.

An interesting by-product of our bursting scenario is that, without imposing any excitation conditions, the problem of ensuring "good" lim sup performance in the presence of arbitrarily slow plant parameter variations is as hard as the problem of ensuring good L_∞ performance (i. e., including adaptation transients) for LTI plants with arbitrary initial conditions in the parameter estimates but with restricted initial conditions on the plant/filter states.

4. EXAMPLE: LTI PLANT WITH DISTURBANCE

Although not formally treated in the present study, similar scenaria can be extended, at least in principle, to the adaptive control case where the parameter drift can cause a temporary destabilization of the closed-loop and, thus, induce even more severe bursting. In the following we illustrate the construction of such bursting scenaria by means of a simple example from model reference adaptive control (MRAC).

Consider the plant with input disturbance d

$$y_p = \frac{b}{s+a}[u_p + d]$$

with nominal parameters $a = 0, b = 1$ and suppose that the control input u_p is designed so that the nominal plant output tracks the output of the reference model

$$y_m = \frac{1}{s+1}[r]$$

for any bounded reference input r. To achieve this objective when the plant parameters are partially unknown we use the controller

$$u_p = [r, y_p]\theta$$

where θ is updated by a gradient-based adaptive law with projection. For simulation convenience, we update θ in discrete-time with sampling interval $T_s = 0.2$, according to the following equations:

$$\theta(k+1) = \Pi_{\mathcal{M}}\left[-\frac{2\epsilon_1(k)\zeta(k)}{m(k) + 2\zeta^\top(k)\zeta(k)}\right].$$

[3] g_t is a scalar, time-varying gain, switching the perturbation off during the bursting intervals in order do simplify the derivations.

The estimation error and regressor vector are taken as the sampled versions of their continuous time counterparts

$$\epsilon_1 = \theta^\top \zeta - \frac{1}{s+1}[u_p]; \qquad \zeta = [y_p, \frac{1}{s+1}[y_p]]^\top$$
$$m = 1 + m_2; \qquad \dot{m}_2 = -0.75 m_2 + u_p^2 + y_p^2.$$

The projection set $\mathcal{M}$ is selected to contain the nominal controller parameter vector ($\theta_* = [1, -1]^\top$); in our simulations we take $\mathcal{M} = [0.3, 3] \times [-4, 4]$.

Guided by the previously presented construction of bursting scenaria, we let

$$\begin{aligned} r &= R_1[\sin(4t) + \sin(t)] + R_2 \\ d &= \mathrm{sat}_{0.5}[-K y_p] \end{aligned}$$

where $\mathrm{sat}_{0.5}$ denotes a saturation nonlinearity with linear region $[-0.5, 0.5]$ (clearly, $||d||_\infty \leq 0.5$). It now follows that whenever R_1, R_2 are sufficiently small so that $K y_p$ is in the linear region of the saturation, and r is PE, the adaptation algorithm drives the parameter estimates towards the point $[1, K-1]$. Thus, if $K - 1 > 0$, the nominal unperturbed closed-loop is unstable, something that becomes evident in the form of a burst as soon as the disturbance is removed and/or the magnitude of the reference input is increased. This behavior is illustrated in Figures 1–6. It should be emphasized that the burst magnitude is essentially independent of the disturbance bound; if the latter is decreased, the same qualitative behavior will be obtained by decreasing R_1 and increasing the length of the drift phase.

5. BURST SUPPRESSION

In this section we briefly discuss the implication of the above results on the design of adaptive laws that provide practically meaningful lim sup performance guarantees in the absence of excitation. For this purpose, we observe that the violation of at least one of the assumptions can be interpreted as a necessary condition for burst suppression. In the following, we use our previous example to discuss the differences between two approaches to burst supression for LTI plants. One is the use of some form of a dead-zone that modifies the optimization objective so that error convergence to zero is not required (violating Assumption 2.1. The other was introduced in [35] and uses a "standard" adaptive scheme (i. e., conforming to the previous bounds) but employs set-membership concepts to estimate (and reduce) the parametric uncertainty set on-line. Notice that in this approach, Assumption 2.1,2 is also violated but in a more subtle way. That is, the parameter estimates do not converge to any point in the initial parametric uncertainty set $\mathcal{M}$, for all initial conditions and all inputs satisfying the PE condition. They do, however, converge to any point in $\mathcal{M}$ that is consistent with the measurements and the "noise" bound.

5.1. Adaptation with dead–zone

For our example, since the plant is LTI, one quick remedy for bursting is the use of a dead-zone in the adaptation. Following [20], such a dead-zone can be introduced in the adaptive law by simply replacing the estimation error ϵ_1 by

$$\epsilon_{dz} = \begin{cases} \epsilon_1 - d_0 & \text{if } \epsilon_1 > d_0 \\ 0 & \text{if } |\epsilon_1| \leq d_0 \\ \epsilon_1 + d_0 & \text{if } \epsilon_1 < -d_0 \end{cases}$$

where d_0 is the dead-zone threshold. For the selection of d_0 we need to know an upper bound on the effective perturbation entering ϵ_1. If such a bound is available, the modified adaptive law guarantees that $\limsup |\epsilon_1| \leq \limsup d_0$. Furthermore, for a threshold that is strictly greater than the perturbation bound (in the limsup), the estimated parameters converge, resulting in a tracking performance bound $\limsup |e_1| \leq (\min_{\mathcal{M}} \theta_1)^{-1} \limsup d_0$. Although in the general case the design of d_0 could be quite involved, in our simple example it turns out that it suffices to choose d_0 to be greater than the input disturbance. In our simulations, shown in Figures 1–6, we choose $d_0 = 0.6$.

5.2. Adaptation with updated projection set

A different approach to burst suppression is to decrease the size of the effective parametric uncertainty ($\operatorname{diam} \mathcal{M}$). This idea was explored in [35], where set-membership (SM) estimation principles were used to reduce the parametric uncertainty set online and establish dead-zone-like limsup performance guarantees.

Set-membership estimation deals with the problem of estimating the smallest possible set that contains the unknown parameters of an affine parametric model, using I/O measurements and an instantaneous bound on the noise. A suboptimal but computationally tractable solution of this problem can be obtained recursively as an ellipsoid that contains the constraint set (see, for example, [3, 4, 6, 9, 28]; see also the Appendix for a brief review of the recursion equations and some basic results). Although the primary interest in this approach is the estimation of the parametric uncertainty set (or constraint set), the center of the ellipsoid could be used as an estimate of the unknown parameters. Employing such an estimator in an adaptive control scheme offers asymptotic performance guarantees that are similar to those obtained with dead-zone algorithms (see Proposition A.1 in the Appendix). There are, however, some differences between the two that should be taken into account in the interpretation of theoretical results as well as practical designs:

- SM algorithms exploit measurement information more efficiently than dead-zones, taking into account both the size and directionality of the excitation. This often results in a significant performance improvement in terms of speed of convergence and asymptotic properties of the estimation and parameter errors.[4]
- Naturally, SM algorithms introduce an increased complexity in the estimator that could be significant when the number of adjustable parameters is large. This added complexity could pose a problem in the selection of sampling rates,

[4] This is analogous to the performance comparison between least-squares and gradient estimators.

since in the initial stages of estimation the SM updates are quite frequent. To a lesser degree, the memory requirements for SM estimation could become prohibitive for large problems.

- A more subtle difference is due to the more restrictive assumptions of SM algorithms. As defined, an SM estimator requires the knowledge of an instantaneous bound of the perturbation, while its dead-zone counterpart needs only a bound of the lim sup value. Among the consequences of this difference is that SM algorithms also require a (possibly time-varying) bound on the transient terms, including contributions of initial conditions, with the corresponding results being only semi-global. On the other hand, dead-zone algorithms need only "steady-state" information and produce global results. Furthermore, violation of the assumptions can induce more severe problems in an SM estimator in that the constraints become inconsistent and the algorithm requires re-initialization.

An interesting use of the SM algorithm, explored in [35], is as a set estimator that attempts to reduce the size of the uncertainty set on-line. In this approach, the adaptive controller employs two estimators: One is responsible for the updating on the controller parameters via a simple gradient scheme with projection. The other is an SM etimator, responsible for updating the parametric uncertainty set used for projecting the parameters of the first estimator. That is, the parameter updates are performed according to

$$\begin{aligned} \dot{\theta} &= \boldsymbol{\Pi}_{E_k \cap \mathcal{M}} \left(\frac{-\gamma \epsilon_1 \zeta}{m_2} \right); \quad \theta(0) \in \mathcal{M} &(8)\\ \theta(t^+) &= \boldsymbol{\Pi}^{o}_{E_k \cap \mathcal{M}}[\theta(t)] \text{ whenever } \operatorname{dist}(\theta, E_k \cap \mathcal{M}) > \epsilon_* &(9) \end{aligned}$$

where the constant ϵ_* denotes the thickness of the boundary layer used in the projection $\boldsymbol{\Pi}_{E_k \cap \mathcal{M}}$ and $\operatorname{dist}(\theta, E_k \cap \mathcal{M})$ is defined as

$$\operatorname{dist}^2(\theta, E_k \cap \mathcal{M}) = \operatorname{dist}^2_{R_o^{-1}}(\theta, \mathcal{M}) + \operatorname{dist}^2_{R_k^{-1}}(\theta, E_k).$$

The sets E_k are updated by the auxiliary set-membership estimator (16)-(21) and (24) (operating in hybrid mode), with the regressor and error signals defined as

$$\begin{aligned} z_k &= \zeta(t_k)/\sqrt{m_2(t_k)} \\ \hat{e}_k &= -\epsilon_1(t_k)/\sqrt{m_2(t_k)} + \theta^{\mathsf{T}}(t_k) z_k - c_{k-1}^{\mathsf{T}} z_k \\ \mu_k &= \mu(t_k) \end{aligned} \qquad (10)$$

($\lim_{k\to\infty} t_k = \infty$; typical initialization $E_0 \supseteq \mathcal{M}$.)

As shown in [35], for LTI plants, this combination of estimators maintains the desirable RMS performance of gradient adaptation (with fixed projection set) while, in addition, it offers dead-zone-like performance guarantees in the lim sup sense. More precisely, under the assumption that the effective perturbation entering the estimation error ϵ_1 has an a priori known bound $\mu(t)$, a MRAC with the adaptive law (8)-(9) has the following properties:

- Signal boundedness is guaranteed for sufficiently small unmodeled dynamics (the sufficient condition for robustness is the same as for a MRAC with fixed projection set [32]).

- The RMS performance guarantees are the same as for a MRAC with fixed projection set (i.e., the upper bounds for $RMS\left[\frac{\epsilon_1}{\sqrt{m_2}}\right]$ and $RMS\left[\frac{e_1}{\sqrt{m_2}}\right]$ are as in [32]).

- The estimation and tracking errors satisfy the following lim sup performance guarantees

$$\lim_{t\to\infty}\sup\left|\frac{\epsilon_1}{\sqrt{m_2}}\right| \le \left(\sqrt{2n(1+\epsilon_*)+1}\right)\lim_{t\to\infty}\sup(\mu)+O\left(\epsilon^{1/4},\sup_k(t_k-t_{k-1})\right)$$
$$\lim_{t\to\infty}\sup\left|\frac{e_1}{\sqrt{m_2}}\right| \le O\left(\sqrt{2n(1+\epsilon_*)+1}\right)\lim_{t\to\infty}\sup(\mu)+O\left(\epsilon^{1/4},\sup_k(t_k-t_{k-1})\right)$$

  where $2n$ is the number of adjustable parameters and ϵ is a small design parameter (see Proposition A.1).

These results should be carefully interpreted, since they only establish the equality of worst-case performance and robustness bounds of the fixed and updated projection set MRACs. The two schemes, however, will not necessarily exhibit the same RMS performance, nor they will preserve signal boundedness for exactly the same class perturbations. Further, for the implementation of the updated projection set algorithm, the a *priori* knowledge required is the combination of the knowledge needed for the fixed-set projection and dead-zone MRACs together with an upper bound on the transient contribution of the perturbation terms, including initial conditions, in the estimation error. Although the latter can be annoying, it should be emphasized that unlike dead-zone algorithms, the RMS performance guarantees of the updated projection set MRAC are not affected by the conservatism in the selection of the threshold $\mu(t)$.

Example. Returning to our example, we simulate the closed-loop response with an adaptive controller employing updated projection sets. As before, we use the discrete-time version of the estimator to adapt the controller parameters, i.e.,

$$\theta(k+1) = \boldsymbol{\Pi}_{E_k\cap\mathcal{M}}\left[-\frac{2\epsilon_1(k)\zeta(k)}{m(k)+2\zeta^\top(k)\zeta(k)}\right]$$

with sampling time 0.2 time units. The SM updates of the parametric uncertainty set E_k (i.e., R_k and c_k) are performed every 0.5 time units, with a threshold $\mu(t) = 0.6/\sqrt{m} + 5e^{-0.25t}$ which is consistent with dead-zone case. In addition, for comparison purposes, we also simulate the closed-loop response where the controller parameters are updated by an SM estimator alone (setting $\theta(k) = c_k$).

5.3. Simulation results

The results of our simulations, shown in Figures 1 − 7, illustrate the main points of our previous analysis and discussion. The first set of figures (1 − 3) shows the closed-loop responses with the four adaptive controllers starting with zero initial conditions while the reference input and the perturbation alternate between the following two phases:

Phase 1: *(Drift phase)* $K = 5$, $R_1 = 0.1$, $R_2 = 0$ (85 time units).

Phase 2: *(Burst phase)* $K = 0$, $R_1 = 0$, $R_2 = 1$ (15 time units).

Figure 1 shows the output responses with MRACs employing adaptation with fixed set projection, dead-zone, SM, and updated projection sets, as well as the "nominal" response of a non-adaptive MRC that was designed for the nominal plant. Figure 2 shows the corresponding tracking error responses; the error plots are clipped to emphasize steady-state details. Figure 3 contains the trajectories of the parameter estimates for each of the four adaptive laws.

In these plots we observe that the response of the MRAC with fixed projection set exhibits large bursts in the beginning of each Phase 2 interval. At these instances the parameter estimates have been driven near the point (1,4) and correspond to an unstable closed-loop system. Although stictly speaking this behavior is not periodic, th parameter trajectories in Figure 4 indicate that these bursts do not disappear with time.

As expected, the tracking error bursts are considerably reduced with dead-zone adaptation. This improvement, however, is obtained at the expense of a deterioration of the RMS performance. The latter is manifested by a significant increase of the error in the Phase 2 intervals where the reference input is large (the peak error is close to two, its estimated worst case value.) Similar conclusions can be drawn for the MRAC with the SM estimator, although in this case the deterioration of the RMS performance is considerably smaller. This can be attributed to the more efficient utilization of information by the SM estimator that results in a steady-state error of approximately 0.2 during the Phase 2 intervals.

On the other hand, the MRAC with updated projection sets exhibits good asymptotic performance in both the RMS and $\limsup$ sense. Its initial transient is similar to the MRAC with fixed projection set since, at that point, the available data offer very little information about the parametric uncertainty set. During the rest of the cycles, however, the reduction of the latter does not permit any significant parameter drift towards the destabilizing region and, consequently, limits the size of the error bursts. Notice that, in contrast to the dead-zone and SM adaptation, this algorithm ensures the "convergence" of the tracking error to zero during the Phase 2 intervals where the plant disturbance is absent.

Finally, it should be emphasized that the non-decaying memory of SM estimators can cause significant differences in the estimates of the parametric uncertainty set for inputs that have the same steady-state but different transients. This is illustrated in the second set of figures (4 − 6) where the reference input and the perturbation again alternate between the same two phases, but starting with Phase 2. In this case, even though the steady-state tracking response remains approximately the same, the updated projection set MRAC results in a tracking error with smaller but more persistent transient contributions. The absence of a "large" error burst offers less information about the parametric uncertainty set, something that is indicated by the much larger final ellipsoid (Figure 7).

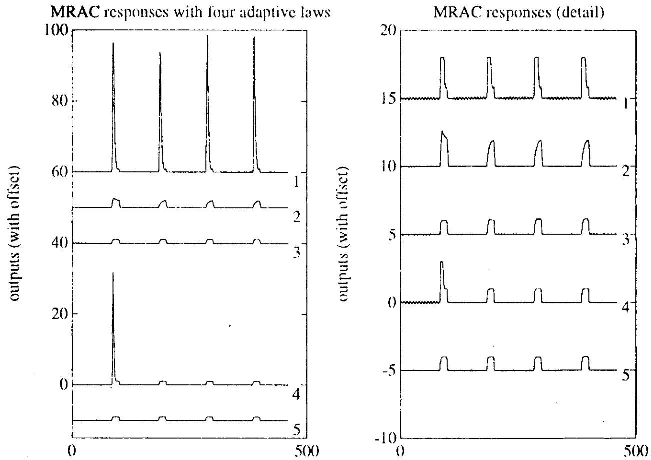

Fig. 1. Output response of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 1 followed by Phase 2.
Key: 1: fixed projection set, 2: dead-zone, 3: SM, 4: updated projection set, 5: nominal.

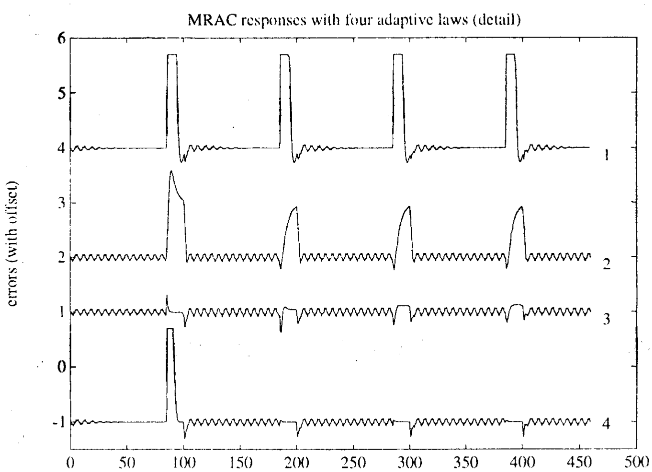

Fig. 2. Tracking error response of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 1 followed by Phase 2.
Key: 1: fixed projection set, 2: dead-zone, 3: SM, 4: updated projection set.

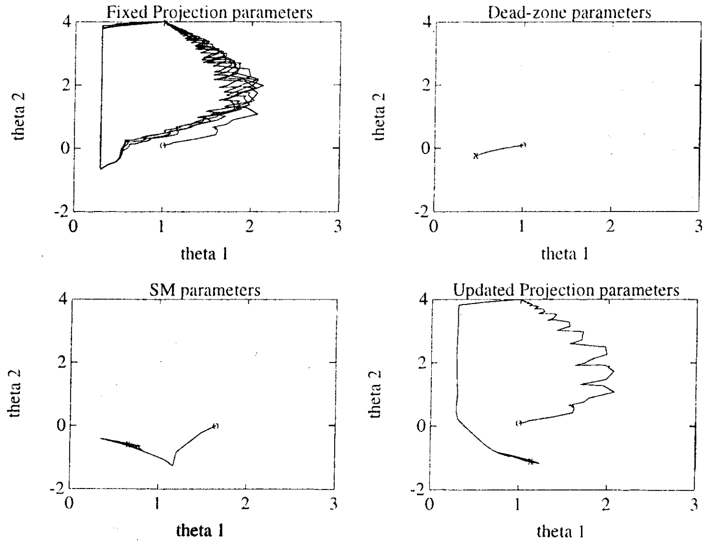

Fig. 3. Parameter trajectories of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 1 followed by Phase 2.

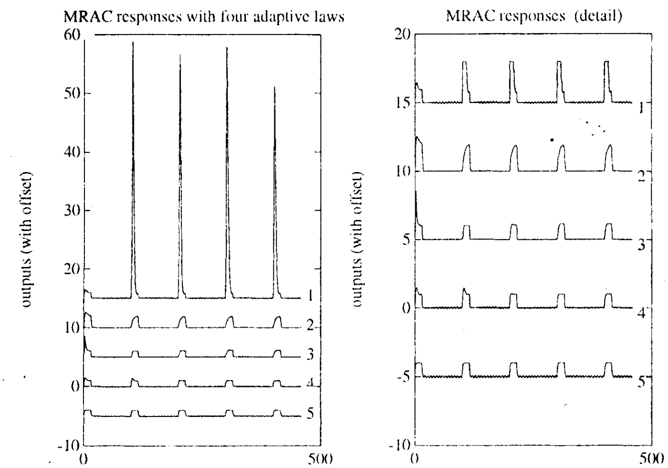

Fig. 4. Output response of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 2 followed by Phase 1.

Key: 1: fixed projection set, 2: dead-zone, 3: SM, 4: updated projection set, 5: nominal.

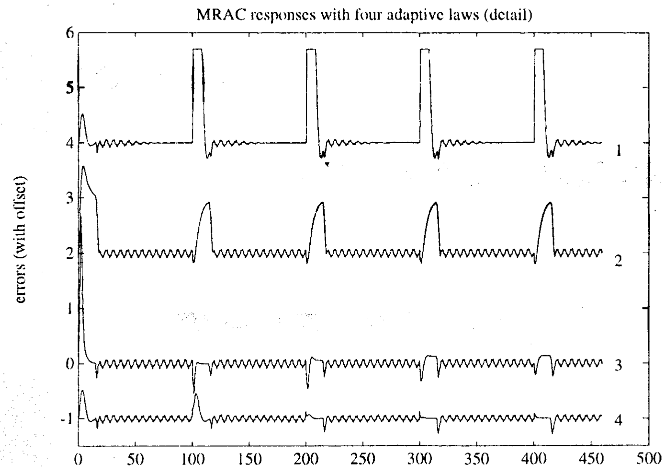

Fig. 5. Tracking error response of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 2 followed by Phase 1.
Key: 1: fixed projection set, 2: dead-zone, 3: SM, 4: updated projection set.

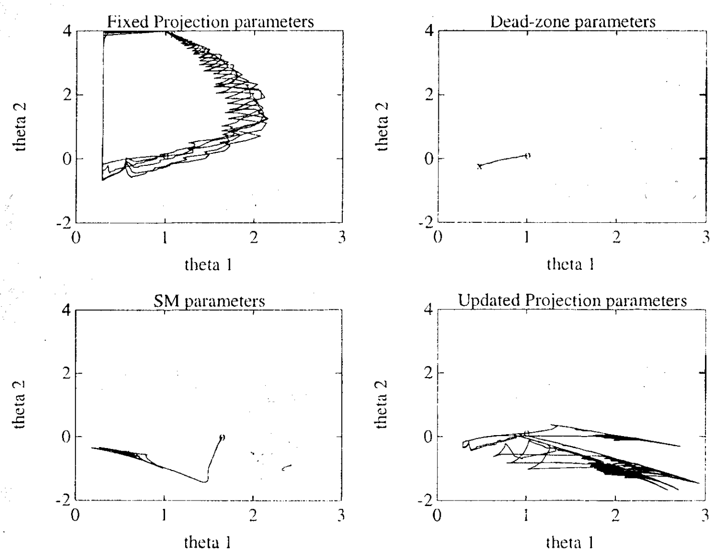

Fig. 6. Parameter trajectories of MRAC with the four adaptive laws. Reference input and disturbance selection: Phase 2 followed by Phase 1.

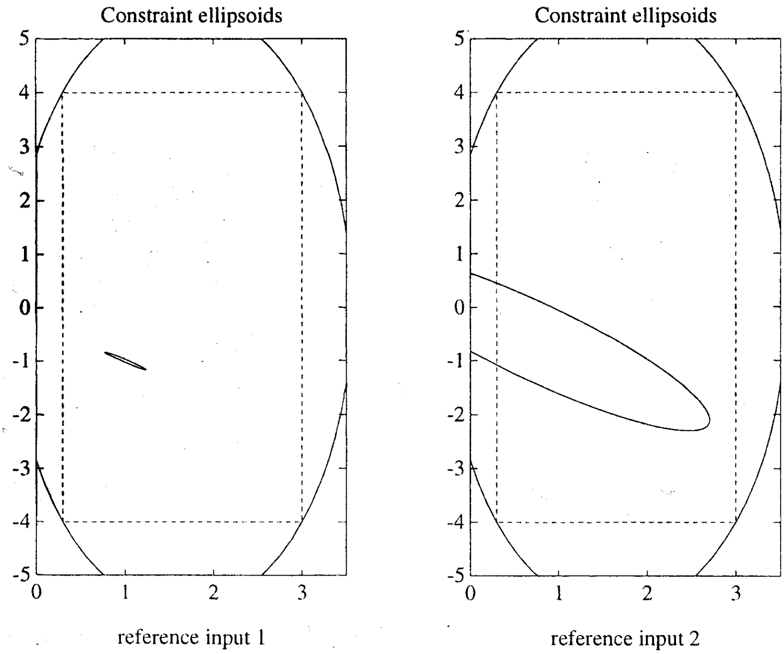

Fig. 7. Initial and final SM estimates of the parametric uncertainty for the MRAC with updated projection set.

6. CONCLUSIONS AND DISCUSSION

With the development of a general bursting scenario we analyzed some fundamental performance limitations of a wide class of adaptive algorithms, operating without sufficient excitation in "noisy" environments. The underlying principle behind these limitations is that, in the absence of sufficient excitation, perturbations may cause significant drifts (or effective drifts, in the time-varying case) of the parameter estimates. The manifestation of such a drift in the error signals is the appearance of a "burst" as soon as the excitation becomes large in a suitable direction. At that point, the size of the burst depends on the speed of the parameter updates relative to the time-scale of the regressor signals.

Further, using our bursting scenario in the adaptive control case, we illustrated some basic "burst" suppression mechanisms. Classical dead-zone adaptation prevents parameter adjustment when the signal-to-noise ratio may be poor. A more efficient, but also more restrictive, use of information can be achieved using set-membership estimation principles. This approach leads to a burst suppression mechanism whereby the parameter estimates are prevented from drifting to regions where bursts have occured. Combining such an estimator with a standard gradient algorithm with projection, we demonstrated that the size of the bursts can be suppressed without sacrificing the desirable RMS performance properties of gradient adaptation. The results, however, are applicable to the LTI case only, while the LTV generalization seems to be susceptible to bursting in manner analogous to dead-zone algorithms. Moreover, since SM algorithms rely heavily on the knowledge of an instantaneous bound of the noise, their practical application would require some "intelligent" on-line management of the estimated parametric uncertainty set.

Other promising approaches include the use of multiple estimators operating on a partition of the original parametric uncertainty set [20, 22]. Roughly, under this approach, the best estimator is selected at each time instant (or short interval) according to a cost objective. Effectively, the switching of estimators implements an adaptive law with infinite adaptation gain and, thus, violates our sufficient conditions for bursting. Still, the stability analysis for such an algorithm is available only for LTI plants.

Yet another idea is the use of "error compensation" [5, 31] which introduces a suitable high gain feedback mechanism in the loop that decreases the size of the error bursts. From our viewpoint, this approach effectively translates the parametric uncertainty into a region where any possible parameter drift has a small contribution in the tracking error. The success of this technique, however, has been limited to MRC problems.

Finally, we should emphasize that even if adaptation bursting turns out to be unavoidable in the general case, there is still a potentially viable adaptive control strategy in the injection of PE signals in the closed-loop. Under this approach, the injected signal should be of sufficiently high strength to provide a "good" signal-to-noise ratio and ensure parameter convergence to a small residual set [18,21,27]. On the other hand, such a signal should be small enough in order to have a minimal interference with the control objective. This basic trade-off between the parameter error residual set and the perturbation due to the injected signal suggests that the achievable lim sup performance should be of the order of the worst-case disturbance magnitude. Although conceptually simple, it seems that the main challenge in this approach is to establish quantitative and non-conservative criteria for the selection and implementation of the injected excitation.

APPENDIX

A SET-MEMBERSHIP ESTIMATION: BOUNDING ELLIPSOIDS

The basic development of the Set-Membership estimation algoritms begins with an affine parametric model of the form

$$y_k = \theta_*^{\mathsf{T}} z_k + n_k \tag{11}$$

where y_k, z_k and n_k are bounded sequences, $\theta_* \in \mathrm{IR}^M$ is the unknown constant parameter vector, y_k, z_k are available for measurement and n_k satisfies

$$|n_k| \leq \mu_k \tag{12}$$

where μ_k is an *a priori* known bounded sequence. [5] The last condition implies that each I/O measurement (z_k, y_k) defines a set containing θ_* as an intersection of two half-spaces, that is

$$\theta_* \in H_k \triangleq \{\theta \in \mathbb{R}^M : |y_k - \theta^\top z_k| \leq \mu_k\}. \tag{13}$$

Clearly, given k I/O pairs, the smallest set that is guaranteed to contain θ_* is the intersection $\cap_{i=1}^k H_i$. Alternatively, $\cap_{i=1}^k H_i$ can be viewed as the set of all parameters compatible with the model (11), the measurements and the constraint (12).

Although conceptually simple, this solution of the set-membership estimation problem is computationally intractable since, in general, the number of parameters required to define $\cap_{i=1}^k H_i$ grows linearly with k. To overcome this problem, we may relax our objective and find an ellipsoid, say E_k that "tightly encloses" the set $\cap_{i=1}^k H_i$. The advantage of this approach is that an ellipsoid is defined in terms of a fixed number of parameters (a center and a generalized radius) as shown below.

$$E_k \triangleq E(R_k, c_k) = \left\{\theta \in \mathbb{R}^M : |\theta - c_k|_{R_k^{-1}} \leq 1\right\} \tag{14}$$

where R_k is a positive definite matrix and $c_k, R_k^{1/2}$ are the center and generalized radius of the ellipsoid, respectively. Next, and in order to obtain a recursive solution, we may further modify our objective as follows: *Given an ellipsoid E_{k-1} containing θ_* and the I/O pair (z_k, y_k), find the "smallest" ellipsoid E_k such that $E_{k-1} \cap H_k \subseteq E_k$*. To quantify the meaning of *the smallest ellipsoid* we can employ various measures of the size of a set. One such measure, leading to relatively simple expressions, is the volume of the set which, for an ellipsoid E_k, is proportional to the determinant of R_k.

It is now a straightforward exercise in geometry to find the ellipsoid E_k of minimum volume, containing the intersection $E_{k-1} \cap H_k$. A simple way to achieve this is to view H_k as a degenerate ellipsoid and define E_k in terms of a nonnegative scalar weight as shown below.

$$E_k(q) = \{\theta \in \mathbb{R}^M : |\theta - c_{k-1}|^2_{R_{k-1}^{-1}} + q|y_k - \theta^\top z_k|^2 \leq 1 + q\mu_k^2\} \tag{15}$$

where $q \geq 0$ is selected so as to minimize the volume of $E_k(q)$. Combining (14) and (15), we obtain the following recursive formulae for the ellipsoid E_k:

$$\begin{aligned} \hat{e}_k &= y_k - c_{k-1}^\top z_k \\ a_k &= 1 + q(\mu_k^2 - \hat{e}_k^2) + q^2 \frac{\hat{e}_k^2 z_k^\top R_{k-1} z_k}{1 + q z_k^\top R_{k-1} z_k} \end{aligned} \tag{16}$$

$$R_k = a_k R_{k-1} - q a_k \frac{R_{k-1} z_k z_k^\top R_{k-1}}{1 + q z_k^\top R_{k-1} z_k} \tag{17}$$

[5] Only the discrete-time version of this estimation problem is considered here since, in order to simplify the analysis, our intention is to operate the set-membership estimator in a hybrid mode. Also notice that the assumption that y_k, z_k, n_k are bounded can easily be met in our MRAC framework via signal normalization.

$$\begin{aligned} \det R_k &= \frac{a_k^M}{1+qz_k^\top R_{k-1}z_k}\det R_{k-1} & (18)\\ R_k^{-1} &= [R_{k-1}^{-1}+qz_kz_k^\top]/a_k & (19)\\ c_k &= c_{k-1}+qR_{k-1}\frac{\hat{e}_kz_k}{1+qz_k^\top R_{k-1}z_k}. & (20) \end{aligned}$$

Further, minimizing $\det R_k(q)$ subject to the constraint $q \geq 0$ we obtain

$$\begin{aligned} q_{opt} &= \begin{cases} 0 & \text{if } \beta_2^2-4\beta_1\beta_3<0\\ \max\left(0, \frac{-\beta_2+\sqrt{\beta_2^2-4\beta_1\beta_3}}{2\beta_1}\right) & \text{otherwise} \end{cases} & (21)\\ \beta_1 &= (M-1)\,\mu_k^2(z_k^\top R_{k-1}z_k)^2\\ \beta_2 &= (2M-1)\,z_k^\top R_{k-1}z_k\mu_k^2-(z_k^\top R_{k-1}z_k)^2+\hat{e}_k^2z_k^\top R_{k-1}z_k\\ \beta_3 &= M(\mu_k^2-\hat{e}_k^2)-z_k^\top R_{k-1}z_k. \end{aligned}$$

In the above derivations, we have tacitly assumed that $M > 1$. The case where $M = 1$ (only one adjustable parameter) must be treated separately. In this case, however, the parametric uncertainty set reduces to an interval and $\cap_{i=1}^k H_i$ can easily be computed recursively. For this reason, in the rest of our discussion we only consider the more interesting case $M > 1$.

Equations (16) – (21) provide the basic recursion for a class of set-membership estimators. We should point out that the so-obtained estimates are more conservative than the optimal polytope solution of the set-membership problem. The sources of conservatism are in the bounding of the optimal polytope by an ellipsoid, the recursive computation of the ellipsoid ($\cap_{i=1}^k H_i$ may be contained in the interior of E_k) and the simplified computation of E_k (the above formulae do not take into account the case where E_k lies strictly in the interior of one of the half-planes defining H_k). Nevertheless, as shown in [9] under some excitation conditions on the regressor z_k and the disturbance n_k the set E_k converges to a single point, namely θ_*. Furthermore, even in the absence of any excitation conditions, the ellipsoid E_k maintains certain desired properties, as stated in the following proposition.

A.1. Proposition. [35] Consider the parametric model (1) and suppose that

$$\begin{aligned} \theta_* &\in E(R_0, c_0) & (22)\\ \mu_k &\geq |n_k|+\epsilon,\ \forall k & (23) \end{aligned}$$

where $c_0 \in \mathrm{IR}^M$, $R_0 = R_0^\top > 0$, $\mu_k > 0$, $\forall k$ are known and ϵ is a positive design constant. Further, consider the ellipsoid E_k defined by (16), (20) where at each sample k, the weight q is selected as:

$$q = \begin{cases} q_{opt} & \text{if } \det R_k(q_{opt}) \leq (1-\epsilon)\det R_{k-1}\\ 0 & \text{otherwise.} \end{cases} \qquad (24)$$

Then,

1. E_k converges in finite steps; that is, there exists $k_0 > 0$ such that $\forall\, k > k_0$, $q = 0$ and $E_k = E_{k_0}$.
2. $\forall\, \theta \in E_k$, $|y_k - \theta^\top z_k| \leq \left(\sqrt{M+1}\right) \mu_k + O\left(\epsilon^{1/4}\right)$.

Finally, it is worthwhile to notice that a necessary and sufficient condition for $q(k) = 0$ is

$$M(\mu_k^2 + \delta - \hat{e}_k^2) - z_k^\top R_{k-1} z_k \geq 0 \tag{25}$$

where $\delta = O(\sqrt{\epsilon})$. Since q_k converges to zero in finite steps, the above inequality (from which Property 2 of Proposition A.1 is derived) indicates that, especially for large regressors z_k, the noise bound is used by this algorithm in a more efficient way than a simple dead-zone.

(Received February 14, 1996.)

REFERENCES

[1] B. D. O. Anderson: Adaptive systems, lack of persistency of excitation and bursting phenomenon. Automatica *21* (1985), 247–248.

[2] K. J. Astrom and B. Wittenmark: Adaptive Control. Addison–Wesley, 1989.

[3] G. Belforte, B. Bona and V. Cerone: Parameter estimation algorithms for a set–membership description of uncertainty. Automatica *26* (1990), 5, 887–898.

[4] D. P. Bertsekas and I. B. Rhodes: Recursive state estimation for a set–membership description of uncertainty. IEEE Trans. Automat. Control *AC–16* (1971), 2, 117–128.

[5] A. Data and P. A. Ioannou: Performance Analysis and Improvement in Model Reference Adaptive Control. U. S. C. Report, EE–Systems Dept., 1992.

[6] J. R. Deller: Set–membership identification in digital signal processing. IEEE ASSP Magazine *6* (1990), 4–20.

[7] B. Egardt: Stability of Adaptive Controllers. Springer–Verlag, New York 1979.

[8] M. D. Espana: Intermittent phenomena in adaptive systems: A case study. Automatica *27* (1991), 717–720.

[9] E. Fogel and Y. F. Huang: On the value of information in system identification – Bounded noise case. Automatica *18* (1982), 2, 229–238.

[10] G. C. Goodwin and K. S. Sin: Adaptive Filtering Prediction and Control. Prentice Hall, NJ 1984.

[11] P. A. Ioannou and P. V. Kokotovic: Adaptive Systems with Reduced Models. Springer Verlag, New York 1983.

[12] P. A. Ioannou and J. Sun: Robust Adaptive Control. Prentice Hall, NJ 1996.

[13] J. M. Krause and P. P. Khargonekar: Parameter identification in the presence of non–parametric dynamic uncertainty. Automatica *26* (1990), 113–123.

[14] G. Kreisselmeier: Adaptive control of a class of slowly time–varying plants. Systems Control Lett. *8* (1986), 97–103.

[15] L. Ljung: System Identification. Theory for the User. Prentice Hall, New Jersey 1987.

[16] R. Lozano–Leal and R. Ortega: Reformulation of the parameter identification problem for systems with bounded disturbances. Automatica *23* (1987), 243–251.

[17] I. M. Y. Mareels and R. R. Bitmead: Nonlinear dynamics in adaptive control: Chaotic and periodic stabilization. Automatica *22* (1986), 641–655.

[18] I. M. Y. Mareels and M. Gevers: Persistency of Excitation Criteria for Linear, Multivariable, Time Varying Systems. Australian National University Report, 1987.

[19] R. H. Middleton and G. C. Goodwin: Adaptive control of time-varying linear systems. IEEE Trans. Automat. Control *33* (1988), 2, 150–155.
[20] R. Middleton, G. C. Goodwin, D. Hill and D. Mayne: Design issues in adaptive control. IEEE Trans. Automat. Control *AC-33* (1988), 50–58.
[21] K. S. Narendra and A. M. Annaswamy: Stable Adaptive Systems. Prentice Hall, NJ 1989.
[22] K. S. Narendra and J. Balakrishnan: Improving transient response of adaptive control systems using multiple models and switching. In: Proceedings 32nd IEEE CDC, San Antonio 1993, pp. 1067–1072. Automatica *23* (1987), 247–251.
[23] B. B. Peterson and K. S. Narendra: Bounded error adaptive control. IEEE Trans. Automat. Control *AC-27* (1982), 1161–1168.
[24] G. J. Rey, C. R. Johnson and R. R. Bitmead: Nonlinear interactions between signals and parameters in robust adaptive control. In: Proceedings American Control Conference, 1990, pp. 1064–1069.
[25] B. D. Riedle and P. V. Kokotovic: Disturbance instabilities in an adaptive system. IEEE Trans. Automat. Control *AC-29* (1984), 822–824.
[26] C. E. Rohrs, L. Valavani, M. Athans and G. Stein: Robustness of continuous-time adaptive control algorithms in the presence of unmodeled dynamics. IEEE Trans. Automat. Control *AC-30* (1985), 881–889.
[27] S. S. Sastry and M. Bodson: Adaptive Control: Stability, Convergence and Robustness. Prentice Hall, NJ 1989.
[28] F. C. Schweppe: Recursive state estimation: Unknown but bounded errors and system inputs. IEEE Trans. Automat. Control *AC-13* (1968), 1, 22–28.
[29] W. A. Sethares, C. R. Johnson, Jr. and C. E. Rohrs: Bursting in adaptive hybrids. IEEE Trans. Comm. *37* (1989), 8, 791–799.
[30] T. Soderstrom and P. Stoica: System Identification. Prentice Hall, NJ 1989.
[31] J. Sun: A modified model reference adaptive control scheme for improved performance. In: Proceedings American Control Conference, 1991, pp. 150–155.
[32] K. S. Tsakalis: Robustness of model reference adaptive controllers: An input-output approach. IEEE Trans. Automat. Control *37* (1992), 5, 556–565.
[33] K. S. Tsakalis: Performance limitations of adaptive parameter estimation and system identification algorithms in the absence of excitation. In: Proceedings American Control Conference, 1994, pp. 1260–1264.
[34] K. S. Tsakalis and P. A. Ioannou: Linear Time-Varying Systems: Control and Adaptation. Prentice Hall, NJ 1993.
[35] K. S. Tsakalis and S. Limanond: Asymptotic performance guarantees in adaptive control. Internat. J. Adapt. Control and Signal Proc. *8* (1994), 173–199.
[36] B. E. Ydstie: Bifurcations and complex dynamics in adaptive control systems. In: Proceedings 25th CDC, Athens 1986.

Professor Kostas S. Tsakalis, Arizona State University, Center for Systems Science and Engineering, Box 877606, Tempe, AZ 85287-7606. U. S. A.