Bursting Scenaria and Performance Limitations of Adaptive Algorithms in the Absence of Excitation *

Kostas S. Tsakalis

Arizona State University
Center for Systems Science and Engineering
Box 877606, Tempe, AZ 85287-7606
E-mail: tsakalis@enuxsa.eas.asu.edu

Abstract

A simple, yet general, bursting scenario is developed 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. Simple examples are constructed, illustrating the implications of these results in adaptive control.

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 [1, 2].

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 [3, 4, 5, 6]. 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, 12, 13]. 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 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, 4, 5, 14] and references therein). A similar result has also been established in the practically interesting case where the plant is slowly time-varying [8, 9, 10].

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 sufficient excitation 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 [15]. Bursting has been studied with a variety of analytical tools relying primarily on geometrical nonlinear systems theory and averaging techniques, e.g., see [16, 17, 18, 19, 20]. Partial remedies include

*This work was supported in part by ARPA under grant F49620-93-1-0062 and in part by NSF under RIA grant ECS-9111346.

the use of dead-zones with linear time-invariant (LTI) plants e.g., [7, 3, 21, 22, 14, 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 [24, 25]. 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, [19, 20] 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. This construction relies on general properties of the estimator and, as such, is applicable to a wide class of adaptive algorithms. Based on this simple bursting mechanism we derive a lower bound 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 present some simple examples illustrating the construction of bursting phenomena in adaptive control and briefly discuss the possibility of designing adaptive laws that offer 'reasonable' *lim sup* performance guarantees, albeit a rigorous analysis is left as a topic of future studies.

2 Bursting in Parameter Estimation

2.1 LTI Models

Let us consider the linear process model with output disturbance

$$y = w^\top \theta_* + d \quad (1)$$

where $y : \mathrm{R}_+ \mapsto \mathrm{R}$ is the output of the process, $w : \mathrm{R}_+ \mapsto \mathrm{R}^n$ is a vector of signals (regressor) available for measurement, $\theta_* \in \mathrm{R}^n$ 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

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

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 [1, 2].

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 [15, 19]. This phenomenon has a simple interpretation from an optimization point of view: the argument of the minimization of, say, ϵ_1^2 over θ for arbitrary w is not a continuous functional of d at $d = 0$. The above simple example can be generalized to yield a constructive proof of estimation error bursting for a class of estimation algorithms. For this purpose, let us consider the following class of parameter estimation algorithms $\mathcal{A}$.

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 $t \geq t_0$,

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

$$y(t) = w^\top(t) \theta_* \quad (4)$$

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 $\mathcal{M} \subset \mathbf{R}^n$ *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* $t \geq 0$.

Loosely speaking, this class of algorithms is characterized by the property that, whenever the I/O data are generated by an ideal linear model and satisfy an excitation condition, the resulting parameter error converges to zero, irrespective of the initial conditions. A basic bursting mechanism and the worst-case *lim sup* performance of an algorithm satisfying Assumption 1 are quantified below.

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 1.1-2. Then for any* $\delta > 0$ *and any* $\theta_0 \in \mathbf{R}^n$*, 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 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* *lim sup* $_{t\to\infty}|\epsilon_1| = \|w\|_\infty \max_{\theta\in\mathcal{M}} |\theta - \theta_*| + \delta$ $\nabla\nabla$

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. Its explanation is indeed quite simple. 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 drifts 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 $y : \mathbf{R}_+ \mapsto \mathbf{R}$ is the output of the process, $w : \mathbf{R}_+ \mapsto \mathbf{R}^n$ is the regressor vector and $\theta_* : \mathbf{R}_+ \mapsto \mathbf{R}^n$ 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:

3 Assumption:] *$\mathcal{A}$ is a parameter estimation algorithm satisfying Assumption 1 with the following modifications:*

(i.) In 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 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 [28] or an error functional over a finite moving window.

4 Proposition: *Consider the case where an algorithm $\mathcal{A}$ is used to estimate the TV parameter vector θ_* of the linear model (5).*

Suppose that $\mathcal{A}$ satisfies Assumption 3. Then for any $\mu > 0$, there exist bounded, piecewise continuous w and $\theta_* : \mathbb{R}_+ \mapsto \mathcal{M}$ with $\|\dot{\theta}_*\|_\infty \le \mu$, such that $\limsup_{t\to\infty} |\epsilon_1| \ge \|w\|_\infty \operatorname{diam}\mathcal{M} - d_z\sqrt{n}$ $\nabla\nabla$

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 $|w_1(t)| \le c_w$, $\forall t$. 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 $\epsilon_1(T_2) \ge c_w \operatorname{diam}\mathcal{M} - c_w(d_z\sqrt{T/\delta_w} + \epsilon)$. Clearly, the sequence can be repeated ad infinitum with θ_* oscillating between θ_1 and θ_2. Hence, $\limsup_{t\to\infty} |\epsilon_1| \ge c_w(\operatorname{diam}\mathcal{M} - d_z\sqrt{T/\delta_w})$. 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 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 $\limsup_{t\to\infty} |\epsilon_1| \ge \frac{1}{2}\|w\|_\infty \operatorname{diam}\mathcal{M}$. Nevertheless, Proposition 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]

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 [10]). Further, using the definitions

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

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 (in the graph topology) 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.

5 Assumption: *In Assumptions 1 and 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 1 and 3 satisfies $\Gamma \le \gamma|\theta(t) - \theta_*(t)|$, 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 [5, 4].)

[1] Note that Proposition 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.

With this observation, conservative but intuitively appealing statements on the performance limitations of a class of adaptive identification algorithms are given below.

3.1 LTI Systems

6 Proposition: *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 1 and 5. Then for any $\delta > 0$, there exist bounded, piecewise continuous u, d with $\|d\|_\infty \leq \delta$ and such that*

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

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. $\{\theta : |\theta - \theta_*| \leq r\} \subseteq \mathcal{M}$. $\nabla\nabla$

The proof of the proposition follows the same basic idea outlined in Proposition 2, 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.

Thus, as in the case of linear-model parameter estimation, the presence of arbitrarily small disturbances 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, given in the proposition, 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 too conservative for large adaptation gains. (Note that $C_u \to 0$ as $\tau_u \to 0$.) In the latter case, the second lower bound[2] offers a qualitatively similar conclusion at the expense of a reduction in the size of the parametric uncertainty.

It is worthwhile to point out that 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

$$d = \Delta_1[u] + \Delta_2[y] \qquad (8)$$

[2] This is derived by estimating the first and second derivatives of e_1 from the corresponding state-space representation.

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 6 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, to emulate the effect of the previous burst-inducing disturbance, we may choose

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

where g_t is a scalar, time-varying gain ($|g_t| \leq 1$) and $G_H(s)$ is a stable, high-pass transfer function with 'cut-off' frequency v_o. Observe that, by choosing v_o to be sufficiently large, the induced gains of the Δ_i's can be made arbitrarily small. Next, during each PE interval, select $g_t = 1$ and u as a sum of high-frequency ($\gg v_o$) sinusoids with enough spectral lines to ensure that w is PE. Moreover, we can further require that at the same frequencies $G_H(jw) = 1$. Thus, inside each PE interval the effective perturbation entering the linear model is the same as in the bounded disturbance case, except that now this is true for a specific input u and modulo an exponentially decaying term.

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 6 remain valid if the term $\max_{\theta\in\mathcal{M}}|\theta - \theta_*|$ is replaced by M_r.

3.2 LTV Systems

Here we consider the adaptive identification problem of an LTV plant of the form (6) where (q, F) is an observable pair and $\theta_* : \mathbf{R}_+ \mapsto \mathcal{M}$. 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

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

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 $\theta \in \mathcal{M}'$.

Under these conditions and as in the case of linear-model parameter estimation, the time-variation of the system parameters alone is sufficient to induce bursting behavior in both the estimation and the identification error.

7 Proposition: *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 $y = w^{\mathsf{T}}\theta_*$. Suppose that $\mathcal{A}$ satisfies Assumptions 3 and 5 and $\theta_* \in \mathcal{M}'$. Then there exists $\mu_0 > 0$ such that for any $\mu \in (0, \mu_0)$, there exist $\theta_* : \mathrm{R}_+ \mapsto \mathcal{M}'$ with $\|\dot{\theta}_*\|_\infty \leq \mu$, and a bounded, piecewise continuous u such that

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

where C_u, τ_u, α are as in Proposition 6, $M' = [\mathrm{diam}\mathcal{M}' - d_z\sqrt{T/\delta_w} - O(\Gamma')]$ and $M'_r = 2\max_{\theta_o \in \mathcal{M}'}[r]$ s.t. $\{\theta : |\theta - \theta_o| \leq r\} \subseteq \mathcal{M}'$.[3] $\nabla\nabla$

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.

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 Examples

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 two simple examples from model reference adaptive control.

8 Example: *(LTI Plant with disturbance)* 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

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

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

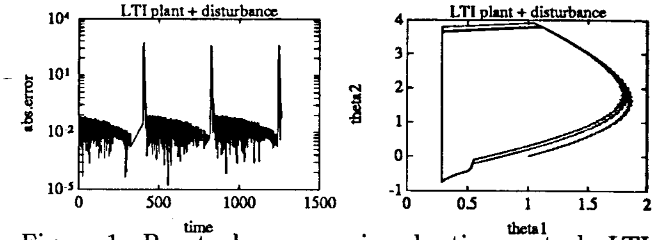


Figure 1: Burst phenomena in adaptive control: LTI plant with a bounded disturbance

for any bounded reference input r. To achieve this objective when the plant parameters are partially unknown we use the following adaptive law with projection:

$$
\begin{aligned}
u_p = [r, y_p]\theta \quad &; \quad \dot{\theta} = \mathcal{P}_{\mathcal{M}}(-20\epsilon_1\zeta/m) \\
\epsilon_1 &= \theta_1 y + \theta_2\zeta_2 - w
\end{aligned}
$$

where $\zeta = [y_p, \frac{1}{s+1}[y_p]]^{\mathsf{T}}$, $w = \frac{1}{s+1}[u_p]$, $m = 1 + \zeta_2^2 + w^2$. The projection set $\mathcal{M}$ is selected to contain the nominal controller parameter vector ($\theta_* = [1, -1]^{\mathsf{T}}$); in our simulations we take $\mathcal{M} = [0.3, 3] \times [-4, 4]$.

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

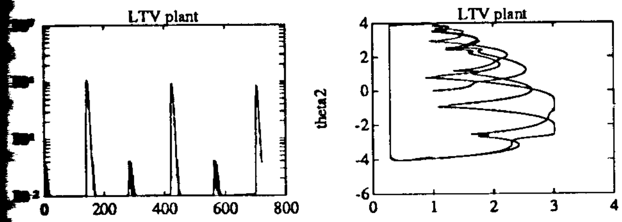
$$
\begin{aligned}
r &= R_1[\sin(4t) + \sin(t)] + R_2 \\
d &= \mathrm{sat}_{0.5}[-Ky_p]
\end{aligned}
$$

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 Ky_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 Fig. 1 where we alternate between the following to phases:

1. *Drift phase:* $K = 5, R_1 = 0.1, R_2 = 0$ (400 time units).
2. *Burst phase:* $K = 0, R_1 = 0, R_2 = 1$ (20 time units).

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. $\nabla\nabla$

9 Example: *(LTV Plant)* In the framework of Example 8, let us now consider the case where the input disturbance is absent but the plant parameters change with time. Again, guided by our construction



gure 2: Burst phenomena in adaptive control: LTV ant

bursting scenaria for LTV plants, we define the ref-ence input and plant parameter variation as:[4]

$$\begin{aligned} r &= R_1[\sin(4t)+\sin(t)] + R_2(-\theta_2 + a - 2)y_p/\theta_1 \\ \dot{a} &= K(-a+P) \end{aligned}$$

bursting behavior is thus obtained by repeating the llowing phases, in sequence:

1. *Drift phase:* $K = 0.25, R_1 = 0, R_2 = 1$ (100 time units).
2. *Burst/Excitation phase:* $K = 0, R_1 = 1, R_2 = 0$ (30 time units).
3. *Regulation phase:* $K = 0, R_1 = 0, R_2 = 1$ (10 time units).

The value of P is switched between 5 and -3 at the beginning of every Drift phase and remains constant during the other two phases. The simulation results for this example are shown in Fig. 2 where the first Drift phase is omitted as unnecessary ($a(0) = 5$). Notice that although error bursts appear in every Burst phase, the ones corresponding to a locally stable closed-loop are "small" (approx. one) while those corresponding to a locally unstable closed-loop are significantly larger (in the thousands).

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. (Note that the discussion is of a speculative, albeit intuitive, nature by necessity since few 'hard' results are currently available.) 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. Based on the available options, the possible avenues with the potential to provide an affirmative result are to introduce appropriate input constraints or use algorithms violating Assumption 5 (i.e., with 'infinite' adaptation gain). Yet another possibility is to use a standard adaptive scheme, conforming to the previous bounds, but employing some form of reduction of the effective parametric uncertainty set.

For example, the relation between the speed of variation of the regressor vector and the parameter adaptation plays a critical role in all of our constructions of bursting scenaria. If we restrict the speed of variation of the regressor vector to be small, relative to the speed of adaptation, the zero-error manifold changes slowly, giving the adaptive law sufficient time to drive the parameters near the manifold and, thus, maintain a small value of the estimation error. It is, in fact, straightforward to demonstrate the validity of this argument. Consider, for example, a standard gradient-based adaptive law with projection applied to any of the previous parameter estimation/system identification problems. That is,

$$\dot{\phi} = \mathcal{P}_{\mathcal{M}}(-\gamma\epsilon_1 w) \quad ; \quad \epsilon_1 = w^{\top}\phi + d \tag{9}$$

where $\mathcal{P}_{\mathcal{M}}$ is a (vector field) projection operator [4] $\gamma > 0$ is the adaptation gain and d denotes the effective perturbation in the estimation error (for adaptive laws using normalization, ϵ_1, w, d are the normalized versions of the respective signals). Further, let us suppose that $|\dot{\theta}_*|, |\dot{w}|, |d|$ are all less than μ and that $\theta_* \in \mathcal{M}$ is at most v_0-close to the boundary of $\mathcal{M}$, i.e., $dist(\theta_*, \partial\mathcal{M}) \geq v_0$. Under these conditions, and after some simple geometry, it follows that whenever $|\phi^{\top} w| \geq c_e > \mu$, $\phi^{\top} w w^{\top} \mathcal{P}_{\mathcal{M}}(-\gamma\epsilon_1 w) \leq -\lambda\gamma\phi^{\top} w\epsilon_1 w^{\top} w$ where $\lambda = v_0^2/(v_0^2 + M^2)$ and $M = \mathrm{diam}\mathcal{M}$. Next, letting $V = (\phi^{\top} w)^2/2$, it follows that whenever $V \geq c_e^2/2$, $\dot{V} \leq 0$ provided that

$$c_e \geq \max\left\{2\mu, 2\sqrt{\frac{\mu M(v_0^2+M^2)}{\gamma v_0^2}}, M\sqrt[3]{\frac{4\mu(v_0^2+M^2)}{\gamma v_0^2}}\right\}$$

implying that, as $t \to \infty$, ϵ_1 converges to a residual set where $|\epsilon_1| \leq c_e + \mu$.

This simple example demonstrates the principle that the *lim sup* performance guarantees of adaptive estimators can be improved by adjusting the ratio of the speed of adaptation versus the variation of the regressor vector. This can be achieved by restricting their environment of operation with respect to the maximum derivative (or frequency content) of the regressor vector, for example, via low-pass filtering the input u.[5] The obvious drawback of this approach is that it limits the excitation to a low-frequency range. Consequently, the identified system can only capture the low-frequency characteristics of the actual one, even if persistent excitation becomes available. Furthermore, the extension of this result to the adaptive control case is not straightforward (if at all possible) since the input is now a closed-loop signal.

[4] Here it is convenient to define the reference signal in a feedback form so as to ensure closed-loop boundedness; this definition does not invalidate the results.

[5] Notice that fast adaptation may also serve to reduce the size of the steady-state estimation error; its usefulness is, however, limited since it has an adverse effect on the tracking error and closed-loop robustness by increasing the size of the perturbation term $\tilde{\eta}$.

A different approach to burst suppression would be to decrease the size of the effective parametric uncertainty (diam$\mathcal{M}$). This, with some directionality considerations, is the basic idea explored by [27]. In that study, set-membership estimation principles were used to reduce the parametric uncertainty set on-line and establish *lim sup* performance guarantees. 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. In the same vein, another promising idea is to employ multiple estimators operating on a partition of the original parametric uncertainty set e.g. along the lines of [22, 30]. 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, as such, is able to compensate for fast variations of the zero error manifold. This strategy violates the previously derived sufficient conditions for bursting and may potentially lead to practical *lim sup* performance guarantees. On the other hand, the analytical verification of this idea must overcome the problems caused by the swapping of the fast-varying parameters in the identification error. (Results are available for LTI plants only, where these terms vanish asymptotically.)

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 [5, 4, 29]. 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, the quantitative aspects of this approach require some further work. For example, the available estimates of the excitation level produced by an external signal depend on the unknown plant in a rather complicated manner. Furthermore, the design of the injected signal should take into account the existence of reference signals so that the excitation properties of the former are not destroyed by the latter (e.g., via some frequency separation conditions).

References

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

[2] T. Soderstrom and P. Stoica, *System Identification.* Prentice Hall, NJ, 1989.

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

[4] S.S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness.* Prentice Hall, NJ, 1989.

[5] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems.* Prentice-Hall, NJ, 1989.

[6] K.J. Astrom and B. Wittenmark, *Adaptive Control.* Addison-Wesley, 1989.

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

[8] G. Kreisselmeier, "Adaptive Control of a Class of Slowly Time-Varying Plants," *Syst. and Cont. Letters*, Dec. 1986.

[9] R.H. Middleton and G.C. Goodwin, "Adaptive Control of Time-Varying Linear Systems," *IEEE Trans. on Autom. Contr.* Vol. 33, No. 2, pp. 150–155, Feb. 1988.

[10] Tsakalis, K.S. and P.A. Ioannou, *Linear Time-Varying Systems: Control and Adaptation.* Prentice Hall, 1993.

[11] C. E. Rohrs, L. Valavani, M. Athans and G. Stein, "Analytical Verification of Undesirable Properties of Direct Model Reference Adaptive Control Algorithms," *Proc. 20th IEEE CDC*, San Diego, Dec. 1981.

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

[13] B. D. Riedle and P. V. Kokotovic, "Disturbance Instabilities in an Adaptive System," *IEEE AC-29*, 1984.

[14] P. A. Ioannou and J. Sun, "Theory and Design of Robust Direct and Indirect Adaptive Control Schemes," *Int. J. Contr.*, V. 47, 775, 1988.

[15] B.D.O. Anderson, "Adaptive Systems, lack of persistency of excitation and bursting phenomenon," *Automatica*, May 1985.

[16] I.M.Y Mareels and R.R. Bitmead, "Nonlinear dynamics in adaptive control: Chaotic and periodic stabilization," *Automatica*, Nov. 1986.

[17] G. J. Rey, C. R. Johnson and R.R. Bitmead, "Nonlinear interactions between signals and parameters in robust adaptive control," *Proceedings 1990 ACC*, May 1990.

[18] B.E. Ydstie, "Bifurcations and complex dynamics in adaptive control systems," *Proc. 25th CDC*, Dec. 1986.

[19] Sethares, W.A., C.R. Johnson, Jr. and C.E. Rohrs, "Bursting in Adaptive Hybrids," *IEEE Trans. Commun.*, 37, 8, 791–799, Aug. 1989.

[20] Espana, M.D., "Intermittent Phenomena in Adaptive Systems: A Case Study," *Automatica*, 27, 1991.

[21] B. B. Peterson and K. S. Narendra, "Bounded Error Adaptive Control," *IEEE AC-27*, Dec. 1982.

[22] R. Middleton, G. C. Goodwin, D. Hill and D. Mayne, "Design Issues in Adaptive Control," *IEEE AC-33*, 1988.

[23] Lozano-Leal, R. and R. Ortega, "Reformulation of the parameter identification problem for systems with bounded disturbances," *Automatica*, Vol. 23, 247–251, 1987.

[24] Sun, J., "A modified model reference adaptive control scheme for improved performance," *Proc. 1991 ACC.*

[25] Data, A. and P.A. Ioannou, "Performance analysis and improvement in model reference adaptive control," *EE-Systems Dept., U.S.C. Report*, 1992.

[26] Fogel, E. and Y.F. Huang, "On the value of information in system identification — Bounded noise case," *Automatica*, Vol. 18, No. 2, 229–238, 1982.

[27] Tsakalis, K.S. and S. Limanond, "Asymptotic Performance Guarantees in Adaptive Control, *Int. J. of Adapt. Contr. and Signal Proc.*, Vol. 8, 173–199, 1994.

[28] J. M. Krause and P. P. Khargonekar, "Parameter identification in the presence of non-parametric dynamic uncertainty," *Automatica*, 1990.

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

[30] K.S. Narendra and J. Balakrishnan, "Improving Transient Response of Adaptive Control Systems Using Multiple Models and Switching," *Proc. 32nd IEEE CDC*, 1067–1072, San Antonio, TX, Dec. 1993.