Proceedings of the American Control Conference
Albuquerque, New Mexico June 1997
0-7803-3832-4/97/$10.00 © 1997 AACC

$\mathcal{L}_p$-Robustness and Exponential Convergence of Some Estimation Schemes in Adaptive Control*

Vítor H. Nascimento and Ali H. Sayed

Department of Electrical Engineering
University of California
Los Angeles, CA 90024

Abstract

In this paper, we suggest a design procedure for $\mathcal{L}_p$-stable estimation schemes (in a sense to be fully defined in the paper), and use the insights gained from this design to highlight new convergence and robustness properties of well-known estimation schemes, such as leakage-based (switching sigma) and parameter projection algorithms. In particular, we show that new robustness and convergence statements can be obtained if we further restrict the noise to lie in $\mathcal{L}_p \cap \mathcal{L}_\infty$, $1 < p < \infty$. We also indicate connections with results in robust statistics.

1. Introduction

Many estimation problems can be written as

$$y(t) = \phi(t)^T\theta + v(t), \qquad (1)$$

where θ is the parameter vector to be estimated, column regression vector, and T denotes transposition. In the absence of the noise vector $v(t)$, several algorithms exist that give a converging estimate $\hat{\theta}(t)$ when $\phi(t)$ satisfies a persistence of excitation (PE) condition. One of these algorithms is the gradient algorithm [5, 1]

$$\dot{\hat{\theta}}(t) = \Gamma\phi(t)\epsilon(t), \quad \hat{\theta}(0) = \hat{\theta}_0 = \text{ initial guess}; \qquad (2)$$

where $\epsilon(t) \triangleq y(t) - \phi^T(t)\hat{\theta}(t) = \phi^T(t)\tilde{\theta}(t) + v(t)$, is the measured estimation error, $\tilde{\theta}(t) \triangleq \theta - \hat{\theta}(t)$ is the parameter estimation error, and Γ is a positive-definite matrix. We can also write (2) in terms of the parameter estimation error only, as follows

$$\begin{cases} \dot{\tilde{\theta}}(t) = -\Gamma\phi(t)\phi^T(t)\tilde{\theta}(t) - \Gamma\phi(t)v(t)\ , \\ e(t) = \phi^T(t)\tilde{\theta}(t)\ , \end{cases} \qquad (3)$$

where we have defined the estimation error e. Note that e is the error we are interested in taking to zero as $t \to \infty$, as opposed to the *measured* estimation error ϵ that is corrupted by noise.

With PE regressors $\phi(t)$ and zero noise $v(t)$, the origin $\tilde{\theta} = 0$ in (3) can be shown to be exponentially stable and, therefore, the mapping from v to $\tilde{\theta}$ is $\mathcal{L}_p$ - stable for $1 \leq p \leq \infty$ [11, 5, 6]. However, it has been long known that in the presence of bounded noise, (2) may give unbounded estimates $\hat{\theta}$ when the regressor ϕ is not PE [4], [5, pp.545-552], [1, pp. 255-258]. Several modifications have been proposed to (2) in order to maintain the boundedness of $\hat{\theta}$ in the presence of bounded noise and plant uncertainties and without the PE property (leakage [5], parameter projection [7], dead-zone [5]). These modifications assume some knowledge of the magnitudes of the noise (dead-zone) or of the parameter vector itself (parameter projection and switching-σ, a form of leakage). schemes establish that the estimates $\hat{\theta}$ will converge only to a *region* containing the true parameter θ.

In this paper we reconsider this problem and show that the above estimation schemes, as well as some new ones we propose here, can still converge to the *true* solution θ if the noise is further restricted to lie in $\mathcal{L}_p \cap \mathcal{L}_\infty$ for $1 < p < \infty$; and give bounds for the $\mathcal{L}_p$ gain from $v(\cdot)$ to $e(\cdot)$. We extended the results of [12] to a general $p > 1$, class of algorithms. For two of the algorithms in this class we give bounds for the $\mathcal{L}_p$-gain from the disturbances $v(t)$ to the estimation error $e(t)$.

The contributions of this work are a design procedure for $\mathcal{L}_p$-stable estimation schemes, bounds for the $\mathcal{L}_p$ gain for two newly proposed algorithms [13]; as well as new convergence and robustness properties of well-known estimation schemes. We also indicate connections with results in robust statistics [3, 10].

1.1. Notation and Definitions

Throughout this paper we denote the usual 2-norm of a vector by $\|x\|_2$, and by $\|x\|_{2,p}$ the norm $\|x\|_{2,p} = \left(\int_0^\infty (\|x(t)\|_2)^p\, dt\right)^{1/p}$ for finite p, and $\|x\|_{2,\infty} = \sup_{t\geq 0} \|x(t)\|_2$ (see [14]). We also employ the following three definitions.

*This material was based on work supported in part by the National Science Foundation under Award No. MIP-9409319. The work of V. Nascimento was supported by a fellowship from CNPq, while on leave from the LPS-DEE - Escola Politécnica da USP - Brazil.

Definition 1 ($\mathcal{L}_p$-Stability) *A (causal) system*

$$\dot{x} = f(t,x) + g(t)u$$

is said to be $\mathcal{L}_p$-stable if there exist nonnegative scalars k and k_1 such that $\|x\|_{2,p} \le k_1 + k\|u\|_{2,p}$ for any input $u \in \mathcal{L}_p$.

Definition 2 (Persistence of Excitation) *A vector-valued function $\phi(t)$ is said to be persistently exciting (PE) if there exist positive scalars γ_0, γ_1, and T such that for all t, $\gamma_0 I \le \int_t^{t+T} \phi(\tau)\phi^T(\tau)d\tau \le \gamma_1 I$.*

2. $\mathcal{L}_2$-Stability of the Gradient Algorithm

We start with the gradient algorithm (2). It is known that when $\phi(t)$ is PE and $v(t) = 0$, the origin $\tilde{\theta} = 0$ in (2) is exponentially stable, and therefore the update law (2) leads to an $\mathcal{L}_p$-stable system for all $p \in [1,\infty]$ [5, p. 236], [6, p. 269]). This result is valid for PE regressors only. However, it has been recently noted in [12] that when the noise is nonzero but has finite energy ($v \in \mathcal{L}_2$), then *even* if ϕ is *not* PE, algorithm (2) still guarantees that the error signal $e = \tilde{\theta}^T\phi$ is in $\mathcal{L}_2$ and that $\tilde{\theta}$ remains bounded. Moreover, the following contraction relation always holds for any $\tau > 0$:

$$\frac{\tilde{\theta}^T(\tau)\Gamma^{-1}\tilde{\theta}(\tau) + \int_0^\tau e^2(t)dt}{\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(0) + \int_0^\tau v^2(t)dt} \le 1. \tag{4}$$

Expression (4) shows that the mapping from the disturbances $\{\Gamma^{-1/2}\tilde{\theta}(0), v(.)\}$ to the estimation errors $\{\Gamma^{-1/2}\tilde{\theta}(\tau), e(.)\}$ is a contraction, and therefore, in the language of H_∞-filtering, the gradient algorithm (2) is a robust filter [2, 12].

For noise in other $\mathcal{L}_p$ spaces (i.e., for $p \neq 2$), the denominator in (4) can be unbounded, and algorithm (2) can lead to unbounded $\hat{\theta}$ (even with $v(t) \to 0$ as $t \to \infty$), examples are given in [4].

3. Design of $\mathcal{L}_p$-Robust Estimators

In this section we show how to modify the gradient algorithm (2) in order to guarantee a bound similar to (4) for $v \in \mathcal{L}_p$, $p > 1$. More specifically, we derive a class of adaptive laws that guarantee bounds of the following general form

$$\frac{\tilde{\theta}^T(\tau)\Gamma^{-1}\tilde{\theta}(\tau) + \alpha\int_0^\tau |e|^p \mathrm{dt}}{\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(0) + \beta\int_0^\tau |v|^p \mathrm{dt}} \le 1 \tag{5}$$

for some positive numbers α and β. Note that it is the noise-free error e (and not ϵ) that we employ in (5) and (4).

The resulting family of adaptive laws that satisfy (5) will turn out to include, as special cases, several algorithms that have been discussed in the literature. in [13], and some algorithms studied in [10] in the context of robust statistics. These algorithms correspond to nonlinear estimation schemes of the form

$$\dot{\hat{\theta}} = \Gamma\phi(t)f(\epsilon(t)) \ , \tag{6}$$

where Γ is as before and $f(\cdot)$ is a Lipschitz-continuous function that we choose. Next we consider which Using the Lyapunov function $V(\tilde{\theta}) = \frac{1}{2}\tilde{\theta}^T(t)\Gamma^{-1}\tilde{\theta}(t)$, we obtain $\dot{V} = -f(\epsilon)(\epsilon - v) = -\epsilon f(\epsilon) + f(\epsilon)v$. Invoking Young's inequality [8] on the product $|f(\epsilon)v|$ and integrating the result, we obtain, for any $K > 0$,

$$V(\tau)+\int_0^\tau \left(\epsilon f(\epsilon) - \frac{K^q}{q}|f(\epsilon)|^q\right) dt \le V(0)+\frac{1}{pK^p}\int_0^\tau |v|^p dt. \tag{7}$$

Now assume we choose f and K such that the term in parentheses is nonnegative for all ϵ, that is,

$$|f(\epsilon)| \ \le \ \frac{q^{p-1}}{K^p}|\epsilon|^{p-1}. \tag{8}$$

In this case we see that when v is in $\mathcal{L}_p$, the error vector $\tilde{\theta}$ will be uniformly bounded. Note that the choice $f(\epsilon) = \epsilon$, which corresponds to the stochastic gradient algorithm (2), does not satisfy the condition above for any $p \neq 2$. Before continuing our discussion, we state the following preliminary result.

Lemma 1 (Choice of f) *Given a particular $p > 1$, if f is chosen such that (8) holds, then the nonlinear estimation scheme (6) guarantees that (7) holds, with the expression inside the parenthesis in the LHS always nonnegative. Moreover, if $v \in \mathcal{L}_p$, $\tilde{\theta}$ will be in $\mathcal{L}_\infty$.*

We now consider some particular choices for f.

3.1. A Polynomial Choice

A straightforward choice for $f(\cdot)$ that fits our framework has been recently considered in [13], and is given by

$$f_1(\epsilon) = \mathrm{sign}(\epsilon)|\epsilon|^{p-1}. \tag{9}$$

This algorithm was developed in [13] in the context of model reference adaptive control design, with the intent of guaranteeing $\epsilon \in \mathcal{L}_\alpha$, for some finite α, in the presence of $\mathcal{L}_p$ disturbance v. Recall that ϵ is in fact an error signal that is corrupted by noise. Using (9) and choosing $K = 1$ we can rewrite (7) as

$$V(\tau) + \frac{1}{p}\int_0^\tau |\epsilon|^p dt \le V(0) + \frac{1}{p}\int_0^\tau |v|^p dt \ .$$

Since $v, \epsilon \in \mathcal{L}_p$, e is also in $\mathcal{L}_p$. If we further assume that ϕ, $\dot{\phi}$, $v \in \mathcal{L}_\infty$, then (2) implies $\dot{\tilde{\theta}} \in \mathcal{L}_\infty$, and thus

$\dot{e} \in \mathcal{L}_\infty$; therefore, by Barbălat's Lemma [6], we get $e \to 0$ as $t \to \infty$.

To better understand the relation between the estimation error e and the it is useful to establish the existence of a contraction relation between e and v as in (4), rather than ϵ and v. The following statement summarizes the desired result, along with the earlier conclusions.

Theorem 1 ($\mathcal{L}_p$-Robustness of (9)) *Consider the update law (6) with the choice (9) and some finite $p \geq 2$. The following facts hold:*

(i) If $v \in \mathcal{L}_p$, then $\tilde{\theta} \in \mathcal{L}_\infty$ and ϵ, $e \in \mathcal{L}_p$.

(ii) If in addition to $v \in \mathcal{L}_p$, we also have v, ϕ, $\dot{\phi} \in \mathcal{L}_\infty$, then e, $\dot{\tilde{\theta}} \in \mathcal{L}_p \cap \mathcal{L}_\infty$ and $e \to 0$ as $t \to \infty$.

(iii) The following contraction relation holds for any $\tau > 0$:

$$\frac{\tilde{\theta}^T(\tau)\Gamma^{-1}\tilde{\theta}(\tau) + \alpha \int_0^\tau |e|^p\,dt}{\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(0) + \beta \int_0^\tau |v|^p\,dt} \leq 1\,, \tag{10}$$

where $\alpha = \frac{1}{2^{p-2}}\left(1 - \frac{1}{2^p}\right)$, and $\beta = \frac{2(p-1)}{p}$.

Proof: The proof of parts (i) and (ii) follows directly from the discussion above. The proof of part (iii) is more involved. Due to the limited space, we only sketch a few steps . Using (9), we obtain

$$\begin{aligned} -\dot{V} &= \text{sign}(\epsilon)|\epsilon|^{p-1}(\epsilon - v) = \text{sign}(e+v)|e+v|^{p-1}e \\ &= |e+v|^{p-2}(e+v)e = (e^2 + ev)|e+v|^{p-2}. \end{aligned}$$

Writing $v = \gamma e$ and minimizing the above expression over γ, we obtain a bound for $\dot{V}$. Integrating both sides of this bound, the desired contraction (10) is obtained.

■

A similar result holds for $p \in (1,2)$, with different expressions for α and β.

3.2. A Modified Gradient Algorithm

For $p \geq 2$, the choice (9) for $f(\cdot)$ has the drawback of having large gains away from $\epsilon = 0$, which can produce poor transients. Note, however, that the linear choice $f(\epsilon) = \epsilon$, satisfies (8) for large ϵ. Therefore, a new choice for $f(\cdot)$ that does not suffer from the high gain problem is

$$f_2(\epsilon) = \begin{cases} \text{sign}(\epsilon)|\epsilon|^{p-1} & \text{if } |\epsilon| \leq 1 \\ \epsilon & \text{otherwise} \end{cases} \tag{11}$$

Note further that this adaptive law reduces to a form of discontinuous dead-zone if we let $p \to \infty$, and to the usual gradient algorithm (2) as $p \to 2$ (this last property is shared by f_1).

Theorem 2 ($\mathcal{L}_p$–Robustness of (11)) *Consider the update law (6) with the choice (11) (with $p \geq 2$), and introduce the function*

$$g(\epsilon) = \begin{cases} \frac{1}{2}|\epsilon|^p & \text{if } |\epsilon| \leq 1 \\ \frac{1}{2}|\epsilon|^2 & \text{otherwise} \end{cases}$$

The following facts hold:

(i) If $v \in \mathcal{L}_p$, with $p \geq 2$, then $\tilde{\theta} \in \mathcal{L}_\infty$ and $g(\epsilon(\cdot)) \in \mathcal{L}_1$.

(ii) If in addition to (i) we also have v, ϕ, $\dot{\phi} \in \mathcal{L}_\infty$, then $\dot{\tilde{\theta}}$, ϵ, $e \in \mathcal{L}_\infty \cap \mathcal{L}_p$ and $e \to 0$ as $t \to \infty$.

(iii) Under conditions (i) and (ii), then

$$\frac{\tilde{\theta}^T(\tau)\Gamma^{-1}\tilde{\theta}(\tau) + \gamma \int_0^\tau |e|^p\,dt}{\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(0) + \eta \int_0^\tau |v|^p\,dt} \leq 1 \tag{12}$$

where $\eta = \frac{2(p-1)}{p}$, and

$$\gamma = \min\left\{\frac{1}{2^{p-2}}\left(1 - \frac{1}{2^p}\right),\ \frac{1}{p},\ \frac{1}{2\nu_e^{p-2}}\right\}\,,$$

$$\nu_e^2 = \|\phi\|_{2,\infty}^2 \|\Gamma\|_2 \left(\|\Gamma^{-1}\|_2 \|\tilde{\theta}(0)\|_2^2 + \frac{1}{p}\|v\|_{2,p}^p\right).$$

Proof: Similar to that of the previous theorem.

The figure below shows this scheme applied to the estimation of the poles of a second order linear plant, with PE regressors and noise in $\mathcal{L}_9$.

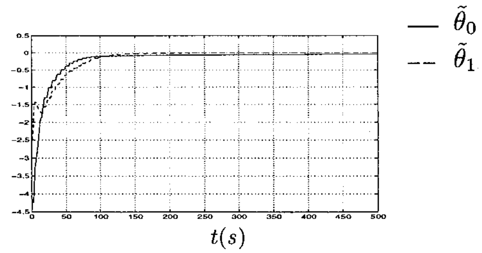

Figure 1: *parameter estimation errors for law (11)*

3.3. Connections with Robust Statistics

The notion of robustness has also been discussed in statistics (e.g., [3, 10]), where the primary concern is the design of optimal estimators that are robust to uncertainties in the noise probability distribution. Here we only show that one such estimator (proposed in [3, 10]) fits our framework; in a future work we will show from a deterministic point of view why this modification improves robustness. Although [10] deals with discrete-time and RLS-type algorithms, we consider here a continuous-time LMS version.

Suppose we have an estimation problem of the form (1), where now the noise is a white noise process, but where the probability distribution $p(v)$ of $v(t)$ is not exactly normal. As an example, take where $\delta \in [0,1]$,

$p_0 \in N(0,\sigma^2)$, and p_1 is an arbitrary density function. References [10, 9] show that an adaptive version of the "best" estimator for the above class of distributions is of the form $\hat{\theta}(t) = \Gamma\phi(t) f_3(\epsilon(t))$, where

$$f_3(\epsilon) = \begin{cases} \epsilon & \text{if } |\epsilon| \leq \Delta \\ \Delta \text{sign}(\epsilon) & \text{otherwise,} \end{cases} \tag{13}$$

and Δ is infinity for $\delta = 0$ and 0 for $\delta = 1$.

We readily see that f_3 satisfies (8) for $p \in (1,2]$, so (13) will give bounded estimates for $\mathcal{L}_p$-noise, $1 < p \leq 2$. We also know from the previous sections that if $v \in \mathcal{L}_2$ then $e \in \mathcal{L}_\infty$ with bound ν_e. If we further assume that v is bounded, we may bound $|e+v|$ by $\nu_e + \|v\|_{2,\infty}$. Thus, if $|\epsilon| > \Delta$, $\dot{V} = -\text{sign}(e+v)e \leq -\frac{1}{\nu_e + \|v\|_{2,\infty}}(e^2 + ev)$.

For $\epsilon \leq \Delta$ we have the gradient algorithm again, so using (4) we obtain the contraction

$$\frac{\tilde{\theta}^T(\tau)\Gamma^{-1}\tilde{\theta}(\tau) + \mu \int_0^\tau e^2 dt}{\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(\tau) + \nu \int_0^\tau v^2 dt} \leq 1, \tag{14}$$

where $\mu = \min\left\{\frac{1}{2}, \frac{1}{2(\nu_e + \|v\|_{2,\infty})}\right\}$, and $\nu = \max\left\{\frac{1}{2}, \frac{1}{2(\nu_e + \|v\|_{2,\infty})}\right\}$.

4. A Leakage-Based Adaptive Algorithm

In fact, there exist bounds for $\|\tilde{\theta}\|$ that depend on $\|v\|_{2,\infty}$, the norm of the true parameter vector $\|\theta\|$, and on some design constants (see [5, Ch. 8]). Most of the convergence results for this that the estimates $\hat{\theta}$ will converge to a *domain* containing the true parameter θ.

In this section, we establish that the scheme can still converge to the *true* solution θ (and not to a domain around it) if the noise is further restricted to lie in $\mathcal{L}_p \cap \mathcal{L}_\infty$ for any finite p.

In the switching-σ scheme, the update law (2) is modified as follows [5, p. 587]:

$$\dot{\hat{\theta}}(t) = \Gamma\phi(t)\epsilon(t) - \Gamma\sigma_s(t)\hat{\theta}(t) , \tag{15}$$

where the switching function $\sigma_s(t)$ is defined according to the following rules: Let σ_0 and M_0 be positive constants selected by the designer. Then

$$\sigma_s(t) = \begin{cases} 0 & \text{if } \|\hat{\theta}\|_2(t) \leq M_0 \\ \sigma_0\left(\frac{\|\hat{\theta}(t)\|_2}{M_0} - 1\right) & \text{if } M_0 < \|\hat{\theta}(t)\|_2 \leq 2M_0 \\ \sigma_0 & \text{if } \|\hat{\theta}(t)\|_2 > 2M_0 \end{cases}$$

In other words, as long as the estimate $\hat{\theta}$ remains inside a disc of size M_0, the switching modification is not activated. The choice of M_0 requires the designer to have some a priori knowledge of the size of the unknown parameter θ.

The resulting error equation becomes

$$\dot{\tilde{\theta}}(t) = -\Gamma\phi(t)\phi^T(t)\tilde{\theta}(t) + \Gamma\sigma_s(t)\hat{\theta}(t) - \Gamma\phi v(t) . \tag{16}$$

That this scheme guarantees boundedness of the estimates for any bounded noise is a well-known result [5, p. 587].

We now prove that the switching-σ modification does not destroy the exponential stability property of the gradient algorithm with PE regressors. This observation is the basis for our proof of the convergence of the parameter estimates when the noise is in $\mathcal{L}_p \cap \mathcal{L}_\infty$, for any finite $p > 0$. In the following statement, $\kappa(\Gamma)$ denotes the condition number of the matrix Γ.

Theorem 3 (Convergence of Switching Sigma) *Consider the switching-σ scheme (15) and assume that the regressor ϕ is PE and M_0 is chosen as $M_0 > \kappa(\Gamma)\|\theta\|_2$. The following facts hold:*

(i) The origin in the error equation (16), $\tilde{\theta} = 0$, is exponentially stable when the noise v is identically zero.

(ii) If $v \in \mathcal{L}_\infty$, the parameter estimation errors converge exponentially fast to the residual set

$$\Theta_s = \left\{\tilde{\theta} : \|\tilde{\theta}\|_2 \leq \lambda\|v\|_{2,\infty}\right\} \tag{17}$$

for some positive constant λ.

(iii) If $v \in \mathcal{L}_p \cap \mathcal{L}_\infty$ for any finite $p > 1$, then there exist positive constants α, γ and γ_v such that

1. $\|\tilde{\theta}\|_{2,p}^p < \frac{\alpha}{\gamma}\|\tilde{\theta}(0)\|_2^p + \frac{\gamma_v}{\gamma}\|v\|_{2,p}^p$,

2. $\dot{\tilde{\theta}} \in \mathcal{L}_p \cap \mathcal{L}_\infty$, *3.* $\lim_{t\to\infty} \tilde{\theta} = 0$.

Here, unlike for the algorithms in the previous section, α, γ and γ_v will depend not only on p, but also on ϕ.

Proof: The argument is based on a Converse Lyapunov Theorem (e.g., [15, p. 244]). Assuming the regressor ϕ is PE, the error equation for the gradient algorithm (2) is exponentially stable. This implies that there exists a Lyapunov function $V(t,\tilde{\theta})$, and α_1, α_2, α, $\gamma > 0$, such that for any $p > 1$ and for any $\tilde{\theta}$ the following relations are satisfied: $\alpha_1\|\tilde{\theta}\|_2^p \leq V(t,\tilde{\theta}) \leq \alpha_2\|\tilde{\theta}\|_2^p$, $\|\frac{\partial V}{\partial \tilde{\theta}}\|_2 \leq \alpha\|\tilde{\theta}\|_2^{p-1}$, and

$$\dot{V} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \tilde{\theta}}\left(-\Gamma\phi\phi^T\right) \leq -\gamma\|\tilde{\theta}\|_2^p.$$

Now introduce the non-negative function

$$W(t,\tilde{\theta}) = \beta V(t,\tilde{\theta}) + \frac{1}{p}\|\tilde{\theta}\|_2^p$$

for some $\beta > 0$. Differentiating W in the direction of the flow of (15) we establish that, in the absence of noise v, W can be taken as a Lyapunov function for the switching-σ error system (16). This enables us to show that the switching modification, under proper conditions on M_0, does not destroy the exponential stability of the original gradient error system (3).

■

Note that the above result does not imply a contraction relation as (5) – if the regressor vector ϕ is not PE, it is possible to have noise in a space $\mathcal{L}_p$ $(p > 2)$ and $e \notin \mathcal{L}_p$. As an example, take a scalar system with $\phi(t) = 1$ for $t \in [e^k, 1 + e^k]$, $k \geq 0$, and $(1+t)^{-\frac{1}{2}}$ otherwise. Let the noise be $v(t) = \frac{1}{\sqrt{t+1}}$. Choosing M_0 large enough, one can show that $e(t) \notin \mathcal{L}_p$ for any $p > 0$, even though $v \in \mathcal{L}_{2+\delta}$ for any $\delta > 0$.

5. Conclusion

We have given a general (sufficient) condition for an algorithm in the form (6) to be $\mathcal{L}_p$-stable (for some given $p > 1$) – a condition that does not require persistantly exciting regressors. We discussed some of the estimation algorithms that satisfy this condition, and in some contraction relations between $v(t)$ and $e(t)$ thus generalizing the results of [12] for the $\mathcal{L}_2$ norm. value of p is explicitly used in the equations, i.e., we must know a priori that the noise $v(t)$ has finite p-norm, for some p. This means that an algorithm designed for $p = 4$ will not be robust with respect to noise in $\mathcal{L}_5$.

The switching-σ algorithm, on the other hand, guarantees boundedness of the estimates for noise in $\mathcal{L}_\infty$. We showed that even though, in general, this algorithm does not satisfy a contraction relation as in (5), under some conditions (one of which is PE), the origin of its error equation will be exponentially stable and therefore the algorithm will give converging parameter estimates even with $\mathcal{L}_p$ noise.

References

[1] K.J. Åström and B. Wittenmark. *Adaptive Control.* Addison-Wesley, 1995.

[2] B. Hassibi, A.H. Sayed, and T. Kailath. H^∞ optimality of the LMS algorithm. *IEEE Trans. on Signal Processing*, 44(2):267–280, February 1996.

[3] P.J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

[4] P.A. Ioannou and P. V. Kokotović. Instability analysis and improvement of robustness of adaptive control. *Automatica*, 20(5):583–594, 1984.

[5] P.A. Ioannou and J. Sun. *Robust Adaptive Control.* PTR Prentice-Hall, NJ, 1996.

[6] H.K. Khalil. *Nonlinear Systems.* Prentice-Hall, 2nd edition, 1996.

[7] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification.* The MIT Press, 1983.

[8] D.S. Mitrinović. *Analytic Inequalities.* Springer-Verlag, 1970.

[9] B.T. Poljak and YA.Z. Tsypkin. Adaptive estimation algorithms: convergence, optimality, robustness. *Automation and Remote Control*, 3:71–84, 1979.

[10] B.T. Poljak and YA.Z. Tsypkin. Robust identification. *Automatica*, 16:53–63, 1980.

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

[12] A.H. Sayed and P.V. Kokotović. A note on a Lyapunov argument for stochastic gradient methods in the presence of noise. In *Proc. American Control Conference*, Seattle, WA, June 1995.

[13] G. Tao. Inherent robustness of MRAC schemes. Preprint.

[14] A. van der Schaft. *$\mathcal{L}_2$-Gain and Passivity Techniques in Nonlinear Control.* Springer, New York, 1996.

[15] M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice-Hall, 2nd edition, 1993.