ELSEVIER

Systems & Control Letters 30 (1997) 83–91

SYSTEMS & CONTROL LETTERS

Stable on-line parameter identification with general knowledge on level of information noise Discrete-time case[1]

Fu-Ming Lee[a], Li-Chen Fu[b,*], I-Kong Fong[c]

[a] *Department of Information Management, Chaoyang Institute of Technology, Taichung, Taiwan, Republic of China*
[b] *Department of Electrical Engineering, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, Republic of China*
[c] *Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, Republic of China*

Received 24 February 1996; revised 20 May 1996

Abstract

An on-line parameter identification problem is posed and solved for discrete-time systems with general knowledge on the level of the inherent information noise. The knowledge can be the bound on either the magnitude or the finite-index ℓ^p norm, $p \in [1, \infty)$, of the noise. Based on the knowledge, a switching type gradient algorithm (or called gradient algorithm with dead zone) is proposed to estimate the parameters of the system from the available input–output data. In spite of the existence of the noise, this on-line algorithm guarantees that the estimation error is monotonically decreasing, and the parameter estimate is convergent to a steady-state value under a mild condition. Furthermore, the algorithm is stable in the sense that the estimation error will converge to zero as the bound on the noise gradually diminishes. © 1997 Published by Elsevier Science B.V.

Keywords: Robust identification; Parameter identification; On-line identification; Gradient technique; Noise

1. Introduction

On-line parameter identification has several advantages over the off-line counterpart. The primary one is that it can be used to progressively generate more accurate models, based on which the associated controllers can be continuously refined to improve the performance of the controlled system. Many theoretical and application results have been developed in this context. For example, [1, 2, 8] consider the identification problem with magnitude information, [3] considers the problem with finite-index ℓ^2 norm information, and [6] considers the problem with either magnitude information or finite-index ℓ^2 norm information on the level of the inherent information noise. However, all the results obtained so far are not applicable to the problem with general finite-index ℓ^p norm information on the noise level, for $p \in [1, \infty)$ and $p \neq 2$. This information, however, is generally available for systems whose nonparametric uncertainties are characterized by bounded induced ℓ^p norms.

[1] A preliminary version of the present work was presented at the 1995 American Control Conference [5].

PII S0167-6911(96)00079-5

In this paper, an on-line parameter identification problem is posed for systems with general knowledge on the level of the inherent information noise. The knowledge can be the bound on either the magnitude or the finite-index ℓ^p norm, $p \in [1, \infty)$, of the noise. The goals of this problem are to find on-line a converging parameter estimate for the system and to reduce monotonically the estimation error. To solve this problem, a concrete switching type gradient algorithm (or called gradient algorithm with dead zone) is developed to estimate the parameters by using the available input–output data and bounding information about the noise.

Besides the fact that this paper considers the general knowledge on the level of the noise, there are several additional differences between the present and the previous works. For instance, in [1–3, 6], complex set-membership identification techniques are used to estimate the admissible parameter set, in which any point can be an admissible estimate of the parameters. When a point estimate is requested, usually the centroid of the set is chosen as the parameter estimate; however, no guarantee is made that the estimation error is monotonically decreasing. In contrast, the class of algorithms of the present work belong to switching type gradient ones. They are simpler than those based on set-membership identification techniques. Furthermore, in spite of the existence of the noise, each of the algorithms guarantees that, even without the sufficient richness condition on the regressor, the parameter estimate is convergent to a steady-state value under a mild condition, and the estimation error is monotonically decreasing. Note that the convergence and monotonicity properties of the algorithm are useful to the subsequent controller design and essential in establishing the performance and stability of identification-based adaptive control system. Finally, with the sufficient richness condition on the regressor, these algorithms are stable in the sense that the estimation errors will converge to zeros when the bound on the noise gradually diminishes.

In summary, the contribution of this paper are as follows. First, it examines and unifies many identification problems in a framework. Second, it develops a switching type gradient parameter identification technique to successfully handle systems with various bounding information on the noise level. Finally, it provides a thorough analysis of the performance of the algorithms developed here.

The layout of this paper is as follows. In Section 2, notations and a preliminary result are established. In Section 3, an on-line parameter identification problem is formulated precisely and an example belonging to this problem are examined. In Section 4, concrete algorithms are presented and analyzed for the problem. At the end, Section 5 gives the conclusion.

2. Notations and preliminaries

Here, some notations and definitions are introduced for subsequent use. Let $\mathbf{I}_0$, $\mathbf{R}_0$, $\mathbf{R}$, and $\mathbf{R}^q$ denote the set of nonnegative integers, nonnegative real numbers, real numbers, and q dimensional real vectors, respectively. Let $|v|_p := (\sum_{i=1}^{q} |v_i|^p)^{1/p}$ denote the p-norm in $\mathbf{R}^q$ space, $p \in [1, \infty)$. For any $k \in \mathbf{I}_0$, $\rho \in [1, \infty)$, and discrete-time function vector $x : \mathbf{I}_0 \to \mathbf{R}^q$, two ρ-exponentially-weighted ℓ^p norms on x are defined:

$$\|x\|_{\ell^p, k, \rho} := \left(\sum_{i=0}^{k} \rho^{-p(k-i)} |x(i)|_p^p \right)^{1/p}; \qquad \|x\|_{\ell^p, \rho} := \left(\sum_{i=0}^{\infty} \rho^{pi} |x(i)|_p^p \right)^{1/p}.$$

Then, an induced norm on a causal system Δ is defined:

$$\|\Delta\|_{\ell^p, \rho} := \sup_{\|x\|_{\ell^p, \rho} \neq 0} \frac{\|\Delta x\|_{\ell^p, \rho}}{\|x\|_{\ell^p, \rho}}.$$

By a causal system we mean a system whose output at index k depends only on the input of the system at indices smaller than or equal to k. Finally, a preliminary result is established for later use in Section 3.

Lemma 1. *For any number $p \in [1, \infty)$, $k \in \mathbf{I}_0$, and $\rho \in [1, \infty)$, and discrete-time function vector $x : \mathbf{I}_0 \to \mathbf{R}^q$, $\|\Delta x\|_{\ell^p, k, \rho} \leqslant \|x\|_{\ell^p, k, \rho}$ if Δ is a causal system with $\|\Delta\|_{\ell^p, \rho} \leqslant 1$.*

Proof. Let $\tilde{x}(i) = x(i)$ for all $i = 0, \ldots, k$, and $\tilde{x}(i) = 0$ for all $i > k$. Because Δ is causal,

$$\| \Delta x \|_{\ell^p,k,\rho} = \| \Delta \tilde{x} \|_{\ell^p,k,\rho}.$$

From definitions, we have

$$\| \Delta \tilde{x} \|_{\ell^p,k,\rho} = \rho^{-k} \left(\sum_{i=0}^{k} \rho^{pi} |(\Delta \tilde{x})(i)|_p^p \right)^{1/p} \leqslant \rho^{-k} \| \Delta \tilde{x} \|_{\ell^p,\rho}$$

$$\leqslant \rho^{-k} \| \tilde{x} \|_{\ell^p,\rho} = \rho^{-k} \left(\sum_{i=0}^{k} \rho^{pi} |\tilde{x}(i)|_p^p \right)^{1/p} = \| \tilde{x} \|_{\ell^p,k,\rho} = \| x \|_{\ell^p,k,\rho}.$$

This completes the proof. □

3. Problem formulation

Consider a class of single-input single-output systems which can be modeled as follows:

$$y(k) = \phi^{\mathrm{T}}(k)\theta^* + \eta(k), \quad \phi(k) \in \mathbf{R}^q, \ \theta^* \in \mathbf{R}^q, \ k \in \mathbf{I}_0, \tag{1}$$

where y is the observed scalar signal, ϕ is the observed regression vector signal, θ^* is the unknown parameter vector, and η is the unknown signal noise. The problem is to on-line identify the parameter θ^* of a system by using the observed signals y and ϕ. To proceed further, some additional information concerning the noise η has to be given.

(IF) Either the bounding information $\alpha_d(k)$ or $\alpha_{\ell^p}(k)$ with $p \in [1, \infty)$ is known such that $|\eta(k)| \leqslant \alpha_d(k)$ or $\| \eta \|_{\ell^p,k,\rho} \leqslant \alpha_{\ell^p}(k)$ at any time index $k \in \mathbf{I}_0$.

Now we are ready to restate the on-line parameter identification problem precisely as follows.

Given (i) the observed signals $\{ y(i), \phi(i) \}_{i=0}^{k}$, $k \in \mathbf{I}_0$, and (ii) either the bounding information $\alpha_d(k)$ or $\alpha_{\ell^p}(k)$ described in (IF);

Find an on-line algorithm $A_k|_{k=0}^{\infty}$ which maps the given information into a parameter estimate $\hat{\theta}(k)$ such that $\hat{\theta}(k)$ is convergent to a steady-state value and the estimation error $e(k) := |\hat{\theta}(k) - \theta^*|_2$ satisfies

$$\begin{cases} e(k+1) - e(k) < 0 & \text{if } \hat{\theta}(k+1) - \hat{\theta}(k) \neq 0, \\ e(k+1) - e(k) = 0 & \text{otherwise}, \end{cases} \quad \forall k \in \mathbf{I}_0. \tag{2}$$

Furthermore, the algorithm is stable in the sense that

$$\lim_{k \to \infty} e(k) = 0 \text{ if } \begin{cases} \lim_{k \to \infty} \alpha_d(k) = 0 & \text{when } \alpha_d(k) \text{ is given.} \\ \lim_{k \to \infty} \alpha_{\ell^p}(k) = 0 & \text{when } \alpha_{\ell^p}(k) \text{ is given.} \end{cases} \tag{3}$$

In other words, in spite of the existence of the noise, the parameter estimate will be settling down to a steady-state value and the estimation error is strictly decreasing whenever the parameter estimate updates. Furthermore, the estimation error converges to zero as the bound on the noise gradually diminishes.

These problems arise in both explicit plant parameter identification and direct adaptive control when the plant contains a perturbation. Here, for illustration purpose, we show an explicit identification case.

Example (*Identification for systems with nonparametric uncertainties*). The problem is to on-line identify a model of a linear plant G. The available input–output data are generated by $y = Gu$, where u is an applied input and y is the observed output. Suppose the transfer function of G is characterized by a factor form as follows:

$$G(z) = \frac{N(z) + \Delta_N(z) W_N(z)}{D(z) + \Delta_D(z) W_D(z)},$$

where

$$D(z) = 1 + a_1^* z^{-1} + \cdots + a_n^* z^{-n}, \qquad N(z) = b_1^* z^{-1} + \cdots + b_m^* z^{-m},$$

both W_N and W_D are (specified) known proper stable weighting functions, and $\Delta = [\Delta_N, \Delta_D]$ forms an uncertain system with $\|\Delta\|_{\ell^p,\rho} \leqslant 1$ for some $p \in [1, \infty)$ and $\rho \in [1, \infty)$. After some manipulation, the system equation $y = Gu$ can be rewritten into a linear regression form as (1) with

$$\eta = \Delta v, \quad v = [W_N u, -W_D y]^{\mathrm{T}},$$

$$\phi = [-z^{-1}y, \ldots, -z^{-n}y, z^{-1}u, \ldots, z^{-m}u]^{\mathrm{T}}, \qquad \theta^* = [a_1^*, \ldots, a_n^*, b_1^*, \ldots, b_m^*]^{\mathrm{T}}.$$

Because $\{u(i), y(i)\}_{i=0}^{k}$, $k \in \mathbf{I}_0$, are available and both W_N and W_D are known, it is clear that the series of the signal triplets $\{y(i), \phi(i), v(i)\}_{i=0}^{k}$, $k \in \mathbf{I}_0$, are available. Furthermore, by Lemma 1,

$$\|\eta\|_{\ell^p,k,\rho} \leqslant \|v\|_{\ell^p,k,\rho} := \alpha_{\ell^p}(k),$$

where $\|v\|_{\ell^p,k,\rho}$ can be constructed on-line by the following way:

$$\|v\|_{\ell^p,k,\rho}^{p} = \rho^{-p}\|v\|_{\ell^p,k-1,\rho}^{p} + |v(k)|_p^p, \quad \|v\|_{\ell^p,0,\rho}^{p} = |v(0)|_p^p.$$

Thus, the identification problem considered in this example belongs to the family of problems formulated above with bounding information $\alpha_{\ell^p}(k)$.

4. Main results

Note that the signal noise $\eta(k)$ in (1) is estimated by

$$\hat{\eta}(k) := y(k) - \phi^{\mathrm{T}}(k)\hat{\theta}(k) = \eta(k) - \phi^{\mathrm{T}}(k)\tilde{\theta}(k), \tag{4}$$

where $\tilde{\theta}(k) := \hat{\theta}(k) - \theta^*(k)$. Recall the problems formulated in Section 3. The following concrete on-line parameter identification algorithms are suggested to treat those corresponding problems. Let $r(\cdot)$ denote an arbitrarily chosen positive function and let $\mathrm{sign}(\cdot)$ be defined by

$$\mathrm{sign}(x) := \begin{cases} 1 & \text{if } x \geqslant 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Algorithm (*Parameter identification with information α_d, $A_k^d|_{k=0}^{\infty}$*)

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \lambda_d(k)\xi_d(k)q_d(k), \qquad \xi_d(k) = \frac{2(\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k))}{|q_d(k)|_2^2 + r(k)},$$

$$q_d(k) = \hat{\eta}(k)\phi(k), \qquad \lambda_d(k) = \begin{cases} 1 & \text{if } |\hat{\eta}(k)| > \alpha_d(k), \\ 0 & \text{otherwise.} \end{cases}$$

Algorithm (*Parameter identification with information α_{ℓ^p}, $A\ell_k^p|_{k=0}^{\infty}$*)

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \lambda_{\ell^p}(k)\xi_{\ell^p}(k)q_{\ell^p}(k), \qquad \xi_{\ell^p}(k) = \frac{2(\|\hat{\eta}\|_{\ell^p,k,\rho}^{p} - \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k))}{|q_{\ell^p}(k)|_2^2 + r(k)},$$

$$q_{\ell^p}(k) = \sum_{i=0}^{k} \rho^{-p(k-i)}|\hat{\eta}(i)|^{p-1}\,\mathrm{sign}(\hat{\eta}(i))\phi(i), \qquad \lambda_{\ell^p}(k) = \begin{cases} 1 & \text{if } \|\hat{\eta}\|_{\ell^p,k,\rho}^{p} > \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) + \beta_{\ell^p}(k), \\ 0 & \text{otherwise,} \end{cases}$$

$$\beta_{\ell^p}(k) = \sum_{i=0}^{k} \rho^{-p(k-i)}|\hat{\eta}(i)|^{p-1}\,\mathrm{sign}(\hat{\eta}(i))\phi^{\mathrm{T}}(i)(\hat{\theta}(k) - \hat{\theta}(i)).$$

Remark. The algorithms proposed above are indeed on-line algorithms since $q_{\ell^p}(k)$, $\|\hat{\eta}\|^p_{\ell^p,k,\rho}$, and $\beta_{\ell^p}(k)$ can be implemented by the following ways:

$$q_{\ell^p}(k) = \rho^{-p} q_{\ell^p}(k-1) + |\hat{\eta}(k)|^{p-1} \operatorname{sign}(\hat{\eta}(k))\phi(k), \quad q_{\ell^p}(0) = |\hat{\eta}(0)|^{p-1} \operatorname{sign}(\hat{\eta}(0))\phi(0),$$

$$\|\hat{\eta}\|^p_{\ell^p,k,\rho} = \rho^{-p}\|\hat{\eta}\|^p_{\ell^p,k-1,\rho} + |\hat{\eta}(k)|^p, \quad \|\hat{\eta}\|^p_{\ell^p,0,\rho} = |\hat{\eta}(0)|^p,$$

$$\beta_{\ell^p}(k) = \rho^{-p}\beta_{\ell^p}(k-1) + \rho^{-p}\lambda_{\ell^p}(k-1)\xi_{\ell^p}(k-1)|q_{\ell^p}(k-1)|_2^2, \quad \beta_{\ell^p}(0) = 0.$$

Remark. The parameter adjustment direction, $q_d(k)$, for algorithm $A^d_k|^\infty_{k=0}$ is apparently the gradient of $-\hat{\eta}^2(k)$ with respect to $\hat{\theta}(k)$. Likewise, the direction $q_{\ell^p}(k)$ for algorithm $A\ell^p_k|^\infty_{k=0}$ is conceptually the gradient of $-\|\hat{\eta}\|^p_{\ell^p,k,\rho}$ with respect to $\hat{\theta}$. Thus, the underlying identification technique is, in fact, a switching type gradient one. If $\lambda_d(k) = 1$ or $\lambda_{\ell^p}(k) = 1$, it can be shown that the estimation error $\tilde{\theta}(k)$ must be nonzero. As is shown in the following theorem, the parameter estimate $\hat{\theta}(k)$ is updated to reduce the estimation error.

Theorem 1. *When the algorithms $A^d_k|^\infty_{k=0}$ or $A\ell^p_k|^\infty_{k=0}$ are applied to system* (1), $p \in [1, \infty)$, *the following properties hold.*

(a) *The error $e(k) = |\hat{\theta}(k) - \theta^*|_2$ satisfies* (2) *and $\lim_{k\to\infty} e(k)$ exists.*

(b) *If $\liminf_{k \geq 0} r(k) > 0$ and*

r and q_d are bounded, for algorithm $A^d_k|^\infty_{k=0}$,
r and q_{ℓ^p} are bounded, for algorithm $A\ell^p_k|^\infty_{k=0}$,

then, given $\varepsilon > 0$, there exist integers $T_d > 0$ or $T_{\ell^p} > 0$ such that

$$|\hat{\eta}(k)| < \alpha_d(k) + \varepsilon, \quad \forall k \geq T_d, \qquad \textit{for algorithm } A^d_k|^\infty_{k=0},$$

$$\|\hat{\eta}\|^p_{\ell^p,k,\rho} < \|\hat{\eta}\|^{p-1}_{\ell^p,k,\rho}\alpha_{\ell^p}(k) + \beta_{\ell^p}(k) + \varepsilon, \quad \forall k \geq T_{\ell^p}, \quad \textit{for algorithm } A\ell^p_k|^\infty_{k=0}.$$

(c) *Let the* consistent parameter set *Θ^c_d be defined as*:

$$\Theta^c_d = \begin{cases} \{\theta : |y(k) - \phi^T(k)\theta| \leq \alpha_d(k), \forall k \in \mathbf{I}_0\} & \textit{for algorithm } A^d_k|^\infty_{k=0}, \\ \{\theta : \|y - \phi^T\theta\|_{\ell^p,k,\rho} \leq \alpha_{\ell^p}(k), \forall k \in \mathbf{I}_0\} & \textit{for algorithm } A\ell^p_k|^\infty_{k=0}, \end{cases}$$

then, for any $\theta \in \Theta^c_d$, the error $e_\theta(k) = |\hat{\theta}(k) - \theta|_2$ satisfies

$$\begin{cases} e_\theta(k+1) - e_\theta(k) < 0 & \textit{if } \hat{\theta}(k+1) - \hat{\theta}(k) \neq 0, \\ e_\theta(k+1) - e_\theta(k) = 0 & \textit{otherwise} \end{cases} \quad \forall k \in \mathbf{I}_0,$$

and $\lim_{k\to\infty} e_\theta(k)$ exists.

(d) *$\lim_{k\to\infty}\hat{\theta}(k)$ exists if the conditions in* (b) *are satisfied and there exist a positive integer j and some $\theta^1, \ldots, \theta^j \in \Theta^c_d$ such that $\bigcap_{i=1}^j \{v : |v - \theta^i|_2 = c_i\}$ consists of at most isolated points for any $c_i \in \mathbf{R}_0$.*

Remark. The conditions about Θ^c_d in (d) of Theorem 1 are almost always satisfied. For example, consider $\theta \in \mathbf{R}^3$, if there exist θ^1, θ^2, and $\theta^3 \in \Theta^c_d$ such that θ^1, θ^2, and θ^3 are not on a line, then $\bigcap_{i=1}^3 \{v : |v - \theta^i|_2 = c_i\}$ consists of at most two isolated points for any $c_i \in \mathbf{R}_0$.

Proof of Theorem 1. (a) For all $k \in \mathbf{I}_0$, let $V(k) = e^2(k) = \tilde{\theta}^T(k)\tilde{\theta}(k)$. First, we consider the algorithm $A^d_k|^\infty_{k=0}$. If $\lambda_d(k) = 0$, it is clear that $\hat{\theta}(k+1) - \hat{\theta}(k) = 0$ which implies $V(k+1) - V(k) = 0$; if $\lambda_d(k) = 1$, i.e.

$|\hat{\eta}(k)| > \alpha_d(k)$, we know that $\xi_d(k) > 0$ and

$$\begin{aligned}
&V(k+1) - V(k)\\
&\quad = (\tilde{\theta}(k) + \xi_d(k)q_d(k))^{\mathrm{T}}(\tilde{\theta}(k) + \xi_d(k)q_d(k)) - \tilde{\theta}^{\mathrm{T}}(k)\tilde{\theta}(k)\\
&\quad = 2\xi_d(k)q_d^{\mathrm{T}}(k)\tilde{\theta}(k) + \xi_d^2(k)|q_d(k)|_2^2\\
&\quad = 2\xi_d(k)\hat{\eta}(k)\phi^{\mathrm{T}}(k)\tilde{\theta}(k) + \xi_d^2(k)|q_d(k)|_2^2\\
&\quad = -2\xi_d(k)\hat{\eta}^2(k) + 2\xi_d(k)\hat{\eta}(k)\eta(k) + \xi_d^2(k)|q_d(k)|_2^2\\
&\quad \leqslant -2\xi_d(k)(\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k) - \tfrac{1}{2}\xi_d(k)|q_d(k)|_2^2)\\
&\quad = -\frac{4(\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k))}{|q_d(k)|_2^2 + r(k)}\left\{\frac{(\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k))(|q_d(k)|_2^2 + r(k)) - (\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k))|q_d(k)|_2^2}{|q_d(k)|_2^2 + r(k)}\right\}\\
&\quad = -\frac{4r(k)}{(|q_d(k)|_2^2 + r(k))^2}(\hat{\eta}^2(k) - |\hat{\eta}(k)|\alpha_d(k))^2 < 0.
\end{aligned}$$

In other words,

$$V(k+1) - V(k) \leqslant -\frac{4r(k)}{(|q_d(k)|_2^2 + r(k))^2}(|\hat{\eta}(k)|[|\hat{\eta}(k)| - \alpha_d(k)]^+)^2 \leqslant 0,$$

where

$$[x]^+ := \begin{cases} x & \text{if } x > 0,\\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Thus, $e(k)$ satisfies (2). Because $e(k)$ is monotonically decreasing and bounded below by zero, $\lim_{k\to\infty} e(k)$ exists [10, Theorem 3.14]. Similarly, considering the algorithm $A\ell_k^p|_{k=0}^{\infty}$, if $\lambda_{\ell^p}(k) = 0$, $\hat{\theta}(k+1) - \hat{\theta}(k) = 0$ which implies $V(k+1) - V(k) = 0$; if $\lambda_{\ell^p}(k) = 1$, i.e. $\|\hat{\eta}\|_{\ell^p,k,\rho}^p > \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) + \beta_{\ell^p}(k)$, we have $\xi_{\ell^p}(k) > 0$ and

$$\begin{aligned}
&V(k+1) - V(k)\\
&\quad = 2\xi_{\ell^p}(k)q_{\ell^p}^{\mathrm{T}}(k)\tilde{\theta}(k) + \xi_{\ell^p}^2(k)|q_{\ell^p}(k)|_2^2\\
&\quad = 2\xi_{\ell^p}(k)\sum_{i=0}^{k}\rho^{-p(k-i)}|\hat{\eta}(i)|^{p-1}\operatorname{sign}(\hat{\eta}(i))\phi^{\mathrm{T}}(i)\tilde{\theta}(i) + 2\xi_{\ell^p}(k)\beta_{\ell^p}(k) + \xi_{\ell^p}^2(k)|q_{\ell^p}(k)|_2^2\\
&\quad = -2\xi_{\ell^p}(k)\|\hat{\eta}\|_{\ell^p,k,\rho}^p + 2\xi_{\ell^p}(k)\sum_{i=0}^{k}\rho^{-p(k-i)}|\hat{\eta}(i)|^{p-1}\operatorname{sign}(\hat{\eta}(i))\eta(i) + 2\xi_{\ell^p}(k)\beta_{\ell^p}(k) + \xi_{\ell^p}^2(k)|q_{\ell^p}(k)|_2^2.
\end{aligned}$$

By Hölder inequality [9, p. 122], we know that

$$\sum_{i=0}^{k}\rho^{-p(k-i)}|\hat{\eta}(i)|^{p-1}\operatorname{sign}(\hat{\eta}(i))\eta(i) \leqslant \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\|\eta\|_{\ell^p,k,\rho},$$

which implies

$$\begin{aligned}
V(k+1) - V(k) &\leqslant -2\xi_{\ell^p}(k)(\|\hat{\eta}\|_{\ell^p,k,\rho}^p - \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k) - \tfrac{1}{2}\xi_{\ell^p}(k)|q_{\ell^p}(k)|_2^2)\\
&= -\frac{4r(k)}{(|q_{\ell^p}(k)|_2^2 + r(k))^2}(\|\hat{\eta}\|_{\ell^p,k,\rho}^p - \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k))^2 < 0.
\end{aligned}$$

Again,

$$V(k+1) - V(k) \leqslant -\frac{4r(k)}{(|q_{\ell^p}(k)|_2^2 + r(k))^2}([\|\hat{\eta}\|_{\ell^p,k,\rho}^p - \|\hat{\eta}\|_{\ell^p,k,\rho}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k)]^+)^2 \leqslant 0,$$

and hence $e(k)$ satisfies (2) and $\lim_{k\to\infty} e(k)$ exists.

(b) For the algorithm $A_k^d|_{k=0}^{\infty}$, let

$$f_d(k) = \frac{4r(k)}{(|q_d(k)|_2^2 + r(k))^2}(|\hat{\eta}(k)|[|\hat{\eta}(k)| - \alpha_d(k)]^+)^2.$$

From the proof of (a), we know that $V(k+1) - V(k) \leqslant -f_d(k) \leqslant 0$ which implies f_d is summable and $\lim_{k\to\infty} f_d(k) = 0$. Because $\liminf_{k\geqslant 0} r(k) > 0$, and both r and q_d are bounded, we have

$$\lim_{k\to\infty} [|\hat{\eta}(k)| - \alpha_d(k)]^+ = 0.$$

Similarly, for the algorithm $A\ell_k^p|_{k=0}^{\infty}$, let

$$f_{\ell^p}(k) = \frac{4r(k)}{(|q_{\ell^p}(k)|_2^2 + r(k))^2}([\|\hat{\eta}\|_{\ell^p,k,p}^p - \|\hat{\eta}\|_{\ell^p,k,p}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k)]^+)^2.$$

It can be shown that $\lim_{k\to\infty} f_{\ell^p}(k) = 0$ and

$$\lim_{k\to\infty} [\|\hat{\eta}\|_{\ell^p,k,p}^p - \|\hat{\eta}\|_{\ell^p,k,p}^{p-1}\alpha_{\ell^p}(k) - \beta_{\ell^p}(k)]^+ = 0.$$

Thus (b) is true.

(c) First, we consider the algorithm $A_k^d|_{k=0}^{\infty}$. For any $\theta \in \Theta_d^c$, it is clear that there exists a signal η_θ such that $y(k) = \phi^{\mathrm{T}}(k)\theta + \eta_\theta(k)$ and $\eta_\theta(k) \leqslant \alpha_d(k)$ for all $k \in \mathbf{I}_0$. From the proof of (a), we know that (c) is true for the algorithm $A_k^d|_{k=0}^{\infty}$. Similarly, (c) is true for the algorithm $A\ell_k^p|_{k=0}^{\infty}$.

(d) For the algorithm $A_k^d|_{k=0}^{\infty}$, it is clear from definitions and (b) that

$$\lim_{k\to\infty} \lambda_d(k)\xi_d(k) = 0,$$

which implies

$$\lim_{k\to\infty} (\hat{\theta}_i(k+1) - \hat{\theta}_i(k)) = 0, \quad \forall i = 1, \ldots, \dim(\hat{\theta}), \tag{5}$$

since q_d is uniformly bounded. From (c) we know that, for any $i = 1, \ldots, j$, $|\hat{\theta}(k) - \theta^i|_2$ is monotonically decreasing and $\lim_{k\to\infty} |\hat{\theta}(k) - \theta^i|_2 = c_i$ for some $c_i \in \mathbf{R}_0$. Thus,

$$\lim_{k\to\infty} \min\left\{ |\hat{\theta}(k) - \theta|_2 : \theta \in \bigcap_{i=1}^{j} \{v : |v - \theta^i|_2 = c_i\} \right\} = 0.$$

Because $\bigcap_{i=1}^j \{v : |v - \theta^i|_2 = c_i\}$ consists of at most isolated points and $\hat{\theta}(k+1) - \hat{\theta}(k)$ converges to zero as described in (5), $\lim_{k\to\infty} \hat{\theta}(k) = \theta$ for some $\theta \in \bigcap_{i=1}^j \{v : |v - \theta^i|_2 = c_i\}$. Similarly, it can be shown that (d) is true for the algorithm $A\ell_k^p|_{k=0}^{\infty}$. □

Remark. The results of (b) in Theorem 1 mean that the steady-state performances of the algorithms $A_k^d|_{k=0}^{\infty}$ or $A\ell_k^p|_{k=0}^{\infty}$ depend on the bounds $\alpha_d(k)$ or $\alpha_{\ell^p}(k)$, respectively. Therefore, if these bounds are smaller because the noise level is lower, then the steady-state performance is better. As is shown in the following theorem, those algorithms are stable.

Theorem 2. *When the algorithms $A_k^d|_{k=0}^{\infty}$ or $A\ell_k^p|_{k=0}^{\infty}$ are applied to system* (1), $p \in [1, \infty)$, *and the conditions of* (b) *in Theorem* 1 *are satisfied, the error $e(k)$ satisfies* (3) *if,*

(a) *for algorithm $A_k^d|_{k=0}^{\infty}$, $\lim_{k\to\infty} \hat{\theta}(k)$ exists and ϕ is sufficiently rich that for any convergent function vector x*

$$\lim_{k\to\infty} |\phi^{\mathrm{T}}(k)x(k)| = 0 \quad \textit{implies} \quad \lim_{k\to\infty} |x(k)|_2 = 0;$$

(b) *for algorithm* $A\ell_k^p|_{k=0}^{\infty}$, $\rho > 1$, *both* y *and* ϕ *are bounded*, $\lim_{k\to\infty}\hat{\theta}(k)$ *exists, and* ϕ *is sufficiently rich that for any convergent function vector* x

$$\lim_{k\to\infty} \|\phi^{\mathrm{T}}x\|^p_{\ell^p,k,\rho} = 0 \quad \textit{implies} \quad \lim_{k\to\infty} |x(k)|_2 = 0.$$

Proof. (a) Because $|\eta(k)| \leqslant \alpha_d(k)$, from (4) and the results of (b) in Theorem 1 we know that

$$|\phi^{\mathrm{T}}(k)\tilde{\theta}(k)| \leqslant |\hat{\eta}(k)| + |\eta(k)| < 2\alpha_d(k) + \varepsilon, \quad \forall k \geqslant T_d,$$

which implies $\lim_{k\to\infty}|\phi^{\mathrm{T}}(k)\tilde{\theta}(k)| = 0$ if $\lim_{k\to\infty}\alpha_d(k) = 0$. Thus, $e(k)$ satisfies (3) since $\lim_{k\to\infty}\hat{\theta}(k)$ exists and ϕ is sufficiently rich.

(b) It can be shown that $\lim_{k\to\infty}\beta_{\ell^p}(k) = 0$ when $\rho > 1$, both y and ϕ are bounded, and $\lim_{k\to\infty}\hat{\theta}(k)$ exists. Because $\|\eta\|_{\ell^p,k,\rho} \leqslant \alpha_{\ell^p}(k)$, from (4) and the results of (b) in Theorem 1 we know that

$$\|\phi^{\mathrm{T}}\tilde{\theta}\|^p_{\ell^p,k,\rho} \leqslant 2^p(\|\hat{\eta}\|^p_{\ell^p,k,\rho} + \|\eta\|^p_{\ell^p,k,\rho}) < 2^p(\|\hat{\eta}\|^{p-1}_{\ell^p,k,\rho}\alpha_{\ell^p}(k) + \beta_{\ell^p}(k) + \varepsilon + \alpha^p_{\ell^p}(k)), \quad \forall k \geqslant T_{\ell^p},$$

which implies $\lim_{k\to\infty}\|\phi^{\mathrm{T}}\tilde{\theta}\|^p_{\ell^p,k,\rho} = 0$ if $\lim_{k\to\infty}\alpha_{\ell^p}(k) = 0$. Thus, $e(k)$ satisfies (3) since $\lim_{k\to\infty}\hat{\theta}(k)$ exists and ϕ is sufficiently rich. □

Remark. In view of (c) in Theorem 1, the size of Θ_d^c depends on ϕ and α_d (or α_{ℓ^p}). When ϕ is sufficiently rich and $\lim_{k\to\infty}\alpha_d(k) = 0$ (or $\lim_{k\to\infty}\alpha_{\ell^p}(k) = 0$), this set reduces to a single point, namely θ^*.

Remark. If, in addition, a compact convex set Θ_0 is given in advance such that $\theta^* \in \Theta_0$, a projection operator can be applied to the algorithms to ensure $\hat{\theta} \in \Theta_0$ while retaining the other properties of the algorithms. Furthermore, the modified algorithms can be successfully applied to identify the parameter of systems with both nonparametric uncertainties and disturbances [4, 7].

Remark. This paper deals with discrete-time case. The results for continuous-time case are shown in [5].

5. Conclusion

An on-line parameter identification problem is formulated and solved for systems with general knowledge on the level of the inherent information noise. A switching type gradient algorithm (or called gradient algorithm with dead zone) is developed for this problem, which guarantees that the estimation error is decreasing strictly whenever the parameter estimate updates, and the parameter estimate is convergent to a steady state value under a mild condition. Furthermore, the algorithm is stable in the sense that the estimation error will converge to zero as the bound on the noise gradually diminishes.

For an identification-based adaptive control system, it is essential that the parameter estimate of the identification algorithm is convergent and the estimation error is monotonically decreasing. With those convergence and monotonicity properties, each of the on-line algorithms proposed in this paper has a great potential to be incorporated in a robust adaptive control discrete-time system as a robust estimator. Further research is underway to establish the stability and performance of an adaptive control system using these identification techniques.

References

[1] G. Belforte and T.T. Tay, Two new estimation algorithms for linear models with unknown but bounded measurement noise, *IEEE Trans. Automat. Control* **38** (1993) 1273–1279.

[2] M.-F. Cheung, S. Yurkovich and K.M. Passino, An optimal volume ellipsoid algorithm for parameter set estimation, *IEEE Trans. Automat. Control* **38** (1993) 1292–1296.

[3] E. Fogel, System identification via membership set constraints with energy constrained noise, *IEEE Trans. Automat. Control* **24** (1979) 752–758.

[4] F.-M. Lee, I-K. Fong and L.-C. Fu, Parameter identification of systems with uncertain dynamics and disturbances, *Proc. Amer. Control Conf.* (1992) 2445–2449.
[5] F.-M. Lee, L.-C. Fu and I-K. Fong, Robust on-line parameter identification with general knowledge on level of information noise: continuous and discrete cases, *Proc. Amer. Control Conf.* (1995) 4077–4081.
[6] F.-M. Lee, I-K. Fong and L.-C. Fu, Robust on-line set-membership identification using a general ellipsoid-bounding technique, *Automatica*, submitted.
[7] F.-M. Lee, I-K. Fong and L.-C. Fu, Stable on-line parameter identification algorithms for systems with nonparametric uncertainties and disturbances, *Internat. J. Control* **65** (1996) 329–345.
[8] R. Lozano-Leal and R. Ortega, Reformulation of the parameter identification problem for systems with bounded disturbances, *Automatica* **23** (1987) 247–251.
[9] H.L. Royden, *Real Analysis* (Macmillan, New York, 3rd ed., 1988).
[10] W. Rudin, *Principles of Mathematical Analysis* (McGraw-Hill, New York, 3rd ed., 1976).