New Convergence Results for Least Squares Identification Algorithm

Xiao-Li Hu, Lennart Ljung
Division of Automatic Control
E-mail: xlhu@amss.ac.cn, ljung@isy.liu.se

13th May 2009

Report no.: LiTH-ISY-R-2904
Accepted for publication in The 17:th IFAC World Congress in Seoul, Korea, 2008

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se
Abstract

The basic least squares method for identifying linear systems has been extensively studied. Conditions for convergence involve issues about noise assumptions and behavior of the sample covariance matrix of the regressors. Lai and Wei proved in 1982 convergence for essentially minimal conditions on the regression matrix: All eigenvalues must tend to infinity, and the logarithm of the largest eigenvalue must not tend to infinity faster than the smallest eigenvalue. In this contribution we revisit this classical result with respect to assumptions on the noise: How much unstructured disturbances can be allowed without affecting the convergence? The answer is that the norm of these disturbances must tend to infinity slower than the smallest eigenvalue of the regression matrix.

Keywords: System Identification
New Convergence Results for the Least Squares Identification Algorithm

Xiao-Li Hu * Lennart Ljung **

* Department of Mathematics, College of Science, China Jiliang University, Hangzhou, 310018, China
** Department of Electrical Engineering, Linköping University, Linköping, 58183, Sweden

Abstract: The basic least squares method for identifying linear systems has been extensively studied. Conditions for convergence involve issues about noise assumptions and behavior of the sample covariance matrix of the regressors. Lai and Wei proved in 1982 convergence for essentially minimal conditions on the regression matrix: All eigenvalues must tend to infinity, and the logarithm of the largest eigenvalue must not tend to infinity faster than the smallest eigenvalue. In this contribution we revisit this classical result with respect to assumptions on the noise: How much unstructured disturbances can be allowed without affecting the convergence? The answer is that the norm of these disturbances must tend to infinity slower than the smallest eigenvalue of the regression matrix.

1. INTRODUCTION

The least squares method for identifying simple dynamical models like

\[ y_n + a_1 y_{n-1} + \ldots + a_p y_{n-p} = b_1 u_{n-1} + \ldots + b_q u_{n-q} + \bar{w}_n \]

is probably the most used, and most extensively analyzed identification method. Its origin in this application is the classical paper by Mann & Wald (1943). There have been many efforts to establish minimal conditions under which the estimates of \( a \) and \( b \) converge to their true values. Since (1) is the archetypal model for adaptive control applications, such convergence results are also tied to the asymptotic behavior of adaptive regulators.

The convergence of the estimates will depend on two factors:
- The nature of the disturbance \( \bar{w} \).
- The properties of the regression vector

\[ \varphi(t) = [-y_n \ldots -y_{n-p} \ u_{n-1} \ldots u_{n-q}]^T \]

associated with (1)

Let

\[ R_n = \sum_{t=1}^{n} \varphi(t)\varphi(t)^T \]

(3)

Classical convergence results were obtained for the case where \( \bar{w} \) is white noise and \( R_n/n \) converges to a non-singular matrix. See, e.g. Åström & Eykhoff (1971). In Ljung (1976) it was shown that it is sufficient that \( \bar{w}_n \) is a martingale difference and that \( \lambda_{\min}(R_n) \to \infty \) (where \( \lambda_{\min}(A) \) denotes the smallest eigenvalue of the matrix \( A \)) in case the estimation is done for a finite collection of parameter values. In the 70’s it was generally believed that these conditions would also suffice for continuous parameterizations, and several attempts were made to prove that. Such a result would have been very welcome for the analysis of adaptive controllers. However, in 1982, Lai & Wei (1982) proved that, in addition, it is necessary that the logarithm of the largest eigenvalue of \( R_n \) does not grow faster than the smallest eigenvalue. Later, important related results have been obtained by e.g. Chen & Guo (1991), Guo (1995).

It is the purpose of the current paper to revisit the celebrated results of Lai and Wei, by examining how to relax the first condition, that \( \epsilon \) is a martingale difference. We shall work with the assumption that

\[ \bar{w}_n = w_n + \delta_n \]

(4)

where \( w_n \) is a martingale difference and \( \delta \) is an arbitrary, not necessarily stochastic disturbance.

2. MOTIVATION AND NUMERICAL EXAMPLES

Let us do some numerical experiments of LS estimation of the parameters for the following SISO linear system

\[ y_{n+1} + ay_n = bu_n + \delta_n + w_{n+1}, \]

(5)

where \( a = 0.5, b = 1 \), with white noise \( w_n \in \mathcal{N}(0, 0.5^2) \), and \( \delta_n \) is a deterministic or random disturbance, that does not necessarily tend to 0.

From Fig. 1, we can see that although there are non-decaying disturbances, the LS algorithm may still work nicely. Thus, we may ask that whether zero mean of the noise is necessary for the convergence of LS algorithm. Clearly in the example, although the disturbance tend to zero it appear more and more seldom, so it impact is limited.

From Fig. 2 we can see that the LS-estimate may still work even with a disturbance with unbounded norm. How to explain the convergence in this case? Clearly, in the example, the growing disturbance is compensated for by an input of increasing amplitude.
The well known Least square estimate (LSE) is convergence results.

Aström & Eykhoff (1971).

\[ \theta = \begin{cases} 1, & \text{if } n = k^2, \ k = 1, 2, \ldots, \\ 0, & \text{otherwise}. \end{cases} \]

where \( z \) and disturbance resp., and
\( \{ F_n \} \) are \( \sigma \)-algebras, satisfying
\[ \sup_{n \geq 0} E[||w_{n+1}||^2 | F_n] \trianglelefteq \sigma < \infty \ a.s., \ \beta \geq 2; \]

H2. \( u_n \) is \( F_n \)-measurable, and \( \delta_n \) is a deterministic signal or \( F_n \)-measurable random variable.

For convenience, by \( M_k = O(\varepsilon) \) (ordo) we mean that there is a constant \( C \geq 0 \) such that
\[ |M_k| \leq C \varepsilon, \ \forall k \geq 0. \]

Also by \( f_n = o(g_n), n \to \infty \) (small ordo) we mean
\[ \frac{f_n}{g_n} \to 0 \text{ as } n \to \infty. \]

Denote \( \lambda_{\max}(n) \) and \( \lambda_{\min}(n) \) as the maximum and minimum eigenvalue of the matrix
\[ P_n^{-1} = \sum_{i=0}^{n} \varphi_i \varphi_i^T + \frac{1}{\alpha_0} I. \] (10)

For simplicity, denote
\[ \rho_\beta(x) \triangleq \frac{1}{(\log \log x)^\beta}, \ \beta > 2, \]
with arbitrary \( c > 1 \).

Then we have the following basic result:

**Theorem 3.1.** Assume that conditions H1 and H2 are satisfied. Let \( \theta_n \) be the LSE (9) and let \( \theta \) be the true value (7). Then the error has the following bound with probability one:
\[ ||\theta_{n+1} - \theta||^2 = O \left( \frac{\log \lambda_{\max}(n) \cdot \rho_\beta(\lambda_{\max}(n)) + \sum_{i=0}^{n} \delta_i^2}{\lambda_{\min}(n)} \right) \] (12)

where \( \rho_\beta \) is defined by (11).

If \( \delta_n = 0 \) for each \( n \), Theorem 3.1 turns out to be Theorem 4.1 in Chen & Guo (1991) for the white noise case. It is also worth pointing out that the bound (or convergence) rate \( \frac{\log \lambda_{\max}(n)}{\lambda_{\min}(n)} \) for estimation error was first shown in the breakthrough paper Lai & Wei (1982). The extended LS identification scheme for ARMA model with errors \( \delta_n \) has been discussed in Chen & Deniau (1994), where a similar (somewhat special) result is established. Also, the proof of Theorem 3.1 that follows, uses some techniques and ideas in Chen & Deniau (1994); Chen & Guo (1991); Lai & Wei (1982).

With \( tr(A) \) denoting the trace of a matrix, we have from (10)
\[ tr(P_n^{-1}) = a_0 + \sum_{i=0}^{n} \varphi_i \varphi_i^T \trianglelefteq r_n. \] (13)

Together with the non-negativeness of \( P_n^{-1} \), then we get a corollary of Theorem 3.1 as follows.

**Corollary 3.1.** Under the same conditions of Theorem 3.1, we have the following bound on the estimation error:
\[ ||\theta_{n+1} - \theta||^2 = O \left( \frac{\log r_n \cdot \rho_\beta(r_n) + \sum_{i=0}^{n} \delta_i^2}{\lambda_{\min}(n)} \right) \ a.s., \] (14)

where \( r_n \) is defined by (13).
We list Theorem 2.8 of Chen & Guo (1991) as a lemma here.

**Lemma 3.1.** Let \( \{x_n, F_n\} \) be a martingale difference sequence and \( \{M_n, F_n\} \) an adapted sequence of random variables \( |M_n| < \infty \) a.s., \( \forall n \geq 0 \). If
\[
\sup_n E[|x_n|^{\alpha}|F_n] < \infty \quad \text{a.s.}
\]
for some \( \alpha \in (0, 2] \), then as \( n \to \infty \)
\[
\sum_{i=0}^{n} M_i x_i = O \left( s_n(\alpha) \cdot \log^{1/\eta} \left( s_n(\alpha) \right) \right) \quad \text{a.s., } \forall \eta > 0,
\]
where
\[
s_n(\alpha) = \left( \sum_{i=0}^{n} |M_i|^\alpha \right)^{1/\alpha}.
\]

**Remark.** For simple notation we use here and in the rest of the paper the convention \( \log x = \max(\log x, 1) \).

**Lemma 3.2.** Let \( \{w_n, F_n\} \) be a martingale difference sequence satisfying H1, then
\[
\sum_{i=0}^{n+1} \varphi_i^T P_i \varphi_i = O \left( \log \lambda_{\max}(n) \right),
\]
(16)
\[
\sum_{i=0}^{n+1} \varphi_i^T P_i \varphi_i w_{i+1}^2 = O \left( \log \lambda_{\max}(n) \cdot \rho_\beta \right),
\]
(17)
where \( P_i \) and \( \delta(\beta) \) are defined by (8a) and (11) respectively.

**Proof.** We first note a basic fact (see Lai & Wei (1982)):
\[
|I + \alpha \beta^T| = 1 + \beta^T \alpha,
\]
(18)
where \( I \) is an \( n \times n \) identity matrix, \( \alpha \) and \( \beta \) are two \( n \times 1 \) vectors, and \( |\cdot| \) is the operator norm. Obviously, if \( \alpha = 0 \), i.e., a zero vector, (18) holds. When \( \alpha \neq 0 \), we have
\[
|I + \alpha \beta^T| \alpha = (1 + \beta^T \alpha) \alpha,
\]
which means that \( 1 + \beta^T \alpha \) is an eigenvalue of the matrix \( I + \alpha \beta^T \). Notice that all the other eigenvalues are all 1. Thus, (18) holds.

Hence, we have
\[
|P_i^{-1}| = |P_{i+1}^{-1} - \varphi_i \varphi_i^T| = |P_{i+1}^{-1} - P_{i+1} \varphi_i \varphi_i^T| = |P_{i+1}^{-1} \varphi_i \varphi_i^T P_{i+1}| \cdot |I - P_{i+1} \varphi_i \varphi_i^T|,
\]
(19)
where \( \alpha = P_i \varphi_i \) and \( \beta = \varphi_i \) by using (18). Thus,
\[
\varphi_i^T P_{i+1} \varphi_i = \frac{|P_{i-1}^T - P_{i+1}^{-1}|}{|P_{i+1}^{-1}|}.
\]
Therefore,
\[
\sum_{i=0}^{n+1} \varphi_i^T P_{i+1} \varphi_i = \sum_{i=0}^{n+1} \frac{|P_{i-1}^T - P_{i+1}^{-1}|}{|P_{i+1}^{-1}|} = \sum_{i=0}^{n+1} \int_{P_{i+1}^{-1}}^{|P_{i+1}^T - P_{i-1}|} \frac{dx}{x} = \log |P_{i+1}^{-1}| + a_0 \log a_0,
\]
Hence, (16) follows.

The proof of (17) is similar to the counterpart of the proof of Theorem 4.1 in Chen & Guo (1991). Taking \( \alpha \in [1, \min(\beta/2, 2)] \) and applying Lemma 3.1 with \( M_i = P_i \varphi_i \), \( x_i = w_{i+1} - E[w_{i+1}|F_i] \), we obtain
\[
\sum_{i=0}^{n+1} \varphi_i^T P_{i+1} \varphi_i w_{i+1}^2 = \sum_{i=0}^{n+1} M_i x_i + \sum_{i=0}^{n+1} \varphi_i^T P_{i+1} E[w_{i+1}^2|F_i] = O \left( \sum_{i=0}^{n+1} M_i^\alpha \right)^{1/\alpha} \cdot \log^{1/\eta} \left( \sum_{i=0}^{n+1} M_i^\alpha \right) + a_0 \log a_0.
\]
(20)

**Proof of Theorem 3.1.** Denote \( \tilde{\theta}_n = \theta - \theta_n \). Obviously, (9a) can be written
\[
\tilde{\theta}_{n+1} = \tilde{\theta}_n + a_n P_n \varphi_n (w_{n+1} - \tilde{\theta}_n \varphi_n),
\]
(21)
Noticing \( P_n^{-1} \geq \lambda_{\min}(n) I \), we see that
\[
||\tilde{\theta}_{n+1}||^2 \leq \frac{1}{\lambda_{\min}(n)} \tilde{\theta}_n^T P_n^{-1} \tilde{\theta}_{n+1}.
\]
(22)
Hence, it is sufficient to analyse \( \tilde{\theta}_n^T P_n^{-1} \tilde{\theta}_{n+1} \).

By (21), we have
\[
\tilde{\theta}_{n+1}^T \varphi_n = \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n) P_n \varphi_n \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n)^2 (\varphi_n P_n \varphi_n)^2.
\]
(23)
Thus,
\[
\tilde{\theta}_n^T P_n^{-1} \tilde{\theta}_{n+1} = \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n) P_n \varphi_n \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n^2) \varphi_n P_n \varphi_n \tilde{\theta}_n^T \varphi_n
\]
+ \( + a_n^2 (w_{n+1} - \tilde{\theta}_n \varphi_n)^2 \varphi_n P_n \varphi_n \tilde{\theta}_n^T \varphi_n \)
\[
= \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n) P_n \varphi_n \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n)^2 \varphi_n P_n \varphi_n \tilde{\theta}_n^T \varphi_n
\]
\[
= \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n) P_n \varphi_n \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n)^2 \varphi_n P_n \varphi_n \tilde{\theta}_n^T \varphi_n
\]
\[
= \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n) P_n \varphi_n \tilde{\theta}_n^T \varphi_n + 2a_n (w_{n+1} - \tilde{\theta}_n \varphi_n)^2 \varphi_n P_n \varphi_n \tilde{\theta}_n^T \varphi_n
\]
\[
= \tilde{\theta}_n^T \varphi_n + 2a_n \tilde{\theta}_n^T \varphi_n P_n \varphi_n w_{n+1} - a_n (\tilde{\theta}_n \varphi_n)^2 + 2a_n \tilde{\theta}_n \varphi_n P_n \varphi_n w_{n+1}.
\]
(24)
\[ \tilde{\theta}^T_{n+1} P_{n+1}^{-1} \tilde{\theta}_{n+1} = \tilde{\theta}^T_0 P_0^{-1} \tilde{\theta}_0 + \sum_{i=0}^{n} [a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 - a_i (\tilde{\theta}^T_i \varphi_i)^2 + 2a_i \tilde{\theta}^T_i \varphi_i \tilde{w}_{i+1}] = O(1) + O \left( \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 \right) - \frac{1}{2} \sum_{i=0}^{n} a_i (\tilde{\theta}^T_i \varphi_i)^2 + 2 \sum_{i=0}^{n} a_i \tilde{\theta}^T_i \varphi_i \tilde{w}_{i+1} \]

\[ \leq O(1) + O \left( \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 \right) - \frac{1}{2} \sum_{i=0}^{n} a_i (\tilde{\theta}^T_i \varphi_i)^2 + 2 \sum_{i=0}^{n} a_i \delta_i^2 \]

\[ = O(1) + O \left( \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 \right) + O \left( \sum_{i=0}^{n} a_i \delta_i^2 \right) \] (25)

It is worth pointing out that we use Lemma 3.1 and the fact 
\[ \frac{1}{2} \delta_i^2 + 2\delta_i \leq \delta_i^2 \]
in the third step of (25). Notice the fact 0 \leq a_i \leq 1 and 0 \leq \varphi_i^T P_i \varphi_i < 1 (by (19)).
\[ \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 \leq 2 \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i (w_{i+1}^2 + \delta_i^2) \]
\[ \leq 2 \sum_{i=0}^{n} a_i \varphi_i^T P_i \varphi_i w_{i+1}^2 + 2 \sum_{i=0}^{n} \delta_i^2. \] (26)

Hence, (12) follows from (22), (25), (26) and Lemma 3.2 directly.

4. CONVERGENCE OF LEAST SQUARES ALGORITHM

In the previous section some upper bounds were established for the estimate error. We shall now apply these results more specifically to the identification case (6). Notice that the inputs of the model may be chosen freely in a pure identification case. Thus, we establish upper bound of estimate error expressed by \{ u_k \}, \{ \delta_k \} and \{ w_k \} in the following. So, the result here may be more applicable to open loop case. And then, the convergence of Figures 1 and 2 are explained.

Some ideas and techniques of Chen & Guo (1991); Guo (1994, 1995) are used in the proof for the result. Especially two key lemmas of Guo (1994) are presented.

Denote the minimum and maximum eigenvalue of a matrix \( A \) as \( \lambda_{\min}(A) \) and \( \lambda_{\max}(A) \) respectively and introduce the further assumptions

**H3.** \( A(z) \) is stable, and \( A(z) \) and \( B(z) \) are coprime;

**H4:** \( u_i \) is weakly persistently exciting of order \( p + q \):

\[ \lambda_{\min} \left( \sum_{i=0}^{n} U_i U_i^T \right) \geq cn \gamma \] for some \( c > 0 \), \( \gamma > 0 \), (27)

where \( U_i = [u_i \ldots u_{i-p+1}]^T \); This condition is similar to Definition 3.4.B of Goodwin & Sin (1984).

**H5:** For the same \( \gamma \) as in H3,

\[ \sum_{i=0}^{n} u_i \tilde{w}_i = o(n^\gamma); \quad \text{for} \ |i-j| \leq p + q \] (28)

Note that this condition means that the noise and the input must not be strongly correlated, thus essentially ruling out closed loop operation.

**H6:**

\[ \sum_{i=0}^{n} \delta_i^2 = O(n^{\gamma_1}), \quad \sum_{i=0}^{n} \tau_i^2 = O(n^{\gamma_2}) \quad \text{for some} \ \gamma_{1,2} > 0. \] (29)

We are now ready to formulate the main result:

**Theorem 4.1.** Assume that conditions H1 – H6 hold. Then the LS algorithm (9) for model (6) has the following estimation error bound:

\[ ||\theta_{n+1} - \theta||^2 = O \left( \frac{\log n \cdot \rho_{\beta}(n)}{n^\gamma} + \sum_{i=0}^{n} \delta_i^2 \right) \] a.s., (30)

where \( \gamma \) is given in H3 and H4 and \( \rho_{\beta}(\cdot) \) is defined by (11) for a \( \beta \) for which H1 holds.

Obviously, \( \theta_n \rightarrow s \) a.s. \( n \rightarrow \infty \) if \( \sum_{i=0}^{n} \delta_i^2 = o(n^{\gamma}). \)

We list Theorem 34.1.1 (Schur’s inequality) of Prasolov (1994) as a lemma as follows.

**Lemma 4.1.** Let \( \lambda_1, \ldots, \lambda_n \) be eigenvalues of \( A = (a_{ij})_{n \times n}. \) Then

\[ \sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i=1}^{n} |a_{ij}|^2 \]

and the equality is attained if and only if \( A \) is a normal matrix.

The following two lemmas are similar to Lemma 2.3 and 2.2 in Guo (1994), respectively. We omit the proofs here. See Hu & Ljung (2007) for some variants of the proofs, that perhaps are simpler.

**Lemma 4.2.** Let \( \{X_k \in \mathbb{R}^d, \ k = 0, 1, \ldots \} \) be a vector sequence where \( d > 0, \) and

\[ F(z) = f_0 + f_1 z + \cdots + f_n z^n \]

be a polynomial with \( M_F \triangleq \left( \sum_{i=0}^{n} |f_i| \right)^2 > 0. \) Set \( \bar{X}_k = F(z) X_k. \) Then,

\[ \lambda_{\min} \left( \sum_{k=0}^{n} X_k X_k^T \right) \geq \frac{1}{M_F} \lambda_{\min} \left( \sum_{k=0}^{n} \bar{X}_k \bar{X}_k^T \right) \quad \forall n \geq 0. \] (31)

**Lemma 4.3.** Let \( G(z) = g_0 + g_1 z + \cdots + g_n z^n, \) \( H(z) = h_0 + \cdots + h_n z^n \) be two coprime polynomials. For any integers \( m \geq 0, \ n \geq 0, \) and any sequence \( \{ \xi_k \}, \) define \( Y_k = [G(z), zG(z), \ldots, z^m G(z)], \)

\[ H(z), zH(z), \ldots, z^n H(z)]^T x_k \]

where \( m < n_h \) and \( n < n_g. \) Then,

\[ \lambda_{\min} \left( \sum_{k=0}^{n} Y_k Y_k^T \right) \geq M_F \lambda_{\min} \left( \sum_{k=0}^{n} X_k X_k^T \right) \quad \forall k \geq 1, \] (32)
Thus, \( \sum \) This can be proved as follows:

\[
\sum_{k=0}^{n} (A^{-1}(z)\xi_k)^2 = \sum_{k=0}^{n} \left( \sum_{i=0}^{k} a_i \xi_{k-i} \right)^2 = \sum_{k=0}^{n} \left( \sum_{i=0}^{k} (i+1)a_i \frac{1}{(i+1)\xi_{k-i}} \right)^2 \leq \sum_{k=0}^{n} \sum_{k=0}^{n} \left( i+1 \right) a_i^2 \sum_{i=0}^{k} \frac{1}{(i+1)^2} \xi_{k-i}^2 = O \left( \sum_{k=0}^{n} \frac{1}{(i+1)^2} \xi_{k-i}^2 \right) = O \left( \sum_{k=0}^{n} \frac{1}{(i+1)^2} \xi_{k-i}^2 \right) = O \left( \sum_{k=0}^{n} \frac{1}{(k-j+1)^2} \xi_{j-k}^2 \right) = O \left( \sum_{j=0}^{n} \frac{1}{(k-j+1)^2} \xi_{j-k}^2 \right) = O \left( \sum_{k=0}^{n} \xi_k^2 \right).\]

Hence, the assertion follows.

**Proof of Theorem 4.1.**

In view of Corollary 3.1, we need only analyse \( \lambda_{\min} \left( \sum_{i=0}^{k} \varphi_i \varphi_i^T \right) \) and \( \tau_n \) respectively.

By the definition of \( \varphi_i \) and (6), it is clear that

\[
\psi_i = A(z) \varphi_i = \Gamma(zB(z), A(z); p - 1, q - 1)U_i + \hat{W}_i = \psi_i + \hat{W}_i,
\]

where \( \Gamma \) is defined by (34) and the \((p + q) \times 1\) vector \( \hat{W}_i \triangleq [\bar{w}_1 \cdots \bar{w}_{p+1}]^T \). By Lemma 4.2 we have

\[
\lambda_{\min} \left( \sum_{i=0}^{n} \varphi_i \varphi_i^T \right) \geq \frac{1}{M_A} \min \left( \sum_{i=0}^{n} \psi_i \psi_i^T \right). \quad (37)
\]

Since \( A(z) \) has no zero root, by assumption \( zB(z) \) and \( A(z) \) are also coprime. Hence, by Lemma 4.3 we have

\[
\lambda_{\min} \left( \sum_{i=0}^{n} \psi_i \psi_i^T \right) \geq M_{\Gamma} \lambda_{\min} \left( \sum_{i=0}^{n} U_i U_i^T \right). \quad (38)
\]

On the other side, by (36) and (38), clearly,

\[
\sum_{i=0}^{n} \psi_i \psi_i^T = \sum_{i=0}^{n} \left( \psi_i^T \psi_i^T + \psi_i^T \hat{W}_i + \hat{W}_i^T \psi_i^T + \hat{W}_i \psi_i^T \right) \geq \sum_{i=0}^{n} (\psi_i^T \hat{W}_i^T + \hat{W}_i^T \psi_i^T) \geq c M_{\Gamma} n^2 I + \sum_{i=0}^{n} (\psi_i^T \hat{W}_i^T + \hat{W}_i^T \psi_i^T). \quad (39)
\]

In view of (28), clearly, each element of the matrix

\[
\sum_{i=0}^{n} (\psi_i^T \hat{W}_i^T + \hat{W}_i^T \psi_i^T)
\]

is \( o(n^\gamma) \) as \( n \) tends to infinity. By Schur’s inequality (Lemma 4.1), we have

\[
\lambda_{\max} \left( \sum_{i=0}^{n} (\psi_i^T \hat{W}_i^T + \hat{W}_i^T \psi_i^T) \right) = o(n^\gamma). \quad (40)
\]

Hence, (39) turns to be

\[
\lambda_{\min} \left( \sum_{i=0}^{n} \psi_i \psi_i^T \right) \geq c_1 n^\gamma \quad (41)
\]

with certain \( c_1 > 0 \) for sufficient large \( n \). Hence, by (37) and (41) we have

\[
\lambda_{\min} \left( \sum_{i=0}^{n} \varphi_i \varphi_i^T \right) \geq \frac{c_1}{M_A} n^\gamma. \quad (42)
\]

Taking \( \alpha = 1 \), applying Lemma 3.1 with \( M_i = E[w_i^2 | F_{i-1}] \) and \( x_i = \frac{w_i^2}{E[w_i^2]} \), we have

\[
\sum_{i=0}^{n} w_i^2 = \sum_{i=0}^{n} M_i x_i + \sum_{i=0}^{n} E[w_i^2 | F_{i-1}] = \sum_{i=0}^{n} E[w_i^2 | F_{i-1}] \cdot \log \left( \sum_{i=0}^{n} E[w_i^2 | F_{i-1}] \right) = O \left( \sum_{i=0}^{n} \log n \right). \quad (43)
\]

Thus, by (6) (29) (43) and Lemma 4.4, we have

\[
\sum_{i=0}^{n} y_i^2 = O \left( n^{\gamma_2} \right) + O \left( n^{\gamma_1} \right) + O \left( n \log n \right) \quad (44)
\]

Hence, by (13), we have
\[ r_n = \alpha_0 + \sum_{i=0}^{n} \sum_{j=0}^{p-1} y_{i-j}^2 + \sum_{i=0}^{n} \sum_{j=0}^{q-1} u_{i-j}^2 \leq \alpha_0 + p \sum_{i=0}^{n} y_i^2 + q \sum_{i=0}^{n} u_i^2 = O(n^{\nu}) + O(n^{\gamma}) + O(n \log n). \tag{45} \]

Therefore, by (42) (45) and Corollary 3.1, the assertion (30) holds.

We are now in a position to verify the convergence in Figures 1 and 2. For convenience, we list a Central Limit Theorem result for martingale difference sequence (Corollary 2.6 of Chen & Guo (1991)) as a lemma here.

**Lemma 4.5.** Let \( \{x_i, F_i\} \) be a martingale difference sequence. If either \( \sup_{p} E|x_i|^p |F_{i-1}| < \infty \) a.s. or \( \sup_{p} E|x_i|^p < \infty \) for some \( p \in [1,2] \), then as \( n \to \infty \) for any \( q > 1 \)

\[
\frac{1}{n^{q/2}} \sum_{i=1}^{n} x_i \to 0 \quad \text{a.s.} \tag{46}
\]

**Remark 4.1.** Consider a special case \( p = q > 1 \) in Lemma 4.5, we have \( \frac{1}{n} \sum_{i=1}^{n} x_i \to 0 \) under assumption \( \sup_{p} E|x_i|^{1+\nu} |F_{i-1}| < \infty \) or \( \sup_{p} E|x_i|^{1+\nu} < \infty \) with \( \nu > 0 \).

For an adaptive sequence \( \{x_i, F_i\} \) satisfying

\[
\sup_{p} E|x_i|^{1+\nu} |F_{i-1}| < \infty
\]

or \( \sup_{p} E|x_i|^{1+\nu} < \infty \) with \( \nu > 0 \), we have

\[
\sum_{i=1}^{n} x_i = O \left( \sum_{i=1}^{n} E|x_i| F_{i-1} \right) + o(n).
\]

Clearly, for both cases in figures 1 and 2, the conditions H3 and H6 of Theorem 4.1 are satisfied.

In Figure 1, by the help of Remark 4.1 we have

\[ \lambda_{\min} \left( \sum_{i=0}^{n} U_i U_i^T \right) \geq cn \text{ for some } c > 0, \]

and

\[ \sum_{i=0}^{n} u_i w_j = o(n) \]

in view of the fact \( \sum_{i=1}^{n} \delta_i^2 = O(\sqrt{n}) \) and the independence of \( \{u_i\} \) and \( \{w_j\} \) (open loop). Thus,

\[ \|\theta_{n+1} - \theta\|^2 = O \left( \frac{1}{\sqrt{n}} \right) \text{ a.s.} \]

In Figure 2, by the help of Remark 4.1 we have

\[ \lambda_{\min} \left( \sum_{i=0}^{n} U_i U_i^T \right) \geq cn^2 \text{ for some } c > 0, \]

and

\[ \sum_{i=0}^{n} u_i w_j = o(n^2) \]

by the help of the fact \( \sum_{i=1}^{n} \delta_i^2 = n \) and the independence of \( \{u_i\} \) and \( \{w_j\} \) (open loop). Thus,

\[ \|\theta_{n+1} - \theta\|^2 = O \left( \frac{1}{n} \right) \text{ a.s.} \]

5. CONCLUSIONS

Some new convergence issues of LS with more general noise or disturbance compared to existing references have been studied in this paper. First, a general result, Theorem 3.1 including some existing classic results as special cases is established. Next, a useful variant (especially for open loop) is given as Theorem 4.1. The results make it possible to find out how much unstructured disturbances can be present without affecting the limit estimates. The essential answer is that the norm of the unstructured disturbance must grow slower than the smallest eigenvalue of the regression matrix. The results can also be used to analyze the properties of the LSE when applied to time-varying systems, that vary “around” a constant system, see Hu & Ljung (2007).

Some techniques and ideas of Chen & Deniau (1994); Chen & Guo (1991); Guo (1994, 1995); Lai & Wei (1982) were of key importance for the proof. The extensions compared to Chen & Deniau (1994) are essentially that an input signal is introduced, thus it becomes important to address the growth of the smallest eigenvalue of the regression matrix.

For further study, it is desirable to generalize the results to the closed loop case and the colored noise case.

REFERENCES


New Convergence Results for Least Squares Identification Algorithm

Abstract
The basic least squares method for identifying linear systems has been extensively studied. Conditions for convergence involve issues about noise assumptions and behavior of the sample covariance matrix of the regressors. Lai and Wei proved in 1982 convergence for essentially minimal conditions on the regression matrix: All eigenvalues must tend to infinity, and the logarithm of the largest eigenvalue must not tend to infinity faster than the smallest eigenvalue. In this contribution we revisit this classical result with respect to assumptions on the noise: How much unstructured disturbances can be allowed without affecting the convergence? The answer is that the norm of these disturbances must tend to infinity slower than the smallest eigenvalue of the regression matrix.