Identification and System Parameter Estimation

Part 2

IV IFAC Symposium

Tbilisi USSR
21-27 September
1976

Preprints

ON MARTINGALES AND LEAST SQUARES LINEAR SYSTEM IDENTIFICATION

J.B. Moore (Australia)
J.M.C. Clark (U.K.)
B.D.O. Anderson (Australia)

In the pioneering work of Mann and Wald [1] on the subject of linear system identification, and in much subsequent work, the assumption of ergodic signals is crucial to the analysis. More recently, Ljung et al [2], seek to relax the requirement of ergodicity in the consistency proofs of a wide class of identification algorithms so that the case of nonstationary noise can be encompassed. Perhaps the key contribution in [2] is the demonstration of the relevance of the martingale convergence theorem and the Kronecker lemma to this topic. Using the martingale theory, almost sure convergence is studied rather than weaker forms as in earlier work.

Two aspects of the work of [2] are as follows. First, the results rely on scalar martingale convergence theorems and the Kronecker lemma in scalar form. The second aspect which is very much tied up with the first aspect is that the signal model is required to be asymptotically stable.

In this paper, the Toeplitz lemma and standard martingale convergence theorems in vector form are applied to yield least squares convergence results with weaker conditions than those of [2]. In particular unstable signal models are permitted. Also in the paper it is observed that the least squares estimation error is in fact a vector martingale normalized by its process variance. The question arises as to convergence conditions for such normalized martingales.

In the most significant theorem of the paper, the convergence properties of a scalar martingale normalized by its scalar process variance [3-6] are generalized to the case of a vector martingale normalized by its matrix process variance. Certain side conditions arise in the vector case which are absent or trivially satisfied in the scalar case. These results are derived using a matrix Toeplitz lemma and standard martingale convergence theorems. As a byproduct, a novel matrix Kronecker lemma is derived with side conditions which are trivially satisfied in the scalar case. The matrix Kronecker lemma results constitute a mild generalization of results given in [7].

CONVERGENCE OF A VECTOR MARTINGALE NORMALIZED BY ITS PROCESS VARIANCE. In this section, the convergence properties of a martingale normalized by its process variance [3-6] are generalized to the matrix case. The main result, Theorem 2.1, will be used to establish convergence of the identification procedure given in later sections.

Definition (2.1) Let X_k be an n-vector martingale with respect to a σ-algebra F_k, such that $X_0 = 0$, $X_k \in L^2$ for all k. Bet B_k denote its n×n matrix process variance, defined from

$$B_0 = 0, \; B_j = \sum_{k=1}^{j} E\{[X_k - X_{k-1}][X_k - X_{k-1}]' | F_{k-1}\} \qquad (2.1)$$

and assume that B_j^{-1} exists for j=1, and therefore for all $j \geq 1$. Let

$$Z_{mk} = \begin{cases} \sum_{j=k}^{m} B_j^{-1}(X_j - X_{j-1}) & m \geq k \\ 0 & m < k \end{cases}$$

Lemma 2.1 (i) With the above definition (2.1), for each fixed k, Z_{mk} is an F_m martingale closed in L^2 and converges almost surely and in the norm of L^2. (ii) Define the almost sure limit $Z_k = \lim_{m\to\infty} Z_{mk}$. Then $E[Z_k Z_k' | F_{k-1}]$

$\leq 2B_k^{-1}$ and $\lim_{k\to\infty} Z_k = 0$ almost surely. (iii) If for p×n matrices C_k, possibly ω-dependent and then F_{k-1} measurable $||\sum_{k=1}^{\infty} C_k B_k^{-1} C_k'|| < M$ for some constant M, then $\lim_{k\to\infty} C_k Z_k = 0$ almost surely.

Proof: (See [11])

Theorem 2.1 With the definitions (2.1) and notation of Lemma 2.1.

i) if $\lim_{m\to\infty} B_m$ exists and is finite, then $\lim_{m\to\infty} B_m^{-1} X_m$ exists and is finite almost surely

ii) If $\lim_{m\to\infty} B_m$ exists and is finite, then $\lim_{m\to\infty} (\operatorname{tr} B_m)^{-1} X_m = 0$ almost surely

iii) If $\lim_{m\to\infty} B_m^{-1} = 0$ and if there exists a sequence of non-singular matrices C_k such that $\lim_{m\to\infty} C_m Z_m = 0$ almost surely and

$$\limsup_{m\to\infty} \sum_{k=1}^{m} ||B_m^{-1}(B_k - B_{k-1})C_k^{-1}|| < \infty \tag{2.2}$$

one has

$$\lim_{m\to\infty} B_m^{-1} X_m = 0 \quad \text{almost surely} \tag{2.3}$$

Remarks: 1. Condition (2.2) is implied by the much stronger condition

$$\limsup_{m\to\infty} \frac{\lambda_{max}(B_m)}{\lambda_{min}(B_m)} < \infty \tag{2.4}$$

2. Approaches to proving Theorem 2.1 might reasonably be based on extensions to the matrix case of proofs applicable to the scalar case when (2.2) is trivially satisfied [4, 5, 6]. The Kronecker Lemma approach [6] is applied to the matrix problem below, since it yields the most general matrix results but first a matrix Toeplitz lemma must be studied.

Toeplitz Lemma (Matrix Case) Let A_{mk} be real p×q matrices and r_k real q-vectors defined for all positive integer m, k and satisfying the following conditions:

$$\lim_{m\to\infty} A_{mk} = 0 \text{ for each fixed } k, \quad \limsup_{m\to\infty} \sum_{k=1}^{\infty} ||A_{mk}|| < \infty \tag{2.5}$$

$$\lim_{k\to\infty} r_k = r < \infty \tag{2.6}$$

Then

$$\lim_{m\to\infty} \sum_{k=1}^{\infty} A_{mk}(r_k - r) = 0 \tag{2.7}$$

Proof: The proof is similar to that for scalar case [10].

Now we turn to the Kronecker lemma. For a scalar version, see e.g. [6, 8]; for some remarks on a matrix version, see [7].

Kronecker's Lemma (Matrix case) Let x_k, $k=1, 2, \ldots$ be a sequence of real *n*-vectors, A_k a sequence of real symmetric nonsingular $n \times n$ matrices with $A_{k+1} \geq A_k > 0$, and C_k a sequence of nonsingular matrices. Suppose that $\bar{r}_k = \sum_{j=k}^{\infty} A_j^{-1} x_j$ exists (and is finite). Then

i) If $A_\infty = \lim_{m\to\infty} A_m$ exists and is finite, $\lim_{m\to\infty} A_m^{-1} \sum_{k=1}^{m} x_k$ exists and is finite.

ii) If $\lim_{m\to\infty} (\operatorname{tr} A_m)^{-1} = 0$, then $\lim_{m\to\infty} (\operatorname{tr} A_m)^{-1} \sum_{k=1}^{m} x_k = 0$

iii) If $\lim_{m\to\infty} A_m^{-1} = 0$, $\lim_{m\to\infty} C_m \bar{r}_m = 0$, and

$$\limsup_{m\to\infty} \sum_{k=1}^{m} ||A_m^{-1}(A_k - A_{k-1})C_k^{-1}|| < \infty, \quad \text{then}$$

$\lim_{m\to\infty} A_m^{-1} \sum_{k=1}^{m} x_k = 0$

Proof: Follows from Toeplitz lemma as discussed in [11].

To prove Theorem 2.1, apply the Kronecker lemma with the identifications $A_m \sim B_m$, $\sum_{k=1}^{m} x_k \sim X_m$ and $C_k \sim C_k$. Notice that the summation $\bar{r}_k = \sum_{j=k}^{\infty} A_j^{-1} x_j$ is for this case $Z_k = \sum_{j=k}^{\infty} B_j^{-1}(X_j - X_{j-1})$ which exists by virtue of Lemma 2.1. The conditions and results of the Kronecker lemma parts (i), (ii) and (iii), under the identification above are the conditions and results of Theorem 2.1 parts (i), (ii) and (iii) respectively.

LEAST SQUARES IDENTIFICATION OF LINEAR SYSTEMS. The parameter estimation error $\tilde{\theta}_m$ in least squares identification of linear systems can be expressed as

$$\tilde{\theta}_m = P_{m+1} \sum_{k=1}^{m} x_k v_{k+1}, \quad P_{m+1}^{-1} = \sum_{k=1}^{m} x_k x_k' \tag{3.1}$$

with x_m measurable and v_k white noise of zero mean and unit variance. The following lemma is readily established.

Lemma 3.1 $\tilde{\theta}_m$ is an F_{m+1} martingale

$X_m = \sum_{k=1}^{m} x_k v_{k+1}$ normalised by its process variance

$B_m = \sum_{k=1}^{m} x_k x_k' = P_{m+1}^{-1}$. That is,

$$\tilde{\theta}_m = B_m^{-1} X_m \tag{3.2}$$

Application of the martingale convergence theorem 2.1 (see part (iii)) yields that $\lim_{m\to\infty} \tilde{\theta}_m = 0$ almost surely if $\lim_{m\to\infty} P_{m+1} = 0$, $\lim_{m\to\infty} C_m Z_m = 0$ and $\lim_{m\to\infty} \sup \sum_{k=1}^{m} ||P_{m+1} x_k x_k' C_k^{-1}|| < \infty$ for some sequence of nonsingular C_k. Recall from Lemma 2.1 that $\lim_{m\to\infty} C_m Z_m = 0$ if $C_k = I$ for all k, or if C_k is F_k-measurable and such that $||\sum_{k=1}^{\infty} C_k P_{k+1} C_k'|| < M < \infty$.

The choice $C_k = k^{-\frac{1}{2}-\epsilon} P_{k+1}^{-\frac{1}{2}}$ for any $\epsilon > 0$ guarantees that $||\sum_{k=1}^{\infty} C_k P_{k+1} C_k'|| < M < \infty$. Accordingly, we have the following convergence result for $\tilde{\theta}_m$:

Theorem 3.1 Sufficient conditions on x_k for the least squares estimate $\hat{\theta}_m$ given from (3.2) to converge almost surely to the true parameter are that $\lim_{m\to\infty} P_{m+1} = 0$ and that either

$$\lim_{m\to\infty} \sup \sum_{k=1}^{m} ||P_{m+1} x_k x_k'|| < \infty \tag{3.3}$$

or, for some $\epsilon > 0$

$$\lim_{m\to\infty} \sup \sum_{k=1}^{m} [k^{\frac{1}{2}+\epsilon} ||P_{m+1} x_k x_k P_{k+1}^{\frac{1}{2}}||] < \infty \tag{3.4}$$

1 Stable Systems Let us note how Theorem 3.1 applies to the identifications of stable systems. It is conventional, see e.g. [2], to make the assumptions

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} (y_k + u_k) \tag{3.5}$$

$$\{ \theta \mid \lim_{m\to\infty} \inf \frac{1}{m} \sum_{k=1}^{m} (\theta' x_k)^2 = 0\} = \{0\} \tag{3.6}$$

as well as assumptions involving the independence of the u_k and v_k sequences.

Consider just the two assumptions (3.5) and (3.6) above. Together, (3.1) and (3.5) imply that trace P_{m+1}^{-1} grows no faster than linearly. Further (3.1) and (3.6) together show that $\lambda_{min}(P_{m+1}^{-1})$ grows linearly. Hence

$$\lim_{m\to\infty} P_{m+1} = 0, \quad \limsup_{m\to\infty} \frac{\lambda_{max}(P_m^{-1})}{\lambda_{min}(P_m^{-1})} < \infty$$

It is not difficult to see that (3.3) is satisfied, and identification is therefore achieved.

<u>2 Unstable Systems without inputs</u> Suppose now that the system has all unstable modes and that inputs are absent from (3.1); we expect then that $P_{m+1}^{-1} \to 0$ with the rate of convergence being exponentially fast.

In case there is one unstable mode only, i.e. P_m is a 1×1 matrix, then (3.3) holds and convergence follows. However, when there is more than one unstable mode, simulations have shown that while (3.3) does not hold, the following inequality, as far as can be ascertained, does hold:

$$\limsup_{m\to\infty} \sum_{k=1}^{m} [k||P_{m+1} \; x_k||] < \infty \tag{3.7}$$

Now this inequality implies (3.4) for $\varepsilon = \frac{1}{2}$, since, as is easily shown,

$$||P_{k+1}^{\frac{1}{2}} \; x_k|| \leq 1$$

Of course, in no sense have we proved that convergence of the least squares algorithm is bound to occur when the system has all unstable modes and inputs are absent. On the other hand, we may have reached a significant half-way house on this route.

<u>3 Mixed-mode case</u> We remark that simulations of unstable systems with inputs and with some stable modes have demonstrated convergence. Moreover, convergence associated with unstable modes is exponentially fast. Conditions for this case have been derived which reduce to (3.3) [(3.4)] as the number of stable [unstable] modes equals the total number of modes. These rely on introducing a transformation T_i such that $T_i' P_i T_i$ is diagonal and $T_i' T_i = I$.

What is really desired however is a simple condition covering the mixed-mode case. In light of the fact that (3.3) suffices for the purely stable case, and (3.4) for the purely unstable case, we conjecture that the condition

$\limsup_{m\to\infty} \sum_{k=1}^{m} [k^{\varepsilon}||P_{m+1}x_k||] < \infty$ for $0 < \varepsilon < \frac{1}{2}$ will establish convergence. The natural question arises as to whether this condition will be satisfied in practice; simulations suggest so, and in the following remark, we argue that it is satisfied when the noise is bounded or gaussian.

<u>4 Bounded or Gaussian Noise</u> The estimation error θ given in (3.5) can be rewritten as

$$\tilde{\theta} = \sum_{k=1}^{m} (P_{m+1}x_k C_k^{-1})(C_k\ v_{k+1})$$

for some scalar C_k such that $C_{k+1} \geq C_k > 0$ for all k. Identifying (for $m \geq k$) $P_{m+1}x_k C_k^{-1}$ and $C_k v_{k+1}$ with A_{mk} and r_k of the Toeplitz lemma, the Toeplitz lemma may be applied to yield that $\tilde{\theta}$ converges almost surely to zero if

$$\lim_{m\to\infty} P_{m+1} = 0,\ \lim_{m\to\infty} C_m v_{m+1} = 0 \text{ almost surely}$$

$$\limsup_{m\to\infty} \sum_{k=1}^{m} ||P_{m+1}x_m C_k^{-1}|| < \infty$$

As for our more general results, with $\sum_{k=1}^{\infty} C_k{}^2 < \infty$, $\lim C_m v_{m+1} = 0$ almost surely since $E[v_k{}^2] = 1$. In the event that $||v_k|| < M$ for some fixed M and all k, then $C_k \to 0$ as $k \to \infty$ *implies* $\lim_{m\to\infty} C_m v_{m+1} = 0$. For the case when v_k is Gaussian, a condition $\sum_{k=1}^{\infty} C_k{}^{2i} < \infty$ for some integer $i \geq 1$ implies using the Borel-Cantelli lemma and the Markov inequality that $\lim_{m\to\infty} C_m v_{m+1} = 0$. In both the bounded and gaussian noise case, $C_k = k^{-\varepsilon}$ for $0 < \varepsilon < \frac{1}{2}$ ensures $\lim_{m\to\infty} C_m v_{m+1}$.

<u>5 Alternative Derivation</u> The proof of Theorem 3.1 depended on the crucial observation of Lemma 3.1 that $\tilde{\theta}_m$ is a martingale normalised by its process variance. This allowed application of ideas developed by others in studying processes of this type. With hindsight however, one can obtain a shorter proof of Theorem 3.1 *by eliminating this observation.* The proof still relies on using a martingale convergence result like that of

Lemma 2.1, as well as the Toeplitz lemma and the following representation of $\tilde{\theta}_m$:

Lemma 3.2 With (A1) satisfied,

$$\tilde{\theta}_m = \sum_{k=1}^{\infty} A_{mk} r_{mk} + P_{m+1} P_1^{-1}(S_m - S_0)$$

$$S_m = \sum_{k=1}^{m} P_{j+1} x_j v_{j=1} \qquad r_{mk} = C_m \sum_{j=k}^{m} P_{j+1} x_j v_{j+1}$$

$$A_{mk} = P_{m+1} x_k x_k' C_k^{-1} \quad k \leq m$$
$$= \qquad\qquad k > m$$

Moreover r_{mk} is an $\mathcal{F}_m$ martingale with $E[r_{mk} r_{mk}'] \leq I$.

6 Other models The ideas used in proving the Theorem 3.1 carry over to the more sophisticated models which arise in considering the regulator problems associated with (3.1). As discussed in [2], a suitable model is for some integer T

$$y_{k+T+1} = \theta' x_k + \beta u_k + \varepsilon_{k+T+1}$$

where ε_k is a moving average of v_k, $v_{k-1} \dots v_{k-T}$. That is, $\varepsilon_{k+T+1} = \sum_{i=0}^{T} \gamma_k^{(i)} v_{k+i+1}$. Then Theorem 3.1 goes through with the following additional requirements. For each i, one has $|\gamma_k^{(i)}| < \bar{\gamma} < \infty$ for all k; with x_k replaced by $\gamma_k^{(i)} x_k$ in the definition of P_{m+1}, $\lim_{m\to\infty} P_{m+1} = 0$ and (3.3) or (3.4) holds with the same replacement of x_k.

CONCLUSIONS. In this paper, we have stated a convergence theorem associated with vector martingales, and applied it to establishing the consistency of least squares parameter estimates of a signal model of known structure. The class of signal model permits controls to be derived by feedback, nonstationarity in the noise statistics, and even unstable signal models; no untoward assumptions seem to be required.

A number of problems arise out of the material of the paper. In relation to the martingale convergence theorem, one can ask if, as for the Kronecker Lemma, one can obtain a counterexample to the convergence result in case the assumption $\limsup_{m\to\infty} \sum_{k=1}^{m} \| B_m^{-1}(B_k - B_{k-1}) C_k^{-1} \| < \infty$ fails. Second, it is evident that the convergence results for unstable plants and mixed-mode plants need improvement, in terms of delineating convergence conditions. linking these with readily discernible properties. of the plant and its input, and pinning down a convergence rate.

Third, we might ask what sort of convergence of $\hat{\theta}_m$ is encountered in case the real plant of which (3.1) is the model has fewer poles and zeroes than that of the model; the work of [2, 3] would suggest that $\hat{\theta}_m$ would converge to a set with the property that all members of the set defined the same transfer function. Finally one could consider in what way the multivariable plant problem might best be tackled.

REFERENCES

[1] H.B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, Vol.11, 1943, pp. 173-220.

[2] L. Ljung and B. Wittenmark, "Asymptotic Properties of Self-Tuning Regulators", Report 7404, Div. of Auto. Control, Lund Inst. of Technology, February, 1974.

[3] L. Ljung, "On Consistency for Prediction Error Identification Methods", Report 7405, Division of Auto. Control, Lund. Inst. of Technology, March, 1974.

[4] L.E. Dubins and D.A. Freedman, "A Sharper Form of the Borel-Cantelli Lemma and the Strong Law", Annals of Math. Stats., Vol. 36, 1965, pp. 800-807.

[5] P. Meyer, Martingales and Stochastic Integrals - I, Lecture Notes in Maths. Series, No. 284, Springer Vertag, New York, 1972.

[6] J. Neveu, Mathematical Foundations of the Calculus of Probability, Holden-Day Inc., San Francisco, 1976.

[7] B.D.O. Anderson and J.B. Moore, "A Matric Kronecker Lemma", Linear Algebra and Its Applications, 1976, to appear.

[8] K.L. Chung, A Course in Probability Theory, Harcourt Brace and World Inc., New York. 1968.

[9] K.J. Astrom and B. Wittenmark, "On Self-Tuning Regulators", Automatica, Vol. 9, No.2, March 1973. pp. 185-200.

[10] W.F. Short, Almost Sure Convergence, Academic Press, 1974.

[11] J.B. Moore and B.D. Anderson, "On Martingales and Least Squares Linear System Identification", Technical Report, No. EE7522, University of Newcastle.