On estimating the parameters of a dynamic model from noisy input and output measurements

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 88-02

On estimating the parameters of a dynamic model from noisy input and output measurements

by
J.M. ten Vregelaar

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Eindhoven, January 1988
The Netherlands

On estimating the parameters of a dynamic model from noisy input and output measurements

Jan ten Vregelaar
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

ABSTRACT

In this paper an algorithm is provided to compute least squares estimates for the parameters of a dynamic model from noisy measurements of inputs and outputs. Furthermore, we prove the consistency property for the corresponding estimators under some assumptions.

Keywords: Parameter estimation, Dynamic model, Noisy measurements, Least squares estimation, Consistency.

January 1988

1. Introduction

In [5], Eising et al discuss an identification method for a prototype situation of an ARMA-model with noisy measurements on inputs and outputs. In order to obtain estimates for the unknown parameters, they propose to use only object function evaluations which are calculated by inverting some matrix with a lot of "structure". Furthermore it is stated in [5] that the corresponding estimators should be consistent. To our knowledge no proof of consistency has appeared up to now.

For a much more general situation of the model (MIMO system, possible different autoregressive and moving average orders, not assuming zero initial conditions) we present in this paper a new algorithm to calculate both object function and its first derivative evaluations. Furthermore, estimators as minimizing solutions are proved to be consistent under some mild assumptions.

We consider the ARMA model

$$\eta_t = \sum_{i=1}^{p} \alpha_i \eta_{t-i} + \sum_{j=0}^{q} \beta_j \xi_{t-j} \ , \quad t = m+1, m+2, \cdots \ , \tag{1.1}$$

with known orders p and q and $m := \max(p,q)$. The inputs ξ_t and outputs η_t, which may be vectors say of length r and s respectively (MIMO system when $r,s > 1$), are supposed to be measured with noise

$$\begin{aligned} y_t &= \eta_t + \varepsilon_t \ , \\ & \qquad\qquad t = 1,2,\cdots,m+N \ . \\ x_t &= \xi_t + \delta_t \ , \end{aligned} \tag{1.2}$$

The problem is to estimate the unknown parameter matrix (with size $s \times [ps + (q+1)r]$)

$$\theta = [\alpha_1 \, \alpha_2 \cdots \alpha_p \, \beta_0 \, \beta_1 \cdots \beta_q] \tag{1.3}$$

from the set of data $\{(x_1,y_1), (x_2,y_2), \cdots, (x_{m+N}, y_{m+N})\}$.

A least squares estimate for θ is obtained as solution of the minimization problem

$$\begin{aligned} &\min \sum_{t=1}^{m+N} \{\|y_t - \eta_t\|^2 + \|x_t - \xi_t\|^2\} \\ &\text{with respect to } \theta, \xi_1, \cdots, \xi_{m+N}, \eta_1, \cdots, \eta_{m+N} \ , \\ &\text{subject to the model equation (1.1) for } t = m+1, \cdots, m+N. \end{aligned} \tag{1.4}$$

The norm $\|\cdot\|$ is the Euclidean vector norm. In order to use a compact notation we rewrite (1.1) for $t = m+1, \cdots, m+N$ as

$$D(\theta)\, \zeta = 0 \ , \tag{1.5a}$$

where (I denoting the identity matrix)

$$D = \left[\begin{array}{cccccc|cccccc} -I & \alpha_1 & \cdots & \alpha_m & & & \beta_0 & \cdots & \beta_m & & & \\ & \ddots & & & \ddots & & & \ddots & & & \ddots & \\ & & -I & \alpha_1 & \cdots & \alpha_m & & & \beta_0 & \cdots & \beta_m & \end{array}\right] \tag{1.5b}$$

(has size $sN \times (s+r)(m+N)$, consists of two block-Toeplitz submatrices; empty space represents zero elements whereas we define $\alpha_k := 0$ for $k > p$, $\beta_k := 0$ for $k > q$), and

$$\zeta = \begin{bmatrix} \eta_{m+N} \\ \vdots \\ \eta_1 \\ \xi_{m+N} \\ \vdots \\ \xi_1 \end{bmatrix} . \tag{1.5c}$$

Let z and e denote the corresponding vectors of observations and noises such that

$$z = \zeta + e \quad , \tag{1.6}$$

is a shorthand notation for (1.2).

Then the minimization problem (1.4) can be written as

$$\min_{\theta,\zeta} \|z - \zeta\|^2 \text{ subject to } D\zeta = 0 \quad . \tag{1.7}$$

For the stationary point $(\tilde{\zeta}, \tilde{\lambda})$ of the Lagrangian

$$L(\zeta,\lambda) = \frac{1}{2}(z - \zeta)^T(z - \zeta) + \lambda^T D\zeta$$

we obtain

$$\tilde{\zeta} + D^T\tilde{\lambda} = z$$

$$D\tilde{\zeta} = 0 \quad ,$$

hence

$$\tilde{\lambda} = (D^+)^T z \tag{1.8}$$

$$\tilde{\zeta} = (I - D^+D)z \quad ,$$

with $D^+ = D^T(DD^T)^{-1}$ being the Moore-Penrose inverse of D since D is of full row rank. Therefore (1.7) reduces to

$$\min_{\theta} z^T P(\theta) z, \qquad P = D^+D \quad . \tag{1.9}$$

As pointed out by Aoki et al [1] for the SISO model, solutions of (1.9) correspond to maximum likelihood estimates of θ when we assume Gaussian white noise. Algorithms and convergence results are given in [1] for *approximate* maximum likelihood methods, whereas in a companion paper [2] they prove convergence results for the *true* maximum likelihood estimators in the special case of no input noise.

2. Algorithm

Since (1.9) has in general no closed-form solution, we need an iterative algorithm to calculate estimates. We propose the Broyden-Fletcher-Goldfarb-Shanno formula (cf. [7], pp. 89-90) which has good numerical properties. It uses object function and gradient evaluations.
Let J_N denote the object function,

$$J_N(\theta) = z^T P(\theta) z \tag{2.1}$$

then for any component of its gradient J_N',

$$\dot{J}_N = 2z^T D^+ \dot{D} P^{\perp} z \tag{2.2}$$

holds, since $\dot{P} = D^+ \dot{D} P^{\perp} + P^{\perp} \dot{D}^T (D^+)^T$
($\cdot$ representing $\frac{\partial}{\partial \theta_i}$ for any element θ_i from θ).

Here, both P and $P^{\perp} := I - P$ are orthogonal projection matrices. To evaluate J_N and J_N' we prefer performing a $Q - R$ decomposition of D^T rather than inverting DD^T. The matrix D^T has full column rank, so there exist orthogonal Q and regular R $(sN \times sN)$ such that

$$Q^T D^T = \begin{bmatrix} R \\ \hline 0 \end{bmatrix} . \tag{2.3}$$

Because of the special form of D^T, R will be lower triangular. If Q_1 is the submatrix of Q consisting of its first sN columns, then

$$D^T = Q_1 R \tag{2.4}$$

holds.
Due to the orthogonality of Q,

$$P = Q_1 Q_1^T \tag{2.5}$$

hence

$$J_N = \|Q_1^T z\|^2 . \tag{2.6}$$

Furthermore, when $\lambda := (D^+)^T z$ (cf. (1.8)) then

$$D^T \lambda = Pz \tag{2.7}$$

and (2.2) implies

$$\dot{J}_N = 2\lambda^T \dot{D}(z - D^T \lambda) . \tag{2.8}$$

Premultiplying (2.7) by Q_1^T gives via (2.4) and (2.5)

$$R\lambda = Q_1^T z . \tag{2.9}$$

Summarizing, when matrix R and vector $u := Q_1^T z$ (length sN) are computed from

$$Q^T[D^T \mid z] = \left[\begin{array}{c|c} R & u \\ \hline 0 & * \end{array}\right] , \tag{2.10}$$

then, from (2.6), (2.8) and (2.9) we obtain

$$J_N = \|u\|^2 \tag{2.11a}$$

and

$$\dot{J}_N = 2\lambda^T \dot{D}(z - D^T\lambda) \tag{2.11b}$$

where λ is easily solved from

$$R\lambda = u \quad , \tag{2.11c}$$

since R is lower triangular.

The matrix Q in (2.10) is not computed explicitly: by means of Householder matrices D^T is transformed into $\begin{bmatrix} R \\ 0 \end{bmatrix}$. Using the special structure this can be done very efficiently. For details we refer to [10]. The computation of R takes $O(N)$ operations, whereas matrix inversion $O(N^2)$, cf. [5].
As another advantage, $Q - R$ decomposition is a numerically stable procedure.

3. Assumptions

From now on the vector of errors e in (1.6) is assumed to be random with zero mean, all its scalar components are stochastically independent and the covariance matrix of e is $\operatorname{var} e = \sigma^2 I$ where σ is unknown. The latter assumption is discussed later, see also [5].
For convenience, the object function which is random now, is multiplied by the term $\frac{1}{sN}$:

$$J_N(\theta) = \frac{1}{sN} z^T P(\theta) z \quad . \tag{3.1}$$

It will be suitable to rewrite θ as a vector of length $\mu := ps^2 + (q+1)sr$,

$$\theta = \begin{bmatrix} (\alpha_1^T)_{*1} \\ \vdots \\ (\alpha_1^T)_{*s} \\ (\alpha_2^T)_{*1} \\ \vdots \\ (\alpha_p^T)_{*s} \\ (\beta_0^T)_{*1} \\ \vdots \\ (\beta_q^T)_{*s} \end{bmatrix} , \tag{3.2}$$

where M_{*j} denotes column j of any matrix M.

We make 5 assumptions which turn out to be sufficient conditions for consistency of any $\hat{\theta}_N$ defined by

$$J_N(\hat{\theta}_N) = \min_{\theta \in \Theta} J_N(\theta) \quad . \tag{3.3}$$

<u>Assumption 1</u>
The set Θ is a known compact subset of $\mathbb{R}^\mu$, which contains the unknown true parameter vector θ_0.

According to Bierens [3], p. 53, Θ is compact and J_N is continuous imply that $\hat{\theta}_N$ defined by (3.3) is indeed a random vector.
In order to state the next assumption we introduce the polynomial matrices

$$\begin{aligned} A(\lambda) &= -I + \sum_{i=1}^{p} \alpha_i \lambda^i \\ B(\lambda) &= \sum_{j=0}^{q} \beta_j \lambda^j \quad . \end{aligned} \tag{3.4}$$

For λ representing the shift-back operator, (1.1) can be written as

$$A(\lambda)\eta_t + B(\lambda)\xi_t = 0, \quad t = m+1, m+2, \cdots \quad . \tag{3.5}$$

Assumption 2

For all $\theta \in \Theta$, $A(\lambda)$ is stable, i.e. the zeros of $\det A(\lambda)$ lie outside the closed unit disk.

We associate with (3.4) matrices A and B (of size $sN \times sN$ and $sN \times rN$ respectively) defined by

$$A = -I + \sum_{k=1}^{p} S^k \otimes \alpha_k$$

$$B = \sum_{k=0}^{q} S^k \otimes \beta_k \tag{3.6}$$

where S is the $N \times N$ shiftmatrix

$$S = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} \tag{3.7}$$

and $\otimes$ denotes the Kronecker product for matrices. From (1.5b) it follows that

$$D = [A\ C_1 \mid B\ C_2] \tag{3.8}$$

with

$$C_1 = \begin{bmatrix} & & \\ & & \\ \alpha_m & & \\ \vdots & \ddots & \\ \alpha_1 & \dots & \alpha_m \end{bmatrix} \quad (sN \times sm \text{ matrix})$$

and

$$C_2 = \begin{bmatrix} & & \\ & & \\ \beta_m & & \\ \vdots & \ddots & \\ \beta_1 & \dots & \beta_m \end{bmatrix} \quad (sN \times rm \text{ matrix}) \ .$$

Now an important consequence of Assumption 2 is given by

Lemma 1.

The matrix AA^T has "uniform bounds": there exist some constants ρ_1 and ρ_2 with $0 < \rho_1 < \rho_2 < \infty$ such that

$$\rho_1 I \leq AA^T \leq \rho_2 I \text{ for all } \theta \in \Theta \text{ and } N \geq p + 1 \ .$$

(By definition: $M_1 \leq M_2$ if $x^T M_1 x \leq x^T M_2 x$ for all x).

Proof Appendix.

Corollary 1

For all $\theta \in \Theta$ and $N \geq p + 1$,

$$\rho_1 I \leq DD^T \leq \rho_3 I$$

holds, with $\rho_2 \leq \rho_3 < \infty$.

Proof

The "lower bound" is obvious from $DD^T = AA^T + BB^T + C_1 C_1{}^T + C_2 C_2{}^T$ and how to define ρ_3 is evident from the proof of Lemma 1. ▯

Assumption 3

The sequence of true inputs $\{\xi_i\}_{i=1}^{\infty}$ is bounded:

there exists a constant M_1 such that $\|\xi_i\| \leq M_1$ for $i = 1, 2, \cdots$.

As a consequence of the well-known BIBO-stability result Assumptions 2 and 3 imply:

Corollary 2

The sequence of true outputs $\{\eta_i\}_{i=1}^{\infty}$ is bounded.

In order to state the next assumption we rewrite the vector $D\zeta$ in (1.5) as

$$D\zeta = (H + K)\,\theta - \begin{bmatrix} \eta_{m+N} \\ \vdots \\ \eta_{m+1} \end{bmatrix} , \tag{3.9}$$

where θ as defined in (3.2) and H and K are $sN \times \mu$ matrices given by

$$H = [(S \otimes I_s)\eta \cdots (S^p \otimes I_s)\eta \mid \xi\ (S \otimes I_s)\xi \cdots (S^q \otimes I_s)\xi] \tag{3.10a}$$

$$K = \left[\begin{array}{cccc|ccccc} & & & & & & & & \\ & & & I_s \otimes \eta_m^T & & & & & I_s \otimes \xi_m^T \\ & & \cdot^{\cdot^{\cdot}} & \vdots & & & & \cdot^{\cdot^{\cdot}} & \vdots \\ I_s \otimes \eta_m^T & \cdots & & I_s \otimes \eta_{m+1-p}^T & 0 & I_s \otimes \xi_m^T & \cdots & & I_s \otimes \xi_{m+1-q}^T \end{array}\right] \tag{3.10b}$$

with I_s is the $s \times s$ identity matrix,

$$\eta = \begin{bmatrix} I_s \otimes \eta_{m+N}^T \\ \vdots \\ I_s \otimes \eta_{m+1}^T \end{bmatrix} (sN \times s^2) \ ,$$

and

$$\xi = \begin{bmatrix} I_s \otimes \xi_{m+N}^T \\ \vdots \\ I_s \otimes \xi_{m+1}^T \end{bmatrix} (sN \times sr) \ .$$

The model equation (1.1) holds for the true value θ_0, so $D\zeta = 0$ for $\theta = \theta_0$. Therefore (3.9) implies

$$D\zeta = (H + K)(\theta - \theta_0) \ . \tag{3.11}$$

Assumption 4

The matrix $\frac{H^T(DD^T)^{-1}H}{sN}$ converges as $N \to \infty$, uniformly on Θ. The uniform limit, say G, is positive definite on Θ.

As will be seen in the next section this assumption implies a convergence result for $\mathbb{E}\ J_N(\theta)$, which is one of the tools for proving consistency.

Generalizing Aoki et al [1], we can give an interpretation for the convergence of $\frac{H^T(DD^T)^{-1}H}{N}$ to a positive definite matrix in the SISO-case $s = r = 1$. The above defined θ, H, η and ξ reduce to

$$\theta = [\alpha_1 .. \alpha_p\ \beta_0 \cdots \beta_q]^T \ ,$$

$$H = [S\eta \cdots S^p\eta \mid \xi\ S\xi \cdots S^q\xi] \ ,$$

$$\eta = [\eta_{m+N} \cdots \eta_{m+1}]^T$$

and

$$\xi = [\xi_{m+N} \cdots \xi_{m+1}]^T \ .$$

Defining

$$\nu = D\zeta - A\eta - B\xi \ ,$$

(1.5) and (3.8) imply

$$\nu = C_1 \begin{bmatrix} \eta_m \\ \vdots \\ \eta_1 \end{bmatrix} + C_2 \begin{bmatrix} \xi_m \\ \vdots \\ \xi_1 \end{bmatrix} \ .$$

For $\theta = \theta_0$ (notation sup index 0) $A^0\eta = -B^0\xi - \nu^0$ holds. Then, using $AS^k = S^kA$ for $k = 1, 2, .., m$,

$$A^0H = [-SB^0\xi \cdots -S^pB^0\xi \mid A^0\xi \cdots S^qA^0\xi] + \omega^0$$

where $\omega := -[S\nu \cdots S^p\nu \mid 0 \cdots 0]$.
Hence

$$H = (A^0)^{-1}(\Xi\, E^0 + \omega^0) \ ,$$

with

$$\Xi = [\xi \ \ S\xi \ \cdots \ S^{p+q}\xi], \quad N \times (p+q+1)$$

and

$$E = \left[\begin{array}{ccc|ccc} 0 & & & -1 & & \\ -\beta_0 & \ddots & & \alpha_1 & \ddots & \\ \vdots & \ddots & 0 & \vdots & \ddots & -1 \\ -\beta_q & & -\beta_0 & \alpha_p & & \alpha_1 \\ & \ddots & \vdots & & \ddots & \vdots \\ & & -\beta_q & & & \alpha_p \end{array} \right] , \quad (p+q+1) \times (p+q+1) \ .$$

The effect of ω^0 in $\frac{H^T(DD^T)^{-1}H}{N}$ vanishes, whence

$$\lim_{N\to\infty} \frac{H^T(DD^T)^{-1}H}{N} = (E^0)^T \lim_{N\to\infty} \frac{\Xi^T(A^0)^{-T}(DD^T)^{-1}(A^0)^{-1}\Xi}{N} E^0 \ . \tag{3.12}$$

The matrices $(A^0)^{-T}(A^0)^{-1}$ and $(DD^T)^{-1}$ can be bounded in the sense of Lemma 1 and its Corollary. Therefore, provided the existence of the limits, the limit in the left hand side of (3.12) is positive definite if and only if

(i) $\lim_{N\to\infty} \frac{\Xi^T\Xi}{N} > 0$

(ii) E^0 is regular .

Condition (i) could be a definition of persistency of excitation of order $p+q$ for the input sequence $\{\xi\}_{i=1}^{\infty}$ (see [1], p. 544), whereas the second condition is equivalent to the statement that the polynomials $A(\lambda)$ and $B(\lambda)$ in (3.4) are coprime (see [9], p. 133).

The last assumption requires the fourth moment of the noises to be uniformly bounded.

Assumption 5

Let e_i denote (scalar) component i of the noise vector e. There exists a constant M_2 such that

$$\mathbb{E}e_i^4 \le M_2 \text{ for all } i = 1,2,\cdots \ .$$

4. Consistency

Consistency is obtained by using a result of Bierens [3], p. 54, 65: when the object function converges *in some sense* uniformly on a compact set to a continuous limit function which is uniquely minimal in the true value of the parameter vector then any minimizing solution converges *in that sense* to the true value.
In the sequel uniform convergence refers to convergence with respect to θ on the compact Θ.
Let us start by proving two lemmas.

Lemma 2

$$\mathbb{E}\ J_N(\theta) \longmapsto J(\theta)\ (N \to \infty),\ \text{uniformly}\quad ,$$

where

$$J(\theta) := \sigma^2 + (\theta - \theta_0)^T G(\theta)\,(\theta - \theta_0)\quad . \tag{4.1}$$

Proof
Observe that the object function defined by (3.1) has mean

$$\begin{aligned}\mathbb{E}\ J_N &= \sigma^2 + \frac{1}{sN}\zeta^T P \zeta \\ &= \sigma^2 + (\theta - \theta_0)^T\ \frac{(H+K)^T (DD^T)^{-1} (H+K)}{sN}\ (\theta - \theta_0)\end{aligned}$$

by (3.11). The lemma follows from Assumption 4, since K has a finite number of nonzero elements. □

Lemma 3

$$\text{var}\ J_N \to 0\ \ (N \to \infty),\ \text{uniformly}\quad .$$

Proof
From

$$z^T P z = e^T P e + \zeta^T P \zeta + 2\zeta^T P e$$

we obtain

$$\text{var}\, z^T P z = \text{var}\ e^T P e + 4\ \text{var}\ \zeta^T P e + 4\ \text{cov}(e^T P e, \zeta^T P e)\quad .$$

Hence

$$\text{var}\ J_N \le \frac{1}{s^2}\ \frac{\text{var}\ e^T P e}{N^2} + \frac{4}{s^2}\ \frac{\text{var}\ \zeta^T P e}{N^2} + \frac{4}{s^2}\sqrt{\frac{\text{var}\ e^T P e}{N^2}}\sqrt{\frac{\text{var}\ \zeta^T P e}{N^2}}$$

by virtue of Cauchy-Schwarz.
Applying a result of Whittle [11] p. 302, gives

$$\text{var } e^T Pe \le k \max_{i=1,2,\ldots,(s+r)(m+N)} \mathbb{E}e_i^4 \, \|P\|^2$$

with k is some constant.
Obviously, var $\zeta^T Pe = \sigma^2 \|P\zeta\|^2$ holds.
The orthogonal projection matrix P obeys $\|P\|^2 = sN$ and $\|P\zeta\|^2 \le \|\zeta\|^2$.
Assumptions 5 and 3 and Corollary 2 imply now that $\frac{\text{var } e^T Pe}{N^2}$ and $\frac{\text{var } \zeta^T Pe}{N^2}$ converge to zero, uniformly, as $N \to \infty$, proving the lemma. □

Remark. As one can see from the proof, Assumption 5 may be weakened to

$$\max_{i=1,2,\ldots,(s+r)(m+N)} \mathbb{E}e_i^4 = o(N) \; .$$

The function J defined by (4.1) is continuous with respect to θ as it is a uniform limit of continuous functions. Now the convergence of the sequence of object functions J_N to J will be shown.

Proposition

$$J_N \overset{p}{\mapsto} J, \text{ uniformly} \; ,$$

(p meaning convergence in probability), i.e.

$$\mathbb{P}(\sup_{\theta \in \Theta} | J_N(\theta) - J(\theta) | > \varepsilon) \to 0 \text{ as } N \to \infty$$

for any $\varepsilon > 0$.

Proof
Since J_N and J are continuous on the compact Θ, there exists a sequence θ_N with $\theta_N \in \Theta$, such that

$$\mathbb{P}(\sup_{\theta \in \Theta} | J_N(\theta) - J(\theta) | > \varepsilon) = \mathbb{P}(| J_N(\theta_N) - J(\theta_N) | > \varepsilon)$$

$$\le \frac{\mathbb{E}(J_N(\theta_N) - J(\theta_N))^2}{\varepsilon^2} = \frac{\text{var } J_N(\theta_N) + (\mathbb{E} \; J_N(\theta_N) - J(\theta_N))^2}{\varepsilon^2}$$

for any $\varepsilon > 0$, applying Chebyshev's inequality. Lemmas 2 and 3 complete the proof. □
We are able to give the main result.

Theorem
Under Assumptions 1-5, any least squares estimator $\hat{\theta}_N$ (defined by (3.3)) is (weakly) consistent for the true parameter vector θ_0, i.e.

$$\mathbb{P}(\|\hat{\theta}_N - \theta_0\| > \varepsilon) \mapsto 0 \text{ as } N \to \infty, \text{ for all } \varepsilon > 0 \; .$$

Proof

According to Lemma 3.1.8 (p. 65) in [3] the result is immediate from the Proposition and two properties of J: it is continuous and has on Θ a unique minimum in θ_0, which is obvious from (4.1) and Assumption 4. □

Remarks.

1. An estimate for the unknown variance σ^2 is given by $J_N(\hat{\theta}_N)$: as a consequence of the Theorem and the Proposition $J_N(\hat{\theta}_N) \overset{P}{\mapsto} J(\theta_0) = \sigma^2$ holds.

2. To discuss the covariance matrix assumption $\operatorname{var} e = \sigma^2 I$, we consider the very special case of no dynamics ($p = q = 0$) and SISO-model ($s = r = 1$). The model and measurements read

$$\eta_t = \beta\xi_t \ , \qquad t = 1,2,\cdots,N$$

$$\left.\begin{aligned} y_t &= \eta_t + \varepsilon_t \ , \\ x_t &= \xi_t + \delta_t \ , \end{aligned}\right\} \qquad t = 1,2,\cdots,N \ .$$

Matrices D and P reduce to

$$D = [-I \ \ \beta I], \quad P = \frac{1}{1+\beta^2}\begin{bmatrix} I & -\beta I \\ -\beta I & \beta^2 I\end{bmatrix}$$

whereas $z = \begin{bmatrix} y \\ x \end{bmatrix}$ and $\theta = \beta$.

When we assume now e.g. $\operatorname{var} e = \operatorname{diag}(\sigma_\varepsilon^2 I, \sigma_\delta^2 I)$ then the object function $J_N(\beta) = \frac{1}{(1+\beta^2)N}\|y - \beta x\|^2$ has mean $\mathbb{E}\ J_N(\beta) = \frac{1}{1+\beta^2}[\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2 + (\beta_0 - \beta)^2 \frac{\xi^T\xi}{N}]$ which converges to $J(\beta) = \frac{1}{1+\beta^2}[\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2 + (\beta_0 - \beta)^2 G]$ by Assumption 4, where β_0 denotes the true value and $\xi = \begin{bmatrix} \xi_N \\ \vdots \\ \xi_1 \end{bmatrix}$.

For $\beta_0 \neq 0$, $J'(\beta_0) = \frac{2\beta_0}{(1+\beta_0^2)^2}(\sigma_\delta^2 - \sigma_\varepsilon^2) \neq 0$ when $\sigma_\delta \neq \sigma_\varepsilon$. Therefore J is not minimal in β_0 and consistency is not obtained for any minimizing solution of $J_N(\beta)$.

Solari [8] has shown that the corresponding likelihood has no maximum, whence additional information is needed (see also Linssen [6], p. 3-4).

5. Conclusion

Estimating the unknown parameters of a *dynamic* model with errors in *all* variables by means of the least squares method, gives an object function which contains some inverse matrix. We propose a $Q - R$ decomposition to evaluate object function and gradient. Those are used in an iterative procedure in order to obtain estimates.
Furthermore, though in general the object function can not be written as a sum of independent random variables (cf. [2]), we are able to prove weak consistency under some mild conditions on input, system and noise.

In future we will report on results of simulations by applying a computer program based on the algorithm described in section 2 and on a proof for asymptotic normality. Generalizations as contraints on the parameters and partial input noise will have our attention as well.

Appendix

Proof of Lemma 1

1. $AA^T \leq \rho_2 I$

For any sN vector x, $x^T AA^T x = \|A^T x\|^2$ holds. By definition of A (see (3.6)) we have

$$A^T x = -x + \sum_{i=1}^{p} [(S^k)^T \otimes \alpha_k{}^T] x \quad ,$$

hence

$$\|A^T x\| \leq (1 + \sum_{k=1}^{p} \|\alpha_k\|) \, \|x\| \quad .$$

Therefore, define

$$\rho_2 = \max_{\theta \in \Theta} (1 + \sum_{k=1}^{p} \|\alpha_k\|)^2 \quad .$$

2. $AA^T \geq \rho_1 I$

It will be proved that $(AA^T)^{-1} \leq \frac{1}{\rho_1} I$.

Generalizing Aoki et al [2] (appendix B), we obtain

$$x^T (AA^T)^{-1} x = \|A^{-1} x\|^2$$

with

$$A^{-1} = \sum_{k=0}^{N-1} S^k \otimes g_k$$

and the $s \times s$ matrices g_k satisfy

$$g_0 = -I$$

$$g_k = \sum_{i=1}^{\min(k,p)} \alpha_i \, g_{k-i} \, , \qquad k = 1,2, \cdots, N-1 \quad .$$

Consequently

$$\|A^{-1} x\| \leq (1 + \sum_{k=1}^{N-1} \|g_k\|) \, \|x\| \quad . \tag{A1}$$

Define for $k = p, p+1, \cdots$ the $ps \times s$ matrix

$$G_k = \begin{bmatrix} g_k \\ g_{k-1} \\ \vdots \\ g_{k-(p-1)} \end{bmatrix} \quad . \tag{A2}$$

Obviously

$$G_{k+1} = \psi\, G_k \;, \quad k = p, p+1, \cdots$$

holds, where ψ is the $ps \times ps$ matrix

$$\psi = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_p \\ I & & & \\ & \ddots & & \\ & & I & 0 \end{bmatrix}$$

with characteristic polynomial

$$p_\psi(\lambda) = p_{\psi^T}(\lambda) = \lambda^{ps} \det(-A(\lambda^{-1})) \text{ for } \lambda \neq 0$$

(with $A(\lambda)$ as defined in (3.4), cf. Chen [4], Section 2.3).

Let $r(\psi)$ denote the spectral radius of ψ, i.e.

$$r(\psi) = \max \{ |\lambda| ; \lambda \text{ eigenvalue of } \psi \} \;,$$

then Assumptions 1 and 2 imply

$$r := \max_{\theta \in \Theta} r(\psi) < 1 \;,$$

since the nonzero eigenvalues of ψ coincide with the reciprocals of zeros of $\det A(\lambda)$ and $r(\psi)$ depends continuously on θ on the compact Θ.

Define $\gamma = \frac{r+1}{2}$, then $r < \gamma < 1$ and Cauchy's formula (cf. Zadeh et al [12], p. 606) gives

$$\psi^k = \frac{1}{2\pi i} \oint_C (zI - \psi)^{-1} z^k dz \;, \quad k = 0, 1, 2, \cdots$$

where C is the circle $\{z \in \mathbb{C} ; |z| = \gamma\}$.
Therefore,

$$\|\psi^k\| \le M\, \gamma^{k+1} \text{ for all } k = 0, 1, \cdots \text{ and all } \theta \in \Theta \qquad \text{(A3)}$$

where $M := \max_{(z,\theta) \in C \times \Theta} f(z, \theta)$ with $f(z, \theta) = \|(z\,I - \psi)^{-1}\|$ is a continuous function on the compact set $C \times \Theta$.

By virtue of (A1), (A2) and (A3) we have now

$$\|A^{-1}x\| \le (1 + \sum_{k=1}^{\infty} \|g_k\|)\, \|x\|$$

with

$$\sum_{k=1}^{\infty} \|g_k\| \le \sum_{k=1}^{p-1} \|g_k\| + \|G_p\|\, M \frac{\gamma}{1-\gamma} \;.$$

Defining

$$\rho_1 = \{\max_{\theta \in \Theta} (1 + \sum_{k=1}^{p-1} \|g_k\| + \|G_p\| \, M \, \frac{\gamma}{1-\gamma})^2\}^{-1}$$

gives

$$x^T (AA^T)^{-1} x \leq \frac{1}{\rho_1} \|x\|^2 \text{ for all } N \geq p+1 \text{ and } \theta \in \Theta \ .$$ []

References

[1] M. Aoki and P.C. Yue, On a priori error estimates of some identification methods, *IEEE Trans. Automatic Control* AC-15 (1970) 541-548.

[2] M. Aoki and P.C. Yue, On certain convergence questions in system identification, *SIAM J. Control* 8 (2) (1970) 239-256.

[3] H.J. Bierens, *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, Lect. Notes Econom. Math. No. 192 (Springer, Berlin, 1981).

[4] H.F. Chen, *Recursive Estimation and Control for Stochastic Systems* (John Wiley, New York, 1985).

[5] F. Eising, H.N. Linssen and H. Rietbergen, System identification from noisy measurements of inputs and outputs, *Systems & Control Letters* 2 (1983) 348-353.

[6] H.N. Linssen, Functional Relationships and Minimum Sum Estimation, Ph.D. Thesis, Eindhoven University of Technology (1980).

[7] L.E. Scales, *Introduction to Non-linear Optimization* (MacMillan, London, 1985).

[8] M.E. Solari, The "maximum likelihood solution" of the problem of estimating a linear functional relationship, *J. R. Statist. Soc. B* 31 (2) (1969) 372-375.

[9] R.M. Staley and P.C. Yue, On system parameter identifiability, *Inform. Sciences 2* (1) (1970) 127-138.

[10] J.M. ten Vregelaar, An algorithm for computing estimates for parameters of an ARMA-model from noisy measurements of inputs and outputs, COSOR-memorandum 87-13, Eindhoven University of Technology (1987).

[11] P. Whittle, Bounds for the moments of linear and quadratic forms in independent variables, *Theory Prob. Applications* 5 (1960) 302-305.

[12] L.A. Zadeh and C.A. Desoer, *Linear Systems Theory, the State-Space Approach* (McGraw-Hill, New York, 1963).