Automatica, Vol. 11, pp. 533–536. Pergamon Press, 1975. Printed in Great Britain

Brief Paper

Continuous Time Systems Identification with Unknown Noise Covariance*

Systèmes d'Identification de Temps Continu avec Co-variation de Bruit (ou Bruits) Inconnus

Identifikation zeitkontinuierlicher Systeme mit unbekannter Rauschkovarianz

ARUNABHA BAGCHI†

Summary—In identifying parameters of a continuous-time dynamical system, a difficulty arises when the observation noise covariance is unknown. The present paper solves this problem in the case of a linear time-invariant system with white noise affecting additively both the state and the observation. The problem is that the likelihood functional cannot be obtained when the observation noise covariance is unknown. A related procedure is suggested, however, and the estimates are obtained by finding roots of an appropriate functional. It is shown that the estimates obtained are consistent.

1. *Introduction*

THE problem of identifying parameters of dynamical systems has been considered by many authors. The excellent survey paper of Åström and Eykhoff [1] discusses various methods and their relative merits. Almost all of the existing literature, however, has been devoted to discrete time dynamical systems, the notable exception being the works of Balakrishnan [2, 3], who extended the maximum likelihood method to continuous-time linear dynamical systems. We consider a continuous-time linear dynamical system with the state and the observation affected by independent Gaussian white noises in an additive manner, i.e. Wiener processes in the integrated form. In this case, the extension of the method of maximum likelihood, as proposed in [3], to estimate the unknown system parameters is possible when the observation noise covariance is known.

In case this is unknown, however, no likelihood functional can apparently be defined. In the present paper, we solve the problem of estimating all the unknown system parameters when the observation noise covariance is unknown. This is done by defining an appropriate functional and showing that minimizing this functional with respect to all the unknown system parameters, including those in the observation noise covariance, yields consistent estimates of those parameters.

* Received 6 November 1974; revised 3 March 1975. The original version of this paper was presented at the 6th IFAC Congress which was held in Boston/Cambridge, Mass., U.S.A., during August 1974. It was recommended for publication in revised form by associate editor K. J. Åström.

† Department of Applied Mathematics, Twente University of Technology, P.O. Box 217, Enschede, The Netherlands.

2. *Problem statement*

Consider the following continuous-time stochastic linear dynamical system

$$x(t;\omega) = \int_0^t Ax(s;\omega)\,ds + \int_0^t Bu(s)\,ds + \int_0^t F\,dW_1(s;\omega), \tag{2.1}$$

$$Y(t;\omega) = \int_0^t Cx(s;\omega)\,ds + \int_0^t Du(s)\,ds + \int_0^t G\,dW_2(s;\omega), \tag{2.2}$$

where $u(t)$ is a p-dimensional 'input' function; $x(t;\omega)$ and $Y(t;\omega)$ are n- and m-dimensional 'state' and 'output' functions respectively; A, B, C, D are $n\times n$, $n\times p$, $m\times n$, $m\times p$ constant but partially unknown matrices; $W_1(t;\omega)$ and $W_2(t;\omega)$ are $n\times 1$ and $m\times 1$ independent Wiener processes and F, G are $n\times n$, $m\times m$ constant but partially unknown matrices. We assume that $GG^* > 0$ to exclude the singular case, where * denotes the transpose. One problem of systems identification is to estimate the unknown parameters in the matrices A, B, C, D, F and G, based on the observation $Y(t;\omega)$ for one realization of the experiment in $0\leq t\leq T$. Let θ denote the vector of all the unknown system parameters and let θ_0 be their true values. Let $\hat{\theta}_T(\omega)$ denote an estimate of θ_0 based on the data $Y(t;\omega)$ for $0\leq t\leq T$. Our objective is to determine a consistent estimate of θ_0; that is, to find an estimate $\hat{\theta}_T(\omega)$ such that

$$\hat{\theta}_T(\omega) \to \theta_0 \quad \text{in probability as } T\to\infty.$$

In the corresponding discrete time dynamical system, the method of maximum likelihood has been proposed by Åström and Bohlin [4] in the scalar-input, scalar-output case. The extension to vector case has been done by Kashyap [5]. Both the papers contain outline of proofs of consistency of the proposed estimates. In the continuous time problem, let $\mathbf{C} = C^m[0, T]$ denote the space of continuous functions from $[0, T]$ into R^m. Let p_Y be the measure induced on $\mathbf{C}$ by the observation process $Y(t;\omega)$, $0\leq t\leq T$. In analogy with the discrete time case, a likelihood functional can be defined if we find a fixed measure on $\mathbf{C}$ such that p_Y is absolutely continuous with respect to that measure. If G is known, the measure p_{GW_2} induced by the process $GW_2(t;\omega)$, $0\leq t\leq T$, is such a measure and in this case, we can define the likelihood functional as the corresponding Radon–Nikodym derivative evaluated at the sample trajectory of the observation. The above brief discussion is included only for the sake of completeness. The concepts introduced are explained in Yeh [6, Chapts. 6 and 8].

Let us now explain the fundamental difference and the associated difficulty in the continuous-time problem. In the above discussion, we assumed G to be known and took the fixed measure to be p_{GW_s} instead of the more obvious measure p_{W_s} on C. The reason is that p_Y is absolutely continuous with respect to p_{GW_s} but if $GG^* \neq I$, p_{GW_s} is singular with respect to p_{W_s}! This last assertion is an obvious generalization of the corresponding well-known result in one dimension [6, Theorem 32.1]. Intuitively, this can be seen as follows. If we sample $GW(t;\omega)$ and $W(t;\omega)$ at a finite number, say k, of time points, the likelihood ratio of the resulting Gaussian random vectors contains an exponential term multiplied by a determinant term $[(2\pi)^m \det(GG^*)]^{-k/2}$. But as $k\to\infty$, this multiplying term goes to zero, indicating that the function space measures will not be absolutely continuous. When G is known, the likelihood functional defined above is given by [3, p. 195]

$$
\begin{aligned}
&H[\theta;\ Y(\cdot\,;\omega);\ T]\\
&= \exp\Big[-\frac{1}{2}\Big(\int_0^T\{(GG^*)^{-1}[C\hat{x}(t;\theta;\omega)+Du(t)],\\
&\quad C\hat{x}(t;\theta;\omega)+Du(t)\}\,dt-2\\
&\quad\times\int_0^T\{(GG^*)^{-1}[C\hat{x}(t;\theta;\omega)+Du(t)],\\
&\quad dY(t;\omega)\}\Big)\Big],
\end{aligned}
\tag{2.3}
$$

where $[\cdot,\cdot]$ denotes inner product in R^m and the second integral is to be interpreted in the Itô sense. Here $\hat{x}(t;\theta;\omega)$ satisfies the stochastic differential equation

$$
\begin{aligned}
\hat{x}(t;\theta;\omega) &= \int_0^t A\hat{x}(s;\theta;\omega)\,ds+\int_0^t Bu(s)\,ds\\
&\quad+\int_0^t P(s)\,C(GG^*)^{-1}\,dZ(s;\theta;\omega),
\end{aligned}
\tag{2.4}
$$

$$
\begin{aligned}
Z(t;\theta;\omega) &= Y(t;\omega)-\int_0^t C\hat{x}(s;\theta;\omega)\,ds\\
&\quad-\int_0^t Du(s)\,ds,
\end{aligned}
\tag{2.5}
$$

where $P(t)$ satisfies the matrix Ricatti equation

$$
\begin{aligned}
\dot{P}(t) &= AP(t)+P(t)A^*+FF^*-P(t)C^*(GG^*)^{-1}CP(t);\\
P(0) &= 0.
\end{aligned}
\tag{2.6}
$$

At the true value θ_0, $Z(t;\theta_0;\omega)$ is a Wiener process, the so-called 'innovation process', with

$$E[Z(t;\theta_0;\omega)\,Z(s;\theta_0;\omega)^*] = (GG^*)_0 \min(t,s), \tag{2.7}$$

where $(GG^*)_0$ denotes the true value of GG^*. Let

$$q[\theta;\ Y(\cdot\,;\omega);\ T] = -(2/T)\log H[\theta;\ Y(\cdot\,;\omega);\ T]. \tag{2.8}$$

It has been shown in [3] that minimizing this functional in an appropriate neighbourhood of θ_0 yields consistent estimates of all the unknown system parameters. In analyzing the proof, one finds that the weights $(GG^*)^{-1}$ occurring in the expression q play no role in the consistency consideration of the estimates. We note further that (GG^*) occurs in q through $\hat{x}(t;\theta;\omega)$ also. These two observations are exploited to solve the estimation problem when G is unknown.

3. *Method of estimation*

Consider the situation when G is unknown in which case the likelihood functional cannot be apparently defined. We propose a functional analogous to the log-likelihood functional expression q as given by (2.8) for known G. Let $\tilde{G}_0$ be an *a priori* guess of G_0, the true value of G, so that $\tilde{G}_0\tilde{G}_0^* > 0$. We replace the weights $(GG^*)^{-1}$ occurring in q by the guessed values $(\tilde{G}_0\tilde{G}_0^*)^{-1}$ but regard q otherwise as a function of all the unknown parameters including those in G. We can take $\tilde{G}_0\tilde{G}_0^*$ to be the identity matrix without affecting the consistency property of the estimates but a value close to $(GG^*)_0$ may be desirable from a numerical point of view. Noting that θ now stands for all the unknown system parameters including those in G, we now define a modified functional

$$
\begin{aligned}
&\tilde{q}[\theta;\ Y(\cdot\,;\omega);\ T]\\
&= \frac{1}{T}\Big(\int_0^T\{(\tilde{G}_0\tilde{G}_0^*)^{-1}[C\hat{x}(t;\theta;\omega)+Du(t)],\\
&\quad C\hat{x}(t;\theta;\omega)+Du(t)\}\,dt-2\\
&\quad\times\int_0^T\{(\tilde{G}_0\tilde{G}_0^*)^{-1}[C\hat{x}(t;\theta;\omega)+Du(t)],\,dY(t;\omega)\}\Big).
\end{aligned}
\tag{3.1}
$$

In the following, we shall be working only with $\tilde{q}$ and, therefore, we replace $\tilde{q}$ by q without any possibility of confusion with the q in (2.8).

We claim that minimizing this modified functional in a neighbourhood of θ_0 will yield consistent estimates of all the unknown parameters of the system under a certain sufficiency condition. To investigate this asymptotic property of the estimates, we need to impose certain conditions on the system and the input to ensure convergence of some expressions as T becomes very large. We consider θ to be in a suitable neighbourhood $\bar{N}$ of θ_0 to be defined later.

Condition 1. For any θ in $\bar{N}$, assume that A is a stable matrix. This condition implies that the initial condition of the state has no effect on the asymptotic behaviour of the system and this justifies our setting the initial condition to be zero in (2.1).

Condition 2. For any θ in $\bar{N}$, assume that the pair (C, A) is completely observable.

This condition implies that $\lim_{t\to\infty} P(t)$ exists; we denote the limit by P. P satisfies the algebraic equation

$$AP+PA^*+FF^*-PC^*(GG^*)^{-1}CP = 0 \tag{3.2}$$

and, furthermore, the matrix $A-PC^*(GG^*)^{-1}C$ is also stable.

Condition 3. We assume that the input $u(\cdot)$ is such that

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \|u(t)\|^2\,dt<\infty \quad \text{(exists and is finite)}$$

and

$$r_u(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T u(s)\,u(s+t)^*\,ds$$

is a continuous function of t in every finite interval. This condition implies the existence of 'time averages' occurring in the proof of consistency of the estimates.

Let θ_i denote the ith component of θ and let

$$\nabla_\theta\, q[\theta;\ Y(\cdot\,;\omega);\ T]$$

be the gradient vector with the ith component

$$q_i[\theta;\ Y(\cdot\,;\omega);\ T] = \frac{\partial}{\partial\theta_i}q[\theta;\ Y(\cdot\,;\omega);\ T].$$

Let $Q[\theta;\ Y(\cdot\,;\omega);\ T]$ be the matrix with the ijth component

$$q_{ij}[\theta;\ Y(\cdot\,;\omega);\ T] = \frac{\partial^2}{\partial\theta_i\,\partial\theta_j}q[\theta;\ Y(\cdot\,;\omega);\ T].$$

We use the following notation

$$q(\theta;\,T) = Eq[\theta;\; Y(\cdot\,;\,\omega);\; T],$$
$$q(\theta) = \lim_{T\to\infty} q(\theta;\,T)$$

and establish similar notations for the partial derivatives of q. Consistency is proved by showing that in any arbitrarily small neighbourhood of θ_0, the equation

$$\nabla_\theta\, q[\theta;\; Y(\cdot\,;\,\omega);\; T] = 0 \qquad (3.3)$$

has a root for T sufficiently large. Now if θ_1 is another point in the parameter space which yields the same response $Y(t;\,\omega)$ for $t \geq 0$ and almost all ω, the above result would hold in a neighbourhood of θ_1 also. To make sense of the above procedure, we have, therefore, to ensure that there at least exists a neighbourhood of θ_0 in which no other value of θ yields the same response as that of θ_0. The following lemma gives a sufficient condition for this to hold. The proof of the lemma is given in [7].

Lemma 1. If $Q(\theta_0)$ is positive definite, there is a neighbourhood of θ_0 such that no other value of θ will yield a response identical to the one observed for all $t > 0$ for any ω, omitting a set of zero probability.

Remark. If $Q(\theta_0)$ is positive definite, there exists a neighbourhood of θ_0 in which no other value of θ yields a response identical to the one observed for $t \geq 0$ for almost all ω. We consider θ in that neighbourhood. Take the neighbourhood to be closed and bounded and denote it by $\bar{N}$.

We now have the following theorem ensuring consistency.

Theorem 1. Assume that $Q(\theta_0)$ is positive definite. Then given any arbitrarily small positive numbers δ and ε, one can always find a number $T_0 = T_0(\delta, \varepsilon)$ such that for all $T > T_0$, there exists an ω-set $\Lambda(\varepsilon)$ of probability less than ε while for all $\omega \notin \Lambda(\varepsilon)$, the equation

$$\nabla_\theta\, q[\theta;\; Y(\cdot\,;\,\omega);\; T] = 0$$

has a root in the sphere $S_\delta(\theta_0)$ of radius δ about θ_0.

Remark. The above theorem asserts that given any neighbourhood of the true value θ_0, however small, and for ω belonging to a set of probability arbitrarily large, we can find T sufficiently large such that equation (3.3) has a solution in that neighbourhood. Equivalently, one root of equation (3.3) converges to θ_0 in probability as T goes to infinity. The proof of this theorem is given in the Appendix.

It should be noted that the estimates are consistent under the sufficient condition that $Q(\theta_0)$ is positive definite. This is called the 'identifiability condition'. The expression for $Q(\theta_0)$ is given in [7].

4. *Conclusion*

The problem of obtaining consistent estimates of the parameters of a linear, time-invariant, continuous-time stochastic dynamical system when the observation noise covariance is unknown has been solved in the present paper. The estimation method proposed can be used to identify the structured parameters of an aircraft in motion when it is subject to random wind disturbance [8], in the case of unknown observation noise covariance.

Acknowledgements—The author wishes to thank Professor A. V. Balakrishnan for suggesting the problem. Some of the referee's comments are deeply appreciated.

References

[1] K. J. ÅSTRÖM and P. EYKHOFF: Systems identification—a survey. *2nd IFAC Symp. Identification and Process Parameter Estimation*, pp. 1–38, Prague, Czechoslovakia, 1970. Also in *Automatica* 7, 123–162 (1971).

[2] A. V. BALAKRISHNAN: *Techniques of System Identification*, Lecture Notes. University of Rome (1968).

[3] A. V. BALAKRISHNAN: *Stochastic Differential Systems*, Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin (1973).

[4] K. J. ÅSTRÖM and T. BOHLIN: Numerical identification of dynamic systems for normal operating records. *2nd IFAC Symp. Theory of Self-adaptive Systems*, pp. 76–111, Teddington, England (1965).

[5] R. L. KASHYAP: Maximum likelihood identification of stochastic linear systems. *IEEE Trans. Aut. Control* AC-15, 25–34 (1970).

[6] J. YEH: *Stochastic Processes and the Wiener Integral.* Marcel Dekker, New York (1973).

[7] A. BAGCHI: Continuous Time Systems Identification with Unknown Noise Covariance. Ph.D. Dissertation in Engineering, University of California, Los Angeles (1975).

[8] A. V. BALAKRISHNAN: Identification and adaptive control: an application to flight control systems. *J. Opt. Theory Appl.* 9, 187–212 (1972).

APPENDIX

Proof of Theorem 1. The proof of Theorem 1 is based on the following two lemmas. The proofs are omitted and can be found in [7].

Lemma 2. $\nabla_\theta\, q(\theta_0) = 0$.

Lemma 3. Assuming that Conditions 1–3 of Section 3 are satisfied,

$$E\{\, \| q[\theta;\; Y(\cdot\,;\,\omega);\; T] - q(\theta) \|^2 \}$$

goes to zero as T goes to infinity and, moreover, the convergence is uniform with respect to θ in compact subsets.

Corollary. An analogous statement holds for all derivatives of $q(\theta;\; Y(\cdot\,;\,\omega);\; T)$.

Proof of Theorem 1. Expanding $\nabla_\theta\, q[\theta;\; Y(\cdot\,;\,\omega);\; T]$ in Taylor series about θ_0, we have

$$\begin{aligned}\nabla_\theta\, q[\theta;\; Y(\cdot\,;\,\omega);\; T] &= \nabla_\theta\, q[\theta_0;\; Y(\cdot\,;\,\omega);\; T] \\ &\quad + Q[\theta_0;\; Y(\cdot\,;\,\omega);\; T]\,(\theta-\theta_0) \\ &\quad + \{\bar{J}[\theta;\; Y(\cdot\,;\,\omega);\; T]\,(\theta-\theta_0)\}\,(\theta-\theta_0), \qquad (*)\end{aligned}$$

where

$$\bar{J}[\theta;\; Y(\cdot\,;\,\omega);\; T] = \int_0^1 \int_0^1 J\{[(1-ts)\,\theta_0 + st\theta];\; Y(\cdot\,;\,\omega);\; T\}\,(1-s)\,\mathrm{d}s\,\mathrm{d}t$$

with $J[\theta;\; Y(\cdot\,;\,\omega);\; T]$ denoting the gradient (Fréchet derivative) of $Q[\theta;\; Y(\cdot\,;\,\omega);\; T]$ with respect to θ.

Let $m > 0$ be the lowest spectral bound for the matrix $Q(\theta_0)$ which is positive definite by assumption. By Lemma 3 and the corollary, all coefficients in (*) converge in mean square sense, uniformly in θ in the compact set $\bar{N}$. Let $M = \sup_{\theta\in\bar{N}} \bar{J}(\theta)$. Then given $\delta_1 > 0$, there exists a set $\Lambda(\varepsilon)$ of measure less than ε such that for all $T > T_0(\delta_1, \varepsilon)$ and for $\omega \notin \Lambda(\varepsilon)$,

$$\begin{aligned}\| \nabla_\theta\, q[\theta_0;\; Y(\cdot\,;\,\omega);\; T] - \nabla_\theta\, q(\theta_0) \| &= \| \nabla_\theta\, q[\theta_0;\; Y(\cdot\,;\,\omega);\; T] \| < \delta_1,\\ \| Q[\theta_0;\; Y(\cdot\,;\,\omega);\; T] - Q(\theta_0) \| &< m/2,\\ \| \bar{J}[\theta;\; Y(\cdot\,;\,\omega);\; T] - \bar{J}(\theta) \| &< M/2 \quad \text{for } \theta \in \bar{N}.\end{aligned}$$

Then $Q[\theta_0;\ Y(\cdot\,;\omega);\ T]$ is also positive definite with the smallest spectral bound greater than $m/2$ implying that

$$\| Q[\theta_0;\ Y(\cdot\,;\omega);\ T]^{-1} \| \leq 2/m.$$

Let $x = \theta - \theta_0$. Define

$$\begin{aligned} f(x) &= x - \{Q[\theta_0;\ Y(\cdot\,;\omega);\ T]\}^{-1} \nabla_\theta q[x+\theta_0;\ Y(\cdot\,;\omega);\ T] \\ &= -[\![Q[\theta_0;\ Y(\cdot\,;\omega);\ T]^{-1} \nabla_\theta q[\theta_0;\ Y(\cdot\,;\omega);\ T] \\ &\quad + Q[\theta_0;\ Y(\cdot\,;\omega);\ T]^{-1}\{J[\theta;\ Y(\cdot\,;\omega);\ T]\,(x)\}\,(x)]\!] \\ &\quad \text{by } (*). \end{aligned}$$

From our previous estimates, we get

$$\| f(x) \| \leq 2\delta_1/m + 3(M/m) \| x \|^2.$$

Choose δ such that $\delta < m/4M$ and then choose $\delta_1 = m\delta/8$.

Then for all x such that $\| x \| \leq \delta$, $\| f(x) \| \leq \delta$. The map f being continuous, by Brouwer fixed point theorem, there exists a point x_r with $\| x_r \| \leq \delta$ such that $f(x_r) = x_r$.

From the definition of f, there is a point θ_r whose distance from θ_0 is less than or equal to δ, such that

$$\nabla_\theta q[\theta_r;\ Y(\cdot\,;\omega);\ T] = 0$$

proving the desired result.