COMBINED FILTERING AND PARAMETER ESTIMATION FOR DISCRETE-TIME SYSTEMS DRIVEN BY APPROXIMATELY WHITE GAUSSIAN NOISE DISTURBANCES

Wolfgang J. Runggaldier
Cinzia Visentin

July 1987
WP-87-60

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

In the problem of combined filtering and parameter estimation one considers a stochastic dynamical system whose state x_t is only partially observed through an observation process y_t. The stochastic model for the process pair (x_t, y_t) depends furthermore on an unknown parameter θ. Given an observation history of the process y_t, the problem then consists in estimating recursively both the current state x_t of the system (filtering) as well as the value θ of the parameter (Bayesian parameter estimation).

The problem is a rather difficult one: Even if, conditionally on a given value of θ, the process pair (x_t, y_t) satisfies a linear–Gaussian model so that the filtering problem for x_t can be solved via the familiar Kalman-Bucy filter, when θ is unknown, the problem becomes a difficult nonlinear filtering problem.

The present paper, partly based on previous joint work of one of the authors, makes a contribution towards the solution of this problem in the case of discrete time and of a (conditionally on θ) linear model for x_t, y_t. The solution that is obtained is shown to be robust with respect to small variations in the a-priori distributions in the model, in particular those of the disturbances.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

ABSTRACT

We consider a partially observable, discrete-time process $\{x_t, \theta_t, y_t\}$ over a finite horizon T. The unobservable components are $\{x_t, \theta_t\}$ and may be interpreted as state of a stochastic dynamical system, partially observed through $\{y_t\}$. Conditionally on $\{\theta_t\}$, the pair $\{x_t\}$, $\{y_t\}$ satisfies a linear model of the form (1) below; $\{\theta_t\}$ itself evolves according to a given joint a-priori distribution $p(\theta_0, \ldots, \theta_T)$.

The purpose of the paper is to determine recursively the joint conditional distribution $p(x_t, \theta_t \mid y^t)$, $(y^t := \{y_0, \ldots, y_t\})$, or, more specifically, $E\{f(x_t, \theta_t) \mid y^t\}$, namely the (mean-square) optimal filter for a given $f(x_t, \theta_t)$.

When θ_t is constant $(\theta_t \equiv \theta)$ and can therefore be interpreted as an unknown parameter, our problem becomes that of the combined filtering and parameter estimation.

The optimal filter is computed for the ideal situation of white Gaussian noises in the model (1) below and it is shown that, when this filter is applied to a more realistic situation where the noises are only approximately (in the sense of weak convergence of measures) white Gaussian and also $\{\theta_t\}$ has only approximately the given distribution $p(\theta_0, \ldots, \theta_T)$, then it remains almost (mean-square) optimal with respect to all alternative filters that are continuous and bounded functions of the past observations.

CONTENTS

COMBINED FILTERING AND PARAMETER ESTIMATION FOR DISCRETE-TIME SYSTEMS DRIVEN BY APPROXIMATELY WHITE GAUSSIAN NOISE DISTURBANCES

Wolfgang J. Runggaldier and Cinzia Visentin

INTRODUCTION

We consider a process triple (x_t, θ_t, y_t), $(x_t \in \mathbf{R}^n, \theta_t \in \mathbf{R}^q, y_t \in \mathbf{R}^m)$ satisfying over a finite horizon $t = 0, 1, \ldots, T$ the model

$$x_t = A_{t-1}(\theta_{t-1})x_{t-1} + B_{t-1}(\theta_{t-1}) + Q_{t-1}(\theta_{t-1})w_t \tag{1.a}$$

$$y_t = C_t(\theta_t)x_t + D_t(\theta_t) + R_t(\theta_t)v_t \tag{1.b}$$

$$x_0 = B_0(\theta_0) + Q_0(\theta_0)w_0 \tag{1.c}$$

where $\{\theta_t\}$ is defined through its joint a-priori distribution $p(\theta_0, \ldots, \theta_T)$ equivalent to assigning

$$p_0(\theta_0) \;\; ; \qquad p(\theta_t \,|\, \theta_{t-1}, \ldots, \theta_0) \tag{1.d}$$

The processes $\{w_t\}$, $\{v_t\}$ $(w_t \in \mathbf{R}^n, v_t \in \mathbf{R}^m)$, also defined through a joint a-priori distribution, are supposed to be independent of $\{\theta_t\}$. Furthermore, $A_{t-1}(\cdot)$, $B_{t-1}(\cdot)$, $Q_{t-1}(\cdot)$, $C_t(\cdot)$, $D_t(\cdot)$, $R_t(\cdot)$ are given (matrix valued) functions of appropriate dimensions and $Q_{t-1}(\cdot)$, $R_t(\cdot)$, are positive definite.

We may think of the pair (x_t, θ_t) as the state of a dynamical system, partially observed through the process y_t. The component θ_t of the state evolves as a (nonlinear), not necessarily Markov process, conditionally on which the process pair (x_t, y_t) satisfies a linear model. If, in particular, θ_t is constant $(\theta_t \equiv \theta)$, then we may think of θ as an unknown parameter in the linear model (1) for (x_t, y_t) with a given a-priori distribution $p(\theta)$.

We shall interpret model (1) in the sense that process y_t is being observed starting from time $t = 1$ and that, at a generic time t, first a transition $(x_{t-1}, \theta_{t-1}) \rightarrow (x_t, \theta_t)$ takes place and then an observation is generated according to (1.b).

Our problem consists in computing recursively for $t = 1, \ldots, T$ the joint conditional distribution $p(x_t, \theta_t \mid y^t)$, given the observations $y^t := \{y_1, \ldots, y_t\}$, letting $p(x_0, \theta_0 \mid y^0) = p(x_0, \theta_0)$ namely the joint initial distribution for (x_0, θ_0) defined through (1.c), (1.d). More specifically, given an integrable function $f(x, \theta)$, instead of considering the entire conditional distribution $p(x_t, \theta_t \mid y^t)$, we are interested in computing recursively

$$\Phi_f(t, y^t) := E\{f(x_t, \theta_t) \mid y^t\} \tag{2}$$

namely the optimal filter for $f(x_t, \theta_t)$ in the sense of the minimal mean square error. In the case when θ_t is constant, and therefore has the interpretation of an unknown parameter for model (1), our problem becomes that of the combined filtering and parameter estimation.

We remark here that the results to be obtained below can easily be seen to hold also if $A_{t-1}(\theta_{t-1})$, $B_{t-1}(\theta_{t-1})$, $Q_{t-1}(\theta_{t-1})$, $C_t(\theta_t)$, $D_t(\theta_t)$, $R_t(\theta_t)$ in model (1) are allowed to depend (continuously) on y^{t-1} and $p(\theta_t \mid \theta_{t-1}, \ldots, \theta_0)$ in (1.d) are continuous functions also of y^{t-1}. Since this more general assumption would make the notation much heavier and less transparent, we prefer to present our derivations for the simpler model as described above.

In the next Section I we shall assume that $\{w_t\}$, $\{v_t\}$ are independent and standard white Gaussian and that θ_t takes only a finite number of possible values. In this case, in analogy to [2], we show that the problem can be solved explicitly via a recursive procedure that becomes also finite-dimensional if θ_t is a constant random variable.

In the following Section II, still retaining the white Gaussian assumption for $\{w_t\}$, $\{v_t\}$, we assume that θ_t is absolutely continuous. In this second case, since an explicit solution is in general impossible, we derive an approximation algorithm along the lines of what is done e.g. in [3] for general nonlinear filtering problems. We discretize the process θ_t to obtain a finite-state process $\theta_t^{(N)}$ so that the algorithm of Section I can be used to compute exactly conditional expectations of the form

$$\Phi_f^{(N)}(t, y^t) := E\{f(x_t, \theta_t^{(N)}) \mid y^t\} \tag{3}$$

The discretization will be such that under suitable assumptions which essentially refer to the continuity of $f(x, \theta)$ and of the coefficients in model (1), one has for all t and all y^t

$$\lim_{N \to \infty} \Phi_f^{(N)}(t, y^t) = \Phi_f(t, y^t) \tag{4}$$

In the last Section III, in analogy to [4], [5], we assume more realistically that $\{w_t\}$, $\{v_t\}$ are only approximately (in the sense of weak convergence of probability measures) white Gaussian and that also θ_t is only approximately distributed according to the given $p_0(\theta_0)$ and $p(\theta_t \mid \theta_{t-1}, \ldots, \theta_0)$ of (1.d). In order that an approximation argument can be used, we embed such a problem into a family of similar problems parameterized by a parameter $\epsilon > 0$. Instead of model (1) we then have

$$x_t^\epsilon = A_{t-1}(\theta_{t-1}^\epsilon)x_{t-1}^\epsilon + B_{t-1}(\theta_{t-1}^\epsilon) + Q_{t-1}(\theta_{t-1}^\epsilon)w_t^\epsilon \tag{5.a}$$

$$y_t^\epsilon = C_t(\theta_t^\epsilon)x_t^\epsilon + D_t(\theta_t^\epsilon) + R_t(\theta_t^\epsilon)v_t^\epsilon \tag{5.b}$$

$$x_0^\epsilon = B_0(\theta_0^\epsilon) + Q_0(\theta_0^\epsilon)w_0^\epsilon \tag{5.c}$$

where we assume that

$$(\{w_t^\epsilon\}, \{v_t^\epsilon\}, \{\theta_t^\epsilon\}) \Rightarrow (\{w_t\}, \{v_t\}, \{\theta_t\}) \tag{5.d}$$

where $\Rightarrow$ denotes weak convergence, $\{w_t\}$, $\{v_t\}$ are independent standard white Gaussian, $\{\theta_t\}$ has a given joint density and is independent of $\{w_t\}$, $\{v_t\}$. For such models it will in general be impossible to derive an analytical expression for $p(x_t^\epsilon, \theta_t^\epsilon \mid (y^\epsilon)^t)$ or to explicitly compute $E\{f(x_t^\epsilon, \theta_t^\epsilon) \mid (y^\epsilon)^t\}$. A natural approximate approach then consists in using the exact or approximate solution obtained for the ideal limiting model (1), namely $\Phi_f(t, y^t)$ or $\Phi_f^{(N)}(t, y^t)$ according to whether θ_t is finite-state or not, and to apply it to the physical model (5), i.e. to compute $\Phi_f(t, (y^\epsilon)^t)$, respectively $\Phi_f^{(N)}(t, (y^\epsilon)^t)$, with the actual observations y_t^ϵ replacing the y_t of the ideal model (1). Under suitable assumptions, essentially those of the previous Section in addition to the boundedness of $f(x, \theta)$, we then show that for all $w.p.$ 1 continuous and bounded functions $F_t(y^t)$ of the past observations $y^t = \{y_1, \ldots, y_t\}$ one has for all t

$$\lim_{\epsilon \downarrow 0} E\{\,|\,f(x_t^\epsilon, \theta_t^\epsilon) - F_t((y^\epsilon)^t)\,|^2\} \geq$$

$$\geq \lim_{\substack{\epsilon \downarrow 0 \\ N \to \infty}} E\{\,|\,f(x_t^\epsilon, \theta_t^\epsilon) - \Phi_f^{(N)}(t, (y^\epsilon)^t)\,|^2\} \tag{6}$$

where the dependence on - and the limit with respect to N can be dropped if θ_t is finite-state. Relation (6) says that the approximate filter $\Phi_f^{(N)}(t, (y^\epsilon)^t)$ is, for small ϵ and large N, an almost (mean square)-optimal filter for $f(x_t^\epsilon, \theta_t^\epsilon)$ with respect to all alternative filters that are continuous and bounded functions of the past observations. This can also be expressed by saying that with respect to the above alternative filters, the filter for the ideal model (1) is robust to small variations in the a-priori distributions of the model.

1. GAUSSIAN WHITE NOISE; θ_t FINITE-STATE

(Exact solution)

In this Section we assume that the sequences $\{w_t\}$, $\{v_t\}$ in model (1) are independent white Gaussian and that θ_t takes only a finite number of possible values. The result to be obtained here is given in Theorem 1.1 below that makes it possible to derive an algorithm for the exact recursive computation of $p(x_t, \theta_t \mid y^t) = \sum_{\theta^{t-1}} p(x_t, \theta^t \mid y^t)$ where $\theta^t := (\theta_0, \dots, \theta_t)$

From the recursive Bayes formula we have

$$p(x_t, \theta^t \mid y^t) \propto p(y_t \mid x_t, \theta^t, y^{t-1}) \cdot$$

$$\cdot \int p(x_t, \theta_t \mid x_{t-1}, \theta^{t-1}, y^{t-1}) p(x_{t-1}, \theta^{t-1} \mid y^{t-1}) \mathrm{d}x_{t-1} \tag{7}$$

where $\propto$ denotes proportionality and where, denoting by $g(x; \mu, \sigma^2)$ the Gaussian density with mean μ and covariance matrix σ^2, from (1.b) we have

$$p(y_t \mid x_t, \theta^t, y^{t-1}) = p(y_t \mid x_t, \theta_t) = g(y_t; C_t(\theta_t)x_t + D_t(\theta_t), R^2(\theta_t)) \tag{8}$$

Furthermore, given the independence of $\{w_t\}$, $\{v_t\}$, $\{\theta_t\}$ from (1.a) we have

$$\begin{aligned} p(x_t, \theta_t \mid x_{t-1}, \theta^{t-1}, y^{t-1}) &= p(x_t \mid x_{t-1}, \theta^t, y^{t-1}) p(\theta_t \mid x_{t-1}, \theta^{t-1}, y^{t-1}) = \\ &= p(\theta_t \mid \theta^{t-1}) p(x_t \mid x_{t-1}, \theta_{t-1}) = \\ &= p(\theta_t \mid \theta^{t-1}) g(x_t; A_{t-1}(\theta_{t-1})x_{t-1} + B_{t-1}(\theta_{t-1}), Q^2_{t-1}(\theta_{t-1})) \end{aligned} \tag{9}$$

Finally, from (1.c), (1.d)

$$p(x_0, \theta_0 \mid y^0) = p(x_0, \theta_0) = p_0(\theta_0) g(x_0; B_0(\theta_0), Q_0^2(\theta_0)) \tag{10}$$

In what follows, given a generic nonsingular matrix M, we let

$$M^{-2} := (M^{-1})' M^{-1}$$

denoting transposition by a prime. Extending a procedure in [2] we now have

THEOREM 1.1 *Under the current assumptions on model (1) we have for all $t = 0, \dots, T$*

$$p(x_t, \theta_t \mid y^t) \propto \sum_{\theta^{t-1}} p(\theta^t) \varphi_t(\theta^t) \cdot$$

$$\cdot \exp\left\{ -\frac{1}{2} x_t' M_t(\theta^t) x_t + x_t' h_t(\theta^t, y^t) + k_t(\theta^t, y^t) \right\} \tag{11}$$

where

$$p(\theta^t) = p_0(\theta_0)p(\theta_1 \mid \theta_0) \cdots p(\theta_t \mid \theta^{t-1}) \tag{12}$$

and where, letting

$$N_t(\theta^{t-1}) := A'_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1})A_{t-1}(\theta_{t-1}) + M_{t-1}(\theta^{t-1}) \ . \tag{13}$$

the other coefficients are obtained recursively as

$$\varphi_t(\theta^t) = \varphi_{t-1}(\theta^{t-1})[(\det Q_{t-1}(\theta_{t-1}))(\det R_t(\theta_t))]^{-1} \bullet \tag{14.a}$$

$$\bullet (\det N_t(\theta^{t-1}))^{-1/2}; \qquad \varphi_0(\theta_0) = (\det Q_0(\theta_0))^{-1}$$

$$M_t(\theta^t) = C'_t(\theta_t)R_t^{-2}(\theta_t)C_t(\theta_t) + Q_{t-1}^{-2}(\theta_{t-1}) - \tag{14.b}$$

$$- Q_{t-1}^{-2}(\theta_{t-1})A_{t-1}(\theta_{t-1})N_t^{-1}(\theta^{t-1})A_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1}); \qquad M_0(\theta_0) = Q_0^{-2}(\theta_0)$$

$$h_t(\theta^t, y^t) = C'_t(\theta_t)R_t^{-2}(\theta_t)(y_t - D_t(\theta_t)) + Q_{t-1}^{-2}(\theta_{t-1})B_{t-1}(\theta_{t-1}) + \tag{14.c}$$

$$+ Q_{t-1}^{-2}(\theta_{t-1})A_{t-1}(\theta_{t-1})N_t^{-1}(\theta^{t-1})[h_{t-1}(\theta^{t-1}, y^{t-1}) -$$

$$- A'_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1})B_{t-1}(\theta_{t-1})]; \qquad h_0(\theta_0) = Q_0^{-2}(\theta_0)B_0(\theta_0)$$

$$k_t(\theta^t, y^t) = k_{t-1}(\theta^{t-1}, y^{t-1}) + \tag{14.d}$$

$$+ \frac{1}{2}[h'_{t-1}(\theta^{t-1}, y^{t-1}) - B'_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1})A_{t-1}(\theta_{t-1})] \bullet N_t^{-1}(\theta^{t-1}) \bullet$$

$$\bullet [h_{t-1}(\theta^{t-1}, y^{t-1}) - A'_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1})B_{t-1}(\theta_{t-1})] -$$

$$- \frac{1}{2}B'_{t-1}(\theta_{t-1})Q_{t-1}^{-2}(\theta_{t-1})B_{t-1}(\theta_{t-1}) - \frac{1}{2}(y_t - D_t(\theta_t))'R_t^{-2}(\theta_t)(y_t - D_t(\theta_t)) \ ;$$

$$k_0(\theta_0) = - \frac{1}{2}B'_0(\theta_0)Q_0^{-2}(\theta_0)B_0(\theta_0)$$

REMARK 1.1 Using the matrix equality $Q^{-2} - Q^{-2}A[A'Q^{-2}A + M]^{-1}A'Q^{-2} = (Q^2 + A\,M^{-1}A')^{-1}$ and the positive definiteness of $Q_t^{-2}(\theta_t)$ it is easily seen by induction that $M_t(\theta^t)$, $N_t(\theta^{t-1})$ are positive definite. $\square$

PROOF It suffices to prove that $p(x_t, \theta^t \mid y^t) \propto p(\theta^t)\varphi_t(\theta^t)\ \exp\ \{- \frac{1}{2} x_t' M_t(\theta^t) x_t + x_t' h_t(\theta^t, y^t) + k_t(\theta^t, y^t)\}$ For simplicity we shall drop the arguments in the various matrices. We proceed by induction: From (10) it follows immediately that the result holds for $t = 0$. Assuming then that it holds for $t - 1 \geq 0$, we show it for t, using the recursive formula (7) with (8) and (9). The induction hypothesis and a straightforward "completion of the square" lead to

$$
\begin{aligned}
& p(\theta_t \mid \theta^{t-1}) \int g(x_t;\, A_{t-1}x_{t-1} + B_{t-1},\, Q_{t-1}^2) p(x_{t-1}, \theta^{t-1} \mid y^{t-1})\, \mathrm{d}x_{t-1} \propto \\
& \propto p(\theta_t \mid \theta^{t-1}) p(\theta^{t-1}) (\det Q_{t-1})^{-1} \varphi_{t-1}(\theta^{t-1}) \bullet \\
& \bullet \exp\left\{- \frac{1}{2} x_t' Q_{t-1}^{-2} x_t - \frac{1}{2} B_{t-1}' Q_{t-1}^{-2} B_{t-1} + x_t' Q_{t-1}^{-2} B_{t-1} + k_{t-1}(\theta^{t-1}, y^{t-1})\right\} \bullet \\
& \bullet \exp\left\{\frac{1}{2} [x_t' Q_{t-1}^{-2} A_{t-1} - B_{t-1}' Q_{t-1}^{-2} A_{t-1} + h_{t-1}(\theta^{t-1}, y^{t-1})] \bullet \right. \\
& \left. \bullet\ [A_{t-1}' Q_{t-1}^{-2} A_{t-1} + M_{t-1}]^{-1} [A_{t-1}' Q_{t-1}^{-2} x_t - A_{t-1}' Q_{t-1}^{-2} B_{t-1} + h_{t-1}(\theta^{t-1}, y^{t-1})] \right\} \bullet \\
& \bullet\ (\det [A_{t-1}' Q_{t-1}^{-2} A_{t-1} + M_{t-1}])^{-1/2}
\end{aligned}
$$

Multiplying this expression according to (7) by $p(y_t \mid x_t, \theta_t)$ given by (8) and collecting terms we get the desired result. □

Theorem 1.1 immediately yields a recursive algorithm for computing $p(x_t, \theta_t \mid y^t)$: It consists in computing recursively at each step t the relations (14) for each of the possible values of θ^t. Notice that the dimension of this procedure increases at each step by a factor equal to the number of possible values of θ. If θ_t is a constant parameter θ, the procedure becomes finite dimensional requiring the relations (14) to be computed at each step t only for the various possible values of θ.

2. GAUSSIAN WHITE NOISE; θ_t ABSOLUTELY CONTINUOUS (Approximate solution)

In this Section we still assume that the sequences $\{w_t\}$, $\{v_t\}$ in model (1) are independent standard white Gaussian, but we let θ_t take a continuum of possible values assuming, without sensible loss of generality, that for all t its distribution has compact support Θ. The main result is the approximation theorem, Theorem 2.1 below, that makes it possible to derive an approximation algorithm for computing $\Phi_f(t, y^t) = E\{f(x_t, \theta_t) \mid y^t\}$.

In the present case we can again use the recursive Bayes formula to obtain

$$p(x_t, \theta^t \mid y^t) \propto p(y_t \mid x_t, \theta^t, y^{t-1}) \cdot$$
$$\cdot \int p(x_t, \theta_t \mid x_{t-1}, \theta^{t-1}, y^{t-1}) p(x_{t-1}, \theta^{t-1} \mid y^{t-1}) \, \mathrm{d}x_{t-1} \tag{15}$$

with the same relations (8), (9), (10) except that this time all quantities appearing in (9), (10) represent densities. Theorem 1.1 also continues to hold with $\sum_{\theta^{t-1}}$ replaced by a (multiple) integral and with $p(\theta^t)$ representing the joint a-priori density of θ^t. Contrary to the previous Section however, the recursions (14) do not yield a computable algorithm as the possible θ^t range over a continuum of values. It makes therefore sense to look for an approximation algorithm.

We shall make the following

ASSUMPTION A.1 *The functions $A_{t-1}(\cdot)$, $B_{t-1}(\cdot)$ $Q_{t-1}(\cdot)$, $C_t(\cdot)$, $D_t(\cdot)$, $R_t(\cdot)$ as well as $p(\theta^t)$ are continuous and $f(x, \theta)$ is a polynomial in x with coefficients that are continuous functions of θ. Furthermore, for all t, the distribution of θ_t has compact support Θ.* □

Given an integer $N > 0$ consider a partition of the compact support Θ into sets $\Theta_i (i = 1, \ldots, r(N);\ \lim_{N \to \infty} r(N) = +\infty)$ such that ($l(\Theta_i)$ denotes the Lebesgue measure of Θ_i)

$$\max_i l(\Theta_i) < 1/N \tag{16}$$

Furthermore, for each $i \leq r(N)$ let θ^i be a fixed element in Θ_i. Given the joint density $p(\theta^t) = p_0(\theta_0) \cdots p(\theta_t \mid \theta^{t-1})$, $(t = 0, 1, \ldots T)$ define

$$p^{(N)}(\theta_0^{i_0}, \ldots, \theta_t^{i_t}) := \int_{\Theta_{i_0}} \cdots \int_{\Theta_{i_t}} p(\theta_0, \ldots, \theta_t) \, \mathrm{d}\theta_0 \cdots \mathrm{d}\theta_t \tag{17}$$

By its definition, $p^{(N)}(\theta_0^{i_0}, \ldots, \theta_t^{i_t})$ represents, for each t and each N, a joint probability distribution for the finite-valued discrete random vector $(\theta_0^{i_0}, \ldots, \theta_t^{i_t})$ $(i_0, \ldots, i_t \leq r(N))$.

Let $\theta_t^{(N)}$ denote the finite-state process with the joint distribution (17) and let $\theta_t^i (i = 1, \ldots, r(N))$ be its values. Furthermore, given y^t, let $p^{(N)}(x_t, \theta_t^i \mid y^t)$ be the joint conditional distribution of x_t, θ_t^i, assuming that the process θ_t in model (1) corresponds to $\theta_t^{(N)}$.

As an immediate consequence of the previous definitions, as well as that of a Riemann-Stieltjes integral, we can now state the following

LEMMA 2.1 *For all t, the discretized random vector $(\theta_0^{i_0}, \ldots, \theta_t^{i_t})$ converges weakly, as $N \to \infty$, to the continuous random vector $(\theta_0, \ldots, \theta_t)$, i.e. for all continuous (and by assumption A.1 also bounded) functions $\Psi(\theta_0, \ldots, \theta_t)$ we have*

$$E\{\Psi(\theta_0^{(N)}, \ldots, \theta_t^{(N)})\} :=$$

$$= \sum_{i_0, \ldots, i_t \leq r(N)} \Psi(\theta_0^{i_0}, \ldots, \theta_t^{i_t}) p^{(N)}(\theta_0^{i_0}, \ldots, \theta_t^{i_t}) \xrightarrow{N \to +\infty}$$

$$\to \int \cdots \int \Psi(\theta_0, \ldots, \theta_t) p(\theta_0, \ldots, \theta_t) \, \mathrm{d}\theta_0 \cdots \mathrm{d}\theta_t =$$

$$= E\{\Psi(\theta_0, \ldots, \theta_t)\} \quad . \qquad \square$$

We are now in a position to prove our approximation theorem that contains an additional result to be used in the next Section.

THEOREM 2.1 *Under assumption A.1 we have for a given $f(x, \theta)$ and all t and y^t*

$$\lim_{N \to \infty} \Phi_f^{(N)}(t, y^t) = \lim_{N \to \infty} \sum_{i=1}^{r(N)} \int f(x_t, \theta_t^i) p(x_t, \theta_t^i \mid y^t) \, \mathrm{d}x_t =$$

$$= \int \int f(x_t, \theta_t) p(x_t, \theta_t \mid y^t) \, \mathrm{d}x_t \mathrm{d}\theta_t = \Phi_f(t, y^t) \tag{18}$$

Furthermore,

$$\Phi_f^{(N)}(t, y^t) \quad \text{and} \quad \Phi_f(t, y^t)$$

are continuous functions of y^t. Finally, given y^t and an arbitrary sequence y_N^t converging to y^t,

$$\lim_{N \to \infty} \Phi_f^{(N)}(t, y_N^t) = \Phi_f(t, y^t) \tag{19}$$

PROOF Let, with the definitions (14),

$$\Psi_f(t; \theta^t, y^t) := \varphi_t(\theta^t) \bullet$$

$$\bullet \int f(x, \theta_t) \exp\left\{- \frac{1}{2} x' M_t(\theta^t) x + x' h_t(\theta^t, y^t) + k_t(\theta^t, y^t)\right\} \mathrm{d}x \tag{20}$$

From (14), assumption A.1 and the positive definiteness of $M_t(\theta^t)$ we have that $\Psi_f(t; \theta^t, y^t)$ is continuous in both θ^t and y^t and is bounded for each given value of y^t. From Lemma 2.1 we then have for all t and y^t

$$\sum_{i=1}^{r(N)} \sum_{i_0,\ldots,i_{t-1} \le r(N)} p^{(N)}(\theta_0^{i_0},\ldots,\theta_t^{i})\varphi_t(\theta_0^{i_0},\ldots,\theta_t^{i}) \cdot$$

$$\cdot \int f(x_t,\theta_t^i)\exp\Bigg\{-\frac{1}{2}x_t^{\prime}M_t(\theta_0^{i_0},\ldots,\theta_t^{i})x_t + x_t^{\prime}h_t(\theta_0^{i_0},\ldots,\theta^{i},y^t) +$$

$$+ k_t(\theta_0^{i_0},\ldots,\theta_t^{i},y^t)\Bigg\}\mathrm{d}x_t =$$

$$= E\{\Psi_f(t;(\theta^{(N)})^t,y^t)\} \overset{N\to\infty}{\longrightarrow} E\{\Psi_f(t;\theta^t,y^t)\} =$$

$$= \int\cdots\int p(\theta^t)\varphi_t(\theta^t)\int f(x_t,\theta_t)\exp\Bigg\{-\frac{1}{2}x_t^{\prime}M_t(\theta^t)x_t + x_t^{\prime}h_t(\theta^t,y^t) +$$

$$+ k_t(\theta^t,y^t)\Bigg\}\mathrm{d}x_t\mathrm{d}\theta^t \tag{21}$$

Notice now that for $f(x,\theta)\equiv 1$ the left- and rightmost terms in (21) are the (inverses of the) normalizing proportionality factors for relation (11) of Theorem 1.1 when applied to the processes $\theta_t^{(N)}$ and θ_t respectively. The first statement of the theorem now follows by using (11) of Theorem 1.1 and then applying (21) both for the given $f(x,\theta)$ and for $f(x,\theta)\equiv 1$ (the latter for the convergence of the normalizing factors).

Concerning the second statement of the Theorem notice that by the continuity of $\Psi_f(t;\theta^t,y^t)$ and the fact that θ^t takes values in a compact set, for y_1^t and y_2^t sufficiently close we have

$$|\Psi_f(t;\theta^t,y_1^t) - \Psi_f(t;\theta^t,y_2^t)| < \epsilon$$

uniformly in θ^t so that for any probability distribution function $p(\theta^t)$ also

$$|\int \Psi_f(t;\theta^t,y_1^t)\,\mathrm{d}p(\theta^t) - \int \Psi_f(t;\theta^t,y_2^t)\,\mathrm{d}p(\theta^t)| < \epsilon$$

From here, recalling that (for the given $p(\theta^t)$)

$$\Phi_f(t,y^t) = \left[\int \Psi_1(t;\theta^t,y^t)\,\mathrm{d}p(\theta^t)\right]^{-1} \cdot \int \Psi_f(t;\theta^t,y^t)\,\mathrm{d}p(\theta^t)$$

we obtain the continuity of $\Phi_f(t;y^t)$; analogously for $\Phi_f^{(N)}(t,y^t)$.

Coming to the last statment of the theorem we recall from [1; Thms. 5.2, 5.5] that, if functions $\{h_N(x)\}_{N\in\mathbb{N}}$ and $h(x)$ as well as probability measures $\{P_N\}$, P are such that

i) $\{x \mid \exists x_N \overset{N\to\infty}{\longrightarrow} x \text{ for which } h_N(x_N) \not\to h(x)\} = \emptyset$

ii) h_N, h are (uniformly in N) bounded

iii) $P_N \Rightarrow P$ (weak convergence)

then

$$\int h_N \mathrm{d}P_N \overset{N\to\infty}{\longrightarrow} \int h \mathrm{d}P \tag{22}$$

Given y^t, let then $y_N^t \overset{N\to\infty}{\longrightarrow} y^t$ and define, with some abuse of notation,

$$\Psi_f^{(N)}(t, \theta^t) := \Psi_f(t; \theta^t, y_N^t); \qquad \Psi_f(t, \theta^t) := \Psi_f(t; \theta^t, y^t) \tag{23}$$

The continuity of $\Psi_f(t; \theta^t, y^t)$ then implies that $\Psi_f^{(N)}(t, \cdot)$ and $\Psi_f(t, \cdot)$ satisfy i) and ii) above. Finally, let P_N and P be the measures induced by the processes $\{\theta_t^{(N)}\}$ and $\{\theta_t\}$ respectively, for which by Lemma 2.1 also iii) above holds. Relation (22) then translates into

$$E\{\Psi_f^{(N)}(t; (\theta^{(N)})^t)\} \overset{N\to\infty}{\longrightarrow} E\{\Psi_f(t, \theta^t)\}$$

From here the result follows recalling that

$$\Phi_f(t, y^t) = [E\{\Psi_1(t, \theta^t)\}]^{-1} \, E\{\Psi_f(t, \theta^t)\} \;\; ;$$

analogously for $\Phi_f^{(N)}(t, y^t)$. □

From Theorem 2.1 we obtain the following approximation algorithm to compute $\Phi_f(t, y^t) = E\{f(x_t, \theta_t) \mid y^t\}$ for a continuous function $f(x, \theta)$ that is a polynomial in x:

Step 1: Given $p_0(\theta_0)$ and $p(\theta_t \mid \theta_{t-1}, \ldots, \theta_0)$, compute according to (17) the joint a-priori distributions for the discretized finite-state process $\theta_t^{(N)}$.

Step 2: Compute $p(x_t, \theta_t^i \mid y^t)$, $(i = 1, \ldots, r(N))$ by means of the algorithm of Section I and determine $\Phi_f^{(N)}(t, y^t)$.

Theorem 2.1 guarantees that, by taking N sufficiently large, the approximation of $\Phi_f(t, y)$ by $\Phi_f^{(N)}(t, y)$ is arbitrarily close.

3. ROBUSTNESS WITH RESPECT TO THE A-PRIORI DISTRIBUTIONS IN THE MODEL

In this Section we consider model (5) with the following assumptions (Recall that $\Rightarrow$ denotes weak convergence)

A.2 $[w_0^\epsilon, \ldots, w_T^\epsilon, v_1^\epsilon, \ldots, v_T^\epsilon, \theta_0^\epsilon, \ldots, \theta_T^\epsilon] \overset{\epsilon \downarrow 0}{\Longrightarrow}$

$$\Rightarrow [w_0, \ldots, w_T, v_1, \ldots, v_T, \theta_0, \ldots, \theta_T]$$

where w_t and v_t are independent standard Gaussian random variables, $\theta_0, \ldots, \theta_T$ have a joint a-priori density $p(\theta^T)$ and $\{w_t\}$, $\{v_t\}$, $\{\theta_t\}$ are mutually independent.

A.3 Same as A.1 in addition to the boundeness of $f(x, \theta)$. □

Assumption A.3 implies that the functions $\Phi_f(t, y^t)$ and $\Phi_f^{(N)}(t, y^t)$ defined in (2) and (3) respectively, are (uniformly in N) bounded in addition to their continuity as shown in Theorem 2.1.

The result to be obtained in this Section is given in Theorem 3.1 below showing that if one uses the (exact or approximate) solution $\Phi_f(t, y^t)$ $(\Phi_f^{(N)}(t, y^t))$ computed for the ideal limit model (1) and uses it for the physical model (5), in the sense of estimating $f(x_t^\epsilon, \theta_t^\epsilon)$ by $\Phi_f(t, (y^\epsilon)^t)$ or $\Phi_f^{(N)}(t, (y^\epsilon)^t)$ depending on whether the limit process $\{\theta_t\}$ is finite-state or not, then this filtered estimate is, for small ϵ and large N, an almost (mean square) – optimal filter for $f(x_t^\epsilon, \theta_t^\epsilon)$ with respect to all alternative filters that are continuous and bounded functions of the past observations. In other words, with respect to the above alternative filters, the filter computed for the ideal model (1) is robust to small variations (in the sense of weak convergence) of the a-priori distributions of the noises $\{w_t\}$, $\{v_t\}$ and of the process $\{\theta_t\}$.

First we have the following Lemma 3.1, which shows that the ideal model (1) is indeed the limit model (for $\epsilon \downarrow 0$) of the family of models (5).

LEMMA 3.1 *Under assumptions A.2 and A.3 we have the weak convergence*

$$[x_0^\epsilon, \ldots, x_T^\epsilon, y_1^\epsilon, \ldots, y_T^\epsilon, \theta_0^\epsilon, \ldots, \theta_T^\epsilon] \overset{\epsilon \downarrow 0}{\Longrightarrow}$$

$$\Rightarrow [x_0, \ldots, x_T, y_1, \ldots, y_T, \theta_0, \ldots, \theta_t]$$

where the vector on the left corresponds to model (5) and the one on the right to model (1).

PROOF Let G be the function from $\mathbf{R}^{(T+1)(n+q)+Tm}$ into itself that according to model (1) expresses the vector $[x_0, \ldots, x_T, \; v_1, \ldots, v_T, \; \theta_0, \ldots, \theta_T]$ as a function of $[w_0, \ldots, w_T, \; v_1, \ldots, v_T, \; \theta_0, \ldots, \theta_T]$. This function remains the same for $[x_t, \; y_t, \; \theta_t]$ and $[w_t, \; v_t, \; \theta_t]$ being replaced respectively by $[x_t^\epsilon, \; y_t^\epsilon, \; \theta_t^\epsilon]$ and $[w_t^\epsilon, \; v_t^\epsilon, \; \theta_t^\epsilon]$ from model (5). The Lemma follows from the continuity of G. $\square$

In the statement of the following Theorem 3.1, to fix the ideas, we assume that the joint limit distribution $p(\theta^t)$ for θ^t is absolutely continuous so that as solution to the ideal limit model (1) one has to take the approximation $\Phi^{(N)}(t, \, y^t)$ computed according to the algorithm of Section II. If the limit process θ_t is already finite-state, then the limit model (1) has the exact solution $\Phi_f(t, \, y^t)$, computed according to the algorithm of Section I, and in the statement of Theorem 3.1 we can drop the dependence on - and the limit with respect to N.

THEOREM 3.1 *Under assumptions A.2 and A.3 we have for all $t = 1, \ldots, T$ and all continuous and bounded functions $F_t(y^t)$*

$$\lim_{\epsilon \downarrow 0} E\{\,|\, f(x_t^\epsilon, \theta_t^\epsilon) - F_t((y^\epsilon)^t) \,|^2\} \geq$$

$$\geq \lim_{\substack{\epsilon \downarrow 0 \\ N \to \infty}} E\{\,|\, f(x_t^\epsilon, \theta_t^\epsilon) - \Phi_f^{(N)}(t, \, (y^\epsilon)^t) \,|^2\}$$

PROOF From Lemma 3.1, assumption A.3, the *w.p.* 1 continuity and boundedness of $F_t(y^t)$, the fact that $\Phi_f(t, \, y^t) = E\{f(x_t, \theta_t) \,|\, y^t\}$ is mean-square optimal, the continuity and (uniform in N) boundedness of $\Phi_f^{(N)}(t, \, y^t)$ and $\Phi_f(t, \, y^t)$ (assumption A.3 and Theorem 2.1), and the last statement of Theorem 2.1 we have

$$\lim_{\epsilon \downarrow 0} E\{\,|\, f(x_t^\epsilon, \theta_t^\epsilon) - F((y^\epsilon)^t) \,|^2\} =$$

$$= E\{\,|\, f(x_t, \theta_t) - F(y^t) \,|^2\} \geq$$

$$\geq E\{\,|\, f(x_t, \theta_t) - \Phi_f(t, \, y^t) \,|^2\} =$$

$$= \lim_{\substack{\epsilon \downarrow 0 \\ N \to \infty}} E\{\,|\, f(x_t^\epsilon, \theta_t^\epsilon) - \Phi_f^{(N)}(t, \, (y^\epsilon)^t) \,|^2\}$$

where for the last equality we use the result from [1; Thms. 5.2, 5.5] recalled in the paragraph above (22). $\square$

REFERENCES

[1] Billingsley, P., *Convergence of Probability Measures*, John Wiley 1968.

[2] Di Masi, G.B., and W.J. Runggaldier, "On Measure Transformations for Combined Filtering and Parameter Estimation in Discrete Time", *Systems & Control Letters*, **2**, 1982, pp. 57–62.

[3] Kushner, H.J., *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press 1977.

[4] Kushner, H.J., and H. Huang, "Approximate and Limit Results for Nonlinear Filters with Wide Bandwidth Observation Noise", *Stochastics*, 1986, **16**, pp. 65–96.

[5] Kushner, H.J., and W. Runggaldier, "Filtering and Control for Wide-Bandwidth Noise Driven Systems". *IEEE Transactions AC-32*, No. 2, 1987, pp. 123–133, also LCDS Rept. 86–8, Brown University.

Authors' permanent address:

Dipartimento di Matematica Pura ed Applicata
Università di Padova
Via Belzoni 7
I-35131 Padova, Italy