Journal of Mathematical Systems, Estimation, and Control
Vol. 6, No. 1, 1996, pp. 1–13

State and Parameter Estimation for Linear Systems*

Robert J. Elliott† John B. Moore

Abstract

The standard linear signal, linear observations model is considered, in which the coefficient matrices depend on unknown parameters. Using the reference probability method an explicit recursive formula is obtained for the unnormalized conditional distribution of the signal and unknown parameters, given the observations.

Key words: recursive filter, unnormalized density, reference probability, parameter estimation

1 Introduction

The standard model for linear, discrete time signal and observations is considered, in which the coefficient matrices depend on unknown, time varying parameters. An explicit recursive expression is obtained for the unnormalized, conditional expectation, given the observations, of the state and the parameters. The method are an adaptation of those in [2], and are based on the introduction of an equivalent probability measure under which the observation random variables are independent of both the signal and unknown parameters. Our construction of the equivalent measure is explicit and the recursion has a simple form.

Finally, we consider the parameter estimation problem for a general ARMAX model. In this case it is remarkable that the recursive formulae for the unnormalized densities do not involve any integration.

*Received June 14, 1993; received in final form September 15, 1993. Summary appeared in Volume 6, Number 1, 1996.

†The first author wishes to thank the Natural Sciences and Engineering Research Council of Canada for its continued support, and the Department of Systems Engineering, Australian National University, for its hospitality.

2 Linear Dynamics and Parameters

All processes are defined initially on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The discrete time model we wish to discuss has the form

$$x_{k+1} = A(\theta_k^1)x_k + B(\theta_k^2)v_{k+1} \quad (1)$$
$$y_k = C(\theta_k^3)x_k + D(\theta_k^4)w_k. \quad (2)$$

Here $k \in \mathbb{N}$ and x_0, or its distribution, are known. The signal process x_k takes values in some Euclidean space $\mathbb{R}^d$ while the observation process y_k takes values in $\mathbb{R}^q$. $\{v_\ell\}$, $\ell \in \mathbb{N}$, is a sequence of *i.i.d.* random variables, with density functions ψ, and v_ℓ has values in $\mathbb{R}^d$. Similarly, $\{w_\ell\}$, $\ell \in \mathbb{N}$, is a sequence of *i.i.d.* random variables with strictly positive density function ϕ, and w_ℓ also takes values in $\mathbb{R}^q$, that is w_ℓ has the same dimensions as y_ℓ. The matrices $A(\theta_\cdot^1)$, $B(\theta_\cdot^2)$, $C(\theta_\cdot^3)$ and $D(\theta_\cdot^4)$ have appropriate dimensions and depend on the parameters $\theta^1, \dots, \theta^4$.

For simplicity we suppose the parameters θ_k^i, $i = 1,2,3,4$, are real valued and satisfy the dynamic equations

$$\theta_{k+1}^i = \alpha^i \theta_k^i + \nu_{k+1}^i.$$

Here either θ_0^i, or its distribution, is known.

The α^i, $i = 1,2,3,4$, are real constants and $\{\nu_\ell^i\}$ is a sequence of *i.i.d.* random variables with densities ρ^i. Finally, we suppose the matrices $B(r)$ and $D(r)$ are nonsingular for all $r \in \mathbb{R}$.

Notation 2.1 *Write* $\mathcal{G}_{k+1}^0 = \sigma\{\theta_\ell^i,\ 1 \le \ell \le k,\ i = 1,2,\ \theta_\ell^j,\ 1 \le \ell \le k+1,\ j = 3,4,\ x_0, x_1, \dots, x_{k+1},\ y_1, \dots, y_k, \theta_0^i\}$, $\mathcal{Y}_k^0 = \sigma\{y_1, \dots, y_k\}$. $\{\mathcal{G}_k\}$ *and* $\{\mathcal{Y}_k\}$, $k \in \mathbb{N}$, *are the complete filtrations generated by the completions of* $\mathcal{G}_k^0$ *and* $\mathcal{Y}_k^0$, *respectively.*

Remarks 2.2 The above conditions can be modified. For example, the parameters θ^i can be vector valued.

2.0.1 Measure change and estimation

Write

$$\gamma_k = \gamma_k(x_k, w_k, \theta_k^i) = |\det D(\theta^4)| \frac{\phi(y_k)}{\phi(w_k)},$$

and

$$\Lambda_k = \prod_{\ell=1}^{k} \gamma_\ell.$$

A new probability measure $\overline{P}$ can be defined on $\left(\Omega, \bigvee_{\ell=1}^{\infty} \mathcal{G}_\ell\right)$ by setting the restriction to $\mathcal{G}_k$ of the Radon-Nikodym derivative $d\overline{P}/dP$ equal to Λ_k.

Lemma 2.3 *Under $\overline{P}$ the random variables $\{y_\ell\}$, $\ell \in \mathbb{N}$, are i.i.d. with density function ϕ.*

Proof: For $t \in \mathbb{R}^q$ the event $\{y_k \le t\} = \{y_k^i \le t^i,\, i = 1, \dots, q\}$. Then

$$\begin{aligned}\overline{P}(y_k \le t \mid \mathcal{G}_k) &= \overline{E}[I(y_k \le t) \mid \mathcal{G}_k] \\ &= E[\Lambda_k I(y_k \le t) \mid \mathcal{G}_k]/E[\Lambda_k \mid \mathcal{G}_k] \\ &= E[\gamma_k I(y_k \le t) \mid \mathcal{G}_k]/E[\gamma_k \mid \mathcal{G}_k].\end{aligned}$$

Now

$$E[\gamma_k \mid \mathcal{G}_k] = \int_{\mathbb{R}^q} |\det D(\theta_k^4)| \phi(y_k) dw_k = 1$$

so

$$\begin{aligned}\overline{P}(y_k \le t \mid \mathcal{G}_k) &= \int_{\mathbb{R}^q} I(y_k \le t) |\det D(\theta_k^4)| \phi(y_k) dw_k \\ &= \int_{-\infty}^{t^1} \dots \int_{-\infty}^{t^q} \phi(y_k) dy_k.\end{aligned}$$

This shows y_k is $\overline{P}$-independent of $\mathcal{G}_k$ and the result follows. ∎

Remarks 2.4 Suppose we now start with a probability measure $\overline{P}$ on $\left(\Omega, \bigvee_{\ell=1}^{\infty} \mathcal{G}_\ell\right)$ such that under $\overline{P}$:

1. $\{y_k\}$, $k \in \mathbb{N}$, is a sequence of *i.i.d.* $\mathbb{R}^q$ valued random variables with positive density function ϕ,

2. $\{\theta_k^i\}$, $k \in \mathbb{N}$, $1 \le i \le 4$, are real variables satisfying $\theta_{k+1}^i = \alpha^i \theta_k^i + \nu_{k+1}^i$ where the $\{\nu_k^i\}$ are sequences of *i.i.d.* random variables with density functions ρ^i,

3. $\{x_k\}$, $k \in \mathbb{N}$, is a sequence of $\mathbb{R}^d$ valued random variables satisfying

$$x_k = A(\theta_{k-1}^1) x_{k-1} + B(\theta_{k-1}^2) v_k$$

   where the $\{v_k\}$, $k \in \mathbb{N}$, is a sequence of *i.i.d.* random variables with density ψ.

Note in particular that under $\overline{P}$ the x_ℓ and y_ℓ are independent. We now construct, by an inverse procedure, a probability measure P, such that under P, $\{w_\ell\}$, $\ell \in \mathbb{N}$, is a sequence of *i.i.d.* random variables with density ϕ, where $w_k := D(\theta_k^4)^{-1}(y_k - C(\theta_k^3) x_k)$.

To construct P from $\overline{P}$ write

$$\bar{\gamma}_k = \bar{\gamma}_k(x_k, y_k, \theta_k^3, \theta_k^4) = |\det D(\theta_k^4)|^{-1} \frac{\phi(w_k)}{\phi(y_k)}$$

and $\overline{\Lambda}_k = \prod_{\ell=1}^k \bar{\gamma}_\ell$. P is defined by putting the restriction to $\mathcal{G}_k$ of the Radon-Nikodym derivative $dP/d\overline{P}$ equal to $\overline{\Lambda}_k$. The existence of $\overline{P}$ is a consequence of Kolmogorov's theorem. (See Shiryaev [4]).

Lemma 2.5 *Under P, $\{w_k\}$, $k \in \mathbb{N}$, is a sequence of i.i.d. random variables with density ϕ.*

Proof: For $t \in \mathbb{R}^q$ the event $\{w_k \leq t\} = \{w_k^i \leq t^i,\ i = 1, \ldots, q\}$. Then

$$\begin{aligned} P(w_k \leq t \mid \mathcal{G}_k) &= E[I(w_k \leq t) \mid \mathcal{G}_k] \\ &= \overline{E}[\overline{\Lambda}_k I(w_k \leq t) \mid \mathcal{G}_k]/\overline{E}[\overline{\Lambda}_k \mid \mathcal{G}_k] \\ &= \overline{E}[\bar{\gamma}_k I(w_k \leq t) \mid \mathcal{G}_k]/\overline{E}[\bar{\gamma}_k \mid \mathcal{G}_k] \end{aligned}$$

and, as before, this is

$$\begin{aligned} &= \overline{E}[\bar{\gamma}_k I(w_k \leq t) \mid \mathcal{G}_k] \\ &= \int_{\mathbb{R}^q} I(w_k \leq t) |\det D(\theta^4)|^{-1} \phi(w_k) dy_k \\ &= \int_{-\infty}^{t^1} \cdots \int_{-\infty}^{t^q} \phi(w) dw. \end{aligned}$$

The result follows. ∎

2.0.2 Unnormalized estimates

A version of Bayes' theorem states that for a $\mathcal{G}$-adapted sequence Φ_k

$$E[\Phi_k \mid \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k \Phi_k \mid \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k \mid \mathcal{Y}_k]}. \tag{3}$$

This identity indicates why the unnormalized, conditional expectation $\overline{E}[\overline{\Lambda}_k \Phi_k \mid \mathcal{Y}_k]$ is investigated. Write $q_k(z, \theta)$, $k \in \mathbb{N}$, for the unnormalized conditional density such that

$$\overline{E}\Big[\overline{\Lambda}_k I(x_k \in dz) \prod_{i=1}^4 I(\theta^i \in d\theta^i) \mid \mathcal{Y}_k\Big] = q_k(z, \theta) dz d\theta^1 d\theta^2 d\theta^3 d\theta^4.$$

(The existence of q_k is discussed in Remarks 2.7)

We now derive a recursive update for q_k. The normalized conditional density

$$p_k(z, \theta) dz d\theta^1 d\theta^2 d\theta^3 d\theta^4 = E\Big[I(x_k \in dz) \prod_{i=1}^4 I(\theta^i \in d\theta^i) \mid \mathcal{Y}_k\Big]$$

is given by

$$p_k(z,\theta) = q_k(z,\theta)/\int_{\mathbb{R}^d}\int_{\mathbb{R}^4} q_k(\xi,\lambda)d\xi d\lambda^1 d\lambda^2 d\lambda^3 d\lambda^4.$$

Theorem 2.6 *For $k \in \mathbb{N}$*

$$\boxed{\begin{array}{l} q_{k+1}(z,\lambda) = \\ \iint \Delta_1(y_{k+1}, z, \lambda, \xi, \sigma)\psi(B(\sigma^2)^{-1}(z - A(\sigma^1)\xi))q_k(\xi,\sigma)d\xi d\sigma \end{array}} \quad (4)$$

where

$$\begin{aligned}
&\Delta_1(y_{k+1}, z, \lambda, \xi, \sigma) = \\
&\qquad |\det D(\lambda^4)|^{-1}\phi(D(\lambda^4)^{-1}(y_{k+1} - C(\lambda^3)z)) \times \\
&\qquad |\det B(\sigma^2)|^{-1}\prod_{i=1}^4 \rho^i(\lambda^i - \alpha^i\sigma^i)\phi(y_{k+1})^{-1}
\end{aligned}$$

Proof: Suppose $f : \mathbb{R}^d \times \mathbb{R}^4 \to \mathbb{R}$ is any Borel test function. Then

$$\begin{aligned}
&\overline{E}[f(x_{k+1}, \theta^1_{k+1}, \theta^2_{k+1}, \theta^3_{k+1}, \theta^4_{k+1})\overline{\Lambda}_{k+1} \mid \mathcal{Y}_{k+1}] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^4} f(z, \lambda^1, \lambda^2, \lambda^3, \lambda^4)q_{k+1}(z, \lambda^1, \lambda^2, \lambda^3, \lambda^4)dz d\lambda^1 d\lambda^2 d\lambda^3 d\lambda^4 \\
&= \overline{E}\big[f(A(\theta^1_k)x_k + B(\theta^2_k)v_{k+1}, \alpha^i\theta^i_k + \nu^i_{k+1})\overline{\Lambda}_k |\det D(\theta^4_{k+1})|^{-1} \\
&\phi(D(\theta^4_{k+1})^{-1}(y_{k+1} - C(\theta^3_{k+1})x_{k+1})) \mid \mathcal{Y}_{k+1}\big]\phi(y_{k+1})^{-1}.
\end{aligned}$$

Substituting for the remaining x_{k+1} and θ^i_{k+1} this is

$$\begin{aligned}
&= \overline{E}\Big[\iiint f(A(\theta^1_k)x_k + B(\theta^2_k)w, \alpha^i\theta^i_k + \nu^i)\overline{\Lambda}_k |\det D(\alpha^4\theta^4_k + \nu^4)|^{-1} \\
&\phi(D(\alpha^4\theta^4_k + \nu^4)^{-1}(y_{k+1} - C(\alpha^3\theta^3_k + \nu^3)(A(\theta^1_k)x_k + B(\theta^2_k)w)) \\
&\psi(w)\rho^1(\nu^1)\rho^2(\nu^2)\rho^3(\nu^3)\rho^4(\nu^4)dw d\nu^1 d\nu^2 d\nu^3 d\nu^4 \mid \mathcal{Y}_{k+1}\Big]\phi(y_{k+1})^{-1}.
\end{aligned}$$

The y_ℓ are independent, so this is

$$\begin{aligned}
&= \iiiint f(A(\lambda^1)z + B(\lambda^2)w, \alpha^i\lambda^i + \nu^i)|\det D(\alpha^4\lambda^4 + \nu^4)|^{-1} \\
&\phi(D(\alpha^4\lambda^4 + \nu^4))^{-1}(y_{k+1} - C(\alpha^3\lambda^3 + \nu^3)(A(\lambda^1)z + B(\lambda^2)w)) \\
&\psi(w)\prod_{i=1}^4 \rho^i(\nu^i)q_k(z, \lambda^1, \lambda^2, \lambda^3, \lambda^4) \\
&dz dw d\lambda^1 d\lambda^2 d\lambda^3 d\lambda^4 d\nu^1 d\nu^2 d\nu^3 d\nu^4 \phi(y_{k+1})^{-1}.
\end{aligned}$$

Write $\xi = A(\lambda^1)z + B(\lambda^2)w$ and $\sigma^i = \alpha^i\lambda^i + \nu^i$, $1 \le i \le 4$.

Then $dzdw\prod_{i=1}^4(d\lambda^i d\nu^i) = |\det B(\lambda^2)|^{-1}dzd\xi\prod_{i=1}^4(d\lambda^i d\sigma^i)$, and the above integral equals

$$\int\!\!\int\!\!\int\!\!\int f(\xi,\sigma)|\det D(\sigma^4)|^{-1}\phi(D(\sigma^4)^{-1}(y_{k+1}-C(\sigma^3)\xi))$$
$$\psi(B(\lambda^2)^{-1}(\xi - A(\lambda^1)z))|\det B(\lambda^2)|^{-1}$$
$$\prod_{i=1}^4 \rho^i(\sigma^i - \alpha^i\lambda^i)q_k(z,\lambda^1,\lambda^2,\lambda^3,\lambda^4)dzd\xi\prod_{i=1}^4(d\lambda^i d\sigma^1)\phi(y_{k+1})^{-1}.$$

This identity holds for all Borel test functions f, so the result follows:

$$q_{k+1}(z,\lambda^1,\lambda^2,\lambda^3,\lambda^4) = |\det D(\lambda^4)|^{-1}\phi(D(\lambda^4)^{-1}(y_{k+1}-C(\lambda^3)z))$$
$$\int\!\!\int \psi(B(\sigma^2)^{-1}(z - A(\sigma^1)\xi))|\det B(\sigma^2)|^{-1}$$
$$\prod_{i=1}^4 \rho^i(\lambda^i - \alpha^i\sigma^i)q_k(\xi,\sigma^1,\sigma^2,\sigma^3,\sigma^4)d\xi d\sigma^1 d\sigma^2 d\sigma^3 d\sigma^4\phi(y_{k+1})^{-1}.$$

■

Remarks 2.7 Suppose $\pi(z)$ is the density of x_0, and $\rho_0(\lambda^1,\lambda^2,\lambda^3,\lambda^4)$ is the density of $(\theta_0^1,\theta_0^2,\theta_0^3,\theta_0^4)$. Then $q_0(z,\lambda^1,\lambda^2,\lambda^3,\lambda^4) = \pi(z)\rho_0(\lambda)$ and updated estimates are obtained by substituting in (4).

Even if the prior estimates for x_0 or θ_0^i, $1 \le i \le 4$, are delta functions, the proof of Theorem 2.6 gives a function for $q_1(z,\lambda)$. In fact, if $\Pi(z) = \delta(x_0)$ and $\rho_0(\lambda) = \delta(\theta_0^1,\theta_0^2,\theta_0^3,\theta_0^4)$ then we see

$$\begin{aligned} q_1(z,\lambda) &= |\det D(\lambda^4)|^{-1}\phi(D(\lambda^4)^{-1}(y_1 - C(\lambda^3)z)) \times \\ &\quad \psi(B(\theta_0^2)^{-1}(z - A(\theta_0^1)x_0)) \\ &\quad |\det B(\theta_0^2)|^{-1}\prod_{i=\ell}^4 \rho^i(\lambda^i - \alpha_i\theta_0^i), \end{aligned}$$

and further updates follow from (4).

If there are no dynamics in one of the parameters, so that $\alpha^i = 1$ and ρ^i is the delta mass at 0 giving $\theta_k^i = \theta_{k-1}^i$, $k \in \mathbb{N}$, then care must be taken with the choice of a prior distribution for θ^i. In fact, if $\rho_0(\theta)^i$ is the prior distribution, the above procedure gives an unnormalized conditional density $q_k^\theta(z,\lambda^j, j \ne i)$ for each possible value of θ, and $q_k(z,\lambda^1,\lambda^2,\lambda^3,\lambda^4) = q_k^{\lambda^i}(z,\lambda^j, j \ne i)\rho_0(\lambda^i)$.

3 The ARMAX Model

We now indicate how the general ARMAX model can be treated. Suppose $\{v_\ell\}$, $\ell \in \mathbb{N}$, is a sequence of (real) *i.i.d.* random variables with density ψ. Write

$$\begin{aligned}\theta^1 &= (a_1, \ldots, a_{r_1}) \in \mathbb{R}^{r_1} \\ \theta^2 &= (w_1, \ldots, w_{r_2}) \in \mathbb{R}^{r_2} \\ \theta^3 &= (c_1, \ldots, c_{r_3}) \in \mathbb{R}^{r_3}, \quad c_{r_3} \neq 0, \end{aligned} \tag{5}$$

for the unknown coefficient vectors, or parameters. An ARMAX system $\{y_\ell\}$ with exogenous inputs $\{u_\ell\}$, $\ell \in \mathbb{N}$, is then given by equations of the form

$$\begin{aligned} & y_{k+1} + a_1 y_k + \ldots + a_{r_1} y_{k+1-r_1} \\ & \quad = w_1 u_k + \ldots + w_{r_2} u_{k+1-r_2} + c_1 v_k + \ldots + c_{r_3} v_{k+1-r_3} + v_{k+1}. \end{aligned} \tag{6}$$

Write x_k for the column vector

$$(y_k, \ldots, y_{k-r_1}, u_k, \ldots, u_{k-r_2}, v_k, \ldots, v_{k-r_3})' \in \mathbb{R}^{r_1+r_2+r_3}.$$

Suppose $A(\theta)$ is the $(r_1+r_2+r_3) \times (r_1+r_2+r_3)$ matrix having $(-\theta^1, \theta^2, \theta^3)$ for its first row and 1 on the subdiagonal, with zeros elsewhere on other rows, except the (r_1+1) and (r_1+r_2+1) rows which are $0 \in \mathbb{R}^{r_1+r_2+r_3}$. B will denote the unit column vector in $\mathbb{R}^{r_1+r_2+r_3}$ having one in the (r_1) position and zeros elsewhere. C will denote the column vector in $\mathbb{R}^{r_1+r_2+r_3}$ having 1 in the first and (r_1+r_2+1) position and zeros elsewhere. The values of the u_ℓ are known exogenously; for example, if the variables u_ℓ are control variables u_k will depend on the values of $y_1, \ldots, y_k$. System (6) can then be written:

$$\begin{aligned} x_{k+1} &= A(\theta) x_k + B u_{k+1} + C v_{k+1} & (7) \\ y_{k+1} &= \langle \theta, x_k \rangle + v_{k+1}. & (8) \end{aligned}$$

Here $\theta = (-\theta^1, \theta^2, \theta^3)$ and $\langle\ ,\ \rangle$ denotes the scalar product in $\mathbb{R}^{r_1+r_2+r_3}$. Representation (7), (8) is not a minimal representation; see for example Anderson and Moore [1]. However, it suffices for our discussion. Notice the same noise term v_{k+1} appears in (8) and (7). This is circumvented by substituting in (7) to obtain

$$x_{k+1} = (A(\theta) - C \otimes \theta) x_k + B u_{k+1} + C y_{k+1} \tag{9}$$

together with

$$y_{k+1} = \langle \theta, x_k \rangle + v_{k+1}. \tag{10}$$

Write $\mathcal{Y}_k^0 = \sigma\{y_1, \ldots, y_k\}$ and $\{\mathcal{Y}_\ell\}$, $\ell \in \mathbb{N}$, for the corresponding complete filtration. Write $\overline{x}_k$ for the column vector $(v_k, \ldots, v_{k-r_3})' \in \mathbb{R}^{r_3}$ so that $x'_k = (y_k, \ldots,\ y_{k-r_1},\ u_k, \ldots, u_{k-r_2},\ \overline{x}')$, and, given $\mathcal{Y}_k$, the $\overline{x}_k$ are the unknown components of x_k. Let $\alpha_k = y_k + \langle\theta^1, (y_k, \ldots, y_{k-r_1})'\rangle - \langle\theta^2, (u_k, \ldots, u_{k-r_2})'\rangle$ and write $\underline{\alpha}_k$ for the vector $\alpha_k \overline{C}$ where $\overline{C} = (1, 0, \ldots, 0)' \in \mathbb{R}^{r_3}$. Then with $, (\theta^3)$ equal to the $r_3 \times r_3$ matrix

$$, (\theta^3) = \begin{pmatrix} -c_1 & -c_2 & \ldots & -c_{r_3} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots 1 & 0 \end{pmatrix}$$

we have $\overline{x}_{k+1} = , (\theta^3)\overline{x}_k + \underline{\alpha}_{k+1}$. Recall the model is chosen so that $c_{r_3} \neq 0$; then

$$, (\theta^3)^{-1} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \ldots \\ 0 & 0 & 0 & \ldots 0 & 1 \\ -\frac{1}{c_{r_3}} & -\frac{c_1}{c_{r_3}} & -\frac{c_2}{c_{r_3}} & \ldots & -\frac{c_{r_3-1}}{c_{r_3}} \end{pmatrix}$$

Given $\mathcal{Y}_k$ we wish to determine the unnormalized conditional density of $\overline{x}_k$ and θ. Again, we suppose the processes are defined on $(\Omega, \mathcal{F}, \overline{P})$ under which $\{y_\ell\}$, $\ell \in \mathbb{N}$, is a sequence of *i.i.d.* random variables with strictly positive densities ϕ. P is defined by putting the restriction of $\dfrac{dP}{d\overline{P}}$ to $\mathcal{G}_k$ equal to $\overline{\Lambda}_k$. Here $\overline{\Lambda}_k = \prod_{\ell=1}^k \bar{\gamma}_\ell$ where $\bar{\gamma}_\ell = \phi(y_{\ell+1} - \langle\theta, x_\ell\rangle)/\phi(y_{\ell+1})$. Write $q_k(\xi, \lambda)$ for the unnormalized conditional density such that

$$\overline{E}[I(\overline{x}_k \in d\xi)I(\theta \in d\lambda)\overline{\Lambda}_k \mid \mathcal{Y}_k] = q_k(\xi, \lambda)d\xi d\lambda.$$

Consider, therefore, any Borel test functions $f : \mathbb{R}^{r_3} \to \mathbb{R}$ and $g : \mathbb{R}^{r_1+r_2+r_3} \to \mathbb{R}$. Write $\bar{y}_k = (y_k, \ldots, y_{k-r_1})'$ and $\overline{u}_k = (u_k, \ldots, u_{k-r_2})'$. The same arguments to those used in Section 4 lead us to consider

$$\begin{aligned} &\overline{E}[f(\overline{x}_{k+1})g(\theta)\overline{\Lambda}_{k+1} \mid \mathcal{Y}_{k+1}] \\ &\quad = \int\int f(\xi)g(\lambda)q_{k+1}(\xi, \lambda)d\xi d\lambda \\ &\quad = \overline{E}\Big[f(, (\theta^3)\overline{x}_k + \underline{\alpha}_{k+1})g(\theta)\overline{\Lambda}_k\phi\big(y_{k+1} + \langle\theta^1, \bar{y}_k\rangle \\ &\quad -\langle\theta^2, \overline{u}_k\rangle - \langle\theta^3, \overline{x}_k\rangle\big) \mid \mathcal{Y}_{k+1}\Big]\phi(y_{k+1})^{-1} \\ &\quad = \int\int f(, (\lambda^3)z + \underline{\alpha}_{k+1})g(\lambda)\phi\big(y_{k+1} + \langle\lambda^1, \bar{y}_k\rangle \\ &\quad -\langle\lambda^2, \overline{u}_k\rangle - \langle\lambda^3, z\rangle\big)q_k(z, \lambda)dzd\lambda\phi(y_{k+1})^{-1}. \qquad (11) \end{aligned}$$

Write

$$\xi = \,,\, (\lambda^3)z + \big(y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle\big)\overline{C}.$$

Then

$$z = \,,\, (\lambda^3)^{-1}\big(\xi - (y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle)\overline{C}\big)$$

and

$$dz d\lambda = dz d\lambda^1 d\lambda^2 d\lambda^3 = \,,\, (\lambda^3) d\xi d\lambda.$$

Substituting in (11) we have

$$\int\!\!\int f(\xi) g(\lambda) q_{k+1}(\xi, \lambda) d\xi d\lambda = \int\!\!\int f(\xi) g(\lambda) \phi\big(y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle$$
$$-\big\langle \lambda^3, \,,\, (\lambda^3)^{-1}\big(\xi - (y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle)\overline{C}\big\rangle\big)\phi(y_{k+1})^{-1}$$
$$q_k\big(,\, (\lambda^3)^{-1}(\xi - (y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle)\overline{C}\big), \lambda\big), \,(\lambda^3)^{-1} d\xi d\lambda$$

We, therefore, have the following remarkable result for updating the unnormalized, conditional density of $\overline{x}_k$ and θ, given $\mathcal{Y}_k$:

Theorem 3.1

$$\boxed{\begin{array}{l} q_{k+1}(\xi, \lambda) = \Delta_2(y_{k+1}, \bar{y}_k, \overline{u}_k, \xi, \lambda) \\ q_k(,\, (\lambda^3)^{-1}(\xi - (y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, u_k\rangle)\overline{C}), \lambda). \end{array}} \tag{12}$$

where

$$\Delta_2(y_{k+1}, \bar{y}_k, \overline{u}_k, \xi, \lambda) = \frac{\phi(\xi_1)}{\phi(y_{k+1})} \,,\, (\lambda^3)^{-1}$$

and

$$\xi_1 = y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle$$
$$-\langle \lambda^3, \,,\, (\lambda^3)^{-1}(\xi - (y_{k+1} + \langle \lambda^1, \bar{y}_k\rangle - \langle \lambda^2, \overline{u}_k\rangle)\overline{C}\rangle$$

Remark 3.2 This does not involve any integration.

If $\pi_0(\xi)$ is the prior density of x_0 and $\rho_0(\lambda)$ that for λ then $q_0(\xi, \lambda) = \pi_0(\xi)\rho_0(\lambda)$. The prior density must reflect information known about x_0 and θ, and not be just a guess. Because no dynamics or noise enter the parameters θ the estimation problem can be treated as though θ is fixed, followed by an averaging over θ using the density $\rho_0(\lambda)$.

4 A Markov Chain Observed in Coloured Noise

In this section we extend the above results to observations with coloured noise. We suppose the signal model parameters depend on some parameter θ which takes values in a measure space (Θ, β, u). The value of θ is unknown, and we suppose it is constant. Then for $1 \leq i, j \leq N$, write $\mathcal{F}_k^0$ for the σ-field generated by $X_0, X_1, \ldots, X_k$ and θ and $\{\mathcal{F}_k\}, k \in \mathbb{N}$, for the filtration generated by $\mathcal{F}_k^0$.

$$\begin{aligned} a_{ij}(\theta) &= P(X_{k+1} = e_i \mid X_k = e_j) \\ &= P(X_1 = e_i \mid X_0 = e_j). \end{aligned}$$

Write $A(\theta)$ for the $N \times N$ matrix $(a_{ij}(\theta))$, $1 \leq i, j \leq N$. Then

$$X_{k+1} = A(\theta) X_k + V_{k+1} \tag{13}$$

where $E[V_{k+1} \mid \mathcal{F}_k] = 0$.

We suppose the chain X is not observed directly; rather there is an observation process $\{y_\ell\}$, $\ell \in \mathbb{N}$, which for simplicity we suppose is real valued. The process y has the form

$$y_{k+1} = c(\theta, X_{k+1}) + d_1(\theta) w_k + \ldots + d_r(\theta) w_{k+1-r} + w_{k+1}. \tag{14}$$

Here $\{w_k\}$, $k \in \mathbb{N}$, is a sequence of *i.i.d.* random variables with nonzero density function ϕ. (The extension to time varying densities ϕ_k is immediate.) Suppose $d_r(\theta) \neq 0$.

$c(\theta, X_k)$ is a function, depending on a parameter θ, and the state X_k. Because X_k is always one of the unit vectors e_i the function $c(\theta, .)$ is determined by a vector

$$c(\theta) = \big(c_1(\theta), c_2(\theta), \ldots, c_N(\theta)\big)$$

and

$$c(\theta, X_k) = \langle c(\theta), X_k \rangle,$$

where $\langle\ ,\ \rangle$ denotes the inner product in Euclidean space.

On $(\Omega, \mathcal{F}, P)$ our observation process, therefore, has the form

$$y_{k+1} = c(\theta, X_{k+1}) + d_1(\theta) w_k + \ldots + d_r(\theta) w_{n-r} + w_{k+1}. \tag{15}$$

Write $\overline{x}_{k+1} = (w_{k+1}, w_k, \ldots, w_{n-r+1})' \in \mathbb{R}^r$, $D = (1, 0, \ldots, 0)' \in \mathbb{R}^r$,

$$, (\theta) = \begin{pmatrix} -d_1(\theta) & -d_2(\theta) & \ldots & -d_r(\theta) \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ 0 & 0 & \cdot\cdot 1 & 0 \end{pmatrix}$$

Then

$$\overline{x}_{k+1} = , (\theta) \overline{x}_k + D(y_{k+1} - \langle c(\theta), X_{k+1} \rangle)$$

and

$$y_{k+1} = \langle c(\theta), X_{k+1} \rangle + \langle d(\theta), \overline{x}_k \rangle + w_{k+1}.$$

The unobserved components are, therefore, X_{k+1}, $\overline{x}_k$, θ.

Again, because $d_r(\theta) \neq 0$

$$, (\theta)^{-1} = d_r^{-1}(\theta) \begin{pmatrix} 0 & d_r(\theta) & 0 & \dots & 0' \\ 0 & 0 & d_r(\theta) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & d_r(\theta) \\ -1 & -d_1(\theta) & -d_2(\theta) & \dots & -d_{r-1}(\theta) \end{pmatrix}$$

Suppose f and h are arbitrary real-valued test functions. The same arguments again lead us to consider

$$\overline{E}[f(\overline{x}_{k+1})h(\theta)\langle X_{k+1}, e_i\rangle\overline{\Lambda}_{k+1} \mid \mathcal{Y}_{k+1}] = \int\int f(\xi)h(u)q^i_{k+1}(\xi, u)d\xi d\lambda(u), \tag{16}$$

where $q^i_{k+1}(\xi, u)$ is the unnormalized conditional density such that

$$\overline{E}[\langle X_{k+1}, e_i\rangle I(\theta \in d\theta)I(\overline{x}_{k+1} \in dz)\overline{\Lambda}_{k+1} \mid \mathcal{Y}_{k+1}] \\ = q^i_{k+1}(z, \theta)dzd\theta$$

Then (16) equals

$$\begin{aligned} & \overline{E}\Big[f\Big(, (\theta)\overline{x}_k + D\big(y_{k+1} - c_i(\theta)\big)\Big)h(\theta) \\ & \langle A(\theta)X_k + V_{k+1}, e_i\rangle\overline{\Lambda}_k \frac{\phi(y_{k+1} - c_i(\theta) - \langle d(\theta), \overline{x}_k\rangle)}{\phi(y_{k+1})} \mid \mathcal{Y}_{k+1}\Big] \\ = \quad & \phi(y_{k+1})^{-1} \int\int \sum_{j=1}^{N} \Big\{f\big(, (u)z + D(y_{k+1} - c_i(u))h(u) \\ & a_{ij}(u)\phi(y_{k+1} - c_i(u) - \langle d(u), z\rangle\big)q^i_k(z, u)\Big\}dzd\lambda(u) \end{aligned} \tag{17}$$

Write

$$\xi = , (u)z + D(y_{k+1} - c_i(u))$$

so

$$z = , (u)^{-1}\big\{\xi - D(y_{k+1} - c_i(u))\big\}$$

and

$$dzd\lambda(u) = , (u)^{-1}d\xi d\lambda(u).$$

The functions f and h are arbitrary so from the equality of (16) and (17) we have the following result:

Theorem 4.1 *Write*

$$\Phi(y_{k+1}, u, \xi) = \phi(y_{k+1})^{-1}\phi\big(y_{k+1} - c_i(u) - \langle d(u), , (u)^{-1}(\xi - D(y_{k+1} - c_i(u)))\rangle;$$

then for $1 \leq i \leq N$

$$\boxed{\begin{array}{l} q^i_{k+1}(\xi, u) = \Phi(y_{k+1}, u, \xi) \\ \sum_{j=1}^N a_{ij}(u) q^j_k\big(,\ (u)^{-1}(\xi - D(y_{k+1} - c_i(u)), u\big). \end{array}} \tag{18}$$

5 A Mixed Case

In this section we consider the situation where a Markov chain influences a linear system which, in turn, is observed linearly in noise. The parameters of the model are supposed unknown. Again a recursive expression is obtained for the unnormalized density of the state and parameters given the observations.

Again, without loss of generality the state space of the Markov chain X is taken to be the set of unit vectors $\{e_1, \ldots, e_N\}$. Then from equation (13)

$$X_{k+1} = A(\theta) X_k + V_{k+1}, \qquad k \in \mathbb{N}.$$

The state of the linear system is given by a process x_k, $k \in \mathbb{N}$, taking values in $\mathbb{R}^d$, and its dynamics are described by the equation

$$x_{k+1} = F(\theta) x_k + G(\theta) X_k + v_{k+1}.$$

Here v_k, $k \in \mathbb{N}$, is a sequence of independent random variables with densities ψ_k.

The observation process has the form

$$y_{k+1} = C(\theta) x_k + w_{k+1}.$$

The w_k are independent random variables having strictly positive densities ϕ_k.

In summary, we have the model

$$\boxed{\begin{array}{l} X_{k+1} = A(\theta) X_k + V_{k+1}, \\ x_{k+1} = F(\theta) x_k + G(\theta) X_k + v_{k+1}, \\ y_{k+1} = C(\theta) x_k + w_{k+1}, \qquad k \in \mathbb{N}. \end{array}} \tag{19}$$

The parameter θ takes values in some measure space (Θ, β, u). Again write $q^i_k(z, \theta)$ for the unnormalized joint conditional density of x_k and θ, given that $X_k = e_i$ such that

$$q^i_k(z, \theta) dz d\theta = \overline{E}[\langle X_k, e_i \rangle I(x_k \in dz) I(\theta \in d\theta) \overline{\Lambda}_k \mid \mathcal{Y}_k].$$

For suitable test functions f and h arguments as before lead us to consider

$$\overline{E}[\langle X_{k+1}, e_i \rangle f(x_{k+1}) h(\theta) \overline{\Lambda}_{k+1} \mid \mathcal{Y}_{k+1}]$$

$$= \int\int f(\xi)h(u)q^i_{k+1}(\xi,u)d\xi d\lambda(u)$$
$$= \overline{E}[\langle A(\theta)X_k + V_{k+1}, e_i\rangle f(F(\theta)x_k + G(\theta)X_k + w_{k+1})$$
$$h(\theta)\overline{\Lambda}_k\phi_{k+1}(y_{k+1} - C(\theta)x_k)/\Phi_{k+1}(y_{k+1}) \mid \mathcal{Y}_k]$$
$$= \phi_{k+1}(y_{k+1})^{-1}\sum_{j=1}^{N}\int\int a_{ij}(u)f(F(u)z + G(u)(e_j + w)$$
$$h(u)\phi_{k+1}(y_{k+1} - C(u)z)\psi_{k+1}(w)q^j_k(z,u)dzd\lambda(u)dw.$$

Substituting $\xi = F(u)z + G(u)e_j + w$, $z = z$, $u = u$, this is

$$= \phi_{k+1}(y_{k+1})^{-1}\sum_{j=1}^{N}\int\int\int a_{ij}(u)f(\xi)h(u)\phi_{k+1}(y_{k+1} - C(u)z)$$
$$\psi_{k+1}(\xi - F(u)z + G(u)e_j)q^i_k(z,u)dzd\lambda(u)d\xi.$$

This identity holds for all test functions f and h, so we have the following result:

Theorem 5.1 *Write* $\psi_{k+1}(\xi,u,z) = \psi_{k+1}(\xi - F(u)z + G(u)e_j)$, *then*

$$\boxed{\begin{array}{l} q^i_{k+1}(\xi,u) = \phi_{k+1}(y_{k+1})^{-1} \\ \int\sum_{j=1}^{N}\{a_{ij}(u)\phi_{k+1}(y_{k+1} - C(u)z)\psi_{k+1}(\xi,u,z)q^j_k(z,u)\}dz \end{array}} \tag{20}$$

References

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering.* Englewood Cliffs, N.J.: Prentice Hall, 1979.

[2] R.J. Elliott. A general recursive discrete time filter. Submitted.

[3] R.J. Elliott and J.B. Moore. Recursive parameter estimation for partially observed Markov chains. Submitted.

[4] A.N. Shiryaev. *Probability.* New York: Springer-Verlag, 1984.

Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, T6G 2G1

Department of Systems Engineering, Cooperative Research Center for Robust and Adaptive Systems, Australian National University, Canberra, A.C.T. 2601, Australia

Communicated by Clyde F. Martin