Journal of Mathematical Systems, Estimation, and Control
Vol. 5, No. 3, 1995, pp. 1–12

A Measure Change Derivation of Continuous State Baum–Welch Estimators*

Lakhdar Aggoun† Robert J. Elliott† John B. Moore†

Abstract

Vector valued signal and observation processes are considered with additive noise. Using changes of measure they are transformed into sequences of independent random variables. The situation where the noise in the signal is singular is discussed. The measure change enables easy recursions to be obtained for the forward and backward unnormalized conditional densities. These are analogs of the Baum–Welch algorithm.

1 Introduction

The Baum–Welch algorithm usually discusses a Markov chain observed in Gaussian noise; see [1]. The forward and backward Baum–Welch estimators are related to considering the observations under an equivalent probability measure; they provide unnormalized filtered and smoothed estimates of the state of the Markov chain, given the observations.

In this paper, like its predecessor [2], nonlinear, vector valued signal and observation dynamics are considered in discrete time, with additive (not necessarily Gaussian) noise. The original work of Baum–Welch considers a Markov chain signal. Novel features of this paper are that possibly singular measures describe the conditional distribution of the state, and that a double measure change is introduced under which both signal and observations become sequences of independent random variables. This facilitates easy derivations of the forward recursion for the 'alpha' unnormalized, conditional density, and the backward recursion for the 'beta' variable. The unnormalized smoothed density is, as in the Baum–Welch situation, the product of alpha and beta.

*Received October 1, 1992; received in final form April 5, 1993. Summary appeared in Volume 5, Number 3, 1995.
†Research partially supported by NSERC Grant A7964

The measure change we use is the discrete time analog of Girsanov's theorem and, in fact, it can be interpreted as a form of Bayes' theorem. We show how to transform the sequences of signal and observation variables into i.i.d. random variables. Calculations then take place in a mathematically nice setting where the variables are i.i.d. The results can then be interpreted back in the 'real world' by an inverse change of measure. We hope this might be considered clearer and more elegant than repeated use of Bayes' rule.

2 State and Observations

All processes are defined initially on a probability space $(\Omega, \mathcal{F}, P)$. Suppose $\{x_\ell\}$, $\ell \in Z^+$, is a discrete-time stochastic state process taking values in some Euclidean space $\mathbb{R}^m$. We suppose that x_0 has a known distribution $\pi_0(x)$. $\{w_\ell\}$, $\ell \in Z^+$, will be a sequence of independent, $\mathbb{R}^m$-valued, random variable with probability distributions $d\psi_\ell$.

For $n \in Z^+$, $F_n : \mathbb{R}^m \to \mathbb{R}^m$ are measurable functions, and we suppose for $n \geq 0$ that

$$x_{n+1} = F_{n+1}(x_n) + w_{n+1}. \tag{2.1}$$

The state process x is not observed directly; rather we suppose there is an observation process $\{y_\ell\}$, $\ell \in Z^+$, related to the state, and taking values in some Euclidean space $\mathbb{R}^d$. $\{b_n\}$ will be a sequence of independent $\mathbb{R}^d$-valued random variables with strictly positive density function ϕ_n.

For $n \in Z^+$, the $G_n : \mathbb{R}^m \to \mathbb{R}^d$ are measurable functions and we suppose for $n \geq 1$ that

$$y_n = G_n(x_n) + b_n. \tag{2.2}$$

Modelling systems by equations of the form (2.1) and (2.2), of course, raises the problem of identifying the parameters in such models. When the x process is a Markov chain a solution is given by the EM algorithm and the Baum–Welch method. (See [1].) A first step in this direction is given by the discussion in Theorem 6.6 of this paper of the conditional joint density function of x_m and x_{m+1}. This might lead to some estimate of the function F in (2.1).

3 Change of Measure for the y Process

Define $\lambda_\ell = \frac{\phi_\ell(y_\ell)}{\phi_\ell(b_\ell)}$, $\ell \in Z^+$. Write G_n, (resp. $\mathcal{Y}_n$), for the completions of the σ-fields

$$\begin{aligned} G_n^0 &= \sigma\{x_0, x_1, \ldots, x_n,\ y_1, \ldots, y_n\}, \\ \mathcal{Y}_n^0 &= \sigma\{y_1, \ldots, y_n\}, \end{aligned}$$

Then $\{G_n\}$, (resp. $\{\mathcal{Y}_n\}$), $n \in Z^+$, will denote the corresponding filtration (that is, increasing family of σ-fields). With $\Lambda_n = \prod_{\ell=1}^{n} \lambda_\ell$ a new probability measure $\overline{P}$ can be defined by setting the restriction of the Radon–Nikodym derivative $\frac{d\overline{P}}{dP}$ to G_n equal to Λ_n. The existence of $\overline{P}$ follows from Kolmogorov's theorem.

Lemma 3.1 *Under $\overline{P}$ the random variables y_ℓ, $\ell \in Z^+$, are independent and the density function of y_ℓ is ϕ_ℓ.*

Proof: Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is any integrable function. Then

$$\begin{aligned}\overline{E}\big[f(y_n) \mid G_{n-1}\big] &= \frac{E[\Lambda_n f(y_n) \mid G_{n-1}]}{E[\Lambda_n \mid G_{n-1}]} \\ &= \frac{\Lambda_{n-1} E[\lambda_n f(y_n) \mid G_{n-1}]}{\Lambda_{n-1} E[\lambda_n \mid G_{n-1}]} \\ &= \frac{E[\lambda_n f(y_n) \mid G_{n-1}]}{E[\lambda_n \mid G_{n-1}]}.\end{aligned}$$

Now

$$\begin{aligned}E[\lambda_n \mid G_{n-1}] &= \int_{\mathbb{R}^d} \frac{\phi_n(y_n)}{\phi_n(b_n)} \phi_n(b_n) db_n \\ &= \int_{\mathbb{R}^d} \phi_n\big(G_n(x_{n-1}) + b_n\big) db_n = 1.\end{aligned}$$

Hence

$$\begin{aligned}\overline{E}[f(y_n) \mid G_{n-1}] &= E[\lambda_n f(y_n) \mid G_{n-1}] \\ &= \int_{\mathbb{R}^d} \frac{\phi_n(y_n)}{\phi_n(b_n)} f(y_n) \phi_n(b_n) db_n \\ &= \int_{\mathbb{R}^d} \phi_n(y_n) f(y_n) dy_n = \overline{E}[f(y_n)],\end{aligned}$$

and the result follows. □

Remark 3.2 We now suppose we start with a probability measure $\overline{P}$ on $\left(\Omega, \bigcup_{n=1}^{\infty} G_n\right)$ such that under $\overline{P}$:

1. $\{x_\ell\}$, $\ell \in Z^+$, still satisfies the dynamics (2.1), that is $x_{n+1} = F_{n+1}(x_n) + w_{n+1}$.
2. $\{y_\ell\}$, $\ell \in Z^+$, is a sequence of independent random variables having density function $\phi_\ell > 0$.

Note that under $\overline{P}$ the y_ℓ are, in particular, independent of the x_ℓ. To represent the situation where the state influences the observations we wish to construct a probability measure P such that, under P, $b_n := y_n - G_n(x_n)$ is a sequence of independent random variables with positive density functions $\phi_n(b)$. To construct P starting from $\overline{P}$ we must proceed in an inverse manner. Write first $\overline{\lambda}_n = \frac{\phi_n(b_n)}{\phi_n(y_n)}$ and $\overline{\Lambda}_n = \prod_{\ell=1}^{n} \overline{\lambda}_\ell$. Then set $\frac{dP}{d\overline{P}}\Big|_{G_n} = \overline{\Lambda}_n$. The existence of P follows from Kolmogorov's extension theorem. It can be shown that under P the $\{b_\ell\}$ are independent random variables having densities ϕ_ℓ using the same argument as Lemma 3.1.

4 Recursive Estimates

We shall work under measure $\overline{P}$, so that the $\{y_\ell\}$, $\ell \in Z^+$, is a sequence of independent $\mathbb{R}^d$-valued random variables with densities ϕ_ℓ and the $\{x_\ell\}$, $\ell \in Z^+$, satisfy the dynamics described in Section 2, that is, $x_{n+1} = F_{n+1}(x_n) + w_{n+1}$. A version of Bayes' theorem states that for a G-adapted sequence $\{\alpha_\ell\}$

$$E[\alpha_\ell \mid \mathcal{Y}_\ell] = \frac{\overline{E}[\overline{\lambda}_\ell \alpha_\ell \mid \mathcal{Y}_\ell]}{\overline{E}[\overline{\Lambda}_\ell \mid \mathcal{Y}_\ell]}. \tag{4.1}$$

Identity (4.1) enables us to obtain the conditional expectation $E[\alpha_\ell \mid \mathcal{Y}_\ell]$ if we know the unnormalized conditional expectation $\overline{E}[\overline{\Lambda}_\ell \alpha_\ell \mid \mathcal{Y}_\ell]$.

Notation 4.1 *Write $dA_n(x)$, $n \in Z^+$, for the unnormalized conditional probability measure of x_n given $\mathcal{Y}_n$ such that $\overline{E}\big[\overline{\Lambda}_n I(x_n \in dx) \mid \mathcal{Y}_n\big] = dA_n(x)$.*

From (4.1), if $d\overline{A}_n(x)$ is the normalized conditional probability measure (distribution) under P of x_n given $\mathcal{Y}_n$, then

$$d\overline{A}_n(x) = dA_n(x)/\Big(\int_{R^m} dA_n(x)\Big) = E[I(x_n \in dx) \mid \mathcal{Y}_n].$$

Theorem 4.2 *For $n \in Z^+$, a recursion for $dA_n(\ )$ is given by $dA_n(z) = \frac{1}{\phi_n(y_n)}\phi_n\big(y_n - G_n(z)\big)\int_{R^m} d\psi_n\big(z - F_n(\xi)\big)dA_{n-1}(\xi)$.*

Proof: Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is any integrable Borel test function. Then $\overline{E}\big[\overline{\Lambda}_n f(x_n) \mid \mathcal{Y}_n\big] = \int_{\mathbb{R}^m} f(z) dA_n(z)$. However,

$$\begin{aligned}
&\overline{E}[\overline{\Lambda}_n f(x_n) \mid \mathcal{Y}_n] \\
&= \overline{E}\Big[\overline{\Lambda}_{n-1}\frac{\phi_n\big(y_n - G_n(F_n(x_{n-1}) + w_n)\big)}{\phi_n(y_n)} f\big(F_n(x_{n-1}) + w_n\big) \mid \mathcal{Y}_n\Big] \\
&= \overline{E}\Big[\overline{\Lambda}_{n-1}\int_{R^m} \frac{\phi_n\big(y_n - G_n(F_n(x_{n-1}) + w)\big)}{\phi_n(y_n)}
\end{aligned}$$

$$f\big(F_n(x_{n-1})+w\big)d\psi_n(w)\mid\mathcal{Y}_n\Big]$$
$$= \frac{1}{\phi_n(y_n)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\phi_n\big(y_n-G_n(F_n(\xi)+w)\big) f\big(F_n(\xi)+w\big)d\psi_n(w)dA_{n-1}(\xi).$$

Write $z=F_n(\xi)+w$. Consequently

$$\int_{\mathbb{R}^m} f(z)dA_n(z) = \frac{1}{\phi_n(y_n)}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m}\phi_n(y_n-G_n(z))f(z)d\psi_n(z-F_n(\xi))dA_{n-1}(\xi).$$

This identity holds for all Borel test functions f, so

$$dA_n(z)=\frac{1}{\phi_n(y_n)}\phi_n(y_n-G_n(z))\int_{R^m}d\psi_n(z-F_n(\xi))dA_{n-1}(\xi).$$

Notation 4.3 *In this section $m,n\in Z^+$ and $m<n$. Write $\overline{\Lambda}_{m,n}=\prod_{\ell=m}^{n}\overline{\lambda}_\ell$ and $dH_{m,n}(x)$ for the unnormalized conditional probability measure of x_m given $\mathcal{Y}_n$ such that $\overline{E}\big[\overline{\Lambda}_n I(x_m\in dx)\mid\mathcal{Y}_n\big]=dH_{m,n}(x)$. Then from (4.1) the normalized conditional probability measure (distribution) under P of x_m given $\mathcal{Y}_n$ is $dH_{m,n}(x)/\int_{R^m}dH_{m,n}(x)$.*

Theorem 4.4 *For $m,n\in Z^+$, $m<n$, $dH_{m,n}(x)=\beta_{m,n}(x)dA_m(x)$ where $dA_m(x)$ is given recursively by Theorem 4.2 and $\beta_{m,n}(x)=\overline{E}\big[\overline{\Lambda}_{m+1,n}\mid x_m=x,\ \mathcal{Y}_n\big]$.*

Proof: For an arbitrary integrable function $f:\mathbb{R}^m\to\mathbb{R}$,

$$\overline{E}[\overline{\Lambda}_n f(x_m)\mid\mathcal{Y}_n]=\int_{\mathbb{R}^m}f(x)dH_{m,n}(x).$$

However, $\overline{E}[\overline{\Lambda}_n f(x_m)\mid\mathcal{Y}_n]=\overline{E}[\overline{\Lambda}_{1,m}f(x_m)\overline{E}[\overline{\Lambda}_{m+1,n}\mid x_0,\dots,x_m,\mathcal{Y}_n]\mid\mathcal{Y}_n]$. Now $\overline{E}[\overline{\Lambda}_{m+1,n}\mid x_m=x,\ \mathcal{Y}_n]:=\beta_{m,n}(x)$. Consequently, $\overline{E}[\overline{\Lambda}_n f(x_m)\mid\mathcal{Y}_n]=\overline{E}[\overline{\Lambda}_{1,m}f(x_m)\beta_{m,n}(x_m)\mid\mathcal{Y}_n]$, and so, from Notation 4.1,

$$\int_{\mathbb{R}^m}f(x)dH_{m,n}(x)=\int_{\mathbb{R}^m}f(x)\beta_{m,n}(x)dA_m(x),$$

$f(x)$ is an arbitrary Borel test function; therefore, we see $dH_{m,n}(x)=\beta_{m,n}(x)dA_m(x)$.

Theorem 4.5 *$\beta_{m,n}(x)$ satisfies the backward recursive equation*

$$\beta_{m,n}(x) = \frac{1}{\phi_{m+1}(y_{m+1})}\int_{R^m}\phi_{m+1}\big(y_{m+1}-G_{m+1}(F_{m+1}(x)+w)\big) \times\beta_{m+1,n}(F_{m+1}(x)+w)d\psi_{m+1}(w)$$

with $\beta_{n,n}=1$.

Proof:

$$
\begin{aligned}
\beta_{m,n}(x) &= \overline{E}[\overline{\Lambda}_{m+1,n} \mid x_m = x, \mathcal{Y}_n] \\
&= \overline{E}\Big[\frac{\phi_{m+1}(y_{m+1} - G_{m+1}(x_{m+1}))}{\phi_{m+1}(y_{m+1})}\overline{\Lambda}_{m+2,n} \mid x_m = x, \mathcal{Y}_n\Big] \\
&= \overline{E}\Big[\frac{\phi_{m+1}(y_{m+1} - G_{m+1}(x_{m+1}))}{\phi_{m+1}(y_{m+1})}\overline{E}[\overline{\Lambda}_{m+2,n} \mid x_m = x, \\
&\qquad x_{m+1},\ \mathcal{Y}_n] \mid x_m = x,\ \mathcal{Y}_n\Big] \\
&= \overline{E}\Big[\frac{\phi_{m+1}\big(y_{m+1} - G_{m+1}(F_{m+1}(x_m) + w_{m+1})\big)}{\phi_{m+1}(y_{m+1})} \\
&\qquad \times\beta_{m+1,n}(F_{m+1}(x_m) + w_{m+1}) \mid x_m = x,\ \mathcal{Y}_n\Big] \\
&= \frac{1}{\phi_{m+1}(y_{m+1})}\int_{R^m} \phi_{m+1}\big(y_{m+1} - G_{m+1}(F_{m+1}(x) + w)\big) \\
&\qquad \times\beta_{m+1,n}(F_{m+1}(x) + w)d\psi_{m+1}(w).
\end{aligned}
$$

Remark 4.6 Of interest in applications is the linear model with a singular matrix coefficient in the noise term of the state dynamics:

$$
\begin{aligned}
x_n &= F(x_{n-1}) + Bw'_n \in \mathrm{IR}^m \\
y_n &= G(x_{n-1}) + b_n \in \mathrm{IR}^d.
\end{aligned}
$$

Recursive estimators are obtained by the methods of this section if we set $w_n = Bw'_n$. Even if B is singular w_n always has probability distribution $d\psi_n(\cdot)$. The support of this distribution is the set of values of Bw'_n.

5 A Change of Measure for the State Process x

In this section we shall suppose that the noise in the state equation (2.1) is not singular, that is, each w_n has a positive density function ψ_n. The observation process y is as given by equation (2.2). Suppose $\overline{P}$ has been constructed as in Section 3. Define $\gamma_\ell = \frac{\psi_\ell(x_\ell)}{\psi_\ell(w_\ell)}$ and $,_n = \prod_{\ell=1}^{n} \gamma_\ell$; set $\frac{d\widehat{P}}{d\overline{P}}\Big|_{G_n} = ,_n$.

Lemma 5.1 *Under $\widehat{P}$ the random variables $\{x_\ell\}$, $\ell \in Z^+$, are independent with density function ψ_ℓ.*

Proof: Suppose $g : \mathrm{IR}^m \to \mathrm{IR}$ is any integrable function. Then

$$
\widehat{E}[g(x_n) \mid G_{n-1}] = \frac{\overline{E}[,_n g(x_n) \mid G_{n-1}]}{\overline{E}[,_n \mid G_{n-1}]}
$$

$$= \frac{\Gamma_{n-1}\overline{E}[\gamma_n g(x_n) \mid G_{n-1}]}{\Gamma_{n-1}\overline{E}[\gamma_n \mid G_{n-1}]}.$$

As in Lemma 3.1, $\overline{E}[\gamma_n \mid G_{n-1}] = 1$, so

$$\begin{aligned}
\widehat{E}[g(x_n) \mid G_{n-1}] &= \overline{E}[\gamma_n g(x_n) \mid G_{n-1}] \\
&= \int_{\mathbb{R}^m} \frac{\psi_n(x_n)}{\psi_n(w_n)} g(x_n)\psi_n(w_n) dw_n \\
&= \int_{R^m} \psi_n(x_n) g(x_n) dx_n = \widehat{E}[g(x_n)].
\end{aligned}$$

The result, therefore, follows.

Remark 5.2 What we now can do is to start with a probability measure $\widehat{P}$ on $\left(\Omega, \bigcup_{n=1}^{\infty} G_n\right)$ under which the process $\{x_\ell\}$ and $\{y_\ell\}$ are two sequences of independent random variables with respective densities ψ_ℓ and ϕ_ℓ. Note the x and y are independent of each other as well. To return to the "real world" model described in Section 2 we must define a probability measure P by setting $\left.\frac{dP}{d\widehat{P}}\right|_{G_n} = \left.\frac{dP}{d\overline{P}}\right|_{G_n} \left.\frac{d\overline{P}}{d\widehat{P}}\right|_{G_n} = \overline{\Lambda}_n \overline{\Gamma}_n$. Here $\overline{\Gamma}_n$ is the inverse of Γ_n, so that $\overline{\Gamma}_n = \prod_{\ell=1}^{n} \overline{\gamma}_\ell$, where $\overline{\gamma}_\ell = \frac{\psi_\ell(w_\ell)}{\psi_\ell(x_\ell)}$. Again the existence of P is guaranteed by Kolmogorov's extension theorem.

6 Recursive Estimates

We shall work under $\widehat{P}$, so that $\{y_\ell\}$, $\ell \in Z^+$, and $\{x_\ell\}$, $\ell \in Z^+$, are two sequences of independent random variables with respective densities ϕ_ℓ and ψ_ℓ. Recall that a version of Bayes' theorem states that for a G-adapted sequence $\{g_\ell\}$, $E[g_\ell \mid \mathcal{Y}_\ell] = \frac{\overline{E}[\overline{\Lambda}_\ell g_\ell \mid \mathcal{Y}_\ell]}{\overline{E}[\overline{\Lambda}_\ell \mid \mathcal{Y}_\ell]}$. Similarly, $\overline{E}[\overline{\Lambda} g_\ell \mid \mathcal{Y}_\ell] = \frac{\widehat{E}[\overline{\Gamma}_n \overline{\Lambda}_n g_\ell \mid \mathcal{Y}_\ell]}{\widehat{E}[\overline{\Gamma}_n \mid \mathcal{Y}_\ell]}$.

Remark 6.1 The x_ℓ sequence is independent of the y_ℓ sequence under $\widehat{P}$. Therefore conditioning on the x's it is easily seen that $\widehat{E}[\overline{\Gamma}_\ell \mid \mathcal{Y}_\ell] = \widehat{E}[\overline{\Gamma}_\ell] = \widehat{E}[\overline{\Gamma}_{\ell-1}] = 1$.

Notation 6.2 *Suppose $\alpha_n(x)$, $n \in Z^+$, is the unnormalized conditional density of x_n given $\mathcal{Y}_n$ such that $\overline{E}[\overline{\Lambda}_n I(x_n \in dx) \mid \mathcal{Y}_n] = \alpha_n(x)dx$. From (4.1) the normalized conditional density, under P, of x_n given $\mathcal{Y}_n$ is $\alpha_n(x)/\int_{R^m} \alpha_n(z)dz$.*

We now re-derive the recursive expression for α_n.

Theorem 6.3 *For $n \in Z^+$, a recursion for $\alpha_n(x)$ is given by*

$$\alpha_n(x) = \frac{\phi_n(y_n - G_n(x))}{\phi_n(y_n)} \int_{R^m} \psi_n(x - F_n(\xi))\alpha_{n-1}(\xi)d\xi. \tag{6.1}$$

Proof: Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is any integrable Borel test function. Then

$$\overline{E}[f(x_n)\overline{\Lambda}_n \mid \mathcal{Y}_n] = \int_{R^m} f(x)\alpha_n(x)dx. \tag{6.2}$$

However, $\overline{E}[f(x_n)\overline{\Lambda}_n \mid \mathcal{Y}_n] = \frac{\widehat{E}[f(x_n)\overline{\Lambda}_n \overline{\Gamma}_n \mid \mathcal{Y}_n]}{\widehat{E}[\overline{\Gamma}_n \mid \mathcal{Y}_n]}$. From Remark 5.2 the denominator equals 1. Using the independence of the x_n's and the y_n's under $\widehat{P}$ we see

$$\begin{aligned} &\overline{E}[f(x_n)\overline{\Lambda}_n \mid \mathcal{Y}_n] \\ &\quad = \widehat{E}[f(x_n)\overline{\Lambda}_n \overline{\Gamma}_n \mid \mathcal{Y}_n] \\ &\quad = \int_{R^m}\int_{R^m} \psi_n(x - F_n(\xi))\frac{\phi_n(y_n - G_n(x))}{\phi_n(y_n)}\alpha_{n-1}(\xi)d\xi f(x)dx. \end{aligned} \tag{6.3}$$

Since f is an arbitrary Borel test function, equations (6.2) and (6.3) yield at once $\alpha_n(x) = \frac{\phi_n(y_n - G_n(x))}{\phi_n(y_n)} \int_{R^m} \psi_n(x - F_n(\xi))\alpha_{n-1}(\xi)d\xi$ which is equation (6.1).

Notation 6.4 *For $m, n \in Z^+$, $m < n$, write $\overline{\Lambda}_{m,n} = \prod_{\ell=m}^{n} \overline{\lambda}_\ell$ and $\overline{\Gamma}_{m,n} = \prod_{\ell=m}^{n} \overline{\gamma}_\ell$. Write $\gamma_{m,n}(x)$ for the unnormalized conditional density of x_m given $\mathcal{Y}_n$, under $\overline{P}$, such that $\overline{E}[\overline{\Lambda}_n I(x_m \in dx) \mid \mathcal{Y}_n] = \gamma_{m,n}(x)dx$.*

From (4.1) the normalized conditional density of x_m given $\mathcal{Y}_n$, $m < n$, is then $\gamma_{m,n}(x)/\int_{R^m} \gamma_{m,n}(z)dz$. As in Theorem 4.4, $\beta_{m,n}(x) = \overline{E}[\overline{\Lambda}_{m+1,n} \mid x_m = x, \mathcal{Y}_n]$. It can be shown as in theorems 4.2, 4.4 and 4.5 that $\gamma_{m,n}(x) = \alpha_m(x)\beta_{m,n}(x)$ where $\alpha_m(x)$ is given recursively by Theorem 6.3 and $\beta_{m,n}(x)$ satisfies the backward recursive equation

$$\begin{aligned} \beta_{m,n}(x) &= \frac{1}{\phi_{m+1}(y_{m+1})} \int_{R^m} \psi_{m+1}(z - F_{m+1}(x)) \\ &\qquad \phi_{m+1}(y_{m+1} - G_{m+1}(z))\beta_{m+1,n}(z)dz. \end{aligned}$$

Notation 6.5 *For $m \in Z^+$, $m < n$, write $\xi_{m,m+1,n}(x^1, x^2)$ for the joint unnormalized conditional density of x_m and x_{m+1} given $\mathcal{Y}_n$, under $\overline{P}$, such that $\overline{E}[\overline{\Lambda}_n I(x_m \in dx^1) I(x_{m+1} \in dx^2) \mid \mathcal{Y}_n] = \xi_{m,m+1,n}(x^1, x^2)dx^1 dx^2$.*

The normalized conditional density $E[I(x_m \in dx^1)I(x_{m+1} \in dx^2) \mid \mathcal{Y}_n]$ is then

$$\frac{\xi_{m,m+1,n}(x^1,x^2)dx^1dx^2}{\int_{R^m}\int_{R^m}\xi_{m,m+1,n}(z^1,z^2)dz^1dz^2}.$$

Also, write $\rho_{n+1,n}(x)$ for the unnormalized conditional density, under $\overline{P}$, of x_{n+1} given $\mathcal{Y}_n$, such that $\overline{E}[\overline{\Lambda}_{n+1}I(x_{n+1} \in dx) \mid \mathcal{Y}_n] = \rho_{n+1,n}(x)dx$. Then $E[I(x_{n+1} \in dx) \mid \mathcal{Y}_n] = \rho_{n+1,n}(x)dx / \int_{R^m}\rho_{n+1,n}(z)dz$.

Theorem 6.6 *For $m, n \in Z^+$, $m < n$,*

$$\begin{aligned}\xi_{m,m+1,n}(x^1,x^2) &= \alpha_m(x^1)\beta_{m+1,n}(x^2)\\ &\quad \psi_{m+1}(x^2 - F_{m+1}(x^1))\frac{\phi_{m+1}(y_{m+1} - G_{m+1}(x^2))}{\phi_{m+1}(y_{m+1})}.\end{aligned} \tag{6.4}$$

Proof: Suppose $f, g : \mathbb{R}^m \to \mathbb{R}$ are arbitrary integrable Borel functions. Then

$$\begin{aligned}&\widehat{E}[f(x_m)g(x_{m+1})\overline{\Lambda}_n\overline{,}_n \mid \mathcal{Y}_n]\\
&= \int_{R^m}\int_{R^m} f(x^1)g(x^2)\xi_{m,m+1,n}(x^1,x^2)dx^1dx^2 \qquad (6.5)\\
&= \widehat{E}\big[\widehat{E}[f(x_m)g(x_{m+1})\overline{\Lambda}_{0,m+1}\overline{\Lambda}_{m+2,n}\\
&\qquad \overline{,}_{0,m+1}\overline{,}_{m+2,n} \mid x_0,\dots,x_{m+1},\mathcal{Y}_n] \mid \mathcal{Y}_n\big]\\
&= \widehat{E}\big[f(x_m)g(x_{m+1})\overline{\Lambda}_{0,m+1}\overline{,}_{0,m+1}\widehat{E}[\overline{\Lambda}_{m+2,n}\overline{,}_{m+2,n} \mid x_{m+1},\mathcal{Y}_n] \mid \mathcal{Y}_n\big]\\
&= \widehat{E}[f(x_m)g(x_{m+1})\overline{\Lambda}_{0,m+1}\overline{,}_{0,m+1}\beta_{m+1,n}(x_{m+1}) \mid \mathcal{Y}_n]\\
&= \widehat{E}\Big[f(x_m)\overline{\Lambda}_{0,m}\overline{,}_{0,m}\int_{\mathbb{R}^m} g(x^2)\psi_{m+1}\big(x^2 - F_{m+1}(x_m)\big)\\
&\qquad \times\frac{\phi_{m+1}(y_{m+1} - G_{m+1}(x^2))}{\phi_{m+1}(y_{m+1})}\beta_{m+1,n}(x^2)dx^2 \mid \mathcal{Y}_n\Big]\\
&= \int_{R^m}\int_{R^m} f(x^1)g(x^2)\psi_{m+1}(x^2 - F_{m+1}(x^1))\\
&\qquad \times\frac{\phi_{m+1}(y_{m+1} - G_{m+1}(x^2))}{\phi_{m+1}(y_{m+1})}\beta_{m+1,n}(x^2)\alpha_m(x^1)dx^1dx^2. \qquad (6.6)\end{aligned}$$

Since $f(x)$ and $g(x)$ are two arbitrary Borel test functions, the comparison of equation (6.5) with the quantity (6.6) gives equation (6.4). □

Theorem 6.7

$$\rho_{n+1,n}(x) = \int_{R^m}\psi_{n+1}(x - F_{n+1}(z))\alpha_n(z)dz. \tag{6.7}$$

Proof: Suppose f is an arbitrary integrable Borel function. Then:

$$\widehat{E}[f(x_{n+1})\overline{\Lambda}_{n+1}\overline{,}_{\,n+1} \mid \mathcal{Y}_n] = \int_{\mathbb{R}^m} f(x)\rho_{n+1,n}(x)dx.$$

However,

$$\begin{aligned}
&\widehat{E}[f(x_{n+1})\overline{\Lambda}_{n+1}\overline{,}_{\,n+1} \mid \mathcal{Y}_n] \\
&\quad = \widehat{E}\big[\widehat{E}[f(x_{n+1})\overline{\Lambda}_{n+1}\overline{,}_{\,n+1} \mid x_0,\dots,x_n,\mathcal{Y}_n] \mid \mathcal{Y}_n\big] \\
&\quad = \widehat{E}\Big[\overline{\Lambda}_n\overline{,}_{\,n}\widehat{E}\Big[\frac{\phi_{n+1}(y_{n+1}-G_{n+1}(x_{n+1}))}{\phi_{n+1}(y_{n+1})}\frac{\psi_{n+1}(x_{n+1}-F_{n+1}(x_n))}{\psi_{n+1}(x_{n+1})} \\
&\qquad\quad \times f(x_{n+1}) \mid x_0,\dots,x_n,\mathcal{Y}_n\Big] \mid \mathcal{Y}_n\Big] \\
&\quad = \widehat{E}\Big[\overline{\Lambda}_n\overline{,}_{\,n}\int_{R^m}\int_{R^d}\phi_{n+1}(y-G_{n+1}(x)) \\
&\qquad\quad \psi_{n+1}(x-F_{n+1}(x_n))f(x)dydx \mid \mathcal{Y}_n\Big] \\
&\quad = \int_{R^m}\int_{R^m}\int_{R^d}\phi_{n+1}(y-G_{n+1}(x)) \\
&\qquad\quad \psi_{n+1}(x-F_{n+1}(z))f(x)\alpha_n(z)dzdydx.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\int_{R^m} f(x)\rho_{n+1,n}(x)dx = &\int_{R^m}\int_{R^d}\int_{R^m}\phi_{n+1}(y-G_{n+1}(x)) \\
&\psi_{n+1}(x-F_{n+1}(z))f(x)\alpha_n(z)dzdydx
\end{aligned}$$

and this identity holds for all Borel test functions. Hence

$$\rho_{n+1,n}(x) = \int_{R^d}\int_{R^m}\phi_{n+1}(y-G_{n+1}(x))\psi_{n+1}(x-F_{n+1}(z))\alpha_n(z)dzdy.$$

However, $\int_{R^d}\phi_{n+1}(y-G_{n+1}(x))dy = 1$, so the result follows.

Remark 6.8 The unnormalized conditional density given by (6.7) can be easily generalized to the k-th step prediction and is given by $\rho_{n+k,n}(x) = \int_{R^{mk}} H(x^1,\dots,x^{k-1},x)\alpha_n(z)dx^1\dots dx^{k-1}dz$ where $H(x^1,\dots,x^{k-1},x) = \psi_{n+1}(x^1-F_{n+1}(z))\ \psi_{n+2}(x^2-F_{n+2}(x^1))\ \dots\psi_{n+k}(x-F_{n+k}(x^{k-1}))$.

7 A Linear Case with Gaussian Noise

Assume here that state and observation processes are given by the dynamics

$$\begin{aligned}
x_{n+1} &= A_{n+1}x_n + w_{n+1} \in \mathbb{R}^m &(7.1)\\
y_{n+1} &= C_{n+1}x_{n+1} + b_{n+1} \in \mathbb{R}^d. &(7.2)
\end{aligned}$$

A_n, C_n are matrices of appropriate dimensions, w_ℓ and b_ℓ are normally distributed with means 0 and respective covariance matrices Σ_ℓ and $,_\ell$. The conditional density of x_n given the observations up to time n is given by $P_n(x) = \alpha_n(x)/\int_{R^m} \alpha_n(x)dx$, where $\alpha_n(x)$ is the unnormalized density given by equation (6.1). The linearity of (7.1) and (7.2) implies that $P_n(x)$ is also normally distributed with mean $\mu_n = E[x_n \mid \mathcal{Y}_n]$ and covariance matrix R_n. Our purpose in this section is to give recursive estimates of μ_n and R_n using the recursion for $\alpha_n(x)$:

$$\begin{aligned}\alpha_n(x) &= K(x)\int_{R^m} \exp\Big(-\frac{1}{2}\Big)\Big\{(x-A_n\xi)'\Sigma_n^{-1}(x-A_n\xi)\\ &\quad +(\xi-\mu_{n-1})'R_{n-1}^{-1}(\xi-\mu_{n-1})\Big\}d\xi\end{aligned}$$

in view of (6.1) and the densities ϕ_n and ψ_n;

$$K(x) = \frac{\phi_n(y_n - C_n x)}{\phi_n(y_n)}(2\pi)^{-m}|\Sigma_n|^{-\frac{1}{2}}|R_{n-1}|^{-\frac{1}{2}}. \tag{7.3}$$

Leaving only terms containing the variable ξ under the integration symbol in (7.3), this is

$$= K_1(x)\int_{R^m} \exp\Big(-\frac{1}{2}\Big)\{\xi' a_n \xi - \beta_n'\xi\}d\xi \tag{7.4}$$

where

$$\begin{aligned}K_1(x) &= K(x)\exp\Big(-\frac{1}{2}\Big)\{x'\Sigma_n^{-1}x + \mu_{n-1}'R_{n-1}^{-1}\mu_{n-1}\} && (7.5)\\ a_n &= A_n'\Sigma_n^{-1}A_n + R_{n-1}^{-1},\\ \beta_n' &= 2(x'\Sigma_n^{-1}A_n + \mu_{n-1}'R_{n-1}^{-1}). && (7.6)\end{aligned}$$

Completing the 'square' in (7.4) this is equal to

$$\begin{aligned}&K_1(x)\exp\Big(-\frac{1}{2}\Big)\Big\{-\frac{\beta'(a_n^{-1})\beta}{4}\Big\}\\ &\qquad\int_{R^m}\exp\Big\{-\frac{1}{2}\Big(\frac{\xi - a_n^{-1}\beta_n}{2}\Big)a_n\Big(\xi - \frac{a_n^{-1}\beta_n}{2}\Big)\Big\}d\xi\\ &= K_1(x)\exp\Big(-\frac{1}{2}\Big)\Big\{-\frac{\beta'(\alpha_n^{-1})\beta}{4}\Big\}|a_n|^{\frac{1}{2}}(2\pi)^{-\frac{m}{2}}. && (7.7)\end{aligned}$$

In view of (7.3), (7.5), (7.6) and (7.7) we have

$$\alpha_n(x) = K_2\exp\Big\{-\frac{1}{2}\Big(x - \frac{\gamma_n^{-1}\delta_n}{2}\Big)'\gamma_n\Big(x - \frac{\gamma_n^{-1}\delta_n}{2}\Big)\Big\} \tag{7.8}$$

where K_2 is a constant independent of x and

$$\begin{aligned} \gamma_n &= \Sigma_n^{-1} + C'_n, {}_n^{-1} C_n - \Sigma_n^{-1} A_n a_n^{-1} A'_n \Sigma_n^{-1} \\ \delta_n &= 2(y'_n, {}_n^{-1} C_n + R_{n-1}^{-1} \mu_{n-1} a_n^{-1} A'_n \Sigma_n^{-1}). \end{aligned}$$

From (7.8) and (7.6) we see

$$\begin{aligned} R_n = \gamma_n^{-1} &= \left[\Sigma_n^{-1} - \Sigma_n^{-1} A_n a_n^{-1} A'_n \Sigma_n^{-1} + C'_n, {}_n^{-1} C_n\right]^{-1} \\ &= \left[(A_n R_{n-1} A'_n + \Sigma_n)^{-1} + C'_n, {}_n^{-1} C_n\right]^{-1} \\ \mu_n = \frac{\gamma_n^{-1}\delta_n}{2} &= R_n\big(y'_n, {}_n^{-1} C_n + R_{n-1}^{-1} \mu_{n-1} a_n^{-1} A'_n \Sigma_n^{-1}\big). \end{aligned}$$

In summary we have

Theorem 7.1 *For the linear model described by equations (7.1) and (7.2), the conditional mean and covariance matrix of the state process x_n are given by*

$$\begin{aligned} \mu_n &= R_n R_{n-1}^{-1} \mu_{n-1} a_n^{-1} A'_n \Sigma_n^{-1} + R_n y'_n, {}_n^{-1} C_n \\ R_n &= \left[(A_n R_{n-1} A'_n + \Sigma_n)^{-1} + C'_n, {}_n^{-1} C_n\right]^{-1} \end{aligned}$$

where $a_n = A'_n \Sigma_n^{-1} A_n + R_{n-1}^{-1}$.

References

[1] L.E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains, *Ann. Math. Stat.* **37** (1966), 1554–1563.

[2] R.J. Elliott. A general recursive, discrete time filter, *Jour. Applied Prob.* Sept. 1993. To appear.

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

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

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

Communicated by Clyde F. Martin