Proceedings of the 36th
Conference on Decision & Control
San Diego, California USA • December 1997

Efficient Estimation for Non-linear and Non-Gaussian State Space Models

Dawei Huang

Centre for Statistical Science and Industrial Mathematics, Queensland University of Technology,

GPO Box 2434, Q4001,Australia,

e-mail: huang@fsc.qut.edu.au

Abstract

A probabilistic approach for state space models with underlying Markov chains x_t and observation sequences y_t is explored. Firstly, we establish a recursive formula for calculating the Cramer-Rao (C-R) lower bound for a general state space model with a transit pdf $p(x_t|x_{t-1})$ and conditional $p(y_t|x_t)$. Secondly, we apply the C-R bound to several models, including FM demodulation models and outlier noise models. Sometimes simple Kalman Filter (KF) can attend the efficiency even for non-Gaussian cases. Sometimes Extended Kalman Filter (EKF) based methods, like Phase Locked Loop (PLL), can achieve the efficiency for non-linear models. To study the performance of non-linear filters like PLL, an algebraic approach is given for calculating the stationary distribution of these filters when it exists. However, there are also cases that both KF and EKF are far away from efficiency. Thirdly, two novel techniques are suggested when conventional filtering methods are inefficient. One is based on Gaussian approximations for $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ by Taylor expansion or maximum entropy method. Then an identity for Gaussian density products can be used to derive non-linear filters. The other is based on so-called Partial Conditional Expectations (PCE) $\hat{y}_{t,n} = E(x_t|y_t, y_{t-1}, ..., y_{t-n+1})$ which can be viewed as a non-linear transform of observations. Then optimal linear filters can be derived for tracking x_t based on $\hat{y}_{t,n}$. Simulation results show that under some circumstances these two approaches really can achieve the efficiency.

Key Words: Cramer-Rao bound, EKF, PLL, Frequency demodulation, Outlier models, Gaussian approximation, Limiting distribution of Markov chains, Gaussian approximation by Taylor expansion or maximum entropy criterion, Partial conditional expectation, Hybrid filter.

1 Introduction.

State space models have been of interest in both theoretical and practical aspects of a variety of areas [1-3]. Usually these models are expressed in their time domain form. In this paper, we consider a probabilistic formulation as well. Following the Hidden Markov Model (HMM) approach [3], a state space model can be stated as: An n−dimensional Markov chain $\{x_t, t = 0, 1, 2, ...\}$ with a transit pdf $p(x_t|x_{t-1})$ is related to an m−dimensional observation sequence $\{y_t, t = 1, 2, ...\}$ by a conditional pdf $p(y_t|x_t)$. In this paper, we assume that x_t and y_t are continuous random variables, although some results, with suitable modifications, can be used for discrete HMM as well. Two examples often discussed in the literature are the outlier model and the Frequency demodulation model: Let T, H, σ_u^2 and σ_k^2 be appropriate matrices, m_k be appropriate vectors, w_k be non-negative weights with sum one, $G(x; \mu, \Sigma)$ be the Gaussian pdf with mean μ and variance Σ, i.e.,

$$G(x; \mu, \Sigma) = \frac{\exp\left[-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right]}{\sqrt{(2\pi)^n} \det(\Sigma)}. \tag{1}$$

The Outlier Model [4] (Gaussian sum [5])

$$x_t = Tx_{t-1} + u_t,\ u_t \sim N\left(0, \sigma_u^2\right); \qquad y_t = Hx_t + v_t, \tag{2}$$

is equivalent to

$$\begin{cases} p(x_t|x_{t-1}) = G\left(x_t; Tx_{t-1}, \sigma_u^2\right), \\ p(y_t|x_t) = \sum_{k=1}^{p} w_k G\left(y_t; Hx_t + m_k, \sigma_k^2\right). \end{cases} \tag{3}$$

The Frequency Demodulation [6, 7] model

$$x_t = x_{t-1} + u_t, \quad y_t = A\begin{bmatrix} \sin x_t \\ \cos x_t \end{bmatrix} + v_t, \tag{4}$$

is equivalent to

$$\begin{cases} p(x_t|x_{t-1}) = G\left(x_t; x_{t-1}, \sigma_u^2\right), \\ p(y_t|x_t) = G\left(y_t; A\begin{bmatrix} \sin x_t \\ \cos x_t \end{bmatrix}, \sigma_v^2 I\right). \end{cases} \tag{5}$$

Based on the observation $Y_t = [y_1', ..., y_t']'$, we want to obtain a fast and efficient algorithm to estimate x_t. The algorithm developed, using conventional computer techniques, should be performed in real time. Also, the estimator $\hat{x}_t$ should be efficient in the sense of minimum Mean Square Error (MSE), i.e., $E\left[(x_t - \hat{x}_t)(x_t - \hat{x}_t)'\right]$ should close to its lower bound.

Theoretically, the optimal estimator is $E(x_t|Y_t)$ which can be calculated by Bayesian formulae (p174 in [1]). However, difficulties arise in calculation. Although computer techniques have developed so rapidly and many numerical methods have been investigated [8], calculating $E(x_t|Y_t)$ is extremely difficult. A recent development in this direction is the Markov Chain Monte Carlo (MCMC) method. Even though MCMC is theoretically sound, the question of how many samples we need to obtain a reliable result is still open [9]. Furthermore, MCMC cannot provide on-line and in real time solutions in general cases which are necessary for applications in communications, automatic control and signal processing.

In this article, we establish a recursive formula for calculating the Cramer-Rao (C-R) bound of general non-linear and non-Gaussian state space models in Section 2. Although we are estimating a stochastic process $\{x_t\}$ instead of a fixed parameter, the C-R bound is still valid [10]. For non-linear but somehow Gaussian noise models, C-R bounds were calculated by Riccati equation [11-13]. Based on the structure of general Fisher information matrices, we derive a more general formula. This formula allows us to calculate the C-R bound of very general state space models recursively. Then, in Section 3, we find that sometimes even a simple Kalman filter is efficient for non-Gaussian noise. Based on a an algebraic approach for calculating limiting distribution of a ergodic Markov chain, we can show that sometimes Extended Kalman Filter (EKF) is also efficient. For example, EKF-based Phase Locked Loop (PLL) is efficient for FM demodulation when the variance of the model noise is small and the Signal to Noise Ratio (SNR) is high. However, sometimes these methods are far away from efficiency. For the later cases, we develop several novel non-linear filtering methods based on probabilistic approaches in Section 4. One is based on Gaussian approximations for $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ by Taylor expansion or maximum entropy criterion. Then an identity for Gaussian density products [17] can be used to derive non-linear filters. The other is based on so-called Partial Conditional Expectations (PCE) $\hat{y}_t = E(x_t|y_t, y_{t-1}, ..., y_{t-k})$ which can be viewed as a non-linear transform of observations $\{y_t, y_{t-1}, ..., y_{t-k}\}$ to obtain the optimal correlation to x_t. Then optimal linear filters can be derived for tracking x_t based on $\{\hat{y}_t\}$. Simulation results show that under some circumstances these approaches really can achieve the efficiency.

2 Cramer-Rao Lower Bound

Consider estimating an n-dimensional random variable X on the basis of an m-dimensional observation Y. Similar to the standard C-R bound, we have [10-13]:

Lemma 1 *Assume that X and Y satisfy the regular condition. Then for all regular estimators of X consisting of an n-dimensional Borel function of Y, namely $f(Y)$, we have*

$$\begin{aligned} & E\{[X - f(Y)][X - f(Y)]'\} \\ \geq\ & I(X|Y)^{-1} \equiv \left\{E\left[\frac{\partial \log p(X|Y)}{\partial X}\frac{\partial \log p(X|Y)}{\partial X'}\right]\right\}^{-1} \\ =\ & -\left\{E\left[\frac{\partial^2 \log p(X|Y)}{\partial X \partial X'}\right]\right\}^{-1}. \end{aligned}$$

Definition 1 *For a state space model, let $X_t = [x_1', x_2', ..., x_t']'$, $J_t = \left[\begin{array}{cc} \underbrace{0,0,...,0}_{(t-1)\times n} & I_{n\times n}\end{array}\right]$, we define the Fisher information matrix and the C-R bound for estimating x_t as*

$$\begin{aligned} I(X_t|Y_t) &\equiv -E\left[\frac{\partial^2 \log p(X_t|Y_t)}{\partial X_t \partial X_t'}\right], \\ B_{t|t-1} &\equiv J_t I(X_t|Y_{t-1})^{-1} J_t',\ B_t \equiv J_t I(X_t|Y_t)^{-1} J_t' \end{aligned}$$

Theorem 2 *For a state space model with $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ satisfying the regular condition, let*

$$\begin{aligned} I_{x_t}(y_t|x_t) &\equiv -E\left[\frac{\partial^2 \log p(y_t|x_t)}{\partial x_t \partial x_t'}\right], \\ I_{x_t}(x_t|x_{t-1}) &\equiv -E\left[\frac{\partial^2 \log p(x_t|x_{t-1})}{\partial x_t \partial x_t'}\right], \\ I_{x_{t-1},x_t}(x_t|x_{t-1}) &\equiv -E\left[\frac{\partial^2 \log p(x_t|x_{t-1})}{\partial x_{t-1} \partial x_t'}\right], \\ I_{x_{t-1}}(x_t|x_{t-1}) &\equiv -E\left[\frac{\partial^2 \log p(x_t|x_{t-1})}{\partial x_{t-1} \partial x_{t-1}'}\right]. \end{aligned}$$

Then, starting from $B_0 = -E\left[\frac{\partial^2 \log p(x_0)}{\partial x_0 \partial x_0'}\right]$, we have the recursive formulae

$$\begin{aligned} & B_{t|t-1} \\ =\ & \{I_{x_t}(x_t|x_{t-1}) - I_{x_{t-1},x_t}(x_t|x_{t-1}) \\ & \times \left[B_{t-1}^{-1} + I_{x_{t-1}}(x_t|x_{t-1})\right]^{-1} I_{x_{t-1},x_t}'(x_t|x_{t-1})\}^{-1}, \end{aligned}$$

$$\begin{aligned} B_t &= \left[I_{x_t}(y_t|x_t) + B_{t|t-1}^{-1}\right]^{-1} \\ &= B_{t|t-1} - B_{t|t-1}\left[I_{x_t}^{-1}(y_t|x_t) + B_{t|t-1}\right]^{-1} B_{t|t-1}. \end{aligned}$$

Corollary 3 *If a state space model has a linear Gaussian state equation*

$$x_t = T_t x_{t-1} + u_t, \quad u_t \sim N\left(0, \sigma_{u,t}^2\right), \tag{6}$$

the C-R bound satisfies the following Riccati equation:

$$B_t = \left\{I_{x_t}(y_t|x_t) + \left[T_t B_{t-1} T_t' + \sigma_{u,t}^2\right]^{-1}\right\}^{-1}. \tag{7}$$

In addition, if a constant dynamic system

$$T_t = T, \quad \sigma_{u,t}^2 = \sigma_u^2, \quad I_{x_t}(y_t|x_t) = I_x,$$

satisfies the stable or completely detectable conditions (p77 in [2]), there is a limit C-R bound, namely B, given by the following equation

$$B = \lim_{t\to\infty} B_t = \left[I_x + \left(TBT' + \sigma_u^2\right)^{-1}\right]^{-1}. \tag{8}$$

3 Efficiency for Outlier and Frequency Demodulation Models, Calculating Limiting Distribution of Ergodic Mackov Chains.

Outlier Model. In (2), let both $\{x_t\}$ and $\{y_t\}$ be scalar,

$$\sum_{k=1}^{p} w_k m_k = 0, \quad \sigma_v^2 \equiv Var\left(\frac{v_t}{H}\right) = \sum_{k=1}^{p} w_k \frac{\sigma_k^2 + m_k^2}{H^2}.$$

the best linear filter is given by the Kalman filter

$$\begin{cases} \hat{x}_t = \left[T\sigma_v^2 \hat{x}_{t-1} + \left(T^2 s_{t-1} + \sigma_u^2\right) y_t\right] / \left(\sigma_v^2 + T^2 s_{t-1} + \sigma_u^2\right), \\ s_t = \sigma_v^2 \left(T^2 s_{t-1} + \sigma_u^2\right) / \left(\sigma_v^2 + T^2 s_{t-1} + \sigma_u^2\right). \end{cases}$$

We can show that

$$s_t \rightarrow S = \frac{\sqrt{\left[\sigma_v^2\left(1-T^2\right)+\sigma_u^2\right]^2 + 4T^2\sigma_v^2\sigma_u^2}}{2T^2} - \frac{\sigma_v^2\left(1-T^2\right)+\sigma_u^2}{2T^2}. \tag{9}$$

On the other hand,

$$\begin{aligned} I_x &= E\left[\frac{\partial p\left(y_t|x_t\right)}{\partial x_t}\frac{1}{p\left(y_t|x_t\right)}\right]^2 \\ &= \int \frac{\left[\sum_{k=1}^{p} \frac{w_k}{\sigma_k^2}\left(v_t - m_k\right) G\left(v_t; m_t, \sigma_k^2\right)\right]^2}{\sum_{k=1}^{p} w_k G\left(v_t; m_t, \sigma_k^2\right)} dv_t. \end{aligned}$$

When $\inf_{i\neq j} \frac{|m_i - m_j|}{\sigma_i + \sigma_j} \geq 3$, there is a partition $\{I_k, k = 1, 2..., p\}$ such that $G\left(v_t; m_j, \sigma_j^2\right) \approx 0$ for all $j \neq k$ when $v_t \in I_k$, then $I_x \approx \sum_{k=1}^{p} \frac{w_k}{\sigma_k^2}$. Thus,

$$B = \lim_{t\rightarrow\infty} B_t = \frac{\sqrt{\left[\left(1-T^2\right)/I_x + \sigma_u^2\right]^2 + 4T^2\sigma_u^2/I_x}}{2T^2} - \frac{\left(1-T^2\right)/I_x + \sigma_u^2}{2T^2}. \tag{10}$$

Calculating S in (9) and B in (10), we can see that Kalman filter is efficient enough and non-linear filtering may not be necessary when $|T| \leq .6$ and $\sigma_u^2 \leq .01$. However, a big difference occurs for large T. When $T = 1$ (random walk state model used in [4]) and $\sigma_u^2 \geq .01$, the C-R bound is only about one-sixth of the linear limit. In such a case, non-linear filtering is necessary.

Frequency Demodulation: The model (4) and its modified versions are very important and have been widely investigated in communications, radar and sonar frequency tracking and other areas [2, 3, 6, 7, 10]. EKF was used for this problem as a successful example [2]. The well-known technique PLL in communications can be derived from EKF [2, 3, 10].

Taking the first order Taylor expansion in (4),

$$y_t \approx A\begin{bmatrix} \sin\hat{x}_{t-1} \\ \cos\hat{x}_{t-1} \end{bmatrix} + A\begin{bmatrix} \cos\hat{x}_{t-1} \\ -\sin\hat{x}_{t-1} \end{bmatrix}\left(x_t - \hat{x}_{t-1}\right) + v_t, \tag{11}$$

then the EKF [2] is

$$\begin{cases} \hat{x}_t = \hat{x}_{t-1} + \frac{A\left(s_{t-1}+\sigma_u^2\right)}{A^2\left(s_{t-1}+\sigma_u^2\right)+\sigma_v^2}\begin{bmatrix} \cos\hat{x}_{t-1} & -\sin\hat{x}_{t-1} \end{bmatrix} y_t, \\ s_t = \frac{\sigma_v^2\left(s_{t-1}+\sigma_u^2\right)}{A^2\left(s_{t-1}+\sigma_u^2\right)+\sigma_v^2}. \end{cases} \tag{12}$$

We can show that

$$s_t \rightarrow S = \frac{\sqrt{1 + 4\frac{\sigma_v^2}{A^2\sigma_u^2}} - 1}{2}\sigma_u^2. \tag{13}$$

Then, let $G = \frac{A\left(S+\sigma_u^2\right)}{A^2\left(S+\sigma_u^2\right)+\sigma_v^2}$, $K = AG$, we have the limit Kalman filter

$$\hat{x}_t = \hat{x}_{t-1} + G\begin{bmatrix} \cos\hat{x}_{t-1} & -\sin\hat{x}_{t-1} \end{bmatrix} y_t. \tag{14}$$

Above the "threshold of linear PLL" [10], $\sin\left(x_t - \hat{x}_{t-1}\right) \approx x_{t-1} - \hat{x}_{t-1} + u_t$, so

$$\hat{x}_t \approx K x_t + (1-K)\hat{x}_{t-1} + G\varepsilon_t, \tag{15}$$

where $\varepsilon_t \equiv \begin{bmatrix} \cos\hat{x}_{t-1} & -\sin\hat{x}_{t-1} \end{bmatrix} v_t$ is Gaussian white noise. Thus, using (13), we can show that under (15) the limit MSE of EKF-PLL is

$$\lim_{t\rightarrow\infty} E\left(x_t - \hat{x}_t\right)^2 = \frac{\sigma_u^2(1-K)^2 + G^2\sigma_v^2}{1-(1-K)^2} = S. \tag{16}$$

On the other hand, it follows from (5) that

$$I_x = E\left[\frac{A}{\sigma_v^2} v_t' \begin{bmatrix} \cos x_t \\ -\sin x_t \end{bmatrix}\right]^2 = \frac{A^2}{\sigma_v^2}.$$

Substituting this into (8), we confirm that the C-R bound for this model is exactly the same as the limit MSE of EKF-PLL given in (16). However, (16) holds only if the approximation (15) is valid. Below the "threshold of linear PLL", we need the following more sophisticated technique to analyze EKF-PLL.

Limiting distribution of ergodic Markov chain. Let $\tilde{x}_t \equiv x_t - \hat{x}_t \pmod{2\pi}$, it follows from (14) that

$$\tilde{x}_t = \tilde{x}_{t-1} + u_t - K\sin\left(\tilde{x}_{t-1} + u_t\right) - G\varepsilon_t. \tag{17}$$

We can show that it is a Harris ergodic Markov chain. So, starting from any initial distribution, we will obtain the same limiting distribution. The problem is how to find its limiting distribution? For continuous time case, a close form solution is showed in [14]. However, there is no close form in our case and only a numerical method based on Fourier transform is investigated [15]. General methods for calculating limiting distribution can be found in [16]. Here we introduce a novel method for calculating the distribution of general ergodic Markov chains so that many non-linear filters can be analyzed more accurately.

For an ergodic discrete time, continuous state Markov chain $\{x_t\}$ with transit pdf $p(x|y)$ on a finite $2D$ support set Π. let $\{a_k\}, -\infty < a_0 < a_k < a_{k+1} < a_n < \infty$, be a partition of x_t's state space such that $a_k - a_{k-1} = \frac{a_n - a_0}{n} \equiv \Delta$. Let $\pi(x)$ be the limiting distribution of $\{x_t\}$ and

$$\pi_k = \int_{a_{k-1}}^{a_k} \pi(x)\, dx, \quad \Pi_n = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{bmatrix},$$

and

$$\begin{aligned}
\hat{x}_j &= \frac{a_{j-1}+a_j}{2}, \quad \hat{y}_k = \frac{a_{k-1}+a_k}{2}, \\
\hat{p}_{kj} &= \frac{p(\hat{x}_j|\hat{y}_k)}{\sum_{j=1}^{n} p(\hat{x}_j|\hat{y}_k)}, \quad \hat{P}_n = [\hat{p}_{kj}]_{k,j=1,2,\ldots,n}, \\
\hat{Q}_n &\equiv \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -\hat{p}_{12} & 1-\hat{p}_{22} & \cdots & -\hat{p}_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ -\hat{p}_{1n} & -\hat{p}_{2n} & \cdots & 1-\hat{p}_{nn} \end{bmatrix}, \\
\hat{\Pi}_n &= \begin{bmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \\ \vdots \\ \hat{\pi}_n \end{bmatrix} = \hat{Q}_n^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\
\delta_n &= \left[\begin{array}{cccc} \frac{\hat{\pi}_1-\pi_1}{\Delta} & \frac{\hat{\pi}_2-\pi_2}{\Delta} & \cdots & \frac{\hat{\pi}_n-\pi_n}{\Delta} \end{array}\right],
\end{aligned}$$

Theorem 4 *Under above notations, assume that $p(x|y)$ is continuous on the finite support set Π, we have $\delta_n \delta_n^T \to 0$ as $n \to \infty$. Then*

$$\begin{cases} \sum_j \hat{x}_j \hat{\pi}_j \to \lim E(x_t) = \int x\pi(x)\, dx, \\ \sum_j [\hat{x}_j - E(x_t)]^2 \hat{\pi}_j \to \lim Var(x_t) \\ \quad = \int [x - E(x_t)]^2 \pi(x)\, dx. \end{cases}$$

Combined with C-R bound, this result can be used for calculating the efficiency of PLL as well as other interesting things such as average jumping times. In general, this approach allows us to analyze the performance of many non-linear filters.

4 Gaussian approximation and hybrid filters based on partial conditional expectation transformation and linear projection

In cases that the conventional methods like KF and EKF, which can be viewed as approximations in time domain, are far away from efficiency, we suggest several ways for approximations in probabilistic domain. First of all, we have the following result [17]:

Lemma 5 *Suppose that x and μ_1 are n-dimensional vectors, μ_2 is a m-dimensional vector, Σ_1 and Σ_2 are $n \times n$ and $m \times m$ positive definite matrices respectively, T is an arbitrary $m \times n$ matrix. Let*

$$\Sigma = \left(\Sigma_1^{-1} + T'\Sigma_2^{-1}T\right)^{-1}, \qquad \mu = \Sigma\left(\Sigma_1^{-1}\mu_1 + T'\Sigma_2^{-1}\mu_2\right)$$

then we have

$$\begin{aligned} & G(x;\mu_1,\Sigma_1)\, G(Tx;\mu_2,\Sigma_2) \\ &= G(x;\mu,\Sigma)\, G(T\mu_1;\mu_2,\Sigma_2 + T\Sigma_1 T'), \end{aligned} \tag{18}$$

$$\Sigma = \Sigma_1 - \Sigma_1 T' \left(\Sigma_2 + T\Sigma_1 T'\right)^{-1} T\Sigma_1, \tag{19}$$

$$\mu = \mu_1 + \Sigma T' \Sigma_2^{-1} (\mu_2 - T\mu_1). \tag{20}$$

Gaussian approximation via Taylor expansion in probabilistic domain. For a non-linear state space model with Gaussian or exponential family distributions, we may obtain a good Gaussian approximation by taking the first three terms in the Taylor expansion for the non-linear function in $p(x_t|x_{t-1})$ and $p(y_t|x_t)$. In contrast to EKF which takes the first two terms of the Taylor expansion in time domain models, this approximation may be more accurate.

For the Frequency Demodulation model (5), let $y_t = \rho_t [\sin\alpha_t, \cos\alpha_t]'$, then

$$p(y_t|x_t) = \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{\rho_t^2 - 2\rho_t A\cos(x_t - \alpha_t) + A^2}{2\sigma_v^2}\right). \tag{21}$$

When $\frac{\rho_t A}{\sigma_v^2}$ is not too small, on each interval $[\alpha_t + (2k-1)\pi, \alpha_t + (2k+1)\pi]$, $k = 0, \pm 1, \ldots$, this function is very close to a normal pdf with the mode at $\alpha_t + 2k\pi$. Since we consider $p(y_t|x_t)\, G(x_t; \hat{x}_{t-1}, s_{t|t-1})$ in the Bayesian formula, if $s_{t|t-1}$ is not too large, we only need to consider the approximation on the interval which covers $\hat{x}_{t-1}$. Choosing α_t such that $|\alpha_t - \hat{x}_{t-1}| \leq \pi$ and approximating $\cos(x_t - \alpha_t)$ by $1 - (x_t - \alpha_t)^2/2$ in (21), when $s_{t|t-1}$ is not too large (<1, say), we obtain the following approximation: For $|x_t - \alpha_t| \leq \pi$,

$$p(y_t|x_t) \approx \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left(-\frac{(\rho_t - A)^2}{2\sigma_v^2}\right) G\left(x_t; \alpha_t, \frac{\sigma_v^2}{\rho_t A}\right). \tag{22}$$

Thus, applying Lemma 5 to the Bayesian formula consisting of $p(x_t|x_{t-1})$ in (5) and $p(y_t|x_t)$ in (22), we have the following algorithm:

$$\begin{cases} s_{t|t-1} = s_{t-1} + \sigma_u^2 \\ \hat{x}_t = \hat{x}_{t-1} + \frac{\rho_t A(\alpha_t - \hat{x}_{t-1})}{\sigma_v^2 + \rho_t A s_{t|t-1}}, \\ s_t = \frac{\sigma_v^2 s_{t|t-1}}{\sigma_v^2 + \rho_t A s_{t|t-1}}. \end{cases} \tag{23}$$

This algorithm can also attain the C-R bound for the model (4) provided (22) holds. The condition (22) is easily to be checked since it does not depend on previous estimator $\hat{x}_{t-1}$ too much. The full details of this method for the FM demodulation case are in [17].

Gaussian approximation via by maximum entropy criterion. For a general state space model with $p(x_t|x_{t-1})$ and $p(y_t|x_t)$, start from initial values $\hat{x}_0$ and s_0, calculate the following estimators recursively:

$$\begin{aligned}
\hat{x}_{t|t-1} &= \int x_t \int p(x_t|x_{t-1})\, G(x_{t-1};\hat{x}_{t-1},s_{t-1})\, dx_{t-1} dx_t,\\
s_{t|t-1} &= \int \left(x_t-\hat{x}_{t|t-1}\right)^2\\
&\quad \times \left[\int p(x_t|x_{t-1})\, G(x_{t-1};\hat{x}_{t-1},s_{t-1})\, dx_{t-1}\right] dx_t,\\
C_t &= \int p(y_t|x_t)\, G\left(x_t;\hat{x}_{t|t-1},s_{t|t-1}\right) dx_t,\\
\hat{x}_t &= \frac{\int x_t p(y_t|x_t)\, G\left(x_t;\hat{x}_{t|t-1},s_{t|t-1}\right) dx_t}{C_t},\\
s_t &= \frac{\int (x_t-\hat{x}_t)^2 p(y_t|x_t)\, G\left(x_t;\hat{x}_{t|t-1},s_{t|t-1}\right) dx_t}{C_t}.
\end{aligned}$$

Here, assume that $p(x_{t-1}|Y_{t-1}) = G(x_{t-1};\hat{x}_{t-1},s_{t-1})$ is Gaussian, we calculate the conditional expectation and variance. Based on this information (constraints), the maximum entropy distribution must be a Gaussian distribution with the same conditional expectation and variance. Thus, we obtain a Gaussian approximation. This method has been applied to the outlier model (2) and had good results.

Hybrid filters based on partial conditional expectation transformation and linear projection. Linear filters are based on "CORRELATION" in time domain instead of "DEPENDENCE" in probabilistic domain. They involve "PROJECTION" that is based on the first and second moments instead of the distributions. This is why linear filters are much simple and fast but may not be always efficient. In such a light, we may use the information in distributions to design a suitable non-linear transform to convert "DEPENDENCE" into "CORRELATION". Then simple and fast linear filters can be applied to the transferred observations. The first question is how to find the best transformation? We have the following answer:

Lemma 6 *Let ξ and η be $p-$vector and $q-$vector r.v.'s with finite second moments. Then for any Borel function $f : R^q \longrightarrow R^p$, no matter linear or non-linear, such that $E|f(\eta)|^2 < \infty$, we have*

$$\frac{|E[\xi' E(\xi|\eta)]|^2}{E|\xi|^2 E|E(\xi|\eta)|^2} \geq \frac{|E[\xi' f(\eta)]|^2}{E|\xi|^2 E|f(\eta)|^2}.$$

This coincides with that the Total Conditional Expectation (TEC) $E(x_t|y_t,y_{t-1},...,y_1)$ is the globally optimal. However, to calculate TCE is usually not possible: Even for the outlier model with p components, Lemma 5 tells us that computational complexity is $O(p^t)$. However, the Partial Conditional Expectation (PCE) $\hat{y}_{t,n} = E(x_t|y_t,y_{t-1},...,y_{t-n+1})$ is usually tractable for not too large n, though PCE does not utilize the information in the whole history. So we take PCE as the non-linear transform to increase correlation between x_t and observations firstly and design a linear filter to track x_t based on the whole history $\{\hat{y}_{t,n}, t=n,n+1,...\}$ secondly. To perform the second step, we need the following results:

Theorem 7 *Suppose that $\{[x_t,y_t], t=0,1,...\}$ is second order stationary and $\{x_t\}$ satisfies*

$$x_t = \rho x_{t-1} + u_t, \quad E(u_t)=0, \tag{24}$$

where u_t is independent to $\{y_{t-k}, k>0\}$. Let $e_{t,n} \equiv \hat{y}_{t,n} - Proj_{H(t-1)}\hat{y}_{t,n}$, $H(t-1)$ be the Hilbert space spanned by $\{\hat{y}_{t-1,n},\hat{y}_{t-2,n},...\hat{y}_{1,n}\}$. The best linear predictor of x_t on the basis of $\{\hat{y}_{k,n}, k \leq t\}$ is given by

$$\hat{x}_{t,n} \equiv Proj_{H(t-1)}\hat{x}_{t,n} = \rho\hat{x}_{t-1,n} + \frac{E(x_t e_{t,n})}{E(e_{t,n}^2)} e_{t,n}.$$

The prediction error is

$$s_{t,n} \equiv E(x_t-\hat{x}_{t,n})^2 = \rho^2 s_{t-1,n} + \sigma_u^2 - \frac{[E(x_t e_{t,n})]^2}{E(e_{t,n}^2)}.$$

In particular, $\hat{x}_{t,t} = E(x_t|Y_1^t)$ is the best predictor among all linear and non-linear functions of Y_1^t. In this case we have

$$e_{t,t} = \hat{y}_{t,t} - \rho\hat{y}_{t-1,t-1}, \qquad E(x_t e_{t,t}) = E(e_{t,t}^2).$$

Corollary 8 *When $\{x_t,y_t\}$ is stationary ($|\rho|<1$ in (24)), we have*

$$E(x_t e_{t,n}) \to S_{xe}(n), \quad E(e_{t,n}^2) \to S_{ee}(n), \quad \text{as } t\to\infty.$$

So, we have an approximately linear optimal time-invariant filter

$$\hat{x}_{t,n} = \rho\hat{x}_{t-1,n} + \frac{S_{xe}(n)}{S_{ee}(n)} e_{t,n}, \tag{25}$$

$$E(x_t-\hat{x}_{t,n})^2 \to S(n) = \frac{\sigma_u^2 - S_{xe}^2(n)/S_{ee}(n)}{1-\rho^2}.$$

Corollary 9 *When $\rho = 1$ in (24), let $v_{t,n} = x_t - \hat{y}_{t,n}$, $z_{t,n} = \hat{y}_{t,n} - \hat{y}_{t-1,n}$, then $\{v_{t,n}, z_{t,n}\}$ is stationary. Then*

$$\begin{aligned}
\hat{x}_{t,n} &\equiv \mathrm{Proj}_{H(\hat{y}_{t,n},\hat{y}_{t-1,n},...)}\hat{x}_{t,n}\\
&= \hat{y}_{t,n} + \mathrm{Proj}_{H(z_{t,n},z_{t-1,n},...)} v_{t,n}.
\end{aligned} \tag{26}$$

Using these results and Lemma 5 to the outlier model, we have

$$p\left(x_t|y_t, y_{t-1}, ..., y_{t-n}\right) = \sum_{i=1}^{p^n} W_{t,i}^n G\left(x_t; \mu_{t,i}^n, s_i^n\right),$$

so

$$\hat{y}_{t,n} = E\left(x_t|y_t, y_{t-1}, ..., y_{t-n}\right) = \sum_{i=1}^{p^n} W_{t,i}^n \mu_{t,i}^n.$$

To calculate the parameters in (25) and (26), we need the covariance functions of $\{x_t, \hat{y}_{t,n}\}$ in Corollary 8 or $\{v_{t,n}, z_{t,n}\}$ in Corollary 9. Two methods can be considered: 1. Numerical integration based on the probabilistic structure; 2. Monte Carlo method to generate the sample covariance functions. Though Method 2 also involves Monte Carlo, comparing with MCMC, we only need that in designing the linear filter of $\{\hat{y}_{t,n}\}$. After the design, on-line in-real time solution will be obtained. On the basis of covariance functions, a variety of time series models, like AR, MA (FIR) or ARMA (IIR) models, can be used to fit $\{x_t, \hat{y}_{t,n}\}$ in Corollary 8 or $\{v_{t,n}, z_{t,n}\}$ in Corollary 9. Simulation was done and show that for very low SNR and rapidly changed signals, this method may be superior than others.

Acknowledgment: This work has been supported by two Large Australian Research Committee's Grants. Prof. B.G. Quinn has been the co-investigator of both Grants with the author. The late Prof. E.J. Hannan was also a co-investigator for the first Grant.

References

[1] A.H. Jazwinski, "Stochastic processes and filtering theory", Academic Press, INC, 1970.

[2] B.D.O. Anderson and J.B. Moore, "Optimal filtering", Prentice-Hall, INC, 1979.

[3] R.J. Elliott, L. Aggoun and J. B. Moore, "Hidden Markov Models: Estimation and Control", Springer-Verlag, 1995.

[4] N. Shephard, "Partial non-gaussian state space", Biometrika, Vol. 81, No. 1, 1994, pp 115-131.

[5] Sorenson, H.W. and D.L. Alspach (1971), Recursive Bayesian estimation using Gaussian sums, Automatica, Vol. 7, pp465-479.

[6] R.S. Bucy and A.J. Mallinckrodt, "An optimal phase demodulator", Stochastics, Vol. 1, 1973, pp3-23.

[7] P.K.S. Tam and J.B. Moore, "A Gaussian sum approach to phase and frequency estimation, IEEE Trans. on Communications", Vol. 25, 1977, pp935-942.

[8] S.C. Kramer and H.W. Sorenson, "Recursive Baysian estimation using piece-wise constant approximations", Automatika, Vol. 24, 1988, pp789-801.

[9] M.K. Cowles and B.P. Carlin, "Markov chain Monte Carlo convergence diagnostics: a comparative review", JASA, 91, 1996, pp883-904.5.

[10] H.L. Van Trees, "Detection, estimation and modulation theory", Part I, 1969, Part II, 1971, John Wiley & Sons.

[11] B.Z. Bobrovsky and M. Zakai, "A lower bound on the estimation error for Markov processes", IEEE Trans. Automat. Contr., Vol. AC-20, no. 6, 1975, pp785-788.

[12] J.I. Galdos, "A Cramer-Rao bound for multidimentional discrete-time dynamical systems", IEEE Trans. Automat. Contr., Vol. AC-25, no. 1, 1980, pp117-119.

[13] P.C. Doerschuk, "Cramer-Rao bounds for discrete-time nonlinear filtering problems", IEEE Trans. Automat. Contr., Vol. AC-40, no. 8, 1995, pp1465-1469.

[14] A. J. Viterbi, "Phase-locked loop dymanics in the presence of noise by Fokker-Planck techniques", Proc. IEEE 51, 1963, pp1737-1753.

[15] C.M. Chie, "Mathematical analogies between first-order digital and analog phase-locked loops", IEEE Trans. Communication, Com-26, 1978, pp860-865.

[16] H. Tong, "Non-linear time series: a dynamical system approach", Oxford Science Publications, 1990.

[17] D. Huang, "A tractable Bayesian filter via Gaussian density product and Gaussian approximation with application to FM demodulation", Proceedings of ISSPA'96, Gold Cost, 1996, pp. 421-422.