A Gaussian Sum Approach to Phase and Frequency Estimation

PETER K. S. TAM AND JOHN B. MOORE, MEMBER, IEEE

Abstract—In this paper, a theory of optimal nonlinear estimation from sampled data signals where the a posteriori probability densities are approximated by Gaussian sums is adapted for application to phase and frequency estimation in high noise. The nonlinear estimators (demodulators) require parallel processing of the received signal. In the limit as the number of parallel processors becomes infinite the FM demodulators become optimum in a minimum mean square error sense and the PM demodulators become optimum in some well defined sense. For the clearly suboptimal case of one processor, the demodulators can be readily simplified to the familiar phase-locked loop. On the other hand, for the intermediate case, significant extension of the phase-locked loop threshold is achieved where (say) six parallel processors are involved.

1. INTRODUCTION

DURING the past decade, the tools of stochastic estimation theory have been applied to the synthesis of phase and frequency estimators (angle demodulators), see for example [1-16]. This paper represents a further contribution in this direction.

In nonlinear estimation [10-24], an attempt is usually made to calculate the relevant a posteriori probability density functions—or at least sufficient statistics of these functions. For the case of interest to us here where the state vector $x(t_k)$ is a first order Markov process, knowledge of the signal model equations, the a posteriori density function $p[x(t_k) \mid t_k]$ at time t_k, and the new measurement at time t_{k+1}, is sufficient for an update of the a posteriori density function to $p[x(t_{k+1}) \mid t_{k+1}]$. With knowledge of the a posteriori density function $p[x(t_k) \mid t_k]$ at each time instant t_k, either a MAP (maximum a posteriori) or MMSE (minimum mean square error) estimate or other type of estimate can be calculated.

With the assumption that the low order moments of the posteriori density function are a set of sufficient statistics [18], optimal (or near optimal) estimators for a limited class of problems can be derived. The assumptions are good ones when the estimation error is small or equivalently in situations characterized by low noise, good prior information, and polynomial type nonlinearities. The extended Kalman filter [18] is the most common form of such a nonlinear estimator.

It is possible to apply extended Kalman filter theory to yield quasi-optimum (in the mean square error sense) angle demodulators [3,4]. When simplified, the continuous time extended Kalman filter for angle demodulation is nothing

other than the familiar phase-locked loop. The approximation inherent in the derivations of the extended Kalman filter for angle demodulation is that the error covariance P_θ of the phase estimate be less than a threshold value (typically we require $P_\theta \leqslant 0.25$ rad^2 [3-5]). When this constraint is violated as it most surely is in a high noise environment, the extended Kalman filter is no longer optimal (or near optimal) and there is in fact a severe degradation of performance as the noise level increases beyond a threshold level determined by the threshold constraint on P_θ.

Similar applications of nonlinear estimation theory for discrete measurements result in digital phase-locked loops and quasi-optimum sampled-data angle demodulators [5-8]. On the one hand, simulations show that these sampled-data demodulators have virtually the same threshold as the counterpart analog schemes. On the other hand, because of recent advances in large-scale integrated semiconductor technology for digital circuits, in some applications the digital demodulators are more attractive; certainly they are easier to simulate on general purpose computers when developing new approaches to demodulation, as in this investigation. Because of the recent advances in technology it appears very reasonable to consider more sophisticated digital circuits than those in [5-8] to improve the performance of angle demodulators in high noise environments.

In this paper the nonlinear Bayesian estimation algorithms using a Gaussian sum approximations for the a posteriori densities [19, 23] are the starting point for the development of novel angle demodulators. In the Gaussian sum approach [19, 23] the key idea is to approximate the a posteriori density as a sum of Gaussian densities where the covariance of each Gaussian density is sufficiently small for the time evolution of its mean and covariance to be calculated accurately using the extended Kalman filter algorithm. The nonlinear estimator of [23] consists of a bank of extended Kalman filters where the signal estimate is a weighted sum of the outputs of the filters. The weightings are calculated from the residuals of the extended Kalman filters. In the examples of [23], the noise is sufficiently low for such an estimator to be near optimal. For high noise situations, [23] broadly outlines predcures which appear to be unnecessarily complicated when applied to FM demodulation—they involve checking the statistics of the estimator measurement residuals for consistency with their theoretical properties as a means for detecting divergence, and a re-initialization of the algorithm to incorporate more filters.

Here attention is focused on the question of how and when to re-initialize the Gaussian sum nonlinear filtering algorithm of [23] in a high noise environment—at least for the FM demodulation case. In particular, to achieve near optimal

performance we study a periodic re-initialization of the algorithm in such a way that the conditional phase error covariance P_{θ_i} associated with each Gaussian density satisfies a constraint—typically $P_{\theta_i} \leqslant 0.25$ rad^2. Relaxing this constraint enables a tradeoff between performance and algorithm complexity. Other ideas are also studied which dramatically reduce algorithm complexity with very small loss in performance.

Other optimal phase and frequency estimation schemes are reported in [9-15]. In [9], the conditional probability distributions of the state vector are approximated by a sum of point mass probabilities in such a way that the Bayesian recursion rules can be readily applied. In the limit as the number of point masses becomes infinite, the estimation algorithm is optimal. However, when the number of terms in the summation is small the approximation does not appear to be as good as that when a Gaussian sum is employed. In [14], the optimal estimate is obtained via a Fourier series representation of the state probability densities. The Fourier coefficients can be updated recursively using a set of infinite dimensional equations. Several techniques to obtain suitable finite dimensional approximations of these equations are discussed in [11,15]. However, except for a few simple cases, these 'truncation' procedures are rather involved.

We also have to mention that in the interesting case when there is only phase drift or frequency shift but *no* additive measurement noise, the optimal phase or frequency estimator consists of a nonlinear preprocessor, a Kalman-Bucy filter and a nonlinear postprocessor. The reader is referred to [12] for a more detailed description of such an estimation.

In Section 2, the state-variable representation of the angle modulated signal [3-16] is reviewed as is a sampling technique of [8] to reduce the signal to a baseband process. Section 3 contains the derivation of our parallel processing demodulator for the model of Section 2, using as a starting point the Gaussian sum approach of [23]. Section 4 studies very useful simplifications to the algorithms. In Sections 5-7, simulation results for a phase tracking example and an FM demodulation example are discussed and and conclusions concerning the results of the paper drawn.

2. STATE-VARIABLE MODEL AND A SAMPLING TECHNIQUE

In order to apply the tools of stochastic estimation theory to the phase and frequency estimation problems, the first step is to set up a model in terms of a finite-dimensional nonlinear stochastic system. For simplicity, we assume that the received angle-modulated signal process is a scalar process contaminated by additive white Gaussian noise, and that the message to be estimated is a sample function from a zero-mean Gaussian random Markov process. An appropriate state-variable model which can be used for a variety of problems is phase tracking, phase demodulation and frequency demodulation, is as follows [3-16]:

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) \qquad (2.1)$$

$$\theta(t) = c'x(t), \qquad c' = [0 \quad 0 \cdots 0 \quad 1] \qquad (2.2)$$

$$z(t) = \sqrt{2}\, a \cos\, [2\pi f_0 t + \theta(t)] + v(t) \qquad (2.3)$$

where $z(\cdot)$ is the received noisy process, $x(t)$ an n-column vector, $(\sqrt{2}a)$ the carrier amplitude and f_0 the carrier frequency; $u(\cdot)$, $v(\cdot)$ are independent white Gaussian processes and $x(t_0)$ is a random vector independent of $u(\cdot)$ and $v(\cdot)$; $x(t_0)$ and $u(\cdot)$ can include the effects of such uncertainties as unknown Doppler shifts.

The various statistical parameters (assumed known) are

$$E\{x(t_0)\} = x_0, \quad E\{x(t_0)x'(t_0)\} = P_0$$

$$E\{u(t)\} = 0, \qquad E\{u(t)u'(\tau)\} = Q(t)\delta(t-\tau), \quad Q(t) \geqslant 0$$

$$E\{v(t)\} = 0, \qquad E\{v(t)v(\tau)\} = r(t)\delta(t-\tau), \quad r(t) > 0$$

$$E\{u(t)v(\tau)\} = 0. \qquad (2.4)$$

Using the in-phase and quadrature-phase sampling technique as described in [8], the scalar noisy FM process described by (2.3) can be represented by the following equivalent 2-vector sampled data process:

$$s(t_k) = h[x(t_k)] + w(t), \qquad k = 0, 1, 2, \cdots \qquad (2.5)$$

where

$$h[x(t_k)] = \sqrt{2}a \begin{bmatrix} \cos\theta(t_k) \\ -\sin\theta(t_k) \end{bmatrix}.$$

Here $h[x(t_k)]$ is the signal sample vector free of noise and $w(t_k)$ is the corresponding noise vector for the measurement. Assuming perfect sampling preceded by an ideal bandpass filter, $w(\cdot)$ is a white Gaussian noise sequence with statistical properties readily determined from those of (2.4) using techniques described in [25]. Specifically, if the measurement noise covariance $r(t)$ can be considered constant during adjacent samples and $t_k = kT$, we have $E\{w(t_k)\} = 0$, $E\{w(t_k)w'(t_l)\} = R(t_k)\delta_{kl}$ where

$$R(t_k) = \begin{bmatrix} r_k & 0 \\ 0 & r_k \end{bmatrix}, \qquad r_k \equiv \frac{r(t_k)}{T}.$$

3. BAYESIAN ESTIMATION USING GAUSSIAN SUM APPROXIMATIONS

In nonlinear filtering via Gaussian sum techniques [23], the a posteriori density $p[x(t_k)/t_k]$ is approximated by a sum of M Gaussian terms. The weighting coefficients, the means, and the covariances of all the terms are updated via a bank of M conditional extended Kalman filters; thereby the entire a posteriori density is updated as time evolves and as more measurements become available. For the Gaussian sum filter to be near optimal, the Gaussian sum approximations have to be good ones and also the covariance of each Gaussian term has to be small.

In this section, this Bayesian Estimation technique of [23] is adapted to yield a solution to our problem of phase and frequency estimation in high noise.

Relaxation of Error Covariance Requirements

It is well known [3-5] that in applying extended Kalman filters for FM demodulation, the usual requirements for extended Kalman filters to be near optimal can be relaxed somewhat. Instead of requiring that the state error covariance be small as is usually the case, since the only linearization is that about the last entry of the state vector which is the signal model phase, than only the phase error variance need be small. That is, P_θ must be small. In practice P_θ must satisfy $P_\theta \leqslant 0.25$ rad^2 (typically) in order that the filter be near optimal. Actually the value of 0.25 is just an arbitrary value in the vicinity of a knee in a relationship between filter performance and optimum performance. For P_θ below 0.25 rad^2 the conditional extended Kalman filter performance is near optimum, but above 0.25 rad^2 the filter performance rapidly falls away from the optimum performance.

As a consequence of the above results concerning extended Kalman filters for FM demodulators, there is an immediate relaxation of the error covariance requirements for the Gaussian sum algorithm to be near optimal. In particular, instead of requiring that the entire error covariance of the i-th conditional extended Kalman filter be small we require only that its phase estimation error covariance $P_{\theta_i} = E\{[\theta(t_k) - \theta_i]^2\}$ be small for $i = 1, 2, \cdots, M$. Usually $P_{\theta_i} \leqslant 0.25$ rad (typically) for near optimum performance of the Gaussian sum estimators when applied to FM demodulation. More explicit derivations of these requirements are studied in [26].

Re-Initialization of the Algorithm

In high noise situations, the conditional error covariances of the individual extended Kalman filters of a Gaussian sum estimation will grow with time, so that eventually the assumptions required for the extended Kalman filters to be near optimum will no longer be valid. That is, P_{θ_i} will be greater than 0.25 rad^2. To overcome this, at each time instant, we check that the conditional error covariances (or in our case the P_{θ_i}'s) remain small. If P_{θ_i} exceeds a certain constraint, the whole state probability density is re-constructed using approximations as another sum of Gaussian densities in such a way that the P_{θ_i}'s of the new conditional densities are within the constraint. It might be thought that to achieve this the number of terms in the Gaussian sum to allow a reasonable approximation will increase, but as discussed in more detail later, it is appropriate to consider Gaussian densities folded around a unit circle, in which case six or so such terms are usually adequate.

In summary, our adaption of the Gaussian sum algorithm of [23] is as follows.

Gaussian Sum Algorithm for PM and FM demodulation

$\{1\}$ = Approximate $p[x(t_k)/t_{k-1}]$ with a sum of M Gaussian densities satisfying the constraint $P_{\theta_i} \leqslant \sigma_c^2$.
$\{2\}$ = Apply measurement update equations (see [23]).
$\{3\}$ = Apply time update equations (see [23]).
$\{4\}$ = Set $k + 1 = k$.
$\{5\}$ = Check that $P_{\theta_i} \leqslant \sigma_c^2$. If the constraint is satisfied, go to $\{2\}$; if not, go to $\{1\}$.
Note: In [23], the step $\{5\}$ is simplified as, $\{5\}' =$ go to

$\{2\}$. Clearly the above algorithm is a minor variation on that of [23] but the variation is absolutely vital for the following theorem and subsequent applications.

A result which is rigorously proved in [26] and heuristically established in [23] is

Convergence Theorem: In the limit as $M \to \infty$ and $\sigma_c^2 \to 0$, the Gaussian sum algorithm gives the exact evolution of the a posteriori density function of $p[x(t_k)/t_k]$.

Remark: Earlier comments suggest that reasonable values for σ_c^2 and M are $\sigma_c^2 = 0.25$ rad^2 and $M = 6$, for then a phase probability density around a unit circle can be reasonably approximated by a Gaussian sum with each Gaussian density satisfying the constraint $P_{\theta_i} \leqslant 0.25$ rad^2. With this constraint satisfied the conditional extended Kalman filters are near optimal. Only simulations can verify just how reasonable these values are. We now move to consider a practical implementation and simulations of a demodulator employing the Gaussian sums idea.

4. PRACTICAL IMPLEMENTATION OF THE PARALLEL PROCESSING DEMODULATOR

In this section, we present a simple way of 'reconstructing' $p[x(t_k)/t_{k-1}]$. We start with four useful simplifications.

1. The measurement update equations can be simplified if we start with all the densities in the Gaussian sum having the same covariance. The simplification is quite dramatic in the case of interest here when the error covariance equations are decoupled from the filter state equations. As noted in [16-29], upon application of extended Kalman filter algorithms to quadrature-phase and in-phase sampled FM (or PM) signals, the resulting state estimate update equations and the error covariance update equations *are decoupled.* Hence, if we re-initialize the covariance update equations with the same value of covariance for each of the M Gaussian densities in the Gaussian sum, the updated covariances would also be of the same value.

2. If the Gaussian sum algorithm is used only in the high noise base, then high noise assumptions lead to the simplification that the error variance need not be updated in step $\{2\}$. Also in the high noise case, the conditional innovations covariance can be approximated by the noise variance, thus simplifying the Kalman gain calculations and the Gaussian density weighting coefficient calculation.

3. With the simplifying assumption that the conditional probability distribution for each specified phase is Gaussian, the grid of μ_i taken in the phase-frequency space collaspes to a one dimensional grid. A considerable reduction in the number of grid points and as a consequence a dramatic reduction in computational effort is thereby achieved. Based on the experimental results for the examples studied later, it appears that the assumption here is a reasonable one. Observe that there is no possibility of achieving any simplifications using this Gaussian assumption when applying the point mass approximations of [9].

4. We note that the conditional innovations associated with the individual extended Kalman filters depend on the phase error modulo 2π rather than on the actual phase error.

Therefore, in approximating $p[x(t_k)/t_{k-1}]$ by a Gaussian sum with covariance constraints, it usually suffices to consider only the folded density function obtained by wrapping the density around a circle $\hat{\theta} - \pi \leq \theta < \hat{\theta} + \pi$ [9-15]. The advantages of considering this folded density on a circle rather than the unfolded density on the whole real line should be clear. A finite grid of, say, six points evenly spaced around the circle, would be sufficient to ensure that the new Gaussian sum is a good approximation of the given density.

Employing the above simplifications, we present the following simple procedure for 'reconstructing' a given density. Suppose the given density $p(x)$ is in the form of a Gaussian sum of M terms

$$p(x) = p(\lambda, \theta) = \sum_{i=1}^{M} \alpha_i N[x - x_i, P] \qquad (4.1)$$

with

$$x = \begin{bmatrix} \lambda \\ \theta \end{bmatrix}, \quad x_i = \begin{bmatrix} \lambda_i \\ \theta_i \end{bmatrix}, \quad P = \begin{bmatrix} D & E \\ E' & P_\theta \end{bmatrix}; \qquad (4.2)$$

θ, θ_i, p_θ are scalars and λ, λ_i are $(n-1)$ vectors; and we wish to approximate $p(x)$ with another Gaussian sum of M terms as

$$p(x) \cong \sum_{i=1}^{M} \tilde{\alpha}_i N[x - \tilde{x}_i, \tilde{P}]; \quad \tilde{x}_i = \begin{bmatrix} \tilde{\lambda}_i \\ \tilde{\theta}_i \end{bmatrix}; \quad P = \begin{bmatrix} \tilde{D} & \tilde{E} \\ \tilde{E}' & \tilde{P}_\theta \end{bmatrix}$$

$$(4.3)$$

such that $\tilde{P}_\theta \leq \sigma_c^2$ for some σ_c. Assuming $p(x)$ has mean

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\theta} \end{bmatrix}$$

and covariance

$$\begin{bmatrix} P_1 & P_2 \\ P_2' & P_3 \end{bmatrix}$$

we begin with setting up a grid space as an arc on the circle, chosen to cover the significant part (say $\hat{\theta} \pm 2\sqrt{P_3}$) of the folded density. The grid space is divided into M equal intervals and a grid point $\tilde{\theta}_i$ is placed in the center of each such interval. The phase covariance $\tilde{P}_\theta$ of each term in the new density is simply taken as a fraction of the interval width. We propose that 'reconstruction' of $p(x)$ is necessary when the parameter $P_\theta > \max(\tilde{P}_\theta, 0.25)$.

When $P_\theta > \max(\tilde{P}_\theta, 0.25)$, the other parameters $\tilde{\alpha}_i$, $\tilde{\lambda}_i$, $\tilde{D}$, $\tilde{E}$ are then calculated as follows

$$\tilde{\alpha}_i = c \int p(\lambda, \tilde{\theta}_i) \, d\lambda, \quad c \text{ is a normalizing constant}$$

$$\tilde{\lambda}_i = \frac{1}{\alpha_i} \int \lambda p(\lambda, \tilde{\theta}_i) \, d\lambda + \left[\hat{\lambda} - \sum_{i=1}^{M} \int \lambda p(\lambda, \tilde{\theta}_i) \, d\lambda \right]$$

$$\tilde{D} = P_1 - \sum_{i=1}^{M} \tilde{\alpha}_i (\hat{\lambda} - \tilde{\lambda}_i)(\hat{\lambda} - \tilde{\lambda}_i)'$$

$$\tilde{E} = 0.$$

Here $p(\lambda, \tilde{\theta}_i)$ is treated as the folded density of $p(x)$ around a circle and all integrations are performed from $\tilde{\theta}_i - (\pi/2)$ to $\tilde{\theta}_i + (\pi/2)$. It can be seen that the calculations are designed such that each interval is approximated by a Gaussian density of appropriate mean, covariance and weighting coefficient. For example, the calculated $\tilde{\lambda}_i$ is roughly the conditional mean of λ given $\theta = \theta_i$. (The correction term in square brackets, usually very small, is to ensure that $\Sigma_{i=1}^{M} \tilde{\alpha}_i \tilde{\lambda}_i = \hat{\lambda}$.)

The integrations for calculating $\hat{\alpha}_i$ and $\tilde{\lambda}_i$ can be further simplified to weighted sums of Gaussian densities around a circle. Thus, we have the following algorithm for 'reconstructing' $p(x)$ or more precisely its parameters $\tilde{\alpha}_i$, $\tilde{x}_i$, $\tilde{P}$.

Algorithm (4.1)

{1} = Set up a grid space $[\hat{\theta} - A, \hat{\theta} + A)$, and let $\tilde{P}_\theta = A/(gM)$ where

$A = \min\{\pi, f\sqrt{P_3}\}$
f = design factor (a number between 1.4 and 3, say)
g = design factor (say 1.1)

{2} = If $P_\theta < \max(\hat{P}_\theta, 0.25)$, set $\tilde{\alpha}_i = \alpha_i \tilde{x}_i = x_i$ and $\tilde{P} = P$. EXIT.

{3} = Select equally spaced grid points $\{\tilde{\theta}_i\}$

$$\tilde{\theta}_i = \hat{\theta} + (2i - M - 1) \frac{A}{M}.$$

{4} = For each j, shift the individual θ_i's around in steps of 2π until $\theta_j - \pi \leq \tilde{\theta}_i < \tilde{\theta}_j + k$ for all $i = 1, \cdots, M$. Then calculate

$$a_j = \sum_{i=1}^{M} \alpha_i N[\tilde{\theta}_j - \theta_i, P_\theta]$$

$$b_j = \frac{1}{a_j} \sum_{i=1}^{M} \alpha_i N[\tilde{\theta}_j - \theta_i, P_\theta] \left(\lambda_i - \frac{\tilde{\theta}_j - \theta_i}{P_\theta} E \right).$$

{5} = Calculate

$$s_1 = \sum_{i=1}^{M} \alpha_i, \qquad s_2 = \sum_{i=1}^{M} b_i.$$

{6} = Set $\tilde{\alpha}_i = a_i/s_1$, $\tilde{\lambda}_i = b_i/s_1 - s_2/s_1 + \hat{\lambda}$

$$\tilde{D} = P_1 - \sum_{i=1}^{M} \tilde{\alpha}_i (\hat{\lambda} - \tilde{\lambda}_i)(\hat{\lambda} - \tilde{\lambda}_i)', \quad \tilde{E} = 0.$$

{7} = Calculate $\tilde{P}$, $\tilde{x}_i$, $\tilde{\alpha}_i$ from (4.1). EXIT.

Remark: For $M < 6$, step 1 effectively sets σ_c^2 to be greater than the normally accepted .25 rad². The exact value of the designed σ_c depends on both the design factors f and g. Simulation results show that for small M (2 or 3), we can only allow about 5% change in f or g without affecting significantly the filter performance. For larger M (4, 5 or 6), the filter sensitivity to f or g is much reduced.

5. PHASE TRACKING EXAMPLE

We consider a phase tracking problem that is essentially the same as that considered in [9, 15]. The phase signal is modeled as integrated zero mean, unit variance white Gaussian noise (i.e., as Brownian motion). This phase signal modulates a carrier signal $\sin 2\pi f_0 t$. The transmitted signal is corrupted by additive white Gaussian noise of two-sided spectral magnitude r. The received signal is translated to a baseband discrete-time process via in-phase and quadrature-phase sampling, where the sampling error is assumed negligible.

The appropriate equations for such a model are

$$\dot{\theta}(t) = u(t)$$

$$s(t_k) = \begin{bmatrix} \cos \theta(t_k) \\ \sin \theta(t_k) \end{bmatrix} + w(t_k) \qquad k = 0, 1, 2, \cdots \qquad (5.1)$$

where $t_k = kT$, $\theta(t_0) = 0$; $u(\cdot)$ and $w(\cdot)$ are independent zero mean white Gaussian noise processes with $E\{u(t)u(\tau)\} = \delta(t - \tau)$ and

$$E\{w(t_k)w'(t_l)\} = \begin{bmatrix} \dfrac{r}{T} & 0 \\ 0 & \dfrac{r}{T} \end{bmatrix} \delta_{kl}$$

We wish to estimate the modular phase, $[\theta(\cdot) \bmod 2\pi]$. Following [9, 15], we choose the estimate to minmize the expected value of the loss function $[1 - \cos \tilde{\theta}]$ where $\tilde{\theta}$ denotes the estimation error. This loss function has the nice properties that it is equivalent to the mean-square error when the error is small but has the appropriate modular 2π periodicity when the error is large. For such a loss function, the optimal estimate $\hat{\theta}(t)$ of $[\theta(t) \bmod 2\pi]$ given measurements up to and including t_k is given by $\hat{\theta}(t) = \tan^{-1}(\hat{s}/\hat{c})$ [9, 15] where $\hat{s} = E\{\sin \theta(t) \mid t_k\}$ and $\hat{c} = E\{\cos \theta(t) \mid t_k\}$. If the a posteriori density $p[\theta(t) \mid t_k]$ is expressed as a sum of Gaussian densities

$$p[\theta(t) \mid t_k] = \sum_{i=1}^{M} \alpha_i N(\theta - \theta_i, P_\theta),$$

then $\hat{s}$ and $\hat{c}$ can be readily calculated as [15]

$$\hat{c} = e^{-(P_\theta/2)} \sum_{i=1}^{M} \alpha_i \cos \theta_i,$$

$$\hat{s} = e^{-(P_\theta/2)} \sum_{i=1}^{M} \alpha_i \sin \theta_i.$$

Our phase estimator consists of a bank of extended Kalman filters operating in parallel as described in algorithm 3.1 and 4.1. The phase estimate is obtained as a nonlinear combination of the outputs of the extended Kalman filters, as described above.

The quantity $\sqrt{r}$ in our model represents the *idealized* steady-state phase error variance of the standard phase-locked
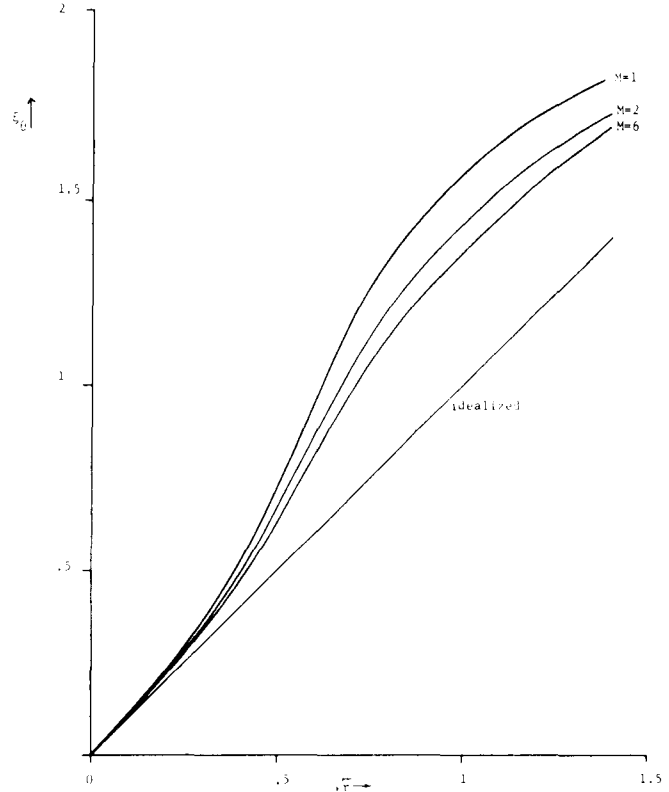

Figure 1. Phase error variance results for phase tracking.

loop for (5.1) *if* the linear assumption holds throughout. It is common practice [4, 9, 15] to take it as the variable to be varied in obtaining performance results.

We apply digital computer simulations to evaluate the performance of our estimator. Following [9], each of our simulation results is obtained through 40 different runs, each of 500 time steps, with each time-step equal to one-tenth of the time-constant of the standard phase-locked loop for (5.1) (i.e., $T = 0.1\sqrt{r}$).

A popular performance measure is of course the mean-square error. In Figure 1, we present plots of ξ_θ versus $\sqrt{r}$ where ξ_θ is the evaluated mean-square phase error, calculated as the average over 40 sample paths of the quantity $1/500 \sum_{k=1}^{500} [\theta(t_k) - \hat{\theta}(t_k)]^2$. M represents the number of extended Kalman filters in the filter bank. The curve $M = 1$ consists of the results of digital phase-locked loop simulations. This is found to be very close to the analytic performance prediction for the phase-locked loop (PLL) computed from the exact error density [4]. In fact, the two curves are nearly identical. The simulated results of the PLL thus indicate the validity of the overall simulation.

Another performance measure, suggested by our choice of the cost function $1 - \cos \tilde{\theta}$ is the average over 40 sample paths of the quantity $1/500 \sum_{k=1}^{500} \{1 - \cos [\theta(t_k) - \hat{\theta}(t_k)]\}$. Table 1 contains some such results.

To compare our results with those of [9, 15], we first reproduce in Table 2 some relevant results of the Fourier Coefficient Filter (FCF) from [15] for comparison of the results in Table 1. It can be seen that while the results for $M = 2$ are about the same as the FCF, the results for $M = 6$ are better than the FCF.

TABLE 1
SIMULATION RESULTS FOR PHASE TRACKING USING
GAUSSIAN SUM $E\{1 - \cos\theta\}$

$\sqrt{r}$	M = 1	M = 2	M = 6	Improvement over corresponding results for M = 1	
				M = 2	M = 6
1.0	0.547	0.508	0.490	7.48%	10.75%
0.7	0.433	0.395	0.375	8.77%	13.39%
0.3	0.226	0.209	0.204	8.25%	9.73%

TABLE 2
SIMULATION RESULTS FROM [15] $E\{1 - \cos\theta\}$

$\sqrt{r}$	PLL	FCF	Improvement of FCF over corresponding results for PLL
1.0	0.5312	0.5002	5.84%
0.7	0.4098	0.3687	10.03%
0.4	0.2447	0.2253	7.93%

TABLE 3
COMPARISON OF SIMULATION RESULTS FOR $\sqrt{r} = 1.0$
MEAN-SQUARE ERROR

		Mean-square error rad^2	Improvement over PLL (or M = 1)
results from [9]	PLL	1.614	
	Bayesian Estimator	1.395	13.57%
results from [15]	PLL	1.575	
	FCF	1.446	8.19%
results using Gaussian estimator	M = 1	1.586	
	M = 2	1.434	9.58%
	M = 6	1.374	13.37%

As a further comparison, we present in Table 3 the percentage improvement of the evaluated mean-square error over the simulated phase-locked loop (PLL) for the three different types of nonlinear phase estimators reported here and in [9, 15] for $\sqrt{r} = 1$. It can be seen that results for $M = 2$ are about the same as the FCF of [15], while results for $M = 6$ are about the same as the Bayesian estimator of [9]. Note that our estimator for $M = 6$ requires fewer computations than the estimator of [9], where the simulation results are obtained using 100 discrete mass points (or 100 filters in the filter bank).

6. FM DEMODULATION EXAMPLE

As an example in FM demodulation, we consider a single-pole message model corresponding to the first-order Butterworth message spectrum. The state $x(t)$ of the FM model is a 2-dimensional vector composed of the message $\lambda(t)$ and the phase $\theta(t)$. The transmitted signal is $\sqrt{2}\cos[2\pi f_0 t + \theta(t)]$ and is corrupted by additive white Gaussian noise of two-sided spectral magnitude r. The received signal is translated to a baseband process via in-phase and quadrature sampling. Sampling error and oscillator instability are assumed negligible.

The appropriate equations for such a model are

$$\dot{x}(t) = \begin{bmatrix} \dot{\lambda}(t) \\ \dot{\theta}(t) \end{bmatrix} = \begin{bmatrix} -\alpha & 0 \\ c & 0 \end{bmatrix} \begin{bmatrix} \lambda(t) \\ \theta(t) \end{bmatrix} + \begin{bmatrix} u(t) \\ 0 \end{bmatrix}$$

$$s(t_k) = \sqrt{2} \begin{bmatrix} \cos\theta(t_k) \\ -\sin\theta(t_k) \end{bmatrix} + w(t_k) \qquad k = 0, 1, 2, \cdots$$

where $t_k = kT$; $\lambda(t_0)$ is a random variable with probability density function $N(0, 1)$; $\theta(t_0)$ is uniformly distributed in $[-\pi, \pi)$; $u(\cdot)$ and $w(\cdot)$ are independent zero mean white Gaussian noise processes with

$$E\{u(t)u(\tau)\} = q\delta(t - \tau) \quad \text{and} \quad E\{w(t_k)w'(t_l)\}$$

$$= \begin{bmatrix} \dfrac{r}{T} & 0 \\ 0 & \dfrac{r}{T} \end{bmatrix} \delta_{kl}$$

respectively, α is the 3-dB frequency of the message process in radians/second.

Our demodulator is a bank of extended Kalman filters which operate according to algorithms (3.1) and (4.1). The state estimate is taken to be the conditional mean, calculated as a weighted sum (3.3) of the outputs of the filters.

In our digital computer simulations, we set $c = 1$, $q = 2$ so that $\lim E\{\lambda^2(t)\} = 1$, the root-mean-square bandwidth of the FM baseband spectrum $= 1$ rad, and the bandwidth expansion ratio $\beta = 1/\alpha$. We select $T = 2\pi/16$ s to permit adequately fast sampling of the FM baseband process.

A commonly used steady-state performance display consists of plots of ξ_λ^{-1}, the inverse of the 'evaluated' mean-square message error, versus CNR $= 2/(r\alpha)$, the carrier-to-noise ratio in the message bandwidth. For our curves, ξ_λ is evaluated as the average over J sample paths of the quantity $\Sigma_{k=1}^K [\lambda(t_k) - \hat{\lambda}(t_k \mid t_k)]^2$ where $J = 40, K = 2000$.

A set of such steady-state performance curves for $\alpha = 0.01$ is presented in Figure 2. The curve $M = 1$ corresponds to the performance of the existing quasi-optimum demodulator as presented in [3-8]. In the region of high CNR, the performance of the demodulators are the same for all M. However, in the region of low CNR, the performance of larger M is improved over that of smaller M. In such a region, ξ_λ^{-1} for $M = 2$ is roughly 1 dB better than that for $M = 1$, while for $M = 6$, the improvement is about 3 dB.

Another set of steady-state performance curves consists of plots of ξ_θ^{-1} v.s. CNR, where ξ_θ is the mean-square error of the one step ahead predicted phase estimate, evaluated as the average over J sample paths of the quantity $\Sigma_{k=1}^K [\theta(t_k) - \theta(t_k \mid t_{k-1})]^2$. Figure 3 shows such a set of performance curves for $J = 10$, $K = 600$. Again, it is clear from these curves that improved phase estimates can be obtained using $M > 1$.

Finally, our demodulators can also be used to improve on the existing transient performance of FM demodulation, even in the region of high CNR, as is illustrated in Figure 4.

7. CONCLUDING REMARKS

(1) Certainly we have demonstrated that significant threshold extension for FM and PM demodulators can be
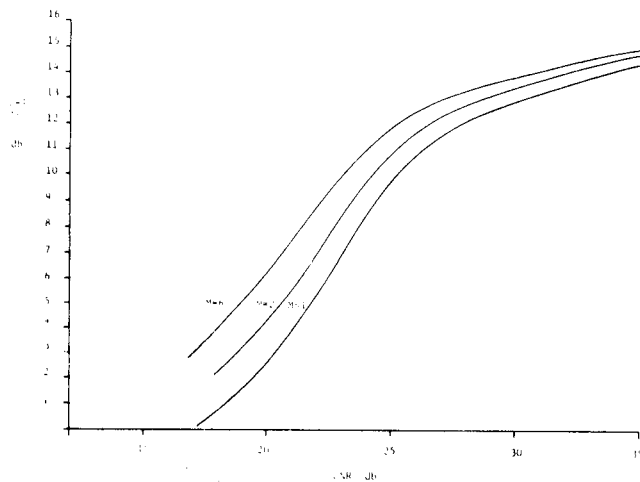
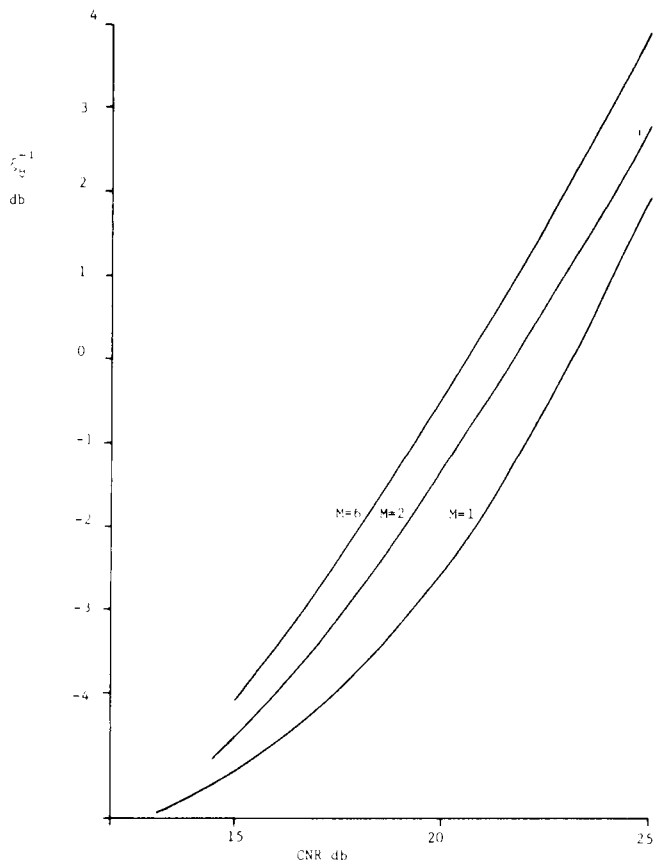Figure 2.   Inverse message error variance for FM.
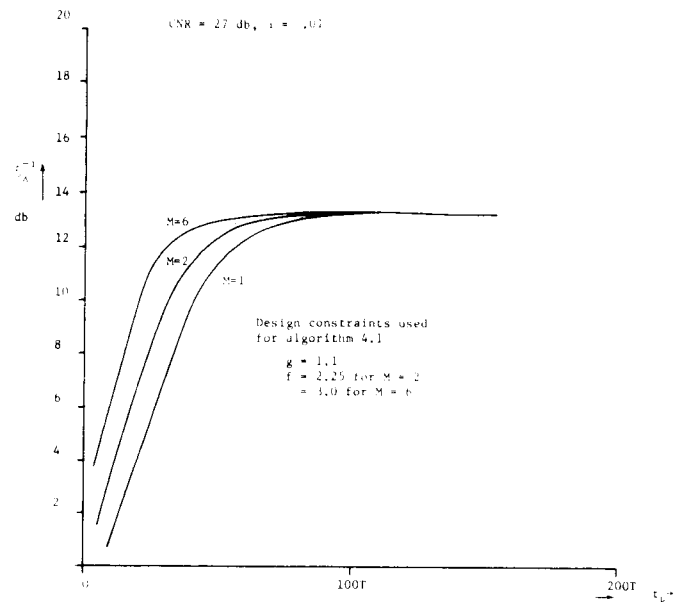

Figure 3.   Inverse angle error variance for FM.

Figure 4.   Transient performance for FM.

demodulation of FM and PM signals—for such demodulators a bank of continuous time filters is required with states periodically reset in the same manner as for the digital filters.

REFERENCES

[1]   F. W. Lehan & R. J. Parks, "Optimum demodulation", *IRE National Convention Record*, Part 8, 1953, pp. 101-103.

[2]   D. C. Youla, "The use of maximum likelihood in estimating continuously modulated intelligence which has been corrupted by noise", *IRE Trans. Inform. Theory*, Vol. IT-3, March 1954, pp. 90-105.

[3]   D. L. Snyder, *The state-variable approach to continuous estimation*, Mass. MIT Press, 1969.

[4]   H. L. Van Trees, *Detection, estimation and modulation theory*, Part II, John Wiley & Sons, 1971.

[5]   C. N. Kelly, "Discrete time demodulation of continuous time signals", Ph.D. Dissertation, Southern Methodist Univ., Dallas, Tex., May 1971.

[6]   C. N. Kelly & S. C. Gupta, "The digital phase-locked loop as a near optimum FM demodulator," *IEEE Trans. Comm.*, Vol. COM-20, June 1972, pp. 406-411.

[7]   D. R. Polk & S. C. Gupta, "Quasi optimum digital phase-locked loops", *IEEE Trans. Comm.*, Vol. COM-21, January 1973, pp. 75-82.

[8]   A. L. McBride, "On optimum samples-data FM demodulation", *IEEE Trans. Comm.*, Vol. COM-21, No. 1, January 1973, pp. 40-50.

[9]   R. S. Bucy & A. J. Mallinckrodt, "An optimal phase demodulator", *Stochastics*, Vol. 1, 1973, pp. 3-23.

[10]   J. T. Lo & A. S. Willsky, "Estimation for rotational processes with one degree of freedom I: Introduction and continuous time processes", *IEEE Trans. Autom. Contr.*, Vol. AC-20, No. 1, February 1975, pp. 10-21.

[11]   A. S. Willsky & J. T. Lo, "Estimation for rotational processes with one degree of freedom II: Discrete time processes", *IEEE Trans. Autom. Contr.*, Vol. AC-20, No. 1, February, 1975, pp. 22-30.

[12]   A. S. Willsky & J. T. Lo, "Estimation for rotational processes with one degree of freedom III: Applications and implementation", *IEEE Trans. Autom. Contr.*, Vol. AC-20, No. 1, February, 1975, pp. 31-33.

[13]   J. T. Lo & A. S. Willsky, "Stochastic control of rotational processes with one degree of freedom", *SIAM J. Contr.*, Vol. 13 No. 4, July 1975, pp. 886-898.

[14]   A. S. Willsky, "Fourier series and estimation on the circle with applications to synchronous communication I: Analysis", *IEEE Trans. Inform. Theory*, Vol. IT-20, No. 5, September 1974, pp. 577-583.

achieved using suitable adaptations of the Gaussian sum nonlinear estimators. The cost associated with increase in filter complexity is however quite considerable and would certainly make one think twice before upgrading the simple phase-locked loop.

(2) It is of theoretical significance that *optimal* demodulation in high noise can be achieved in the limit as the number of filters in our filter bank becomes infinite.

(3) The ideas of the present paper constitute an extension of the Gaussian sum nonlinear filtering theory and thus extend in their application to other nonlinear filtering problems. For example, the ideas of this paper have application to the analog

[15] A. S. Willsky, "Fourier series and estimation on the circle with applications to synchronous communication II: Implementation", *IEEE Trans. Inform Theory*, Vol. IT-20, No. 5, September 1974, pp. 584-590.

[16] D. E. Gustafson & J. L. Speyer, "Linear and asymptotic minimum variance filters applied to phase-lock loops", to appear in *IEEE Trans. Comm.*

[17] Y. C. Ho & R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control", *IEEE Trans. Autom. Contr.*, Vol. AC-9, 1964, pp. 333-339.

[18] A. H. Jazwinski, *Stochastic processes and filtering theory*, Academic Press, New York and London, 1970.

[19] H. W. Sorenson & D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums", *Automatica*, Vol. 7, 1971, pp. 465-479.

[20] R. S. Bucy & K. D. Senne, "Digital synthesis of non-linear filters", *Automatica*, Vol. 7, 1971, pp. 287-298.

[21] D. G. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaptation', *IEEE Trans. Autom. Contr.*, Vol. AC-16, No. 2, April 1971, pp. 160-170.

[22] D. G. Lainiotis, "Optimal non-linear estimation", *Int. J. Control*, Vol. 14, No. 6, 1971, pp. 1137-1148.

[23] D. L. Alspach & H. W. Sorenson, "Nonlinear Bayesian estimation using Guassian sum approximations", *IEEE Trans. Autom. Contr.*, Vol. AC-17, No. 4, August 1972, pp. 439-448.

[24] J. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering", *IEEE Trans. Inform. Theory*, Vol. IT-18, No. 5, September 1972, pp. 583-588.

[25] J. M. Wozencraft & I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965.

[26] P. K. S. Tam, "Parallel processing and smoothing techniques in filtering applications", Ph.D. Thesis, University of Newcastle, January, 1975.

[27] P. K. S. Tam and J. B. Moore, "Improved demodulation of sampled FM signals in high noise", to appear.

[28] A. J. Mallinckrodt, R. S. Bucy and S. Y. Cheng, "Final report for a design study for an optimal nonlinear receiver—demodulation", NASA Contract NAS 5-10789, Goddard Space Flight Center, Maryland, 1970.

[29] R. S. Bucy, C. Hecht and K. D. Senne, "An application of Bayes-law estimation to nonlinear phase demodulation", *Proc. 3rd Symposium on non-linear estimation and its applications*, San Diego, September, 1972.

Peter Kwong-Shun Tam was born in Hong Kong on February 12, 1944. In 1967 he obtained the Higher Diploma in Electrical Engineering from Hong Kong Technical College. From 1967 to 1971, he was employed as a staff member in the electrical department of the B.H.P. Newcastle Steelworks, Australia.

In 1971, he graduated B.E. with 1st class honors from the University of Newcastle, receiving a prize from the Institution of Engineers, Australia. From the same university, he obtained the M.E. degree in 1972 and the Ph.D. in 1975.

In 1975 he lectured at the Footscray Institute of Technology. In 1976 he worked as a Forecasting Research Officer at the Gas and Fuel Corporation of Victoria. He is currently a lecturer in Electrical Engineering at the Gippsland Institute of Advanced Education.

His interests include the extension and application of Kalman filtering techniques.

★

John B. Moore (S'66-M'67) was born in China in 1941. He received his bachelor and masters degrees in Electrical Engineering in 1963 and 1964 respectively, and his doctorate in Electrical Engineering by the University of Santa Clara, California in 1967. He was appointed Senior Lecturer at the University of Newcastle in 1967, promoted to Associate Professor in 1968 and Full Professor in 1973, and has held other visiting academic appointments at the University of Santa Clara in 1968 and at the University of Maryland 1970, Colorado State University and Imperial College 1974. He has spent periods in industry as a design engineer and as a consultant.

Dr. Moore's current research in control and communication systems is supported by the Australian Research Grants Committee and the Radio Research Board. He is co-author with Brian D. Anderson of a book *Linear Optimal Control* (Prentice-Hall 1971).