NONLINEAR PHASE ESTIMATORS BASED ON THE KULLBACK DISTANCE

José M. N. Leitão[1] and José M. F. Moura[2]

[1] Centro de Análise e Processamento de Sinais, and
Departamento de Engenharia Electrotécnica e de Computadores.
Instituto Superior Técnico, Lisboa, PORTUGAL

[2] LASIP, Department of Electrical and Computer Engineering,
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

ABSTRACT

This paper considers the design of phase estimators by combining concepts of stochastic nonlinear filtering and information theory. To propagate the involved probability density functions, adequate finite representations are needed. This is accomplished in this work by adopting minimum Kullback distance criteria. Applied to the important and paradigmatic cyclic phase estimation problem, our approach leads to consistent and systematic design methods. The resulting simple and parallelizable structure outperforms the commonly used extended Kalman-Bucy filter in tracking and acquisition situations. These features make the developed nonlinear filter suited to digital communications (carrier synchronization).

I. INTRODUCTION

Phase estimation problems have been extensively studied within the stochastic nonlinear filtering framework [1]. In [2] an algorithm has been developed that explores the prominent features of the problem and forms the kernel of different specialized estimators. Besides cyclic estimation, absolute phase acquisition and tracking have been addressed. In absolute phase tracking, the nonlinear filter clearly outperforms the *phase locked loop* (PLL) which is the steady state *extended Kalman-Bucy filter* (EKBF). In tracking Brownian motion, an improvement of about 50% in the mean time between cycle slips has been reported in [3]. Reference [4] studies the application of that algorithm to real signals propagating under the Arctic ice crust, showing a much better performance than the standard procedures, with additional flexibility. Absolute phase acquisition, an inherently global nonlinear filtering problem for which no local estimator such as the EKBF is adequate, is the main subject of paper [5], which deals with ranging in radar/sonar systems. Cyclic (modulo 2π) phase estimation is the classical test bed for optimal nonlinear filtering. Studies have been published showing the superior performance of nonlinear filtering over the PLL [1], [6].

To propagate the involved probability density functions, appropriate finite representations are needed. The algorithm developed in [2] used *ad hoc* fitting and matching techniques. This paper reconsiders the design of suboptimal nonlinear phase estimators by adopting a minimum Kullback distance criterion [7]. The paradigmatic and practically important cyclic phase estimation problem is mainly addressed. The resulting simple scheme achieves in steady state (cyclic tracking), and under strong noise condition, almost the maximal achievable performance gain over the PLL. For high signal to noise ratios, both filters exhibit the same cyclic tracking variances; they behave, however, differently in acquiring the initial phase value. In fact, the nonlinear filter pays more attention to the observations than the PLL does, being able to converge to steady state in a much smaller number of iterations. Also, and contrarily to the PLL, the proposed estimator is *hang up* free [8].

II. PROBLEM FORMULATION

A. Model

Consider a phase modulated carrier in an additive white Gaussian noise channel. The receiver signal is $z(t) = \cos(\omega_0 t + x_1(t)) + v(t)$, where ω_0 is the nominal carrier frequency, $x_1(t)$ is the modulating signal and $v(t)$ is a white Gaussian noise. The carrier (known) amplitude is normalized to one. Further assume that the received signal is down converted to baseband with a local generator of nominal frequency ω_0.

The sampled (integration and dump) in-phase and quadrature components of $z(t)$, $\{\mathbf{z}_n = [z_{1,n}\,,\, z_{2,n}]^T\}$, are given by

$$\begin{bmatrix} z_{1,n} \\ z_{2,n} \end{bmatrix} = \begin{bmatrix} \cos(x_{1,n}) \\ \sin(x_{1,n}) \end{bmatrix} + \begin{bmatrix} v_{1,n} \\ v_{2,n} \end{bmatrix} \qquad (1)$$

where $\{v_{1,n}\}$ and $\{v_{2,n}\}$ are mutually independent zero mean white Gaussian sequences with variance r, and $\{x_{1,n}\}$ is the first component of the discrete vector Markov process $\{\mathbf{x}_n \in \mathbb{R}^K\}$, described by the stochastic difference equation

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{B}\mathbf{u}_n \qquad n = 1, 2, \ldots \qquad (2)$$

where $\{\mathbf{u}_n \in \mathbb{R}^m\}$ is a vector zero mean white Gaussian sequence with covariance matrix $\mathbf{Q}$.

B. Optimal Bayesian Solution

Taking into account the preceding models and assumptions, consider the problem of estimating $\mathbf{x}_n$, based on the

0-7803-1775-0/94 $3.00 © 1994 IEEE

set of present and past observations $\mathbf{Z}_n = \{\mathbf{z}_k,\ 1 \le k \le n\}$. The main task of a (global) nonlinear filter is to propagate the conditional probability density function $F_n = p(\mathbf{x}_n|\mathbf{Z}_n)$ herein referred to as the *filtering density*. The solution consists in the recursive application of Bayes law (filtering step) and Chapman-Kolmogorov equation (prediction step) [1]:

$$P_n = S_n * F_{n-1} \quad \text{(prediction)}, \tag{3}$$

$$F_n = C_n H_n \bullet P_n \quad \text{(filtering)}, \tag{4}$$

where $*$ denotes convolution, $\bullet$ means pointwise multiplication, and C_n is a normalizing constant. The *convolution kernel*, $S_n = p(\mathbf{x}_{n+1}|\mathbf{x}_n)$, reflects equation (2) and the assumptions therein; it is Gaussian and given by

$$S_n \propto \mathcal{N}(\mathbf{x}_{n+1} - \mathbf{A}\mathbf{x}_n, \mathbf{BQB}^T), \tag{5}$$

with the notation $\mathcal{N}(\mathbf{s}, \mathbf{V}) = \exp\{-(1/2)\ \mathbf{s}^T\mathbf{V}^{-1}\mathbf{s}\}$. This kernel actuates on the preceding filter density F_{n-1} to give the *prediction density* $P_n = (\mathbf{x}_n|\mathbf{Z}_{n-1})$, which is updated by the multiplicative effect of the *observation factor* $H_n = p(\mathbf{z}_n|\mathbf{x}_n)$. According to model (1), H_n is given by

$$H_n \propto e^{\lambda_n \cos(x_{1,n} - b_0^{H_n})}, \tag{6}$$

with

$$\lambda_n = \frac{1}{r}\sqrt{z_{1,n}^2 + z_{2,n}^2}, \qquad \text{and} \qquad b_0^{H_n} = \arctan\frac{z_{2,n}}{z_{1,n}}.$$

The optimal estimate $\hat{\mathbf{x}}_n$ is obtained by minimizing the conditional expectation of a suitable cost function $L(\mathbf{x}_n - \hat{\mathbf{x}}_n)$.

III. Proposed Approach

To implement equations (3) and (4), adequate finite representations of operands are required. The nonlinear filter (NLF), whose structure is schematized in Fig. 2, incorporates the representation procedures developed along this section.

A. Sensor factor representation

Each normalized period of H_n is a Tikhonov function. It approximates a Gaussian function for large values of λ_n, becoming flat as λ_n tends to zero. These features suggest the substitution of H_n by a train of Gaussian functions, centered on $b_i^{H_n} = b_0^{H_n} + 2\pi i$, all having a common variance σ^{H_n},

$$\tilde{H}_n \propto \sum_{i=-\infty}^{+\infty} \mathcal{N}\left(x_{1,n} - b_i^{H_n}, \sigma^{H_n}\right). \tag{7}$$

This representation should reproduce, as much as possible, the shape of H_n for all values of λ_n.

Take, for simplicity, $b_0^{H_n} = 0$; the normalized central periods of H_n and $\tilde{H}_n$ are, respectively,

$$h(x,\lambda) = \frac{e^{\lambda\cos(x)}}{2\pi I_0(\lambda)} \tag{8}$$

$$\tilde{h}(x,\sigma) = \frac{\sum_{i=-\infty}^{+\infty} e^{-(x-2\pi i)^2/(2\sigma)}}{(2\pi\sigma)^{\frac{1}{2}}}, \tag{9}$$

where $I_0(\cdot)$ is the first kind modified zero order Bessel function.

The Kullback distance, or divergence [9],

$$D(h\|\tilde{h}) = \int_{-\pi}^{\pi} h(x,\lambda)\log\frac{h(x,\lambda)}{\tilde{h}(x,\sigma)}\,dx \tag{10}$$

is a measure of dissimilarity between the probability density functions $h(x,\lambda)$ and $\tilde{h}(x,\sigma)$. The adopted representation criterion consists in finding, for a given λ, the value of σ that minimizes $D(h\|\tilde{h})$. This implies the equation

$$\frac{\partial D}{\partial\sigma} = -\int_{-\pi}^{\pi} h(x,\lambda)\,\frac{\partial\tilde{h}(x,\sigma)/\partial\sigma}{\tilde{h}(x,\sigma)}\,dx = 0. \tag{11}$$

Inserting (8) and (9) into (11) yields

$$\sigma = \int_{-\pi}^{\pi}\frac{e^{\lambda\cos(x)}}{2\pi I_0(\lambda)}\,\Phi(x,\sigma)\,dx, \tag{12}$$

with

$$\Phi(x,\sigma) = \frac{\sum_{-\infty}^{+\infty}(x-2\pi i)^2\, e^{-\frac{(x-2\pi i)^2}{2\sigma}}}{\sum_{-\infty}^{+\infty} e^{-\frac{(x-2\pi i)^2}{2\sigma}}}. \tag{13}$$

Fig. 1(a) shows a map of σ versus λ, obtained by numerically solving equation (13). A lookup table containing this map provides the optimal value σ^{H_n} corresponding to any value of λ_n, with a minor computational effort. Fig. 1(b) plots the minimal Kullback distance as function of λ. The quality of the representation can be assessed from Fig. 1(c); notice the apparent coincidence of h and $\tilde{h}$, when the Kullback distance is practically zero (λ_1 and λ_3), and the similarity between h and $\tilde{h}$ in the worst case situation (for λ_2).

Finally, it should be stressed that only two parameters are needed to represent H_n: $b_0^{H_n}$ and σ^{H_n}, both easy to compute from the data.

B. Filtering

B.1 Multiplication

To implement equation (4), H_n is substituted by its representation (7). Consider a Gaussian prediction density

$$P_n \propto \mathcal{N}\left(\mathbf{x}_n - \mathbf{b}^{P_n}, \mathbf{V}^{P_n}\right),$$

and assume that only the J nearest modes of $\tilde{H}_n$ contribute significantly to the product $P_n \bullet \tilde{H}_n$. Typically $J = 3$, or even $J = 1$ for small values of r. The result is

$$F_n = \sum_{i=1}^{J} k_i^{F_n}\,\mathcal{N}\left(\mathbf{x}_n - \mathbf{b}_i^{F_n}, \mathbf{V}^{F_n}\right), \tag{14}$$

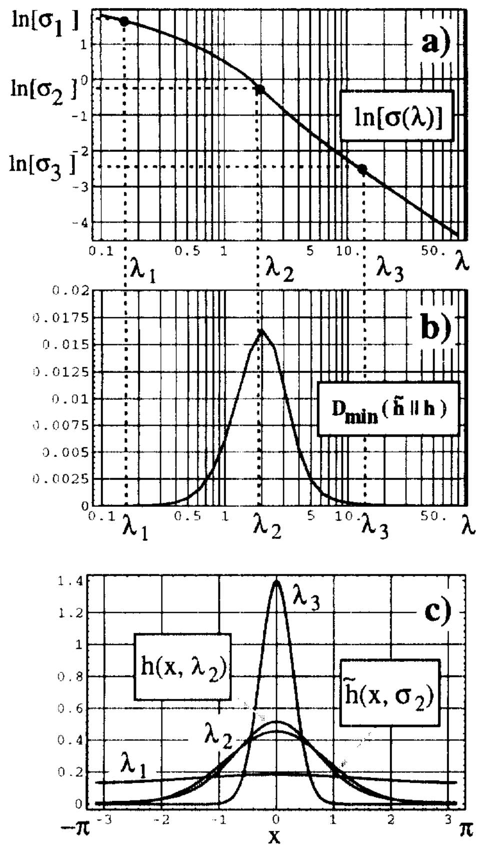

Figure 1: Graphical information concerning the sensor factor representation. See subsection III.A

with

$$\mathrm{b}_i^{F_n} = \mathrm{b}^{P_n} - \frac{b_1^{P_n} - b_i^{H_n}}{V_{11}^{P_n} + \sigma^{H_n}} \begin{bmatrix} V_{11}^{P_n} \\ \vdots \\ V_{1k}^{P_n} \end{bmatrix} \tag{15}$$

$$\mathbf{V}^{F_n} = \mathbf{V}^{P_n} - \frac{1}{V_{11}^{P_n} + \sigma^{H_n}} \begin{bmatrix} V_{11}^{P_n} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ V_{1k}^{P_n} & 0 & \cdots & 0 \end{bmatrix} \mathbf{V}^{P_n} \tag{16}$$

$$k_i^{F_n} \propto \mathcal{N}\left(b_1^{P_n} - b_i^{H_n}, V_{11}^{P_n} + \sigma^{H}\right) \tag{17}$$

where $k_i^{F_n}$ are normalizing weight factors.

Equations (15) and (16) are discrete Kalman-Bucy filtering (updating) steps working with observations given by $b_i^{H_n} = x_{1,n} + v_n$ where $\{v_n\}$ is a white Gaussian sequence of zero mean and adaptive variance σ^{H_n}. The means $b_i^{H_n}$ of each mode of $\tilde{H}_n$ play the role of pseudo-measurements.

B.2 Gaussian matching

We consider now cyclic phase estimation problems (e.g. carrier synchronization in digital communications) in which only modulo 2π phase values are sought ($\hat{x}_{1,n} \in [-\pi, +\pi]$); as for the other components of $\mathbf{x}_n$, evolutions along the real line ($\hat{x}_{l,n} \in \mathbb{R}$, $l = 2, 3, \ldots, K$) are meaningful. Accordingly, the adopted global cost function is

$$L(\mathbf{x}_n - \hat{\mathbf{x}}_n) = \sum_{l=1}^{K} L_l(x_{l,n} - \hat{x}_{l,n}), \tag{18}$$

with $L_1(x_{1,n} - \hat{x}_{1,n}) = 2[1 - \cos(x_{1,n} - \hat{x}_{1,n})]$ (the cyclic cost function as in [6]), and $L_l(x_{l,n} - \hat{x}_{l,n}) = (x_{l,n} - \hat{x}_{l,n})^2$ for $l = 2, \ldots, K$.

Let $\hat{\mathbf{x}}_n$ be the estimate stemming from cost function (18) and density (14). To get a regular, constant (low) dimension structure, the following strategy is adopted: replace, at each iteration, the density F_n by the single Gaussian

$$\tilde{F}_n = (2\pi)^{-K/2} |\mathbf{V}^{\tilde{F}_n}|^{-1/2} \mathcal{N}\left(\mathbf{x}_n - \hat{\mathbf{x}}_n, \mathbf{V}^{\tilde{F}_n}\right), \tag{19}$$

with covariance matrix $V^{\tilde{F}_n}$ obtained by minimizing the Kullback divergence $D(F_n \| \tilde{F}_n)$. Density $\tilde{F}_n$ still leads to the same estimate $\hat{\mathbf{x}}_n$, while preserving the information content of F_n in a Kullback sense.

C. Prediction

Given the form of expressions (5) and (19), prediction (3) is a standard Kalman-Bucy prediction step. This closes the loop in the block diagram of Fig. 2.

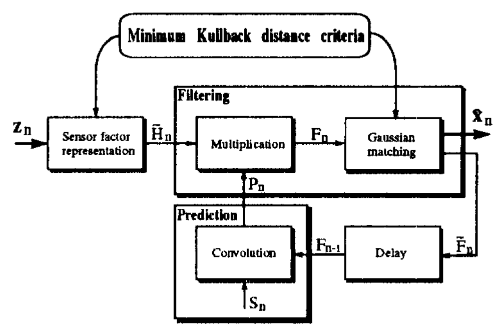

Figure 2: Schematic structure of the NLF.

D. Simulation results

The results herein reported concern the estimation of Brownian motion. The noise condition is expressed in terms of the steady state error variance that would be achieved if the observations were linear. This variance, provided by the corresponding Ricatti equation, is denoted as V_∞. The NLF exhibits, in steady state and under strong noise condition, essentially the same performance as the point mass filter studied in [6], which is known to reach the maximal achievable gain over the PLL. As the signal to noise ratio increases, both estimators tend to the same asymptotic steady state error variance predicted by the linear theory. Transient behavior (acquisition) is however very different, as can be concluded from the following experimental data.

For a given value of V_∞, 10^4 runs (of length $N = 150$) are performed and the mean squared errors (modulo 2π) $\overline{\epsilon_n^2}$ are computed at each iteration. For comparative purposes the PLL, the EKBF, and the NLF are driven by the same noise sequences.

Fig. 3(a) shows, for $V_\infty = -20\,dB$, the evolutions of $\overline{\epsilon_n^2}$ for the 3 estimators under study. Two different initial conditions are considered:

1. Initial phase value uniformely distributed on the interval $[-\pi\,,\,+\pi]$ (curves 1, 2 and 3). For the NLF and the EKBF, this condition is expressed by assuming that P_1 is Gaussian with zero mean and infinite variance. The PLL, being the steady state EKBF, starts with P_1 being zero mean Gaussian with variance V_∞.

2. Initial phase value generated according to $P_1 = p(x_1) = \frac{1}{2}\delta(x_1 - \pi) + \frac{1}{2}\delta(x_1 + \pi)$ (curves 4, 5, and 6). This distribution simulates the conditions where the PLL typically suffers from the *hang up* phenomenon (phase errors close to $\pm\pi$) [8]. All filters are initialized as above.

There are remarkable differences concerning the convergence times. As expected, the EKBF exhibits better performance than the PLL in both conditions. The NLF, besides showing much better performance than the EKBF and the PLL, also reveals independence from the initial phase errors (notice the coincidence of curves 1 and 4); as a corollary, the NLF is *hang up* free.

Fig. 3(b) is equivalent to Fig. 3(a), now for $V_\infty = -10\,dB$. Notice the significantly faster convergence of the EKBF and the PLLL, when compared with the $V_\infty = -20\,dB$ situation; as for the NLF, its convergence pattern remains practically unchanged.

IV. Concluding Remarks

In this paper, we considered the design of nonlinear phase estimators according to minimum Kullback distance criteria. This approach, applied to cyclic phase estimation, leads to a simple and parallelizable (open-loop) structure suitable for parallel architectures and VLSI implementation. Furthermore, it is a very fast acquisition and *hang up* free algorithm. A bank of such filters has been proposed in [10] to perform simultaneous phase estimation and symbol detection in digital communications.

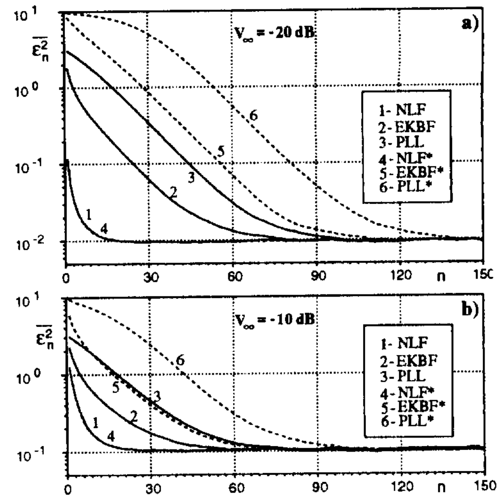

Figure 3: Mean squared error evolution for the 3 compared filters. See section IV for details.

References

[1] R. Bucy, C. Hecht, and K. D. Senne, "An engineer's guide to building nonlinear filters", Technical Report, Frank J. Seiler Research Laboratory, 1989.

[2] J. M. N. Leitão, *Nonlinear Optimal Phase Estimation. Acquisition and Tracking*, PhD thesis, Instituto Superior Técnico – Technical University of Lisbon, Lisbon, Portugal, 1983, (In Portuguese).

[3] J. M. N. Leitão and J. M. F. Moura, "Reducing the error diffusion rate in absolute phase estimation", in H. W. Schüssler, editor, *Signal Processing II. Theory and Applications*, pp. 507–510. EURASIP, Elsevier Science Publishers, 1983.

[4] J. M. F. Moura and A. B. Baggeroer, "Phase unwarping of signals propagated under the artic ice crust: A statistical approach", *I.E.E.E. Trans. on Acoust. Speech and Signal Proc.*, vol. 36, pp. 617–630, 1988.

[5] J. M. N. Leitão and J. M. F. Moura, "Acquisition is phase demodulation: Application to ranging in radar/sonar systems", Submited to *I.E.E.E. Transactions on Aerospace and Electronic Systems*, 1992.

[6] R.S.Bucy and A.J.Mallinckrodt, "An optimal phase demodulator", *Stochastics*, vol. 1, pp. 3–23, 1973.

[7] S. Kullback, *Information Theory and Statistics*, Peter Smith, 1978.

[8] H.Meyr and G. Ascheid, *Synchronization in Digital Communications*, John Wiley & Sons, New York, 1990.

[9] T.M.Cover and J.A.Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[10] J. M. N. Leitão and F. D. Nunes, "Nonlinear phase estimation and detection in digital communications", in J. Moura and I. Lourtie, editors, *Proceedings of the NATO-ASI on Acoustic Signal Processing for Ocean Exploration*. Kluwer Academic Publishers, 1992.