A Concept of Approximated Densities for Efficient Nonlinear Estimation

Virginie F. Ruiz

Department of Cybernetics, The University of Reading, Whiteknights, Reading RG6 6AY, UK
Email: v.f.ruiz@reading.ac.uk

This paper presents the theoretical development of a nonlinear adaptive filter based on a concept of filtering by approximated densities (FAD). The most common procedures for nonlinear estimation apply the extended Kalman filter. As opposed to conventional techniques, the proposed recursive algorithm does not require any linearisation. The prediction uses a maximum entropy principle subject to constraints. Thus, the densities created are of an exponential type and depend on a finite number of parameters. The filtering yields recursive equations involving these parameters. The update applies the Bayes theorem. Through simulation on a generic exponential model, the proposed nonlinear filter is implemented and the results prove to be superior to that of the extended Kalman filter and a class of nonlinear filters based on partitioning algorithms.

Keywords and phrases: nonlinear estimation, nonlinear adaptive filter, exponential distribution, maximum entropy, non-Gaussian signal processing.

1. INTRODUCTION

This paper describes a recursive algorithm based on a nonlinear approach to parameter estimation. The most common procedure applied for nonlinear estimation is the extended Kalman filter (EKF) [1, 2, 3].

It is known that the nonlinear estimator for nonlinear models does not provide a finite solution. By linearising the state equations about the conditional means $x_{k|k-1}$ and $x_{k|k}$, the EKF provides a solution in terms of minimal mean square error to the original nonlinear problem. A number of other procedures have also been applied including partitioning approaches, statistical linearisation, maximum a posteriori, least square criteria, and functional approximation of conditional state density [3, 4, 5, 6, 7].

Crucially, finite models or filters that completely define physical systems are rare [8, 9]. Nevertheless, the state distributions depend on a finite number of parameters, which are recursively computed as the observed data arrive [10, 11, 12]. The computation of a finite number of *pertinent* parameters is a common procedure for the definition of *approximated* probability distributions. As nonlinear state space models rarely yield explicit or analytic distributions from output measurements, the distribution normally has to be approximated [12].

This paper presents the theoretical development of a simple method of approximation [13]. In the following, the proposed nonlinear adaptive filter may be referred to as the FAD filter. The approach uses a maximum entropy principle to approximate the filtering equations arising from a state model with nonlinear equations. The probability density functions (pdf) created by applying the entropy principle are of an exponential type and depend on a finite number of parameters. The nonlinear filtering leads to recursive equations involving these parameters. The use of the maximum entropy approximation closes the nonlinear equations of the filter; that is the exponential-type family of distributions under consideration is stable for nonlinear equations.

At each time instant k, the FAD filter estimates the consecutive a priori and a posteriori state probability density functions given past and current observations. The logarithms of the pdfs are linear combinations of several functions, $\varphi_1, \ldots, \varphi_n$, chosen according to some specific criterion. The prediction uses maximum entropy subject to constraints related to the functions φ. The update applies the Bayes theorem.

2. MAXIMUM ENTROPY PRINCIPLE

Let μ be a probability over the measurable space $(R^n, \mathfrak{R}^n)$ and P a probability law absolutely continuous with respect to μ and of Radon-Nikodym density

$$\frac{dP}{d\mu} = p(x), \quad \text{almost surely.} \tag{1}$$

Proposition 1. *Let $\varphi_0 = 1$, $\varphi_1, \ldots, \varphi_n$ be $(n+1)$ real-valued functions in the Hilbert space. Then, the maximum entropy*

$$H(x) = -\int \ln p(x)\, p(x)\, d\mu(x) \tag{2}$$

subject to the real-valued constraints, for all $i = 1, \ldots, n$,

$$l_i = \int \varphi_i(x)\, p(x)\, d\mu(x) \tag{3}$$

gives rise to an exponential-type density function,

$$p(x) = \exp\left[\lambda_0 + \sum_{i=1}^{n} \lambda_i \varphi_i(x)\right], \tag{4}$$

where the real coefficients $\lambda_0, \ldots, \lambda_n$ are Lagrange multipliers that are evaluated from the constraints.

Proof. We use the method of the Lagrange multipliers and introduce $\lambda_1, \ldots, \lambda_n$ adjustable constants called the Lagrange multipliers in order to maximise the expression

$$
\begin{aligned}
K &= H(x) + \sum_{i=1}^{n} \lambda_i l_i \\
&= \int -p(x) \ln p(x)\, d\mu(x) + \sum_{i=1}^{n} \int \lambda_i \varphi_i(x)\, p(x)\, d\mu(x) \\
&= \int \left[-\ln p(x) + \lambda_1 \varphi_1(x) + \cdots + \lambda_n \varphi_n(x) \right] p(x)\, d\mu(x).
\end{aligned}
\tag{5}
$$

Using the calculus of variations, the solution for $p(x)$ is given by

$$\frac{\partial}{\partial p(x)}\left[-\ln p(x) + \lambda_1 \varphi_1(x) + \cdots + \lambda_n \varphi_n(x) \right] p(x) = 0. \tag{6}$$

Thus,

$$-\ln p(x) + \lambda_1 \varphi_1(x) + \cdots + \lambda_n \varphi_n(x) - 1 = 0. \tag{7}$$

Observing that $p(x)$ is a probability density function and denoting the constant-valued function equal to 1 by φ_0, a new constraint appears

$$l_0 = \int \varphi_0(x)\, p(x)\, d\mu(x) = 1. \tag{8}$$

Therefore, we can substitute $\lambda_0 \varphi_0(x)$ for -1 in (7) and solving for $p(x)$ gives an exponential type density function

$$p(x) = \exp\left[\sum_{i=0}^{n} \lambda_i \varphi_i(x)\right]. \tag{9}$$

$\square$

Due to the maximum entropy principle, the method inherently approximates all probability distributions with exponential type density functions. The Lagrange multipliers λ_i vary with the family of densities under consideration and are computed at each update and prediction.

3. NONLINEAR MODEL

We choose to pursue the derivation in the scalar case and discrete time, $k \in \mathbb{N}$, for simplicity purposes. Nonlinear state models are often given by equations of the form

$$
\begin{aligned}
z_{k+1} &= g(z_k) + w_k, \\
y_k &= h(z_k) + v_k,
\end{aligned}
\tag{10}
$$

where $\{z_k\}$ and $\{y_k\}$ are the state and observation processes, g and h are some nonlinear functions. The noise sequences $\{w_k\}$ and $\{v_k\}$ are assumed to be white having any distribution.

Let x_k be a state defined by

$$x_k = h[z_k] \tag{11}$$

with h locally injective, we can write at time instant $k+1$

$$
\begin{aligned}
x_{k+1} &= f(x_k, w_k), \tag{12} \\
y_k &= x_k + v_k, \tag{13}
\end{aligned}
$$

where the nonlinear function f is defined by

$$f(x_k, w_k) = h\{g[h^{-1}(x_k)] + w_k\}. \tag{14}$$

Thus, the filter uses the nonlinear model defined by (12) and (13) where x_k and y_k are the state and observation, respectively. The noise processes are mutually independent, and independent of the state, for all $k \in N$.

4. NONLINEAR FILTER

The FAD filter approximates, at each time instant k, the conditional pdfs of the state x_k given the past or present observations through expression (4). We start with the prediction that uses the state equation (12) to predict the conditional density $p_{k+1|k}(x|\mathbf{y})$ of the state x_{k+1} given the observation up to time instant k, $\mathbf{y} = (y_1, \ldots, y_k)$.

4.1. Prediction

The prediction introduces several constraints l_i, $i \in N$,

$$l_i = E[\varphi_i(x_{k+1})] = \int_R \varphi_i(x)\, p_{k+1|k}(x|\mathbf{y})\, dx \tag{15}$$

upon the state x_{k+1} given $\mathbf{y} = (y_1, \ldots, y_k)$.

The functions φ_i define the information retained to describe the state distribution at time instant $k+1$. For example, a function indicator over an interval defines a constraint that is the probability of observing the state x_{k+1} in that interval; a function power of x_{k+1} defines a constraint that is a moment of the state variable.

The state equation (12) is used to evaluate the constraints. It precisely defines the transfer of random variables from time instant k to instant $k+1$. Thus,

$$l_i = \iint \varphi_i[f(x_k, w_k)]\, p_{k|k}(x|\mathbf{y})\, dx\, p_w(w)\, dw, \tag{16}$$

where p_w denotes the exponential pdf of the state noise expressed as in (4).

The determination of the constraints relies on the calculation of multiple integrals. These integrals can easily be computed by numerical integration based on polynomial interpolation. However, in most applications analytical expressions can be derived.

Note from the state equation that the dynamic behaviour is not retained by this mean in its entirety. A study of the state equation without noise reveals the properties of the asymptotic or limit distribution as k approaches infinity. In most cases, the density has several maxima (even infinite), that is, it is multimodal. In practice, we pay particular attention to the number of local maxima in order to elaborate the a priori exponential-type density from the functions φ_i that must be chosen accordingly. Some practical examples of bi-, tri-, and five-modal density functions may be found in areas of application such as motion compensated video compression and medical image analysis [13, 14, 15].

The density predicted is the density that maximises the entropy subject to the constraints resulting from (16). The density created is of an exponential type as in (4). The Lagrange multipliers in the density are determined from the constraints by solving the system of $(n + 1)$ nonlinear equations

$$l_i = E[\varphi_i(x_{k+1})]. \tag{17}$$

Depending on the nature of the distributions, analytical expressions may exist. The case of Gaussian approximation is derived in Section 5. The Lagrange multipliers vary with the family of densities under consideration and are computed at each prediction and update.

4.2. Update

At each time instant k, the update uses the observation equation (13) to estimate the a posteriori density $p_{k|k}(x|\mathbf{y})$ of the state x_k given the past observations up to y_{k-1} and the actual observation $y_k = \eta$, $\mathbf{y} = ((y_1, \ldots, y_{k-1}), y_k)$.

By hypothesis, the density of the state x_k and the noise w_k given the past observations, are of an exponential type as in (4). Applying the Bayes rule to the observation equation (13) gives the conditional density of the state, which remains as an exponential-type without modification of the functions φ_i as, in this particular case, the observation equation is linear.

The conditional density $p_{k|k}(x|\mathbf{y})$, given the observations up to $y_k = \eta$, is defined by

$$p_{k|k}(x|\mathbf{y}) = C p_{k|k-1}(x|\mathbf{y}) p_v(\eta - x), \tag{18}$$

where C is the normalisation constant

$$C^{-1} = \iint p_{k|k-1}(\alpha, \beta) p_v(\eta - \alpha) \, d\alpha \, d\beta. \tag{19}$$

The densities in (18) being as in (4), equating the coefficient of the like terms in φ_i provides a very simple update of the Lagrange multipliers in the state density. The complete derivation is shown in Section 5 using Gaussian approximation.

5. IMPLEMENTATION

In its simplest form and for comparison purposes with techniques such as extended Kalman, the approach can be applied to the particular case of Gaussian approximation. In this case, the Lagrange multipliers in the pdf are easily obtained from the constraint. The Gaussian case is presented as a didactic example.

Gaussian approximation

A Gaussian distribution has a density function

$$
\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x - m}{\sigma}\right)^2\right) \\
&= \exp\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{m^2}{2\sigma^2} + \frac{m}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right) \\
&= \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2),
\end{aligned} \tag{20}
$$

where m and σ denote the mean and standard deviation, respectively.

It can be seen from the above expression that the Gaussian density is of an exponential type. The logarithm of the pdf is linearly developed from the basis functions

$$\varphi_0(x) = 1, \qquad \varphi_1(x) = x, \qquad \varphi_2(x) = x^2. \tag{21}$$

In that particular case, the constraints, l_i, correspond to the moments up to second order of the Gaussian density. Thus, numerical computation of integrals of the form $E[\varphi_i(X)]$, where X is a random variable, is avoided.

Therefore, expressions for the constraints in terms of the Lagrange multipliers exist and are given by

$$
\begin{aligned}
l_0 &= \int e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = \sqrt{-\frac{2\pi}{2\lambda_2}} e^{\lambda_0 - \lambda_1^2/4\lambda_2}, \quad \lambda_2 < 0, \\
l_1 &= \int x \cdot e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = -\frac{\lambda_1}{2\lambda_2} \cdot l_0, \\
l_2 &= \int x^2 \cdot e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2} dx = \left(\frac{\lambda_1^2}{4\lambda_2^2} - \frac{1}{2\lambda_2}\right) \cdot l_0.
\end{aligned} \tag{22}
$$

Normalising the constraints yields the statistical characteristics of the Gaussian distribution

$$m = \frac{l_1}{l_0}, \qquad \sigma^2 = \frac{l_2}{l_0} - m^2, \tag{23}$$

where m is the mean and σ^2 is the variance.

Thus, the Lagrange multipliers in the pdf are defined in terms of the moments by

$$
\begin{aligned}
\lambda_2 &= -\frac{1}{2\sigma^2}, \\
\lambda_1 &= \frac{m}{\sigma^2}, \\
\lambda_0 &= -\frac{1}{2}\ln(2\pi\sigma^2) + \ln l_0 - \frac{m^2}{2\sigma^2}.
\end{aligned} \tag{24}
$$

Generic exponential model

The algorithm is implemented on a generic exponential model. The results are compared to those given in [4] for the same exponential model where the extended Kalman filter and a class of adaptive nonlinear filter (ANLF) based on partitioning algorithms were implemented.

The state model is given by

$$z_{k+1} = 1.7 \exp\left(-2z_k^2\right) + w_k,$$
$$y_k = z_k^3 + v_k. \tag{25}$$

Before implementing the FAD filter, we modify the above model so as to obtain a linear observation equation. Following the transformation given in Section 3, the resulting nonlinear model is

$$x_{k+1} = \left(1.7 \exp\left(-2x_k^{2/3}\right) + w_k\right)^3, \tag{26}$$
$$y_k = x_k + v_k, \tag{27}$$

where the noise processes are mutually independent and independent of the state. The state and measurement noises are taken to be Gaussian distributed with zero mean and variance $Q = 1$, $R = 0.5$, and densities

$$p_w(w) = N(0, 1) = \exp\left(\beta_0 + \beta_1 w + \beta_2 w^2\right),$$
$$p_v(v) = N(0, 0.5) = \exp\left(\mu_0 + \mu_1 v + \mu_2 v^2\right), \tag{28}$$

where β and μ are the Lagrange multipliers of the respective densities.

For comparison purposes, the initial state x_0 is set to zero and is Gaussian with density

$$p_{0|0}(x) = N(0, 0.25) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2\right). \tag{29}$$

Note that, due to the model transformation used, the x-state trajectory will span over a range of values that is equal to that of the z-state trajectory to the power 3. Therefore, the nonlinear filter will have to track huge jumps in state value.

5.1. Prediction

The prediction of the conditional density $p_{k+1|k}(x|\mathbf{y})$, given past observations $\mathbf{y} = (y_1, \ldots, y_k)$, requires the determination of the constraints

$$l_0(k+1|k) = \int p_{k+1|k}(x|\mathbf{y})\, dx,$$
$$l_1(k+1|k) = \int x \cdot p_{k+1|k}(x|\mathbf{y})\, dx, \tag{30}$$
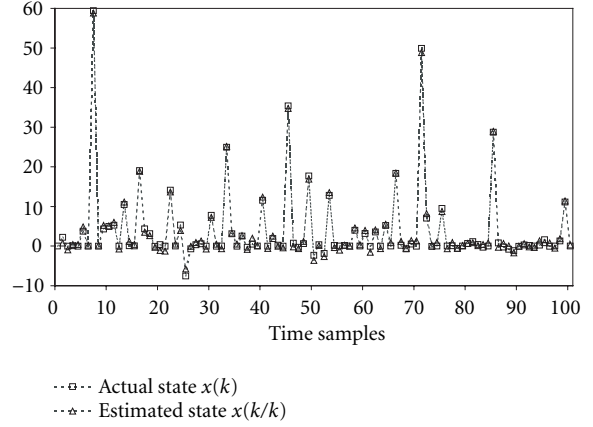$$l_2(k+1|k) = \int x^2 \cdot p_{k+1|k}(x|\mathbf{y})\, dx.$$

FIGURE 1: State estimation. Generic exponential model.

Using the state equation (26), the constraints are evaluated in terms of the actual form

$$l_0(k+1|k) = \iint p_{k|k}(x|\mathbf{y}) p_W(w)\, dx\, dw,$$
$$l_1(k+1|k) = \iint \left(1.7 \exp\left(-2x^{2/3}\right) + w\right)^3 \\ \times p_{k|k}(x|\mathbf{y}) p_W(w)\, dx\, dw, \tag{31}$$
$$l_2(k+1|k) = \iint \left(\left(1.7 \exp\left(-2x^{2/3}\right) + w\right)^3\right)^2 \\ \times p_{k|k}(x|\mathbf{y}) p_W(w)\, dx\, dw.$$

5.2. Update

Considering the actual observation $y_k = \eta$ and (27), the conditional density $p_{k|k}(x|\mathbf{y})$ given the observations up to time instant k results from (18)

$$p_{k|k}(x|\mathbf{y}) = C p_{k|k-1}(x|\mathbf{y}) p_v(\eta - x). \tag{32}$$

The expansion of the densities gives,

$$p_{k|k}(x|\mathbf{y}) = C e^{\lambda_0(k|k-1) + \lambda_1(k|k-1)x + \lambda_2(k|k-1)x^2} e^{\mu_0 + \mu_1(\eta-x) + \mu_2(\eta-x)^2}. \tag{33}$$

A simple addition of the coefficient of the like powers in x yields the updated Lagrange multipliers of the conditional density function,

$$\lambda_0(k|k) = \lambda_0(k|k-1) + \mu_0 + \mu_1\eta + \mu_2\eta^2,$$
$$\lambda_1(k|k) = \lambda_1(k|k-1) - \mu_1 - 2\mu_2\eta, \tag{34}$$
$$\lambda_2(k|k) = \lambda_2(k|k-1) + \mu_2.$$

5.3. Results

The x-state trajectory is shown in Figure 1, where the actual states x_k and their estimate $x_{k|k}$ are displayed. It is evident from Figure 1 that the FAD filter can easily track jumps in the state trajectory where an EKF would fail.
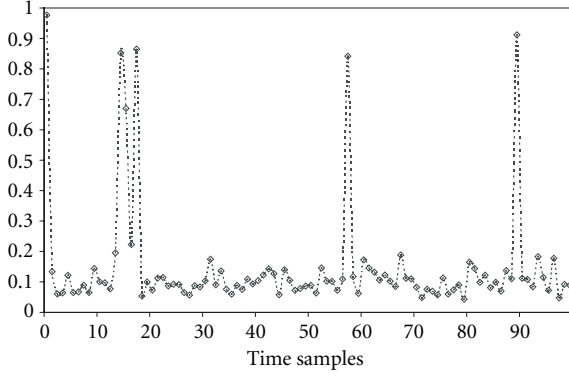
FIGURE 2: Error performance of the FAD filter. Generic exponential model (100 time samples over 50 runs).

TABLE 1: Average NRMS error comparison.

	EKF	ANLF (pr)	ANLF (fil)	FAD
ANRMS	1.2	0.73	0.47	0.15

In [4], the authors compare the EKF to an ANLF formulation with predicted estimates for the innovations and one with filtered estimates. Both involved a bank of 10 subfilters. The performances were evaluated in terms of normalised root mean square error (NRMSE) of 100 time samples over 50 runs

$$\text{NRMS}(k) = \sqrt{\frac{\text{MSE}(k)}{(1/50) \sum_{50} x_k^2}}, \tag{35}$$

where the MSE is defined as

$$\text{MSE}(k) = \frac{1}{50} \sum_{50} \left[x_k - x_{k|k} \right]^2. \tag{36}$$

In [4], the authors found that the NRMS error for the EKF was always greater than 0.6. For the ANLF with predicted estimates for the innovation, the NRMS error varies within the interval 0.4 and 1. Finally, for the ANLF with filtered estimated estimates, the NRMS error varies within the interval 0.3 and 0.6.

For comparison purposes, the performance of the filter is evaluated in terms of normalised root mean-square error as is done in [4]. Figure 2 displays the NRMS(k) resulting from 50 runs of the proposed filter. Over all, the performance of the proposed filter is superior to the performance of the EKF and both ANLFs.

In [4], the authors also give the values of the average normalised RMS error (average over 100 values) for the EKF, the ANLF (predicted) and the ANLF (filtered). We compare these values to the average NRMS error obtained with the FAD filter.

The values shown in Table 1, again, confirm the superiority of the FAD filter over the others.

6. CONCLUSION

This paper has presented the theoretical development of a nonlinear adaptive filter based on a concept of filtering by approximated densities (FAD). The proposed recursive algorithm uses a maximum entropy principle subject to constraints to approximate the filtering equations arising from nonlinear state space models.

The approach approximates the state distributions through parameterised exponential density functions. The filter updates and predicts the conditional densities given past and actual observations. The prediction uses the entropy principle, and the update of the Bayes theorem.

The simulation results presented on a generic exponential model prove the ability of the proposed approach to work with highly nonlinear system dynamics. The performance of the filter is superior to the performance found in [4] by the authors on the same exponential model for the extended Kalman filter and two adaptive nonlinear filters based on partitioning algorithms, one formulated with predicted estimates for the innovations and the other with filtered estimates.

The proposed approach has also been applied to nonlinear models involving time-varying Markovian parameters for applications such as motion estimation [14] in motion compensated video compression, and mammography [15], both involving multimodal Gaussian distributions. Further research is currently being conducted for synthetic aperture radar processing.

Other developments not yet published demonstrate the applicability of the approach in a purely nonlinear, non-Gaussian, nonstationary context.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[2] J. V. Candy, *Signal Processing: The Model-Based Approach*, McGraw-Hill, New York, NY, USA, 1986.

[3] A. Gelb, Ed., *Applied Optimal Estimation*, MIT Press, Cambridge, Mass, USA, 1974.

[4] D. G. Lainiotis and P. Papaparaskeva, "A new class of efficient adaptive nonlinear filters (ANLF)," *Trans. on Signal Processing*, vol. 46, no. 6, pp. 1730–1737, 1998.

[5] J. B. Pearson, "On nonlinear least-squares filtering," *Automatica*, vol. 4, pp. 97–105, 1967.

[6] H. W. Sorenson and A. R. Stubberud, "Nonlinear filtering by approximation of the a posteriori density," *International Journal of Control*, vol. 18, pp. 33–51, 1968.

[7] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[8] V. Benes, "Exact finite dimensional filters for certain diffusions with non-linear drift," *Stochastisc*, vol. 5, pp. 65–92, 1981.

[9] M. L. Andrade, L. Gimeno, and M. J. Mendes, "On the optimal and suboptimal nonlinear filtering problem for discrete-time systems," *IEEE Trans. Automatic Control*, vol. 23, no. 6, pp. 1062–1067, 1978.

[10] A. M. Makowski, "Results on filtering problem for linear systems with non-Gaussian initial conditions," in *Proc. IEEE Int. Conf. on Decision and Control*, pp. 101–104, Orlando, Fla, USA, 1982.

[11] U. G. Haussmann and E. Pardoux, "A conditionally almost linear filtering problem with non-Gaussian initial condition," *Stochastisc*, vol. 23, pp. 241–275, 1988.

[12] F. E. Daum, "New exact nonlinear filters," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed., chapter 8, pp. 199–226, Marcel Dekker, New York, USA, 1988.

[13] V. F. Ruiz and D. de Brucq, "Modèle d'état à coefficient aléatoire: Filtrage par densités approchées," *J. Traitement du signal*, vol. 11, no. 2, pp. 133–140, 1994.

[14] V. F. Ruiz and A. Skodras, "Approximated densities for block-based motion estimation," *Electronics Letters*, vol. 33, no. 16, pp. 1364–1365, 1997.

[15] V. F. Ruiz, "Statistical modelling for breast screening abnormalities: a concept of approximated densities," in *Proc. 7th IASTED Int. Conf. Signal and Image Processing, SIP 2000*, pp. 257–262, Las Vegas, Nev, USA, 2000.

Virginie F. Ruiz received her B.S. in 1989, M.S. in 1990, and Ph.D. in signal processing in 1993 from the University of Rouen, France. She has the honour of being a recipient of the French Foreign Office, Lavoisier programme 1993. Her research interests are in the development of fundamental principle for non-Gaussian signal processing. Particular areas of applications are in medical signal and image processing, synthetic aperture radar, communications. She has published over 30 papers. She has been with the Department of Cybernetics at University of Reading since 1998. She is the Director of the Biomedical Engineering and Cybernetics programme and is currently involved in a number of international research projects and industrial projects. She is a member of many technical programme committees for international conferences and serves as reviewer for a number of international journals. She is a member of the IEEE and an associate member of IEE.