World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:7, 2009

Unscented Grid Filtering and Smoothing for Nonlinear Time Series Analysis

Nikolay Nikolaev, and Evgueni Smirnov

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

Abstract— **This paper develops an unscented grid-based filter and a smoother for accurate nonlinear modeling and analysis of time series. The filter uses unscented deterministic sampling during both the time and measurement updating phases, to approximate directly the distributions of the latent state variable. A complementary grid smoother is also made to enable computing of the likelihood. This helps us to formulate an expectation maximisation algorithm for maximum likelihood estimation of the state noise and the observation noise. Empirical investigations show that the proposed unscented grid filter/smoother compares favourably to other similar filters on nonlinear estimation tasks.**

I. INTRODUCTION

Nonlinear time series modeling has practical applications in various fields, such as automatic control, signal processing, econometrics, etc. [17]. The rationale is that many real-world time series assume descriptions by nonlinear discrete state-space models. Such latent state models are learned through filtering and smoothing. The filtering pass involves two steps: a time step, which generates a state prior, and a measurement step, which updates the posterior state distribution. The smoothing pass computes backwards improved estimates of the state posterior using information that has not been available during the forward processing. Having algorithms to compute the state posterior enables us to find the noise hyperparameters, so as to obtain accurate forecasts from unseen inputs.

Nonlinear models are often processed using linearization with the derivatives of the observation equation, and application of the standard equations from the Extended Kalman filter (EKF) [6]. Such model linearization through the output derivatives with respect to the state, however, produces large errors as it does not reflect the uncertainty in the hidden state, and guarantees achieving only first-order accuracy. Higher accuracy can be achieved using derivative free methods that rely on sampling to approximate the state distribution. Among the stochastic and deterministic sampling methods, current research directs more attention to the deterministic sampling methods as more efficient and accurate. The deterministic sampling filtersing methods include Sigma Point Filters (SPF) [19], like Unscented Kalman filters (UKF) [18], and Central Difference Filters (CDF) [12], as well as Quadrature Kalman Filters (QKF) [9], [1].

Nikolay Nikolaev with the Department of Computing, Goldsmiths College, University of London, New Cross, London SE14 6NW, UK (phone: +44 (0)207 919 7854; fax: +44 (0)207 919 7853; email: n.nikolaev@gold.ac.uk).

Evgueni Smirnov is with the MICC-IKAT, Maastricht University, Maastricht 6200 MD, The Netherlands (phone: +31 (0)433 882 023; email: smirnov@micc.unimaas.nl).

Sigma point filters, relying on the unscented sampling technique [7], describe the model nonlinearities with accuracy up to the second-order when the state density is Gaussian. During the second measurement update step, however, SPF filters still compute the posterior with the standard Kalman equations through linearization of the observation model, which may cause two problems: 1) as the model nonlinearities increase the accuracy decreases; and 2) the linearization may lead to a breakdown in the correlation between the state and the observation, which prevents the state updating [20]. Moreover, the SPF typically assume that the data are normally distributed, while if they are skewed their accuracy decreases.

These issues can be addressed by designing grid filters [14] that use numerical approximation through sampling during both the time and the measurement updating steps. Grid-based methods provide more flexibility as they attempt to match directly the moments of the distributions of interest and can perform well even on skewed distributions. The recent one-step unscented Kalman filter (OUKF) [20] implements this idea for low dimensional systems using Gauss-Hermite quadratures. A shortcoming of the OUKF is the use of Gaussian quadratures to determine the sampling points limits its usefulness because it requires a lot of points to achieve accurate approximation.

This paper elaborates an Unscented Grid Filter with a Smoother (UGFS) for accurate estimation of the state mean and variance. A distinguishing feature of UGFS is that during both the time and measurement updates, as well as during backward smoothing it approximates directly the distributions using deterministic unscented sampling [7]. There are two main contributions of the presented work: 1) it develops an unscented grid filter and smoother for multidimensional inputs and outputs, and 2) it derives an Expectation Maximisation (EM) algorithm [16] for learning the state and observation noise parameters. The formulation of the EM algorithm involves design of a backward smoother to calculate the complete data likelihood, which is necessary to find the noises. The empirical investigations show that UGFS compares favorably to similar filters on modeling nonstationary series and option price modeling.

This paper is organized as follows. Section two introduces the basics of nonlinear estimation and grid-based filtering. Section three presents the unscented transform and elaborates the novel filter. The next section four gives the EM algorithm with the smoother. Section five offers the empirical study. Finally a brief discussion and conclusions are provided.

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

II. NONLINEAR ESTIMATION

The task of nonlinear estimation is to infer the dynamical characteristics of a system that models given series of discrete noisy observations $D = \{\mathbf{y}_t\}_{t=1}^{T}$. The latent component of such a system that describes its dynamics of the unobserved underlying process is the state. We consider the following nonlinear model:

$$\begin{array}{ll} \mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}) + \mathbf{q}_{t-1} & \text{/state equation/} \\ \mathbf{y}_t = \mathbf{f}(\mathbf{x}_t) + \mathbf{r}_t & \text{/observation equation/} \end{array} \quad (1)$$

where $\mathbf{x}_t \in \mathcal{R}^n$ is the state, $\mathbf{y}_t \in \mathcal{R}^m$ is the observation at time t, $\mathbf{g}$ is the state transition function, $\mathbf{q}_{t-1}$ is the state noise, and $\mathbf{f}$ is a nonlinear measurement function driven by observational noise $\mathbf{r}_t$. The noises are assumed to be Gaussian with unknown variances $\mathbf{q}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{t-1})$ and $\mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$.

A. Recursive Bayesian Inference

The analysis of sequential data described by such state-space models in a probabilistic setting is performed using the transition density $p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim \mathcal{N}(\mathbf{x}_t|\mathbf{g}(\mathbf{x}_{t-1}), \mathbf{Q}_{t-1})$ and the measurement density $p(\mathbf{y}_t|\mathbf{x}_t) \sim \mathcal{N}(\mathbf{y}_t|\mathbf{f}(\mathbf{x}_t), \mathbf{R}_{t-1})$. Nonlinear estimation of this probabilistic representation is carried out following the principles of recursive Bayesian inference. According to it, the posterior distribution of the latent state variable $\mathbf{x}_t$ can be obtained through filtering using the Bayes rule in the following way [8]:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \quad (2)$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = C_t^{-1} p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \quad (3)$$

$$\text{where } C_t = \int p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) d\mathbf{x}_t$$

where $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is the state posterior (filtering distribution), $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is the state prior (predictive distribution), and $p(\mathbf{y}_t|\mathbf{x}_t)$ is the data likelihood.

Improved estimates of the posterior density $p(\mathbf{x}_t|\mathbf{y}_{1:T})$, that use subsequently arrived information (not available during the forward pass), can be obtained through a backward smoothing pass from the series end T down to the current point $t < T$. The smoothing equations are deduced from the integral of the joint distribution $p(\mathbf{x}_t, \mathbf{x}_{t+1}|\mathbf{y}_{1:T})$ [8]:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:T}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t+1}|\mathbf{y}_{1:T}) d\mathbf{x}_{t+1} \\ &= \int p(\mathbf{x}_t|\mathbf{x}_{t+1}, \mathbf{y}_{1:T}) p(\mathbf{x}_{t+1}|\mathbf{y}_{1:T}) d\mathbf{x}_{t+1} \end{aligned} \quad (4)$$

which follows from the Markovian nature of the latent state sequence. The smoothed posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:T})$ is assumed Gaussian.

The problem is that these integrals can not be solved analytically for nonlinear models because they lead to untractable nonstandard distributions (without a predominant mode). This difficulty can be alleviated using sampling or linearization of the observation model (using the Taylor's expansion). The linearization, however, tends to prodice large errors and does not reflect the state uncertainty.

B. Grid-Based Estimation

Integral approximation in recursive modeling is often accomplished by deterministic or stochastic sampling [4], [14]. The deterministic sampling, considered as more accurate, computes the integrals by discrete summation and averaging over a finite set of carefully chosen grid points $\mathbf{x}_t^{(i)}$, $i = 1, ..., N_{(t-1)}$. This enables us to evaluate recursively the probability densities of interest by drawing sample points at each algorithmic step, as they induce point-mass representations of these densities.

Following the principles of sequential Bayesian inference, a grid filter can be defined by replacing the recursively computed densities with their numerical approximation. This approach leads to a grid filter that calculates iteratively the predictive state distribution as follows:

$$\widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \sum_{i=1}^{N_{(t-1)}} w_t^{(i)} p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}) \widehat{p}(\mathbf{x}_{t-1}^{(i)}|\mathbf{y}_{1:t-1}) \quad (5)$$

and the corresponding filtering distribution also by weighted summation:

$$\begin{aligned} \widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}) &= \widehat{C}_t^{-1} w_t^{(i)} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \widehat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}) \quad (6) \\ \text{where } \widehat{C}_t &= \sum_{i=1}^{N_{(t)}} w_t^{(i)} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \widehat{p}(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t-1}) \end{aligned}$$

where $N_{(t)}$ is the number of sampled states at time t, and $w_t^{(i)}$ are the weights for the samples $\mathbf{x}_t^{(i)}$.

There are various filtering techniques using finite sum approximations of integrals, like Gauss-Hermite quadrature filters [9], [20], [1], and Quasi-Monte Carlo filters. Although these techniques place the samples on optimal locations, they need a large number of points which grows exponentially with the increase of the state dimension. An effective approach to deterministic sampling for approximation that requires less points is provided by the unscented transform [7].

III. UNSCENTED GRID FILTERING

The popular nonlinear filters relying on numerical integration are similar in doing the time updating by discrete weighted summation of sampled states, obtained after passing them through the particular state equation. They implement differently however the measurement updating either in one or in two consequtive steps: 1) with direct simulation of the state posterior distribution; and 2) tackling the joint input output density first, and next handling the likelihood and the prior.

Two step measurement updating is used by the filters from the SPF family, like UKF, CDF and QKF. They draw sample states (sigma-points) in order to approximate the joint input output density, which in case of Gaussian noise leads to tractable recursions for state adaptation. However when the state density is highly skewed it does not perform well. Moreover, if the inputs and the outputs are uncorrelated such filters fail to update the state [20]. One-step measurement updating is a more general strategy for filtering as it can handle flexibly distributions of various forms [9].

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

A. The Unscented Transform

The unscented transform [7] is a technique that enables us to estimate the statistics of nonlinearly transformed random variables, like the state in our nonlinear state-space model. It suggests to pick samples from carefully selected locations aroung the particular variable. The spread of the points is determined in such a way so as to obtain a density estimate with the same properties as the true unknown distribution. This helps us to achieve higher accuracy (up to second order) of modeling the mean and variance than in the case of attempting Taylor approximations.

The numerically stable scaled version of the unscented transform attains high fitting accuracy by computing symmetric sigma-points by the following algebraic operations [7]:

$$\begin{aligned} \mathbf{x}_{t-1}^{(0)} &= \widehat{\mathbf{x}}_{t-1} \\ \mathbf{x}_{t-1}^{(i)} &= \widehat{\mathbf{x}}_{t-1} + \left[\sqrt{(L+\lambda)\,\mathbf{P}_{t-1}}\right]_i, \; i = 1, ..., L \\ \mathbf{x}_{t-1}^{(i)} &= \widehat{\mathbf{x}}_{t-1} - \left[\sqrt{(L+\lambda)\,\mathbf{P}_{t-1}}\right]_i, \; i = L+1, ..., 2L \end{aligned} \tag{7}$$

where L is the state dimension, $\widehat{\mathbf{x}}_{t-1}$ is the previous mean state, $\mathbf{P}_{t-1}$ is the covariance, and the index i indicates the i-th row of the matrix square root.

The state distribution is approximated by weighted averaging over these sigma-points. Each sigma point is associated with a corresponding weight $\mathcal{S}_{i,t-1} = \{W_i, \mathbf{x}_{t-1}^{(i)}\}$, $i = 0, ..., 2L$, and they are normalised $\sum_{i=0}^{2L} W_i = 1$. The weights are computed with the following equations:

$$\begin{aligned} W_m^{(0)} &= \lambda/(\lambda+1) \\ W_c^{(0)} &= \lambda/(\lambda+1) + (1 - \alpha^2 + \beta) \\ W_m^{(i)} &= W_c^{(i)} = 1/(2(L+\lambda)), \; i = 1, ..., 2L \end{aligned} \tag{8}$$

where α, β, and λ are scaling parameters.

The parameters α and β are chosen to control the spread of the distribution. The spread is scaled with respect to the mean of the distribution with the specific variable $\lambda = \alpha^2(L+\kappa) - L$, using another parameter κ. This scaling parameter λ has to be greater than or equal to zero in order to achieve positive definite terms under the square root.

B. Time Updating

The time updating phase computes the predictive state distribution $\widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\widehat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1})$ which can be written in matrix notation as follows [15]:

$$\begin{aligned} \mathbf{X}_{t-1|t-1} &= \left[\mathbf{x}_{t-1}^{(0)}, \mathbf{x}_{t-1}^{(1)}, ..., \mathbf{x}_{t-1}^{(2L)}\right] \\ \mathbf{X}_{t-1} &= \mathbf{g}\left(\mathbf{X}_{t-1|t-1}\right) \\ \widehat{\mathbf{x}}_{t|t-1} &= \mathbf{X}_{t-1}\mathbf{w}_m \qquad (9) \\ \mathbf{P}_{t|t-1} &= \mathbf{X}_{t-1}\mathbf{W}\mathbf{X}_{t-1}^T \qquad (10) \\ \mathbf{W} &= \left(\mathbf{I} - [\mathbf{w}_m, ...\mathbf{w}_m]\right)\mathbf{Z}\left(\mathbf{I} - [\mathbf{w}_m, ...\mathbf{w}_m]\right)^T \\ \mathbf{Z} &= diag\left(W_c^{(0)}, ..., W_c^{(2L)}\right) \end{aligned}$$

where $\widehat{\mathbf{x}}_{t|t-1}$ is the mean state (filtering prior), $\mathbf{P}_{t|t-1}$ is the prior state covariance, $\mathbf{I}$ is the identity matrix, and T denotes transpose of a matrix.

C. Measurement Grid Updating

A specific feature of the grid filter is that it performs measurement updating by direct approximation of the state posterior integral [9], [20], [10] using deterministic sampling. The state posterior is obtained as a linear mixture in one step, not in the typical two steps of linearized approximation of the joint density through the derivatives of the observation model, and, next, linear filtering with the Kalman equations. One-step measurement updating is a more general filtering strategy filtering as it can handle distributions of various forms.

We implement this idea through sampling from the prior $\widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ using again the unscented transform. Symmetric points $\mathcal{S}_{i,t|t-1} = \{W_i, \mathbf{x}_{t|t-1}^{(i)}\}$, $i = 0, ..., 2L$, are drawn around the predicted mean $\widehat{\mathbf{x}}_{t|t-1}$ as follows:

$$\begin{aligned} \mathbf{x}_{t|t}^{(0)} &= \widehat{\mathbf{x}}_{t|t-1} \\ \mathbf{x}_{t|t}^{(i)} &= \widehat{\mathbf{x}}_{t|t-1} + \left[\sqrt{(L+\lambda)\,\mathbf{P}_{t|t-1}}\right]_i, \; i = 1, ..., L \\ \mathbf{x}_{t|t}^{(i)} &= \widehat{\mathbf{x}}_{t|t-1} - \left[\sqrt{(L+\lambda)\,\mathbf{P}_{t|t-1}}\right]_i, \; i = L+1, ..., 2L \end{aligned} \tag{11}$$

where $\mathbf{P}_{t|t-1}$ is obtained in the time updating phase.

The effect from the target at the particular moment on the state posterior is absorbed through the likelihood. The likelihood integral may be envisioned analytically tractable for measurement functions with linear dependance on the noise [20]. Therefore, we can evaluate the likelihood with the outputs $\mathbf{Y}_t$ produced by passing the sigma-points $\mathbf{x}_{t|t}^{(i)}$ from the state matrix $\mathbf{X}_t$ through the measurement function:

$$\begin{aligned} \mathbf{X}_t &= \left[\mathbf{x}_{t|t}^{(0)}, \mathbf{x}_{t|t}^{(1)}, ..., \mathbf{x}_{t|t}^{(2L)}\right] \\ \mathbf{Y}_t &= \mathbf{f}\left(\mathbf{X}_t\right)\mathbf{w}_m \qquad (12) \\ p(\mathbf{y}_t|\mathbf{x}_{t|t}) &= \int \delta\left(\mathbf{Y}_t - y_t\right) p(\mathbf{r}_t) d\mathbf{r}_t \qquad (13) \end{aligned}$$

where δ is the Kronecher delta. There are standard density functions foe computing the probability $p(\mathbf{y}_t|\mathbf{x}_{t|t})$.

The mean and variance of the state posterior distribution $\widehat{p}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\widehat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t})$ are finally obtained with the following expressions:

$$\begin{aligned} \widehat{\mathbf{x}}_{t|t} &= \widehat{C}_t^{-1} p(\mathbf{y}_t|\mathbf{x}_{t|t})\mathbf{X}_t\mathbf{w}_m \qquad (14) \\ \mathbf{P}_{t|t} &= \mathbf{X}_t\mathbf{W}\mathbf{X}_t^T \qquad (15) \\ \mathbf{W} &= \left(\mathbf{I} - [\mathbf{w}_m, ...\mathbf{w}_m]\right)\mathbf{Z}\left(\mathbf{I} - [\mathbf{w}_m, ...\mathbf{w}_m]\right)^T \\ \mathbf{Z} &= \widehat{C}_t^{-1} p(\mathbf{y}_t|\mathbf{x}_{t|t}) diag\left(W_c^{(0)}, ..., W_c^{(2L)}\right) \end{aligned}$$

where the normalizing constant is $\widehat{C}_t = p(\mathbf{y}_t|\mathbf{x}_{t|t})\mathbf{w}_m$.

The UGF is a reliable algorithm as it can deal with data coming from heavier than the normal (Gaussian) distribution tails as well as from skewed distributions. It can handle also situations with uncorrelated states and observations, therefore it can learn a larger class of models than the original UKF.

IV. NONLINEAR EXPECTATION MAXIMISATION

The dynamic expectation Maximisation (EM) algorithm [16] searches for the maximum of the log likelihood $\log p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\mathbf{Q}, \mathbf{R})$, so as to find optimal noise parameters $\mathbf{Q}$ and $\mathbf{R}$. The algorithm alternates between expectation and

maximisation steps. Since the likelihood has to be evaluated with the complete data, the expectation step has to carry out a forward pass followed by a backward pass over the series.

A. Unscented Grid Smoothing

The unscented Rauch-Tung-Striebel type algorithm computes the smoothed density $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_{1:T})$ running backwards down to the begining of the time interval [13]. The novelty here is that it evaluates the conditional distribution $p(\mathbf{x}_t, \mathbf{x}_{t+1}|\mathbf{y}_{1:T})$ through the use of deterministic sampling. After performing unscented sampling:

$$\begin{aligned} \mathbf{X}_{t-1} &= \left[\mathbf{x}_{t-1}^{(0)}, \mathbf{x}_{t-1}^{(1)}, ..., \mathbf{x}_{t-1}^{(2L)}\right] \\ \mathbf{X}_t &= \mathbf{g}\left(\mathbf{X}_{t-1}\right) \quad (16) \\ \mathbf{S}_{t-1} &= \mathbf{X}_{t-1}\mathbf{W}\mathbf{X}_t^T, \mathbf{P}_{t|t-1} = \mathbf{X}_t\mathbf{W}\mathbf{X}_t^T \quad (17) \end{aligned}$$

the smoothed mean state and covariance matrix are computed as follows:

$$\begin{aligned} \mathbf{J}_{t-1} &= \mathbf{S}_{t-1}\left(\mathbf{P}_{t|t-1}\right)^{-1} \quad (18) \\ \widehat{\mathbf{x}}_{t-1|T} &= \widehat{\mathbf{x}}_{t-1|t-1} + \mathbf{J}_{t-1}(\widehat{\mathbf{x}}_{t|T} - \widehat{\mathbf{x}}_{t|t-1}) \quad (19) \\ \mathbf{P}_{t-1|T} &= \mathbf{P}_{t-1|t-1} + \mathbf{J}_{t-1}(\mathbf{P}_{t|T} - \mathbf{P}_{t|t-1})\mathbf{J}_{t-1}^T \quad (20) \end{aligned}$$

which starts with $\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{T|T}$ and $\mathbf{P}_{t|t} = \mathbf{P}_{T|T}$.

In order to apply the EM algorithm it is necessary to compute also the cross-covariance:

$$\begin{aligned} \mathbf{H}_{t-1} &= \mathbf{P}_{t,t-1|T} - \mathbf{P}_{t-1|t-1} \\ \mathbf{P}_{t-1,t-2|T} &= \mathbf{P}_{t-1|t-1}\mathbf{J}_{t-2}^T + \mathbf{J}_{t-1}\mathbf{H}_{t-1}\mathbf{J}_{t-1}^T \quad (21) \end{aligned}$$

which starts with $\mathbf{P}_{T,T-1|T} = \mathbf{P}_{t-1|t-1}$.

B. Maximisation Step

The maximisation step aims at finding such state and observation noises that maximise the expected log-likelihood of the complete data set. The complete likelihood is[1]:

$$p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\mathbf{Q}, \mathbf{R}) = p(\mathbf{x}_0)\prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1})\prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t) \quad (22)$$

assuming that the noises are uncorrelated.

The optimisation is performed by taking the expectation $\langle\cdot\rangle$ of the logarithm of the complete likelihood:

$$\begin{aligned} &2\log p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\mathbf{Q}, \mathbf{R}) = \\ &-T\log|\mathbf{Q}| - \sum_{t=1}^{T}(\mathbf{x}_t - \mathbf{g}(\mathbf{x}_{t-1}))\mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{g}(\mathbf{x}_{t-1})) \\ &-T\log|\mathbf{R}| - \sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{f}(\mathbf{x}_t))\mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{f}(\mathbf{x}_t)) \quad (23) \end{aligned}$$

where the constants are omitted.

[1]The theory of maximum likelihood estimation assumes that the available data for training $(\mathbf{x}_t, \mathbf{y}_t), 1 \leq t \leq T$ are independent identically distributed [16].

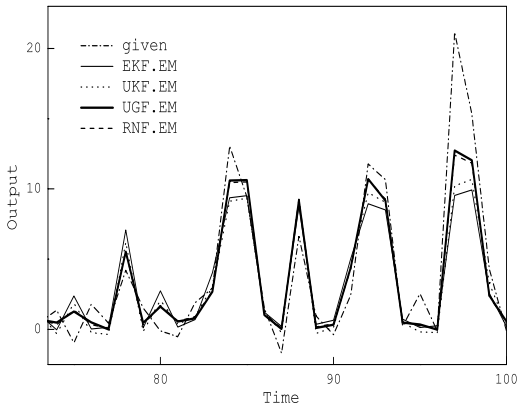

Fig. 1. Approximations of the UNGM series by the studied filters.

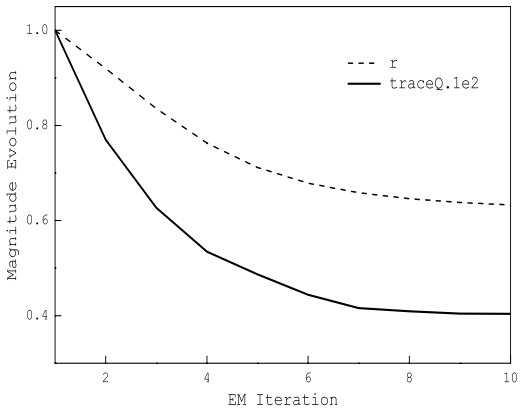

Fig. 2. Evolution of the weight noise covariance and the output noise parameter during training.

Taking the derivatives of the expectation of the likelihood $\langle\log p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\mathbf{Q}, \mathbf{R})\rangle$, equating to zero, and solving for $\mathbf{Q}$ and $\mathbf{R}$ yields corresponding formulae for their updates:

$$\begin{aligned} \mathbf{A} &= \sum_{t=1}^{T}\left(\widehat{\mathbf{x}}_{t|T}\widehat{\mathbf{x}}_{t|T}^T + \mathbf{P}_{t|T}\right) \\ \mathbf{B} &= \sum_{t=1}^{T}\left(\widehat{\mathbf{x}}_{t|T}\widehat{\mathbf{x}}_{t-1|T}^T + \mathbf{P}_{t,t-1|T}\right) \\ \mathbf{Q} &= T^{-1}\left(\mathbf{A} - \mathbf{B}\mathbf{B}^T\right) \quad (24) \end{aligned}$$

where $\widehat{\mathbf{x}}_{t|T}$ and $\widehat{\mathbf{x}}_{t-1|T}$ are the smoothed state vectors.

The likelihoods $p(\mathbf{y}_t|\mathbf{x}_t)$ are approximated directly by finite sum approximations using the unscented sigma-points generated by equation (11). This leads to the following formaula for the observational noise covariance:

$$\mathbf{R} = T^{-1}\sum_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t)\mathbf{w}_m \quad (25)$$

where $\mathbf{w}_m = [W_m^{(0)}, W_m^{(1)}, ..., W_m^{(2L)}]^T$, and $p(\mathbf{y}_t|\mathbf{x}_t) = \int \delta\left(\mathbf{Y}_t - y_t\right) p(\mathbf{r}_t) d\mathbf{r}_t$ with $\mathbf{Y}_t$ computed by (12).

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

TABLE I
AVERAGED ERRORS AND STANDARD DEVIATIONS FROM 50 RUNS OVER THE UNGM SERIES USING THE FILTERS TUNED TO MAKE 10 EM ITERATIONS.

Filter	NSE	StDev
EKF	53.94	0.0522
UKF	46.75	0.0481
UGF	25.54	0.0267
RNF	26.21	0.0289

It should be noted that although the EM algorithm features proven convergence to a maximum of the likelihood function, this can be either local or global minimum when nonlinear models are manipulated.

V. EMPIRICAL INVESTIGATIONS

Experiments were conducted to find out how the learning potential of the developed UGF relates to similar filters from previous research. There were considered tasks which were already found to require nonlinear modeling and nonlinear estimation. In order to facilitate comparisons with relevant research we designed and implemented the following four filters: classical EKF [6] as a baseline method, the UKF [18], the UGF, and a Robust Nonlinear Filter (RNF) [2]. The RNF also uses EM for hyperparameter reestimation, and can deal with heavy tailed output noise.

Modelling Nonstationary Dynamics. The univariate nonstationary growth model (UNGM) is a higly nonlinear benchmarking model which is challenging for learning by standard filtering algorithms [3]:

$$\begin{aligned} x_t &= \alpha x_{t-1} + \beta \frac{x_{t-1}}{1+x_{t-1}^2} + \gamma \cos(1.2(t-1)) + u_t \\ y_t &= 0.05 x_t^2 + r_t \end{aligned} \quad (26)$$

where the noises are Gaussian $u_t \in \mathcal{N}(0, \sigma_u^2)$ and $r_t \in \mathcal{N}(0, \sigma_r^2)$ with variances set to $\sigma_u^2 = 0.1$ and $\sigma_r^2 = 3\sin(0.05t)$, the initial state is $x_0 = 0.1$, $\alpha = 0.5$, $\beta = 25$, $\gamma = 8$. A sequence of $T = 250$ values was generated. The cosine is independent of x_t but depends on t, and so it simulates time-varying noise.

We designed a multilayer perceptron (MLP) network with 6 hidden sigmoidal units and one summation output. The filters were applied to find the weights (19 in total) using only one input value x_t to predict the output y_t. All filters were initialised with $P_0 = 1$, $[Q]_{ii} = 10^{-2}$, and $r = 1$. The initial weights were randomly drawn from a zero-mean Gaussian with covariance one.

Table 1 provides the normalised squared errors (NSE) and the corresponding variances averaged over independent 50 runs conducted with each filter. These normalized errors were calculated with the formula: $NSE = \sqrt{\sum_t (y_t - \widehat{y}_t)^2}$. The results in Table 1 demonstrate that UGF outperforms the other algorithms on this task, although it is only slightly better than the RNF. It should be pointed out that RNF uses approximations through linearisations with derivatives like EKF, so one is inclined to think that avoiding simply linearisations of the nonlinear equation in filtering as done in UKF is not a sufficient condition for accurate modelling.

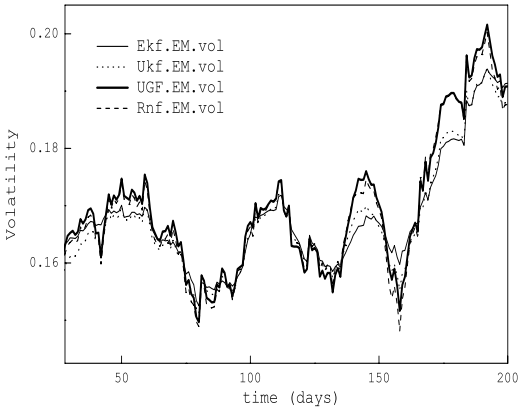

Fig. 3. Sequentially estimated interest rates by the studied filters, recorded during runs with price K=2925.

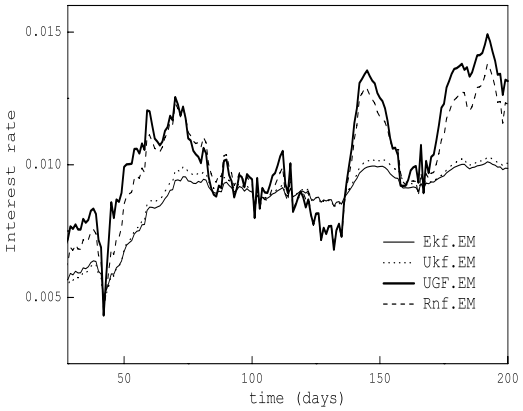

Fig. 4. Sequentially estimated implied volatility by the studied filters, recorded during runs with price K=2925.

Segments from all curves of the estimated mappings by the studied filters are plotted in Figure 1. These curves are recorded after one run with each of the filters. They show that the UGF mapping is quite close the one generated by the RNF. One can see that the UGF and RNF curves are closer to the given curve especially at the peaks around the 85, 93 and 98 time instants, that is them seem to capture better the series fluctuations.

Figure 2 depicts the changes of the weight covariance matrix and the output noise parameter r_t being re-estimated with the EM algorithm. The curves are obtained during one run with the UGF algorithm. Figure 2 may be envisoned as an illustration of the convergence of the proposed dynamic EM algorithm.

Option Price Modelling. Option pricing is an problem whose accurate solutions help to work efficiently with various financial derivatives and hedge against risks [5]. It is a difficult problem because of the nonstationary and stochastic behaviour of the market price series.

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

TABLE II
ACCURACY OF MODELING THE CALL/STRIKE PRICES, ACHIEVED BY THE STUDIED FILTERS AFTER 10 EM ITERATIONS.

Call	2925	3025	3125	3225	3325
EKF	0.0509	0.0735	0.0614	0.0493	0.0289
UKF	0.0507	0.0733	0.0612	0.0492	0.0288
UGF	0.0482	0.0724	0.0602	0.0487	0.0283
RNF	0.0486	0.0728	0.0605	0.0485	0.0284

TABLE III
ACCURACY OF MODELING THE PUT/STRIKE PRICES, ACHIEVED BY THE STUDIED FILTERS AFTER 10 EM ITERATIONS.

Put	2925	3025	3125	3225	3325
EKF	0.0363	0.0545	0.0624	0.0728	0.0752
UKF	0.0362	0.0544	0.0623	0.0726	0.0751
UGF	0.0345	0.0535	0.0613	0.0714	0.0748
RNF	0.0357	0.0541	0.0614	0.0713	0.0749

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf

Assuming that prices follow a geometric Brownian motion, the fair prices of European call $\mathcal{C}$ and put $\mathcal{P}$ options are given by the Black-Scholes formulae[2]:

$$\begin{aligned} \mathcal{C} &= S\Phi(d_1) - Ke^{-rT_m}\Phi(d_1 - \sqrt{vT_m}) \\ \mathcal{P} &= \mathcal{C} + Ke^{-rT_m} - S \qquad (27) \\ d_1 &= \frac{\ln(S/K) + (r + 0.5v)T_m}{\sqrt{vT_m}} \qquad (28) \end{aligned}$$

where S is the stock price, K is the strike price, r is the risk-free interest rate, T_m is the time to maturity, v is the stock return variance (volatility), and $\Phi(\cdot)$ is the cumulative normal distribution function.

The objective here is to model option prices by treating the implied volatility and the interest rate as unobservables [11]. The volatility, the interest rate and the prices are assumed Gaussian, but their noises are unknown. The filters were applied with two inputs: the stock price divided by the strike price, and the time to maturity. The outputs were the call and put prices normalised also by the strike price.

There were taken five pairs of publicly available call and put option contracts from the FTSE-100 index from February till December 1994, with time to maturity December [3]. The corresponding strike prices were $2925, 3025, 3125, 3225$ and 3325. The initial noise covariance matrices were with entries $[R]_{ii} = 10^{-5}$ and $[Q]_{ii} = 10^{-6}$.

The NSE errors on the call and put prices are given in Tables 2 and 3. They were measured as one-step ahead prediction errors over the last 180 points from the series in order to allow the state to mature.

These results show that UGF compares favourably to the other studied here filters on this task, although on some series it is not the best one. The RNF exhibits a very close behaviour to UGF and it is better on the fourth series. The UKF and EKF filters are competitive but clearly inferior in NSE sense.

Figure 3 and Figure 4 plot the interest rate and implied volatility inferred using the call and put series with strike price $K = 2925$. These figures reveal that the UGF and RNF curves tend to oscillate more than the other. One reason for achieving close results by UGF and RNF is that both are Bayesian approaches whose application to seemingly peaked posterior densities yields similar accuracy.

[2]Strictly speaking the Black-Scholes formulae are valid upon several conditions: no arbitrage opportunities, continuous trading, no dividents, constant volatility and risk-free interest rate [5].

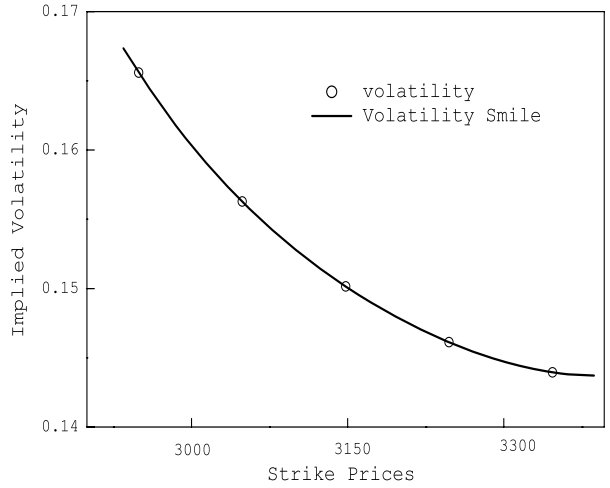

Fig. 5. Volatility smile (smirk) made by a 4-th order polynomial fit of the UGF estimates taken at t=100.

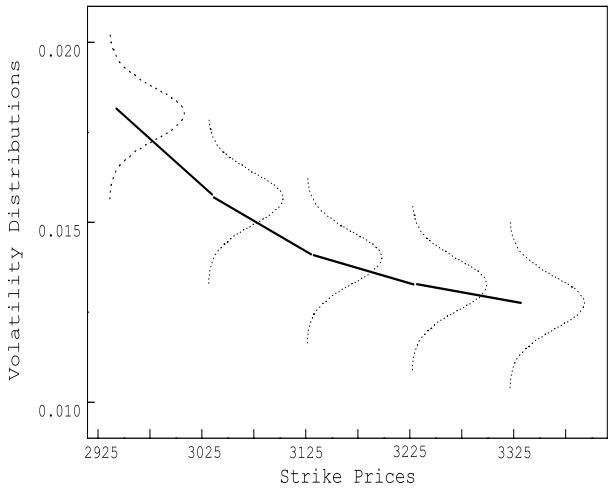

Fig. 6. Volatility distribution smile (smirk) made as piecewise-line of the UGF estimates taken at t=200.

Figure 5 plots the implied volatility recorded at time step $t = 100$ against the strike prices known as volatility smile. Since UGF infers the distribution of the volatility, we can also obtain the probability smile as a temporal sequence of the densities. Figure 6 shows the volatility distributions computed at each strike price at the end of training at time $t = 200$, whose connection resembles a probability smile.

VI. CONCLUSION

This paper presented an unscented grid filter and smoother that perform finite sum approximations using function evaluations, and do not require expensive computation of derivatives. Being a grid-based method the unscented transform samples the points evenly, and so it may eventually fail to achive very high accuracy in regions of high density. Another problem of UGFS is that the number of sampled points increases dramatically with the state dimension.

The UGFS approximates both nonlinear functions in the model by numerical integration, which is similar to Gaussian Sum filters so it can be used to make such mixture filters. The UGFS is general and it can be implemented also with Gauss-Hermite quadratures, which may improve it by sampling more points in cases when the state is low dimensional. UGFS can be be successfull on practical tasks because it can work well with nonlinear models and in relaxed circumstances, like mild non-Gaussianity and non-stationarity.

REFERENCES

[1] I. Arasaratnam, S. Haykin and R.J. Elliott, "Discrete-Time Nonlinear Filtering Algorithms using Gauss-Hermite Quadrature", *Proc. of the IEEE*, vol.95, pp.953-977, 2007.

[2] T. Briegel and V. Tresp, "Robust Neural Network Regression for Offline and Online Learning", in *Advances in NIPS 12*, Solla,S. et. al, Eds., Cambridge, MA: The MIT Press, 2000, pp.407-413.

[3] J.F.G. de Freitas, M. Niranjan and A.G. Gee, "Hierarchical Bayesian Models for Regularization in Sequential Learning", *Neural Computation*, vol.12, pp.933-953, 2000.

[4] A. Doucet, N. de Freitas and N. Gordon, *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag, 2001.

[5] J.C. Hull, *Options, Futures and Other Derivatives*, New Jersey: Prentice Hall, 2000.

[6] A. Jazwinsky, *Stochastic Processes and Filtering Theory*, New York: Academic Press, 1970.

[7] S.J. Julier and J.K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems", *Proc. SPIE Int. Soc. Opt. Eng.*, vol.3068, pp.182-193, 1997.

[8] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-space Models", *Journal of Computational and Graphical Statistics*, vol.5, pp.1-25, 1996.

[9] H.J. Kushner and A.S. Budhiraja, "A Nonlinear Filtering Algorithm Based on an Approximation of the Conditional Distribution", *IEEE Trans. on Automatic Control*, vol.45, pp.580-585, 2000.

[10] N. Nikolaev and E. Smirnov, "A One-Step Unscented Particle Filter for Nonlinear Dynamical Systems", in *Proc. Int. Conf. on Artificial Neural Networks, ICANN-2007*, LNCS 4668, Springer-Verlag, 2007, pp.747-756.

[11] M. Niranjan, "Sequential Tracking in Pricing Financial Options using Model Based and Neural Network Approaches", in *Advances in NIPS-8*, M.C.Mozer et al., Eds., Cambridge, MA: The MIT Press, 2000, pp.960-966.

[12] M. Norgaard, N.K. Poulsen and O. Ravn, "New Developments in State Estimation for Nonlinear Systems", *Automatica*, vol.36, pp.1627-1638, 2000.

[13] M.L. Psiaki and M. Wada, "Derivation and Simulation Testing of a Sigma-points Smoother", *Journal of Guidance, Control and Dynamics*, vol.30, pp.78-86, 2007.

[14] B. Ristic, M.S. Arulampalam and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.

[15] S. Särkkä, "Recursive Bayesian Inference on Stochastic Differential Equations", PhD Thesis, Dept. El. and Comm. Eng., Helsinki Univ. of Technology, 2006.

[16] R.H. Shumway and D.S. Stoffer, "An Approach to Time Series Smoothing and Forecasting using the EM Algorithm", *Journal of Time Series Analysis*, vol.3, pp.253-264, 1982.

[17] H. Tanizaki, *Nonlinear Filters: Estimation and Applications*, 2nd Ed., Springer, 1996.

[18] E.A. Wan and R. van der Merwe, "The Unscented Kalman Filter", in *Kalman Filtering and Neural Networks*, S.Haykin Ed., New York: John Wiley and Sons, 2001, pp.221-282.

[19] R. van der Merwe, "Sigma-point Kalman Filters and Probabilistic Inference in Dynamic State-Space Models", PhD Thesis, OGI School of Science and Engineering, Oregon Health and Science University, 2004.

[20] O. Zoeter, A. Ypma and T. Heskes, "Improved Unscented Kalman Smoothing for Stock Volatility Estimation", in *Machine Learning for Signal Processing, Proc. of the 14th IEEE Signal Proc. Society Workshop*, A.Barros et al. Eds., New Jersey: IEEE Press, 2004, pp.143-152.

Open Science Index, Mathematical and Computational Sciences Vol:3, No:7, 2009 publications.waset.org/9787.pdf