Technical Report CUED/F-INFENG/TR 380

Cambridge University Engineering Department

THE UNSCENTED PARTICLE FILTER

Rudolph van der Merwe⋆ Arnaud Doucet† Nando de Freitas‡ Eric Wan⋆

August 16, 2000

⋆ Oregon Graduate Institute
Electrical and Computer Engineering
20000 NW Walker Rd.,Beaverton, OR 97006, USA
{rvdmerwe,ericwan}@ece.ogi.edu

† Cambridge University
Engineering Department
Cambridge CB2 1PZ, England
ad2@eng.cam.ac.uk

‡ UC Berkeley, Computer Science
387 Soda Hall, Berkeley
CA 94720-1776 USA
jfgf@cs.berkeley.edu

Abstract

In this paper we propose a novel method for nonlinear, non-Gaussian, on-line estimation. The algorithm consists of a particle filter that uses an unscented Kalman filter (UKF) to generate the importance proposal distribution. The UKF allows the particle filter to incorporate the latest observations into a prior updating routine. In addition, the UKF generates proposal distributions that match the true posterior more closely and also has the capability of generating heavier tailed distributions than the well known extended Kalman filter. As a result, the convergence results predict that the new filter should outperform standard particle filters, extended Kalman filters and unscented Kalman filters. A few experiments confirm this prediction.

1 Introduction

Filtering is the problem of estimating the state of a system as a set of observations becomes available on-line. This problem is of paramount importance in many fields of science, engineering and finance. To solve it, one begins by modeling the evolution of the system and the noise in the measurements. The resulting models typically exhibit complex nonlinearities and non-Gaussian distributions, thus precluding analytical solution.

The best known algorithm to solve the problem of non-Gaussian, nonlinear filtering (filtering for short) is the extended Kalman filter (Anderson and Moore 1979)[1]. This filter is based upon the principle of linearizing the measurements and evolution models using Taylor series expansions. The series approximations in the EKF algorithm can, however, lead to poor representations of the nonlinear functions and probability distributions of interest. As as result, this filter can diverge.

Recently, Julier and Uhlmann (Julier and Uhlmann 1996, Julier and Uhlmann 1997b) have introduced a filter founded on the intuition that it is easier to approximate a Gaussian distribution than it is to approximate arbitrary nonlinear functions. They named this filter the unscented Kalman filter (UKF). They have shown that the UKF leads to more accurate results than the EKF and that in particular it generates much better estimates of the covariance of the states (the EKF often seems to underestimate this quantity). Wan and van der Merwe (Wan, van der Merwe and Nelson 2000, Wan and van der Merwe 2000) extended the use of the UKF to parameter estimation as well as dual estimation[2]. They reported a significant improvement in performance over that which is achieved by using an EKF for the same problem. The UKF has, however, the limitation that it does not apply to general non-Gaussian distributions.

Another popular solution strategy for the general filtering problem is to use sequential Monte Carlo methods, also known as particle filters: see for example (Doucet 1998, Doucet, de Freitas and Gordon 2000, Gordon, Salmond and Smith 1993). These methods allow for a complete representation of the posterior distribution of the states, so that any statistical estimates, such as the mean, modes, kurtosis and variance, can be easily computed. They can therefore, deal with any nonlinearities or distributions.

[1] We should point out that there are many other finite dimensional filters for specialized cases, including the HMM filter for discrete state-spaces, filters for counting observations (Smith and Miller 1986), filters for dynamic models with a time-varying, unknown process noise covariance matrix (West and Harrison 1997) and filters applicable to classes of the exponential family state-space models (Vidoni 1999). We, however, restrict the presentation to the most popular and general filters for continuous state-spaces.

[2] Dual estimation is the problem of simultaneously estimating the state of a system as well as the model parameters that define the dynamics of the system.

Particle filters rely on importance sampling and, as a result, require the design of proposal distributions that can approximate the posterior distribution reasonably well. In general, it is hard to design such proposals. The most common strategy is to sample from the probabilistic model of the states evolution (transition prior). This strategy can, however, fail if the new measurements appear in the tail of the prior or if the likelihood is too peaked in comparison to the prior. This situation does indeed arise in several areas of engineering and finance, where one can encounter sensors that are very accurate (peaked likelihoods) or data that undergoes sudden changes (non-stationarities): see for example (Pitt and Shephard 1999, Thrun 2000). To overcome this problem, several techniques based on linearization have been proposed in the literature (de Freitas 1999, de Freitas, Niranjan, Gee and Doucet 2000, Doucet 1998, Pitt and Shephard 1999). For example, in (de Freitas et al. 2000), the EKF Gaussian approximation is used as the proposal distribution for a particle filter. In this paper, we follow the same approach, but replace the EKF proposal by a UKF proposal. The resulting filter should perform better not only because the UKF is more accurate, but because it also allows one to control the rate at which the tails of the proposal distribution go to zero. That is, the UKF can be used to generate proposal distributions with larger high order moments and with means that are close to the true mean of the target distribution.

The last remark is the crux of our approach. *We will show theoretically and empirically that particle filters with a proposal distribution obtained using the UKF outperform other existing filters.* For comparison purposes, we will also present particle filters that use the EKF to generate the proposal distribution.

The remainder of this paper is organized as follows. Section 2 introduces the notation and the general state-space model formulation. Section 3 introduces the EKF and UKF, while sections 4 and 5 are devoted to the theory and implementation details of particle filters. After discussing the shortcomings of standard particles in Section 6, we propose the new unscented particle filter. Section 7 treats the convergence aspects of this filter. Some experimental results are discussed in Section 8. Finally, Section 9 contains some concluding remarks and pointers for future research.

2 Dynamic State-Space Model

The general state-space model (neglecting control inputs for the sake of clarity) can be broken down into a state transition and state measurement model

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{1}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^{n_x}$ denotes the states (hidden variables or parameters) of the system at time t and $\mathbf{y}_t \in \mathbb{R}^{n_y}$ the observations. The states follow a first order Markov process and the observations are assumed to be independent given the states. For example, if we are interested in nonlinear, non-Gaussian regression, the model can be expressed as follows

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}, \mathbf{v}_{t-1}) \tag{3}$$

$$\mathbf{y}_t = \mathbf{h}(\mathbf{u}_t, \mathbf{x}_t, \mathbf{n}_t) \tag{4}$$

where, in this case, $\mathbf{y}_t \in \mathbb{R}^{n_y}$ denotes the output observations, $\mathbf{u}_t \in \mathbb{R}^{n_u}$ the input observations, $\mathbf{x}_t \in \mathbb{R}^{n_x}$ the state of the system, $\mathbf{v}_t \in \mathbb{R}^{n_v}$ the process noise and $\mathbf{n}_t \in \mathbb{R}^{n_n}$ the measurement noise. The mappings $\mathbf{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \mapsto \mathbb{R}^{n_x}$ and $\mathbf{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_n} \mapsto \mathbb{R}^{n_y}$ represent the deterministic process and measurement models. To complete the specification of the model, the prior distribution (at $t = 0$) is denoted by $p(\mathbf{x}_0)$.

The posterior density $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$, where $\mathbf{x}_{0:t} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t\}$ and $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\}$, constitutes the complete solution to the sequential estimation problem. In many applications, such as tracking, it is of interest to estimate one of its marginals, namely the filtering density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. By computing the filtering density recursively, we do not need to keep track of the complete history of the states. Thus, from a storage point of view, the filtering density is more parsimonious than the full posterior density function. If we know the filtering density, we can easily derive various estimates of the system's states including means, modes, medians and confidence intervals. This will be our goal.

3 The EKF and Unscented Kalman Filters

In this section, we shall present the EKF and unscented filters, which provide Gaussian approximations to $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. These algorithms will be incorporated into the particle filtering framework in Section 6.

3.1 The Extended Kalman Filter

The EKF is a minimum mean-square-error (MMSE) estimator based on the Taylor series expansion of the nonlinear functions $\mathbf{f}$ and $\mathbf{h}$ around the estimates $\bar{\mathbf{x}}_{t|t-1}$ of the states $\mathbf{x}_t$

(Anderson and Moore 1979). For example

$$\mathbf{f}(\mathbf{x}_t) = \mathbf{f}(\bar{\mathbf{x}}_{t|t-1}) + \frac{\partial \mathbf{f}(\mathbf{x}_t)}{\partial \mathbf{x}_t}\Big|_{(\mathbf{x}_t=\bar{\mathbf{x}}_{t|t-1})}(\mathbf{x}_t - \bar{\mathbf{x}}_{t|t-1}) + \cdots$$

Using only the linear expansion terms, it is easy to derive the following update equations for the mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}$ of the Gaussian approximation to the posterior distribution of the states

$$\begin{aligned}
\bar{\mathbf{x}}_{t|t-1} &= \mathbf{f}(\bar{\mathbf{x}}_{t-1}, 0) \\
\mathbf{P}_{t|t-1} &= \mathbf{F}_t \mathbf{P}_{t-1} \mathbf{F}_t^T + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}_t^T \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1} \mathbf{H}_t^T [\mathbf{U}_t \mathbf{R}_t \mathbf{U}_t^T + \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T]^{-1} \\
\bar{\mathbf{x}}_t &= \bar{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{h}(\bar{\mathbf{x}}_{t|t-1}, 0)\right) \\
\mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}
\end{aligned} \tag{5}$$

where $\mathbf{K}_t$ is known as the Kalman gain, $\mathbf{Q}$ is the variance of the process noise (assumed to be zero-mean Gaussian), $\mathbf{R}$ is the variance of the measurement noise (also assumed to be zero-mean Gaussian), $\mathbf{F}_t \triangleq \frac{\partial \mathbf{f}(\mathbf{x}_t)}{\partial \mathbf{x}_t}\Big|_{(\mathbf{x}_t=\bar{\mathbf{x}}_{t|t-1})}$ and $\mathbf{G}_t \triangleq \frac{\partial \mathbf{f}(\mathbf{v}_t)}{\partial \mathbf{v}_t}\Big|_{(\mathbf{v}_t=\bar{\mathbf{v}})}$ are the Jacobians of the process model and $\mathbf{H}_t \triangleq \frac{\partial \mathbf{h}(\mathbf{x}_t)}{\partial \mathbf{x}_t}\Big|_{(\mathbf{x}_t=\bar{\mathbf{x}}_{t|t-1})}$ and $\mathbf{U}_t \triangleq \frac{\partial \mathbf{h}(\mathbf{n}_t)}{\partial \mathbf{n}_t}\Big|_{(\mathbf{n}_t=\bar{\mathbf{n}})}$ are the Jacobians of the measurements model,

3.2 The Unscented Kalman Filter

The unscented Kalman filter (UKF) is a recursive MMSE estimator that addresses some of the approximation issues of the EKF (Julier and Uhlmann 1997b). Because the EKF only uses the first order terms of the Taylor series expansion of the nonlinear functions, it often introduces large errors in the estimated statistics of the posterior distributions of the states. This is especially evident when the models are highly nonlinear and the local linearity assumption breaks down, i.e., the effects of the higher order terms of the Taylor series expansion becomes significant. Unlike the EKF, the UKF does not approximate the non-linear process and observation models, it uses the *true* nonlinear models and rather approximates the distribution of the state random variable. In the UKF the state distribution is still represented by a Gaussian random variable (GRV), but it is specified using a minimal set of deterministically chosen sample points. These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true nonlinear system, captures the posterior mean and covariance accurately to the 2nd order for any nonlinearity, with errors only introduced in the 3rd and higher orders. To elaborate on this, we start by first explaining the *unscented transformation*. After this the *scaled unscented transformation* (SUT) is introduced and discussed. The scaled unscented transformation is a generalizing extension of the unscented transformation and forms the algorithmic core of the unscented Kalman filter.

3.2.1 The unscented transformation

The unscented transformation (UT) is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation and builds on the principle that it is easier to approximate a probability distribution than an arbitrary nonlinear function (Julier and Uhlmann 1996). Consider propagating a n_x dimensional random variable $\mathbf{x}$ through an arbitrary nonlinear function $\mathbf{g} : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$ to generate $\mathbf{y}$,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \tag{6}$$

Assume $\mathbf{x}$ has mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}_x$. To calculate the statistics (first two moments) of $\mathbf{y}$ using the UT, we proceed as follows: First, a set of $2n_x + 1$ weighted samples or *sigma points* $\mathcal{S}_i = \{W_i, \boldsymbol{\mathcal{X}}_i\}$ are deterministically chosen so that they completely capture the true mean and covariance of the prior random variable $\mathbf{x}$. A selection scheme that satisfies this requirement is

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_0 &= \bar{\mathbf{x}} & W_0 &= \kappa/(n_x + \kappa) & i &= 0 \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}} + \left(\sqrt{(n_x + \kappa)\mathbf{P}_x}\right)_i & W_i &= 1/\{2(n_x + \kappa)\} & i &= 1, \dots, n_x \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}} - \left(\sqrt{(n_x + \kappa)\mathbf{P}_x}\right)_i & W_i &= 1/\{2(n_x + \kappa)\} & i &= n_x + 1, \dots, 2n_x
\end{aligned} \tag{7}$$

where κ is a scaling parameter and $\left(\sqrt{(n_x + \kappa)\mathbf{P}_x}\right)_i$ is the ith row or column of the matrix square root of $(n_x + \kappa)\mathbf{P}_x$. W_i is the weight associated with the ith point such that $\sum_{i=0}^{2n_x} W_i = 1$. Each sigma point is now propagated through the nonlinear function

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{g}(\boldsymbol{\mathcal{X}}_i) \qquad i = 0, \dots, 2n_x \tag{8}$$

and the estimated mean and covariance of $\mathbf{y}$ are computed as follows

$$\bar{\mathbf{y}} = \sum_{i=0}^{2n_x} W_i \boldsymbol{\mathcal{Y}}_i \tag{9}$$

$$\mathbf{P}_y = \sum_{i=0}^{2n_x} W_i \left(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right)\left(\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right)^T \quad . \tag{10}$$

These estimates of the mean and covariance are accurate to the second order (third order for Gaussian priors) of the Taylor series expansion of $\mathbf{g}(\mathbf{x})$ for *any* nonlinear function. Errors are introduced in the third and higher order moments but are scaled by the choice of the parameter κ. In comparison, the EKF only calculates the posterior mean and covariance accurately to the first order with all higher order moments truncated. For a detailed proof

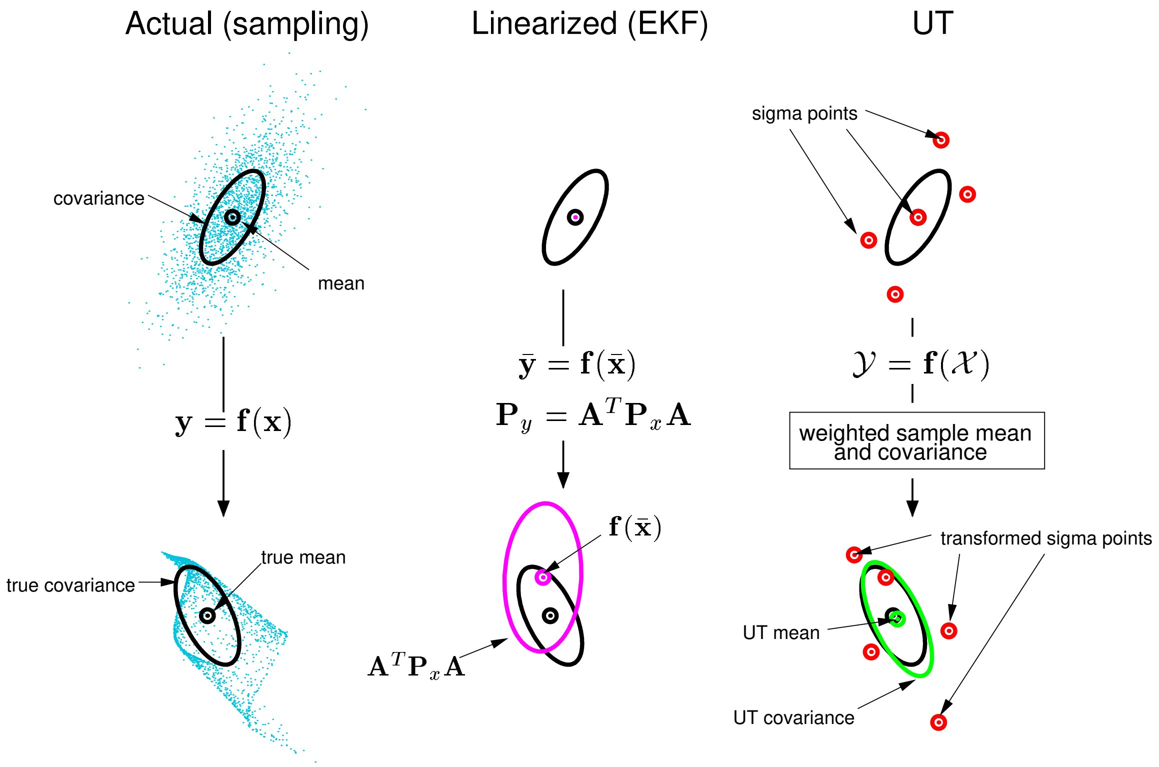

Figure 1: Schematic diagram of the Unscented Transformation: A cloud of 5000 samples drawn from a Gaussian prior is propagated through an arbitrary highly nonlinear function and the true posterior sample mean and covariance are calculated. This reflects the truth as calculated by a Monte Carlo approach and is shown in the left plot. Next, the posterior random variable's statistics are calculated by a linearization approach as used in the EKF. The middle plot shows these results. The errors in both the mean and covariance as calculated by this "first-order" approximation is clearly visible. The right plot shows the results of the estimates calculated by the unscented transformation. There is almost no bias error in the estimate of the mean and the estimated covariance is also much closer to the true covariance. The superior performance of the UT is clearly evident.

of this, see (Julier and Uhlmann 1996). A comparison of the performance of the UT versus that of the linearization approach used in the EKF is shown in Figure 1.

The sigma point selection scheme used in the UT has the property that as the dimension of the state-space increases, the radius of the sphere that bounds all the sigma points increases as well. Even though the mean and covariance of the prior distribution are still captured correctly, it does so at the cost of sampling non-local effects. If the nonlinearities in question are very severe, this can lead to significant difficulties. In order to address this problem, the sigma points can be scaled towards or away from the mean of the prior distribution by a proper choice of κ. The distance of the ith sigma point from $\bar{\mathbf{x}}$, $|\boldsymbol{\mathcal{X}}_i - \bar{\mathbf{x}}|$, is proportional to $\sqrt{(n_x + \kappa)}$. When $\kappa = 0$, the distance is proportional to $\sqrt{n_x}$. When $\kappa > 0$ the points are scaled further from $\bar{\mathbf{x}}$ and when $\kappa < 0$ the points are scaled *towards* $\bar{\mathbf{x}}$. For the special case of $\kappa = 3 - n_x$, the desired dimensional scaling invariance is achieved by canceling the effect of n_x. However, when $\kappa = 3 - n_x < 0$ the weight $W_0 < 0$ and the calculated covariance can be *non-positive semidefinite*. The *scaled unscented transformation* was developed to address this problem (Julier 2000).

3.2.2 The scaled unscented transformation

The scaled unscented transformation (SUT) replaces the original set of sigma points with a transformed set given by

$$\boldsymbol{\mathcal{X}}_i^{'} = \boldsymbol{\mathcal{X}}_0 + \alpha(\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_0) \qquad i = 0 \ldots 2n_x, \tag{11}$$

where α is a positive scaling parameter which can be made arbitrarily small to minimize higher order effects. This formulation gives an extra degree of freedom to control the scaling of the sigma points without causing the resulting covariance to possibly become non-positive semidefinite. This is achieved by applying the UT to an *auxiliary random variable* propagation problem which is related to the original nonlinear model of equation (6) by

$$\mathbf{z} = \mathbf{g}'(\mathbf{x}) = \frac{\mathbf{g}\left[\bar{\mathbf{x}} + \alpha(\mathbf{x} - \bar{\mathbf{x}})\right] - \mathbf{g}(\bar{\mathbf{x}})}{\alpha^2} + \mathbf{g}(\bar{\mathbf{x}}). \tag{12}$$

The Taylor series expansion of $\bar{\mathbf{z}}$ and $\mathbf{P}_z$ agrees with that of $\bar{\mathbf{y}}$ and $\mathbf{P}_y$ exactly up to the second order, with the higher order terms scaling geometrically with a common ratio of α. The same second order accuracy of the normal UT is thus retained with a controllable scaling of the higher order errors by a proper choice of α. The auxiliary random variable formulation of the SUT is identical to applying the original UT on a *pre-scaled* set of sigma points (Julier 2000). A set of sigma points $\mathcal{S} = \{\mathbf{W}, \boldsymbol{\mathcal{X}}\}$ is calculated using equation (7) and then transformed into the scaled set $\boldsymbol{\mathcal{S}}^{'} = \{\mathbf{W}^{'}, \boldsymbol{\mathcal{X}}^{'}\}$ by

$$\begin{aligned} \boldsymbol{\mathcal{X}}_i^{'} &= \boldsymbol{\mathcal{X}}_0 + \alpha(\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_0) \\ W_i^{'} &= \begin{cases} W_0/\alpha^2 + (1 - 1/\alpha^2) & i = 0 \\ W_i/\alpha^2 & i \neq 0 \end{cases} \end{aligned} \tag{13}$$

where α is the new sigma point *scaling* parameter. The sigma point selection and scaling can also be combined into a single step (thereby reducing the number of calculations) by setting

$$\lambda = \alpha^2(n_x + \kappa) - n_x \tag{14}$$

and selecting the sigma point set by:

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_0 &= \bar{\mathbf{x}} \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}} + \left(\sqrt{(n_x+\lambda)\mathbf{P}_x}\right)_i & i &= 1,\dots,n_x \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}} - \left(\sqrt{(n_x+\lambda)\mathbf{P}_x}\right)_i & i &= n_x+1,\dots,2n_x \\
W_0^{(m)} &= \lambda/(n_x+\lambda) \\
W_0^{(c)} &= \lambda/(n_x+\lambda) + (1-\alpha^2+\beta) \\
W_i^{(m)} &= W_i^{(c)} = 1/\{2(n_x+\lambda)\} & i &= 1,\dots,2n_x
\end{aligned} \tag{15}$$

The weighting on the zeroth sigma point directly affects the magnitude of the errors in the fourth and higher order terms for symmetric prior distributions (Julier 2000). A third parameter, β, is thus introduced which affects the weighting of the zeroth sigma point for the calculation of the covariance. This allows for the minimization of higher order errors if prior knowledge (i.e. kurtosis, etc.) of the distribution of $\mathbf{x}$ is available.

The complete scaled unscented transformation is thus given by the following:

1. Choose the parameters κ, α and β. Choose $\kappa \geq 0$ to guarantee positive semi-definiteness of the covariance matrix. The specific value of kappa is not critical though, so a good default choice is $\kappa = 0$. Choose $0 \leq \alpha \leq 1$ and $\beta \geq 0$. α controls the "size" of the sigma point distribution and should ideally be a small number to avoid sampling non-local effects when the nonlinearities are strong. β is a non-negative weighting term which can be used to incorporate knowledge of the higher order moments of the distribution. For a Gaussian prior the optimal choice is $\beta = 2$. This parameter can also be used to control the error in the kurtosis which affects the 'heaviness' of the tails of the posterior distribution.

2. Calculate the set of $2n_x + 1$ scaled sigma points and weights $\boldsymbol{\mathcal{S}} = \{\mathbf{W}, \boldsymbol{\mathcal{X}}\}$ by setting $\lambda = \alpha^2(n_x + \kappa) - n_x$ and using the combined selection/scaling scheme of equation (15). As mentioned earlier, n_x is the dimension of $\mathbf{x}$.

3. Propagate each sigma point through the nonlinear transformation

$$\boldsymbol{\mathcal{Y}}_i = \mathbf{g}\left(\boldsymbol{\mathcal{X}}_i\right) \qquad i = 0,\dots,2n_x$$

4. The mean $\bar{\mathbf{y}}$ and covariance $\mathbf{P}_y$ are computed as follows

$$\begin{aligned}
\bar{\mathbf{y}} &= \sum_{i=0}^{2n_x} W_i^{(m)} \boldsymbol{\mathcal{Y}}_i \\
\mathbf{P}_y &= \sum_{i=0}^{2n_x} W_i^{(c)} \left\{\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right\}\left\{\boldsymbol{\mathcal{Y}}_i - \bar{\mathbf{y}}\right\}^T
\end{aligned}$$

3.2.3 Implementing the Unscented Kalman Filter

The Unscented Kalman Filter (UKF) is a straightforward application of the scaled unscented transformation to recursive minimum mean-square-error (RMMSE) estimation (Julier and Uhlmann 1997b), where the state random variable (RV) is redefined as the concatenation of the original state and noise variables: $\mathbf{x}_t^a = [\mathbf{x}_t^T \mathbf{v}_t^T \mathbf{n}_t^T]^T$. The SUT sigma point selection scheme is applied to this new augmented state RV to calculate the corresponding sigma matrix, $\boldsymbol{\mathcal{X}}_t^a$. The complete UKF algorithm that updates the mean $\bar{\mathbf{x}}$ and covariance $\mathbf{P}$ of the Gaussian approximation to the posterior distribution of the states is given by:

1. Initialize with:

$$\begin{aligned}
\bar{\mathbf{x}}_0 &= E[\mathbf{x}_0] \\
\mathbf{P}_0 &= E[(\mathbf{x}_0 - \bar{\mathbf{x}}_0)(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T] \\
\bar{\mathbf{x}}_0^a &= E[\mathbf{x}^a] = [\bar{\mathbf{x}}_0^T \ \mathbf{0} \ \mathbf{0}]^T
\end{aligned}$$

$$\mathbf{P}_0^a = E[(\mathbf{x}_0^a - \bar{\mathbf{x}}_0^a)(\mathbf{x}_0^a - \bar{\mathbf{x}}_0^a)^T] = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}$$

2. For $t \in \{1, \dots, \infty\}$,

   (a) Calculate sigma points:

$$\boldsymbol{\mathcal{X}}_{t-1}^a = \left[\bar{\mathbf{x}}_{t-1}^a \ \ \bar{\mathbf{x}}_{t-1}^a \pm \sqrt{(n_a + \lambda)\mathbf{P}_{t-1}^a}\right]$$

   (b) Time update:

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_{t|t-1}^x &= \mathbf{f}\left(\boldsymbol{\mathcal{X}}_{t-1}^x, \boldsymbol{\mathcal{X}}_{t-1}^v\right) \\
\bar{\mathbf{x}}_{t|t-1} &= \sum_{i=0}^{2n_a} W_i^{(m)} \mathcal{X}_{i,t|t-1}^x \\
\mathbf{P}_{t|t-1} &= \sum_{i=0}^{2n_a} W_i^{(c)} [\mathcal{X}_{i,t|t-1}^x - \bar{\mathbf{x}}_{t|t-1}][\mathcal{X}_{i,t|t-1}^x - \bar{\mathbf{x}}_{t|t-1}]^T \\
\boldsymbol{\mathcal{Y}}_{t|t-1} &= \mathbf{h}\left(\boldsymbol{\mathcal{X}}_{t|t-1}^x, \boldsymbol{\mathcal{X}}_{t-1}^n\right) \\
\bar{\mathbf{y}}_{t|t-1} &= \sum_{i=0}^{2n_a} W_i^{(m)} \mathcal{Y}_{i,t|t-1}
\end{aligned}$$

(c) Measurement update equations:

$$\begin{aligned}
\mathbf{P}_{\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t} &= \sum_{i=0}^{2n_a} W_i^{(c)}[\mathcal{Y}_{i,t|t-1} - \bar{\mathbf{y}}_{t|t-1}][\mathcal{Y}_{i,t|t-1} - \bar{\mathbf{y}}_{t|t-1}]^T \\
\mathbf{P}_{\mathbf{x}_t\mathbf{y}_t} &= \sum_{i=0}^{2n_a} W_i^{(c)}[\mathcal{X}_{i,t|t-1} - \bar{\mathbf{x}}_{t|t-1}][\mathcal{Y}_{i,t|t-1} - \bar{\mathbf{y}}_{t|t-1}]^T \\
\mathbf{K}_t &= \mathbf{P}_{\mathbf{x}_t\mathbf{y}_t}\mathbf{P}^{-1}_{\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t} \\
\bar{\mathbf{x}}_t &= \bar{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \bar{\mathbf{y}}_{t|t-1}) \\
\mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{K}_t\mathbf{P}_{\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_t}\mathbf{K}_t^T
\end{aligned}$$

where, $\mathbf{x}^a = [\mathbf{x}^T\ \mathbf{v}^T\ \mathbf{n}^T]^T$, $\boldsymbol{\mathcal{X}}^a = [(\boldsymbol{\mathcal{X}}^x)^T\ (\boldsymbol{\mathcal{X}}^v)^T\ (\boldsymbol{\mathcal{X}}^n)^T]^T$, λ=composite scaling parameter, $n_a = n_x + n_v + n_n$, $\mathbf{Q}$=process noise cov., $\mathbf{R}$= measurement noise cov., $\mathbf{K}$=Kalman gain,W_i=weights as calculated in Eqn. 15.

Note that no explicit calculation of Jacobians or Hessians are necessary to implement this algorithm. The UKF requires computation of a matrix square root which can be implemented directly using a Cholesky factorization in order $n_x^3/6$. However, the covariance matrices can be expressed recursively, and thus the square-root can be computed in order n_x^2 by performing a recursive update to the Cholesky factorization. So, not only does the UKF outperform the EKF in accuracy and robustness, it does so at no extra computational cost. The superior performance of the UKF over that of the EKF have been reported in numerous publications including (Wan et al. 2000, Wan and van der Merwe 2000, Chong and Kleeman 1997, Julier and Uhlmann 1997b, Julier and Uhlmann 1997a, Clark 1999).

This is the most general form of the unscented Kalman filter. For the special (but often found) case where the process and measurement noise are purely additive, the computational complexity of the UKF can be reduced. In such a case, the system state need not be augmented with the noise RV's. This reduces the dimension of the sigma points as well as the total number of sigma points used. The covariances of the noise sources are then incorporated into the state covariance using a simple additive procedure. For more details, see (Julier and Uhlmann 1997b).

4 Particle Filtering

We have so far presented two nonlinear filtering strategies that rely on Gaussian approximation. In this section, we shall present a filtering method (particle filtering) that does not require this assumption. However, it has other problems as we will point out in Section 6. In that section, we will show that it is possible to overcome some of the problems inherent to particle filters by combining them with the EKF and UKF strategies in a theoretically valid setting.

In recent years, many researchers in the statistical and signal processing communities have, almost simultaneously, proposed several variations of particle filtering algorithms. In recent years, many researchers in the statistical and signal processing communities have, almost simultaneously, proposed several variations of particle filtering algorithms. As pointed out in (Liu, Chen and Logvinenko 2000), basic sequential Monte Carlo methods, based on sequential importance sampling, had already been introduced in the physics and statistics literature in the fifties! (Hammersley and Morton 1954, Rosenbluth and Rosenbluth 1955). These methods were also introduced in the automatic control field in the late sixties (Doucet 1998, Handschin and Mayne 1969). In the seventies, various researchers continued working on these ideas (Akashi and Kumamoto 1977, Handschin 1970, Zaritskii, Svetnik and Shimelevich 1975). However, all these earlier implementations were based on plain sequential importance sampling, which, as we shall see later, degenerates with time. The major contribution towards allowing this class of algorithm to be of any practical use was the inclusion of a resampling stage in the early nineties (Gordon et al. 1993). Since then many new improvements have been proposed (Doucet et al. 2000).

Before presenting particle filtering algorithms, we need to review perfect Monte Carlo simulation and importance sampling. This will allow us to present particle filters in a very general setting.

4.1 Perfect Monte Carlo Simulation

In Monte Carlo simulation, a set of weighted particles (samples), drawn from the posterior distribution, is used to map integrals to discrete sums. More precisely, the posterior can be approximated by the following empirical estimate

$$\widehat{p}(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\mathbf{x}_{0:t}^{(i)}}(\mathrm{d}\mathbf{x}_{0:t})$$

where the random samples $\{\mathbf{x}_{0:t}^{(i)}; i = 1, \ldots, N\}$, are drawn from the posterior distribution and $\delta(\mathrm{d}\cdot)$ denotes the Dirac delta function. Consequently, any expectations of the form

$$\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big) = \int \mathbf{g}_t(\mathbf{x}_{0:t})p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})\mathrm{d}\mathbf{x}_{0:t}$$

may be approximated by the following estimate

$$\overline{\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_t(\mathbf{x}_{0:t}^{(i)})$$

where the particles $\mathbf{x}_{0:t}^{(i)}$ are assumed to be independent and identically distributed (i.i.d.) for the approximation to hold. According to the law of large numbers, we have $\overline{\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)} \xrightarrow[N\to\infty]{a.s.} \mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)$, where $\xrightarrow[N\to\infty]{a.s.}$ denotes almost surely convergence. Moreover, if the posterior variance of $\mathbf{g}_t(\mathbf{x}_{0:t})$ is bounded, that is $var_{p(\cdot|\mathbf{y}_{1:t})}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big) < \infty$, then the following central limit theorem holds

$$\sqrt{N}\left(\overline{\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)} - \mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)\right) \underset{N\to\infty}{\Longrightarrow} \mathcal{N}\left(0, var_{p(\cdot|\mathbf{y}_{1:t})}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)\right)$$

where $\underset{N\to\infty}{\Longrightarrow}$ denotes convergence in distribution.

4.2 Bayesian Importance Sampling

As mentioned in the previous section, one can approximate the posterior distribution with a function on a finite discrete support. Consequently, it follows from the strong law of large numbers that as the number or samples N increases, expectations can be mapped into sums. Unfortunately, it is often impossible to sample directly from the posterior density function. However, we can circumvent this difficulty by sampling from a known, easy-to-sample, proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ and making use of the following substitution

$$\begin{aligned}
\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big) &= \int \mathbf{g}_t(\mathbf{x}_{0:t}) \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t} \\
&= \int \mathbf{g}_t(\mathbf{x}_{0:t}) \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t} \\
&= \int \mathbf{g}_t(\mathbf{x}_{0:t}) \frac{w_t(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t}
\end{aligned}$$

where the variables $w_t(\mathbf{x}_{0:t})$ are known as the unnormalized importance weights

$$w_t = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \tag{16}$$

We can get rid of the unknown normalizing density $p(\mathbf{y}_{1:t})$ as follows

$$\begin{aligned}
\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big) &= \frac{1}{p(\mathbf{y}_{1:t})} \int \mathbf{g}_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t} \\
&= \frac{\int \mathbf{g}_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t}}{\int p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}) \frac{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \mathrm{d}\mathbf{x}_{0:t}} \\
&= \frac{\int \mathbf{g}_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t}}{\int w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \mathrm{d}\mathbf{x}_{0:t}} \\
&= \frac{\mathbb{E}_{q(\cdot|\mathbf{y}_{1:t})}\big(w_t(\mathbf{x}_{0:t}) \mathbf{g}_t(\mathbf{x}_{0:t})\big)}{\mathbb{E}_{q(\cdot|\mathbf{y}_{1:t})}\big(w_t(\mathbf{x}_{0:t})\big)}
\end{aligned}$$

where the notation $\mathbb{E}_{q(\cdot|\mathbf{y}_{1:t})}$ has been used to emphasize that the expectations are taken over the proposal distribution $q(\cdot|\mathbf{y}_{1:t})$. Hence, by drawing samples from the proposal function

$q(\cdot|\mathbf{y}_{1:t})$, we can approximate the expectations of interest by the following estimate

$$\begin{aligned}\overline{\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)} &= \frac{1/N\sum_{i=1}^{N}\mathbf{g}_t(\mathbf{x}_{0:t}^{(i)})w_t(\mathbf{x}_{0:t}^{(i)})}{1/N\sum_{i=1}^{N}w_t(\mathbf{x}_{0:t}^{(i)})} \\ &= \sum_{i=1}^{N}\mathbf{g}_t(\mathbf{x}_{0:t}^{(i)})\widetilde{w}_t(\mathbf{x}_{0:t}^{(i)}) \qquad (17)\end{aligned}$$

where the normalized importance weights $\widetilde{w}_t^{(i)}$ are given by

$$\widetilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}$$

The estimate of equation (17) is biased as it involves a ratio of estimates. However, it is possible to obtain asymptotic convergence and a central limit theorem for $\overline{\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)}$ under the following assumptions (Doucet 1998, Geweke 1989):

1. $\mathbf{x}_{0:t}^{(i)}$ corresponds to a set of i.i.d. samples drawn from the proposal distribution, the support of the proposal distribution includes the support of the posterior distribution and $\mathbb{E}\big(\mathbf{g}_t(\mathbf{x}_{0:t})\big)$ exists and is finite.

2. The expectations of w_t and $w_t\mathbf{g}_t^2(\mathbf{x}_{0:t})$ over the posterior distribution exist and are finite.

A sufficient condition to verify the second assumption is to have bounds on the variance of $\mathbf{g}_t(\mathbf{x}_{0:t})$ and on the importance weights (Geweke 1989, Crisan and Doucet 2000). Thus, as N tends to infinity, the posterior density function can be approximated arbitrarily well by the point-mass estimate

$$\widehat{p}(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N}\widetilde{w}_t^{(i)}\delta_{\mathbf{x}_{0:t}^{(i)}}(\mathrm{d}\mathbf{x}_{0:t})$$

4.3 Sequential Importance Sampling

In order to compute a sequential estimate of the posterior distribution at time t without modifying the previously simulated states $\mathbf{x}_{0:t-1}$, proposal distributions of the following form can be used,

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}) \quad , \qquad (18)$$

Here we are making the assumption that the current state is not dependent on future observations, i.e., we're doing filtering and not smoothing. It needs to be emphasized that more general proposals, which modify previously simulated trajectories, might be necessary in some scenarios (Pitt and Shephard 1999). This issue is, however, beyond the scope of this paper. Under our assumptions that the states correspond to a Markov process and that the observations are conditionally independent given the states, we get

$$p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0)\prod_{j=1}^{t} p(\mathbf{x}_j|\mathbf{x}_{j-1}) \qquad \text{and} \qquad p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) = \prod_{j=1}^{t} p(\mathbf{y}_j|\mathbf{x}_j) \tag{19}$$

By substituting equations (18) and (19) into equation (16), a recursive estimate for the importance weights can be derived as follows

$$\begin{aligned} w_t &= \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t})} \\ &= w_{t-1}\frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t-1}|\mathbf{x}_{0:t-1})p(\mathbf{x}_{0:t-1})}\frac{1}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t})} \\ &= w_{t-1}\frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t})} \end{aligned} \tag{20}$$

Equation (20) provides a mechanism to sequentially update the importance weights, given an appropriate choice of proposal distribution, $q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t})$. The exact form of this distribution is a critical design issue and is usually approximated in order to facilitate easy sampling. The details of this is discussed in the next section. Since we can sample from the proposal distribution and evaluate the likelihood and transition probabilities, all we need to do is generate a prior set of samples and iteratively compute the importance weights. This procedure, known as sequential importance sampling (SIS), allows us to obtain the type of estimates described by equation (17).

4.3.1 Choice of proposal distribution

The choice of proposal function is one of the most critical design issues in importance sampling algorithms and forms the main issue addressed in this paper. The preference for proposal functions that minimize the variance of the importance weights is advocated by (Doucet 1997). The following result has been proved:

Proposition 1 [Proposition 3 of (Doucet, Gordon and Krishnamurthy 1999)] *The proposal distribution* $q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}) = p(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t})$ *minimizes the variance of the importance weights conditional on* $\mathbf{x}_{0:t-1}$ *and* $\mathbf{y}_{1:t}$.

This choice of proposal distribution has also been advocated by other researchers (Kong, Liu and Wong 1994, Liu and Chen 1995, Zaritskii et al. 1975). Nonetheless, the distribution

$$q(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}) \triangleq p(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{21}$$

(the transition prior) is the most popular choice [3] of proposal function (Avitzour 1995, Beadle and Djurić 1997, Gordon et al. 1993, Isard and Blake 1996, Kitagawa 1996). Although

[3] $A \triangleq B$ implies that we *choose* B to approximate A.

it results in higher Monte Carlo variation than the optimal proposal $p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$, as a result of it not incorporating the most recent observations, it is usually easier to implement (Berzuini, Best, Gilks and Larizza 1997, Doucet 1998, Liu and Chen 1998). The transition prior is defined in terms of the probabilistic model governing the states' evolution (3) and the process noise statistics. For example, if an additive Gaussian process noise model is used, the transition prior is simply,

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(f\left(\mathbf{x}_{t-1}, 0\right), Q_{t-1}\right) \quad . \tag{22}$$

As illustrated in Figure 2, if we fail to use the latest available information to propose new values for the states, only a few particles will have significant importance weights when their likelihood are evaluated. *It is therefore of paramount importance to move the particles towards the regions of high likelihood. This problem also arises when the likelihood function is too narrow compared to the prior.* In Sections 6 and 7, we shall describe several algorithms, based on linearization and the unscented transformation, to implement the optimal importance function.

4.3.2 Degeneracy of the SIS algorithm

The SIS algorithm discussed so far has a serious limitation: the variance of the importance weights increases stochastically over time. In order to show this we begin by expanding Equation (16),

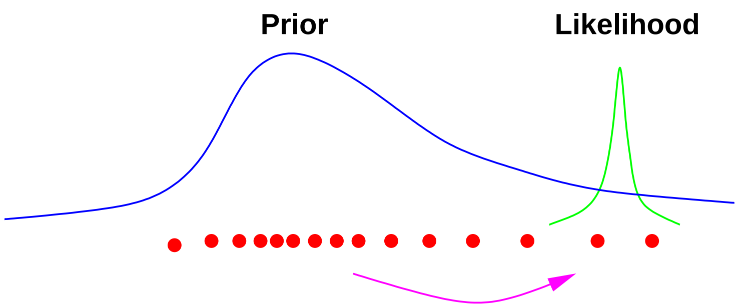

Figure 2: The optimal importance distribution allows us to move the samples in the prior to regions of high likelihood. This is of paramount importance if the likelihood happens to lie in one of the tails of the prior distribution, or if it is too narrow (low measurement error).

$$
\begin{aligned}
w_t &= \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\
&= \frac{p(\mathbf{y}_{1:t},\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\
&= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})p(\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\
&\propto \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}
\end{aligned} \tag{23}
$$

The ratio in the last line[4] of Equation (23) is called the *importance ratio* and it can be shown that its variance increases over time. For a proof of this, see (Kong et al. 1994) and (Doucet et al. 1999). We thus state (without proof):

Proposition 2 [Page 285 of (Kong et al. 1994), proposition 4 of (Doucet et al. 1999)] *The unconditional variance (that is, when the observations are regarded as random) of the importance ratios increases over time.*

To understand why the variance increase poses a problem, suppose that we want to sample from the posterior. In that case, we want the proposal density to be very *close*[5] to the posterior density. When this happens, we obtain the following results for the mean and variance (see (Doucet 1997) for a proof)

$$
\mathbb{E}_{q(\cdot|\mathbf{y}_{1:t})}\left(\frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}\right) = 1
$$

and

$$
var_{q(\cdot|\mathbf{y}_{1:t})}\left(\frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}\right) = \mathbb{E}_{q(\cdot|\mathbf{y}_{1:t})}\left(\left(\frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} - 1\right)^2\right) = 0
$$

In other words, we want the variance to be close to zero in order to obtain reasonable estimates. Therefore, a variance increase has a harmful effect on the accuracy of the simulations. In practice, the degeneracy caused by the variance increase can be observed by monitoring the importance weights. Typically, what we observe is that, after a few iterations, one of the normalized importance weights tends to 1, while the remaining weights tend to zero. A large number of samples are thus effectively removed from the sample set because their importance weights become numerically insignificant. The next section presents a strategy to reduce this *degeneration* or depletion of samples.

[4]The proportionality in the last line of the equation follows from the fact that $p(\mathbf{y}_{1:t})$ is a constant.

[5]Closeness is defined over the full support of the true posterior. This implies that the best possible (but not practical) choice for the proposal is $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$

4.4 Selection

To avoid the degeneracy of the SIS simulation method, a selection (resampling) stage may be used to eliminate samples with low importance weights and multiply samples with high importance weights. It is possible to see an analogy to the steps in genetic algorithms (Higuchi 1997).

A selection scheme associates to each particle $\mathbf{x}_{0:t}^{(i)}$ a number of "children", say $N_i \in \mathbb{N}$, such that $\sum_{i=1}^{N} N_i = N$. Several selection schemes have been proposed in the literature. These schemes satisfy $\mathbb{E}(N_i) = N\widetilde{w}_t^{(i)}$ but their performance varies in terms of the variance of the particles $var(N_i)$. Results in (Kitagawa 1996) and (Crisan, Del Moral and Lyons 1999) indicate that the restriction $\mathbb{E}(N_i) = N\widetilde{w}_t^{(i)}$ is unnecessary to obtain convergence results. So it is possible to design biased but computationally inexpensive selection schemes.

We will now present a number of selection or resampling schemes, namely: *sampling-importance resampling (SIR)*, *residual resampling* and *minimum variance sampling*. We found that the specific choice of resampling scheme does not significantly affect the performance of the particle filter, so we used residual resampling in all of the experiments in Section 9.

4.4.1 Sampling-importance resampling (SIR) and multinomial sampling

Many of the ideas on resampling have stemmed from the work of Efron (Efron 1982), Rubin (Rubin 1988) and Smith and Gelfand (Smith and Gelfand 1992). Resampling involves mapping the Dirac random measure $\{\mathbf{x}_{0:t}^{(i)}, \widetilde{w}_t^{(i)}\}$ into an equally weighted random

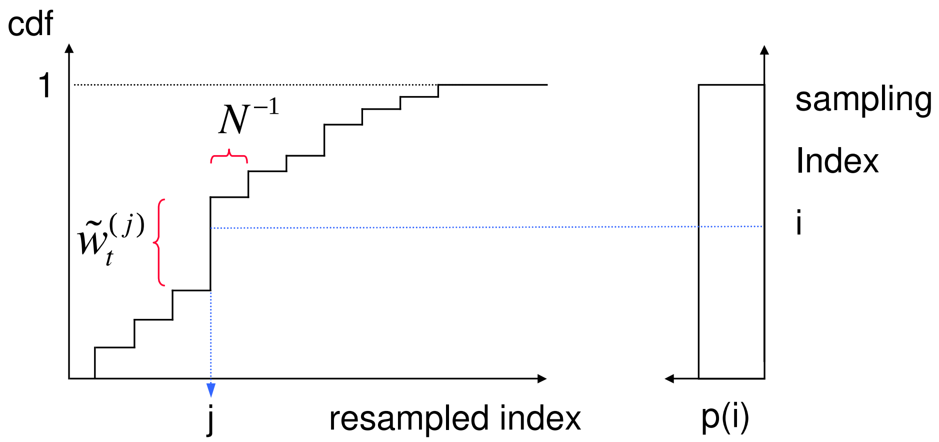

Figure 3: Resampling process, whereby a random measure $\{\mathbf{x}_{1:t}^{(i)}, \widetilde{w}_t^{(i)}\}$ is mapped into an equally weighted random measure $\{\mathbf{x}_{1:t}^{(j)}, N^{-1}\}$. The index i is drawn from a uniform distribution.

measure $\{\mathbf{x}_{0:t}^{(j)}, N^{-1}\}$. This can be accomplished by sampling uniformly from the discrete set $\{\mathbf{x}_{0:t}^{(i)}; i = 1, \ldots, N\}$ with probabilities $\{\widetilde{w}_t^{(i)}; i = 1, \ldots, N\}$ as proposed in the seminal paper of Gordon, Salmond and Smith (1993). A mathematical proof of this can be found on pages 111–112 of (Gordon 1994). Figure 3 shows a way of sampling from this discrete set. After constructing the cumulative distribution of the discrete set, a uniformly drawn sampling index i is projected onto the distribution range and then onto the distribution domain. The intersection with the domain constitutes the new sample index j. That is, the vector $\mathbf{x}_{0:t}^{(j)}$ is accepted as the new sample. Clearly, the vectors with the larger sampling weights will end up with more copies after the resampling process.

Sampling N times from the cumulative discrete distribution $\sum_{i=1}^{N} \widetilde{w}_t^{(i)} \delta_{\mathbf{x}_{0:t}^{(i)}}(\mathrm{d}\mathbf{x}_{0:t})$ is equivalent to drawing $(N_i; i = 1, \ldots, N)$ from a multinomial distribution with parameters N and $\widetilde{w}_t^{(i)}$. This procedure can be implemented in $\mathcal{O}(N)$ operations (Doucet 1998, Pitt and Shephard 1999) following the work of (Ripley 1987, pp. 96). As we are sampling from a multinomial distribution, the variance is $var(N_i) = N\widetilde{w}_t^{(i)}(1 - \widetilde{w}_t^{(i)})$. As pointed out in (Carpenter, Clifford and Fearnhead 1999) and (Liu and Chen 1998), it is possible to design selection schemes with lower variance.

4.4.2 Residual resampling

This procedure involves the following steps (Higuchi 1997, Liu and Chen 1998). Firstly, set $\widetilde{N}_i = \left\lfloor N\widetilde{w}_t^{(i)} \right\rfloor$. Secondly, perform an SIR procedure to select the remaining $\overline{N}_t = N - \sum_{i=1}^{N} \widetilde{N}_i$ samples with new weights $w_t'^{(i)} = \overline{N}_t^{-1}\left(\widetilde{w}_t^{(i)} N - \widetilde{N}_i\right)$. Finally, add the results to the current $\widetilde{N}_i$. For this scheme, the variance $\left(var(N_i) = \overline{N}_t w_t'^{(i)}(1 - w_t'^{(i)})\right)$ is smaller than the one given by the SIR scheme. Moreover, this procedure is computationally cheaper.

4.4.3 Minimum variance sampling

This strategy includes the stratified/systematic sampling procedures introduced in (Kitagawa 1996) and the Tree Based Branching Algorithm presented in (Crisan 2000). One samples a set of N points U in the interval $[0, 1]$, each of the points a distance N^{-1} apart. The number of children N_i is taken to be the number of points that lie between $\sum_{j=1}^{i-1} \widetilde{w}_t^{(j)}$ and $\sum_{j=1}^{i} \widetilde{w}_t^{(j)}$. This strategy introduces a variance on N_i even smaller than the residual resampling scheme, namely $var(N_i) = \overline{N}_t w_t'^{(i)}(1 - \overline{N}_t w_t'^{(i)})$. Its computational complexity is $\mathcal{O}(N)$.

5 The Particle Filter Algorithm

We have so far explained how to compute the importance weights sequentially and how to improve the sample set by resampling. The pseudo-code of a generic particle filter can now be presented.

Generic Particle Filter

1. Initialization: $t = 0$
   - For $i = 1, \ldots, N$, draw the states $\mathbf{x}_0^{(i)}$ from the prior $p(\mathbf{x}_0)$.

2. For $t = 1, 2, \ldots$

   (a) Importance sampling step

   - For $i = 1, \ldots, N$, sample $\widehat{\mathbf{x}}_t^{(i)} \sim q(\mathbf{x}_t | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ and set $\widehat{\mathbf{x}}_{0:t}^{(i)} \triangleq \left(\mathbf{x}_{0:t-1}^{(i)}, \widehat{\mathbf{x}}_t^{(i)}\right)$
   - For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p(\mathbf{y}_t | \widehat{\mathbf{x}}_t^{(i)}) p(\widehat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\widehat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \tag{24}$$

   - For $i = 1, \ldots, N$, normalize the importance weights:

$$\widetilde{w}_t^{(i)} = w_t^{(i)} \left[\sum_{j=1}^{N} w_t^{(j)}\right]^{-1}$$

   (b) Selection step (resampling)

   - Multiply/Suppress samples $\widehat{\mathbf{x}}_{0:t}^{(i)}$ with high/low importance weights $\widetilde{w}_t^{(i)}$, respectively, to obtain N random samples $\mathbf{x}_{0:t}^{(i)}$ approximately distributed according to $p(\mathbf{x}_{0:t}^{(i)} | \mathbf{y}_{1:t})$.
   - For $i = 1, \ldots, N$, set $w_t^{(i)} = \widetilde{w}_t^{(i)} = \frac{1}{N}$

   (c) Output: The output of the algorithm is a set of samples that can be used to approximate the posterior distribution as follows

$$p\left(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}\right) \approx \widehat{p}\left(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}\right) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\left(\mathbf{x}_{0:t}^{(i)}\right)}\left(d\mathbf{x}_{0:t}\right)$$

   One obtains straightforwardly the following estimate of $\mathbb{E}\left(\mathbf{g}_t\left(\mathbf{x}_{0:t}\right)\right)$

$$\mathbb{E}\left(\mathbf{g}_t\left(\mathbf{x}_{0:t}\right)\right) = \int \mathbf{g}_t\left(\mathbf{x}_{0:t}\right) p\left(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}\right) d\mathbf{x}_{0:t} \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}_t\left(\mathbf{x}_{0:t}^{(i)}\right)$$

   for some function of interest $\mathbf{g}_t : \left(\mathbb{R}^{n_x}\right)^{(t+1)} \rightarrow \mathbb{R}^{n_{g_t}}$ integrable with respect to $p\left(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}\right)$. Examples of appropriate functions include the marginal conditional mean of $\mathbf{x}_{0:t}$, in which case $\mathbf{g}_t\left(\mathbf{x}_{0:t}\right) = \mathbf{x}_t$, or the marginal conditional covariance of $\mathbf{x}_{0:t}$ with $\mathbf{g}_t\left(\mathbf{x}_{0:t}\right) = \mathbf{x}_t \mathbf{x}_t' - \mathbb{E}_{p(\mathbf{x}_t | \mathbf{y}_{1:t})}\left[\mathbf{x}_t\right] \mathbb{E}'_{p(\mathbf{x}_t | \mathbf{y}_{1:t})}\left[\mathbf{x}_t\right]$. The marginal conditional mean is often the quantity of interest, because it is the optimal MMSE estimate of the current state of the system.

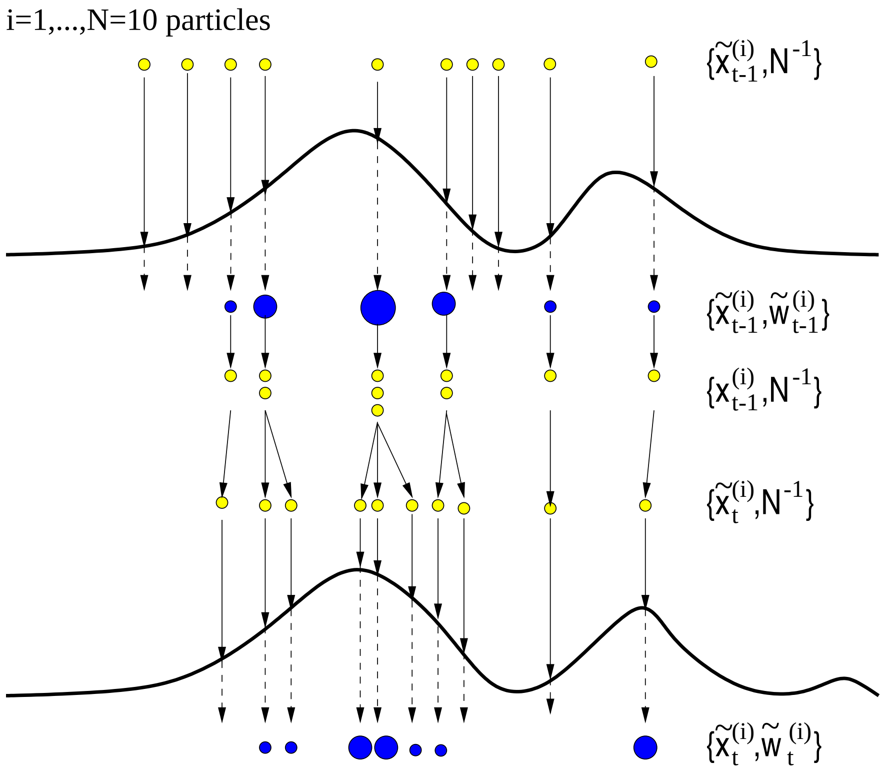

Figure 4: In this example, a standard particle filter starts at time $t-1$ with an unweighted measure $\{\widetilde{\mathbf{x}}^{(i)}_{t-1}, N^{-1}\}$, which provides an approximation of $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-2})$. For each particle we compute the importance weights using the information at time $t-1$. This results in the weighted measure $\{\widetilde{\mathbf{x}}^{(i)}_{t-1}, \widetilde{w}^{(i)}_{t-1}\}$, which yields an approximation $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$. Subsequently, the resampling step selects only the "fittest" particles to obtain the unweighted measure $\{\widetilde{\mathbf{x}}^{(i)}_{t-1}, N^{-1}\}$, which is still an approximation of $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$. Finally, the sampling (prediction) step introduces variety, resulting in the measure $\{\widetilde{\mathbf{x}}^{(i)}_{t}, N^{-1}\}$, which is an approximation of $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$.

A graphical representation of the algorithm is shown in Figure 4. The generic PF algorithm is rather straightforward to implement, but to make it robust, we need to consider some improvements discussed in the following section.

6 Improving Particle Filters

The success of the PF algorithm depends on the validity of the following underlying assumptions:

Monte Carlo (MC) assumption : The Dirac point-mass approximation provides an adequate representation of the posterior distribution.

Importance sampling (IS) assumption : It is possible to obtain samples from the posterior by sampling from a suitable proposal distribution and applying importance sampling corrections.

If any of these conditions are not met, the PF algorithm can perform poorly. The discreteness of the approximation poses a resolution problem. In the resampling stage, any particular sample with a high importance weight will be duplicated many times. As a result, the cloud of samples may eventually collapse to a single sample. This degeneracy will limit the ability of the algorithm to search for lower minima in other regions of the error surface. In other words, the number of samples used to describe the posterior density function will become too small and inadequate. A brute force strategy to overcome this problem is to increase the number of particles. A more refined strategy is to implement a Markov chain Monte Carlo (MCMC) step after the selection step as discussed in the following subsection.

6.1 MCMC Move Step

After the selection scheme at time t, we obtain N particles distributed marginally approximately according to $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$. Since the selection step favors the creation of multiple copies of the "fittest" particles, it enables us to track time varying filtering distributions. However, many particles might end up having no children ($N_i = 0$), whereas others might end up having a large number of children, the extreme case being $N_i = N$ for a particular value i. In this case, there is a severe depletion of samples. We, therefore, require a procedure to introduce sample variety after the selection step without affecting the validity of the approximation.

A strategy for solving this problem involves introducing MCMC steps of invariant distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ on each particle (Andrieu, de Freitas and Doucet 1999b, Carpenter et al. 1999, Doucet and Gordon 1999, Gilks and Berzuini 1998, MacEachern, Clyde and Liu 1999). The basic idea is that if the particles are distributed according to the posterior $p(\widetilde{\mathbf{x}}_{0:t}|\mathbf{y}_{1:t})$, then applying a Markov chain transition kernel $\mathcal{K}(\mathbf{x}_{0:t}|\widetilde{\mathbf{x}}_{0:t})$, with invariant distribution $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ such that $\int \mathcal{K}(\mathbf{x}_{0:t}|\widetilde{\mathbf{x}}_{0:t})p(\widetilde{\mathbf{x}}_{0:t}|\mathbf{y}_{1:t}) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$, still results in a set of particles distributed according to the posterior of interest. However, the new particles might have been moved to more interesting areas of the state-space. In fact, by applying a Markov transition kernel, the total variation of the current distribution with respect to the invariant distribution can only decrease. Note that we can incorporate any of the standard MCMC methods, such as the Gibbs sampler and Metropolis Hastings algorithms, into the filtering framework, but we no longer require the kernel to be ergodic. The MCMC move step can also be interpreted as sampling from the finite mixture distribution $N^{-1}\sum_{i=1}^{N}\mathcal{K}(\mathbf{x}_{0:t}|\widetilde{\mathbf{x}}_{0:t}^{(i)})$.

Convergence results for this type of algorithm are presented in (Gilks and Berzuini 1998).

One can generalize this idea by introducing MCMC steps on the product space with invariant distribution $\prod_{i=1}^{N} p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})$, that is to apply MCMC steps on the entire population of particles. It should be noted that independent MCMC steps spread out the particles in a particular mode more evenly, but do not explore modes devoid of particles, unless "clever" proposal distributions are available. By adopting MCMC steps on the whole population, we can draw upon many of the ideas developed in parallel MCMC computation. In this work, however, we limit ourselves to the simpler case of using independent MCMC transitions steps on each particle. In the case of standard particle filters, we propose to sample from the transition prior and accept according to a Metropolis-Hastings (MH) step as follows.

Smoothing MH step

- Sample $v \sim \mathcal{U}_{[0,1]}$.
- Sample the proposal candidate $\mathbf{x}_t^{\star(i)} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$
- If $v \leq \min\left\{1, \frac{p(\mathbf{y}_t|\mathbf{x}_t^{\star(i)})}{p(\mathbf{y}_t|\widetilde{\mathbf{x}}_t^{(i)})}\right\}$
  - then accept move:

$$\mathbf{x}_{0:t}^{(i)} = \left(\widetilde{\mathbf{x}}_{0:t-1}^{(i)}, \mathbf{x}_t^{\star(i)}\right)$$

  - else reject move:

$$\mathbf{x}_{0:t}^{(i)} = \widetilde{\mathbf{x}}_{0:t}^{(i)}$$

End If.

It is possible, however, to use more complex proposals such as mixtures of Metropolis-Hastings steps to ensure an efficient exploration of the sample space (de Freitas 1999). It is even possible to implement reversible jump MCMC steps (Green 1995) so as to allow the particles to move from one subspace to other subspaces of, possibly, different dimension (Andrieu, de Freitas and Doucet 1999a). Later, we shall describe MCMC steps that use the EKF and unscented filters to generate the proposal distributions.

6.2 Designing Better Importance Proposals

The importance sampling approximation depends on how close the proposal distribution is to the posterior distribution. As illustrated in Figure 2, if the likelihood is too peaked or if there is little overlap between the prior and the likelihood, one needs to move the samples to regions of high likelihood. Various approaches have been proposed to solve this problem. We present some of them.

6.2.1 Prior editing, rejection methods and auxiliary particle filters

Prior editing (Gordon et al. 1993) is an *ad-hoc* acceptance test for proposing particles in regions of high likelihood. After the prediction step, the residual error $\mathbf{e}_t = \mathbf{y}_t - \mathbf{h}_t(\widehat{\mathbf{x}}_t^{(i)})$ is computed. If $|\mathbf{e}_t| > K_l\sqrt{r}$, where r is the scale of the measurement error model and K_l is a constant chosen to indicate the region of non-negligible likelihood, then the sample $\widehat{\mathbf{x}}_t^{(i)}$ is rejected. The procedure is repeated until a specified number of particles is accepted. The problem with this approach is that it is too heuristic and can be computationally intensive unless the rejection rate is small. In addition, it introduces a bias on the distribution of the particles.

Rejection methods: If the likelihood is bounded, say $p(\mathbf{y}_t|\mathbf{x}_t) < M_t$, it is possible to sample from the optimal importance distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)$ using an accept/reject procedure. Firstly, we obtain a sample from the prior $\widehat{\mathbf{x}} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and a uniform variable $u \sim \mathcal{U}_{[0,1]}$. Subsequently, the sample from the prior is accepted if $u \leq p(\mathbf{y}_t|\widehat{\mathbf{x}}_t)/M_t$. Otherwise, we reject the proposed sample and repeat the process until N samples are accepted. Unfortunately, the rejection sampler requires a random number of iterations at each time step. This proves to be computationally expensive in high-dimensional spaces (Doucet 1998, Müller 1991, Pitt and Shephard 1999).

The **auxiliary particle filter** (Pitt and Shephard 1999) allows us to obtain approximate samples from the optimal importance distribution by introducing an auxiliary variable k. Specifically, the aim of the algorithm is to draw samples from the joint distribution

$$q(\mathbf{x}_t, k|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mu_t^{(k)})p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})p(\mathbf{x}_{1:t-1}^{(k)}|\mathbf{y}_{1:t-1})$$

where $\mu_t^{(k)}, k = 1, \ldots, N$ is the mean, mode, draw, or some other value associated with the transition prior. One way to accomplish this objective is to evaluate the marginal auxiliary variable weights $g(k|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mu_t^{(k)})p(\mathbf{x}_{1:t-1}^{(k)}|\mathbf{y}_{1:t-1})$ and use them to select M particles from the transition prior. Typically, one boosts the sample set so that $M > N$. The particle filter then proceeds to evaluate the correction weights

$$w_t = \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(j)})}{p(\mathbf{y}_t|\mu_t^{(k_j)})}$$

where $j = 1, \ldots, M$ and k_j denotes the k-th "parent" of particle j. Finally, the correction weights are used to perform a second selection step to obtain N particles approximately distributed according to the posterior distribution.

In comparison to the SIR filter, the auxiliary particle filter can generate better estimates of the posterior whenever the likelihood is situated in one of the priors tails. On the other hand, if the likelihood and prior coincide, the SIR filter can produce more accurate estimates. The latter behavior is a consequence of the extra variance introduced by the additional selection step.

One alternative way of viewing the auxiliary particle filter is to interpret the distribution $q(\mathbf{x}_t, k|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$ as the importance proposal. In doing so, the following importance weights are obtained

$$\begin{aligned} w_t &\propto \frac{p(\mathbf{x}_{0:t}^{(k)}|\mathbf{y}_{1:t})}{p(\mathbf{y}_t|\mu_t^{(k)})p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})p(\mathbf{x}_{1:t-1}^{(k)}|\mathbf{y}_{1:t-1})} \\ &\propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(k)})p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})p(\mathbf{x}_{1:t-1}^{(k)}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mu_t^{(k)})p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(k)})p(\mathbf{x}_{1:t-1}^{(k)}|\mathbf{y}_{1:t-1})} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(k)})}{p(\mathbf{y}_t|\mu_t^{(k)})} \end{aligned}$$

The three methods presented above for designing better proposal distributions have numerous inefficiencies as discussed in the literature. For this reason we didn't include them in the filter set used for the experiments in Section 9. We presented them here for completeness though and refer the reader to the literature for more detail and experimental results.

6.2.2 Local linearization

This is a popular method for devising proposal distributions that approximate the optimal importance distribution, by incorporating the most current observation with the optimal Gaussian approximation of the state: see (Doucet 1998, Pitt and Shephard 1999) for example. It relies on the first order Taylor series expansions of the likelihood and transition prior as described in Section 3.1, as well as a Gaussian assumption on all the random variables in question. In this framework, the EKF approximates the optimal MMSE estimator of the system state by calculating the conditional mean of the state, given all of the observations. This is done in recursive framework, by propagating the Gaussian approximation of the posterior distribution through time, combining it at each time step with the new observation. In other words, the EKF calculates the following recursive approximation to the true posterior filtering density,

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx p_{\mathcal{N}}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}\left(\bar{\mathbf{x}}_t, \widehat{\mathbf{P}}_t\right) \tag{25}$$

Within the particle filter framework, a separate EKF is used to generate and propagate a Gaussian proposal distribution for each particle, i.e.,

$$q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}) \triangleq \mathcal{N}\big(\bar{\mathbf{x}}_t^{(i)}, \widehat{\mathbf{P}}_t^{(i)}\big) \qquad i = 1, \ldots, N. \tag{26}$$

That is, at time $t-1$ one uses the EKF equations, with the new data, to compute the mean and covariance of the importance distribution for each particle. Next, we sample the i-th particle from this distribution. The method requires that we propagate the covariance $\widehat{P}^{(i)}$ and specify the EKF process and measurement noise covariances. This new filter is called the *extended Kalman particle filter*.

Since the EKF is an MMSE estimator, this local linearization method leads to an improved annealed sampling algorithm, whereby the variance of each proposal distribution changes with time. Ideally, we start searching over a large region of the error surface and as time progresses, we concentrate on the regions of lower error.

Although the EKF moves the prior towards the likelihood, thus possibly creating a better proposal distribution, this is done at the cost of making a Gaussian assumption on the form of the posterior as well as introducing inaccuracies due to linearization. When we compare the form of Equation (25) to the Gaussian transition prior of Equation (22), we see that EKF generated proposal distribution does indeed include the effect of the most current observation at time t. In general though (even with additive Gaussian process and measurement noise models), the true form of this density will not be Gaussian. This can easily be shown using a Bayes rule expansion of the proposal distribution. Because of this, we have to experimentally determine if we are gaining more than we lose in filter performance. The results of this is shown in Section 9. The *unscented particle filter* was developed to address some of the short-comings of the extended Kalman particle filter and is presented in Section 7.

The pseudo-code for the extended Kalman particle filter follows.

Extended Kalman particle filter

1. Initialization: $t = 0$

   - For $i = 1, \ldots, N$, draw the states (particles) $\mathbf{x}_0^{(i)}$ from the prior $p(\mathbf{x}_0)$.

2. For $t = 1, 2, \ldots$

   (a) Importance sampling step

   - For $i = 1, \ldots, N$:
     - Compute the Jacobians $\mathbf{F}_t^{(i)}$ & $\mathbf{G}_t^{(i)}$ and $\mathbf{H}_t^{(i)}$ & $\mathbf{U}_t^{(i)}$ of the process and measurement models.
     - Update the particles with the EKF:

$$
\begin{aligned}
\bar{\mathbf{x}}_{t|t-1}^{(i)} &= \mathbf{f}(\mathbf{x}_{t-1}^{(i)}) \\
\mathbf{P}_{t|t-1}^{(i)} &= \mathbf{F}_t^{(i)}\mathbf{P}_{t-1}^{(i)}\mathbf{F}_t^{T(i)} + \mathbf{G}_t^{(i)}\mathbf{Q}_t\mathbf{G}_t^{T(i)} \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1}^{(i)}\mathbf{H}_t^{T(i)}[\mathbf{U}_t^{(i)}\mathbf{R}_t\mathbf{U}_t^{T(i)} + \mathbf{H}_t^{(i)}\mathbf{P}_{t|t-1}^{(i)}\mathbf{H}_t^{T(i)}]^{-1} \\
\bar{\mathbf{x}}_t^{(i)} &= \bar{\mathbf{x}}_{t|t-1}^{(i)} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}(\bar{\mathbf{x}}_{t|t-1}^{(i)})) \\
\widehat{\mathbf{P}}_t^{(i)} &= \mathbf{P}_{t|t-1}^{(i)} - \mathbf{K}_t\mathbf{H}_t^{(i)}\mathbf{P}_{t|t-1}^{(i)}
\end{aligned}
$$

     - Sample $\widehat{\mathbf{x}}_t^{(i)} \sim q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}) = \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}, \widehat{\mathbf{P}}_t^{(i)})$
     - Set $\widehat{\mathbf{x}}_{0:t}^{(i)} \triangleq (\mathbf{x}_{0:t-1}^{(i)}, \widehat{\mathbf{x}}_t^{(i)})$ and $\widehat{\mathbf{P}}_{0:t}^{(i)} \triangleq (\mathbf{P}_{0:t-1}^{(i)}, \widehat{\mathbf{P}}_t^{(i)})$
   - For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant:

$$
w_t^{(i)} \propto \frac{p(\mathbf{y}_t|\widehat{\mathbf{x}}_t^{(i)})p(\widehat{\mathbf{x}}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\widehat{\mathbf{x}}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}
$$

   - For $i = 1, \ldots, N$, normalize the importance weights.

   (b) Selection step

   - Multiply/Suppress particles $(\widehat{\mathbf{x}}_{0:t}^{(i)}, \widehat{\mathbf{P}}_{0:t}^{(i)})$ with high/low importance weights $\widetilde{w}_t^{(i)}$, respectively, to obtain N random particles $(\widetilde{\mathbf{x}}_{0:t}^{(i)}, \widetilde{\mathbf{P}}_{0:t}^{(i)})$.

   (c) MCMC step (optional)

   - Apply a Markov transition kernel with invariant distribution given by $p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})$ to obtain $(\mathbf{x}_{0:t}^{(i)}, \mathbf{P}_{0:t}^{(i)})$.

   (d) Output: The output is generated in the same manner as for the generic particle filter.

The optional MCMC step consists of the MH algorithm witch uses the EKF to generate a proposal distribution, as follows.

EKF MH step

- Sample v from a uniform distribution: $v \sim \mathcal{U}_{[0,1]}$.
- Compute the Jacobians $\mathbf{F}_t^{\star(i)}$ & $\mathbf{G}_t^{\star(i)}$ and $\mathbf{H}_t^{\star(i)}$ & $\mathbf{U}_t^{\star(i)}$ of the process and measurement models.
- Update the states (particles) with the EKF:

$$
\begin{aligned}
\bar{\mathbf{x}}_{t|t-1}^{\star(i)} &= \mathbf{f}(\widetilde{\mathbf{x}}_{t-1}^{(i)}) \\
\mathbf{P}_{t|t-1}^{\star(i)} &= \mathbf{F}_t^{\star(i)}\widetilde{\mathbf{P}}_{t-1}\mathbf{F}_t^{\star T(i)} + \mathbf{G}_t^{\star(i)}\mathbf{Q}_t\mathbf{G}_t^{\star T(i)} \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1}^{\star(i)}\mathbf{H}_t^{\star T(i)}[\mathbf{U}_t^{\star(i)}\mathbf{R}_t\mathbf{U}_t^{\star T(i)} + \mathbf{H}_t^{\star(i)}\mathbf{P}_{t|t-1}^{\star(i)}\mathbf{H}_t^{\star T(i)}]^{-1} \\
\bar{\mathbf{x}}_t^{\star(i)} &= \bar{\mathbf{x}}_{t|t-1}^{\star(i)} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}(\bar{\mathbf{x}}_{t|t-1}^{\star(i)})) \\
\mathbf{P}_t^{\star(i)} &= \mathbf{P}_{t|t-1}^{\star(i)} - \mathbf{K}_t\mathbf{H}_t^{\star(i)}\mathbf{P}_{t|t-1}^{\star(i)}
\end{aligned}
\tag{27}
$$

- Sample the candidate $\mathbf{x}_t^{\star(i)} \sim q(\mathbf{x}_t|\widetilde{\mathbf{x}}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}) = \mathcal{N}\left(\bar{\mathbf{x}}_t^{\star(i)}, \mathbf{P}_t^{\star(i)}\right)$
- If $v \leq \min\left\{1, \frac{p(\mathbf{y}_t|\mathbf{x}_t^{\star(i)})p(\mathbf{x}_t^{\star(i)}|\widetilde{\mathbf{x}}_{t-1}^{(i)})q(\widetilde{\mathbf{x}}_t|\widetilde{\mathbf{x}}_{0:t-1}^{(i)},\mathbf{y}_{1:t})}{p(\mathbf{y}_t|\widetilde{\mathbf{x}}_t^{(i)})p(\widetilde{\mathbf{x}}_t^{(i)}|\widetilde{\mathbf{x}}_{t-1}^{(i)})q(\mathbf{x}_t^{\star(i)}|\widetilde{\mathbf{x}}_{0:t-1}^{(i)},\mathbf{y}_{1:t})}\right\}$
  - then accept move:

$$
\begin{aligned}
\mathbf{x}_{0:t}^{(i)} &= \left(\widetilde{\mathbf{x}}_{0:t-1}^{(i)}, \mathbf{x}_t^{\star(i)}\right) \\
\mathbf{P}_{0:t}^{(i)} &= \left(\widetilde{\mathbf{P}}_{0:t-1}^{(i)}, \mathbf{P}_t^{\star(i)}\right)
\end{aligned}
$$

  - else reject move:

$$
\begin{aligned}
\mathbf{x}_{0:t}^{(i)} &= \widetilde{\mathbf{x}}_{0:t}^{(i)} \\
\mathbf{P}_{0:t}^{(i)} &= \widetilde{\mathbf{P}}_{0:t}^{(i)}
\end{aligned}
$$

End If.

In the following section, we introduce one technique that should in general perform better than rejection methods or extended Kalman filter expansions.

7 The Uscented Particle Filter

As shown in Section 3, the unscented Kalman filter (UKF) is able to more accurately propagate the mean and covariance of the Gaussian approximation to the state distribution, than the EKF. In comparison to the EKF, the UKF tends to generate more accurate estimates of the true covariance of the state. Distributions generated by the UKF generally have a bigger support overlap with the true posterior distribution than the overlap achieved by the EKF estimates. This is in part related to the fact that the UKF calculates the posterior covariance accurately to the 3rd order, whereas the EKF relies on a first order biased approximation. This makes the UKF a better candidate for more accurate proposal distribution[6] generation within the particle filter framework. The UKF also has the ability to scale the approximation errors in the higher order moments of the posterior distribution, eg. kurtosis, etc., allowing for heavier tailed distributions. Because the sigma point set used in the UKF is deterministically designed to capture certain characteristic of the prior distribution, one can explicitly optimize the algorithm to work with distributions that have heavier tails than Gaussian distributions, i.e. Cauchy or Student-t distributions. This characteristic makes the UKF very attractive for the generation of proposal distributions.

The new filter that results from using a UKF for proposal distribution generation within a particle filter framework is called the *Unscented Particle Filter* (UPF), and is the major new contribution of this paper.

The pseudo-code of the UPF follows

Unscented Particle Filter

1. Initialization: $t = 0$

   - For $i = 1, \ldots, N$, draw the states (particles) $\mathbf{x}_0^{(i)}$ from the prior $p(\mathbf{x}_0)$ and set,

$$\begin{aligned}
\bar{\mathbf{x}}_0^{(i)} &= E[\mathbf{x}_0^{(i)}] \\
\mathbf{P}_0^{(i)} &= E[(\mathbf{x}_0^{(i)} - \bar{\mathbf{x}}_0^{(i)})(\mathbf{x}_0^{(i)} - \bar{\mathbf{x}}_0^{(i)})^T] \\
\bar{\mathbf{x}}_0^{(i)a} &= E[\mathbf{x}^{(i)a}] = [(\bar{\mathbf{x}}_0^{(i)})^T \ \mathbf{0} \ \mathbf{0}]^T
\end{aligned}$$

$$\mathbf{P}_0^{(i)a} = E[(\mathbf{x}_0^{(i)a} - \bar{\mathbf{x}}_0^{(i)a})(\mathbf{x}_0^{(i)a} - \bar{\mathbf{x}}_0^{(i)a})^T] = \begin{bmatrix} \mathbf{P}_0^{(i)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}$$

CONTINUED ...

[6]Like the EKF, the UKF also incorporates the latest observations, but this is done in a more accurate way.

2. For $t = 1, 2, \ldots$

(a) Importance sampling step

- For $i = 1, \ldots, N$:
  - Update the particles with the UKF:
    - Calculate sigma points:

$$\boldsymbol{\mathcal{X}}_{t-1}^{(i)a} = \left[\bar{\mathbf{x}}_{t-1}^{(i)a} \;\; \bar{\mathbf{x}}_{t-1}^{(i)a} \pm \sqrt{(n_a + \lambda)\mathbf{P}_{t-1}^{(i)a}} \right]$$

    - Propagate particle into future (time update):

$$\boldsymbol{\mathcal{X}}_{t|t-1}^{(i)x} = \mathbf{f}\left(\boldsymbol{\mathcal{X}}_{t-1}^{(i)x}, \boldsymbol{\mathcal{X}}_{t-1}^{(i)v}\right) \qquad \bar{\mathbf{x}}_{t|t-1}^{(i)} = \sum_{j=0}^{2n_a} W_j^{(m)} \mathcal{X}_{j,t|t-1}^{(i)x}$$

$$\mathbf{P}_{t|t-1}^{(i)} = \sum_{j=0}^{2n_a} W_j^{(c)} [\mathcal{X}_{j,t|t-1}^{(i)x} - \bar{\mathbf{x}}_{t|t-1}^{(i)}][\mathcal{X}_{j,t|t-1}^{(i)x} - \bar{\mathbf{x}}_{t|t-1}^{(i)}]^T$$

$$\boldsymbol{\mathcal{Y}}_{t|t-1}^{(i)} = \mathbf{h}\left(\boldsymbol{\mathcal{X}}_{t|t-1}^{(i)x}, \boldsymbol{\mathcal{X}}_{t-1}^{(i)n}\right) \qquad \bar{\mathbf{y}}_{t|t-1}^{(i)} = \sum_{j=0}^{2n_a} W_j^{(m)} \mathcal{Y}_{j,t|t-1}^{(i)}$$

    - Incorporate new observation (measurement update):

$$\mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t} = \sum_{j=0}^{2n_a} W_j^{(c)} [\mathcal{Y}_{j,t|t-1}^{(i)} - \bar{\mathbf{y}}_{t|t-1}^{(i)}][\mathcal{Y}_{j,t|t-1}^{(i)} - \bar{\mathbf{y}}_{t|t-1}^{(i)}]^T$$

$$\mathbf{P}_{\mathbf{x}_t \mathbf{y}_t} = \sum_{j=0}^{2n_a} W_j^{(c)} [\mathcal{X}_{j,t|t-1}^{(i)} - \bar{\mathbf{x}}_{t|t-1}^{(i)}][\mathcal{Y}_{j,t|t-1}^{(i)} - \bar{\mathbf{y}}_{t|t-1}^{(i)}]^T$$

$$\mathbf{K}_t = \mathbf{P}_{\mathbf{x}_t \mathbf{y}_t} \mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t}^{-1} \qquad \bar{\mathbf{x}}_t^{(i)} = \bar{\mathbf{x}}_{t|t-1}^{(i)} + \mathbf{K}_t(\mathbf{y}_t - \bar{\mathbf{y}}_{t|t-1}^{(i)})$$

$$\widehat{\mathbf{P}}_t^{(i)} = \mathbf{P}_{t|t-1}^{(i)} - \mathbf{K}_t \mathbf{P}_{\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t} \mathbf{K}_t^T$$

  - Sample $\widehat{\mathbf{x}}_t^{(i)} \sim q(\mathbf{x}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)},, \mathbf{y}_{1:t}) = \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}, \widehat{\mathbf{P}}_t^{(i)})$
  - Set $\widehat{\mathbf{x}}_{0:t}^{(i)} \triangleq (\mathbf{x}_{0:t-1}^{(i)}, \widehat{\mathbf{x}}_t^{(i)})$ and $\widehat{\mathbf{P}}_{0:t}^{(i)}(\mathbf{P}_{0:t-1}^{(i)}, \widehat{\mathbf{P}}_t^{(i)})$
- For $i = 1, \ldots, N$, evaluate the importance weights up to a normalizing constant.

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t | \widehat{\mathbf{x}}_t^{(i)}) p(\widehat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\widehat{\mathbf{x}}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})}$$

- For $i = 1, \ldots, N$, normalize the importance weights.

(b) Selection step

- Multiply/Suppress particles $(\widehat{\mathbf{x}}_{0:t}^{(i)}, \widehat{\mathbf{P}}_{0:t}^{(i)})$ with high/low importance weights $\widetilde{w}_t^{(i)}$, respectively, to obtain N random particles $(\widetilde{\mathbf{x}}_{0:t}^{(i)}, \widetilde{\mathbf{P}}_{0:t}^{(i)})$.

(c) MCMC step (optional)

- Apply a Markov transition kernel with invariant distribution $p(\mathbf{x}_{0:t}^{(i)} | \mathbf{y}_{1:t})$ to obtain $(\mathbf{x}_{0:t}^{(i)}, \mathbf{P}_{0:t}^{(i)})$.

(d) Output: The output is generated in the same manner as for the generic particle filter.

8 Theoretical Convergence

Let $B(\mathbb{R}^n)$ be the space of bounded, Borel measurable functions on $\mathbb{R}^n$. We denote $\|f\| \triangleq \sup_{x \in \mathbb{R}^n} |f(x)|$. The following theorem is a straightforward consequence of Theorem 1 in (Crisan and Doucet 2000), which is an extension of previous results in (Crisan et al. 1999).

Theorem 1 *If the importance weight*

$$w_t \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})} \tag{28}$$

is upper bounded for any $(\mathbf{x}_{t-1}, \mathbf{y}_t)$ *and if one uses one of the selection schemes described previously, then, for all* $t \geq 0$*, there exists* c_t *independent of* N *such that for any* $f_t \in B\left(\mathbb{R}^{n_x \times (t+1)}\right)$

$$\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} f_t\left(\mathbf{x}_{0:t}^{(i)}\right) - \int f_t(\mathbf{x}_{0:t})\, p(d\mathbf{x}_{0:t}|\mathbf{y}_{1:t})\right)^2\right] \leq c_t \frac{\|f_t\|^2}{N}. \tag{29}$$

The expectation in equation (29) is with respect to the randomness introduced by the particle filtering algorithm. This convergence result shows that, under very loose assumptions, convergence of the (unscented) particle filter is ensured and that the convergence rate of the method is independent of the dimension of the state-space. The only crucial assumption is to ensure that w_t is upper bounded, that is that the proposal distribution $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$ has heavier tails than $p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{x}_{t-1})$. Considering this theoretical result, it should not be surprising that the UKF, which provides a better approximation to the higher moments of the filtering distribution than the EKF, yields better proposal distributions than the EKF within the particle filtering framework. In the following section, we present a few experiments that confirm this conjecture.

9 Experiments

We compared the performance of the Unscented Particle Filter to that of the other nonlinear filters on two estimation problems. The first problem is a synthetic, scalar estimation problem and the second is a real world problem concerning the pricing of financial instruments.

9.1 Synthetic Experiment

For this experiment, a time-series was generated by the following process model

$$x_{t+1} = 1 + sin(\omega \pi t) + \phi_1 x_t + v_t \tag{30}$$

where v_t is a Gamma $\mathcal{G}a(3,2)$ random variable modeling the process noise, and $\omega = 4e-2$ and $\phi_1 = 0.5$ are scalar parameters. A non-stationary observation model,

$$y_t = \begin{cases} \phi_2 x_t^2 + n_t & t \leq 30 \\ \phi_3 x_t - 2 + n_t & t > 30 \end{cases} \tag{31}$$

is used, with $\phi_2 = 0.2$ and $\phi_3 = 0.5$. The observation noise, n_t, is drawn from a Gaussian distribution $\mathcal{N}(0, 0.00001)$. Given only the noisy observations, y_t, the different filters were used to estimate the underlying clean state sequence x_t for $t = 1 \ldots 60$. The experiment was repeated 100 times with random re-initialization for each run. All of the particle filters used 200 particles and residual resampling (see Section 4.4 for details on resampling). The SUT parameters were set to $\alpha = 1$, $\beta = 0$ and $\kappa = 2$. These parameters are optimal for the scalar case. Table 1 summarizes the performance of the different filters. The table shows the means and variances of the mean-square-error (MSE) of the state estimates. Figure 5 compares the estimates generated from a single run of the different particle filters. The superior performance of the *unscented particle filter* (UPF) is clearly evident. Figure 6 shows the estimates of the state covariance generated by a stand-alone EKF and UKF for this problem. Notice how the EKF's estimates are consistently smaller than those generated by the UKF. This property makes the UKF better suited than the EKF for proposal distribution generation within the particle filter framework.

Algorithm	**MSE**	
	mean	var
Extended Kalman Filter (EKF)	0.374	0.015
Unscented Kalman Filter (UKF)	0.280	0.012
Particle Filter : generic	0.424	0.053
Particle Filter : MCMC move step	0.417	0.055
Particle Filter : EKF proposal	0.310	0.016
Particle Filter : EKF proposal and MCMC move step	0.307	0.015
Particle Filter : UKF proposal ("*Unscented Particle Filter*")	0.070	0.006
Particle Filter : UKF proposal and MCMC move step	0.074	0.008

Table 1: State estimation experiment results. This plot shows the mean and variance of the MSE calculated over 100 independent runs.

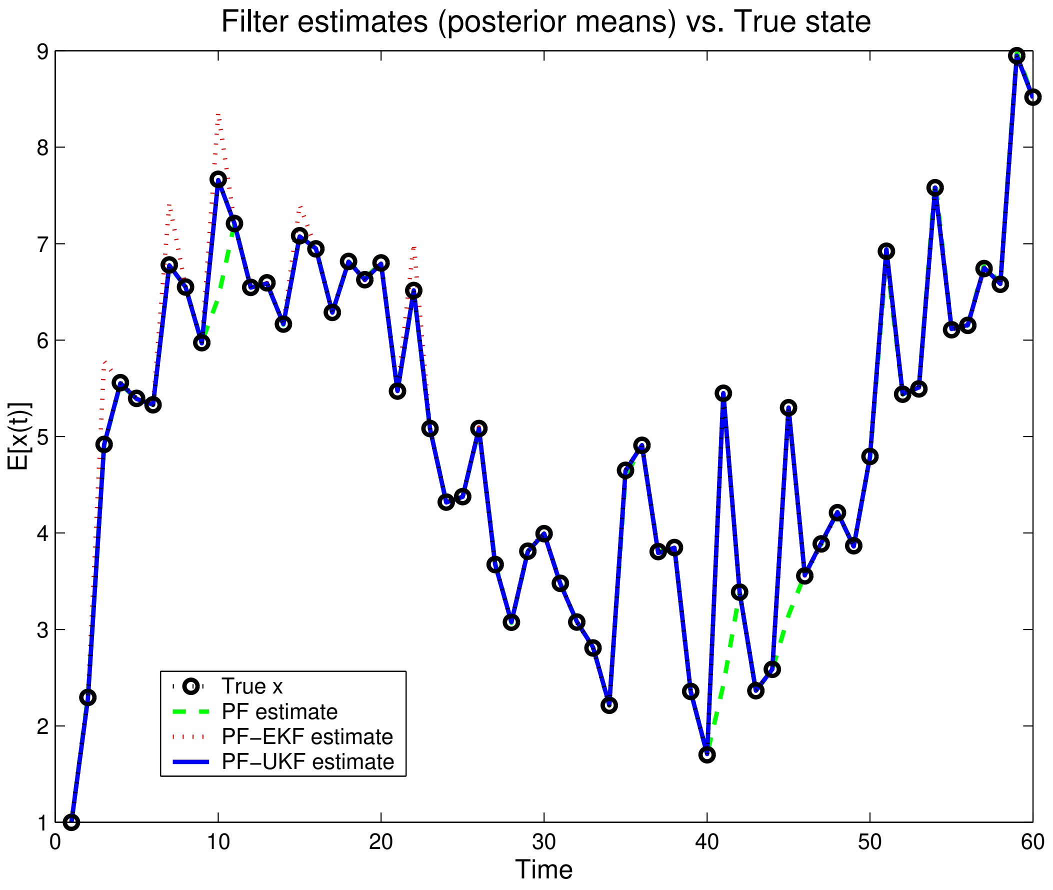

Figure 5: Plot of estimates generated by the different filters on the synthetic state estimation experiment.

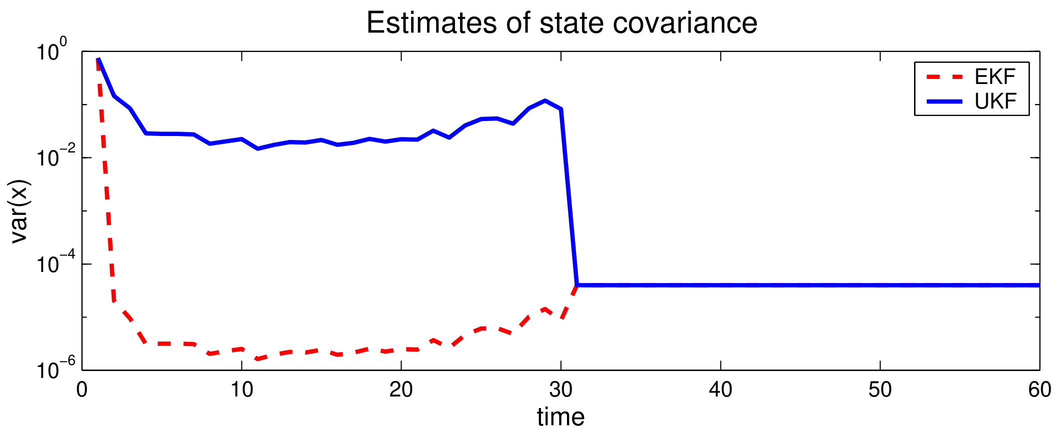

Figure 6: EKF and UKF estimates of the state covariance.

9.2 Pricing financial options

Derivatives are financial instruments whose value depends on some basic underlying cash product, such as interest rates, equity indices, commodities, foreign exchange or bonds

(Hull 1997). An option is a particular type of derivative that gives the holder the right to do something. For example, a call option allows the holder to buy a cash product, at a specified date in the future, for a price determined in advance. The price at which the option is exercised is known as the strike price, while the date at which the option lapses is often referred to as the maturity time. Put options, on the other hand, allow the holder to sell the underlying cash product.

The Black Scholes partial differential equation is, essentially, the main industry standard for pricing options (Hull 1997). It relates the current value of an option (f) to the current value of the underlying cash product (S), the volatility of the cash product (σ) and the risk-free interest rate (r) as follows

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf$$

This basic equation is only valid under several conditions, namely no risk-less arbitrage opportunities, an instantaneous risk-less portfolio, continuous trading, no dividends, constant volatility and risk-free interest rate. In addition, the cash product is assumed to be dictated by the following geometric Brownian motion model

$$\frac{\mathrm{d}S}{S} = \mu \mathrm{d}t + \sigma\epsilon \mathrm{d}t$$

where μ is the expected return and ϵ corresponds to a random sample from a standardized normal distribution (with mean zero and unit variance). In their seminal work (Black and Scholes 1973), Black and Scholes derived the following solutions for pricing European call and put options

$$C = S\mathcal{N}_c(d_1) - Xe^{-rt_m}\mathcal{N}_c(d_2) \quad (32)$$

$$P = -S\mathcal{N}_c(-d_1) + Xe^{-rt_m}\mathcal{N}_c(-d_2) \quad (33)$$

where C denotes the price of a call option, P the price of a put option, X the strike price, t_m the time to maturity, $\mathcal{N}_c(.)$ is the cumulative normal distribution, and d_1 and d_2 are given by

$$\begin{aligned} d_1 &= \frac{\ln(S/X) + (r + \sigma^2/2)t_m}{\sigma\sqrt{t_m}} \\ d_2 &= d_1 - \sigma\sqrt{t_m} \end{aligned}$$

The volatility is usually estimated from a small moving window of data over the most recent 50 to 180 days (Hull 1997). The risk-free interest rate is often estimated by monitoring interest rates in the bond markets. In our approach, which follows from (Niranjan 1996), we use the state-space representation to model the system given by equations (32) and (33).

We treat r and σ as the hidden states and C and P as the output observations. t_m and S are treated as known control signals (input observations). We believe that this approach is better since it constitutes a more natural way of dealing with the sequential behavior and non-stationarity of the data. In the end, we are able to compute daily complete probability distributions for r and σ and to decide whether the current value of an option in the market is being either over-priced or under-priced.

Typically, options on a particular equity and with the same exercise date are traded with several strike prices. For example, in our experiments, we used five pairs of call and put option contracts on the British FTSE100 index (from February 1994 to December 1994) to evaluate the pricing algorithms. For each option on this set one can estimate a different volatility. By plotting the Black-Scholes estimates of the volatilities against their respective strike prices, we obtain a curve which is known as the volatility smile (Hull 1997). A well known pricing strategy is to leave one of the options out and then determine the volatility smile given by the other options. If the option that was left out is below the curve, it could mean that it is under-priced by the market.

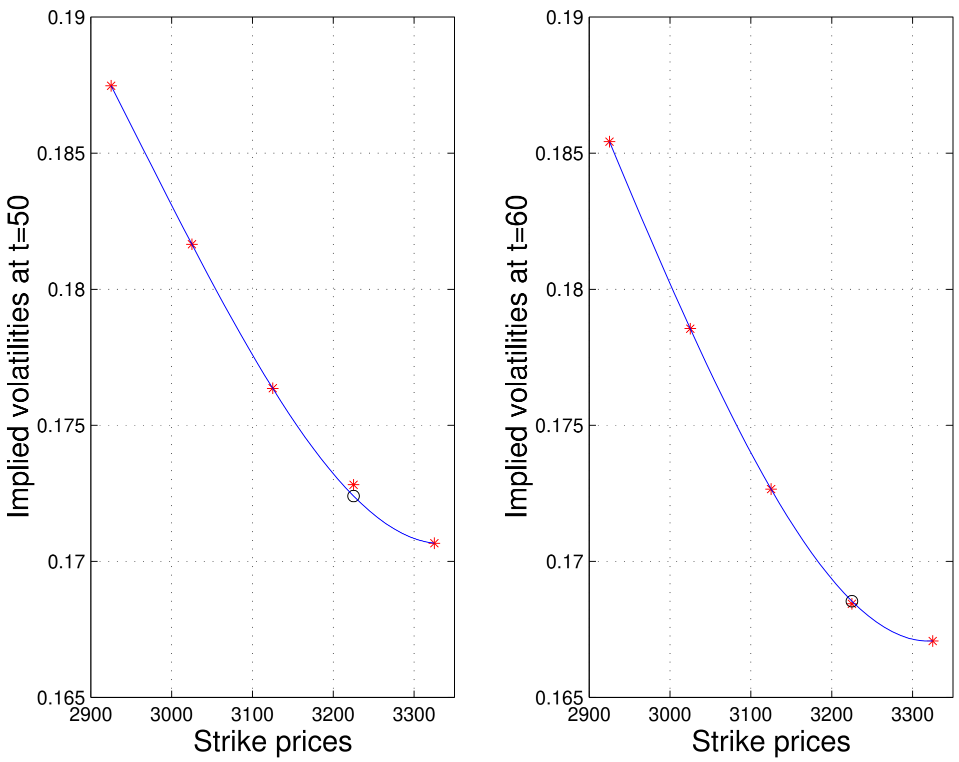

Figure 7: Volatility smile indicating that an option on the FTSE-100 index was over-priced. The option value 10 days later confirmed this hypothesis. Estimates obtained with a particle filter [*], 4-th order polynomial fit [—] and hypothesized volatility [o].

Figure 7 shows an example of this phenomenon obtained by tracking 5 pairs of call and put option contracts on the FTSE-100 index (1994) with a particle filter. On the 50th day, option 4 seems to be over-priced. The state of this option 10 days later confirms this hypothesis. However, depending on the state of the particular equity, some options might remain under-priced or over-priced during their entire life-time. For example, if an option on a company product seems to be over-priced according to its volatility smile, but investors know that the company is being bought by a larger company with better management, the option price will remain higher than the smile prediction (Haugen 1990).

In the sequential Monte Carlo framework, we can improve this trading strategy. Instead of plotting a volatility smile, we plot a probability smile. That is, we can plot the probability density function of each implied volatility against their respective strike prices, as shown in Figure 8. This plot, clearly, conveys more information than a simple plot of the posterior mean estimates.

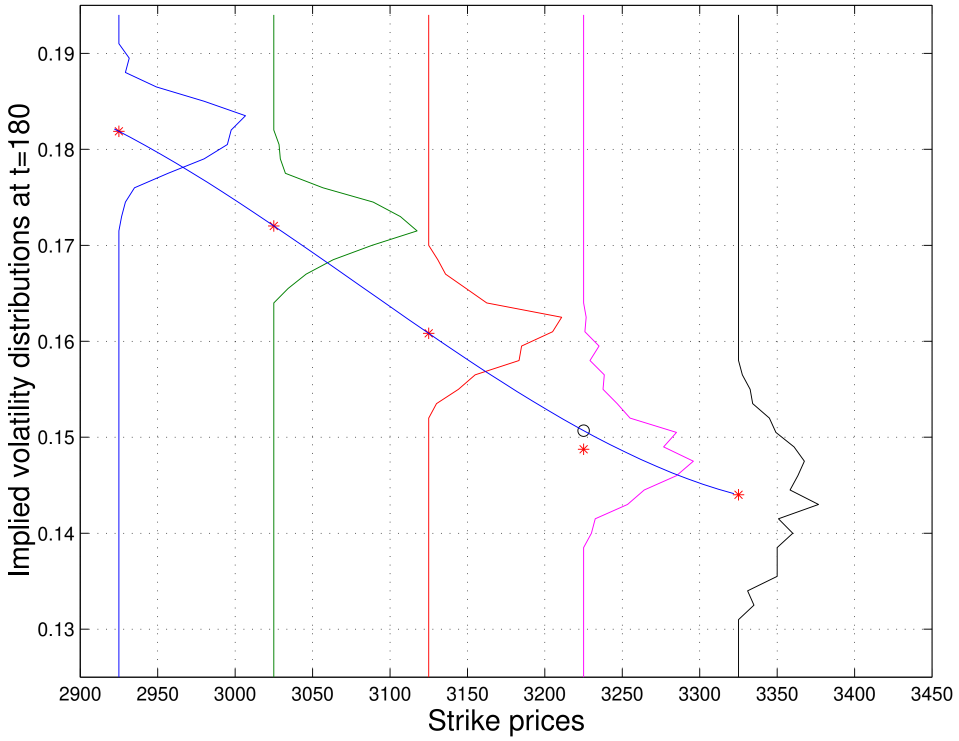

Figure 8: Probability smile for options on the FTSE-100 index (1994). Although the volatility smile indicates that the option with strike price equal to 3225 is under-priced, the shape of the probability gives us a warning against the hypothesis that the option is under-priced. Posterior mean estimates obtained with Black-Scholes model and particle filter [*], 4-th order polynomial fit [—] and hypothesized volatility [o].

The type of predictions obtained with the Unscented Particle Filter were very close

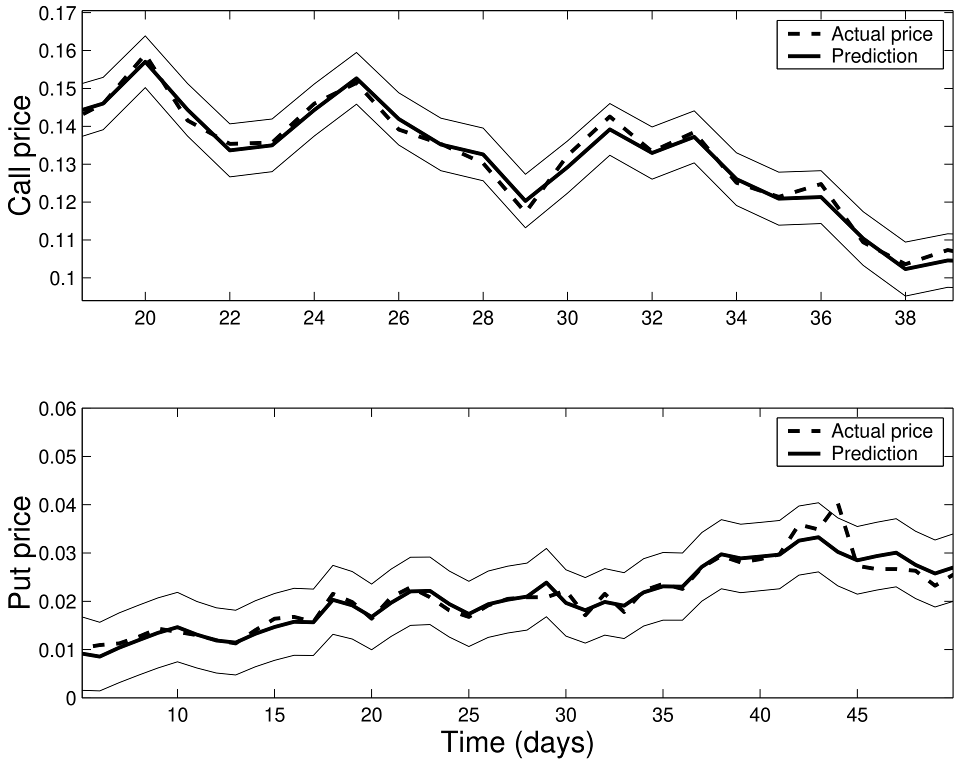

Figure 9: UPF one-step-ahead predictions on the call and put option's prices with confidence intervals (2 standard deviations).

to the measured data as evidenced by Figure 9. Figure 10 shows the estimated volatility and interest rate for a contract with a strike price of 3225. Plots of the evolution of the probability distributions of the interest rate and volatility are depicted in Figures 11 and 12.

In Table 2, we compare the one-step-ahead normalized square errors obtained with each method on a pair of options with strike price 2925. The normalized square errors are defined as follows

$$NSE_C = \sqrt{\sum_t (C_t - \widehat{C}_t)^2}$$

$$NSE_P = \sqrt{\sum_t (P_t - \widehat{P}_t)^2}$$

where $\widehat{C}_t$ and $\widehat{P}_t$ denotes the one-step-ahead predictions of the call and put prices. The square errors were only measured over the last 100 days of trading, so as to allow the algorithms to converge. The experiment was repeated 100 times and we used 100 particles in each particle filter. This table shows that, this time both the EKF and UKF led to

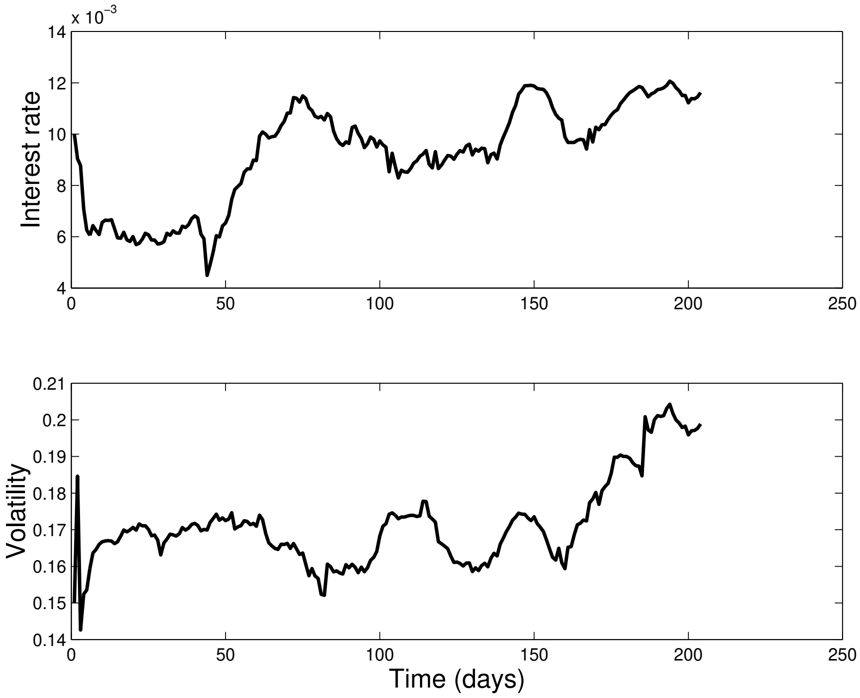

Figure 10: Estimated interest rate and volatility.

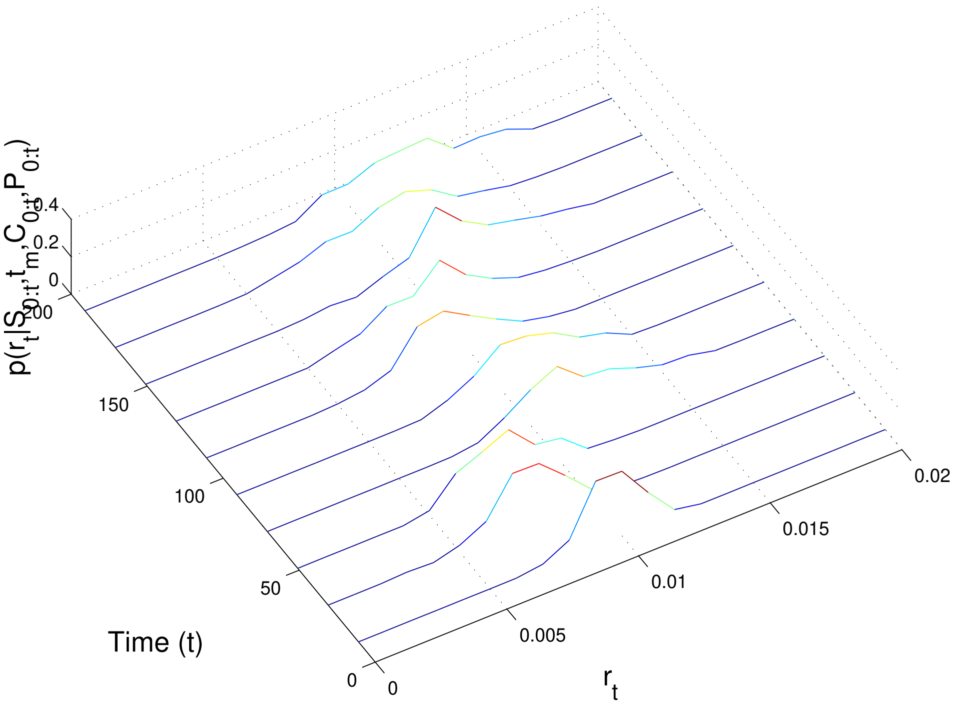

Figure 11: Probability distribution of the implied interest rate.

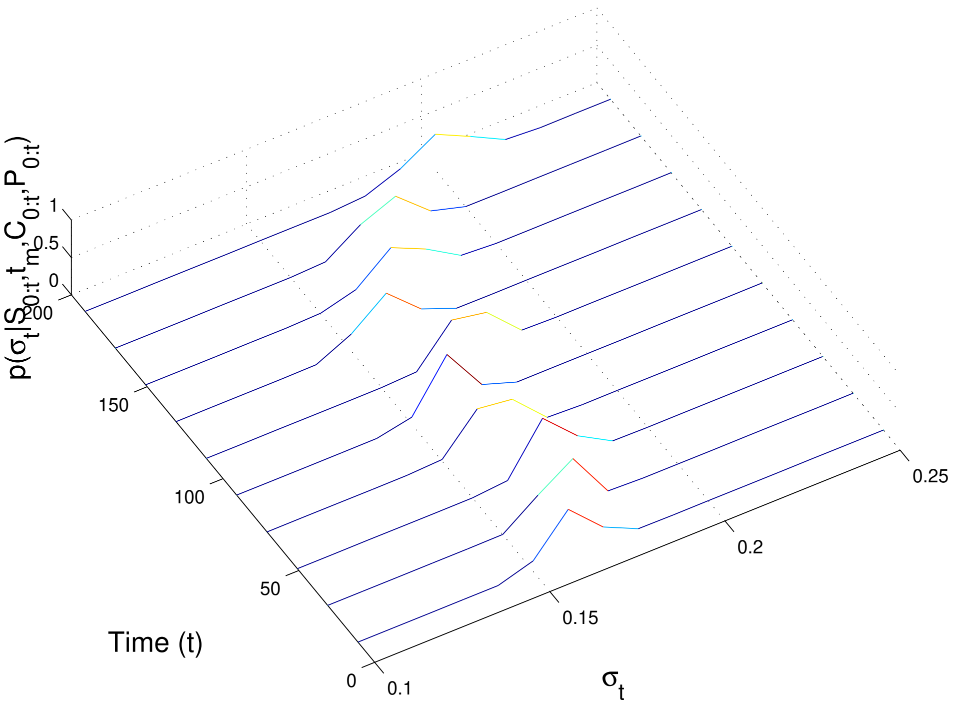

Figure 12: Probability distribution of the implied volatility.

Option type	**Algorithm**	**NSE** mean	**NSE** var
Call	Trivial	0.078	0.000
	Extended Kalman Filter (EKF)	0.037	0.000
	Unscented Kalman Filter (UKF)	0.037	0.000
	Particle Filter : generic	0.037	0.000
	Particle Filter : EKF proposal	0.009	0.000
	Unscented Particle Filter	0.009	0.000
Put	Trivial	0.035	0.000
	Extended Kalman Filter (EKF)	0.023	0.000
	Unscented Kalman Filter (UKF)	0.023	0.000
	Particle Filter : generic	0.023	0.000
	Particle Filter : EKF proposal	0.007	0.000
	Unscented Particle Filter	0.008	0.000

Table 2: One-step-ahead normalized square errors over 100 runs. The trivial prediction is obtained by assuming that the price on the following day corresponds to the current price.

the same improvement over standard particle filtering. This is because in this particular instance the process model is linear Gaussian and the nonlinearity of the measurement model is not too severe. This, however, will not be the case in general. The important thing is to notice that the UKF, used as a mechanism to generate the proposal distribution, works well with a difficult real data set.

10 Conclusions

In this paper, we proposed a new particle filter that uses the UKF to generate the proposal distribution. When the process and measurement models are either highly nonlinear or contain heavy tailed noise, the UKF produces proposal distributions that exhibit a larger support overlap with the true posterior than the EKF proposal distributions, making it better suited for proposal distribution generation. Since the UKF can also theoretically have heavier tails than the EKF, while still incorporating the latest information before the evaluation of the importance weights, the theory predicts that this filter can perform very well in situations where the likelihood is peaked or when one finds outliers in the data. A synthetic experiment and an experiment with real financial data showed that the unscented particle filter can perform better than other sequential estimation algorithms. We hope in the future to extend the range of applications of the unscented particle filter. Towards this purpose, we have made the software freely available at `http://www.cs.berkeley.edu/~jfgf` and `http://varsha.ece.ogi.edu/~rvdmerwe`.

Acknowledgements

We would like to thank Mahesan Niranjan for providing the options data and for many interesting discussions on this problem.

References

Akashi, H. and Kumamoto, H. (1977). Random sampling approach to state estimation in switching environments, *Automatica* **13**: 429–434.

Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, New Jersey.

Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999a). Sequential Bayesian estimation and model selection applied to neural networks, *Technical Report CUED/F-INFENG/TR 341*, Cambridge University Engineering Department.

Andrieu, C., de Freitas, J. F. G. and Doucet, A. (1999b). Sequential MCMC for Bayesian model selection, *IEEE Higher Order Statistics Workshop*, Ceasarea, Israel, pp. 130–134.

Avitzour, D. (1995). A stochastic simulation Bayesian approach to multitarget tracking, *IEE Proceedings on Radar, Sonar and Navigation* **142**(2): 41–44.

Beadle, E. R. and Djurić, P. M. (1997). A fast weighted Bayesian bootstrap filter for nonlinear model state estimation, *IEEE Transactions on Aerospace and Electronic Systems* **33**(1): 338–343.

Berzuini, C., Best, N. G., Gilks, W. R. and Larizza, C. (1997). Dynamic conditional independence models and Markov Chain Monte Carlo methods, *Journal of the American Statistical Association* **92**(440): 1403–1412.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–659.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999). Building robust simulation-based filters for evolving data sets, Unpublished. Department of Statistics, Oxford University.

Chong, K. S. and Kleeman, L. (1997). Sonar based map building for a mobile robot, *Proc. of 1997 IEEE International Conference on Robotics and Automation*.

Clark, S. (1999). *Autonomous Land Vehicle Navigation using Millimetre Wave Radar*, PhD thesis, University of Sydney.

Crisan, D. (2000). Particle filters - a theoretical perspective, *in* A. Doucet, J. F. G. de Freitas and N. J. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.

Crisan, D. and Doucet, A. (2000). Convergence of generalized particle filters, *Technical Report CUED/F-INFENG/TR 381*, Cambridge University Engineering Department.

Crisan, D., Del Moral, P. and Lyons, T. (1999). Discrete filtering using branching and interacting particle systems, *Markov Processes and Related Fields* **5**(3): 293–318.

de Freitas, J. F. G. (1999). *Bayesian Methods for Neural Networks*, PhD thesis, Department of Engineering, Cambridge University, Cambridge, UK.

de Freitas, J. F. G., Niranjan, M., Gee, A. H. and Doucet, A. (2000). Sequential Monte Carlo methods to train neural network models, *Neural Computation* **12**(4): 955–993.

Doucet, A. (1997). *Monte Carlo Methods for Bayesian Estimation of Hidden Markov Models. Application to Radiation Signals*, PhD thesis, University Paris-Sud, Orsay, France. Chapters 4 and 5 in English.

Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering, *Technical Report CUED/F-INFENG/TR 310*, Department of Engineering, Cambridge University.

Doucet, A. and Gordon, N. J. (1999). Simulation-based optimal filter for manoeuvring target tracking, *SPIE Signal and Data Processing of Small Targets*, Vol. SPIE 3809. To appear.

Doucet, A., de Freitas, J. F. G. and Gordon, N. (2000). Introduction to sequential Monte Carlo methods, *in* A. Doucet, J. F. G. de Freitas and N. J. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.

Doucet, A., Gordon, N. J. and Krishnamurthy, V. (1999). Particle filters for state estimation of jump Markov linear systems, *Technical Report CUED/F-INFENG/TR 359*, Cambridge University Engineering Department.

Efron, B. (1982). *The Bootstrap, Jacknife and other Resampling Plans*, SIAM, Philadelphia.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* **24**: 1317–1399.

Gilks, W. R. and Berzuini, C. (1998). Monte Carlo inference for dynamic Bayesian models, Unpublished. Medical Research Council, Cambridge, UK.

Gordon, N. J. (1994). *Bayesian Methods for Tracking*, PhD thesis, Imperial College, University of London.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F* **140**(2): 107–113.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**: 711–732.

Hammersley, J. M. and Morton, K. W. (1954). Poor man's Monte Carlo, *Journal of the Royal Statistical Society B* **16**: 23–38.

Handschin, J. E. (1970). Monte Carlo techniques for prediction and filtering of non-linear stochastic processes, *Automatica* **6**: 555–563.

Handschin, J. E. and Mayne, D. Q. (1969). Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering, *International Journal of Control* **9**(5): 547–559.

Haugen, R. A. (1990). *Modern Investment Theory*, Prentice-Hall International, New Jersey.

Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators, *Journal of Statistical Computation and Simulation* **59**(1): 1–23.

Hull, J. C. (1997). *Options, Futures, and Other Derivatives*, third edn, Prentice Hall.

Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density, *European Conference on Computer Vision*, Cambridge, UK, pp. 343–356.

Julier, S. J. (2000). The Scaled Unscented Transformation, *To appear in Automatica.* (preprint).

Julier, S. J. and Uhlmann, J. K. (1996). A General Method for Approximating Nonlinear Transformations of Probability Distributions, *Technical report*, RRG, Dept. of Engineering Science, University of Oxford.

Julier, S. J. and Uhlmann, J. K. (1997a). A consistent, debiased method for converting between polar and cartesian coordinate systems., *Proc. of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Orlando, Florida.*, Vol. Acquisition, Tracking and Pointing XI.

Julier, S. J. and Uhlmann, J. K. (1997b). A new extension of the Kalman filter to nonlinear systems, *Proc. of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, Orlando, Florida.*, Vol. Multi Sensor Fusion, Tracking and Resource Management II.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**: 1–25.

Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems, *Journal of the American Statistical Association* **89**(425): 278–288.

Liu, J., Chen, R. and Logvinenko, T. (2000). A theoretical framework for sequential importance sampling with resampling, *in* A. Doucet, J. F. G. de Freitas and N. J. Gordon (eds), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag.

Liu, J. S. and Chen, R. (1995). Blind deconvolution via sequential imputations, *Journal of the American Statistical Association* **90**(430): 567–576.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association* **93**: 1032–1044.

MacEachern, S. N., Clyde, M. and Liu, J. S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation, *Canadian Journal of Statistics* **27**: 251–267.

Müller, P. (1991). Monte Carlo integration in general dynamic models, *Contemporary Mathematics* **115**: 145–163.

Niranjan, M. (1996). Sequential tracking in pricing financial options using model based and neural network approaches, *in* M. C. Mozer, M. I. Jordan and T. Petsche (eds), *Advances in Neural Information Processing Systems*, Vol. 8, pp. 960–966.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* **94**(446): 590–599.

Ripley, B. D. (1987). *Stochastic Simulation*, Wiley, New York.

Rosenbluth, M. N. and Rosenbluth, A. W. (1955). Monte Carlo calculation of the average extension of molecular chains, *Journal of Chemical Physics* **23**: 356–359.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions, *in* J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds), *Bayesian Statistics 3*, Oxford University Press, Cambridge, MA, pp. 395–402.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective, *American Statistician* **46**(2): 84–88.

Smith, R. L. and Miller, J. E. (1986). Predictive records, *Journal of the Royal Statistical Society* **B 36**: 79–88.

Thrun, S. (2000). Monte Carlo POMDPs, *in* S. Solla, T. Leen and K.-R. Müller (eds), *Advances in Neural Information Processing Systems 12*, MIT Press, pp. 1064–1070.

Vidoni, P. (1999). Exponential family state space models based on a conjugate latent process, *Journal of the Royal Statistical Society* **B 61**: 213–221.

Wan, E. A. and van der Merwe, R. (2000). The Unscented Kalman Filter for nonlinear estimation, *Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC)*, IEEE, Lake Louise, Alberta, Canada.

Wan, E. A., van der Merwe, R. and Nelson, A. T. (2000). Dual Estimation and the Unscented Transformation, *in* S. Solla, T. Leen and K.-R. Müller (eds), *Advances in Neural Information Processing Systems 12*, MIT Press, pp. 666–672.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, Springer Series in Statistics, second edn, Springer-Verlag, New York.

Zaritskii, V. S., Svetnik, V. B. and Shimelevich, L. I. (1975). Monte-Carlo techniques in problems of optimal information processing, *Automation and Remote Control* **36**(3): 2015–2022.