# IMPROVED UNSCENTED KALMAN SMOOTHING FOR STOCK VOLATILITY ESTIMATION

Onno Zoeter, Alexander Ypma, and Tom Heskes
SNN, University of Nijmegen
Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands
E-mail: orzoeter,ypma,tomh@snn.kun.nl
Web: www.snn.kun.nl

**Abstract. We introduce a novel approximate inference algorithm for non-linear dynamical systems. The algorithm is based upon expectation propagation and Gaussian quadrature. The first forward pass is strongly related to the unscented Kalman filter. It improves upon unscented Kalman filtering by only making Gaussian approximations in the latent and not in the observation space.**

**Smoothed estimates can be found without inverting latent space dynamics and can be improved by iteration. Multiple forward and backward passes make it possible to improve local approximations and make them as consistent as possible.**

**We demonstrate the validity of the approach with an interesting inference problem in stochastic stock volatility models. The traditional unscented Kalman filter is ill suited for this problem: it can be proven that the traditional filter effectively never updates prior beliefs. The novel algorithm gives good results and improves with iteration.**

## INFERENCE IN STOCHASTIC VOLATILITY MODELS

In 1973, Black, Scholes and Merton [1, 7] reasoned that under certain idealized market assumptions the prices of stocks and derivatives on those stocks are coupled. A derivative is a financial product whose pay-off is determined by the price of another asset. A European call option for instance entitles the holder the right to buy a certain stock for a specific price, the strike price, at a specific moment in the future, the maturity time. The effective pay-off at maturity time is the difference between the stock price and the strike price if the former exceeds the latter, and zero otherwise.

If all the market assumptions from [1, 7] hold, the price of such an option is a deterministic function of the current price of the underlying stock, the

stock's volatility, the risk-free interest rate, the strike price and the maturity time of the option. Any other price allows traders to sell over priced and buy under priced assets and make a risk-free profit. One of the crucial assumptions is that the underlying stock $S$ follows a geometric Brownian motion

$$\frac{dS}{S} = \mu dt + \sqrt{V} dz. \tag{1}$$

In (1) $dz$ is a Brownian motion, $\mu$ is a drift and $\sqrt{V}$ is the volatility. The latter two are constant or a deterministic function of time.

It is mainly the assumption of constant volatility that does not seem to hold in practice. Equation (1) implies that daily log returns are normally distributed with constant mean and standard deviation. What is observed for most stocks is that this standard deviation (the volatility) is not constant, but is auto-correlated and mean reverting. Also the returns do not appear to come from a normal distribution but from a distribution with heavier tails.

These observations have led many researchers to formulate stochastic volatility models; models where the volatility itself follows an (unobserved) stochastic process. In our experiments we will use a discrete time model that is inspired by the model from [3].

We denote the log returns with $y_t = \log \frac{S_t}{S_{t-1}}$, where $t$ ranges over exchange closing times. As mentioned previously, if the volatility would be constant, the $y_t$'s would be identically, independently and normally distributed. We keep the mean of this distribution fixed at $\mu$, but treat the volatility as a random variable itself. We define $x_t$ to be the log of the volatility at time $t$. It follows an AR process with a base level $l$ to which it reverts with rate $a$. The complete model reads

$$x_t = a(x_{t-1} - l) + l + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, q), \tag{2}$$

$$y_t = \mathrm{e}^{x_t} \eta_t + \mu, \quad \eta_t \sim \mathcal{N}(0, 1)\ . \tag{3}$$

In the above $\mathcal{N}(m, v)$ denotes the Gaussian probability distribution with mean $m$ and variance $v$. All disturbances $\epsilon_t$ and $\eta_t$ are assumed to be independently drawn. At $t = 1$, $x_1 \sim \mathcal{N}(m_1, v_1)$. Figure 1 shows an artificial dataset generated from this model.

The latent state dynamics are linear and Gaussian, but the observation model is non-linear. As a result exact inference of filtered and smoothed posteriors, $p(x_t|y_{1:t})$ and $p(x_t|y_{1:T})$, with $T > t$, is infeasible. In this article we will introduce an approximate inference technique that is closely related to the unscented Kalman filter [4] but significantly improves upon it.

We will consider the following general class of non-linear models

$$x_t = f(x_{t-1}, \epsilon_t), \quad \epsilon_t \sim \mathcal{N}(0, q) \tag{4}$$

$$y_t = g(x_t, \eta_t), \quad \eta_t \sim \mathcal{N}(0, r)\ . \tag{5}$$

The only requirement on $g$ is that $p(y_t|x_t) = \int g(x_t, \eta_t) p(\eta_t)$ with $p(\eta_t)$ Gaussian, can be computed analytically. For this to hold it is sufficient (but not necessary) that $g$ is linear in $\eta_t$.

## GAUSSIAN QUADRATURE

In both the traditional unscented Kalman filter and our proposed algorithm local integrals are approximated using Gaussian quadrature. Gaussian quadrature is a general technique to approximate integrals of the form $\int h(x)K(x)dx$, where $K(x)$ is a known non-negative function. In the inference algorithms $K(x)$ will be a Gaussian kernel.

Based on $K(x)$, $n$ points $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and corresponding weights $W_1, \ldots, W_n$ are chosen such that

$$\int K(x)h(x)dx \approx \sum_{i=1}^{n} h(\mathcal{X}_i)W_i \ ,$$

is exact if $h(x)$ is a polynomial of degree at most $2n-1$. The constraint that the approximation is exact for polynomials results in a set of coupled non-linear equations. The position of the points is determined up to a common scale factor which determines the locality of the approximation. In the canonical unscented filter 3 points are used with a scale factor of $\sqrt{3}$. In our experiments we will use the same scale, but 5 points since the non-linearity seems to be too severe to be correctly approximated using 3 points.

Multi-dimensional integrals can be computed on a grid with positions dictated by the one-dimensional points, or more sophisticated rules resulting from correctness constraints on monomials can be used. See e.g. [6] for a general introduction.

## THE UNSCENTED KALMAN FILTER

The traditional unscented filter is a recursive algorithm based upon Gaussian approximations $\tilde{p}(x_t|y_{1:t})$ of the exact filtered state posteriors $p(x_t|y_{1:t})$. Throughout this article we will use the notational convention that $\tilde{p}(X)$ is an approximation of $p(X)$. Given an approximation $\tilde{p}(x_{t-1}|y_{1:t-1}) = \mathcal{N}(m_{t-1|t-1}, v_{t-1|t-1})$, a new observation $y_t$ is incorporated using a prediction step and a measurement update step.

**Prediction step** $\tilde{p}(x_t|y_{1:t-1}) = \mathcal{N}(m_{t|t-1}, v_{t|t-1})$, with

$$\begin{aligned} m_{t|t-1} &= \iint f(x_{t-1}, \epsilon_t)p(\epsilon_t)\tilde{p}(x_{t-1}|y_{1:t-1})dx_{t-1}d\epsilon_t \ , \\ v_{t|t-1} &= \iint (f(x_{t-1}, \epsilon_t) - m_{t|t-1})^2 p(\epsilon_t)\tilde{p}(x_{t-1}|y_{1:t-1})dx_{t-1}d\epsilon_t . \end{aligned}$$

The above integrals are of the form $\int h(x)K(x)dx$, where $K(x)$ is a Gaussian kernel and can be approximated using Gaussian quadrature.

This is done by determining points and weights $\{\mathcal{X}_i, W_i\}$ for the state $x_{t-1}$ augmented with the Gaussian disturbance $\epsilon_t$

$$\begin{bmatrix} x_{t-1|t-1} \\ \epsilon_t \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} m_{t-1|t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} V_{t-1|t-1} & 0 \\ 0 & q \end{bmatrix} \right)$$

using one of the methods described above. The predicted mean and variance are then

$$\begin{aligned} m_{t|t-1} &= \sum_i W_i f(\mathcal{X}_i(1), \mathcal{X}_i(2)) \\ v_{t|t-1} &= \left[\sum_i W_i f(\mathcal{X}_i(1), \mathcal{X}_i(2))^2\right] - \left(m_{t|t-1}\right)^2 . \end{aligned}$$

In the above $\mathcal{X}_i(1)$ denotes the state component and $\mathcal{X}_i(2)$ the noise disturbance component in the vector $\mathcal{X}_i$.

**Measurement update** Follow the linear filtering paradigm approximately.

1. Compute $\tilde{p}(y_t|x_t) = \mathcal{N}(m^y_{t|t-1}, v^y_{t|t-1})$ in the same way as the latent state prediction with points taken from the augmented state $[x_t, \eta_t]^\top$.
2. Also compute the covariance

$$v^{xy}_{t|t-1} \equiv \iint (x_t - m_{t|t-1})(g(x_t, \eta_t) - m^y_{t|t-1})\tilde{p}(x_t|y_{1:t-1})p(\eta_t)dx_t d\eta_t \tag{6}$$

   using points from the augmented state.
3. Compute the Kalman gain $K_t \equiv \frac{v^{xy}_{t|t-1}}{v^y_{t|t-1}}$ and update the latent state prediction as in the Kalman filter

$$\begin{aligned} m_{t|t} &= m_{t|t-1} + K_t(y_t - m^y_{t|t-1}) && (7) \\ v_{t|t} &= v_{t|t-1} - (K_t)^2 v^y_{t|t-1} . && (8) \end{aligned}$$

In the measurement update step $p(y_t|y_{1:t-1})$ and $p(x_t, y_t|y_{1:t-1})$ are approximated by Gaussians and the update follows the linear filtering paradigm. For models in which $x_t$ and $y_t$ are uncorrelated in the predictive distribution $p(x_t, y_t|y_{1:t-1})$, this will lead to poor results. In such models a Gaussian approximation of $p(x_t, y_t|y_{1:t-1})$ will render $x_t$ and $y_t$ independent. As a result the Kalman gain will be 0 and the unscented filter effectively never updates prior beliefs.

This phenomenon occurs in the stochastic volatility model. It falls in a class of models where the observation model has the form

$$g(x, \eta) = g_x(x)g_\eta(\eta) + c , \tag{9}$$

with $\int g_\eta(\eta)p(\eta) = 0$, and $c$ a constant. For this class we have that

$$v^{xy}_{t|t-1} \equiv \langle xy \rangle - \langle x \rangle \langle y \rangle = \langle x \rangle \langle y \rangle - \langle x \rangle \langle y \rangle = 0 , \tag{10}$$

and hence that (6) is 0. As a result, the Kalman gain $K_t$ is 0 and the updates (7) and (8) effectively do not take place. Writing out the integrals implied by (10) easily gives the required results.

The straight line in the top plot in Figure 1 shows the break down of the unscented filter in this model.

## ONE-STEP UNSCENTED KALMAN FILTERING

The extra Gaussian approximation of $p(x_t, y_t|y_{1:t-1})$ in the traditional unscented filter is not necessary. We give the measurement update for univariate problems below but extensions to multivariate problems are straightforward.

**Measurement update** $\tilde{p}(x_t|y_{1:t}) = \mathcal{N}(m_{t|t}, v_{t|t})$, with

$$m_{t|t} = \int x_t \frac{p(y_t|x_t)\tilde{p}(x_t|y_{1:t-1})}{Z_t} dx_t \ , \tag{11}$$

$$Z_t = \int p(y_t|x_t)\tilde{p}(x_t|y_{1:t-1}) dx_t \ , \tag{12}$$

$$v_{t|t} = \int (x_t - m_{t|t})^2 \frac{p(y_t|x_t)\tilde{p}(x_t|y_{1:t-1})}{Z_t} dx_t \ . \tag{13}$$

These integrals can again be approximated using Gaussian quadrature. Monomial points and weights $\{\mathcal{X}_i, W_i\}$ are determined for the state $x_t \sim \mathcal{N}(m_{t|t-1}, v_{t|t-1})$, and the mean and variance are updated as

$$Z_t = \sum_i W_i p(y_t|\mathcal{X}_i) \tag{14}$$

$$m_{t|t} = \sum_i W_i \mathcal{X}_i \frac{p(y_t|\mathcal{X}_i)}{Z_t} \tag{15}$$

$$v_{t|t} = \left[\sum_i W_i \mathcal{X}_i^2 \frac{p(y_t|\mathcal{X}_i)}{Z_t}\right] - \left(m_{t|t}\right)^2 \ . \tag{16}$$

In (11) to (13) (and hence (14) to (16)) we have assumed that the integral

$$p(y_t|x_t) = \int p(y_t|x_t, \eta_t) p(\eta_t) d\eta_t,$$

with $p(y_t|x_t, \eta_t) \equiv \delta_{y_t = g(x_t, \eta_t)}$ a Kronecker delta function, can be done analytically. As mentioned previously, this holds if $g$ is linear in $\eta_t$. So the important class of a non-linear mapping with additive Gaussian noise can be treated in this way. More complex observation models leading to $\chi^2$, $t$, or $F$ distributions are on the boundary of what can be handled by the one-step filter.

The product of all the normalization constants

$$\prod_{t=1}^{T} Z_t = \prod_{t=1}^{T} \tilde{p}(y_t|y_{1:t-1}) \approx p(y_{1:t}) \ ,$$

gives an approximation of the likelihood.

Figure 1 presents a result of the one-step filter using 5 monomial points.

## ITERATIVE UNSCENTED KALMAN SMOOTHING

Using the expectation propagation framework [8] we can formulate a symmetric smoothing pass for the general class of models we are considering, without inverting the latent state dynamics.

We will give a brief introduction to expectation propagation and introduce some notation, but refer the interested reader to [8] and [2] for more details.

The required joint posterior over all latent states can be represented as a product over factors $\Psi_t$ defined as

$$\begin{aligned} \Psi_1(x_1) &\equiv p(y_1|x_1)p(x_1) \\ \Psi_t(x_{t-1,t}) &\equiv p(y_t|x_t)p(x_t|x_{t-1}) \ , \end{aligned}$$

such that

$$p(x_{1:T}|y_{1:T}) \propto \prod_{t=1}^{T} \Psi_t(x_{t-1,t}) \ . \tag{17}$$

Any required marginal can be computed from this joint by integration. However, computing the product in (17) explicitly is computationally too intensive and the required integrals cannot be done analytically.

To get approximate results, an approximating family $q(x_{1:T})$ is introduced. For the unscented smoother we choose $q(x_{1:T}) = \prod_t q(x_t)$, $q(x_t) \equiv \mathcal{N}(m_{t|T}, v_{t|T})$, a fully factorized Gaussian distribution. The algorithm is initialized with arbitrary approximations of the factors $\Psi_t$ such that their product is a member of $q(x_{1:T})$. Since $q(x_{1:T})$ factors, the approximation of $\Psi_t(x_{t-1,t})$ factors into a contribution to $q(x_{t-1})$ and a contribution to $q(x_t)$. We use the notation

$$\Psi_t(x_{t-1,t}) \approx c_t \beta_{t-1}(x_{t-1})\alpha_t(x_t) \ ,$$

such that $q(x_t) \propto \alpha_t(x_t)\beta_t(x_t)$. In the above $c_t$ emphasizes that the product of $\alpha_t$ and $\beta_t$ need not be normalized. The $\alpha_t$ and $\beta_t$ are often referred to as messages, and are general Gaussian potentials. Readers familiar with the HMM forward-backward algorithm can keep in mind that the $\alpha_t$ and $\beta_t$ messages have a similar interpretation here as they do in the HMM algorithms.

In every update step the approximation of one of the factors $\Psi_t$ is removed by division and replaced by the exact factor. As a result the new combination is not in the chosen approximating family. The intermediate result $r(x_{t-1,t})$ is projected back onto the family by minimizing the Kullback-Leibler divergence

$$q^{\text{new}}(x_{t-1,t}) = \min_q \mathrm{KL}\left(r(x_{t-1,t})||q(x_{t-1,t})\right) \equiv \min_q \int r(x_{t-1,t}) \log \frac{r(x_{t-1,t})}{q(x_{t-1,t})} \ .$$

The Gaussian distribution is a member of the exponential family. It is a general result (see e.g. [5]) that minimizing the KL divergence then boils down to matching the moments, i.e. we have to compute the mean and variance of $r(x_{t-1,t})$. The new approximation of $\Psi_t$ is then inferred by dividing $q^{\text{new}}$ by the old approximation with the approximation of $\Psi_t$ removed: $\frac{q^{\text{new}}}{\beta_{t-1}\alpha_t}$.

In principle the updates can be done in any order, but an iteration of forward-backward passes seems most logical. The updates are done as follows.

**Update of the approximation of $\Psi_t$**

1. Remove $\beta_{t-1}(x_{t-1})\alpha_t(x_t)$, the old approximation of $\Psi_t(x_{t,t-1})$, by division

$$\begin{aligned}\frac{q(x_{t-1,t})}{\beta_{t-1}(x_{t-1})\alpha_t(x_t)} &= \frac{\alpha_{t-1}(x_{t-1})\beta_{t-1}(x_{t-1})\alpha_t(x_t)\beta_t(x_t)}{\beta_{t-1}(x_{t-1})\alpha_t(x_t)} \\ &= \alpha_{t-1}(x_{t-1})\beta_t(x_t)\ .\end{aligned}$$

2. Put in the exact factor $\Psi_t(x_{t,t-1})$

$$r(x_{t-1,t}) = \frac{\alpha_{t-1}(x_{t-1})\Psi_t(x_{t-1,t})\beta_t(x_t)}{Z_{t|T}}\ ,$$

with $Z_{t|T} \equiv \int\int \alpha_{t-1}(x_{t-1})\Psi_t(x_{t-1,t})\beta_t(x_t)dx_{t-1}dx_t$.

3. Project back onto the chosen approximating family. For $q(x_{t-1})$ this becomes

$$\begin{aligned}q^{\text{new}}(x_{t-1}) &= \mathcal{N}(m^{\text{new}}_{t-1|T}, V^{\text{new}}_{t-1|T}), \text{with} \\ m^{\text{new}}_{t-1|T} &= \iint x_{t-1}\frac{\alpha_{t-1}(x_{t-1})\Psi_t(x_{t,t-1})\beta_t(x_t)}{Z_{t|T}}dx_{t-1}dx_t\ , \\ Z^{\text{new}}_{t|T} &= \iint \alpha_{t-1}(x_{t-1})\Psi_t(x_{t,t-1})\beta_t(x_t)dx_{t-1}dx_t\ , \\ v^{\text{new}}_{t-1|T} &= \iint (x_{t-1}-m_{t-1|T})^2\frac{\alpha_{t-1}(x_{t-1})\Psi_t(x_{t,t-1})\beta_t(x_t)}{Z_{t|T}}dx_{t-1}dx_t\ ,\end{aligned}$$

and analogously for $q(x_t)$.

4. Infer the contribution of $\Psi_t$ by division

$$\begin{aligned}\beta^{\text{new}}_{t-1}(x_{t-1}) &= \frac{q^{\text{new}}(x_{t-1})}{\alpha_{t-1}(x_{t-1})} \\ \alpha^{\text{new}}_t(x_t) &= \frac{q^{\text{new}}(x_t)}{\beta_t(x_t)}\ .\end{aligned}$$

Combining the above steps 1–3 we get

$$m^{\text{new}}_{t-1|T} = \iint x_{t-1}\frac{q(x_{t-1})q(x_t)\Psi_t(x_{t-1,t})}{\beta_{t-1}(x_{t-1})\alpha_t(x_t)Z_{t|T}}dx_{t-1}dx_t\ , \tag{18}$$

and similarly for $Z^{\text{new}}_{t|T}$, $v^{\text{new}}_{t-1|T}$, $m^{\text{new}}_{t|T}$ and $v^{\text{new}}_{t|T}$. In (18) we can identify the required form of a Gaussian integral $y = \int\int h(x_{t-1,t})K(x_{t-1,t})dx_{t-1}dx_t$ with

$$h(x_{t-1,t}) = \frac{\Psi_t(x_{t-1,t})}{\beta_{t-1}(x_{t-1})\alpha_t(x_t)Z_{t|T}} \tag{19}$$

$$K(x_{t-1,t}) = q(x_{t-1})q(x_t)\ . \tag{20}$$

So the required local approximations can be done using Gaussian quadrature.

For simplicity we have assumed that the transition model is "easy" (additive Gaussian noise). In general $\Psi_t$ would also be a function of $\epsilon_t$ and the required moments such as (18) are found by also integrating over $\epsilon_t$. In (20) we would then have $K(x_{t-1,t}, \epsilon_t) = q(x_{t-1})q(x_t)p(\epsilon_t)$.

One can verify that the filtering algorithm from the previous section is a first forward pass of the algorithm described above with a suitable choice for the messages $\alpha_t$ and $\beta_t$, namely with $\beta_t = 1$ and $\alpha_t$ the prediction as computed in the filtering algorithm.

## EXPERIMENTS

Figure 1 presents a typical result from an experiment with artificially generated data. The bottom plot shows the daily closing values of an artificial stock. The middle plot shows the corresponding log returns as they were drawn from the model (solid) and the predictive mean and 2 standard deviations errors bars. The top plot shows the drawn log volatilities, the traditional unscented and one-step unscented filtered estimates, and smoothed posteriors after 10 iterations. The traditional unscented filter gives non-sensical results, the one-step version gives quite accurate approximations. The smoothed iterated estimates result in a slight improvement over the filtered results.

For single slices it is feasible to compute near exact results by using a very fine grid. The left plot in Figure 2 shows a typical measurement update with 5 monomial points of a prior with mean 0 and variance 1. The observed value in the example of the left plot of Figure 2 was 3 which is slightly over 2 times the standard deviation of the predictive distribution away from its predicted value. Note that the posterior is slightly skewed but that a Gaussian approximation seems to be valid for the current application. The right plot in Figure 2 gives a general picture for several observations (only positive values are shown, the plot is symmetric around 0). On the x-axis are observations 0 to 5 times the standard deviation away from the mean of $p(y_1)$. In this particular case, the method seems to be valid at least for observations lying 5 standard deviations away from their expected value. However, for extreme outliers the method degrades. This is due to the fact that the quadrature points lie in an area of the posterior that gets negligible weight. It is possible to detect such a degradation by checking the variance in $\frac{W_i \mathcal{Y}_i}{Z_{t|T}}$, if only one or a few of these points get non-negligible weights this indicates a degradation. We are currently investigating possibilities to make the algorithm robust against such extreme events.

## SUMMARY

We have presented a one-step unscented Kalman filter, an analogous backward pass and an iteration scheme to improve smoothed posteriors. The

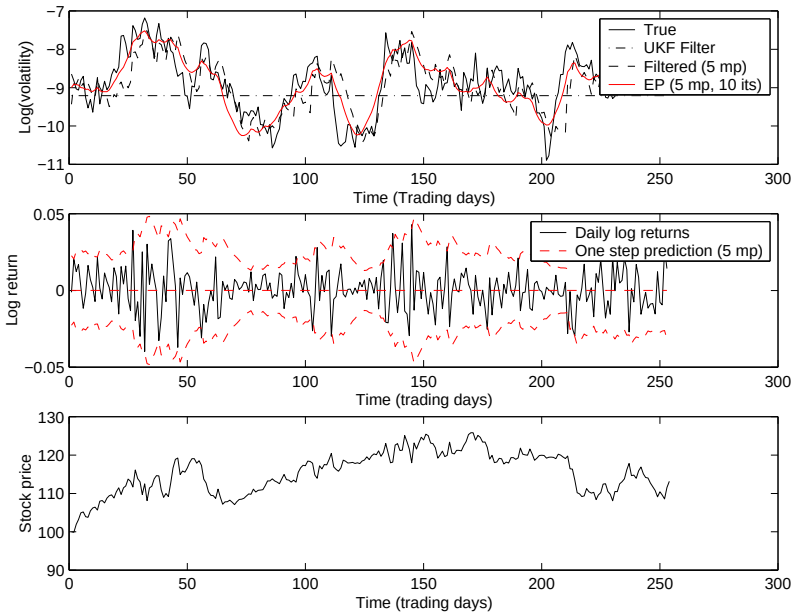


Figure 1: Results from inference on an artificial dataset. The bottom plot shows closing prices of an artificial stock. The middle plot shows the log-ratios of the closing prices (solid line) and one step ahead predictions (dashed line) and 2 standard deviations error bars. The prediction is correctly constant at $\mu$, the error bars show that the proposed filter correctly captures the heteroskedasticity of the series. The top plot shows the artificially generated true volatilities and the various approximate inference results. The traditional unscented filter only propagates the prior and gives nonsensical filtered results.

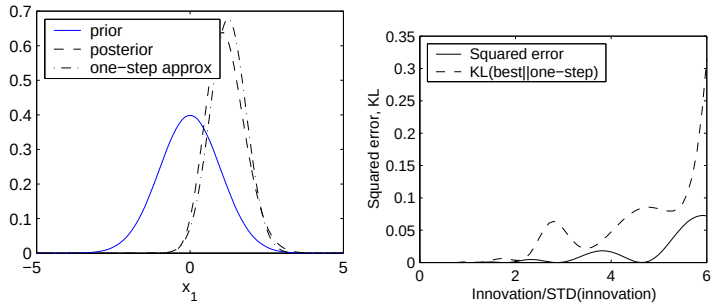


Figure 2: The left plot shows a typical measurement update error; prior $x_1 \sim \mathcal{N}(0,1)$ (solid), exact posterior $p(x_1|y_1 = 3)$ computed using a very fine grid (dashed), Gaussian approximation from the one-step unscented filter using 5 points (dash-dotted). The mean squared error in the posterior means of such updates and the KL-divergence between best and approximated Gaussian is shown in the right plot as a function of the observation $y_t$. The x-axis is normalized by the standard deviation of the predictive distribution. The value $y = 3$ from the left plot corresponds to little over 2 sd.

approaches seem to work very well for the inference problem in stochastic volatility models that we considered. Interestingly enough, the factored form of the observation dynamics in the volatility model makes the traditional unscented Kalman filter break down.

Given the success of the novel algorithm on simple artificial problems we hope to apply these techniques to real exchange data and extend the model to incorporate stock volatility correlation in higher dimensional models.

In principle the introduced expectation propagation variant works for models with an arbitrary topology. It can therefore be seen as an analog to Laplace propagation [9]. We aim to investigate the nature of fixed points of such an algorithm and test the scalability of the algorithm on higher dimensional problems with wilder non-linearities and complexer structure.

For a specific model it should be possible to adapt the general approach we have described and replace the Gaussian kernels in the quadrature approximation with kernels that take advantage of properties of the transition and observation models. Also, the best position of the quadrature points could be determined with knowledge of the observation $y_t$ to make the measurement update robust against outliers.

## REFERENCES


[1] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," **Journal of Political Economy**, vol. 81, pp. 637–654, 1973.

[2] T. Heskes and O. Zoeter, "Generalized belief propagation for approximate inference in hybrid Bayesian Networks," in **Proceedings AISTATS**, 2003.

[3] J. Hull and A. White, "The Pricing of Options on Assets with Stochastic Volatilities," **The Journal of Finance**, vol. 42, no. 2, pp. 281–300, June 1987.

[4] S. Julier and J. K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in **Proceedings of AeroSense: The 11th Int Symp on Aerospace/Defence Sensing, Simulation and Controls**, 1997.

[5] S. Lauritzen, **Graphical Models**, Oxford University Press, 1996.

[6] U. Lerner, **Hybrid Bayesian Networks for Reasoning about Complex Systems**, Ph.D. thesis, Stanford University, 2002.

[7] R. Merton, "On the Pricing of Corporate Debt; the Risk Structure of Interest Rates," **Journal of Finance**, vol. 29, pp. 449–470, 1974.

[8] T. Minka, "Expectation propagation for approximate Bayesian inference," in **Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2001)**, Morgan Kaufmann Publishers, 2001.

[9] A. Smola, V. Vishwanathan and E. Eskin, "Laplace Propagation," in S. Thrun, L. Saul and B. Schölkopf (eds.), **Advances in Neural Information Processing Systems 16**, Cambridge, MA: MIT Press, 2004.