Generalized Gauss-Hermite Filtering for Multivariate Diffusion Processes

Hermann Singer
FernUniversität in Hagen *

November 29, 2006

Abstract

The generalized Gauss-Hermite-filter (GGHF) is implemented in the multivariate case. We utilize a Hermite expansion of the filter density and Gauss-Hermite integration for the computation of expectation values in the time and measurement update (moment equations and Bayes formula). The algorithm is successfully applied to the Bayesian estimation of a volatility parameter, where filters based on two moments (EKF, UKF, GHF) fail. Moreover, the stochastic volatility model of Scott (1987) is treated.

Key Words: Multivariate stochastic differential equations; Nonlinear systems; Discrete time measurements; Continuous-discrete state space model; Gaussian filter; Hermite expansion; Stochastic volatility.

1 Introduction

In a recent article the Gaussian filter was generalized by using a scalar Hermite expansion of the filter density with leading Gaussian term (Singer 2006a). Thus, integrals appearing in the time and measurement update can be computed by Gauss-Hermite integration, as in the Gaussian filter (cf. Ito and Xiong, 2000). The restrictive assumption of a Gaussian filter density is dropped and arbitrary functional forms can be modeled by inclusion of 3rd and higher order moments (skewness, kurtosis etc.). The Gaussian filter is contained as a special case. In this paper, the general multivariate case is

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany

derived. Similar algorithms have been developed by Challa et al. (2000), but we formulate the time update as integro-differential equations solved stepwise by using Gauss-Hermite integration. Moreover, computation of the measurement update (Bayes formula) is improved. We use the normal correlation update as Gaussian weight function in the Gauss-Hermite quadrature to achieve higher numerical accuracy. Finally it is shown how the volatility parameter of an Ornstein-Uhlenbeck process can be estimated sequentially. In this problem Gaussian filters fail since they cannot model the strong non-gaussianity of the posterior density of the volatility. Similar considerations apply to the stochastic volatility models of Scott (1987) or Hull and White (1987).

2 State Space Model and Filter Equations

2.1 Nonlinear Continuous-Discrete State Space Model

The nonlinear *continuous-discrete state space model* is defined as (Jazwinski, 1970)

$$dy(t) = f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t) \tag{1}$$

where discrete measurements $z_i := z(t_i)$ are taken at times $\{t_0, t_1, \ldots, t_T\}$ and $t_0 \leq t \leq t_T$ according to the measurement equation

$$z_i = h(y(t_i), t_i, \psi) + \epsilon_i. \tag{2}$$

In state equation (1), $W(t)$ denotes an r-dimensional Wiener process and the state is described by the p-dimensional state vector $y(t)$. It fulfils a system of stochastic differential equations in the sense of Itô (cf. Arnold, 1974) with random initial condition $y(t_0) \sim p_0(y, \psi)$. The functions $f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^r$ are called drift and diffusion coefficients, respectively. In measurement equation (2), $\epsilon_i \sim N(0, R(t_i, \psi))i.d.$ is a k-dimensional discrete time white noise process (measurement error). Parametric estimation is based on the u-dimensional parameter vector ψ. For notational simplicity, deterministic control variables $x(t)$ are absorbed in the time argument t. Moreover, the functions f and g may also depend on nonanticipative measurements $Z^t = \{z(s)|s \leq t\}$ and h, R may depend on lagged measurements $Z^{i-1} := Z^{t_{i-1}} = \{z(s)|s \leq t_{i-1}\}$ allowing continuous time ARCH specifications. In the linear case, the system is conditionally gaussian (cf. Liptser and Shiryayev, 2001, ch. 11). This dependence will be dropped in the sequel.

2.2 Exact Continuous-Discrete Filter

The exact time and measurement updates of the continuous-discrete filter are given by the recursive scheme (Jazwinski, 1970) for the conditional density $p(y,t|Z^i)$:

Time update:

$$\begin{aligned}\frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i)\;; t \in [t_i, t_{i+1}] \qquad (3)\\ p(y,t_i|Z^i) &:= p_{i|i}\end{aligned}$$

Measurement update:

$$\begin{aligned}p(y_{i+1}|Z^{i+1}) &= \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} \qquad (4)\\ &:= p_{i+1|i+1}\\ p(z_{i+1}|Z^i) &= \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)dy_{i+1}, \qquad (5)\end{aligned}$$

$i = 0, \ldots, T-1$, where

$$F(\cdot) = -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t,\psi)\cdot] + \tfrac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t,\psi)\cdot] \qquad (6)$$

is the Fokker-Planck operator, $\Omega = gg'$, $Z^i = \{z(t)|t \leq t_i\}$ are the observations up to time t_i and $p(z_{i+1}|Z^i)$ is the likelihood function of observation z_{i+1}. The first equation describes the time evolution of the conditional density $p(y,t|Z^i)$ given information up to the last measurement and the measurement update is a discontinuous change due to new information using the Bayes formula. The above scheme is exact, but can be solved explicitly only for the linear case where the filter density is Gaussian with conditional moments $\mu(t|t_i) = E[y(t)|Z^i]; \Sigma(t|t_i) = \mathrm{Var}[y(t)|Z^i]$ and for some special cases (Daum filter; Daum, 1986).

2.3 Exact Moment Equations

In the generalized Gauss-Hermite filter, instead of solving the time update equations (3) for the conditional density directly, the moment equations for K moments are solved approximately. The moments can be used to compute the expansion coefficients of the density function (sect. 4). Using the Euler approximation for the SDE (1), one obtains in a short time interval δt $(\delta W(t) := W(t+\delta t) - W(t))$

$$y(t+\delta t) = y(t) + f(y(t),t)\delta t + g(y(t),t)\delta W(t). \qquad (7)$$

Taking the expectation $E[...|Z^i]$ one gets the moment equation

$$\mu(t+\delta t|t_i) \quad = \quad \mu(t|t_i) + E[f(y(t),t)|Z^i]\delta t \tag{8}$$

or in the limit $\delta t \to 0$

$$\dot{\mu}(t|t_i) = E[f(y(t),t)|Z^i]. \tag{9}$$

The 2nd central moment $m_2 = \Sigma$ (dropping the condition Z^i)

$$m_{2,jk}(t) \quad := \quad E[(y_j(t) - \mu_j(t))(y_k(t) - \mu_k(t))] := E[M_{2,jk}(t)], \tag{10}$$

$j, k = 1, ..., p$ fulfils

$$\begin{aligned} m_{2,jk}(t+\delta t) \quad = \quad & E[(y_j + f_j\delta t - \mu_j(t+\delta t) + g_{jj'}\delta W_{j'}) \times \\ & (y_k + f_k\delta t - \mu_k(t+\delta t) + g_{kk'}\delta W_{k'})], \end{aligned} \tag{11}$$

where $y_j := y_j(t), f_j := f_j(y(t),t)$ etc. and $\mu_j(t+\delta t) = E[y_j + f_j\delta t] := \mu_j + E[f_j]\delta t$. Since the increments of the Wiener process are independent of the terms at time t one obtains, introducing centered variables $y_j^* := y_j - E(y_j), f_j^* := f_j - E(f_j)$,

$$\begin{aligned} m_{2,jk}(t+\delta t) \quad = \quad & m_{2,jk}(t) + E[y_j^* f_k^* + y_k^* f_j^* + \Omega_{jk}]\delta t + \\ + \quad & E[f_j^* f_k^*]\delta t^2. \end{aligned} \tag{12}$$

In the limit $\delta t \to 0$ we have

$$\dot{m}_{2,jk}(t) \quad = \quad E[y_j^* f_k^*] + E[y_k^* f_j^*] + E[\Omega_{jk}]. \tag{13}$$

The exact moment equations (9, 13) are not differential equations, however, since they depend on the unknown conditional density $p(y,t|Z^i)$. For the Gaussian filter, $K = 2$ moments are used, and the density is approximated by $p(y) = \phi(y;\mu,\Sigma)$. For the generalized Gaussian filter, $K > 2$ moments with a density

$$p(y) = \phi(y;\mu,\Sigma)\sum_{k=0}^{K} c_k H_k(\Sigma^{-1/2}(y-\mu)) \tag{14}$$

(Hermite expansion) are utilized (sect. 4). In both cases, the expectation values can be computed by Gauss-Hermite integration. An Euler approximation of the moment equation (13) yields (12), but without quadratic terms $O(\delta t^2)$. The update (12) is numerically more stable, since it is positive semidefinite. Higher order moments, as required by the Hermite expansion (14), can be computed as follows. The third moment (skewness) is given as, using the abbreviations $\gamma_j := g_{jj'}\delta W_{j'}$, $a_j := y_j^* + f_j^*\delta t = y_j - E(y_j) + (f_j - E(f_j))\delta t$

$$m_{3,jkl}(t+\delta t) \quad = \quad E[(a_j + \gamma_j)(a_k + \gamma_k)(a_l + \gamma_l)]. \tag{15}$$

Collecting the terms yields

$$
\begin{aligned}
m_{3,jkl}(t+\delta t) &= E[a_j a_k a_l] + E[a_j \gamma_k \gamma_l] + E[\gamma_j a_k \gamma_l] + E[\gamma_j \gamma_k a_l] \qquad (16) \\
&= E[a_j a_k a_l] + (E[a_j \Omega_{kl}] + E[a_k \Omega_{jl}] + E[a_l \Omega_{jk}])\delta t,
\end{aligned}
$$

since $E[a_j \gamma_k \gamma_l] = E[a_j g_{kk'} g_{ll'}]E[\delta W_{k'} \delta W_{l'}]$ and $E[\delta W_{k'} \delta W_{l'}] = \delta_{k'l'}\delta t$ with the Kronecker delta symbol $\delta_{jk} = 1$ if $j = k$ and 0 elsewhere. Furthermore, $E[a_i a_j \gamma_k] = E[a_i a_j g_{kk'}]E[\delta W_{k'}] = 0$, since the increments of the Wiener process are independent of y_j.
In the limit $\delta t \to 0$ only terms of order $O(\delta t)$ survive yielding

$$
\begin{aligned}
\dot m_{3,jkl}(t) &= E[y_j^* y_k^* f_l^*] + E[y_j^* f_k^* y_l^*] + E[f_j^* y_k^* y_l^*] + \qquad (17) \\
&+ E[y_j^* \Omega_{kl}] + E[y_k^* \Omega_{jl}] + E[y_l^* \Omega_{jk}].
\end{aligned}
$$

In the scalar case the general formula is

$$
\dot m_k(t) = kE[(y^*)^{k-1} f^*] + \tfrac{k(k-1)}{2} E[(y^*)^{k-2}\Omega], \qquad (18)
$$

(Singer 2006b), e.g.

$$
\dot m_3(t) = 3E[(y^*)^2 f^*] + 3E[y^* \Omega], \qquad (19)
$$

thus the multivariate formula can be obtained by combining all different indices with the symbolic notation

$$
\dot m_{3,jkl}(t) = (3)E[y_j^* y_k^* f_l^*] + (3)E[y_j^* \Omega_{kl}] \qquad (20)
$$

(cf. Stratonovich, 1992, p. 27). The number in parantheses is the number of similar terms which differ only in the order of subscripts. The general formula reads

$$
\begin{aligned}
\dot m_{k,j_1 j_2,\ldots,j_k}(t) &= (k)E[y_{j_1}^* \ldots y_{j_{k-1}}^* f_{j_k}^*] + \\
&+ \left(\tfrac{k(k-1)}{2}\right) E[y_{j_1}^* \ldots y_{j_{k-2}}^* \Omega_{j_{k-1} j_k}] \qquad (21)
\end{aligned}
$$

and

$$
\begin{aligned}
m_{k,j_1 j_2,\ldots,j_k}(t+\delta t) &= E\left[\prod_{l=1}^{k} (a_{j_l} + \gamma_{j_l})\right] \\
&= \sum_{l=0}^{k} \left(\binom{k}{l}\right) E[a_{j_1} \ldots a_{j_{k-l}} \gamma_{j_{k-l+1}} \ldots \gamma_{j_k}]. \qquad (22)
\end{aligned}
$$

For example, the term $k = 3, l = 1$ is $\left(\binom{3}{1}\right) E[a_{j_1} a_{j_2} \gamma_{j_3}] = E[a_{j_1} a_{j_2} \gamma_{j_3} + a_{j_1} a_{j_3} \gamma_{j_2} + a_{j_2} a_{j_3} \gamma_{j_1}] = 0$. The expectations $E[a_{j_1} \ldots a_{j_{k-l}} \gamma_{j_{k-l+1}} \ldots \gamma_{j_k}]$ can be simplified by using the independence of δW_j from terms containing y_k.

Inserting the higher order Gaussian moments of δW_j, all expectations can be expressed in terms of $a_j(y)$ and $\Omega_{kl}(y)$. For example, $E[a_j a_k \gamma_l \gamma_m] = E[a_j a_k \Omega_{lm}]\delta t$ and

$$E[\gamma_j \gamma_k \gamma_l \gamma_m] \;=\; E[\Omega_{jk}\Omega_{lm} + \Omega_{jl}\Omega_{km} + \Omega_{jm}\Omega_{kl}]\delta t^2. \tag{23}$$

The higher order Gaussian moments of δW may be computed using the characteristic function $\chi(t) = E[\exp(it'\delta W)] = \exp(-\frac{1}{2}t't\delta t)$. Formula (21) is obtained from (22) in the limit $\delta t \to 0$ keeping terms of order $O(\delta t)$.

3 Gauss-Hermite integration

The moment equations of the last section require the computation of expectations of the type $E[f(Y)]$, where Y is a random variable with density $p(y)$. For the Gaussian filter, one may assume that the true $p(y)$ is approximated by a Gaussian distribution $\phi(y; \mu, \sigma^2)$ with the same mean μ and variance σ^2. Then, the Gaussian integral

$$E_\phi[f(Y)] \;=\; \int f(y)\phi(y;\mu,\sigma^2)dy = \int f(\mu + \sigma z)\phi(z; 0, 1)dz \tag{24}$$

$$\approx \; \sum_{l=1}^{m} f(\mu + \sigma\zeta_l)w_l = \sum_{l=1}^{m} f(\eta_l)w_l \tag{25}$$

may be approximated by Gauss-Hermite quadrature (cf. Ito and Xiong, 2000). Here, (ζ_l, w_l) are quadrature points and weights, respectively. If such an approximation is used, one obtains the Gauss-Hermite filter (GHF). Filters using Gaussian densities are called Gaussian filters (GF). More generally, the density may be approximated by the Hermite series $p(y) = \phi(y; \mu, \sigma^2) * H(z); z = (y - \mu)/\sigma$ (14) which again yields integrals w.r.t. the Gaussian density i.e. $E[f(Y)] = \int f(y)H(y)\phi(y)dy$.
In the multivariate case, the integration is performed using standardization

$$E_\phi[f(Y)] \;=\; \int f(y)\phi(y;\mu,\Sigma)dy \tag{26}$$

$$= \; \int f(\mu + \Sigma^{1/2}z)\phi(z; 0, I)dz_1...dz_p \tag{27}$$

$$\approx \; \sum_{l_1,...,l_p} f(\mu + \Sigma^{1/2}\{\zeta_{l_1}, ..., \zeta_{l_p}\})w_{l_1,...,l_p} \tag{28}$$

$$= \; \sum_{l_1,...,l_p} f(\eta_{l_1}, ..., \eta_{l_p})w_{l_1,...,l_p}, \tag{29}$$

since $\phi(z; 0, I) = \phi(z_1; 0, 1)...\phi(z_p; 0, 1)$ allows stepwise application of the univariate quadrature formula and $\{\zeta_{l_1}, ..., \zeta_{l_p}\}, l_j = 1, ..., m$, is the p-tupel of Gauss-Hermite quadrature points with weights $w_{l_1,...,l_p} = w_{l_1}...w_{l_p}$.

4 Hermite Expansion

4.1 Univariate Hermite Expansion

If the filter density strongly deviates from normality, a Fourier expansion in terms of Hermite polynomials may be utilized (Edgeworth series; cf. Kuznetsov et al., 1960, Courant and Hilbert, 1968, ch. II, 9, Abramowitz and Stegun, 1965, ch. 22, Aït-Sahalia, 2002). The filter density $p(x)$ can be expanded by using the complete set of Hermite polynomials which are orthogonal with respect to the weight function $w(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ (standard Gaussian density), i.e.

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)w(x)dx \quad = \quad n!\delta_{nm} \tag{30}$$

The Hermite polynomials $H_n(x)$ are defined by

$$\phi^{(n)}(x) \quad := \quad (d/dx)^n\phi(x) = (-1)^n\phi(x)H_n(x). \tag{31}$$

and are given explicitly as $H_0 = 1, H_1 = x, H_2 = x^2 - 1, H_3 = x^3 - 3x, H_4 = x^4 - 6x^2 + 3$ etc. Therefore, the density function $p(x)$ can be expanded as [1]

$$p(x) \quad = \quad \phi(x)\sum_{n=0}^{\infty} c_n H_n(x). \tag{32}$$

and the Fourier coefficients are given by

$$c_n \quad := \quad (1/n!)\int_{-\infty}^{\infty} H_n(x)p(x)dx = (1/n!)E[H_n(X)] \tag{33}$$

where X is a random variable with density $p(x)$. The c_n are called *quasi-moments* by Kuznetsov et al. (1960), since the characteristic function corresponding to (32) is the product of a Gaussian and a power series expansion with c_n as expansion coefficients.

The Hermite polynomials contain powers of x, so the expansion coefficients can be expressed in terms of moments $\mu_k = E[X^k]$. Since the expansion has a leading standard Gaussian density, it is more efficient to expand a standardized variable first and afterwards transform to the unstandardized density.

[1] Actually, the expansion is in terms of the orthogonal system $\psi_n(x) = \phi(x)^{1/2}H_n(x)$ (oscillator eigenfunctions), i.e. $q(x) := p(x)/\phi(x)^{1/2} = \sum_{n=0}^{\infty} c_n\psi_n(x)$, so the expansion of $q = p/\phi^{1/2}$ must converge. The function to be expanded must be square integrable in the interval $(-\infty, +\infty)$, i.e. $\int q(x)^2dx = \int \exp(x^2/2)p^2(x)dx < \infty$ (Courant and Hilbert, 1968, p. 81–82).

Using the standardized variables $Z = (X - \mu)/\sigma$ with $\mu = E[X], \sigma^2 = E[X^2] - \mu^2$, $E[Z] = 0, E[Z^2] = 1, E[Z^k] := \nu_k$ one obtains the simplified expressions $c_0 = 1, c_1 = 0, c_2 = 0$,

$$c_3 \;:=\; (1/3!)E[Z^3] = (1/3!)\nu_3 \tag{34}$$
$$c_4 \;:=\; (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(\nu_4 - 3) \tag{35}$$

and the standardized density expansion

$$p_z(z) \;:=\; \phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + ...] \tag{36}$$

which shows that the leading Gaussian term is corrected by higher order contributions containing skewness and kurtosis excess. For a standard Gaussian random variable $p_z(z) = \phi(z)$, so the coefficients $c_k, k \geq 3$ all vanish. For example, the kurtosis of Z is $E[Z^4] = 3$, so $c_4 = 0$.
Using the expansion for the standardized variable and the change of variables formula $p_x(x) = (1/\sigma)p_z(z); z = (x - \mu)/\sigma$ one obtains the desired Hermite expansion for $p_x(x)$

$$p_x(x) \;=\; \phi(x; \mu, \sigma^2) \sum_{n=0}^{\infty} c_n H_n((x - \mu)/\sigma) \tag{37}$$
$$\;:=\; \phi(x; \mu, \sigma^2) H(x) \tag{38}$$

The standardized moments $\nu_k = E[Z^k] = E[(X - \mu)^k]/\sigma^k := m_k/\sigma^k$ necessary for c_k can be expressed in terms of *centered moments*

$$m_k := E[M_k] := E[(X - \mu)^k]. \tag{39}$$

4.2 Multivariate Hermite Expansion

In the case of random vectors $X \in \mathbb{R}^p$ the multivariate (standardized) Hermite expansion

$$p_x(x) \;=\; \phi(x; \mu, \Sigma) \sum_{|n|=0}^{\infty} c_n H_n(z) \tag{40}$$
$$\;:=\; \phi(x; \mu, \Sigma) H(x), \tag{41}$$

$z = \Sigma^{-1/2}(x - \mu)$, with multiindex $n = \{n_1, ..., n_p\}; \sum n_l = |n|$ can be used. The Hermite functions $H_n(x)$ are products $H_{n_1}(x_1)...H_{n_p}(x_p)$ which are orthogonal w.r.t. the Gaussian $\phi(x; 0, I)$, i.e.

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) \phi(x; 0, I) dx \;=\; n! \delta_{nm} \tag{42}$$

or more explicitly

$$\int_{-\infty}^{\infty} H_{n_1}(x_1)...H_{n_p}(x_p)H_{m_1}(x_1)...H_{m_p}(x_p)\times$$
$$\times \quad \phi(x_1;0,1)...\phi(x_p;0,1)dx_1...dx_p = (n_1!...n_p!)\delta_{n_1m_1}...\delta_{n_pm_p}. \tag{43}$$

Thus the expansion coefficients are given by

$$\begin{aligned}
\int_{-\infty}^{\infty} p_x(x)H_m(z)dx &= \sum_{|n|=0}^{\infty} c_n \int_{-\infty}^{\infty} \phi(x;\mu,\Sigma)H_n(z)H_m(z)dx \\
&= \sum_{|n|=0}^{\infty} c_n \int_{-\infty}^{\infty} \phi(z;0,I)H_n(z)H_m(z)dz \\
&= m!c_m = (m_1!...m_p!)c_m.
\end{aligned} \tag{44}$$

Since the Hermite functions $H_{m_1}(z_1)...H_{m_p}(z_p)$ contain powers of maximum order $|m|$, the coefficients can be expressed by the standardized moments of the same order, i.e.

$$\nu_{|m|}^{m_1...m_p} = \int_{-\infty}^{\infty} p_z(z)z_1^{m_1}...z_p^{m_p}dz. \tag{45}$$

The moment is indexed by the order $|m| = m_1 + ... + m_p$ and by the exponents m_l of the several variables. The standardized moments can be computed from the centered moments of section (2.3) by using the relation $z = \Sigma^{-1/2}(y-\mu) := \Gamma(y-\mu)$ with some matrix square root, e.g. the Cholesky decomposition. One obtains

$$\begin{aligned}
\nu_{|m|}^{m_1...m_p} &= E[z_1^{m_1}...z_p^{m_p}] \\
&= (\Gamma_{1i_{11}}...\Gamma_{1i_{1m_1}})...(\Gamma_{pi_{p1}}...\Gamma_{pi_{pm_p}}) \\
&\times \; E[(y^*_{i_{11}}...y^*_{i_{1m_1}})...(y^*_{i_{p1}}...y^*_{i_{pm_p}})]
\end{aligned} \tag{46}$$

with the centered variables $y^*_j = y_j - \mu_j$. One the right hand side, we have the centered moments in index notation (lower subscript)

$$m_{|m|,(i_{11}...i_{1m_1})...(i_{p1}...i_{pm_p})} := E[(y^*_{i_{11}}...y^*_{i_{1m_1}})...(y^*_{i_{p1}}...y^*_{i_{pm_p}})] \tag{47}$$

4.2.1 Example: third moment ($p = 2$ variables)

$$\begin{aligned}
\nu_3^{21} &= E[z_1^2 z_2^1] \\
&= \Gamma_{1i_{11}}\Gamma_{1i_{12}}\Gamma_{2i_{21}}E[(y^*_{i_{11}}y^*_{i_{12}})y^*_{i_{21}}] \\
&= \Gamma_{1i_{11}}\Gamma_{1i_{12}}\Gamma_{2i_{21}}m_{3,i_{11}i_{12}i_{21}}.
\end{aligned} \tag{48}$$

We use the Einstein sum convention (sum over double indices) and drop the summation symbols $\sum$.

4.2.2 Example: quasi moment c_3

In the case $p = 2$, the coefficient c_3 with $n_1 + n_2 = |n| = 3$ contains the terms $\{n_1, n_2\} = \{3, 0\}, \{2, 1\}, \{1, 2\}, \{0, 3\}$. Thus

$$\begin{aligned}
3!0!\, c_{30} &= E[H_3(z_1)] = E[z_1^3 - 3z_1] = E[z_1^3] = \nu_3^{30} &(49)\\
2!1!\, c_{21} &= E[H_2(z_1)H_1(z_2)] = E[(z_1^2 - 1)z_2] = E[z_1^2 z_2] = \nu_3^{21} &(50)\\
1!2!\, c_{12} &= E[H_1(z_1)H_2(z_2)] = E[z_1(z_2^2 - 1)] = E[z_1 z_2^2] = \nu_3^{12} &(51)\\
0!3!\, c_{03} &= E[H_3(z_2)] = E[z_2^3 - 3z_2] = E[z_2^3] = \nu_3^{03}. &(52)
\end{aligned}$$

5 Generalized Gauss-Hermite filtering

5.1 Time update

Extending the Gaussian filter, the densities are represented by the truncated Hermite series (40) and expectation values occuring in the update equations are computed by Gauss-Hermite integration, including the nongaussian term

$$H(y; \{\mu, m_2, ..., m_K\}) = \sum_{|n|=0}^{K} c_n H_n(z) := H(y, K), \tag{53}$$

$n = \{n_1, ..., n_p\}$. For example, the mean equation (9) is

$$\begin{aligned}
\dot{\mu}(t|t_i) &= E[f(y(t), t)|Z^i] = \int f(y, t)p(y, t)dy &(54)\\
&\approx \sum w_l f(\eta_l, t) * H(\eta_l, K), &(55)
\end{aligned}$$

$l = \{l_1, ..., l_p\}$. In lowest order $K = 2$, $H(y; \{\mu, m_2\}) = 1$, so the usual GHF is a special case. The time update of the kth moment $m_{k,j_1 j_2,...,j_k}(t)$, $k = 2, ..., K$ (eqn. 22) can be computed analogously. Since the density expansion is given by K moments, one obtains a closed system of moment equations. In contrast, a Taylor expansion of the functions f and Ω occuring in the moment equations does produce higher order moments which must be truncated or approximated otherwise (e.g. Gaussian factorization; cf. (Singer 2006b)).

5.2 Measurement update

5.2.1 Exact measurement update

The exact measurement update is given by the Bayes formula

$$\begin{aligned}
p(y_{i+1}|Z^{i+1}) &= p(z_{i+1}|y_{i+1})p(y_{i+1}|Z^i)/L_{i+1} &(56)\\
L_{i+1} &= \int p(z_{i+1}|y_{i+1})p(y_{i+1}|Z^i)dy_{i+1} &(57)\\
p(y_{i+1}|Z^i) &= \phi(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)) * H(y_{i+1}) &(58)
\end{aligned}$$

and using Gauss-Hermite integration the likelihood is

$$
\begin{aligned}
L_{i+1} &= \sum_{l_1=1,\ldots l_p=1}^{m} w_l p(z_{i+1}|\eta_l) H(\eta_l) && (59)\\
\eta_l &= \eta_l(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)). && (60)
\end{aligned}
$$

By the same token, the a posteriori moments are given as

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= L_{i+1}^{-1} \sum_{l_1=1,\ldots l_p=1}^{m} w_l p(z_{i+1}|\eta_l) H(\eta_l)\eta_l && (61)\\
m_{|k|}^{k_1,\ldots,k_p}(t_{i+1}|t_{i+1}) &= L_{i+1}^{-1} \sum_{l_1=1,\ldots l_p=1}^{m} w_l p(z_{i+1}|\eta_l) H(\eta_l) \times \\
&\times \; (\eta_l - \mu(t_{i+1}|t_{i+1}))^k, && (62)
\end{aligned}
$$

$|k| = 2, ..., K$. For the powers in the kth moment, we used the abbreviation $x^k = x_1^{k_1}...x_p^{k_p}$. From these updated moments, a new Hermite representation of the filter density with leading a posteriori Gaussian

$$
\begin{aligned}
p(y_{i+1}|Z^{i+1}) &= \phi(y_{i+1}; \mu(t_{i+1}|t_{i+1}), m_2(t_{i+1}|t_{i+1})) \\
&\times \; H(y_{i+1}; \{\mu(t_{i+1}|t_{i+1}), .., m_K(t_{i+1}|t_{i+1})\}) && (63)
\end{aligned}
$$

can be computed and inserted into the time update equations.

5.2.2 Approximate measurement update

The Bayes formula (56)

$$
\begin{aligned}
p(y_{i+1}|Z^{i+1}) &\propto \phi(z_{i+1}; h(y_{i+1}), R_{i+1}) \times \\
&\times \; \phi(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)) H(y_{i+1}). && (64)
\end{aligned}
$$

can be approximated by using the normal correlation update as follows: The product of the two Gaussians is written approximately as (the formula is exact for linear measurements)

$$
L_{0,i+1} * \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})), \qquad (65)
$$

where (setting $h(y_{i+1}) := h_{i+1}$ etc.)

$$
\begin{aligned}
\mu_0(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \; (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^{-}(z_{i+1} - E[h_{i+1}|Z^i]) && (66)\\
\Sigma_0(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \; (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^{-}\mathrm{Cov}(h_{i+1}, y_{i+1}|Z^i) && (67)\\
L_{0,i+1} &= \phi(z_{i+1}; E[h_{i+1}|Z^i], \mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1})) && (68)
\end{aligned}
$$

is the normal correlation update and the approximate likelihood of the Gaussian part. Therefore the complete update is the product of the Gaussian a posteriori density and the a priori Hermite part

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})) \times \\ &\times \; H(y_{i+1}; \{\mu(t_{i+1}|t_i), .., m_K(t_{i+1}|t_i)\})/L_{1,i+1} \qquad (69) \\ L_{1,i+1} &= \int \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})) \times \\ &\times \; H(y_{i+1}; \{\mu(t_{i+1}|t_i), .., m_K(t_{i+1}|t_i)\}) dy_{i+1} \qquad (70) \end{aligned}$$

and the complete likelihood is $L = L_0 * L_1$. If the Hermite correction is $H = 1$, we have $L_1 = 1$ and $L = L_0$ coincides with the Gaussian part. Again, all integrals involving $p = \phi * H$ can be computed using Gauss-Hermite integration, e.g. the a posteriori moments. They are simpler to compute than (61–62), since they involve only polynomials and not the exponential $p(z|y)$. In the case of linear measurements, (69) is exact.

5.2.3 Improved exact measurement update

The approximate update (66–67) can be used to improve the numerical properties of the Bayes update (56). One replaces the integration with respect to $\phi(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i))$ by integration over the linear posteriori density $\phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1}))$, analogously to importance sampling. This is more efficient if the measurements are nonlinear and far from the mean of the a priori density, since more Gauss-Hermite sample points are in regions of large $\phi(z_{i+1}; h(y_{i+1}), R_{i+1})$.

6 Examples

6.1 Bayesian estimation of volatilities

The filtering of unknown parameters is a convenient method of estimation, since it avoids numerical optimization and yields recursive estimates. (e.g Gelb, 1974, Ljung, 1979). For example, the volatility σ in the Ornstein-Uhlenbeck process

$$dy = \lambda y dt + \sigma dW(t) \qquad (71)$$

with measurement equation

$$z_i = y_i + \epsilon_i; \mathrm{Var}(\epsilon_i) = R \qquad (72)$$

can be estimated by exact ML or as Bayes estimator, using an extended state vector $\eta = \{y, \sigma\}$

$$dy = \lambda y dt + \sigma dW(t) \tag{73}$$
$$d\sigma = 0 \tag{74}$$

with trivial dynamics for the volatility parameter σ. Thus we obtain a non-linear filtering problem, although the system is linear in the actual states y. Applying the EKF, UKF or GHF to the system yields the result, that σ is not filtered by these algorithms (fig. 1). On the contrary, the GGHF can filter the volatility (fig. 2). Sampling the data more densely leads to a more rapid convergence of the Bayes estimator (fig. 3). [2]

The problem can be attributed to the fact that in the GHF only two moments are involved and the measurement information is carried by the normal correlation update which involves the covariance of σ with z. If these quantities are not correlated a priori, no information on the volatility will be contained in the measurements. More exactly, the mentioned filters always use Gaussian a priori and a posteriori densities.

The following simple model can be used to explain the problem: If $p(y|\sigma) = \phi(y; \mu, \sigma)$, the posterior density $p(\sigma|y) = p(y|\sigma)p(\sigma)/p(y)$ is not Gaussian as a function of σ, but strongly deviates from normality. It is skewed and the mode depends on y (see fig. 4). The bivariate distribution $p(y|\sigma)p(\sigma) \propto p(\sigma|y)$ (fig. 5) is not Gaussian as well and observation of y yields information about σ. In the Gaussian approximation used by the GHF (and, implicitly, by the EKF and UKF), $p(\sigma|y)$ does not depend on y. This is the reason why filters based on two moments cannot filter the volatility. We must consider higher order moments. This problem occurs although the measurement is linear and the normal correlation update (69) is exact here. However, the a priori Hermite part $H(y)$ carries higher order dependencies which show up in the a posteriori moments. This is demonstrated by using the bivariate GGHF. In the case $K = 2$ (GHF) we have Gaussian densities and no filtering of σ (fig. 6), whereas for the generalized Gauss-Hermite filter the joint density is nongaussian with increasing dispersion for higher σ-values (fig. 7). Thus, measurements of z lead to corrections in the filtered state $\hat{\sigma}(t) = E[\sigma(t)|Z^i]$.

[2] The parameter values were $\theta = \{\lambda, R, \sigma\} = \{-1, 0.1, 2\}$ and the assumed a priori distribution was $\{y_0, \sigma_0\} \sim N(\{0, 4\}, \text{diag}\{1, 2\}$. Data were simulated in the interval [0,20] with true initial values $\{0, 2\}$, discretization interval $\delta t = 0.1$ and sampled at irregular times $\tau = \{0, 4, 6, 8, 10, 11, 12, 13.5, 13.7, 15, 15.1, 17, 19, 20\}$.

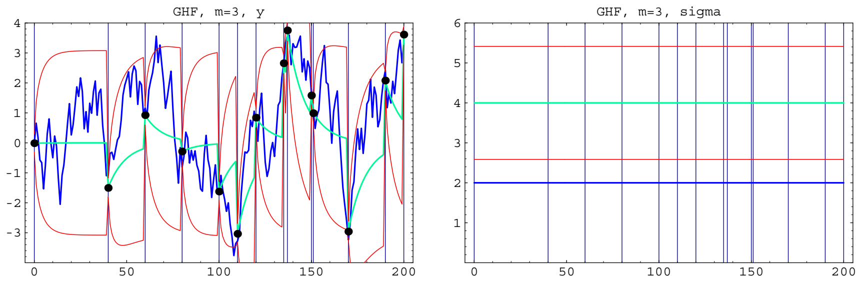

Figure 1: GHF: filtered states and 67% HPD confidence intervals. y (left), σ (right).

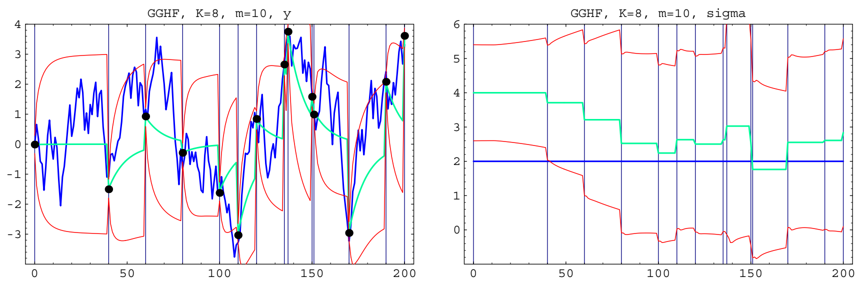

Figure 2: GGHF($K = 10, m = 10$): filtered states and 67% HPD confidence intervals. y (left), σ (right).

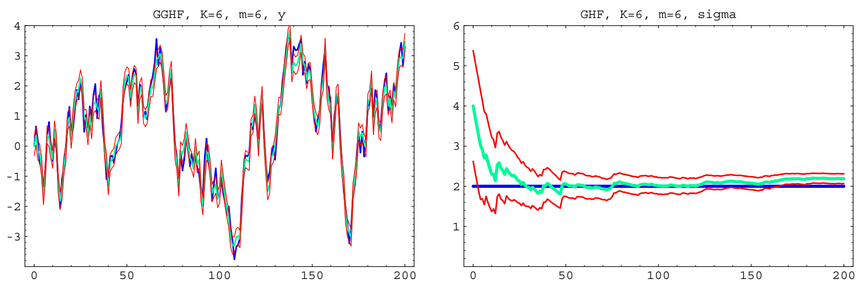

Figure 3: GGHF($K = 6$): Quasi continuous sampling at times $t\delta t; t = 0, ...200$.

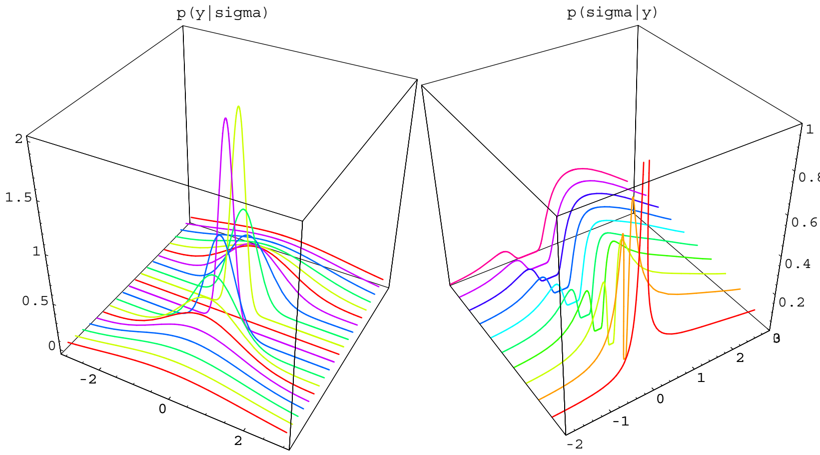

Figure 4: Conditional densities $p(y|\sigma)$ (left) and $p(\sigma|y)$ (right) with prior $p(\sigma) = N(2, \mathrm{Var}(\sigma) = 4)$.

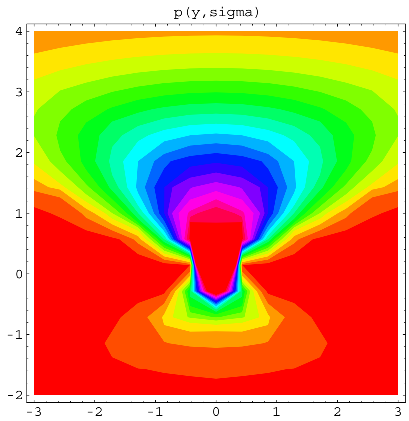

Figure 5: Joint density $p(y|\sigma)p(\sigma)$ which is not Gaussian.

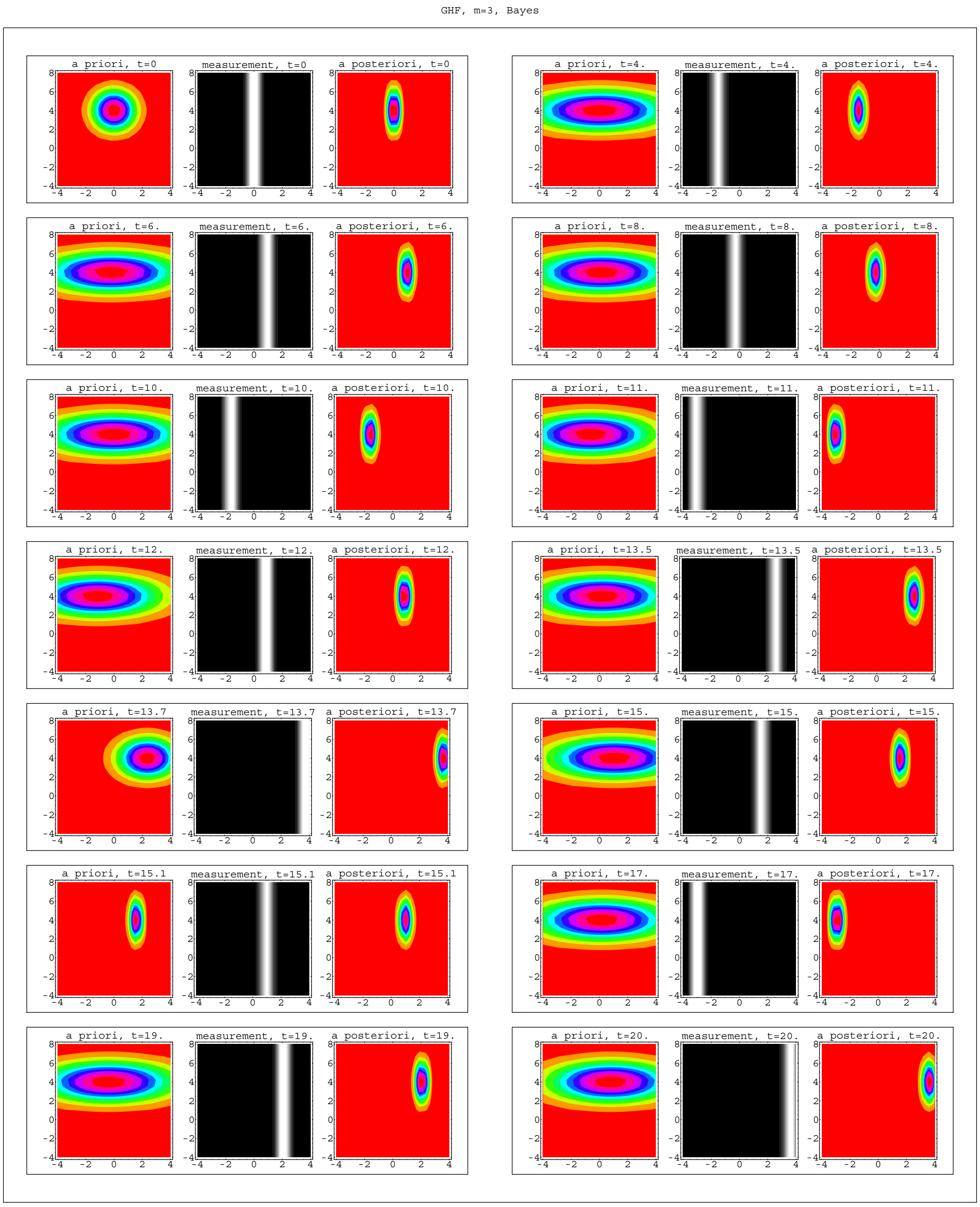

Figure 6: GHF: Gaussian a priori, measurement and a posteriori density for the measurement times.

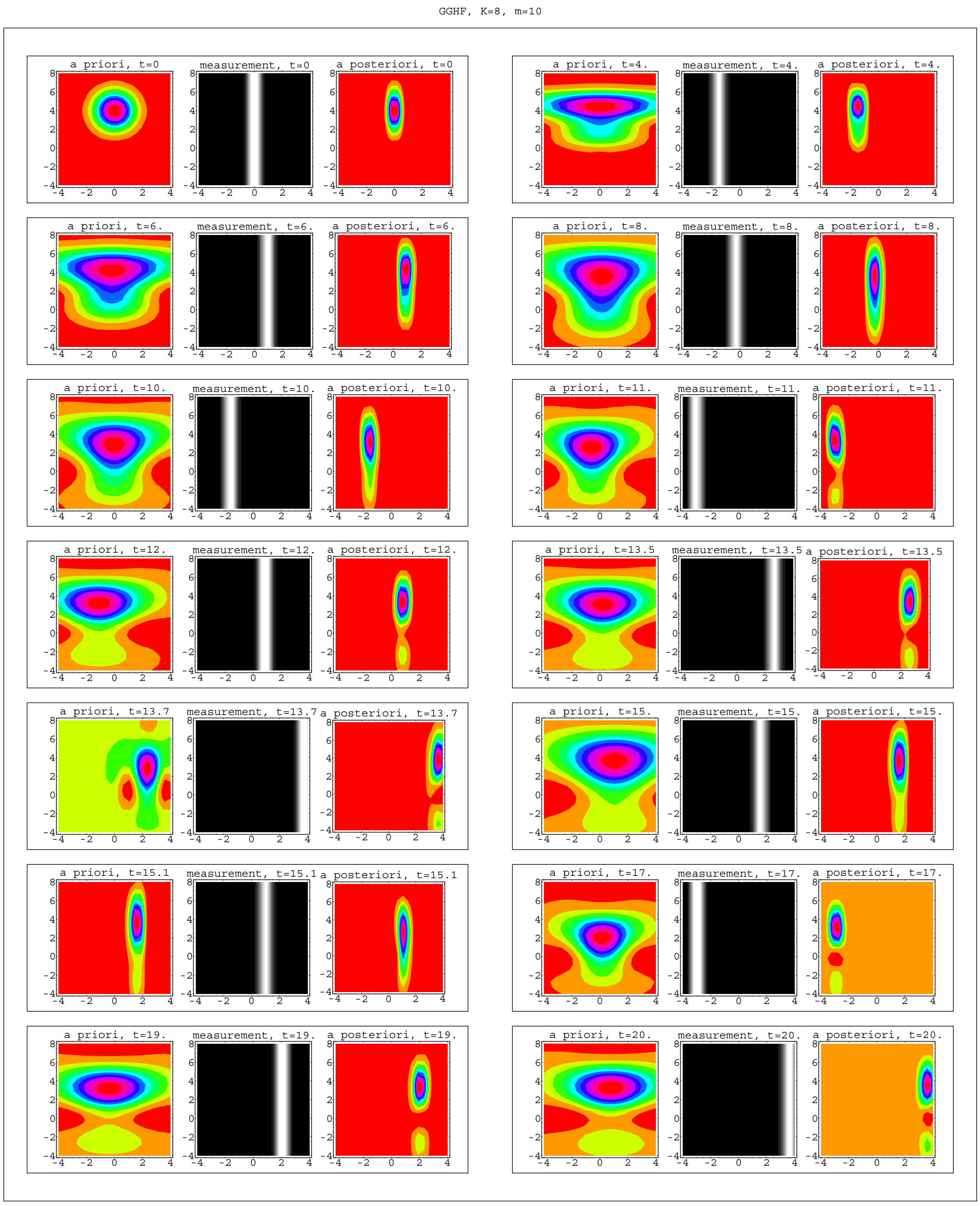

Figure 7: GGHF($K = 8$): Nongaussian a priori, measurement and a posteriori density for the measurement times.

6.2 Stochastic Volatilities

The stochastic volatility model of Scott (1987), cf. also Hull and White (1987)

$$\begin{aligned}
d\log S(t) &= [\mu - \sigma(t)^2/2]dt + \sigma(t)dW(t) \\
d\sigma(t) &= \lambda[\sigma(t) - \bar{\sigma}]dt + \gamma dV(t) \\
z_i &= \log S(t_i)
\end{aligned}$$

leads to similar problems as in the last section. Due to the lemma of Itô

$$d\log S(t) = dS/S - \tfrac{1}{2}S^{-2}dS^2,$$

the volatility process $\sigma(t)$ is also part of the drift, but the filtering with EKF, UKF, GHF etc. does not lead to satisfactory results (fig. 8). Using higher order moment information as in the GGHF(4), GGHF(6) yields estimates of the latent volatility trajectory (figs. 9–10) similar to Monte Carlo filtering (functional integral filter FIF, $N = 10000$ replications; fig. 11; cf. Singer, 2003).

7 Conclusion

The generalized Gauss-Hermite filter (GGHF) is a natural extension of the usual Gauss filter with leading Gaussian and higher order corrections in a Hermite expansion of the filter density. All expectation values occuring in the time and measurement updates can be computed by Gauss-Hermite quadrature and the moment equations are closed. The Bayes update also allows the treatment of strongly nonlinear measurements such as threshold models (ordinal data; cf. (Singer 2006c)). In a linear Ornstein-Uhlenbeck system, the nongaussian bivariate filter density of the extended state could be well approximated by a higher order expansion leading to a sequential Bayesian estimation of the volatility parameter. The same applies to the stochastic volatility model of Scott.

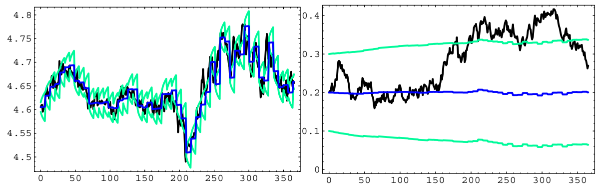

Figure 8: GGHF($K = 2$) = GHF: Stochastic volatility model. Stock price (left) and volatility (right). Similar results are obtained for the EKF, SNF and UKF.

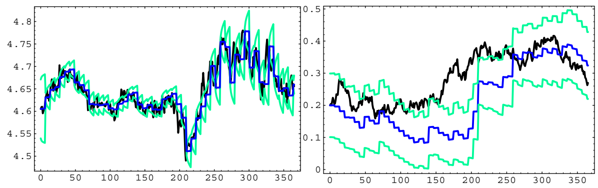

Figure 9: GGHF($K = 4$): Stochastic volatility model. Stock price (left) and volatility (right).

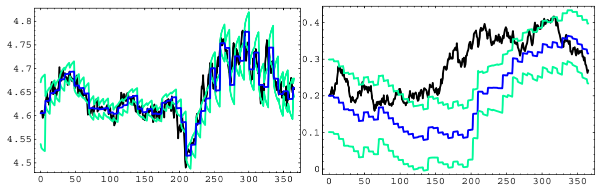

Figure 10: GGHF($K = 6$): Stochastic volatility model. Stock price (left) and volatility (right).

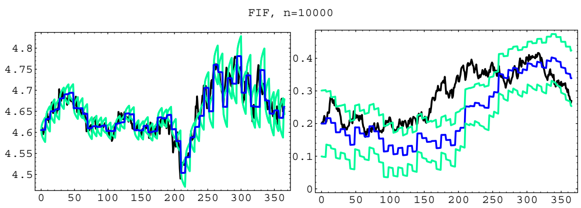

Figure 11: Monte Carlo filter ($N = 10000$ replications): Stochastic volatility model. Stock price (left) and volatility (right).

References

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions*, Dover, New York.

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach, *Econometrica* **70,1**: 223–262.

Arnold, L. (1974). *Stochastic Differential Equations*, John Wiley, New York.

Challa, S., Bar-Shalom, Y. and Krishnamurthy, V. (2000). Nonlinear Filtering via Generalized Edgeworth Series and Gauss-Hermite Quadrature, *IEEE Transactions On Signal Processing* **48**: 1816–1820.

Courant, R. and Hilbert, D. (1968). *Methoden der Mathematischen Physik*, Springer, Berlin, Heidelberg, New York.

Daum, F. (1986). Exact Finite-Dimensional Nonlinear Filters, *IEEE Transactions on Automatic Control* **31**: 616–622.

Gelb, A. (ed.) (1974). *Applied Optimal Estimation*, MIT Press, Cambridge, Mass.

Hull, J. and White, A. (1987). The Pricing of Options with Stochastic Volatilities, *Journal of Finance* **XLII,2**: 281–300.

Ito, K. and Xiong, K. (2000). Gaussian filters for nonlinear filtering problems, *IEEE Transactions on Automatic Control* **45, 5**: 910–927.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Kuznetsov, P., Stratonovich, R. and Tikhonov, V. (1960). Quasi-moment functions in the theory of random processes, *Theory of Probability and its Applications* **V, 1**: 80–97.

Liptser, R. and Shiryayev, A. (2001). *Statistics of Random Processes, Volumes I and II*, second edn, Springer, New York, Heidelberg, Berlin.

Ljung, L. (1979). Asymptotic Behavior of the Extendend Kalman Filter as a Parameter Estimator for Linear Systems, *IEEE Transactions on Automatic Control* **24**: 36–50.

Scott, L. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application, *Journal of Financial and Quantitative Analysis* **22**: 419–438.

Singer, H. (2003). Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing, *Computational Statistics* **18,1**: 79–106.

Singer, H. (2006a). Generalized Gauss-Hermite Filtering, *Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 391*, FernUniversität in Hagen. http://www.fernuni-hagen.de/FBWIWI/forschung/beitraege/pdf/db391.pdf.

Singer, H. (2006b). Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models, *Computational Statistics* **21,3**: in print.

Singer, H. (2006c). Stochastic Differential Equation Models with Sampled Data, *in* K. van Montfort, H. Oud and A. Satorra (eds), *Longitudinal Models in the Behavioral and related Sciences*, Erlbaum publishers.

Stratonovich, R. (1992). *Nonlinear Nonequilibrium Thermodynamics I*, Springer, Berlin.