

Generalized Gauss-Hermite Filtering

Hermann Singer
FernUniversität in Hagen *

May 3, 2006

Abstract

We consider a generalization of the Gauss-Hermite filter (GHF), where the filter density is represented by a Hermite expansion with leading Gaussian term. Thus the usual GHF is included as a special case. The moment equations for the time update are solved stepwise by Gauss-Hermite integration, and the measurement update is computed by the Bayes formula, again using numerical integration.

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

1 Introduction

The Gaussian filter (GF) assumes, that the true filter density may be approximated by a Gaussian distribution parameterized by the conditional mean and variance. Expectation values occurring in the time and measurement updates can be computed numerically by using Gauss-Hermite quadrature (GHF; cf. Ito and Xiong, 2000). Alternatively, such expectations are treated by truncated Taylor expansion, leading to the well known extended Kalman filter (EKF) or the second order nonlinear filter (SNF). It can be shown that the Gaussian filter is equivalent to an infinite Taylor expansion and subsequent Gaussian factorization of moments higher than two. Another method of numerically computing expectations is the unscented Kalman filter (UKF) of Julier and Uhlmann (1997, 2000, 2004). Again only two moments are

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

required and the measurement update rests on the optimal linear estimate (normal correlation).

In this paper higher order moments are explicitly computed and inserted into the Hermite expansion of the filter density. Since the leading term is Gaussian, Gauss-Hermite integration can again be used for the update steps. Thus the method yields closed moment equations. It is demonstrated that the algorithm can successfully model the nongaussian bimodal density of a Ginzburg-Landau system and leads to smaller filtering error than the Gaussian filter.

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)

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

where discrete measurements z_i are taken at times $\{t_0, t_1, \dots, t_T\}$ and $t_0 \leq t \leq t_T$ according to the measurement equation

$$z_i = h(y(t_i), t_i, \psi) + \epsilon_i. \quad (2)$$

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 fixed initial condition $y(t_0)$. The functions $f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \rightarrow \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \rightarrow \mathbb{R}^p \times \mathbb{R}^r$ 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 .

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:

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

Measurement update:

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

$$p(z_{i+1}|Z^i) = \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)dy_{i+1}, \quad (5)$$

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

$$\begin{aligned} Fp &= - \sum_i \frac{\partial}{\partial y_i} [f_i(y, t, \psi)p(y, t|x, s)] \\ &\quad + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} [\Omega_{ij}(y, t, \psi)p(y, t|x, s)] \end{aligned} \quad (6)$$

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

$$\mu(t|t_i) = E[y(t)|Z^i] \quad (7)$$

$$\Sigma(t|t_i) = \text{Var}[y(t)|Z^i]. \quad (8)$$

2.3 Exact Moment Equations

In the Gauss-Hermite filter, instead of solving the time update for the conditional density, the moment equations for the first and second moments are solved approximately. Using the Euler approximation for the SDE (1), one obtains in a short time interval δt ($\delta W(t) := W(t + \delta t) - W(t)$)

$$y(t + \delta t) = y(t) + f(y(t), t)\delta t + g(y(t), t)\delta W(t). \quad (9)$$

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

$$\mu(t + \delta t|t_i) = \mu(t|t_i) + E[f(y(t), t)|Z^i]\delta t \quad (10)$$

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

$$\dot{\mu}(t|t_i) = E[f(y(t), t)|Z^i]. \quad (11)$$

The higher order central moments

$$m_k(t|t_i) := E[(y(t) - \mu(t|t_i))^k|Z^i] := E[M_k(t|t_i)|Z^i] \quad (12)$$

fulfil (scalar notation, dropping the condition)

$$\begin{aligned} m_k(t + \delta t) &= E[y(t) + f(y(t), t)\delta t - \mu(t + \delta t) + g(y(t), t)\delta W(t)]^k \\ &:= E[a + bc]^k \end{aligned} \quad (13)$$

Using the binomial formula we obtain, utilizing the independence of $y(t)$ and $\delta W(t)$

$$E[a + bc]^k = \sum_{j=0}^k \binom{k}{j} E[a^{k-j}b^j] * E[c^j] \quad (14)$$

$$E[c^j] = \begin{cases} (j-1)!!\delta t^{j/2}; & j \text{ is even} \\ 0; & j \text{ is odd} \end{cases} \quad (15)$$

since odd powers of $E[\delta W(t)]^j$ vanish and $E[\delta W(t)^{2j}] = (2j-1)!!\delta t^j$. For example, the second moment (variance) $m_2 = \sigma^2$ fulfils

$$E[a + bc]^2 = E[a^2] + E[b^2]\delta t \quad (16)$$

$$\begin{aligned} &= E[y(t) + f(y(t), t)\delta t - \mu(t + \delta t)]^2 + \\ &+ E[\Omega(y(t), t)]\delta t, \end{aligned} \quad (17)$$

$\Omega = g^2$. Inserting the first moment (10) and setting $a := \alpha + \beta = (y - E(y)) + (f - E(f))\delta t$ one obtains

$$\begin{aligned} m_2(t + \delta t) &= m_2(t) + 2E[(y - E(y))(f - E(f))]\delta t \\ &+ E[f - E(f)]^2\delta t^2 + E[\Omega]\delta t \end{aligned} \quad (18)$$

In general, up to $O(\delta t)$ we have ($M_k := (y - \mu)^k$)

$$\begin{aligned} m_k(t + \delta t) &= E[a^k] + \frac{k(k-1)}{2}E[b^{k-2}]\delta t + O(\delta t^2) \\ &= m_k(t) + kE[(y - E(y))^{k-1}(f - E(f))]\delta t \\ &+ \frac{k(k-1)}{2}E[(y - E(y))^{k-2}\Omega]\delta t + O(\delta t^2) \\ &= m_k(t) + kE[f(y, t) * (M_{k-1}(t) - m_{k-1}(t))]\delta t \\ &+ \frac{k(k-1)}{2}E[M_{k-2}(t)\Omega(y, t)]\delta t + O(\delta t^2). \end{aligned} \quad (19)$$

The exact moment equations (11, 19) are not differential equations, 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 $p(x) = \phi(x; \mu, \sigma^2)$. For the generalized Gaussian filter, $K > 2$ moments with a density $p(x) = \phi(x; \mu, \sigma^2) \sum_{k=0}^K c_k H_k((x - \mu)/\sigma)$ (Hermite expansion) is utilized. In both cases, the integrals can be computed by Gauss-Hermite integration.

3 Gauss-Hermite integration

The moment equations of the last section require the computation of expectations of the type $E[f(X)]$, where X is random variable with distribution $p(x)$. If $p(x)$ is known, the expectation can be computed using (numerical) integration. For the Gaussian filter, one may assume that the true $p(x)$ is approximated by a Gaussian distribution $\phi(x; \mu, \sigma^2)$ with the same mean μ and variance σ^2 . Then, the Gaussian integral

$$E_\phi[f(X)] = \int f(x)\phi(x; \mu, \sigma^2)dx = \int f(\mu + \sigma z)\phi(x; 0, 1)dz \quad (20)$$

$$\approx \sum_{l=1}^m f(\mu + \sigma\zeta_l)w_l = \sum_{l=1}^m f(\xi_l)w_l \quad (21)$$

may be approximated by Gauss-Hermite quadrature (cf. Ito and Xiong, 2000). If such an approximation is used, one obtains the Gauss-Hermite filter (GHF). Generally, filters using Gaussian densities are called Gaussian filters (GF). The GHF can be interpreted in terms of the singular density $p_{GH}(x) = \sum_{l=1}^m w_l \delta(x - \xi_l)$ concentrated at the quadrature points ξ_l . The Gaussian filter is equivalent to a Taylor expansion of f to higher orders L

$$E[f(X)] \approx \sum_{l=0}^L \frac{1}{l!} f^{(l)}(\mu) E[X - \mu]^l = \sum_{l=0}^L \frac{1}{l!} f^{(l)}(\mu) m_l \quad (22)$$

(higher order nonlinear filter HNF(2, L)) and factorization of the moments according to the Gaussian assumption $m_l := E[X - \mu]^l = (l - 1)!!\sigma^l$ (l even) and $m_l = 0$ (l odd). This leads to an exact computation of (20) for $L \rightarrow \infty$. In this limit, the HNF and GF coincide. In the EKF=HNF(2,1) and SNF=HNF(2,2), the higher order corrections are neglected. Also, third and higher order moments could be used (HNF(K, L); cf. Singer, 2006[14]).

4 Gauss-Hermite filtering

4.1 Continuous-discrete filtering scheme

The GHF is a recursive sequence of time and measurement updates for the conditional moments μ and Σ , where expectation values are computed according to (20) using Gauss-Hermite quadrature:

Initial condition: $t = t_0$

$$\begin{aligned}\mu(t_0|t_0) &= \mu + \text{Cov}(y_0, h_0) \times \\ &\quad \times (\text{Var}(h_0) + R(t_0))^{-1} (z_0 - E[h_0]) \\ \Sigma(t_0|t_0) &= \Sigma - \text{Cov}(y_0, h_0) \times \\ &\quad \times (\text{Var}(h_0) + R(t_0))^{-1} \text{Cov}(h_0, y_0) \\ L_0 &= \phi(z_0; E[h_0], \text{Var}(h_0) + R(t_0))\end{aligned}$$

$$\text{quadrature points} : \eta_l = \eta_l(\mu, \Sigma); \mu = E[y_0], \Sigma = \text{Var}(y_0).$$

$i = 0, \dots, T - 1$:

Time update: $t \in [t_i, t_{i+1}]$

$$\begin{aligned}\tau_j &= t_i + j\delta t; j = 0, \dots, J_i - 1 = (t_{i+1} - t_i)/\delta t - 1 \\ \mu(\tau_{j+1}|t_i) &= \mu(\tau_j|t_i) + E[f(y(\tau_j), \tau_j)|Z^i]\delta t \\ \Sigma(\tau_{j+1}|t_i) &= \Sigma(\tau_j|t_i) + \\ &\quad + \{\text{Cov}[f(y(\tau_j), \tau_j), y(\tau_j)|Z^i] + \\ &\quad + \text{Cov}[y(\tau_j), f(y(\tau_j), \tau_j)|Z^i] + \\ &\quad + E[\Omega(y(\tau_j), \tau_j)|Z^i]\}\delta t\end{aligned}$$

$$\text{quadrature points} : \eta_l = \eta_l(\mu(\tau_j|t_i), \Sigma(\tau_j|t_i))$$

Measurement update: $t = t_{i+1}$

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

$$\text{quadrature points} : \eta_l = \eta_l(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)).$$

4.2 Discussion

1. The discretization interval δt is a small value controlling the accuracy of the Euler scheme implicit in the time update. Since the quadrature

points are functions of the mean and variance, the moment equations (10,18) are a coupled system of nonlinear differential equations for the sample points of the Gauss-Hermite scheme. Therefore, other approximation methods such as the Heun scheme or higher order Runge-Kutta schemes could be used.

2. The time update is a multivariate version of eqn. (18) and neglects second order terms. Inclusion of $E[f - E(f)][f - E(f)]'\delta t^2$ leads to a positive semidefinite update, which is numerically more stable.
3. The measurement update is the optimal linear update (normal correlation; Liptser and Shiriyayev, 1978, ch. 13, theorem 13.1, lemma 14.1)

$$\begin{aligned}\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \text{Cov}(y_{i+1}, z_{i+1}|Z^i)\text{Var}(z_{i+1}|Z^i)^{-} \times \\ &\quad \times (z_{i+1} - E[z_{i+1}|Z^i]) \quad (23) \\ \Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \text{Cov}(y_{i+1}, z_{i+1}|Z^i)\text{Var}(z_{i+1}|Z^i)^{-} \times \\ &\quad \times \text{Cov}(z_{i+1}, y_{i+1}|Z^i). \quad (24)\end{aligned}$$

with measurement equation (2) inserted and covariances computed by Gauss-Hermite integration. A direct implementation of the Bayes formula (4) would lead to the asymmetric a posteriori density

$$p(y_{i+1}|Z^{i+1}) = \sum_{l=1}^m w_l^* \delta(y_{i+1} - \eta_l) \quad (25)$$

$$w_l^* = w_l p(z_{i+1}|\eta_l) / \sum_{l=1}^m w_l p(z_{i+1}|\eta_l) \quad (26)$$

where the a priori density is $p(y_{i+1}|Z^i) = \sum_{l=1}^m w_l \delta(y_{i+1} - \eta_l)$ with Gauss-Hermite sample points $\eta_l = \eta_l(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i))$. Computing the a posteriori moments

$$\mu(t_{i+1}|t_{i+1}) = \int y_{i+1} p(y_{i+1}|Z^{i+1}) dy_{i+1} = \sum_{l=1}^m w_l^* \eta_l \quad (27)$$

$$\begin{aligned}\Sigma(t_{i+1}|t_{i+1}) &= \int (y_{i+1} - \mu(t_{i+1}|t_{i+1})) \\ &\quad \times (y_{i+1} - \mu(t_{i+1}|t_{i+1}))' p(y_{i+1}|Z^{i+1}) dy_{i+1} \quad (28)\end{aligned}$$

$$= \sum_{l=1}^m w_l^* [\eta_l - \mu(t_{i+1}|t_{i+1})][\eta_l - \mu(t_{i+1}|t_{i+1})]' \quad (29)$$

one can construct a symmetric a posteriori distribution with the same first and second moments (cf. sect. 5.2.1).

4. Taylor expansion of f , Ω and h around μ leads to the usual EKF and SNF. Using sigma points instead of Gaussian quadrature points yields the unscented Kalman filter UKF (cf. Julier and Uhlmann, 1997, 200, 2004, Singer, 2006[13]).

5 Generalized Gauss-Hermite filtering

5.1 Hermite Expansion

If the filter density strongly deviates from normality, a Fourier expansion in terms of Hermite polynomials may be utilized. The filter density $p(y)$ can be expanded into a Fourier series (cf. Courant and Hilbert, 1968, ch. II, 9, Abramowitz and Stegun, 1965, ch. 22, Ait-Sahalia, 2002) 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.

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

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

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

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

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

and the Fourier coefficients are given by

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

where X is a random variable with density $p(x)$. 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.

¹Actually, 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. $\int q(x)^2 dx = \int \exp(x^2/2)p^2(x)dx < \infty$ (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 $c_0 = 1$, $c_1 = 0$, $c_2 = 0$,

$$c_3 := (1/3!)E[Z^3] = (1/3!)\nu_3 \quad (34)$$

$$c_4 := (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(\nu_4 - 3) \quad (35)$$

and the standardized density expansion

$$p_z(z) := \phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + \dots] \quad (36)$$

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 $p_x(x)$

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

$$= \phi(x; \mu, \sigma^2) H(x) \quad (38)$$

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*

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

5.2 Generalized Gauss-Hermite filtering

Extending the Gaussian filter, the densities are represented by the truncated Fourier series

$$p(y) = \phi(y; \mu, \sigma^2) \sum_{n=0}^K c_n H_n((y - \mu)/\sigma) = \phi * H, \quad (40)$$

and expectation values occurring in the update equations are computed by Gauss-Hermite integration, including the nongaussian term

$$H(y; \{\mu, m_2, \dots, m_K\}) = \sum_{n=0}^K c_n H_n((y - \mu)/\sigma) := H(y, K). \quad (41)$$

For example, the mean equation (11) is

$$\dot{\mu}(t|t_i) = E[f(y(t), t)|Z^i] = \int f(y, t)p(y, t)dy \quad (42)$$

$$\approx \sum w_l f(\eta_l, t) * H(\eta_l, K). \quad (43)$$

In lowest order $K = 2$, $H(y; \{\mu, m_2\}) = 1$, so the usual GHF is a special case. The time update of the k th moment

$$\begin{aligned} m_k(t + \delta t|t_i) &= m_k(t|t_i) + kE[f(y, t) * (M_{k-1}(t|t_i) - m_{k-1}(t|t_i))] \delta t \\ &+ \frac{k(k-1)}{2} E[M_{k-2}(t|t_i) \Omega(y, t)] \delta t + O(\delta t^2) \end{aligned} \quad (44)$$

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 occurring in the moment equations does produce higher order moments which must be truncated or otherwise approximated (e.g. Gaussian factorization; cf. Singer, 2006[14]).

5.2.1 Exact measurement update

The exact measurement update is given by the Bayes formula

$$p(y_{i+1}|Z^{i+1}) = p(z_{i+1}|y_{i+1})p(y_{i+1}|Z^i)/L_{i+1} \quad (45)$$

$$L_{i+1} = \int p(z_{i+1}|y_{i+1})p(y_{i+1}|Z^i) dy_{i+1} \quad (46)$$

$$p(y_{i+1}|Z^i) = \phi(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)) * H(y_{i+1}) \quad (47)$$

and using Gauss-Hermite integration the likelihood is

$$L_{i+1} = \sum_{l=1}^m w_l p(z_{i+1}|\eta_l) H(\eta_l) \quad (48)$$

$$\eta_l = \eta_l(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)). \quad (49)$$

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

$$\mu(t_{i+1}|t_{i+1}) = L_{i+1}^{-1} \sum_{l=1}^m w_l p(z_{i+1}|\eta_l) H(\eta_l) \eta_l \quad (50)$$

$$m_k(t_{i+1}|t_{i+1}) = L_{i+1}^{-1} \sum_{l=1}^m w_l p(z_{i+1}|\eta_l) H(\eta_l) [\eta_l - \mu(t_{i+1}|t_{i+1})]^k, \quad (51)$$

$k = 2, \dots, K$. From these updated moments, a new Hermite representation of the filter density with leading a posteriori Gaussian

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \phi(y_{i+1}; \mu(t_{i+1}|t_{i+1}), m_2(t_{i+1}|t_{i+1})) \\ &\times H(y_{i+1}; \{\mu(t_{i+1}|t_{i+1}), \dots, m_K(t_{i+1}|t_{i+1})\}) \end{aligned} \quad (52)$$

can be computed and inserted into the time update equations.

5.2.2 Approximate measurement update

The Bayes formula (45)

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &\propto \phi(z_{i+1}; h(y_{i+1}), R_{i+1}) \times \\ &\times \phi(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i))H(y_{i+1}). \end{aligned} \quad (53)$$

can be approximated by using the normal correlation (23) as follows: The product of the two Gaussians is written approximately as (the formula is exact for linear measurements)

$$L_{0,i+1} * \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})), \quad (54)$$

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

$$\begin{aligned} \mu_0(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \text{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\ &\times (\text{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^{-1} (z_{i+1} - E[h_{i+1}|Z^i]) \end{aligned} \quad (55)$$

$$\begin{aligned} \Sigma_0(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \text{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\ &\times (\text{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^{-1} \text{Cov}(h_{i+1}, y_{i+1}|Z^i) \end{aligned} \quad (56)$$

$$L_{0,i+1} = \phi(z_{i+1}; E[h_{i+1}|Z^i], \text{Var}(h_{i+1}|Z^i) + R(t_{i+1})) \quad (57)$$

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

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})) \times \\ &\times H(y_{i+1}; \{\mu(t_{i+1}|t_i), \dots, m_K(t_{i+1}|t_i)\})/L_{1,i+1} \end{aligned} \quad (58)$$

$$\begin{aligned} L_{1,i+1} &= \int \phi(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})) \times \\ &\times H(y_{i+1}; \{\mu(t_{i+1}|t_i), \dots, m_K(t_{i+1}|t_i)\}) dy_{i+1} \end{aligned} \quad (59)$$

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 (50), since they involve only polynomials and not the exponential $p(z|y)$. In the case of linear measurements, (58) is exact.

The approximate update can be used to improve the numerical properties of the Bayes update (45). 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 Example: Ginzburg-Landau Model

The algorithm was tested using a nonlinear system which strongly deviates from Gaussian behavior (cf. Miller et al., 1994, Singer, 2002). The Ginzburg-Landau model is a diffusion process in a double well potential $\Phi(y, \{\alpha, \beta\}) = \frac{\alpha}{2}y^2 + \frac{\beta}{4}y^4$ with vector field $f = -\partial\Phi/\partial y$ and state independent diffusion coefficient $g = \sigma$. The SDE reads explicitly (Ginzburg-Landau equation)

$$dy = -[\alpha y + \beta y^3]dt + \sigma dW(t) \quad (60)$$

with measurement equation

$$z_i = y_i + \epsilon_i, \quad (61)$$

$\text{Var}(\epsilon_i) = R$. A physical picture is the strongly damped random movement of a sphere in a landscape defined by the potential Φ . The potential Φ can exhibit a Hopf bifurcation when α becomes negative ($\beta > 0$). In the double wells, the transition density is Gaussian for short times, but for long time intervals a bimodal shape tending to the stationary density $p_0(y) = \lim_{t \rightarrow \infty} p(t, y|x, s) \propto \exp(-\frac{2}{\sigma^2}\Phi(y))$ occurs. The parameters are chosen as $\psi = \{\alpha, \beta, \sigma, R\} = \{-1, 0.1, 2, 1\}$. Figs. (1–3) show the evolution of the filter density from a Gaussian initial condition to time $t = 20$ (time steps $\delta t = 0.1$). Using more terms in the Hermite expansion, the algorithm can model the bimodal shape of the density. Fig. (4) displays the comparison of the true stationary density with the Hermite expansion ($K = 10$) produced by the moment equations at $t = 20$ and Hermite expansions ($K = 10, 20, 30$) computed from the true p_0 . It is seen that higher expansion orders than 10 are necessary for a good approximation. The sequence of measurement updates are displayed in figs. (5–6). The GHF always uses Gaussian densities, whereas the GGHF ($K = 10$) can model the more realistic bimodal shape. Since the a priori distribution has probability mass in both potential wells, the a posteriori density is more sharply peaked and located nearer to the measurements (e.g. second row, $t = 4$, 10th row, $t = 15$, etc.). The measurements are linear, so the normal correlation update (58) was used, which is exact here.

The performance of the filters was tested in a simulation study comparing true and filtered trajectory (table 1). In $M = 100$ replications the following quantities were computed:

- filter error: $\nu_t = y(t) - \hat{y}(t)$
- squared filter error: $A = \sum_t \nu_t^2$
- error mean : $B = 1/T \sum_t \nu_t$

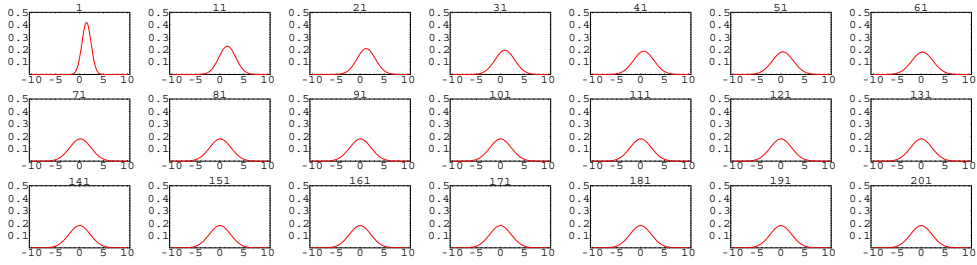


Figure 1: Time evolution of filter density $p(y, t)$ ($K = 2 = \text{GHF}$).

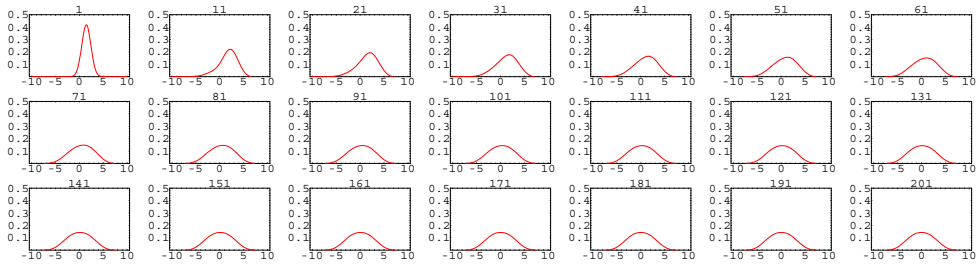


Figure 2: Time evolution using $K = 4$ moments.

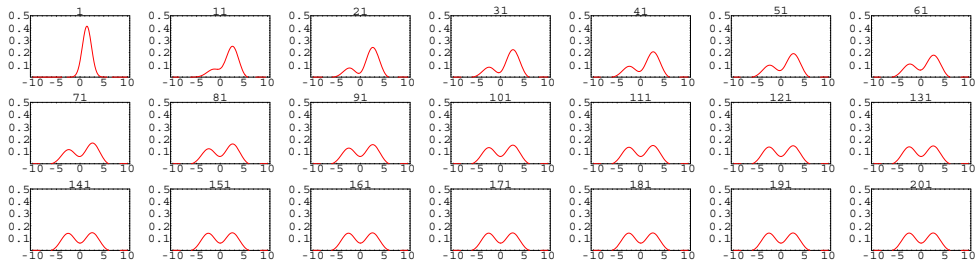


Figure 3: Time evolution using $K = 10$ moments.

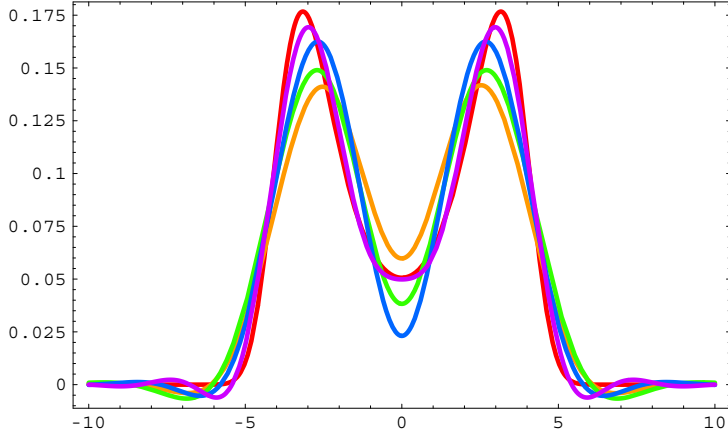


Figure 4: Comparison of the exact stationary density p_0 (red), the Hermite expansion produced by the moment equations (yellow; $K = 10$) and computed from p_0 (green: $K = 10$; blue: $K = 20$, purple: $K = 30$). The density approximation may become negative locally.

- error standard deviation: $C = [1/T \sum_t (\nu_t - B)^2]^{0.5}$
- mean and standard deviation of A, B, C in M samples:
 $\bar{X} = 1/M \sum_m X_m$; $\text{std}(X) = [1/M \sum_m (X_m - \bar{X})^2]^{0.5}$; $X \in \{A, B, C\}$

The table compares the extended Kalman filter (EKF), the unscented Kalman filter (UKF) with the Gauss-Hermite filter (GHF) and the generalized Gauss-Hermite filter (GGHF) for several expansion orders $K = 2, 4, \dots, 14$. For $K = 2$, GHF($m = 4$) and GGHF($m = 4, K = 2$) coincide. The GHF($m = 3$) is equivalent to the UKF($\kappa = 2$) (cf. Ito and Xiong, 2000). Since the drift f is of $O(y^3)$ and terms up to $O(f^2)$ are considered in the time update, one must use $2m - 1 \geq 6$ or $m \geq 4$ in the Gauss-Hermite sum. For the GGHF($K > 2$), also the Hermite correction $H(y; K) \sim O(y^K)$ must be considered in the choice of integration order m . If the drift is $O(y^L)$, the highest order term is $E[y + f\delta t]^K$ leading to terms $O(y^{LK}y^K) = O(y^{(L+1)K})$. Thus, the integration order must be $2m - 1 \geq (L + 1)K$ or $m \geq 1/2[(L + 1)K + 1]$. Setting $L = 3, f \sim O(y^L)$, we find the table with minimal integration order

$$\begin{bmatrix} K & 2 & 4 & 6 & 8 & 10 & 12 & 14 \\ m & 4 & 9 & 13 & 17 & 21 & 25 & 29 \end{bmatrix}. \quad (62)$$

These integration orders were used for the GGHF. The performance of the GGHF in terms of squared filter error improves with the degree of the density approximation (order K of Hermite series $H(y, K)$) and is better than the

GGHF, $K=2$

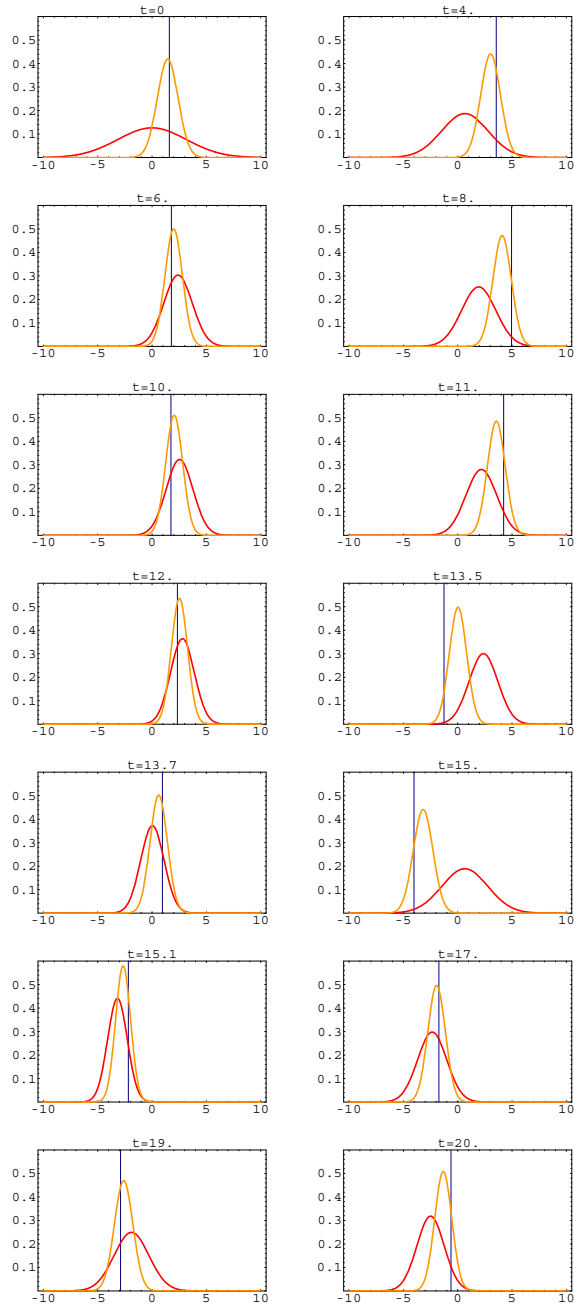


Figure 5: A priori (red) and a posteriori densities (yellow). $K = 2$. The measurements are plotted by a vertical line.

GGHF, $K=10$

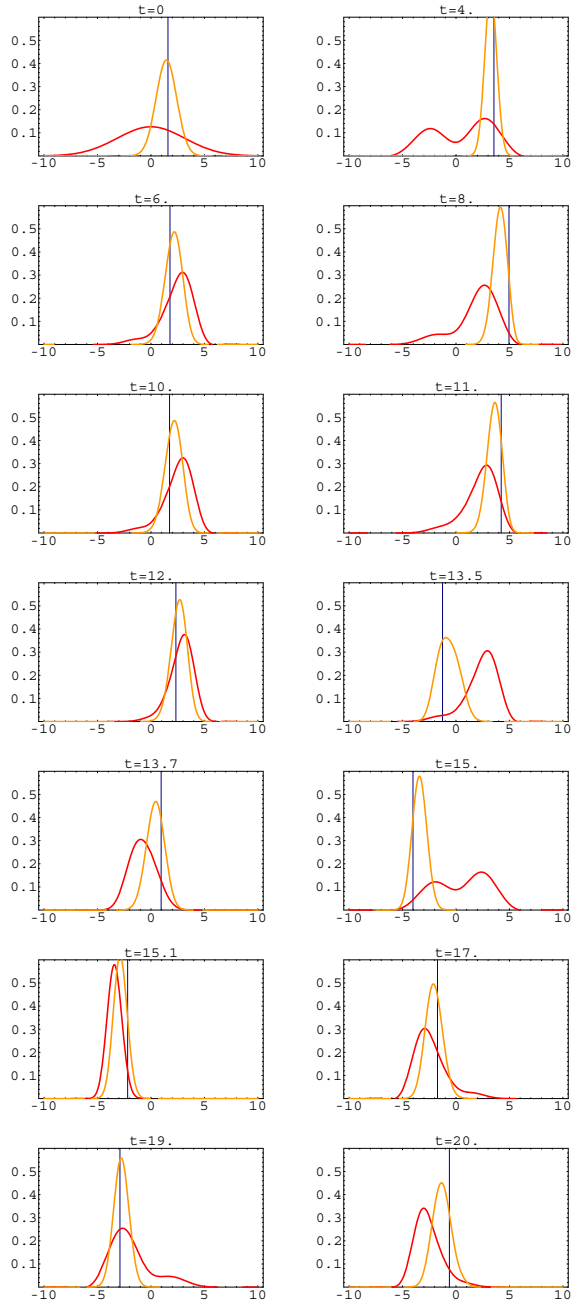


Figure 6: A priori (red) and a posteriori densities (yellow). $K = 10$. The measurements are plotted by a vertical line.

usual GHF ($K = 2$) as well as the UKF. The EKF is outperformed by all algorithms.

numerical remarks Since the Hermite expansion may become locally negative (cf. fig. 4), the Hermite part $H(y, K)$ (41) was replaced by $H^+ := H\theta(H) + \epsilon$, where $\theta(H)$ is the Heaviside unit step function. This is important, if the measurement update takes part in regions with negative or oscillating a priori densities. The addition of a small number $\epsilon = 10^{-1}$ causes a smoothing of oscillating a posteriori densities. The truncation of the polynomial $H(y, K)$ may require higher integration order m , however. In order to improve the performance of the Gauss-Hermite quadrature, only sample points with weights above a certain threshold (10^{-4}) were used. Alternatively, one can exclude points outside a range of k standard deviations, i.e. $|\zeta_l| > k$. Usually, $k = 3$ or 4 is sufficient.

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 occurring in the time and measurement updates can be computed by Gauss-Hermite quadrature and the moment equations are closed. The Bayes update allows the treatment of strongly nonlinear measurements such as threshold models (ordinal data). In a model system, the nongaussian bimodal filter density could be well approximated by higher order expansions leading to a better filter performance. Further work will implement the algorithm for multivariate Hermite expansions and moment equations.

Table 1: Distribution of filter error in $M = 100$ samples. $\bar{A}, \text{std}(A)$: mean and standard deviation of squared filter error; $\bar{B}, \text{std}(B)$: mean and standard deviation of averaged filter error (bias); $\bar{C}, \text{std}(C)$: mean and standard deviation of filter error standard deviation (see text).

	\bar{A}	$\text{std}(A)$	\bar{B}	$\text{std}(B)$	\bar{C}	$\text{std}(C)$
EKF						
	30.4303	9.29225	-0.0172362	0.851867	1.98594	0.605503
UKF						
$\kappa = 0$	24.3465	3.62642	0.0228194	0.643741	1.58941	0.270704
$\kappa = 1$	24.3543	4.01249	0.0197056	0.632016	1.59548	0.292646
$\kappa = 2$	24.4874	4.56068	0.0168696	0.629528	1.60828	0.321831
$\kappa = 3$	24.7165	5.12368	0.0144765	0.635766	1.62544	0.351386
GHF						
$m = 3$	24.4874	4.56068	0.0168696	0.629528	1.60828	0.321831
$m = 4$	24.4653	4.49457	0.0171205	0.628488	1.60677	0.318386
GGHF						
$K = 2$	24.4653	4.49457	0.0171205	0.628488	1.60677	0.318386
$K = 4$	24.2210	4.25883	0.0155556	0.607959	1.5966	0.299874
$K = 6$	24.1879	4.32234	0.0174935	0.595958	1.59868	0.30402
$K = 8$	24.1949	4.45003	0.0190297	0.593306	1.60069	0.310411
$K = 10$	24.2486	4.65799	0.0193939	0.592476	1.60574	0.321725
$K = 12$	24.2309	4.64232	0.0199682	0.590062	1.60533	0.320391
$K = 14$	24.1987	4.46655	0.020575	0.585903	1.6038	0.310969

References

- [1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- [2] Y. Aït-Sahalia. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70,1:223–262, 2002.
- [3] L. Arnold. *Stochastic Differential Equations*. John Wiley, New York, 1974.
- [4] R. Courant and D. Hilbert. *Methoden der Mathematischen Physik*. Springer, Berlin, Heidelberg, New York, 1968.
- [5] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45, 5:910–927, 2000.
- [6] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [7] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control*, Orlando, Florida, 1997.
- [8] S. Julier and J. Uhlmann. Unscented Kalman Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92, 3:401–422, 2004.
- [9] S. Julier, J. Uhlmann, and H. F. Durrant-White. A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactions on Automatic Control*, 45, 3:477–482, 2000.
- [10] R.S. Liptser and A.N. Shiriyayev. *Statistics of Random Processes, Volumes I and II*. Springer, New York, Heidelberg, Berlin, 1977, 1978.
- [11] R.N. Miller, M. Ghil, and F. Gauthiez. Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. *Journal of the Atmospheric Sciences*, 51, 8:1037–1056, 1994.
- [12] H. Singer. Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood vs. Extended Kalman Filter and Itô-Taylor Expansion. *Journal of Computational and Graphical Statistics*, 11,4:972–995, 2002.

- [13] H. Singer. Continuous-Discrete Unscented Kalman Filtering. Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 384, FernUniversität in Hagen, 2006.
- [14] H. Singer. Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. *Computational Statistics*, 21, 3, 2006. in print.