AStA (2008) 92: 179–195
DOI 10.1007/s10182-008-0068-z

ORIGINAL PAPER

Generalized Gauss–Hermite filtering

Hermann Singer

Received: 14 February 2007 / Accepted: 11 February 2008 / Published online: 5 March 2008

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 (GGHF). 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. The performance of the filter is compared numerically with the GHF, the UKF (unscented Kalman filter) and the EKF (extended Kalman filter) and leads to a lower mean squared filter error.

Keywords 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 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), the second-order nonlinear filter (SNF) or higher order nonlinear filters (HNF). 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

H. Singer (✉)
Lehrstuhl für Angewandte Statistik und Methoden der Empirischen Sozialforschung, FernUniversität in Hagen, 58084 Hagen, Germany
e-mail: hermann.singer@fernuni-hagen.de

the unscented transform (UT) leading to the unscented Kalman filter (UKF) of Julier and Uhlmann (1997, 2000, 2004). Closely related is the difference filter of Nørgaard et al. (2000), which uses Stirling interpolation. Again only two moments are 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. Related algorithms were developed by Srinivasan (1970) and Challa et al. (2000), but we formulate the nonlinear 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. In contrast, Challa et al. (2000) used linear time updates and EKF measurement updates in the Bayes formula. It is demonstrated that the new algorithm can successfully model the non-Gaussian bimodal density of a Ginzburg–Landau system and leads to smaller filtering error than the GHF, EKF and UKF.

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)

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

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

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

In state equation (1), $W(t)$ denotes an r-dimensional standard Wiener process where the increments $dW(t)$ are independent of $y(t)$, and the state is described by the p-dimensional state vector $y(t)$. It fulfills a system of stochastic differential equations in the sense of Itô (cf. Arnold 1974, Chap. 6) with random initial condition $y(t_0)$ independent of $dW(t_0)$. 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. To ensure existence and uniqueness of the solution, a Lipschitz and growth condition for f and g must be fulfilled (Arnold 1974, Theorem 6.2.2). 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) independent of $W(t)$. 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 $Z^i = \{z(t_j)|j \le i\}, t_i \le t$, and h, R may depend on lagged measurements $Z^{i-1} = \{z(t_j)|j \le i-1\}$ allowing continuous time ARCH specifications. In the linear case, the system is conditionally Gaussian (cf. Liptser and Shiryayev 2001, Chap. 11). This dependence will be dropped in the following.

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 $p(y,t|Z^i)$ (dropping the parameter vector ψ):

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

Time update:

$$\frac{\partial p(y,t|Z^i)}{\partial t} = F(y,t)p\big(y,t|Z^i\big); \quad t \in [t_i, t_{i+1}], \tag{3}$$
$$p\big(y_i, t_i|Z^i\big) := p_{i|i}.$$

Measurement update:

$$p\big(y_{i+1}, t_{i+1}|Z^{i+1}\big) = \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}, t_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} := p_{i+1|i+1}, \tag{4}$$

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

where

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

is the Fokker–Planck operator, $\Omega = gg'$ and $L_{i+1}(\psi) = p(z_{i+1}|Z^i)$ is the likelihood function of observation z_{i+1}. The conditional density of the measurements is given by the Gaussian $p(z_{i+1}|y_{i+1}, Z^i) = \phi(z_{i+1}; h(y_{i+1}, t_{i+1}), R(t_{i+1}))$. 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, 2005).

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. For simplicity, here we will discuss only the scalar (for the multivariate case, see Singer 2006b). Using the Euler–Maruyama approximation for the SDE (1), one obtains in a short time interval δt

$$y(t+\delta t) = y(t) + f\big(y(t),t\big)\delta t + g\big(y(t),t\big)\delta W(t), \tag{7}$$

where $\delta W(t) := W(t+\delta t) - W(t)$. Taking the expectation $E[\dots|Z^i]$ one gets the moment equation

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

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

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

The higher order central moments

$$m_k(t|t_i) := E\big[\big(y(t)-\mu(t|t_i)\big)^k|Z^i\big] := E\big[M_k(t|t_i)|Z^i\big] \tag{10}$$

fulfill (dropping the condition)

$$\begin{aligned} m_k(t+\delta t) &= E\big\{\big[y(t)+f\big(y(t),t\big)\delta t-\mu(t+\delta t)\big]+g\big(y(t),t\big)\delta W(t)\big\}^k \\ &:= E\{a+bc\}^k, \end{aligned} \tag{11}$$

where $a := y(t) + f(y(t),t)\delta t - \mu(t+\delta t)$, $b := g(y(t),t)$ and $c := \delta W(t) \sim N(0,\delta t)$. Using the binomial formula one obtains, using the independence of $y(t)$ and $\delta W(t)$ (cf. (1))

$$E[a+bc]^k = \sum_{j=0}^{k}\binom{k}{j}E\big[a^{k-j}b^j\big]\,E\big[c^j\big], \tag{12}$$

$$E\big[c^j\big] = \begin{cases}(j-1)!!\delta t^{j/2}; & j \text{ is even,}\\ 0; & j \text{ is odd.}\end{cases} \tag{13}$$

For example, the second moment (variance) $m_2 = \sigma^2$ fulfills

$$\begin{aligned} E[a+bc]^2 &= E\big[a^2\big]+E\big[b^2\big]\delta t \\ &= E\big[y(t)+f\big(y(t),t\big)\delta t-\mu(t+\delta t)\big]^2+E\big[\Omega\big(y(t),t\big)\big]\delta t, \end{aligned} \tag{14}$$

$\Omega := g^2$. Inserting the first moment (8) and using $a = y + f\delta t - \mu(t+\delta t) = (y-E(y)) + (f-E(f))\delta t$ with the shorthand notation $y(t) = y$; $f(y(t),t) = f$, $\Omega(y(t),t) = \Omega$ one obtains

$$\begin{aligned} m_2(t+\delta t) = m_2(t) &+ 2E\big[\big(y-E(y)\big)\big(f-E(f)\big)\big]\delta t \\ &+ E\big[f-E(f)\big]^2\delta t^2 + E[\Omega]\delta t, \end{aligned} \tag{15}$$

and in general, up to $O(\delta t)$ we have, setting $M_k := (y-E(y))^k$,

$$\begin{aligned} m_k(t+\delta t) &= E\big[a^k\big] + \frac{k(k-1)}{2}E\big[a^{k-2}b^2\big]\delta t + O\big(\delta t^2\big) \\ &= m_k(t) + kE\big[\big(y-E(y)\big)^{k-1}\big(f-E(f)\big)\big]\delta t \end{aligned}$$

$$
\begin{aligned}
&+\frac{k(k-1)}{2} E\big[\big(y-E(y)\big)^{k-2}\Omega\big]\delta t + O\big(\delta t^2\big) \\
&= m_k(t) + kE\big[f\left(M_{k-1}(t) - m_{k-1}(t)\right)\big]\delta t \\
&+\frac{k(k-1)}{2} E\big[M_{k-2}(t)\Omega\big]\delta t + O\big(\delta t^2\big).
\end{aligned} \tag{16}
$$

The exact moment (9) and (16) 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 $p(x)=\phi(x;\mu,\sigma^2)$. For the generalized Gaussian filter we use $K>2$ moments with a density $p(x)=\phi(x;\mu,\sigma^2)\sum_{k=0}^{K} c_k H_k((x-\mu)/\sigma)$ (Hermite expansion).

3 Generalized Gauss–Hermite filtering

3.1 Time update

Extending the Gaussian filter (Appendix B), the densities are represented by the truncated Fourier series (Hermite expansion, Appendix C)

$$
p(y) = \phi\big(y;\mu,\sigma^2\big)\sum_{n=0}^{K} c_n H_n\big((y-\mu)/\sigma\big) := \phi(y)\,H(y), \tag{17}
$$

and expectation values occurring in the update equations are computed by Gauss–Hermite integration (Appendix A), including the non-Gaussian term

$$
H\big(y,\{\mu,m_2,\dots,m_K\}\big) = \sum_{n=0}^{K} c_n H_n\big((y-\mu)/\sigma\big) := H(y,K). \tag{18}
$$

For example, the mean equation (9) is (cf. (17))

$$
\dot{\mu}(t|t_i) = E\big[f\big(y(t),t\big)|Z^i\big] = \int f(y,t)p(y,t)\,dy \tag{19}
$$

$$
\approx \int f(y,t)\,\phi(y)\,H(y)\,dy \tag{20}
$$

$$
\approx \sum f(\eta_l,t)w_l H(\eta_l,K). \tag{21}
$$

Thus, the approximation of the density is followed by numerical integration. This may be interpreted as using a singular density $p(y)\approx\sum w_l H_l\delta(y-\eta_l)$ concentrated at certain points ("particles"). To 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

$$
\begin{aligned}
m_k(t+\delta t|t_i) &= m_k(t|t_i) + kE\big[f(y,t)\left(M_{k-1}(t|t_i) - m_{k-1}(t|t_i)\right)\big]\delta t \\
&+\frac{k(k-1)}{2} E\big[M_{k-2}(t|t_i)\Omega(y,t)\big]\delta t + O\big(\delta t^2\big)
\end{aligned} \tag{22}
$$

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 occurring in the moment equations does produce higher order moments which must be truncated (Challa et al. 2000) or approximated otherwise, e.g. Gaussian factorization (Singer 2006c).

3.2 Measurement update

3.2.1 Exact measurement update

The exact measurement update is given by the Bayes formula

$$p\big(y_{i+1}|Z^{i+1}\big) = p(z_{i+1}|y_{i+1})p\big(y_{i+1}|Z^{i}\big)/L_{i+1}, \tag{23}$$

$$L_{i+1} = \int p(z_{i+1}|y_{i+1})p\big(y_{i+1}|Z^{i}\big)\,dy_{i+1}, \tag{24}$$

$$\begin{aligned} p\big(y_{i+1}|Z^{i}\big) &= \phi\big(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)\big) \\ &\quad\times H\big(y_{i+1}, \big\{\mu(t_{i+1}|t_i), m_2(t_{i+1}|t_i), \ldots, m_K(t_{i+1}|t_i)\big\}\big) \end{aligned} \tag{25}$$

with a priori moments $\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i), m_k(t_{i+1}|t_i), k = 3, \ldots, K$, and using Gauss–Hermite integration the likelihood is

$$L_{i+1} = \sum_{l=1}^{m} p(z_{i+1}|\eta_l)w_l H(\eta_l), \tag{26}$$

$$\eta_l = \eta_l\big(\mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)\big). \tag{27}$$

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

$$\mu(t_{i+1}|t_{i+1}) = L_{i+1}^{-1}\sum_{l=1}^{m} \eta_l\, p(z_{i+1}|\eta_l)w_l H(\eta_l), \tag{28}$$

$$m_k(t_{i+1}|t_{i+1}) = L_{i+1}^{-1}\sum_{l=1}^{m} \big[\eta_l - \mu(t_{i+1}|t_{i+1})\big]^k p(z_{i+1}|\eta_l)w_l H(\eta_l), \tag{29}$$

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

$$\begin{aligned} p\big(y_{i+1}|Z^{i+1}\big) &= \phi\big(y_{i+1}; \mu(t_{i+1}|t_{i+1}), m_2(t_{i+1}|t_{i+1})\big) \\ &\quad\times H\big(y_{i+1}, \big\{\mu(t_{i+1}|t_{i+1}), \ldots, m_K(t_{i+1}|t_{i+1})\big\}\big) \end{aligned} \tag{30}$$

can be computed and inserted into the time update equations.

3.2.2 Approximate measurement update

The Bayes formula (23)

$$p\big(y_{i+1}|Z^{i+1}\big) \propto \phi\big(z_{i+1}; h(y_{i+1}), R_{i+1}\big) \\ \times \phi\big(y_{i+1}; \mu(t_{i+1}|t_i), \Sigma(t_{i+1}|t_i)\big) H(y_{i+1}) \tag{31}$$

can be approximated by using the normal correlation ((44), see Appendix B) as follows: The product of the two Gaussians is written approximately as (the formula is exact for linear measurements)

$$L_{0,i+1} \times \phi\big(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})\big), \tag{32}$$

where (setting $h(y_{i+1}) := h_{i+1}$ etc. and $^-$ for the generalized inverse)

$$\mu_0(t_{i+1}|t_{i+1}) = \mu(t_{i+1}|t_i) + \mathrm{Cov}\big(y_{i+1}, h_{i+1}|Z^i\big) \\ \times \big(\mathrm{Var}\big(h_{i+1}|Z^i\big) + R(t_{i+1})\big)^- \big(z_{i+1} - E\big[h_{i+1}|Z^i\big]\big), \tag{33}$$

$$\Sigma_0(t_{i+1}|t_{i+1}) = \Sigma(t_{i+1}|t_i) - \mathrm{Cov}\big(y_{i+1}, h_{i+1}|Z^i\big) \\ \times \big(\mathrm{Var}\big(h_{i+1}|Z^i\big) + R(t_{i+1})\big)^- \mathrm{Cov}\big(h_{i+1}, y_{i+1}|Z^i\big), \tag{34}$$

$$L_{0,i+1} = \phi\big(z_{i+1}; E\big[h_{i+1}|Z^i\big], \mathrm{Var}\big(h_{i+1}|Z^i\big) + R(t_{i+1})\big) \tag{35}$$

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

$$p\big(y_{i+1}|Z^{i+1}\big) = \phi\big(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})\big) \\ \times H\big(y_{i+1}, \big\{\mu(t_{i+1}|t_i), \dots, m_K(t_{i+1}|t_i)\big\}\big)/L_{1,i+1}, \tag{36}$$

$$L_{1,i+1} = \int \phi\big(y_{i+1}; \mu_0(t_{i+1}|t_{i+1}), \Sigma_0(t_{i+1}|t_{i+1})\big) \\ \times H\big(y_{i+1}, \big\{\mu(t_{i+1}|t_i), \dots, m_K(t_{i+1}|t_i)\big\}\big)\, dy_{i+1} \tag{37}$$

and the complete likelihood is $L = L_0 \times L_1$. 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 (28), since they involve only polynomials and not the exponential $p(z|y)$. In the case of linear measurements, (36) is exact. In contrast, Challa et al. (2000) used the (iterated) EKF update to obtain approximate a posteriori means and variances used in the Gauss–Hermite integration.

3.2.3 Improved exact measurement update

The approximate update can be used to improve the numerical properties of the Bayes update (23). 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 $\phi(z_{i+1}; h(y_{i+1}), R_{i+1})$ (cf. Tam and Hatzinakos 1997).

4 Example: Ginzburg–Landau model

The algorithm was tested by 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)

$$dy = -\left[\alpha y + \beta y^3\right] dt + \sigma\, dW(t) \tag{38}$$

with measurement equation

$$z_i = y_i + \epsilon_i, \tag{39}$$

$\mathrm{Var}(\epsilon_i) = R$. 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 $p_0(y) = \lim_{t\to\infty} p(t, y|x, s) \propto \exp(-\frac{2}{\sigma^2}\Phi(y))$ occurs. The parameters are chosen as $\psi = \{\alpha, \beta, \sigma, R\} = \{-1, 0.1, 2, 1\}$. Figure 1 shows 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. Figure 2 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 expansion orders higher than 10 are necessary for a good approximation. The sequence of measurement updates are displayed in Fig. 3. The GHF (left) always uses Gaussian densities, whereas the GGHF ($K = 10$; right) 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, $t = 4$, tenth row, $t = 15$, etc.). The measurements are linear, so the normal correlation update (36) 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: $B = 1/T \sum_t \nu_t$
- Error standard deviation: $C = [1/T \sum_t (\nu_t - B)^2]^{0.5}$

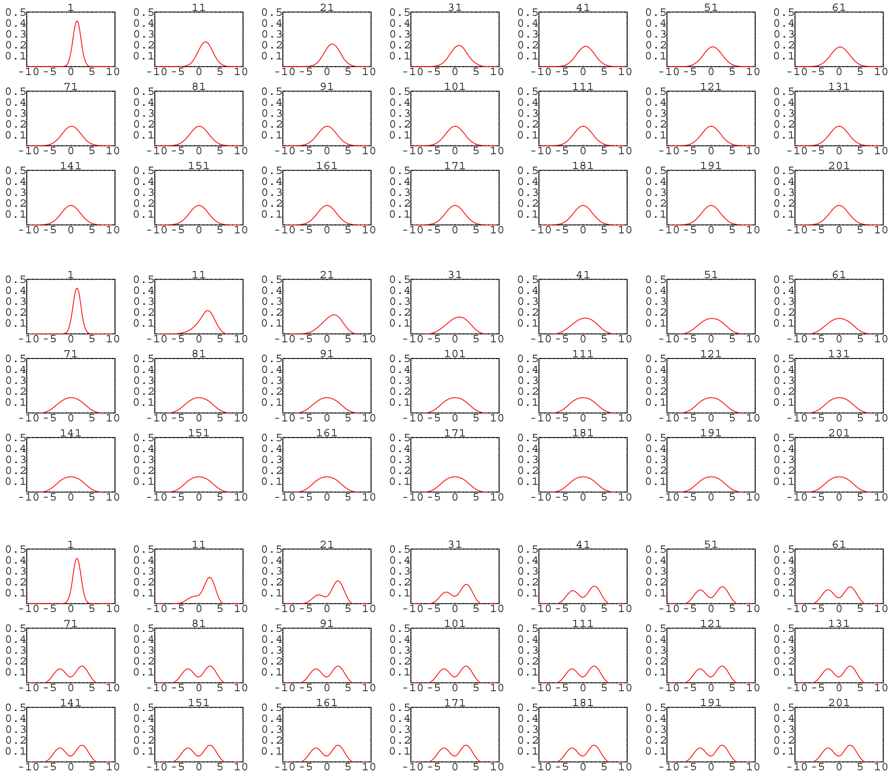

Fig. 1 Time evolution of the filter density $p(y,t)$: $K = 2$ (= GHF) (*top*), $K = 4$ (*middle*), and $K = 10$ moments (*bottom*)

- Mean and standard deviation of A, B, C in M samples: $\bar{X} = 1/M \sum_m X_m$; $\mathrm{std}(X) = [1/M \sum_m (X_m - \bar{X})^2]^{0.5}$; $X \in \{A, B, C\}$.

Table 1 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 $K = 2, 4, \ldots, 14$. 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($K > 2$), the Hermite correction $H(y, K) \sim O(y^K)$ must also 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

$$\begin{bmatrix} K & 2 & 4 & 6 & 8 & 10 & 12 & 14 \\ m & 4 & 9 & 13 & 17 & 21 & 25 & 29 \end{bmatrix}. \tag{40}$$

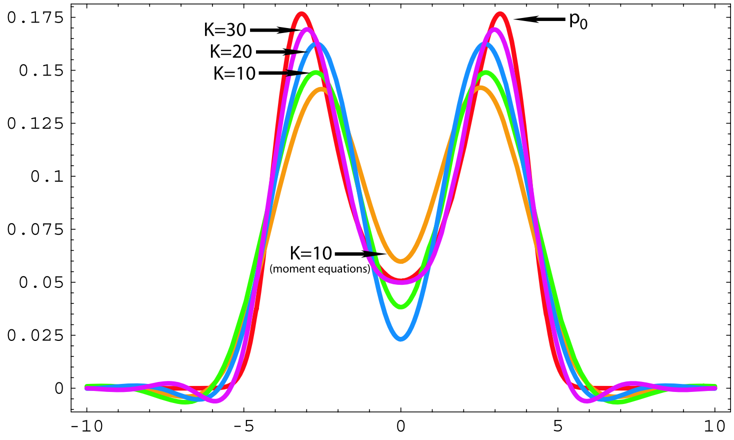

Fig. 2 Comparison of the exact stationary density p_0 (*red*), the Hermite expansion produced by the moment equations (*yellow*; $K = 10$) and computed from p_0 (*green*: $K = 10$; *blue*: $K = 20$, *purple*: $K = 30$). The density approximation may become negative locally

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

	$\bar{A}$	std(A)	$\bar{B}$	std(B)	$\bar{C}$	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

These integration orders were used for the GGHF. The performance of the GGHF in terms of squared filter error first improves with the degree of the density approxima-

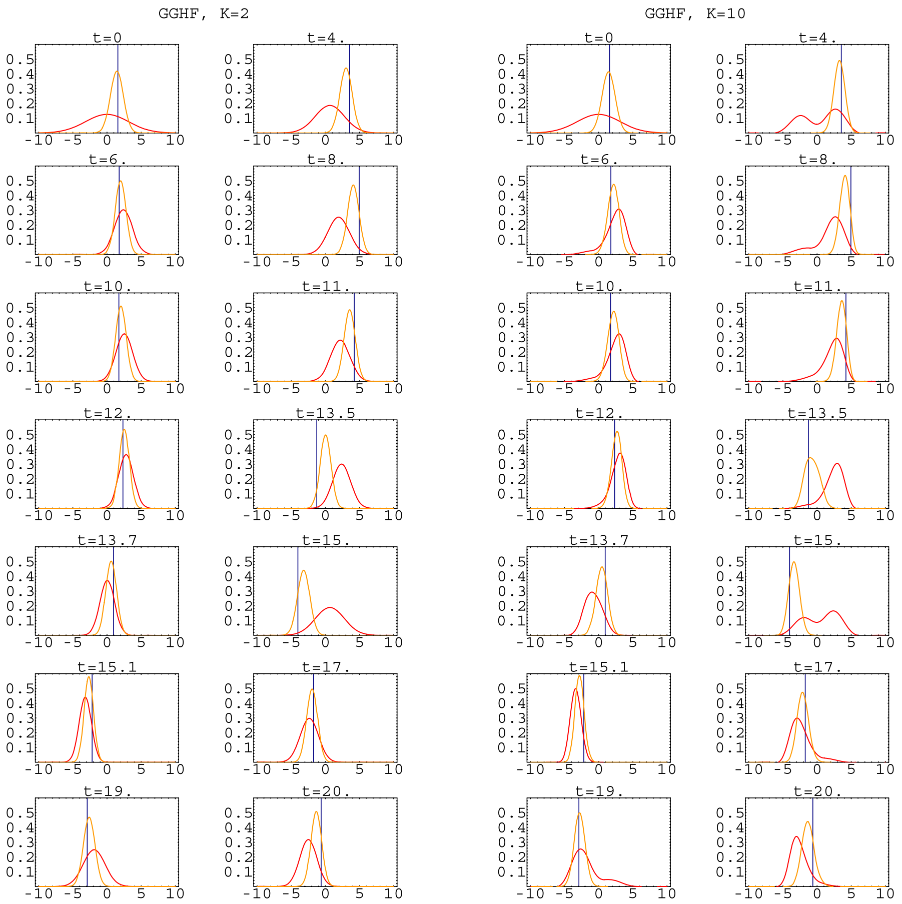

Fig. 3 A priori (*dark grey*) and a posteriori densities (*light grey*). $K = 2$ (*left*), $K = 10$ (*right*). The measurements are plotted by a *vertical line*

tion (order K of Hermite series $H(y, K)$), then deteriorates somewhat ($K = 8, 10$) and then improves again ($K = 12, 14$). Generally, the performance is better than the usual GHF ($K = 2$) as well as the UKF, although the differences are small in terms of the standard error. The EKF is outperformed by all algorithms.

Numerical remarks Since the Hermite expansion may become negative locally (cf. Fig. 2), the Hermite part $H(y, K)$ (18) 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.

5 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 were 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, cf. Singer 2007). In a model system, the non-Gaussian 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.

Acknowledgement I thank the editor and a referee for their detailed comments which helped me to improve the presentation of the paper.

Appendix A: Gauss–Hermite integration

The moment equations of Sect. 2 require the computation of expectations of the type $E[f(X)]$, where X is a 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

$$E_\phi\big[f(X)\big] = \int f(x)\phi\big(x;\mu,\sigma^2\big)\,dx = \int f(\mu+\sigma z)\phi(z;0,1)\,dz \tag{41}$$

$$\approx \sum_{l=1}^{m} f(\mu+\sigma\zeta_l)w_l = \sum_{l=1}^{m} f(\xi_l)w_l \tag{42}$$

may be approximated by Gauss–Hermite quadrature (cf. Ito and Xiong 2000). If such a double 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 $p_{\mathrm{GH}}(x) = \sum_{l=1}^{m} w_l\delta(x-\xi_l)$ concentrated at the quadrature points ξ_l. The Gaussian filter is equivalent to a Taylor expansion of f to higher orders L

$$E\big[f(X)\big] \approx \sum_{l=0}^{L} \tfrac{1}{l!} f^l(\mu)E[X-\mu]^l = \sum_{l=0}^{L} \tfrac{1}{l!} f^l(\mu)m_l \tag{43}$$

(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 (41) for $L\to\infty$. 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 2006c]. Challa et al. (2000) used truncated moment equations where moments higher than K are neglected.

Appendix B: Gauss–Hermite filtering

B.1 Continuous-discrete filtering scheme

The Gauss–Hermite filter GHF is a recursive sequence of time and measurement updates for the conditional moments μ and Σ, where expectation values are computed according to (41) using Gauss–Hermite quadrature ($^-$ denoting the generalized inverse; multivariate notation):

Initial condition: $t = t_0$

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

$i = 0, \dots, T-1$:
Time update: $t \in [t_i, t_{i+1}]$

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

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

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

B.2 Remarks

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 (8) and (15) 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 (15) 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 Shiryayev 2001, Chap. 13, Theorem 13.1, Lemma 14.1)

$$\mu(t_{i+1}|t_{i+1}) = \mu(t_{i+1}|t_i) + \mathrm{Cov}\big(y_{i+1}, z_{i+1}|Z^i\big)\mathrm{Var}\big(z_{i+1}|Z^i\big)^{-} \times \big(z_{i+1} - E\big[z_{i+1}|Z^i\big]\big), \tag{44}$$

$$\Sigma(t_{i+1}|t_{i+1}) = \Sigma(t_{i+1}|t_i) - \mathrm{Cov}\big(y_{i+1}, z_{i+1}|Z^i\big)\mathrm{Var}\big(z_{i+1}|Z^i\big)^{-} \times \mathrm{Cov}\big(z_{i+1}, y_{i+1}|Z^i\big), \tag{45}$$

with measurement (2) inserted and covariances computed by Gauss–Hermite integration. It is linear in z but includes the nonlinear measurements $z = h(y) + \epsilon$ in the expectation values and covariance terms. It does not require any Taylor expansions and can be used for discontinuous measurement functions as in threshold models (ordinal data). A direct implementation of the Bayes formula (4) would lead to the asymmetric a posteriori density

$$p\big(y_{i+1}|Z^{i+1}\big) = \sum_{l=1}^{m} w_l^* \delta(y_{i+1} - \eta_l), \tag{46}$$

$$w_l^* = w_l\, p(z_{i+1}|\eta_l) \Big/ \sum_{l=1}^{m} w_l\, p(z_{i+1}|\eta_l), \tag{47}$$

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

$$\mu(t_{i+1}|t_{i+1}) = \int y_{i+1}\, p\big(y_{i+1}|Z^{i+1}\big)\, dy_{i+1} = \sum_{l=1}^{m} w_l^* \eta_l, \tag{48}$$

$$\begin{aligned}\Sigma(t_{i+1}|t_{i+1}) &= \int \big(y_{i+1} - \mu(t_{i+1}|t_{i+1})\big) \\ &\quad \times \big(y_{i+1} - \mu(t_{i+1}|t_{i+1})\big)'\, p\big(y_{i+1}|Z^{i+1}\big)\, dy_{i+1} \\ &= \sum_{l=1}^{m} w_l^* \big[\eta_l - \mu(t_{i+1}|t_{i+1})\big]\big[\eta_l - \mu(t_{i+1}|t_{i+1})\big]'\end{aligned} \tag{49}$$

one can construct a symmetric a posteriori distribution with the same first and second moments (cf. Sect. 3.2.1).

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, 2004; Julier et al. 2000; Singer 2006a).

Appendix C: Hermite expansion

If the filter density strongly deviates from normality, a Fourier expansion in terms of Hermite polynomials may be used (Edgeworth series; cf. Kuznetsov et al. 1960; Abramowitz and Stegun 1965, Chap. 22; Courant and Hilbert 1968, Chap. II, 9; Srinivasan 1970; Challa et al. 2000; 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.

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) w(x)\, dx = n!\delta_{nm}. \tag{50}$$

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

$$\phi^{(n)}(x) := (d/dx)^n \phi(x) = (-1)^n \phi(x) H_n(x). \tag{51}$$

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

$$p(x) = \phi(x) \sum_{n=0}^{\infty} c_n H_n(x), \tag{52}$$

and the Fourier coefficients are given by

$$c_n := (1/n!) \int_{-\infty}^{\infty} H_n(x) p(x)\, dx = (1/n!) E\big[H_n(X)\big], \tag{53}$$

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 transform to the unstandardized density afterwards.

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 $c_0 = 1, c_1 = 0, c_2 = 0$,

$$c_3 := (1/3!) E\big[Z^3\big] = (1/3!)\nu_3,$$
$$c_4 := (1/4!) E\big[Z^4 - 6Z^2 + 3\big] = (1/24)(\nu_4 - 3),$$

[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. $\int q(x)^2\, dx = 2\pi \int \exp(x^2/2) p^2(x)\, dx < \infty$ (Courant and Hilbert 1968, pp. 81–82). An expansion of $p(x)$ in terms of $\psi_n(x)$ is convergent for square integrable p, but leads to complicated coefficients $b_n = E[\phi(x)^{1/2} H_n(x)]$.

and the standardized density expansion

$$p_z(z) := \phi(z)\big[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + \cdots\big] \tag{54}$$

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 $p_x(x)$

$$p_x(x) = \phi\big(x; \mu, \sigma^2\big) \sum_{n=0}^{\infty} c_n H_n\big((x - \mu)/\sigma\big) := \phi\big(x; \mu, \sigma^2\big)\, H(x). \tag{55}$$

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*

$$m_k := E[M_k] := E\big[(X - \mu)^k\big]. \tag{56}$$

References

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

Aït-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica **70**(1), 223–262 (2002)

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

Challa, S., Bar-Shalom, Y., Krishnamurthy, V.: Nonlinear filtering via generalized edgeworth series and Gauss-Hermite quadrature. IEEE Trans. Signal Process. **48**, 1816–1820 (2000)

Courant, R., Hilbert, D.: Methoden der Mathematischen Physik. Springer, Berlin (1968)

Daum, F.: Exact finite-dimensional nonlinear filters. IEEE Trans. Autom. Control **31**, 616–622 (1986)

Daum, F.: Nonlinear filters: beyond the Kalman filter. IEEE A&E Syst. Mag **20**(8), 57–69 (2005)

Ito, K., Xiong, K.: Gaussian filters for nonlinear filtering problems. IEEE Trans. Autom. Control **45**(5), 910–927 (2000)

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

Julier, S., Uhlmann, J.: 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)

Julier, S., Uhlmann, J.: Unscented Kalman filtering and nonlinear estimation. Proc. IEEE **92**(3), 401–422 (2004)

Julier, S., Uhlmann, J., Durrant-White, H.F.: A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Trans. Autom. Control **45**(3), 477–482 (2000)

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

Liptser, R., Shiryayev, A.: Statistics of Random Processes, vols. I, II, 2nd edn. Springer, New York (2001)

Miller, R., Ghil, M., Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems. J. Atmos. Sci. **51**(8), 1037–1056 (1994)

Nørgaard, M., Poulsen, N., Ravn, O.: New developments in state estimation for nonlinear systems. Automatica **36**, 1627–1638 (2000)

Singer, H.: Parameter estimation of nonlinear stochastic differential equations: simulated maximum likelihood vs. extended Kalman filter and Itô-Taylor expansion. J. Comput. Graph. Stat. **11**(4), 972–995 (2002)

Singer, H.: Continuous-discrete unscented Kalman filtering. Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 384, FernUniversität in Hagen. http://www.fernuni-hagen.de/FBWIWI/forschung/beitraege/pdf/db384.pdf (2006a)

Singer, H.: Generalized Gauss-Hermite filtering for multivariate diffusion processes. Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 402, FernUniversität in Hagen. http://www.fernuni-hagen.de/FBWIWI/forschung/beitraege/pdf/db402.pdf (2006b)

Singer, H.: Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models. Comput. Stat. **21**(3), 385–397 (2006c)

Singer, H.: Stochastic differential equation models with sampled data. In: van Montfort, K., Oud, H., Satorra, A. (eds.) Longitudinal Models in the Behavioral and Related Sciences. The European Association of Methodology (EAM) Methodology and Statistics Series, vol. II, pp. 73–106. Lawrence Erlbaum Associates, Mahwah (2007)

Srinivasan, K.: State estimation by orthogonal expansion of probability distributions. IEEE Trans. Autom. Control **15**(1), 3–10 (1970)

Tam, W., Hatzinakos, D.: An efficient radar tracking algorithm using multidimensional Gauss-Hermite quadratures. IEEE Int. Conf. Acoust. Speech Signal Process. **5**, 3777–3780 (1997)