FernUniversität in Hagen

Conditional Gauss-Hermite Filtering with Application to Volatility Estimation

Hermann Singer

Diskussionsbeitrag Nr. 430
Oktober 2008

Diskussionsbeiträge der Fakultät für Wirtschaftswissenschaft
der FernUniversität in Hagen

Herausgegeben von der Dekanin der Fakultät

Conditional Gauss–Hermite Filtering with Application to Volatility Estimation

Hermann Singer

FernUniversität in Hagen *

Abstract

The conditional Gauss–Hermite filter (CGHF) utilizes a decomposition of the filter density by conditioning on an appropriate part of the state vector. In contrast to the usual Gauss–Hermite filter (GHF) it is only assumed that the terms in the decomposition can be approximated by Gaussians. Due to the nonlinear dependence on the condition, quite complicated densities can be modeled, but the advantages of the normal distribution are preserved. For example, in stochastic volatility models, the joint density of return and volatility strongly deviates from a bivariate Gaussian, whereas the conditional density can be well approximated by a normal distribution. As in the GHF, integrals in the time and measurement updates can be computed by Gauss–Hermite quadrature.

Key Words: Multivariate stochastic differential equations; Nonlinear systems; Discrete time measurements; Continuous-discrete state space model; Conditionally gaussian densities; Stochastic volatility.

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany, hermann.singer@fernuni-hagen.de

1 Introduction

Diffusion processes with discrete time, noisy measurements (continuous-discrete state space models) have found applications in many fields, including engineering (Jazwinski; 1970; Daum; 1986; Ito and Xiong; 2000), finance and economics (Black and Scholes; 1973; Bergstrom; 1976; Gandolfo; 1981), physics (Haken; 1977; Stratonovich; 1989), and the social sciences (Coleman; 1968; Hamerle et al.; 1991).

A key tool to compute recursively the unobserved states and parameters is the Kalman filter algorithm (Kalman; 1960). Nonlinear systems usually do not allow exact solutions, except in special cases such as the Daum filter (Daum; 1986). Then, approximations relating to the nonlinear drift and diffusion functions or to the exact filter density must be employed.

The Gaussian filter (GF) assumes, that the true filter density $p(y)$ can be approximated by a Gaussian distribution $\phi(y)$. Thus, expectation values occuring in the time and measurement update can be computed numerically by Gauss–Hermite integration (GHF, cf. Ito and Xiong; 2000; Arasaratnam et al.; 2007). There are important applications, however, where the joint Gaussian assumption does not lead to satisfactory results. For example, if the volatility parameter of an Ornstein-Uhlenbeck process is filtered (Bayesian estimation), the measurements do not lead to any change in the conditional volatility state. This stems from the fact, that the state vector $(y(t), \sigma(t))$ strongly deviates from a bivariate Gaussian, since the process $y(t)$ is driven by the product of the Gaussian volatility and the Wiener process (cf. fig. 5). Similarly, stochastic volatility models (Scott; 1987; Hull and White; 1987; Harvey et al.; 2004; Jimenez et al.; 2006) are not satisfactorily filtered by the GHF and other filters relying on 2 moments, such as the extended Kalman filter EKF, or the unscented Kalman filter UKF (cf. Julier and Uhlmann; 1997; Julier et al.; 2000; Julier and Uhlmann; 2004).

One can solve the problem by using analytic density approximations, e.g. the Gaussian sum filter (Alspach and Sorenson; 1972; Ito and Xiong; 2000; Arasaratnam et al.; 2007), or by expanding the density into a Fourier series,

e.g. the Hermite expansion (Kuznetsov et al.; 1960; Challa et al.; 2000; Aït-Sahalia; 2002; Singer; 2008). Alternatively, Monte Carlo methods have found widespread applications (e.g. Elerian et al.; 2001; Beskos et al.; 2006; Cappé et al.; 2007; Fearnhead et al.; 2008).

In this paper, it is simply proposed to factorize the joint density of all states y by using the partitioned state $y = (y_1, y_2)$ such that $p(y_1|y_2)$ is (approximately) conditionally Gaussian. Then, $p(y_1|y_2)p(y_2) \approx \phi(y_1|y_2)\phi(y_2)$ and the numerical methods for the usual GHF can be adapted. [1]

In section 2, static conditionally Gaussian models are discussed. This is extended to dynamic models in sect. 3 and illustrated by the Ornstein-Uhlenbeck process (Bayesian estimation of the volatility parameter). Section 4 develops the general conditional Gaussian filter whereas in section 5 recursive ML estimation is compared with sequential filtering of σ using several approximate nonlinear filters.

2 Conditionally Gaussian models

2.1 Example 1

For example, if $y|\mu, \sigma \sim N(\mu, \sigma)$, the joint density

$$p(y, \mu, \sigma) \;=\; \phi(y|\mu, \sigma)p(\mu, \sigma) \propto (2\pi\sigma^2)^{-1/2} \exp[-\tfrac{1}{2}(y-\mu)^2)/\sigma^2] \quad (1)$$

is of Gaussian shape as a function of μ, but not for σ [cf. figs. 1– 2 where we set $\mu = 0$ and used a prior $p(\sigma) = \phi(\sigma; 2, 1)$]. The joint distribution $p(y, \sigma) = p(y|\sigma)p(\sigma)$ displays the variability in the variance of y. It cannot be well approximated by a bivariate Gaussian $\phi(y, \sigma)$, as would be the case for the Gaussian filter. From fig. 2 (right), it can be seen that the posterior mean $E[\sigma|y]$ depends on y and thus we obtain estimates of σ from observations y, although the covariance $\mathrm{Cov}(y, \sigma) = E[y\sigma] - E[y]E[\sigma] = E[E[y|\sigma]\sigma] - E[y]E[\sigma] = 0$, since $E[y|\sigma] = E[y] = \mu = 0$. In contrast, figs. 3–4 [setting

[1] I use the notation $\phi(y_1|y_2) = \phi(y_1; E[y_1|y_2], \mathrm{Var}[y_1|y_2])$ for the Gaussian density with conditional moments.

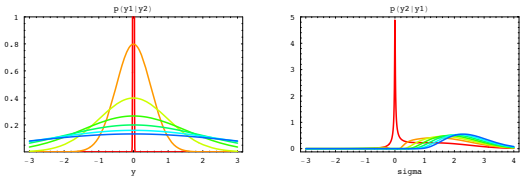


Figure 1: Conditional density $\phi(y|\sigma)$ and posterior density $p(\sigma|y)$ with prior $p(\sigma) = \phi(\sigma; 2, 1)$.

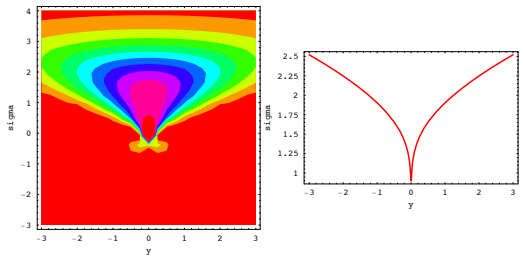


Figure 2: Joint density $\phi(y|\sigma)p(\sigma)$ and posterior mean $E[\sigma|y]$. It depends on y although $\mathrm{Cov}(y, \sigma) = 0$ (see text).

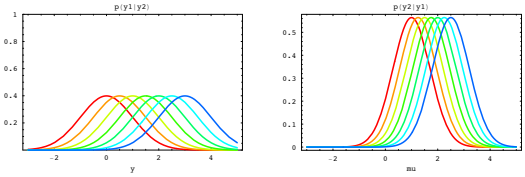


Figure 3: Conditional density $\phi(y|\mu)$ and posterior density $p(\mu|y)$ with prior $p(\mu) = \phi(\mu; 2, 1)$.

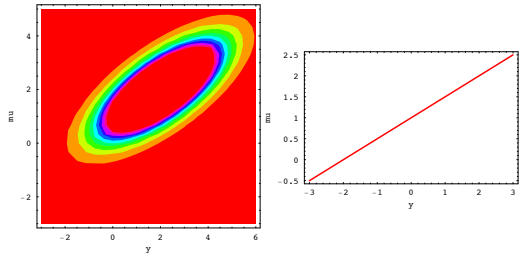


Figure 4: Joint density $\phi(y|\mu)p(\mu)$ and posterior mean $E[\mu|y]$.

$\sigma = 1$ and using a prior $p(\mu) = \phi(\mu; 2, 1)]$ display the joint distribution of y and μ. Since the role of these variables is symmetric in eqn. 1, the Gaussian shape is preserved and the regression $E[\mu|y]$ is linear. Thus, the estimation of parameters related to the mean (e.g. drift coefficients) is much simpler as compared to volatility parameters, where the regression $E[\sigma|y]$ is nonlinear (cf. fig. 2). In a bivariate Gaussian setting $\phi(y, \sigma)$, only a linear relation is possible (cf. 3) ■

Therefore, the idea is put forward, to represent the joint distribution of states and volatilites $p(y, \sigma)$ not by a joint Gaussian $\phi(y, \sigma)$, as in the Gaussian filter (or EKF, UKF), but by the product $\phi(y|\sigma)\phi(\sigma)$. This allows a fully nonlinear specification of the conditional moments

$$\begin{aligned} E[y|\sigma] &= \mu_1(\sigma) \\ \mathrm{Var}[y|\sigma] &= \Sigma_1(\sigma). \end{aligned} \tag{2}$$

In contrast, the joint Gaussian assumption only allows the normal correlation structure (Liptser and Shiryayev; 2001, ch. 13, theorem 13.1, lemma 14.1)

$$\begin{aligned} E[y|\sigma] &= E[y] + \mathrm{Cov}(y, \sigma)\mathrm{Var}(\sigma)^{-}(\sigma - E[\sigma]) \\ \mathrm{Var}[y|\sigma] &= \mathrm{Var}(y) - \mathrm{Cov}(y, \sigma)\mathrm{Var}(\sigma)^{-}\mathrm{Cov}(\sigma, y) \end{aligned} \tag{3}$$

which is linear in the conditional mean and independent of σ for the conditional variance ($^{-}$ denotes the generalized inverse). Put the other way round, a bivariate Gaussian can be obtained by a linear $\mu_1(\sigma)$ and constant $\Sigma_1(\sigma) = \Sigma_1$.

More generally, the distribution $p(y_1, y_2)$ of the vectors y_1, y_2 is not approximated by $\phi(y_1, y_2)$, but by

$$\begin{aligned} p(y_1, y_2) &= p(y_1|y_2)p(y_2) \approx \phi(y_1|y_2)\phi(y_2) \\ &= \phi(y_1; \mu_1(y_2), \Sigma_1(y_2))\, \phi(y_2; \mu_2, \Sigma_2), \end{aligned} \tag{4}$$

where the conditional moments $\mu_1(y_2) = E[y_1|y_2], \Sigma_1(y_2) = \mathrm{Var}(y_1|y_2)$ are nonlinear functions of the conditioning states y_2. Of course, the choice of y_2 depends on the form of the true distribution $p(y_1, y_2)$. It is chosen such, that

$p(y_1|y_2)$ is well approximated by a Gaussian with parameters $\mu_1(y_2), \Sigma_1(y_2)$. In example 1, we must condition on $y_2 = (\mu, \sigma)$ to get exactly $p(y_1|y_2) = \phi(y_1|y_2)$. A jointly Gaussian $\phi(y_1, y_2)$ is included as a special case (linear moments)

$$\begin{aligned} E[y_1|y_2] &= E[y_1] + \mathrm{Cov}(y_1, y_2)\mathrm{Var}(y_2)^-(y_2 - E[y_2]) \\ \mathrm{Var}[y_1|y_2] &= \mathrm{Var}(y_1) - \mathrm{Cov}(y_1, y_2)\mathrm{Var}(y_2)^-\mathrm{Cov}(y_2, y_1). \end{aligned} \tag{5}$$

3 State space models

We want to filter the *continuous-discrete state space model* (Jazwinski; 1970)

$$dy(t) = f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t) \tag{6}$$

where discrete time 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

$$z_i = h(y(t_i), t_i, \psi) + \epsilon_i. \tag{7}$$

In state equation (6), $W(t)$ 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 (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 (7), $\epsilon_i \sim N(0, R(t_i, \psi))i.d.$ 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 $Z^i = \{z(t_j)|j \leq i\}, t_i \leq t$ and h, R may depend on lagged measurements $Z^{i-1} = \{z(t_j)|j \leq i-1\}$ 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 below.

3.1 Example 2: Ornstein-Uhlenbeck process

The linear Gauss-Markov process is given by the SDE

$$dy(t) = \lambda y(t)dt + \sigma dW(t) \tag{8}$$

with measurement equation ($i = 0, \ldots, T$)

$$z_i = y_i + \epsilon_i \tag{9}$$

where $\psi = \{\lambda, \sigma, R = \mathrm{Var}(\epsilon_i)\}$ are unknown (nonrandom) parameters. For simplicity, let $\lambda = -1, R = 0.1$ be known. Then, σ can be estimated by exact ML or as Bayes estimator, using an extended state vector $\eta = \{y, \sigma\}$

$$dy = \lambda y dt + \sigma dW(t) \tag{10}$$

$$d\sigma = 0 \tag{11}$$

$$z_i = y_i + \epsilon_i. \tag{12}$$

The state extension leads to a nonlinear model in the variables, and thus to a nonlinear filtering problem.

As shown in fig. 5 (top), the usual GHF using the moments $\mu = E(\{y, \sigma\})$ and $\Sigma = \mathrm{Var}(\{y, \sigma\})$ cannot filter the volatility state $\sigma(t)$. The same applies to the EKF and UKF methods, since they use the first and second moments as well.

Now, if we note that $dy|\sigma = \lambda y dt + \sigma dW$ is Gaussian, the idea of the last section turns over to the dynamic context. Using the exact discrete model (EDM) at the measurement times t_i, setting $y_i = y(t_i)$ etc., we obtain

$$y_{i+1} = \lambda_i y_i + \sigma_i u_i \tag{13}$$

$$\sigma_{i+1} = \sigma_i \tag{14}$$

$$z_i = y_i + \epsilon_i \tag{15}$$

with the Gaussian error term $u_i = \int_{t_i}^{t_{i+1}} \exp[\lambda(t_{i+1} - s)]dW(s)$ and the AR(1) parameter $\lambda_i = \exp[\lambda(t_{i+1} - t_i)]$.

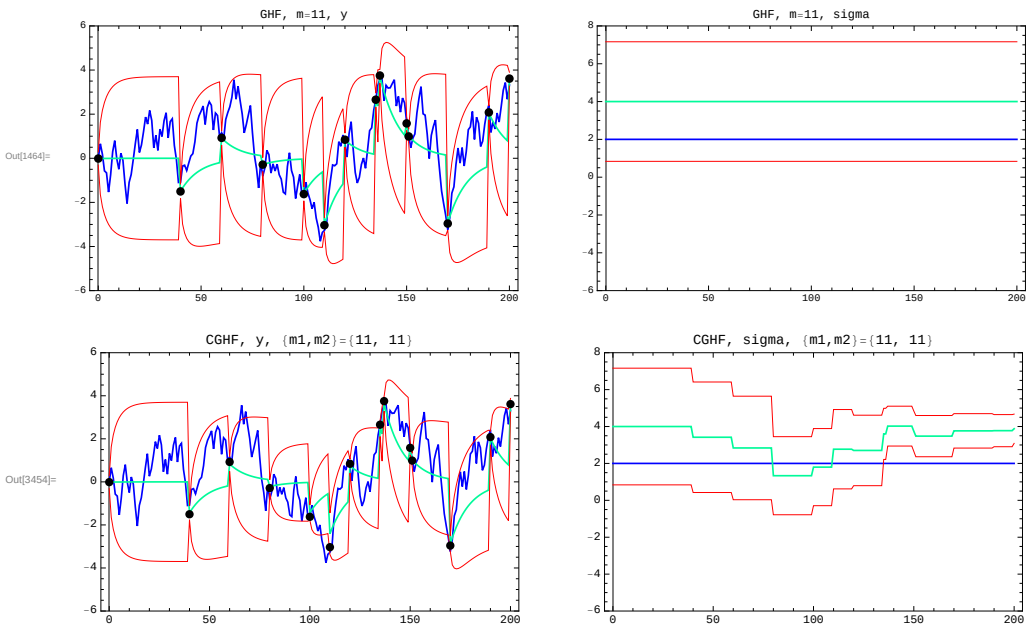


Figure 5: Ornstein-Uhlenbeck process $y(t)$ with volatility parameter σ. Top: Gauss–Hermite filter. Bottom: Conditional Gauss–Hermite filter. Left: true (blue) and filtered state (green) $\pm$ 1 standard deviation (red). Right: true and filtered volatility $\pm$ 1 standard deviation. Measurements are shown as dots.

3.1.1 Time update:

Assume that the posteriori density after measurement z_i is conditionally Gaussian, i.e. $p(y_i|\sigma_i, Z^i) = \phi(y_i; E[y_i|\sigma_i, Z^i], \mathrm{Var}[y_i|\sigma_i, Z^i])$ and $p(\sigma_i|Z^i) = \phi(\sigma_i|Z^i)$. Then, the time update $p(y_{i+1}|\sigma_{i+1}, Z^i)$ is again Gaussian with parameters

$$E[y_{i+1}|\sigma_{i+1}, Z^i] = \lambda_i E[y_i|\sigma_i, Z^i] \quad (16)$$

$$\mathrm{Var}[y_{i+1}|\sigma_{i+1}, Z^i] = \lambda_i \mathrm{Var}[y_i|\sigma_i, Z^i]\lambda_i' + \sigma_i \mathrm{Var}(u_i)\sigma_i' \quad (17)$$

since $\sigma_{i+1} = \sigma_i$ in this simple example.

3.1.2 Measurement update:

At the time of measurement t_{i+1} the Bayes formula

$$p(y_{i+1}, \sigma_{i+1}|z_{i+1}, Z^i) = \frac{p(z_{i+1}|y_{i+1}, \sigma_{i+1}, Z^i)p(y_{i+1}, \sigma_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} \quad (18)$$

can be evaluated easily due to the Gaussian densities (measurement and a priori density)

$$\begin{aligned} p(z_{i+1}|y_{i+1},\sigma_{i+1},Z^i) &= \phi(z_{i+1};y_{i+1},R) && (19)\\ p(y_{i+1},\sigma_{i+1}|Z^i) &= \phi(y_{i+1}|\sigma_{i+1},Z^i)\phi(\sigma_{i+1}|Z^i). && (20)\end{aligned}$$

Since the measurements are linear, the normal correlation update (3) is exact and one obtains

$$\begin{aligned} p(y_{i+1},\sigma_{i+1}|Z^{i+1}) &= \phi(y_{i+1}|\sigma_{i+1},Z^{i+1})p(\sigma_{i+1}|Z^{i+1}) && (21)\\ p(\sigma_{i+1}|Z^{i+1}) &= \phi(z_{i+1}|\sigma_{i+1},Z^i)\phi(\sigma_{i+1}|Z^i)/p(z_{i+1}|Z^i). && (22)\end{aligned}$$

Thus, the posterior of σ_{i+1} is nongaussian due to the nonlinear dependence of $\mathrm{Var}(z_{i+1}|\sigma_{i+1},Z^i)=\mathrm{Var}(y_{i+1}|\sigma_{i+1},Z^i)+R$ from σ_{i+1} (cf. 17). This nonlinear dependence is the reason why the posterior mean

$$\begin{aligned} E[\sigma_{i+1}|Z^{i+1}] &= \int \sigma_{i+1}p(\sigma_{i+1}|Z^{i+1})d\sigma_{i+1} && (23)\\ &= \int \sigma_{i+1}\phi(z_{i+1}|\sigma_{i+1},Z^i)\phi(\sigma_{i+1}|Z^i)d\sigma_{i+1}/p(z_{i+1}|Z^i)\end{aligned}$$

is a function of the measurements, in contrast to the usual GHF. The integral can be computed by Gauss–Hermite integration (see appendix A). From the posteriori moments $E[\sigma_{i+1}|Z^{i+1}]$ and $\mathrm{Var}(\sigma_{i+1}|Z^{i+1})$ one can construct a Gaussian distribution and proceed in the recursive filter algorithm with the next time update.

For the posterior mean of the state y_{i+1} we simply obtain the usual normal correlation update

$$\begin{aligned} E[y_{i+1}|\sigma_{i+1},Z^{i+1}] &= E[y_{i+1}|\sigma_{i+1},Z^i] && (24)\\ &+ \mathrm{Var}(y_{i+1}|\sigma_{i+1},Z^i)(\mathrm{Var}(y_{i+1}|\sigma_{i+1},Z^i)+R)^{-1}\\ &\times (z_{i+1}-E[y_{i+1}|\sigma_{i+1},Z^i])\end{aligned}$$

etc. The a priori terms are given in (16). Figs. 5 (top, bottom) display the difference in the performance of the GHF and the CGHF. In this picture, an Ornstein-Uhlenbeck process was simulated according to (8) with parameters $\psi=(\lambda,\sigma,R=\mathrm{Var}(\epsilon_i))=(-1,2,0.1)$ and sampling interval $\delta t=0.1$. I used a

simple Euler-Maruyama scheme (cf., e.g. Kloeden and Platen; 1992) on a grid $\tau_j = t_0 + j\delta t, j = 0, \ldots, J = (t_T - t_0)/\delta t$ with discretization interval $\delta t = 0.1$. The measurements were taken at times $\tau = \{0, 4, 6, 8, 10, 11, 12, 13.5, 13.7, 15, 15.1, 17, 19, 20\}$. Clearly, the Gauss-Hermite filter (fig. 5, top) does not filter the volatility process (Bayesian parameter) $d\sigma = 0$, whereas the CGHF, due to the conditional Gaussian filter density, yields estimates of σ from the observations $y(t_i)$ ■

4 Conditional Gauss–Hermite filtering

In this section we derive a sequence of time update and measurement update steps for the filter density $p(y_1, y_2, t|Z^i)$ which is approximated by the product of Gaussians

$$p(y_1, y_2, t|Z^i) \approx \phi(y_1, t|y_2, t, Z^i)\phi(y_2, t|Z^i). \tag{25}$$

The densities are evaluated at the time points $\tau_j = t_0 + j\delta t, j = 0, \ldots, J = (t_T - t_0)/\delta t$, and δt is an arbitrary (but small) discretization interval. The times of measurement are given by $t_i = \tau_{j_i}$. The filter proceeds in a recursive sequence of time update (dynamic moment equations) and measurement updates (Bayes formula).

According to the Gaussian assumption (25) one has to consider the conditional moments

$$E[y_1(t)|y_2(t), Z^i] = \mu_1(y_2(t), Z^i) \tag{26}$$

$$E[y_2(t)|Z^i] = \mu_2(t, Z^i) \tag{27}$$

$$\mathrm{Var}(y_1(t)|y_2(t), Z^i) = \Sigma_1(y_2(t), Z^i) \tag{28}$$

$$\mathrm{Var}(y_2(t)|Z^i) = \Sigma_2(t, Z^i) \tag{29}$$

and we seek recursive equations for their time evolution.

The state space model 6–7 is written in partitioned form ($y_1 : p_1 \times 1, g_1 : p_1 \times r$ etc.; dropping ψ)

$$dy_1(t) = f_1(y_1, y_2, t)dt + g_1(y_1, y_2, t)dW(t) \tag{30}$$

$$dy_2(t) = f_2(y_1, y_2, t)dt + g_2(y_1, y_2, t)dW(t) \tag{31}$$

with measurements at t_i

$$z_i = h(y_1(t_i), y_2(t_i), t_i) + \epsilon_i. \tag{32}$$

4.1 Time update

In a short time step δt, the Euler-Maruyama approximation for the Itô equations (30–31) is

$$\begin{aligned} y_1(t+\delta t) &= y_1(t) + f_1(y_1, y_2, t)\delta t + g_1(y_1, y_2, t)\delta W(t) && (33)\\ y_2(t+\delta t) &= y_2(t) + f_2(y_1, y_2, t)\delta t + g_2(y_1, y_2, t)\delta W(t) && (34) \end{aligned}$$

and we find the moment equations (dropping the dependence on Z^i)

$$\begin{aligned} E[y_1(t+\delta t)|y_2(t)] &= E[y_1(t)|y_2(t)] + E[f_1(y_1, y_2, t)|y_2(t)]\delta t && (35)\\ E[y_2(t+\delta t)] &= E[y_2(t)] + E[f_2(y_1, y_2, t)]\delta t && (36) \end{aligned}$$

The second moments read

$$\begin{aligned} \mathrm{Var}[y_1(t+\delta t)|y_2(t)] &= \mathrm{Var}[y_1(t)|y_2(t)] + \\ &\quad \mathrm{Cov}[y_1(t), f_1(y_1, y_2, t)|y_2(t)]\delta t + \\ &\quad \mathrm{Cov}[f_1(y_1, y_2, t), y_1(t)|y_2(t)]\delta t + \\ &\quad E[g_1 g_1'(y_1, y_2, t)|y_2(t)]\delta t && (37)\\ \mathrm{Var}[y_2(t+\delta t)] &= \mathrm{Var}[y_2(t)] + \\ &\quad \mathrm{Cov}[y_2(t), f_2(y_1, y_2, t)]\delta t + \\ &\quad \mathrm{Cov}[f_2(y_1, y_2, t), y_2(t)]\delta t + \\ &\quad E[g_2 g_2'(y_1, y_2, t)]\delta t. && (38) \end{aligned}$$

The expectation values on the right hand sides are with respect to the distributions $\phi(y_1(t)|y_2(t), Z^i)$ and $\phi(y_2(t)|Z^i)$ and can be evaluated using Gauss–Hermite quadrature (appendix A). For example

$$\begin{aligned} E[f_1(y_1, y_2, t)|y_2(t)] &= \int f_1(y_1, y_2, t)\phi(y_1; \mu_1(y_2), \Sigma_1(y_2))dy_1 && (39)\\ &\approx \sum_{l=1}^{L} f_1(\eta_{1lm}, \eta_{2m}, t)w_{1l} && (40) \end{aligned}$$

where

$$\eta_{2m} = \mu_2 + \Sigma_2^{1/2}\zeta_{2m} : p_2 \times 1 \tag{41}$$

$$\eta_{1lm} = \mu_1(\eta_{2m}) + \Sigma_1^{1/2}(\eta_{2m})\zeta_{1l} : p_1 \times 1 \tag{42}$$

are Gauss–Hermite sample points for the integration over y_2 and y_1 (conditional on the values $y_2 = \eta_{2m}$). The sample points depend on the conditional moments $\mu_2 = E[y_{2,t}|Z^i]$ and $\mu_1(y_2) = E[y_{1,t}|y_{2,t}, Z^i]$. Thus, one has y_1-sample points η_{1lm} for each y_2-coordinate $\eta_{2m}; l = 1, \ldots, L; m = 1 \ldots M$ (cf. fig. 6). Similarly,

$$\begin{aligned} E[f_2(y_1, y_2, t)] &= \int\int f_2(y_1, y_2, t)\phi(y_1; \mu_1(y_2), \Sigma_1(y_2)) \\ &\times \phi(y_2; \mu_2, \Sigma_2)dy_1dy_2 \\ &\approx \sum_{l,m=1}^{L,M} f_2(\eta_{1lm}, \eta_{2m}, t)w_{1l}w_{2m}. \end{aligned} \tag{43}$$

Now it is assumed that $E[y_1(t+\delta t)|y_2(t)] \approx E[y_1(t+\delta t)|y_2(t+\delta t)]$ etc. and using this approximation the time update is continued over the complete time interval $[t_i, t_{i+1}]$.

4.2 Measurement update

At time t_{i+1}, new measurements z_{i+1} come in, which are incorporated by using the Bayes formula (setting $y_{i+1} := y(t_{i+1})$ etc.)

$$p(y_{1,i+1}, y_{2,i+1}|z_{i+1}, Z^i) = \frac{p(z_{i+1}|y_{1,i+1}, y_{2,i+1})p(y_{1,i+1}, y_{2,i+1}|Z^i)}{p(z_{i+1}|Z^i)}. \tag{44}$$

The product of the measurement density

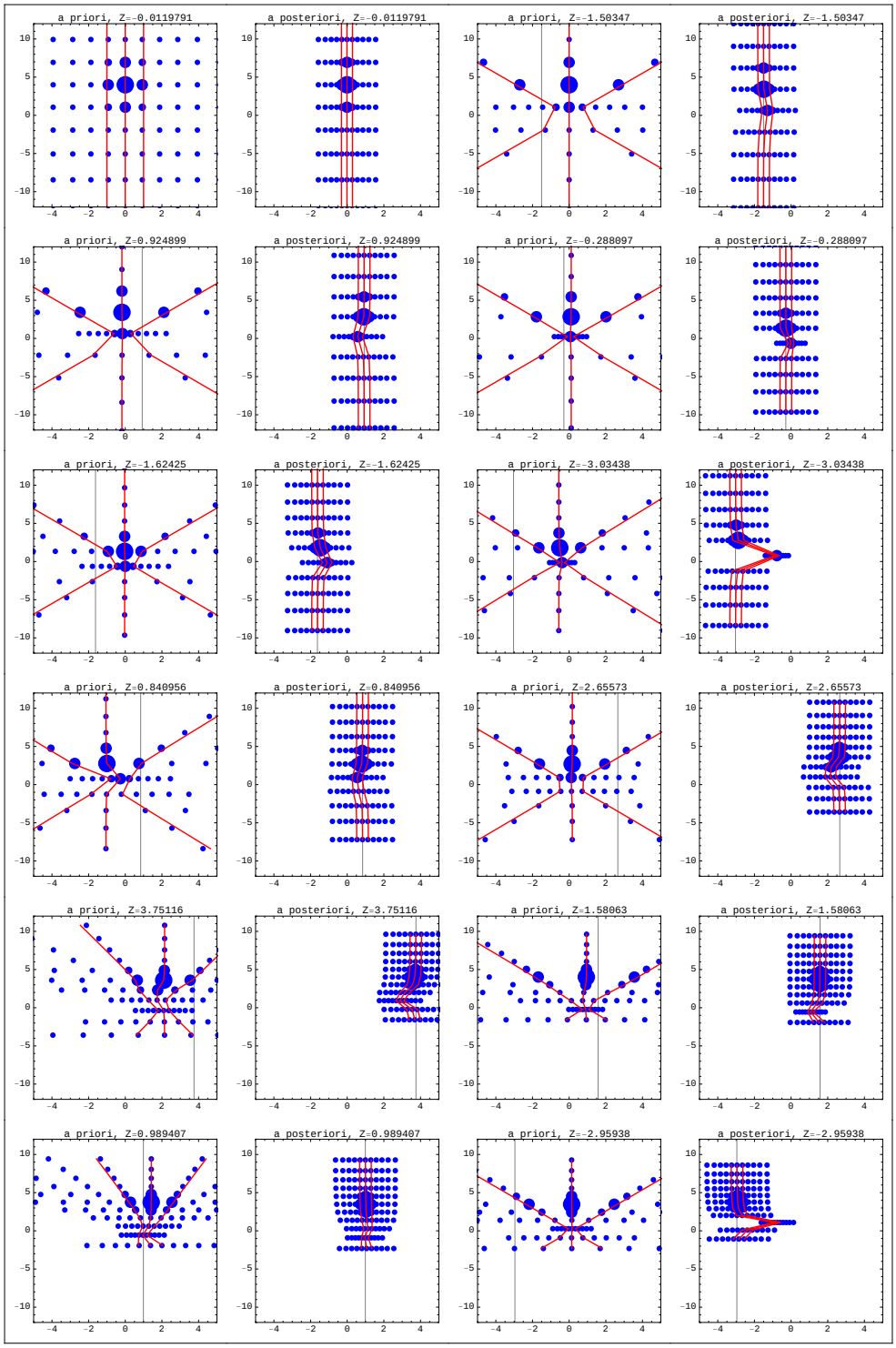
$$p(z_{i+1}|y_{1,i+1}, y_{2,i+1}) = \phi(z_{i+1}; h(y_{1,i+1}, y_{2,i+1}, t_{i+1}), R_{i+1}) \tag{45}$$

with the a priori distribution

$$p(y_{1,i+1}, y_{2,i+1}|Z^i) = \phi(y_{1,i+1}|y_{2,i+1}, Z^i) * \phi(y_{2,i+1}|Z^i) \tag{46}$$

can be evaluated approximately by the normal correlation update as

$$\phi(y_{1,i+1}|y_{2,i+1}, Z^{i+1}) * \phi(z_{i+1}|y_{2,i+1}, Z^i) * \phi(y_{2,i+1}|Z^i) \tag{47}$$



Out[3403]=

Figure 6: Conditional Gauss–Hermite sample points for the Ornstein-Uhlenbeck process ($L = 11, M = 11$) before and after the measurements. The weights are symbolized through the thickness of the dots. Also displayed is the conditional mean and standard deviation $\mu_1(\eta_{2m}) \pm \Sigma_1^{1/2}(\eta_{2m})$ (red).

where $\phi(z_{i+1}|y_{2,i+1}, Z^i) = \phi(z_{i+1}; E[h|y_{2,i+1}, Z^i], \mathrm{Var}[h|y_{2,i+1}, Z^i] + R_{i+1})$ is the conditional likelihood of z_{i+1} given $y_{2,i+1}$. The moments of the posterior of $y_1|y_2$ are given by

$$\begin{aligned} E[y_{1,i+1}|y_{2,i+1}, Z^{i+1}] &= E[y_{1,i+1}|F_{i+1}] + \mathrm{Cov}[y_{1,i+1}, z_{i+1}|F_{i+1}] \\ &\times \mathrm{Var}[z_{i+1}|F_{i+1}]^{-}(z_{i+1} - E[z_{i+1}|F_{i+1}]) \\ \mathrm{Var}[y_{1,i+1}|y_{2,i+1}, Z^{i+1}] &= \mathrm{Var}[y_{1,i+1}|F_{i+1}] - \mathrm{Cov}[y_{1,i+1}, z_{i+1}|F_{i+1}] \\ &\times \mathrm{Var}[z_{i+1}|F_{i+1}]^{-}\mathrm{Cov}[z_{i+1}, y_{1,i+1}|F_{i+1}] \end{aligned}$$

where $F_{i+1} = \{y_{2,i+1}, Z^i\}$ is shorthand for the conditioning variables. Now the moments of the priori distribution ($E[y_{1,i+1}|y_{2,i+1}, Z^i]$, $\mathrm{Var}[y_{1,i+1}|y_{2,i+1}, Z^i]$, $E[y_{2,i+1}|Z^i]$, $\mathrm{Var}[y_{2,i+1}|Z^i]$) are known from the time update and the expectations can be evaluated by Gauss–Hermite integration again. For example

$$\begin{aligned} E[z_{i+1}|y_{2,i+1}, Z^i] &= E[h(y_{1,i+1}, y_{2,i+1}, t_{i+1})|y_{2,i+1}, Z^i] &(48)\\ &\approx \sum_{l=1}^{L} h(\eta_{1lm}, \eta_{2m}, t_{i+1}) w_{1l} &(49) \end{aligned}$$

where again

$$\begin{aligned} \eta_{2m} &= \mu_2 + \Sigma_2^{1/2}\zeta_{2m} &(50)\\ \eta_{1lm} &= \mu_1(\eta_{2m}) + \Sigma_1^{1/2}(\eta_{2m})\zeta_{1l} &(51) \end{aligned}$$

are the Gauss–Hermite sample points evaluated at the a priori moments ($\mu_2 = E[y_{2,i+1}|Z^i]$, $\mu_1(y_2) = E[y_{1,i+1}|y_{2,i+1}, Z^i]$) etc.
The posteriori distribution for y_2 is given by (cf. 47)

$$p(y_{2,i+1}|Z^{i+1}) = \phi(z_{i+1}|y_{2,i+1}, Z^i)\phi(y_{2,i+1}|Z^i)/p(z_{i+1}|Z^i). \quad (52)$$

Now, since

$$\begin{aligned} E[z_{i+1}|y_{2,i+1}, Z^i] &= E[h(y_{1,i+1}, y_{2,i+1}, t_{i+1})|y_{2,i+1}, Z^i] &(53)\\ \mathrm{Var}[z_{i+1}|y_{2,i+1}, Z^i] &= \mathrm{Var}[h(y_{1,i+1}, y_{2,i+1}, t_{i+1})|y_{2,i+1}, Z^i] + R_{i+1} &(54) \end{aligned}$$

are in general nonlinear functions of $y_{2,i+1}$ (cf. example 2; 17), the measurement z_{i+1} is informative for the 'volatility state' $y_{2,i+1}$ and one obtains the

likelihood of observation z_{i+1} and the posterior mean of y_2

$$\begin{aligned} p(z_{i+1}|Z^i) &= \int \phi(z_{i+1}|y_{2,i+1}, Z^i)\phi(y_{2,i+1}|Z^i)dy_{2,i+1} \\ &\approx \sum_{m=1}^{M} \phi(z_{i+1}|\eta_{2m}, Z^i)w_{2m} \qquad (55) \\ E[y_{2,i+1}|Z^{i+1}] &= p(z_{i+1}|Z^i)^{-1} \int y_{2,i+1}\phi(z_{i+1}|y_{2,i+1}, Z^i) \\ &\times \phi(y_{2,i+1}|Z^i)dy_{2,i+1} \\ &\approx p(z_{i+1}|Z^i)^{-1} \sum_{m=1}^{M} \eta_{2m}\phi(z_{i+1}|\eta_{2m}, Z^i)w_{2m} \qquad (56) \end{aligned}$$

(analogously for $\mathrm{Var}[y_{2,i+1}|Z^{i+1}]$).

Starting from the a priori moments $\mu_1(y_2(t_0)) = E[y_1(t_0)|y_2(t_0)], \mu_2 = E[y_2(t_0)]$ etc. one obtains a recursive sequence of measurement and time updates for the moments and the Gauss–Hermite sample points (cf. fig. 6).

4.3 Complete filter

Putting together the measurement update and the time update, one obtains a recursive sequence of moments (26-29) at the measurement times t_i and for the time points $\tau_j = t_i + j\delta t, j = 1, \ldots, (t_{i+1} - t_i)/\delta t$ in between. The unconditional moments (w.r.t. y_2, dropping Z^i) can be computed from the filter terms as

$$\begin{aligned} E[y_1(t)] &= E[E[y_1(t)|y_2(t)]] \qquad (57) \\ \mathrm{Var}[y_1(t)] &= E[(y_1(t) - \mu_1(t))(y_1(t) - \mu_1(t))'] \\ &= E[\mathrm{Var}(y_1(t)|y_2(t))] + \mathrm{Var}(E[y_1(t)|y_2(t)]) \qquad (58) \end{aligned}$$

(residual variance + explained variance).

For the starting values I used $\mu_1(y_2(t_0)) = \mu_1, \Sigma_1(y_2(t_0)) = \Sigma_1$ (independent of y_2) and $\mu_2 = E[y_2(t_0)], \Sigma_2 = \mathrm{Var}[y_2(t_0)]$. Thus, the prior $p_0 = p(y_1(t_0)|y_2(t_0))p(y_2(t_0))$ is a Gaussian distribution with uncorrelated states $y_1(t_0), y_2(t_0)$. After the first measurement update one obtains η_{2m} and $\mu_1(\eta_{2m})$, i.e. the unknown function $\mu_1(y_2)$ is determined on the sample points η_{2m} (same for $\Sigma_1(y_2)$). Iterating, one obtains the regression functions $\mu_1(y_2)$ etc. in a nonparametric way.

5 Example 3: ML vs. recursive Bayesian estimation

5.1 Log volatility model

In example 2, the Ornstein-Uhlenbeck process was discussed. It is interesting to compare exact ML estimation of $\sigma = \exp(h)$ with (approximate) recursive Bayesian filtering.[2] As noted, the likelihood of the parameter h can be computed exactly by using the Kalman filter, i.e.

$$\begin{aligned}
\mu_{i+1|i} &= \lambda_i \mu_{i|i} \\
\Sigma_{i+1|i} &= \lambda_i \Sigma_{i|i} \lambda_i' + \Omega_i \\
\mu_{i+1|i+1} &= \mu_{i+1|i} + K_i(z_{i+1} - \mu_{i+1|i}) \\
K_i &= \Sigma_{i+1|i}(\Sigma_{i+1|i} + R)^- \\
\Sigma_{i+1|i+1} &= (I - K_i)\Sigma_{i+1|i} \\
L_{i+1}(z_{i+1}) &= \phi(z_{i+1}; \mu_{i+1|i}, \Sigma_{i+1|i} + R).
\end{aligned}$$

In the formulae above, $\lambda_i = \exp[\lambda(t_{i+1} - t_i)]$, K_i is the Kalman gain and $\Omega_i = \mathrm{Var}(\sigma u_i) = \sigma^2(1 - \exp(2\lambda \Delta t_i))/(2\lambda_i)$ is the variance of the system error u_i (cf. 13). As usual, $\mu_{i+1|i} = E[y_{i+1}|Z^i]$ etc. denotes the conditional expectations. Starting from a flat prior with $\Sigma_{0|-1} = 10$ one obtains the ML estimator by maximizing $l(h) = \sum_{i=0}^{t} \log(L_i)$. The ML estimator $\hat{h}(t)$ was computed recursively for the sampled data set $z(t_0), \ldots, z(t), t = t_0, \ldots, t_T$, where $t_0 = 0$ and $t_T = 20$. Actually, the computations were done on a grid with spacing (discretization interval) $\delta t = 0.1$ and $z_j = z(\tau_j), \tau_j = t_0 + j\delta t, t_i = \tau_{j_i}$. The values z_j between the measurements $z(t_i)$ are considered as missing (cf. Singer; 1995). The true parameter values were $\lambda = -1, \sigma = 2, R = 0.1$.

Fig. 7 shows a comparision of the sequential estimates of $h = \log(\sigma)$ using maximum likelihood (ML), the CGHF ($(L, M) = (3, 3), \ldots, (21, 21)$ sample points), the Gaussian sum filter implemented with EKF, GHF and UKF

[2] Here I use the log volatility $h = \log(\sigma)$ parametrization in order to avoid bimodality in the marginal distribution of σ.

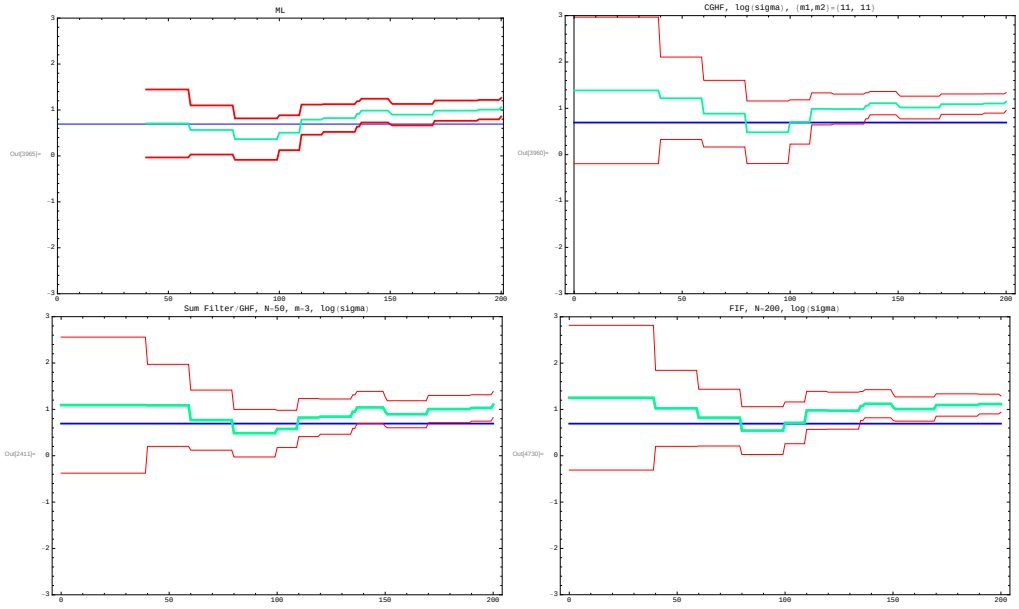


Figure 7: Discrete sampling: Sequential estimation of the parameter $h = \log \sigma$: Maximum likelihood (top, left), CGHF (top, right), Sum filter/GHF (bottom, left), Monte Carlo filter (bottom, right). Estimates $\hat{h}(t) \pm \mathrm{std}(\hat{h}(t))$ (see text).

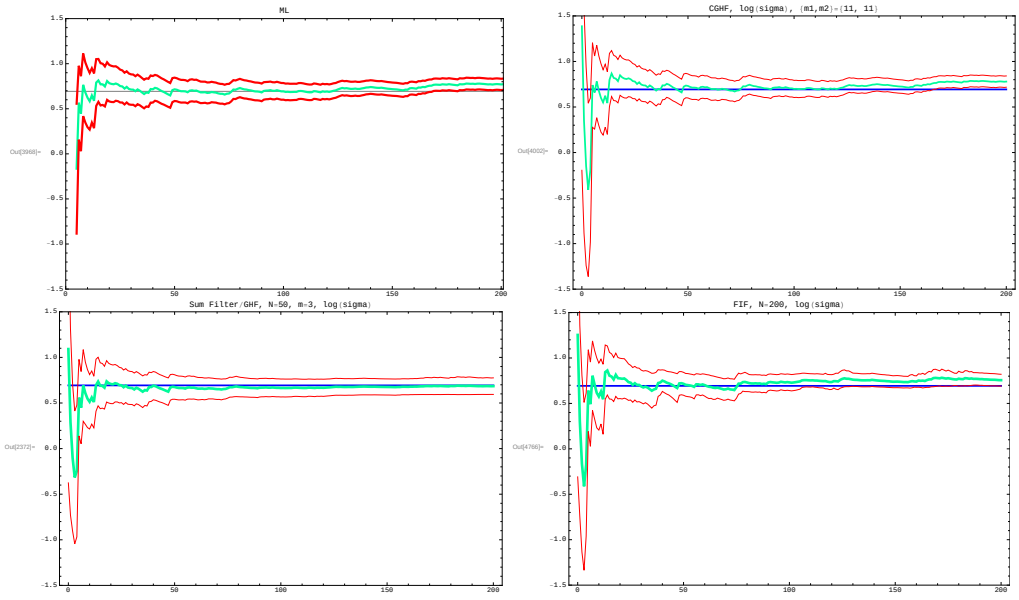


Figure 8: Quasi-continuous sampling: Sequential estimation of the parameter $h = \log \sigma$: Maximum likelihood (top, left), CGHF (top, right), Sum filter/GHF (bottom, left), Monte Carlo filter (bottom, right). Estimates $\hat{h}(t) \pm \mathrm{std}(\hat{h}(t))$ (see text).

discrete sampling					
method	CPU time (sec)	bias ($t=10$)	std ($t=10$)	bias ($t=20$)	std ($t=20$)
ML, linearized	19.11	0.1349	0.4111	0.0126	0.2150
nonlinear	18.25	0.1080	0.4112	-0.0184	0.2156
CGHF					
$(M,L)=(3,3)$	0.27	-0.2014	0.5188	-0.1499	0.2855
$(M,L)=(4,4)$	0.35	-0.0506	0.6166	-0.1017	0.3011
$(M,L)=(5,5)$	0.46	-0.1079	0.4803	-0.087	0.2256
$(M,L)=(11,11)$	1.30	-0.0216	0.4702	-0.0554	0.2190
$(M,L)=(21,21)$	3.71	0.0001	0.4639	-0.0491	0.2181
CUKF ($\kappa=1$)	0.27	-0.0961	0.4791	-0.0913	0.2404
CUKF ($\kappa=2$)	0.27	-0.2014	0.5188	-0.1499	0.2855
Sum filter $N=50$					
EKF	1.91	0.0752	0.4366	0.0133	0.2012
GHF ($M=3$)	5.79	0.0818	0.4358	0.0199	0.2001
UKF ($\kappa=1$)	5.69	0.0818	0.4358	0.0199	0.2001
Monte Carlo					
$N=50$	0.42	0.1085	0.5805	-0.0107	0.4392
$N=50$, antithetic, sorted	0.44	0.0383	0.4693	-0.0053	0.3466
$N=100$	0.88	0.0946	0.4839	-0.0079	0.2802
$N=100$, antithetic, sorted	0.93	0.0544	0.4632	-0.0391	0.2837

Table 1: Simulation study (discrete sampling): CPU times, bias and standard deviation of $\hat{h}(t), t=10,20$ for several algorithms. (M,L) denotes the number of Gauss-Hermite sample points for y and h, κ is the UKF tuning parameter, and N is the number of terms in the sum approximation (or Monte Carlo sample size).

updates and $N=50$ terms in the density expansion (GSF/EKF, GHF, UKF, cf. Ito and Xiong; 2000; Arasaratnam et al.; 2007, and appendix B), and a Monte Carlo (MC) filter with sample size $N=50,100$ (sample-importance-resample (SIR); for details, cf. Singer; 2002). Variance reduction (antithetic sampling) and sorting of the sample is used also (Kitagawa; 1996).

The data were measured at times $\tau=\{0,4,6,8,10,11,12,13.5,13.7,15,15.1,17,19,20\}$.

If denser measurements are used ($\tau=\{0,0.1,0.2,\ldots,20\}$), the convergence is more quickly (cf. fig. 8, using quasi-continuous sampling). Sequential ML works well, but for each estimate $\hat{h}(t)$, a nonlinear optimization problem has

continuous sampling					
method	CPU time (sec)	bias ($t=10$)	std ($t=10$)	bias ($t=20$)	std ($t=20$)
ML, linearized	93.86	0.0178	0.1067	0.0103	0.0690
nonlinear	151.71	-0.0296	0.1073	-0.0369	0.0694
CGHF					
$(M,L)=(3,3)$	0.77	0.0045	0.1153	0.0029	0.0778
$(M,L)=(4,4)$	1.04	0.0045	0.1110	0.0037	0.0707
$(M,L)=(5,5)$	1.35	0.0045	0.1095	0.0038	0.0708
$(M,L)=(11,11)$	3.91	0.0061	0.1099	0.0045	0.0707
$(M,L)=(21,21)$	11.32	0.0062	0.1099	0.0046	0.0706
CUKF ($\kappa=1$)	0.75	0.0080	0.1276	0.0082	0.0910
CUKF ($\kappa=2$)	0.79	0.0045	0.1153	0.0029	0.0778
Sum filter $N=50$					
EKF	4.24	0.0347	0.1126	0.0297	0.0703
GHF $M=3$	9.67	0.0416	0.1119	0.0348	0.0684
UKF $\kappa=1$	8.70	0.0416	0.1119	0.0348	0.0684
Monte Carlo					
$N=50$	3.04	0.0442	0.2636	0.0326	0.2476
$N=50$, antithetic, sorted	3.25	-0.0061	0.1862	-0.0109	0.1849
$N=100$	7.04	0.0137	0.1475	-0.0143	0.1306
$N=100$, antithetic, sorted	7.76	-0.0161	0.1497	-0.0069	0.1455

Table 2: Simulation study (continuous sampling): CPU times, bias and standard deviation of $\hat{h}(t), t=10,20$ for several algorithms. (M,L) denotes the number of Gauss-Hermite sample points for y and h, κ is the UKF tuning parameter, and N is the number of terms in the sum approximation (or Monte Carlo sample size).

to be solved. I used a quasi Newton algorithm with BFGS secant update and numerical score (Dennis Jr. and Schnabel; 1983). The estimate $\hat{h}(\tau_{j-1})$ was used as starting value for the next maximization at time τ_j. In order to avoid numerical problems, the shortest data set was $\{z_0,\ldots,z_{80}\}; z(\tau_{80})=8$. In contrast, the considered nonlinear filters work sequentially without iterative optimization.

5.2 Simulation study

In a simulation study with $M=100$ replications, the ML estimates and filter solutions were computed. CPU times, bias and standard errors for the

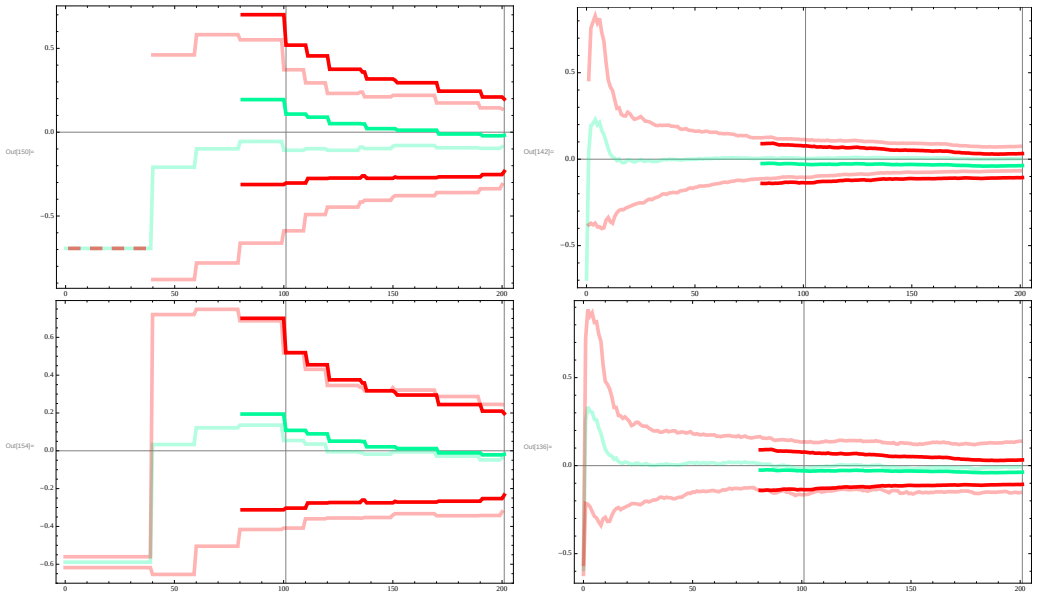


Figure 9: Simulation study ($M = 100$ replications):
Top: Bias (green) and standard error (red) for ML and CGHF(5,5) estimates (opaque colors). Left: discrete sampling, right: continuous sampling. The ML estimates were computed sequentially starting from $\tau_{80} = 8$ ($\delta t = 0.1$; see text).
Bottom: Bias (green) and standard error (red) for ML and MC ($N = 100$; antithetic, sorted) estimates (opaque colors). Left: discrete sampling, right: continuous sampling.

several algorithms are shown in tables 1 and 2 for discrete and continuous sampling (Mathematica 7, Intel Core 2 Duo 2.4 GHz).

The entries in column 3–6 are the 2nd components of the estimated bias $\bar{\nu}_j = M^{-1} \sum_m \nu_{mj}$ with $\nu_{mj} := h - E[y_{mj2}|Z_m^j]$ (filter) and $\nu_{mj} := h - \hat{h}_{mj}$ (ML), where $h = \log(2)$ is the true value of the log volatility. $E[y_{mj2}|Z_m^j]$ is the 2nd component of the filter solution at time j ($\tau_j = t_0 + j\delta t$) for replication m, whereas $\hat{h}_{mj}$ is the respective ML estimate. The tabulated times are $j = 100, 200$ ($t = 10, 20$). The standard deviation is defined as $\mathrm{std}_j = \sqrt{M^{-1} \sum_m (\nu_{mj} - \bar{\nu}_j)^2}$.

The prior distribution of the nonlinear filters was taken as $N(\mu_0, \Sigma_0), \mu_0 = \{0, \log(4)\}, \Sigma_0 = \mathrm{diag}(1, 10/4)$. This is motivated as follows: The true value of $\sigma = \exp(h) = 2$ is not known and was set to the wrong value 4. The prior variance Σ_0 should be set to $\mathrm{diag}(K_1, K_2)$ with $K_{1,2} \to \infty$. For the prior variance of σ, a (not too large) value of 10 was chosen, leading to

$\mathrm{Var}(\log(\sigma)) \approx \mathrm{Var}(\sigma)/\sigma^2 = 10/4$. Large K-values may lead to numerical instabilities in the filter algorithms.

Comparison of methods:

- discrete sampling

    1. ML

       The ML estimator is positively biased for the short data set ($t =$ 10, 5 data points), but the exact ML estimator performs better as the linearized ($\exp(\lambda\delta t) \approx 1 + \lambda\delta t$). At $t = 20$ (14 data points), the bias is negligible.

    2. CGHF

       The CGHF performed stable and fast (for a small number of Gauss-Hermite sample points), but is negatively biased for the short data set ($t = 10$).

    3. CUKF

       Alternatively, the integrations can be done by using the unscented transform (UT) with sigma points, leading to a conditional unscented Kalman filter (CUKF). The choice of the UKF scale parameter κ is somewhat arbitrary, however. The choice $\kappa = 2$ is equivalent to $M = 3$ in one dimension (Ito and Xiong; 2000).

    4. GSF

       The Gaussian sumfilter ($N = 50$) is considerably slower, especially using GHF and UKF updates, and the bias is comparable to CGHF(5,5).

    5. MC

       The Monte Carlo filter is fast and displays a bias comparable to ML, but the standard error is slightly higher.

- continuous sampling

In this case all algorithms perform well in terms of bias. However, the MC approach shows higher standard errors. This might be attributed to simulation sampling error.

6 Conclusion

We have shown how the filtering of volatility parameters can be achieved by a simple probabilistic assumption. Instead of taking the joint Gaussian $\phi(y_1, y_2)$ as approximate filter density, the conditional Gaussian product $\phi(y_1|y_2)\phi(y_2)$ was used, leading to a nonlinear dependence of $E[y_1|y_2]$ and $\mathrm{Var}(y_1|y_2)$ on y_2. In contrast, a joint Gaussian assumption can only accomodate a linear regression $E[y_1|y_2] = a + by_2$ and a constant variance $\mathrm{Var}[y_1|y_2] = \mathrm{Var}[y_1] - \mathrm{Cov}(y_1, y_2)\mathrm{Var}(y_2)^{-}\mathrm{Cov}(y_2, y_1)$. However, in stochastic volatility models, the variance of y_1 is dependent on y_2. The Gaussian product is the simplest assumption for this type of nonlinear problems and leads to an efficient and stable algorithm using Gauss-Hermite quadrature. Alternatively, the integrations can be done using the unscented transform (UT) leading to a conditional unscented Kalman filter (CUKF). Since the conditional Gauss-Hermite sample points η_{1lm}, η_{2m} are computed recursively, there is no need to specify a parametric form for the conditional expectations $E[y_1|y_2]$ and $\mathrm{Var}[y_1|y_2]$.

The simple algorithm performs well in relation to sum filters and Monte Carlo approaches, although these methods are more general. Especially, the MC approach is exact in the limit of large MC sample size.

Appendix A: Gauss–Hermite integration

The moment equations of the (C)GHF 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

$$\begin{aligned} E_\phi[f(Y)] &= \int f(y)\phi(y;\mu,\sigma^2)dy = \int f(\mu+\sigma z)\phi(z;0,1)dz \\ &\approx \sum_{l=1}^{m} f(\mu+\sigma\zeta_l)w_l = \sum_{l=1}^{m} f(\eta_l)w_l \end{aligned}$$

may be approximated by Gauss–Hermite quadrature (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 product of conditionally Gaussian densities $\phi(y_1|y_2)\phi(y_2)$ (CGHF) which again yields integrals w.r.t. the Gaussian density i.e. $E[f(Y)] = \int f(y_1,y_2)\phi(y_1|y_2)\phi(y_2)dy_1dy_2$.
In the multivariate case, the integration is performed using standardization with some matrix square root (e.g. the Cholesky decomposition)

$$\begin{aligned} E_\phi[f(Y)] &= \int f(y)\phi(y;\mu,\Sigma)dy \\ &= \int f(\mu+\Sigma^{1/2}z)\phi(z;0,I)dz_1...dz_p \\ &\approx \sum_{l_1,...,l_p} f(\mu+\Sigma^{1/2}\{\zeta_{l_1},...,\zeta_{l_p}\})w_{l_1,...,l_p} \\ &= \sum_{l_1,...,l_p} f(\eta_{l_1},...,\eta_{l_p})w_{l_1,...,l_p}, \end{aligned}$$

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-tuple of Gauss–Hermite quadrature points with weights $w_{l_1,...,l_p} = w_{l_1}...w_{l_p}$.

Appendix B: Sum filter

The filter density $p(y,t|Z^t) = p(y)$ may be approximated by conditioning on a discrete random variable I, so that

$$p(y) = \sum_i p(y|i)p_i \approx \sum_i \phi(y|\mu_i,\Sigma_i)p_i.$$

Thus it is assumed that the conditional distribution of Y in subsamples (groups) indexed by $I = i$ can be well described by gaussians. The (approx-

imate) mean and variance of y can be expressed as weighted sums

$$\begin{aligned} E[Y] &= \sum_i \mu_i p_i \\ \mathrm{Var}[Y] &= \sum_i \Sigma_i p_i + (\mu_i - \mu)(\mu_i - \mu)' p_i = SSW + SSB. \end{aligned}$$

Thus the total variance can be decomposed in a variance within and between groups. The exact equations for the moments $\mu(t) = E[Y(t)], \Sigma(t) = \mathrm{Var}[Y(t)]$ read (time update; $t \in [t_i, t_{i+1}]$)

$$\begin{aligned} \dot{\mu} &= E[f(Y)] \\ \dot{\Sigma} &= \mathrm{Cov}(f, Y) + \mathrm{Cov}(Y, f) + E[\Omega]. \end{aligned}$$

Inserting the expressions for the moments and the density expansion one gets

$$\sum_i \dot{\mu}_i p_i = \sum_i E_i[f(Y)] p_i \Rightarrow \dot{\mu}_i = E_i[f(Y)]$$

and

$$\dot{\Sigma} = \sum_i \dot{\Sigma}_i p_i + (d/dt)[(\mu_i - \mu)(\mu_i - \mu)'],$$

where the notation $E_i[Y] = \int y \phi(y|\mu_i, \Sigma_i) dy$ means averaging in the ith group. Writing

$$\begin{aligned} \mathrm{Cov}(f, Y) &= \sum_i E_i[f(Y)(Y - \mu_i + \mu_i - \mu)] p_i \\ &= \sum_i \mathrm{Cov}_i(f, Y) + E_i(f)(\mu_i - \mu) \end{aligned}$$

and inserting $\dot{\mu}_i = E_i[f]$ one obtains the exact equation

$$\dot{\Sigma}_i = \mathrm{Cov}_i(f, Y) + \mathrm{Cov}_i(Y, f) + E_i[\Omega]$$

for the second moments. These moment equations can be approximated in the EKF style (Taylor expansion of f and Ω, by using Gauss-Hermite integration or the unscented transform. In this way one obtains the Gaussian sum filter (GSF/EKF) of Alspach and Sorenson (1972) or variants such as GSF/GHF or GHF/UKF (Ito and Xiong; 2000).

References

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

Alspach, D. and Sorenson, H. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control* **17**: 439–448.

Arasaratnam, I., Haykin, S. and Elliott, R. J. (2007). Discrete-Time Nonlinear Filtering Algorithms Using Gauss Hermite Quadrature, *Proceedings of the IEEE* **95**: 953–977.

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

Bergstrom, A. (1976). Non Recursive Models as Discrete Approximations to Systems of Stochastic Differential Equations (1966), *in* A. Bergstrom (ed.), *Statistical Inference in Continuous Time Models*, North Holland, Amsterdam, pp. 15–26.

Beskos, A., Papaspiliopoulos, O., Roberts, G. and Fearnhead, P. (2006). Exact and efficient likelihood-based inference for discretely observed diffusion processes (with discussion), *Journal of the Royal Statistical Society Series B* **68**: 333–382.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–654.

Cappé, O., Godsill, S. J. and Moulines, E. (2007). An overview of existing methods and recent advances in sequential Monte Carlo, *Proc. IEEE*.

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.

Coleman, J. (1968). The mathematical study of change, *in* H. Blalock and B. A.B. (eds), *Methodology in Social Research*, McGraw-Hill, New York, pp. 428–478.

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

Dennis Jr., J. and Schnabel, R. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs.

Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood Inference for Discretely Observed Nonlinear Diffusions, *Econometrica* **69, 4**: 959–993.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2008). Particle filters for partially-observed diffusions, *Journal of the Royal Statistical Society, Series B* **70**: 755–777.

Gandolfo, G. (1981). *Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models*, North Holland, Amsterdam.

Haken, H. (1977). *Synergetics*, Springer, Berlin.

Hamerle, A., Nagl, W. and Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models, *Journal of Mathematical Sociology* **16, 3**: 201–220.

Harvey, A., Koopman, S. J. and Shephard, N. (2004). *State Space and Unobserved Component Models*, Cambridge University Press, Cambridge, U.K.

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.

Jimenez, J., Biscay, R. and Ozaki, T. (2006). Inference Methods for Discretely Oserved Continuous-Time Stochastic Volatility Models: A Commented Overview, *Asia-Pacific Financial Markets* **12**: 109–141.

Julier, S. and Uhlmann, J. (1997). A new extension of the Kalman filter to nonlinear systems, *The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control*, Orlando, Florida.

Julier, S. and Uhlmann, J. (2004). Unscented Kalman Filtering and Nonlinear Estimation, *Proceedings of the IEEE* **92, 3**: 401–422.

Julier, S., Uhlmann, J. and Durrant-White, H. F. (2000). A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators, *IEEE Transactions on Automatic Control* **45, 3**: 477–482.

Kalman, R. (1960). A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D: J. Basic Eng.* **82**: 35–45. http://www.cs.unc.edu/ ∼welch/media/pdf/Kalman1960.pdf.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, *Journal of Computational and Graphical Statistics* **5,1**: 1–25.

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin.

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.

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. (1995). Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values, *Econometric Theory* **11**: 721–735.

Singer, H. (2002). 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.

Singer, H. (2008). Generalized Gauss-Hermite Filtering, *Advances in Statistical Analysis* **92, 2**: 179–195.

Stratonovich, R. (1989). Some Markov methods in the theory of stochastic processes in nonlinear dynamic systems, *in* F. Moss and P. McClintock (eds), *Noise in nonlinear dynamic systems*, Cambridge University Press, pp. 16–71.