Langevin and Kalman Importance Sampling for Nonlinear Continuous-Discrete State Space Models

Hermann Singer

Diskussionsbeitrag Nr. 504

Februar 2017

Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung
FernUniversität in Hagen
Universitätsstraße 41
58084 Hagen
http://www.fernuni-hagen.de/ls_statistik/
Sekretariat.Statistik@FernUni-Hagen.de

Langevin and Kalman Importance Sampling for Nonlinear Continuous-Discrete State Space Models

Hermann Singer
FernUniversität in Hagen *

first draft, December 23, 2016

Abstract

The likelihood function of a continuous-discrete state space model with state dependent diffusion function is computed by integrating out the latent variables with the help of a Langevin sampler. The continuous time paths are discretized on a time grid in order to obtain a finite dimensional integration and densities w.r.t. Lebesgue measure. We use importance sampling, where the exact importance density is the conditional density of the latent states, given the measurements. This unknown density is either estimated from the sampler data or approximated by an estimated normal density. Then, new trajectories are drawn from this Gaussian measure. Alternatively, a Gaussian importance density is directly derived from an extended Kalman smoother with subsequent sampling of independent trajectories (extended Kalman sampling EKS). We compare the Monte Carlo results with numerical methods based on extended, unscented and Gauss-Hermite Kalman filtering (EKF, UKF, GHF) and a grid based solution of the Fokker-Planck equation between measurements. We use the repeated multiplication of transition matrices based on Euler transition kernels, finite differences and discretized integral operators. The methods are illustrated for the geometrical Brownian motion and the Ginzburg-Landau model.

Key Words

Continuous-discrete state space models; Stochastic differential equations; Simulated maximum likelihood; Langevin path sampling; Nonlinear Kalman filtering and smoothing; Kalman sampling

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany, Hermann.Singer@FernUni-Hagen.de

1 Introduction

Maximum likelihood estimation of sampled linear and nonlinear stochastic differential equations is of high practical importance since most data sets are only collected at discrete, often large sampling times, especially in economics and the social sciences. Nevertheless, exact ML estimation is possible, if one can compute the transition density between the sampling intervals.

In the linear case[1], the approach was pioneered by Bergstrom (1988), who introduced the 'exact discrete model', an exact autoregressive scheme valid at the measurement times with nonlinear parameter restrictions due to discrete sampling. The likelihood function can be computed recursively with Kalman filter methods (Jones; 1984; Jones and Tryon; 1987; Harvey and Stock; 1985; Hamerle et al.; 1991; Singer; 1990, 1993, 1995, 1998), or by representing the dynamic state space model as a structural equation model (SEM) (Oud and Jansen; 2000; Oud and Singer; 2008; Singer; 2012). In both approaches, the parameter matrices of the discrete time model are nonlinear matrix functions of the original parameters of the continuous time model. One can also formulate the EDM on a time grid which is finer than the sampling interval. Then, the nonlinear parameter functions can be linearized and their values over the whole sampling interval are implicitly generated by the filter (cf. Singer; 1995) or the SEM equations (Singer; 2012). With this so-called oversampling device, also linear (w.r.t. the parameters) SEM software like LISREL can be used.

In the nonlinear case, the transition density is explicitly known only for some special cases, e.g. the square root model (Feller; 1951). There are a variety of approaches to obtain approximations, e.g. Kalman filtering (Singer; 2002, 2011; Särkkä et al.; 2013), analytical approximations (Aït-Sahalia; 2002, 2008; Chang et al.; 2011; Li; 2013), Monte Carlo methods (Pedersen; 1995; Elerian et al.; 2001; Singer; 2002; Beskos et al.; 2006; Stramer et al.; 2010; Girolami and Calderhead; 2011; Särkkä et al.; 2013) and numerical solutions of the Fokker-Planck equation (Risken; 1989; Wei et al.; 1997).

In this paper, the likelihood function is computed by integrating out the latent variables of the state space model, such that only the marginal distribution of the measurements remains. This task is performed by using a Langevin sampler (Langevin; 1908; Roberts and Stramer; 2001, 2002; Hairer et al.; 2005, 2007; Apte et al.; 2007, 2008; Hairer et al.; 2009) combined with importance sampling. The unknown importance density is estimated from the sampler data in several ways.

We derive a conditional Markov representation which is estimated using kernel density and regression methods. Alternatively, the true importance density is replaced by an estimated Gaussian density. From this, new data are generated which have the same second order properties (mean and covariance function) as the orig-

[1] In order to avoid misunderstandings, one must distinguish between (non)linearity in the continuous time dynamical specification (differential equation) w.r.t. the state variables and in the derived 'exact discrete model' w.r.t. the parameters.

inal Langevin data. As a variant, the mean and covariance function are directly computed by a Kalman smoother, avoiding Langevin sampling. This approach will be called extended Kalman sampling (EKS).

In section 2, the continuous-discrete state space model is introduced. Then, Langevin path sampling is discussed in a finite dimensional approach. We analytically compute the negative gradient of the potential $-\log p(\eta|z)$ [log probability density of latent states given the data], which serves as the drift function of a Langevin equation. It is not assumed, that the diffusion function of the state space model is state independent. Section 4 discusses the maximum likelihood approach and the determination of the likelihood function. In section 4.1, two methods for computing the importance density are discussed, namely an estimation approach and a reference measure method. Then, in section 4.2, the likelihood is calculated by numerical integration, using Euler transition kernels and transition kernels derived from the Fokker-Planck equation (finite differences and integral operator approach). In section 5, applications such as the model of geometrical Brwonian motion and the bimodal Ginzburg-Landau model are considered, and the Fokker-Planck as well as Monte Carlo approach are compared with each other. In the appendix, a continuum limit for the Langevin sampler is considered.

2 Continuous-discrete state space model

Continuous time system dynamics and discrete time measurements (at possibly irregular measurement times $t_i, i = 0, ..., T$) can be unified by using the *nonlinear continuous-discrete state space model* (Jazwinski; 1970, ch. 6.2)

$$\begin{aligned} dY(t) &= f(Y(t), x(t), \psi)dt + g(Y(t), x(t), \psi)dW(t) && (1) \\ Z_i &= h(Y(t_i), x(t_i), \psi) + \epsilon_i;\ i = 0, ..., T. && (2) \end{aligned}$$

In (1), the state vector $Y(t) \in \mathbb{R}^p$ is a continuous time random process and the nonlinear drift and diffusion functions $f : \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^u \to \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^r$ depend on a u-dimensional parameter vector ψ. Furthermore, $x(t) \in \mathbb{R}^q$ are deterministic exogenous (control) variables. The system errors in (1) are increments of the Wiener process $W(t) \in \mathbb{R}^r$. Its formal derivative is Gaussian white noise $\zeta(t) = dW(t)/dt$ with zero mean and covariance function $E[\zeta_\rho(t)\zeta_{\rho'}(s)] = \delta_{\rho\rho'}\delta(t - s), \rho = 1, ..., r$, where $\delta(t - s)$ is the Dirac delta function (cf. Lighthill; 1958), and $\delta_{\rho\rho'}$ is the Kronecker delta symbol. Thus the process errors are independent for the times $t \neq s$ and components $\rho \neq \rho'$. The random initial condition $Y(t_0)$ is assumed to have a density $p_0(y, \psi)$ and is independent of $dW(t)$. The nonlinear state equation (1) is interpreted in the sense of Itô (see e.g., Arnold; 1974). Finally, the error term $\epsilon_i \sim N(0, R(x(t_i), \psi))$ in the measurement equation (2) is a discrete time white noise, independent for the times of measurement. It is assumed to be independent of the system error $dW(t)$ (cf. Jazwinski; 1970, p. 209-210).

3 Langevin path sampling: Finite dimensional approach

3.1 Likelihood and Langevin equation

In order to obtain finite dimensional densities w.r.t. Lebesgue measure and a numerically feasible approach, the SDE (1) is first replaced by the Euler-Maruyama approximation

$$\begin{aligned} \eta_{j+1} &= \eta_j + f(\eta_j, x_j, \psi)\delta t + g(\eta_j, x_j, \psi)\delta W_j \\ &:= \eta_j + f_j \delta t + g_j \delta W_j \end{aligned} \tag{3}$$

on a temporal grid $\tau_j = t_0 + j\delta t; j = 0, ..., J = (t_T - t_0)/\delta t$ (cf. Kloeden and Platen 1992, chs. 10.2, 14.1 and Stuart et al. 2004). The process noise is given as $\delta W_j = z_j\sqrt{\delta t}, z_j \sim N(0, I_r), i.i.d.$ and $x_j = x(\tau_j)$. The state variables are measured at times $t_i = \tau_{j_i}, j_i = (t_i - t_0)/\delta t$ according to

$$Z_i = h(\eta_{j_i}, x_{j_i}, \psi) + \epsilon_i. \tag{4}$$

The approximation error of the Euler scheme could be displayed by a superscript $\eta_j^{\delta t}$, but this is dropped for simplicity. Since $\eta_{j+1}|\eta_j$ is conditionally Gaussian, the finite dimensional density of $\eta = \{\eta_J, ..., \eta_0\}$ is given by [for clarity dropping the parameter and the exogenous variables; setting $\tilde{\eta} := \{\eta_J, ..., \eta_1\}$]

$$p(\eta) = \prod_{j=0}^{J-1} p(\eta_{j+1}|\eta_j)\; p(\eta_0) = p(\tilde{\eta}|\eta_0)p(\eta_0) \tag{5}$$

$$:= Z^{-1}e^{-S}\; p(\eta_0) \tag{6}$$

$$S = \tfrac{1}{2}\sum_{j=0}^{J-1}(\eta_{j+1} - \eta_j - f_j\delta t)'(\Omega_j\delta t)^{-1}(\eta_{j+1} - \eta_j - f_j\delta t) \tag{7}$$

$$Z = \prod_{j=0}^{J-1} |2\pi\Omega_j\delta t|^{1/2}, \tag{8}$$

where $\Omega_j = g_j g_j'$ is assumed to be nonsingular. Otherwise, one can use the singular normal distribution (cf. Mardia et al.; 1979, ch. 2.5.4, p. 41). In order to compute the likelihood function of the measured data $z = \{z_T, ..., z_0\}$, one can augment the density function with imputed variables, e.g. the latent states η, leading to

$$p(z) = \int p(z|\eta)p(\eta)d\eta. \tag{9}$$

The resulting high dimensional integration will be accomplished by Monte Carlo methods, and in comparision, by numerical integration. A direct approximation of (9) by the mean

$$p(z) \approx N^{-1}\sum_{l=1}^{N} p(z|\eta_l)$$

is extremely inefficient, since the unconditional random vectors $\eta_l \sim p(\eta)$ in the sum are mostly suppressed by the measurement density $p(z|\eta) = \prod_{i=0}^{T} \phi(z_i, h(\eta_{j_i}),\, R_i)$. This may be thought of defining windows in phase space where the trajectories η_l must pass through (cf. Kac; 1980). Using the idea of importance sampling, one can rewrite (9) as

$$p(z) \;=\; \int \frac{p(z|\eta)p(\eta)}{p(\eta|z)} p(\eta|z) d\eta, \tag{10}$$

where $p(\eta|z)$ is the optimal importance density (Kloeden and Platen; 1999, p. 519). Thus one must be able to draw conditional trajectories $\eta_l|z$, a task which can be accomplished by the Langevin approach (Parisi and Wu; 1981).

Introducing a $(J+1)p$-dimensional random process $\eta(u) = \eta_{j\alpha}(u); j = 0, ..., J, \alpha = 1, ..., p$ and a potential $\Phi(\eta) := -\log p(\eta)$, one may consider a Langevin equation in the fictious time dimension u

$$d\eta(u) \;=\; -\partial_\eta \Phi(\eta) du + \sqrt{2}\; dW(u), \tag{11}$$

where W is a $(J+1)p$-dimensional Wiener process and $\partial_\eta := \frac{\partial}{\partial \eta}$. Under the assumption of ergodicity, the (autocorrelated) trajectory $\eta(u)$ asymptotically ($u \to \infty$) draws data from the stationary distribution

$$p_{stat}(\eta) \;=\; e^{-\Phi(\eta)} = p(\eta).$$

This may be seen by considering the stationary Fokker-Planck equation

$$\partial_u p(\eta, u) \;=\; \sum_{j\alpha} \partial_{\eta_{j\alpha}} [(\partial_{\eta_{j\alpha}} \Phi(\eta)) p(\eta, u) + \partial_{\eta_{j\alpha}}\, p(\eta, u)] = 0 \tag{12}$$

for the density $p(\eta, u)$ (see, e.g. Risken; 1989, chs. 5.2, 6.0). Of course, one can directly draw independent vectors η from (3). The advantage of (11) is the possibility of drawing from $p(\eta|z)$, by using

$$\Phi(\eta|z) \;=\; -\log p(\eta|z) = -[\log p(z|\eta) + \log p(\eta) - \log p(z)] \tag{13}$$

as potential. The last term, which is the desired quantity (10), drops out by computing the gradient. Keeping a continuum limit $\delta t \to 0$ for eqn. (3) in mind (see appendix), the partial derivatives in the Langevin equation are scaled by the time discretization interval δt, leading to

$$d\eta(u) \;=\; -\delta_\eta \Phi(\eta|z) du + \sqrt{2}\; dW(u)/\sqrt{\delta t}. \tag{14}$$

Here we set $\delta_\eta := \frac{\partial}{\partial \eta \delta t} := \frac{\delta}{\delta \eta}$ in view of the functional derivative $\frac{\delta}{\delta \eta(t)}$.[2]

[2] The functional $\Phi(y)$ may be expanded to first order by using a linear functional $(\delta\Phi/\delta y)(h) = \int (\delta\Phi/\delta y(s)) h(s) ds$ (functional derivative). One has $\Phi(y+h) - \Phi(y) = (\delta\Phi/\delta y)(h) + O(||h||^2)$.

A discrete version is $\Phi(\eta) = \Phi(\eta_0, ..., \eta_J)$ and $\Phi(\eta + h) - \Phi(\eta) = \sum_j [\partial\Phi(\eta)/\partial(\eta_j \delta t)] h_j \delta t + O(||h||^2)$. As a special case, consider the functional $\Phi(\eta) = \eta_j$. Since $\eta_j + h_j - \eta_j = \sum (\delta_{jk}/\delta t) h_k \delta t$ one has the continuous analogue $y(t) + h(t) - y(t) = \int \delta(t-s) h(s) ds$, thus $\delta y(t)/\delta y(s) = \delta(t-s)$. See appendix

3.2 Drift function

In order to obtain an efficient sampling algorithm, the drift function in (14)

$$\delta_\eta \log p(\eta|z) \quad = \quad \delta_\eta[\log p(z|\eta) + \log p(\eta)]$$

must be computed analytically.[3]

Consider the term $\log p(\eta)$ (cf. eqn. 5). The Lagrangian (7), also called Onsager-Machlup functional (Onsager and Machlup; 1953) may be rewritten as

$$\begin{aligned} S &= \tfrac{1}{2}\sum_{j=0}^{J-1} \delta\eta_j'(\Omega_j \delta t)^{-1}\delta\eta_j && (15)\\ &- \sum_{j=0}^{J-1} f_j'\Omega_j^{-1}\delta\eta_j + \tfrac{1}{2}\sum_{j=0}^{J-1} f_j'\Omega_j^{-1}f_j\delta t && (16)\\ &:= S_0 + S_1 + S_2, \end{aligned}$$

$\delta\eta_j := \eta_{j+1} - \eta_j$. In a system without drift ($f = 0$), only the first term S_0 remains, corresponding to a random walk[4]. Therefore, the density $p(\eta)$ can be factorized as $p(\eta) = p_0(\eta)\ \alpha(\eta)$ where

$$\alpha(\eta) \quad = \quad \frac{p(\eta)}{p_0(\eta)} = \exp\left\{\sum_{j=0}^{J-1} f_j'\Omega_j^{-1}\delta\eta_j - \tfrac{1}{2}\sum_{j=0}^{J-1} f_j'\Omega_j^{-1}f_j\delta t\right\} \qquad (17)$$

is the density ratio and $p_0(\eta) = Z^{-1}\exp(-S_0)p(\eta_0)$ is the density of the driftless process including the initial condition. Thus one has the decomposition $\log p(\eta) = -\log Z - S_0 - S_1 - S_2 + \log p(\eta_0)$.

3.2.1 State independent diffusion coefficient

First we assume a state independent diffusion coefficient $\Omega_j = \Omega$, but later we set $\Omega_j = \Omega(\eta_j, x_j)$. This is important, if the Lamperti transformation does not lead to

[3] **Notation:** In the following, the components of vectors and matrices are denoted by greek letters, e.g. $f_\alpha, \alpha = 1, ..., p$, and partial deriatives by commas, i.e. $f_{\alpha,\beta} := \partial f_\alpha/\partial\eta_\beta = \partial_\beta f_\alpha = (f_\eta)_{\alpha\beta}$. The Jacobian matrix $\partial f/\partial\eta$ is written as f_η and its βth column as $(f_\eta)_{\bullet\beta}$. Likewise, $\Omega_{\alpha\bullet}$ denotes row α of matrix $\Omega_{\alpha\beta}$ and $\Omega_{\bullet\bullet} = \Omega$ for short.

Latin indices denote time, e.g. $f_{j\alpha} = f_\alpha(\eta_j)$. Furthermore, a sum convention is used for the greek indices (i.e. $f_\alpha g_\alpha = \sum_\alpha f_\alpha g_\alpha$). The difference operators $\delta = B^{-1} - 1$, $\nabla = 1 - B$, with the backshift $B\eta_j = \eta_{j-1}$ are used frequently. One has $\delta\cdot\nabla = B^{-1} - 2 + B := \Delta$ for the central second difference.

[4] In the case of a state dependent diffusion matrix, $\eta_{j+1} = \eta_j + g(\eta_j, x_j, \psi)\delta W_j$ generates a more general martingale process. Expression (16) remains finite in a continuum limit (see appendix).

constant coefficients in multivariate models.[5] In components, the term (15) reads

$$S_0 \;=\; \tfrac{1}{2}\sum_{j=0}^{J-1}(\eta_{j+1;\beta}-\eta_{j\beta})(\Omega_{\beta\gamma}\delta t)^{-1}(\eta_{j+1;\gamma}-\eta_{j\gamma}),$$

Note that $(\Omega_{\beta\gamma}\delta t)^{-1} \equiv [(\Omega\delta t)^{-1}]_{\beta\gamma}$ and the semicolon in $\eta_{j+1;\beta}$ serves to separate the indices, it is not a derivative. Differentiation w.r.t. the state $\eta_{j\alpha}$ yields ($j = 1, ..., J-1$)

$$\partial S_0/\partial(\eta_{j\alpha}\delta t) \;=\; -\Omega^{-1}_{\alpha\gamma}\delta t^{-2}(\eta_{j+1;\gamma}-2\eta_{j\gamma}+\eta_{j-1;\gamma}) \tag{18}$$

In vector notation, we have $\partial S_0/\partial(\eta_j\delta t) = -\Omega^{-1}\delta t^{-2}\Delta\eta_j$. On the boundaries $j = 0, j = J$ we obtain

$$\begin{aligned}\partial S_0/\partial(\eta_{0\alpha}\delta t) &= -\Omega^{-1}_{\alpha\gamma}\delta t^{-2}(\eta_{1\gamma}-\eta_{0\gamma})\\ \partial S_0/\partial(\eta_{0\alpha}\delta t) &= \Omega^{-1}_{\alpha\gamma}\delta t^{-2}(\eta_{J\gamma}-\eta_{J-1;\gamma})\end{aligned}$$

Next, the derivatives of $\log\alpha(\eta)$ are needed. One gets

$$\partial S_1/\partial(\eta_{j\alpha}\delta t) \;=\; -\delta t^{-1}[f_{j\beta,\alpha}\Omega^{-1}_{\beta\gamma}\delta\eta_{j\gamma}-\Omega^{-1}_{\alpha\gamma}(f_{j\gamma}-f_{j-1;\gamma})]$$

or in vector form, using difference operators

$$\partial S_1/\partial(\eta_{j\alpha}\delta t) \;=\; -\delta t^{-1}[f'_{j\bullet,\alpha}\Omega^{-1}\delta\eta_j-\Omega^{-1}\delta f_{j-1}], \tag{19}$$

where $f_{j\bullet,\alpha}$ is column α of the Jacobian $f_\eta(\eta_j)$. The second term yields

$$\begin{aligned}\partial S_2/\partial(\eta_{j\alpha}\delta t) &= \partial/\partial\eta_{j\alpha}\tfrac{1}{2}[f_{j\beta}\Omega^{-1}_{\beta\gamma}f_{j\gamma}] = f_{j\beta,\alpha}\Omega^{-1}_{\beta\gamma}f_{j\gamma}\\ &= f'_{j\bullet,\alpha}\Omega^{-1}f_j.\end{aligned} \tag{20}$$

Finally, one has to determine the drift component corresponding to the measurements, which is contained in the conditional density $p(z|\eta)$. Since it was assumed that the error of measurement is Gaussian (see 2), we obtain

$$p(z|\eta) \;=\; \prod_{i=0}^{T}p(z_i|\eta_{j_i}) = \prod_{i=0}^{T}\phi(z_i;h_i,R_i),$$

where $\phi(y;\mu,\Sigma)$ is the multivariate Gaussian density, $h_i = h(\eta_{j_i}, x_{j_i})$ is the output function and $R_i = R(x_{j_i})$ is the measurement error covariance matrix. Thus the

[5]These are called irreducible diffusions. A transformation $z = h(y)$ leading to unit diffusion for z must fulfil the system of differential equations $h_{\alpha,\beta}g_{\beta\gamma} = \delta_{\alpha\gamma}, \alpha, \beta = 1, ..., p; \gamma = 1, ..., r$. The inverse transformation $y = v(z)$ fulfils $v_{\alpha,\gamma}(z) = g_{\alpha\gamma}(v(z))$. Thus $v_{\alpha,\gamma\delta} = g_{\alpha\gamma,\epsilon}v_{\epsilon,\delta} = v_{\alpha,\delta\gamma} = g_{\alpha\delta,\epsilon}v_{\epsilon,\gamma}$. Inserting v one obtains the commutativity condition $g_{\alpha\gamma,\epsilon}\; g_{\epsilon\delta} = g_{\alpha\delta,\epsilon}\; g_{\epsilon\gamma}$. which is necessary and sufficient for reducibility. See Kloeden and Platen (1992, ch. 10, p. 348), Aït-Sahalia (2008).

derivative reads (matrix form in the second line)

$$\begin{aligned}\partial \log p(z|\eta)/\partial(\eta_{j\alpha}\delta t) &= \sum_{i=0}^{T} h_{i\gamma,\alpha} R^{-1}_{i\beta\gamma}(z_{i\beta} - h_{i\beta})(\delta_{jj_i}/\delta t) \\ &= \sum_{i=0}^{T} h'_{i\bullet,\alpha} R^{-1}_{i}(z_i - h_i)(\delta_{jj_i}/\delta t) \end{aligned} \tag{21}$$

The Kronecker symbol δ_{jj_i} only gives contributions at the measurement times $t_i = \tau_{j_i}$. Together we obtain for the drift of the Langevin equation (14)

$$\begin{aligned}\delta_\eta \log p(\eta|z) &= \delta_\eta[\log p(z|\eta) + \log p(\eta)] \\ &= (21) - (18 + 19 + 20) + \delta_\eta \log p(\eta_0). \end{aligned} \tag{22}$$

Here, $p(\eta_0)$ is an arbitrary density for the initial latent state.

3.2.2 State dependent diffusion coefficient

In the case of $\Omega_j = \Omega(\eta_j, x_j)$ the expressions get more complicated. The derivative of S_0 now reads

$$\partial S_0/\partial(\eta_{j\alpha}\delta t) = \delta t^{-2}[\Omega^{-1}_{j-1;\alpha\beta}\delta\eta_{j-1;\beta} - \Omega^{-1}_{j\alpha\beta}\delta\eta_{j\beta} + \tfrac{1}{2}\delta\eta_{j\beta}\Omega^{-1}_{j\beta\gamma,\alpha}\delta\eta_{j\gamma}], \tag{23}$$

$\Omega^{-1}_{j\beta\gamma,\alpha} \equiv (\Omega^{-1})_{j\beta\gamma,\alpha}$. A closer relation to expression (18) may be obtained by the Taylor expansion

$$\Omega^{-1}_{j-1;\alpha\beta} = \Omega^{-1}_{j\alpha\beta} + \Omega^{-1}_{j\alpha\beta,\gamma}(\eta_{j-1;\gamma} - \eta_{j\gamma}) + O(||\delta\eta_{j-1}||^2) \tag{24}$$

leading to

$$\begin{aligned}\partial S_0/\partial(\eta_{j\alpha}\delta t) &= -\Omega^{-1}_{j\alpha\beta}\delta t^{-2}(\eta_{j+1;\beta} - 2\eta_{j\beta} + \eta_{j-1;\beta}) \\ &- \Omega^{-1}_{j\alpha\beta,\gamma}\delta t^{-2}\delta\eta_{j-1;\beta}\delta\eta_{j-1;\gamma} + O(\delta t^{-2}||\delta\eta_{j-1}||^3) \\ &+ \tfrac{1}{2}\Omega^{-1}_{j\beta\gamma,\alpha}\delta t^{-2}\delta\eta_{j\beta}\delta\eta_{j\gamma}. \end{aligned} \tag{25}$$

In the state dependent case also the derivative of the Jacobian term $\log Z^{-1} = -\frac{1}{2}\sum_j \log|2\pi\Omega_j\delta t|$ is needed. Since the derivative of a log determinant is

$$\partial \log|\Omega|/\partial\Omega_{\alpha\beta} = \Omega^{-1}_{\beta\alpha},$$

one obtains

$$\partial \log Z^{-1}/\partial(\eta_{j\alpha}\delta t) = -\tfrac{1}{2}\delta t^{-1}\Omega^{-1}_{j\beta\gamma}\Omega_{j\beta\gamma,\alpha} = -\tfrac{1}{2}\delta t^{-1}\mathrm{tr}[\Omega_j^{-1}\Omega_{j,\alpha}],$$

$\Omega_{j,\alpha} = \Omega_{j\bullet\bullet,\alpha}$ for short. Using the formula $\Omega_j\Omega_j^{-1} = I; \Omega_{j,\alpha} = -\Omega_j\Omega^{-1}_{j,\alpha}\Omega_j$ we find

$$\partial \log Z^{-1}/\partial(\eta_{j\alpha}\delta t) = \tfrac{1}{2}\delta t^{-1}\mathrm{tr}[\Omega^{-1}_{j,\alpha}\Omega_j]. \tag{26}$$

The contributions of S_1 and S_2 are now (see 16)

$$\begin{aligned}&\partial S_1/\partial(\eta_{j\alpha}\delta t) = \\ &\quad -\delta t^{-1}[f_{j\beta,\alpha}\Omega^{-1}_{j\beta\gamma}\delta\eta_{j\gamma} - (\Omega^{-1}_{j\alpha\gamma}f_{j\gamma} - \Omega^{-1}_{j-1;\alpha\gamma}f_{j-1;\gamma}) + f_{j\beta}\Omega^{-1}_{j\beta\gamma,\alpha}\delta\eta_{j\gamma}]\end{aligned} \tag{27}$$

$$\partial S_2/\partial(\eta_{j\alpha}\delta t) \;=\; f_{j\beta,\alpha}\Omega^{-1}_{j\beta\gamma}f_{j\gamma} + \tfrac{1}{2}f_{j\beta}\Omega^{-1}_{j\beta\gamma,\alpha}f_{j\gamma}. \tag{28}$$

It is interesting to compare the terms in (23, 27, 28) depending on the derivative $\Omega^{-1}_{j\beta\gamma,\alpha}$, which read in vector form

$$\tfrac{1}{2}\delta t^{-2}\mathrm{tr}[\Omega^{-1}_{j,\alpha}\delta\eta_j\delta\eta_j'] - \delta t^{-1}\mathrm{tr}[\Omega^{-1}_{j,\alpha}\delta\eta_j f_j'] + \tfrac{1}{2}\mathrm{tr}[\Omega^{-1}_{j,\alpha}f_j f_j'],$$

and the Jacobian derivative (26). The terms can be collected to yield

$$\tfrac{1}{2}\delta t^{-2}\mathrm{tr}\{\Omega^{-1}_{j,\alpha}[\Omega_j\delta t - (\delta\eta_j - f_j\delta t)(\delta\eta_j - f_j\delta t)']\}, \tag{29}$$

as may be directly seen from the Lagrangian (7).

In summary, the Langevin drift component $(j\alpha), j = 0, ...J; \alpha = 1, ..., p$ is in vector-matrix form

$$\begin{aligned}\delta_{\eta_{j\alpha}}\log p(\eta|z) &= \delta_{\eta_{j\alpha}}[\log p(z|\eta) + \log p(\eta)] \\ &= \sum_{i=0}^{T} h'_{i\bullet,\alpha}R_i^{-1}(z_i - h_i)(\delta_{jj_i}/\delta t) \\ &+ \delta t^{-2}[\Omega^{-1}_{j\alpha\bullet}\delta\eta_j - \Omega^{-1}_{j-1;\alpha\bullet}\delta\eta_{j-1}] \\ &+ \delta t^{-1}[f'_{j\bullet,\alpha}\Omega_j^{-1}\delta\eta_j - (\Omega^{-1}_{j\alpha\bullet}f_j - \Omega^{-1}_{j-1;\alpha\bullet}f_{j-1})] \\ &- f'_{j\bullet,\alpha}\Omega_j^{-1}f_j \\ &+ \tfrac{1}{2}\delta t^{-2}\mathrm{tr}\{\Omega^{-1}_{j,\alpha}[\Omega_j\delta t - (\delta\eta_j - f_j\delta t)(\delta\eta_j - f_j\delta t)']\} \\ &+ \delta_{\eta_{j\alpha}}\log p(\eta_0).\end{aligned}$$

Here, $h_{i\bullet,\alpha}$ is column α of Jacobian $h_\eta(\eta_{j_i})$, $\Omega^{-1}_{j\alpha\bullet}$ is row α of $\Omega(\eta_j)^{-1}$, $\Omega^{-1}_{j,\alpha} := \Omega^{-1}_{j\bullet\bullet,\alpha}$ and $f_{j\bullet,\alpha}$ denotes column α of Jacobian $f_\eta(\eta_j)$.

4 Maximum likelihood estimation

4.1 Monte Carlo approach

In section (3.1), conditional path sampling was motivated by computing the density function (9) of the data z efficiently. Considering $L(\psi) = p(z;\psi)$ as a function of the parameter vector ψ, the maximum likelihood estimator is obtained as $\hat{\psi} = \arg\max_\psi L(\psi)$ (cf. Basawa and Prakasa Rao; 1980). In equation (10), namely

$$p(z) \;=\; \int \frac{p(z|\eta)p(\eta)}{p(\eta|z)}p(\eta|z)d\eta := E[g(\eta,z)|z], \tag{30}$$

the expectation over the invariant (stationary) distribution may be estimated as fictious-time average

$$\hat{p}(z;U) \;=\; U^{-1}\int_0^U g(\eta(u),z)du \approx N^{-1}\sum_{l=0}^{N-1} g(\eta_l,z) := \hat{p}(z;\delta u,N) \tag{31}$$

if $\eta(u)$ is ergodic. The sum results as an approximation of the integral with grid size $\delta u, U = N\delta u, u_l = l\delta u, \eta_l = \eta(u_l)$. In (31), the importance (smoothing) density $p(\eta|z)$ is unknown, whereas conditional trajectories $\eta_l \sim p(\eta|z)$ can be drawn from the Langevin equation (14) in equilibrium. One could also simulate independent replications $\eta_l(u), l = 0, ..., N-1$, evaluated at large u.

The Langevin equation for the simulation of $\eta(u)$ in the fictious time dimension u is a nonlinear Itô equation with $(J+1)p$ components. It can be solved approximately using several numerical schemes, e.g. the Heun method or a 4th order Runge-Kutta scheme (Kloeden and Platen; 1999; Rümelin; 1982; Gard; 1988). Here we use an Ozaki scheme (Ozaki; 1985) with Metropolis mechanism in order to ensure a correct stationary distribution. This scheme is exact for linear drift functions.

In order to improve the relaxation of the Langevin equation to equilibrium, it can be scaled with a positive definite kernel matrix K. The transformed equation

$$d\eta(u) \;=\; -K\delta_\eta\Phi(\eta|z)du + \sqrt{2}K^{1/2}\,dW(u)/\sqrt{\delta t}, \tag{32}$$

$K = K^{1/2}(K^{1/2})'$ yields the same stationary distribution, as may be seen from the stationary Fokker-Planck equation (12). The kernel may be even time and state dependent, i.e. $K = K(\eta, u)$ (cf. Parisi and Wu; 1981; Okano et al.; 1993), but then additional terms and convolution integrals are required. In the constant case one has (setting $\partial_{j\alpha} := \partial_{\eta_{j\alpha}}$ and summing over double indices)

$$\begin{aligned} 0 \;&=\; \partial_{j\alpha}K_{j\alpha;l\beta}(\partial_{l\beta}\Phi)p + \partial_{j\alpha}\partial_{l\beta}K^{1/2}_{j\alpha;u\gamma}K^{1/2}_{u\gamma;l\beta}p \\ &=\; \partial_{j\alpha}\Big\{K_{j\alpha;l\beta}(\partial_{l\beta}\Phi)p + K_{j\alpha;l\beta}\partial_{l\beta}p\Big\} = \mathrm{div}J \end{aligned}$$

Therefore, setting the probability current $J = 0$ one obtains the equilibrium distribution $p_{stat}(\eta|z) = e^{-\Phi(\eta|z)}$, independent of K (see, e.g. Risken; 1989, chs. 5.2, 6.0). If the Langevin equation (14) were linear with drift $H\eta\delta t$ (i.e. $\Phi = \frac{1}{2}\eta'(-H)\eta\delta t^2$ with symmetric negative definite H), one can use $K = -H^{-1}$ as kernel to obtain an equation with decoupled drift components $d\eta = -(\eta\delta t)du + \sqrt{2}K^{1/2}\,dW(u)/\sqrt{\delta t}$. For example, the leading term (18) is linear. If the linear drift is differentiated w.r.t η, one obtains the matrix $H = \delta_\eta(H\eta\delta t)$. In the nonlinear case, the idea is to compute the Hessian matrix $H_{j\alpha;j'\alpha'} = -\delta_{\eta_{j\alpha}}\delta_{\eta_{j'\alpha'}}\Phi(\eta|z)$ at a certain point $\eta(u_0)$ and use $K = -H^{-1}$ as kernel.

4.1.1 Determination of the importance density

For computing the likelihood function (30), the importance density $p(\eta|z)$ must be determined. Several approaches have been used:

1. Approximation of the optimal $p(\eta|z)$ by a density estimator $\hat{p}(\eta|z)$, using kernel density or regression methods.

   One can use a density estimate

$$\hat{p}(\eta|z) \quad = \quad N^{-1} \sum_{l=1}^{N} \kappa(\eta - \eta_l; H), \tag{33}$$

   where κ is a kernel function and H is a suitably chosen smoothing parameter. In this article a multivariate gaussian kernel $\kappa(y, H) = \phi(y; 0, h^2 S)$ is used, where $e = 1/(p+4), A = (4/(p+2))^e; h = An^{-e}$ and S is the sample covariance matrix (Silverman; 1986, p. 78 ff.).

   The density estimate (33) seems to be natural, but the dimension of η is very high, namely $(J+1)p, J = (t_T - t_0)/\delta t$. It turns out that the estimation quality is not sufficient. Therefore the Markov structure of the state space model must be exploited first. We use the Euler discretized state space model (3)

$$\begin{aligned} \eta_{j+1} &= f(\eta_j)\delta t + g(\eta_j)\delta W_j \\ z_i &= h(y_i) + \epsilon_i, \end{aligned}$$

   (where the dependence on x_j and ψ is dropped for simplicity) in combination with the Bayes formula

$$p(\eta|z) = p(\eta_J|\eta_{J-1}, ..., \eta_0, z)\ p(\eta_{J-1}, ..., \eta_0|z).$$

   Now it can be shown that η_j is a conditional Markov process

$$p(\eta_{j+1}|\eta_j, ..., \eta_0, z) = p(\eta_{j+1}|\eta_j, z). \tag{34}$$

   To see this, we use the conditional independence of the *past* $z^i = (z_0, ..., z_i)$ and *future* $\bar{z}^i = (z_{i+1}, ...z_T)$ given $\eta^j = (\eta_0, ..., \eta_j)$. One obtains

$$\begin{aligned} p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i) &= p(\eta_{j+1}|\eta^j, \bar{z}^i) = p(\eta_{j+1}|\eta_j, \bar{z}^i) \\ p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i) &= p(\eta_{j+1}|\eta_j, \bar{z}^i) \\ j_i \leq \quad j \quad &< j_{i+1} \end{aligned}$$

   since

   (i) the transition density $p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i)$ is independent of past measurements, given the past true states, and only the last state η_j must be considered (Markov process).

   (ii) the transition density $p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i)$ is independent of past measurements.

Thus we have proved $p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i) = p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i)$.

With the representation (34) it is sufficient to estimate the density function $p(\eta_{j+1}|\eta_j, z)$ with a low dimensional argument η_j instead of the full $\eta = (\eta_0, ..., \eta_J)$. The estimation can be accomplished by using either

(a) density estimation methods as in (33), or by

(b) regression methods.

In the latter case, the Euler density is modified to the form

$$p(\eta_{j+1}, \delta t|\eta_j, z) \approx \phi(\eta_{j+1}; \eta_j + (f_j + \delta f_j)\delta t, (\Omega_j + \delta\Omega_j)\delta t) \tag{35}$$

where the correction terms are estimated using the data $\eta_j \equiv \eta_j|z$ from the Langevin sampler. It is assumed that the conditional states fulfil the Euler discretized SDE (cf. 34)

$$\delta\eta_{j+1} \quad = \quad [f(\eta_j) + \delta f_j(\eta_j)]\delta t + [g(\eta_j) + \delta g_j(\eta_j)]\delta W_j \tag{36}$$

with modified drift and diffusion functions. This functions can be estimated by parametric specifications (e.g. $\delta f(x) = \alpha + \beta x + \gamma x^2$) or nonparametrically. The introduction of drift corrections was derived analytically in Singer (2002, 2014).

2. Another approach is the choice of a (in general suboptimal) reference density $p_0(\eta|z) = p_0(z|\eta)p_0(\eta)/p_0(z)$, where the terms on the right hand side are known explicitly. This yields the representation

$$p(z) \quad = \quad p_0(z) \int \frac{p(z|\eta)}{p_0(z|\eta)} \frac{p(\eta)}{p_0(\eta)} \, p_0(\eta|z) d\eta. \tag{37}$$

In this paper we use a conditional gaussian density

$$p_0(\eta|z) = \phi(\eta|z) = \phi(\eta; E[\eta|z], V[\eta|z]) \tag{38}$$

where the conditional moments are estimated from the

(a) Langevin sampler data $\eta_l = \eta_l|z$. Alternatively, one can use a

(b) Kalman smoother

to obtain approximations of the conditional moments. We use algorithms based on the Rauch-Tung-Striebel smoother with Taylor expansions or using integration (unscented transform or Gauss-Hermite integration; see, e.g. Jazwinski 1970; Särkkä 2013, ch. 9).

In both cases, one must (re)sample the data from $p_0 = \phi(\eta|z)$ to compute the likelihood estimate

$$\begin{aligned} \hat{p}(z) &= N^{-1} \sum_l \frac{p(z|\eta_l)p(\eta_l)}{p_0(\eta_l|z)} \\ \eta_l &\sim p_0. \end{aligned} \tag{39}$$

4.1.2 Score function

The score function $s(z) := \partial_\psi \log p(z)$ (dropping ψ) can be estimated by using a well known formula of Louis (1982). One can write

$$\begin{aligned} s(z) &= \partial_\psi \log \int p(z,\eta)d\eta = p(z)^{-1} \int s(z,\eta)p(z,\eta)d\eta \\ &= E[s(z,\eta)|z], \end{aligned} \tag{40}$$

$s(z,\eta) := \partial_\psi \log p(z,\eta) = \partial_\psi p(z,\eta)/p(z,\eta)$, with the estimate

$$\hat{s}(z) = N^{-1} \sum_l s(z,\eta_l). \tag{41}$$

From this a quasi Newton algorithm $\psi_{k+1} = \psi_k + F_k^{-1}\hat{s}_k(z)$ can be implemented. For example, one can use BFGS secant updates for F_k.

4.1.3 Bayesian estimation

A full Bayesian solution to the estimation problem can be obtained by the decomposition of the posterior

$$p(\eta,\psi|z) = p(z,\eta,\psi)/p(z) = p(z|\eta,\psi)p(\eta|\psi)p(\psi)/p(z). \tag{42}$$

From $\log p(\eta,\psi|z)$ one obtains a system of Langevin equations of the form (see eqn. 14)

$$d\eta(u) = \delta_\eta \log p(\eta,\psi|z)du + \sqrt{2}\, dW_1(u)/\sqrt{\delta t} \tag{43}$$

$$d\psi(u) = \partial_\psi \log p(\eta,\psi|z)du + \sqrt{2}\, dW_2(u), \tag{44}$$

where $W_1 : (J+1)p \times 1$ and $W_2 : u \times 1$ are independent Wiener processes. For large "time" u one obtains correlated random samples from $p(\eta,\psi|z)$. The drift term $\partial_\psi \log p(\eta,\psi|z)$ coincides with the score $s(z,\eta)$ except for the prior term $\partial\psi \log p(\psi)$, since $p(\eta,\psi|z) = p(z,\eta|\psi)p(\psi)/p(z)$.

4.2 Numerical integration approach

The Monte Carlo approach will be compared with a method based on Riemann integration combined with transition kernels, which are computed in three different ways, namely by using i) the Euler transition kernel and the matrix exponential of the Fokker-Planck operator, which is represented ii) as a matrix of finite differences or iii) as an integral operator.

4.2.1 Transition kernel approach

The $(J+1)p$-dimensional integral in the likelihood $p(z) = \int p(z|\eta)p(\eta)d\eta$ can be computed without Monte Carlo integration, at least for small dimensions p of η_j. One can write

$$p(z) = \int \prod_{i=0}^{T-1} \left[p(z_{i+1}|\eta_{j_{i+1}}) \prod_{j=j_i}^{j_{i+1}-1} p(\eta_{j+1}|\eta_j) \right] p(z_0|\eta_0)p(\eta_0)d\eta,$$

by using the conditional independence of $Z_i|\eta$ and the Markov property of η (see equations 3–4). The likelihood expression can be represented recursively by the Kalman updates (time update, measurement update, conditional likelihood)

$$\begin{aligned} p(\eta_{j_{i+1}}|Z^i) &= \left[\prod_{j=j_i}^{j_{i+1}-1} \int d\eta_j \; p(\eta_{j+1}|\eta_j)d\eta_j \right] p(\eta_{j_i}|Z^i) \\ p(\eta_{j_{i+1}}|Z^{i+1}) &= p(z_{i+1}|\eta_{j_{i+1}})p(\eta_{j_{i+1}}|Z^i)/p(z_{i+1}|Z^i) \\ p(z_{i+1}|Z^i) &= \int p(z_{i+1}|\eta_{j_{i+1}})p(\eta_{j_{i+1}}|Z^i)d\eta_{j_{i+1}} \\ i &= 0, ..., T \end{aligned}$$

with initial condition $p(\eta_0|Z^0) = p(z_0|\eta_0)p(\eta_0)/p(z_0); p(z_0) = \int p(z_0|\eta_0)p(\eta_0)d\eta_0$ and $Z^i := \{z_i, ..., z_0\}$ (data up to time t_i). Thus one has

$$p(z) = \prod_{i=0}^{T-1} p(z_{i+1}|Z^i)p(z_0)$$

by the Bayes formula. Actually, the Kalman representation is more general since it is also valid for densities $p(\eta_{j+1}|\eta_j, Z^i), p(z_{i+1}|\eta_{j_{i+1}}, Z^i)$ depending on lagged measurements (cf. Liptser and Shiryayev; 2001, vol. II, ch. 13).

The p-dimensional integrals will be approximated as Riemann sums (or using Gauss-Legendre integration)

$$\int p(\xi|\eta)p(\eta|\zeta)d\eta \approx \sum_k p(\xi|\eta_k)p(\eta_k|\zeta)\delta\eta \approx \sum_k p(\xi|\eta_k)p(\eta_k|\zeta)w_k$$

on a p-dimensional grid of supporting points η_k, i.e. $k = \{k_1, ..., k_p\}$, $k_\alpha = 0, ..., K_\alpha$ is a multi index, $\eta_k = \{\eta_{k_1}, ..., \eta_{k_p}\}$ is a p-vector inside a p-dimensional interval $[a, b] = [a_1, b_1] \times ... \times [a_p, b_p]$, with coordinates

$$\eta_{k_\alpha} = a_\alpha + k_\alpha \delta\eta_\alpha \tag{45}$$

$\eta_{K_\alpha} = b_\alpha, \alpha = 1, ..., p$, and volume element $\delta\eta = \prod_{\alpha=1}^p \delta\eta_\alpha$.

4.2.2 Euler transition kernel

The Euler transition kernel $p(\xi_l|\eta_k)$ for the difference equation (3) can be viewed as a matrix T_{lk}. It is given by the normal density $\phi(\xi; \eta + f(\eta)\delta t, \Omega(\eta)\delta t)$ on the grid points ξ_l, η_k. The dependence on the time index j, the exogenous variables $x_j = x(\tau_j)$ and the lagged data Z^i is omitted here.

A better approximation of the true transition kernel $p(y_{j+1}, \tau_{j+1}|y_j, \tau_j)$ of the original SDE (1) can be achieved through an expansion of the drift function using Itô's lemma (Shoji and Ozaki; 1997, 1998; Singer; 2002)

$$f(y) \approx f(y_j) + f_y(y_j)(y - y_j) + \tfrac{1}{2} f_{yy}(y_j)\Omega(y_j)(t - t_j). \tag{46}$$

The approach is known under the label local linearization (LL). Inserting (46) into the moment equations one obtains the linear differential equations $\dot{\mu} = E[f] \approx f_j + A_j(\mu - y_j) + c_j(t - t_j)$, $\dot{\Sigma} = A_j\Sigma + \Sigma A_j' + \Omega_j$ with the Jacobian terms $A_j := f_y(y_j), c_j := \frac{1}{2} f_{yy}(y_j)\Omega(y_j)$ (for details, see Singer; 2002). The second order term c_j only leads to contributions of order δt^2. Thus an improved transition kernel is given by

$$\begin{aligned} p(\eta_{j+1}|\eta_j) &= \phi(\eta_j; \mu_{j+1}, \Sigma_{j+1}) \\ \mu_{j+1} &= \eta_j + A_{1j} f_j + A_j^{-1}[-I\delta t + A_{1j}]c_j \\ \text{row } \Sigma_{j+1} &= (A_j \otimes I + I \otimes A_j)^{-1}(A_j^* \otimes A_j^* - I)\text{row } \Omega_j, \end{aligned}$$

$A_j^* := \exp(A_j\delta t)$, $A_{1j} := A_j^{-1}(A_j^* - I)$, where row is the row-wise vector operator. For small δt, one recovers the Euler terms $\mu_{j+1} = \eta_j + f_j\delta t; \Sigma_{j+1} = \Omega_j\delta t$. The approaches are denoted as ETK and LLTK in the sequel.

4.2.3 Fokker-Planck equation

The transition kernel can also be obtained as a short time approximation to the solution of the Fokker-Planck equation

$$\begin{aligned} \partial_t p(y, t|y_j, \tau_j) &= -\partial_\alpha[f_\alpha p(y, t|y_j, \tau_j)] + \tfrac{1}{2}\partial_\alpha\partial_\beta[\Omega_{\alpha\beta} p(y, t|y_j)] \\ &:= F(y)p(y, t|y_j, \tau_j) \end{aligned}$$

with initial condition $p(y, \tau_j|y_j, \tau_j) = \delta(y - y_j)$. A summation convention over doubly occuring indices α, β is supposed. Formally, one has $p(y_{j+1}, \tau_{j+1}|\ y_j, \tau_j) = \exp[F(y_{j+1})\delta t]\ \delta(y_{j+1} - y_j)$, if the system is autonomous, i.e. $x(t) = x$. In this case, the measurement time interval $\Delta t_i := t_{i+1} - t_i$ may be bridged in one step $p(y_{i+1}, t_{i+1}|y_i, t_i) = \exp[F(y_{j+1})\Delta t_i]\ \delta(y_{i+1} - y_i)$. For equally spaced data, only one kernel must be computed. In the nonautonomous case, a time ordered exponential must be considered. In this paper, it is assumed that $x(t) = x_j$ is constant in the interval $[\tau_j, \tau_{j+1})$, so $p(y_{j+1}, \tau_{j+1}|y_j, \tau_j) = \exp[F(y_{j+1}, x_j)\delta t]\ \delta(y_{j+1} - y_j)$. In this section, two approximations are considered:

1. First, the spatial (y) derivatives are replaced by finite differences on a multi-dimensional grid y_k.

2. Second, the differential operator F is rewritten as an integral operator and the integrals are approximated as sums.

The time dimension t is kept continuous. Such mixed approaches are called lattice approximation, semi-discretization or method of lines (cf. Schiesser; 1991; Jetschke; 1991; Yoo; 2000). In both cases, one obtains a linear matrix differential equation which can be solved by the matrix exponential function.

Finite differences. Using finite differences, one may write

$$-\partial_\alpha[f_\alpha p(y)] \approx -\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\delta_{k_p l_p}[f_\alpha(y_l)p(y_l)]$$

with multi index $l = \{l_1, ..., l_p\}$, $y_l = \{y_{l_1}, ..., y_{l_p}\} = a + l\delta y$ (see 45) and two sided differences $\nabla_{k_\alpha l_\alpha} = (\delta_{k_\alpha+1;l_\alpha} - \delta_{k_\alpha-1;l})/(2\delta y_\alpha)$. The diffusion term is

$$\tfrac{1}{2}\partial_\alpha\partial_\beta[\Omega_{\alpha\beta}p(y)] \approx \tfrac{1}{2}\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\nabla_{k_\beta l_\beta}...\delta_{k_p l_p}[\Omega_{\alpha\beta}(y_l)p(y_l)]$$

with the replacement (diagonal terms $\alpha = \beta$): $\nabla^2_{k_\alpha l_\alpha} \to \Delta_{k_\alpha l_\alpha} = (\delta_{k_\alpha+1;l_\alpha} - 2\delta_{k_\alpha l_\alpha} + \delta_{k_\alpha-1;l_\alpha})/(\delta y^2_\alpha)$ (centered second difference). Thus the Fokker-Planck operator is replaced by the matrix

$$F_{kl} = -\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\delta_{k_p l_p} f_\alpha(y_l) + \tfrac{1}{2}\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\nabla_{k_\beta l_\beta}...\delta_{k_p l_p}\Omega_{\alpha\beta}(y_l). \quad (47)$$

Usually the multi indices are flattened to a $K = \prod_\alpha K_\alpha$-dimensional index. Clearly, one obtains a high dimensional matrix problem. The transition kernel on the grid points is written as matrix (possibly depending on x_j)

$$p(\eta_{j+1;k}|\eta_{jl}) = [\exp(F(x_j)\delta t)]_{kl}/\delta\eta$$

where the matrix exponential function may be evaluated using several methods, including Taylor series and eigen methods (Moler and VanLoan; 2003)

Integral operator. Alternatively, the differential operator $F(y)$ can be represented as integral operator

$$F(y)p(y) = \int F(y, y')p(y')dy'$$

with integral kernel $F(y, y') = F(y)\delta(y - y') = L(y')\delta(y - y') = L(y', y)$ (Risken; 1989, p. 69). Here $L(y) = F^*(y)$ is the backward operator and $\delta(y - y')$ is the Dirac delta function (cf. Lighthill; 1958). The differential operation $q(y) = F(y)p(y) = \int F(y, y')p(y')dy$ is thus transformed to an integral operation. It may be approximated on the p-dimensional grid y_k as matrix-vector product

$$\begin{aligned} q(y_k) &= \sum_l F(y_k, y_l)p(y_l)\delta y \\ F(y_k, y_l) &= F(y_k)\delta(y_k - y_l) := F_{kl}/\delta y. \end{aligned}$$

Explicitly, one obtains the distribution (cf. 47)

$$\begin{aligned} F(y,y') &= -\delta(y_1-y_1')...\partial_{y_\alpha}\delta(y_\alpha-y_\alpha')...\delta(y_p-y_p')f_\alpha(y) \\ &+ \tfrac{1}{2}\delta(y_1-y_1')...\partial_{y_\alpha}\delta(y_\alpha-y_\alpha')...\partial_{y_\beta}\delta(y_\beta-y_\beta')...\delta(y_p-y_p')\Omega_{\alpha\beta}(y). \end{aligned}$$

The delta function is interpreted as a function sequence $\delta_n(y)$ with the property $\lim_{n\to\infty}\int\delta_n(y-y')\phi(y')dy' = \phi(y)$ for any test function ϕ (Lighthill; 1958). In numerical applications, one must use a certain term of this delta sequence with appropriate smoothness properties, leading to a free parameter n of the numerical procedure. For example, one can use the series $\delta_n(y-y') = \sum_{m=0}^{n}\chi_m(y)\chi_m(y')$ for a complete orthonormal system χ_m or the truncated Fourier transform $\delta_n(y-y') = \int_{-n}^{n}\exp(2\pi i(y-y')k)dk$.

If one writes $F(y)p(y) = F(y)\int\delta(y-y')p(y')dy$, the term $p(y,\delta_n) = \int\delta_n(y-y')p(y')dy$ may be interpreted as a functional. In this guise the procedure was called DAF (distributed approximating functional; cf. Wei et al.; 1997) using Hermite functions (oscillator eigenfunctions) χ_m.

If the delta functions on the grid points $y_{k_\alpha}, y_{l_\alpha}$ are replaced by $\delta(y_{k_\alpha}-y_{l_\alpha}) \to \delta_{k_\alpha l_\alpha}/\delta y_\alpha$, $\delta'(y_{k_\alpha}-y_{l_\alpha}) \to (\delta_{k_\alpha+1,l_\alpha}-\delta_{k_\alpha-1,l_\alpha})/(2\delta y_\alpha^2)$, $\delta''(y_{k_\alpha}-y_{l_\alpha}) \to (\delta_{k_\alpha+1,l_\alpha} - 2\delta_{k_\alpha l_\alpha}+\delta_{k_\alpha-1,l_\alpha})/\delta y_\alpha^3$ one recovers the finite difference approximation (47). This choice corresponds to the delta sequence $\delta_{\delta y}(y) = \chi_A(y)/\delta y, A = [-\delta y/2, \delta y/2]$ where $\chi_A(x)$ is the indicator function of the set A. In this case the free parameter $n = \delta y$ is naturally given by the spatial discretization interval δy. The Euler transition kernel does not require the choice of free parameters (it is naturally given by the time discretization δt). The spatial discretization should be of order $\sqrt{\Omega\delta t}$.

5 Applications

5.1 Geometrical Brownian Motion (GBM)

The SDE

$$dy(t) = \mu y(t)dt + \sigma y(t)\,dW(t) \tag{48}$$

is a popular model for stock prices, used by Black and Scholes (1973) for modeling option prices. It contains a multiplicative noise term $y\,dW$ and is thus bilinear. The form $dy(t)/y(t) = \mu dt + \sigma\,dW(t)/dt$ shows, that the simple returns are given by a constant value μdt plus white noise. For the log returns, we set $x = \log y$, and use Itô's lemma to obtain $dx = dy/y + 1/2(-y^{-2})dy^2 = (\mu-\sigma^2/2)dt + \sigma dW$. This shows, that the log returns contain the Wong-Zakai correction, i.e. $\tilde{\mu} = \mu - \sigma^2/2$. From this, we obtain the exact solution

$$y(t) = y(t_0)e^{(\mu-\sigma^2/2)(t-t_0)+\sigma[W(t)-W(t_0)]}, \tag{49}$$

which is a *multiplicative* exact discrete model with log-normal distribution. The exact transition density is thus

$$p(y,t|y_0,t_0) = y^{-1}\phi\left(\log(y/y_0);(\mu-\sigma^2/2)(t-t_0),\sigma^2(t-t_0)\right) \tag{50}$$

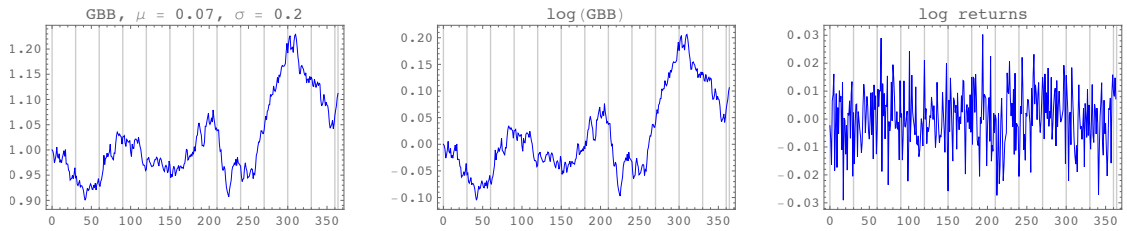

Figure 1: Geometrical Brownian motion GBM: Trajectory and log returns. Vertical lines: measurement times.

The model was simulated using $\mu = 0.07, \sigma = 0.2$ and $\delta t = 1/365$. Only monthly data were used (fig. 1).

The Langevin sampler output is displayed in fig. 2. Due to a convergence kernel (inverse Hessian of the potential Φ), all trajectories relax to equilibrium at about the same rate and the autocorrelation is small (1st row, 2nd columnn). We sampled $M = 2000$ trajectories, dropping 30% at the beginning to avoid nonstationarity and used only every second one (thinning). This reduces autocorrelation in the sample. Thus the effective sample size is $M' = 700$.

We obtain a smooth likelihood surface with small approximation error (figs. 5, 6, 8, 9, 10). Clearly, the usage of the full kernel density (33) yields bad results (fig. 4). In contrast, the conditional Markov representation (34) works well (fig. 3). One can also use a conditional gaussian density in (34) or a linear GLS estimation of the drift correction δf_j and diffusion correction $\delta\Omega_j$ (see eqn. 35). If the diffusion matrix is not corrected, biased estimates occur (fig. 7).

Importance sampling using an extended Kalman smoother (extended Kalman sampling EKS, see (38)) yields very good results (fig. 9). Finally, the transition kernel filter (TKF) with an Euler transition kernel is displayed, where the integration range is $\{y_0, y_1, dy\} = \{0.7, 1.5, .0025\}$. This leads to a 321×321 transition matrix with 103 041 elements (see section 4.2.2). However, entries smaller than 10^{-8} where set to 0, and only 15972 values were stored in a sparse array.

In this one dimensional example, the TKF is most efficient, while the EKS takes 4 times more CPU time. The methods based on the conditional Markov property take much more CPU time, especially when using a kernel density estimate for $p(\eta_{j+1}|\eta_j, z)$ (see 34). In a higher dimensional state space $\mathbb{R}^p$, however, the grid based TKF soon gets infeasible, since we have K^p supporting points $\eta_k, k = (k_1, ..., k_p)$ in $\mathbb{R}^p$. The transition kernel matrix is even of dimension $K^p \times K^p$.

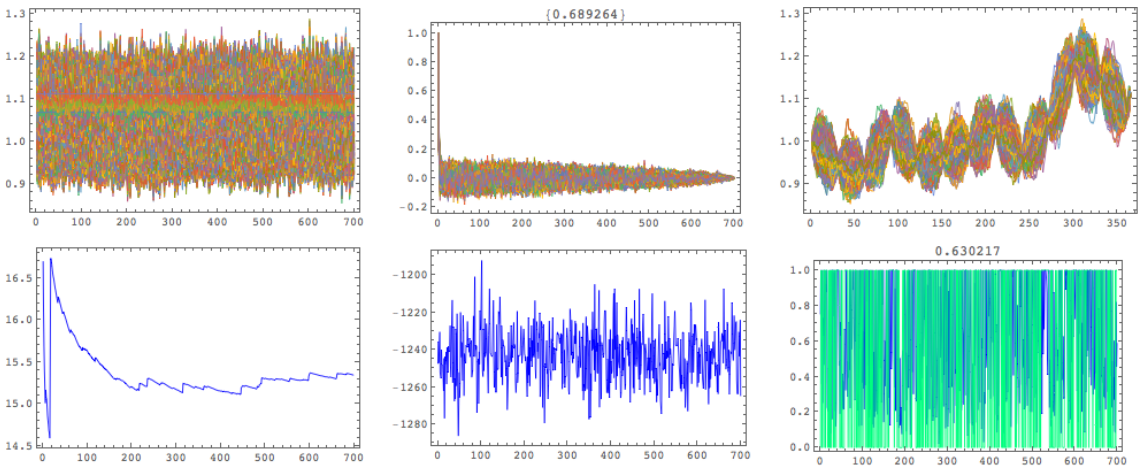

Figure 2: Langevin sampler. Importance density $\hat{p}_2 = \prod_j \phi(\eta_{j+1}, \eta_j|z)/\phi(\eta_j|z)$. From top, left: (1,1): trajectory η_{jl} over l (replications), (1,2): autocorrelation of η_{jl}, (1,3): trajectories η_{jl} over j, (2,1): convergence of estimator $\hat{p}(z)$, (2,2): potential $\log p(\eta_{jl})$ over l, (2,3): average acceptance probability and rejection indicator for Metropolis algorithm.

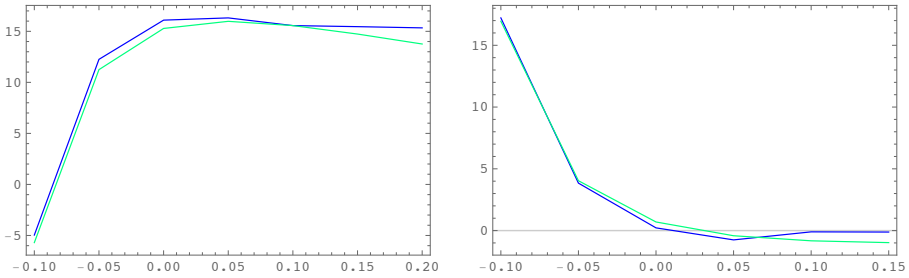

Figure 3: Likelihood (left) and score (right) as a function of $\sigma - 0.2$, $\hat{p}_2 =$ conditional kernel density. Green lines: exact log-likelihood.

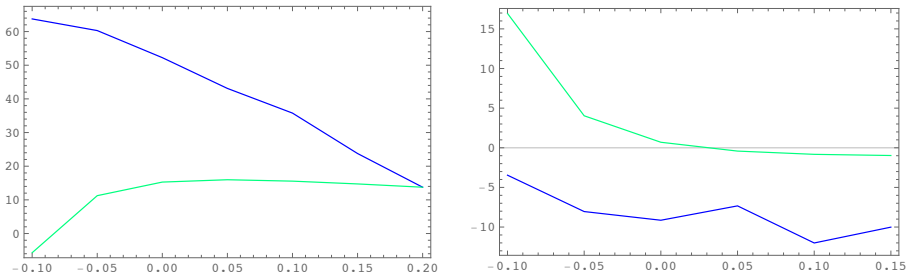

Figure 4: Likelihood and score, $\hat{p}_2 =$ full kernel density.

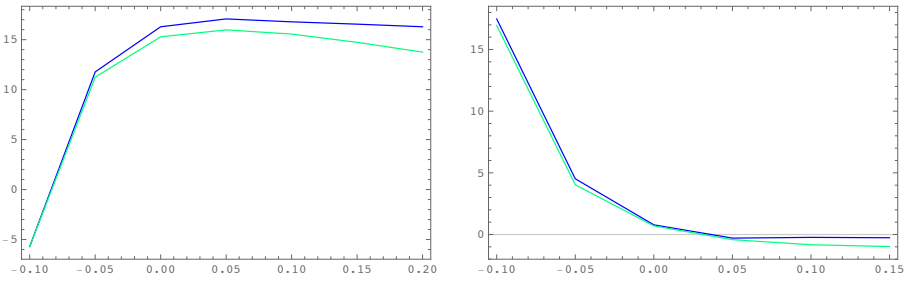

Figure 5: Likelihood and score, $\hat{p}_2 =$ conditionally Gaussian transition density.

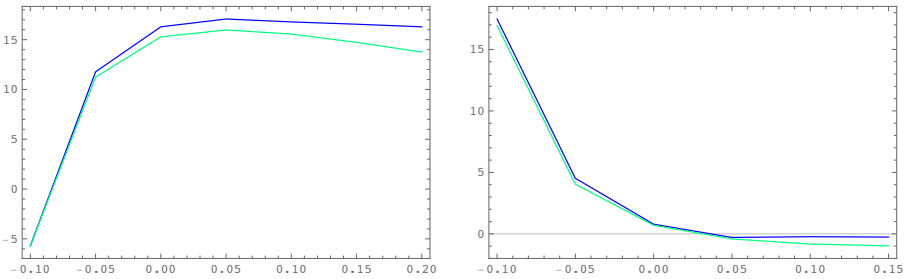

Figure 6: Likelihood and score, $\hat{p}_2 =$ linear GLS estimation of drift and diffusion corrections $\delta f_j, \delta\Omega_j$ (eqn. 35).

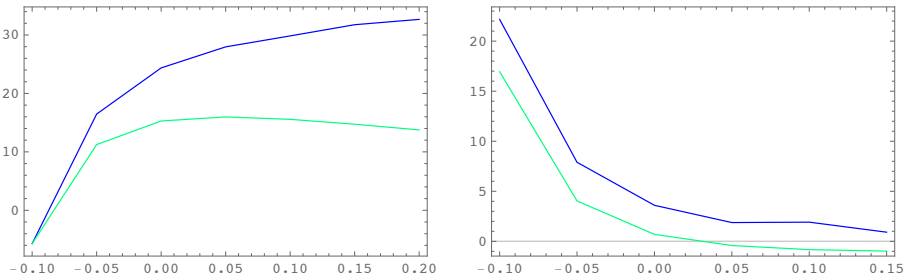

Figure 7: Likelihood and score, $\hat{p}_2 =$ linear GLS, constant diffusion matrix.

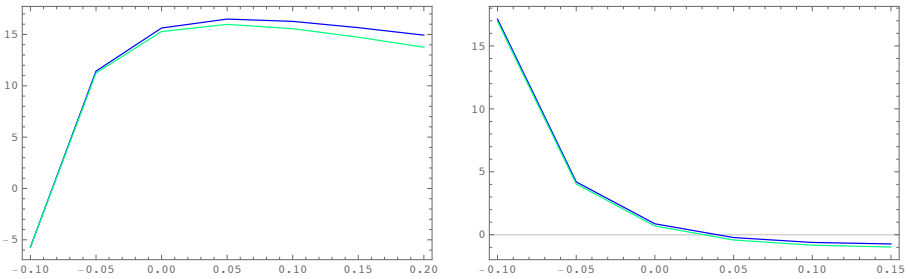

Figure 8: Likelihood and score, Langevin/Gauss-Resampling.

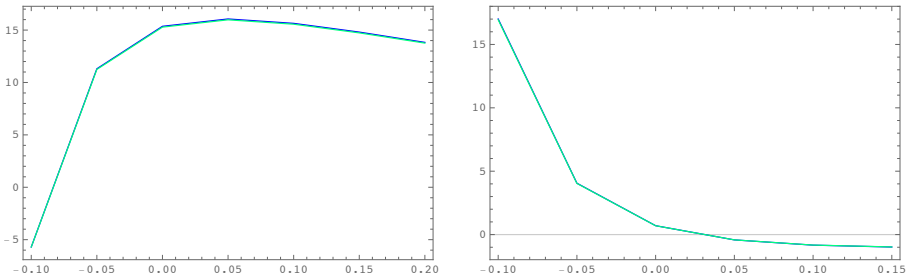

Figure 9: Likelihood and score, extended Kalman sampling.

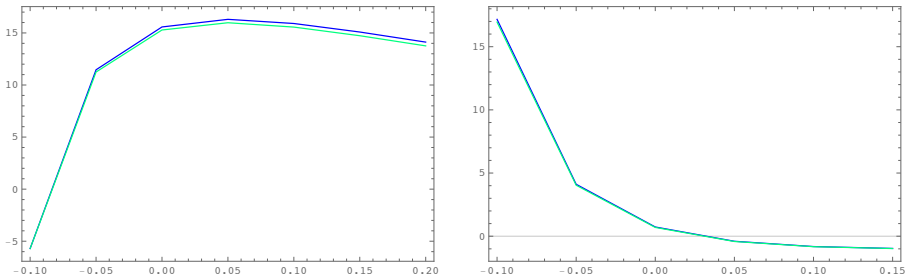

Figure 10: Likelihood and score, transition kernel filter with Euler transition kernel.

5.2 Ginzburg-Landau model

The Ginzburg-Landau model

$$dY = -[\alpha Y + \beta Y^3]dt + \sigma dW(t)$$

with measurement equation

$$Z_i = Y_i + \epsilon_i,$$

at times $t_i \in \{0, 2, 4, 6, 8, 10\}$ is a well known nonlinear benchmark model, since the stationary distribution

$$\begin{aligned} p_0(y) &\propto \exp[-(2/\sigma^2)\Phi(y)], \\ \Phi(y) &= \tfrac{1}{2}\alpha y^2 + \tfrac{1}{4}\beta y^4 \end{aligned}$$

can exhibit a bimodal shape (Singer; 2002, 2011; Särkkä et al.; 2013). The true parameters were set to the values $\{\alpha_0, \beta_0, \sigma_0\} = \{-1, 0.1, 2\}$, $R = \mathrm{Var}(\epsilon_i) = 0.01$ and the trajectories were simulated with a discretization interval $\delta t = 0.1$ using an Euler-Maruyama approximation (cf. Kloeden and Platen; 1999)

Fig. 11–12 shows the simulated likelihood surface as a function of parameter $\sigma - \sigma_0$ using a Gaussian distribution as approximate importance density. The mean $E(\eta_j|Z)$ and covariance matrix $\mathrm{Cov}(\eta_j|Z)$ were computed either using an extended or an unscented Kalman smoother. For the conditional covariance matrix

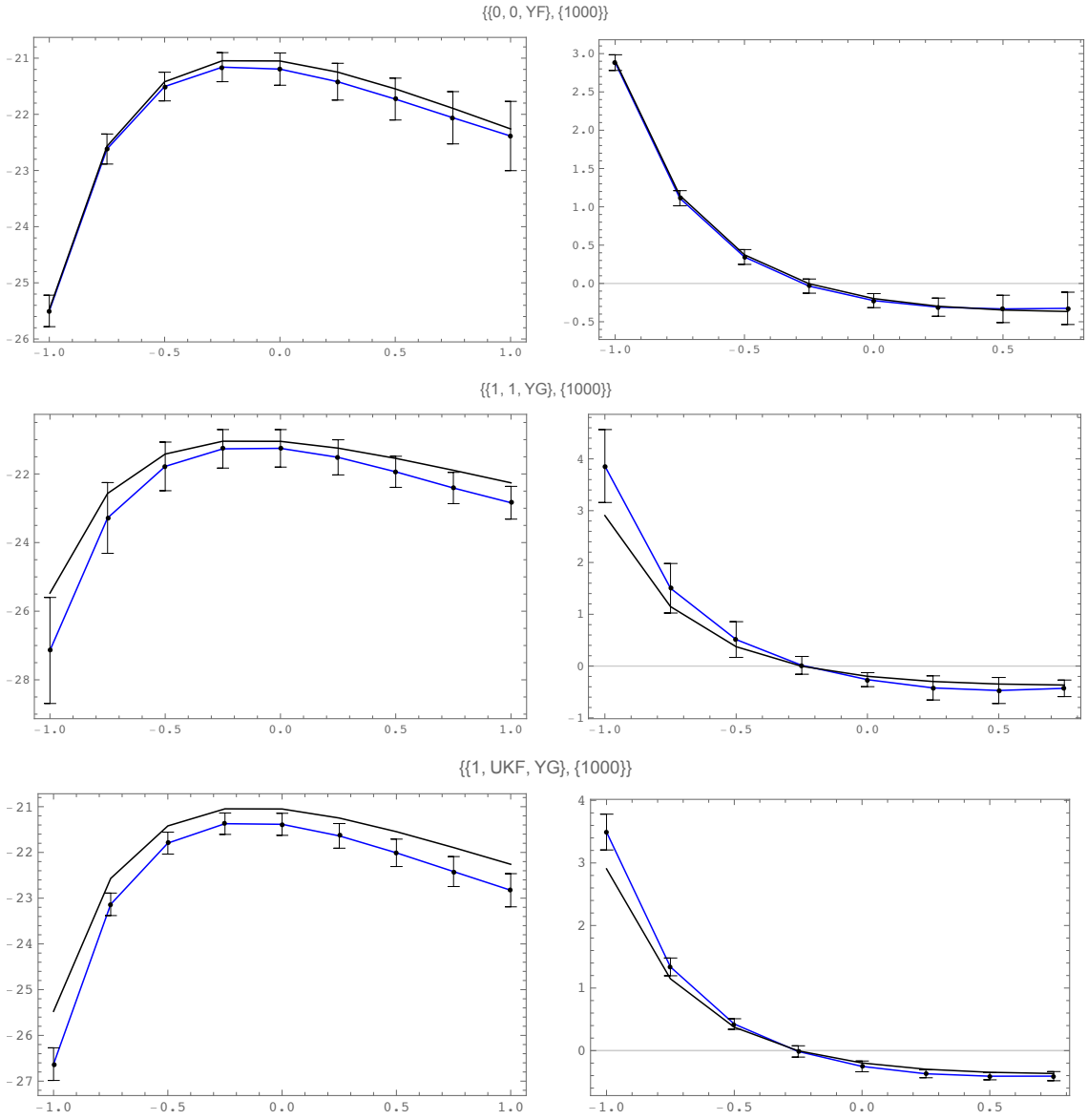

Figure 11: Ginzburg-Landau model. Extended Kalman sampling, $M = 1000$. Likelihood (left) and score (right) as function of $\sigma - \sigma_0$. Reference trajectory = smoother solution. First row: Brownian bridge. 2nd row: Extended Kalman smoother. 3rd row: Unscented Kalman smoother. Black line: transition kernel filter with Euler transition kernel.

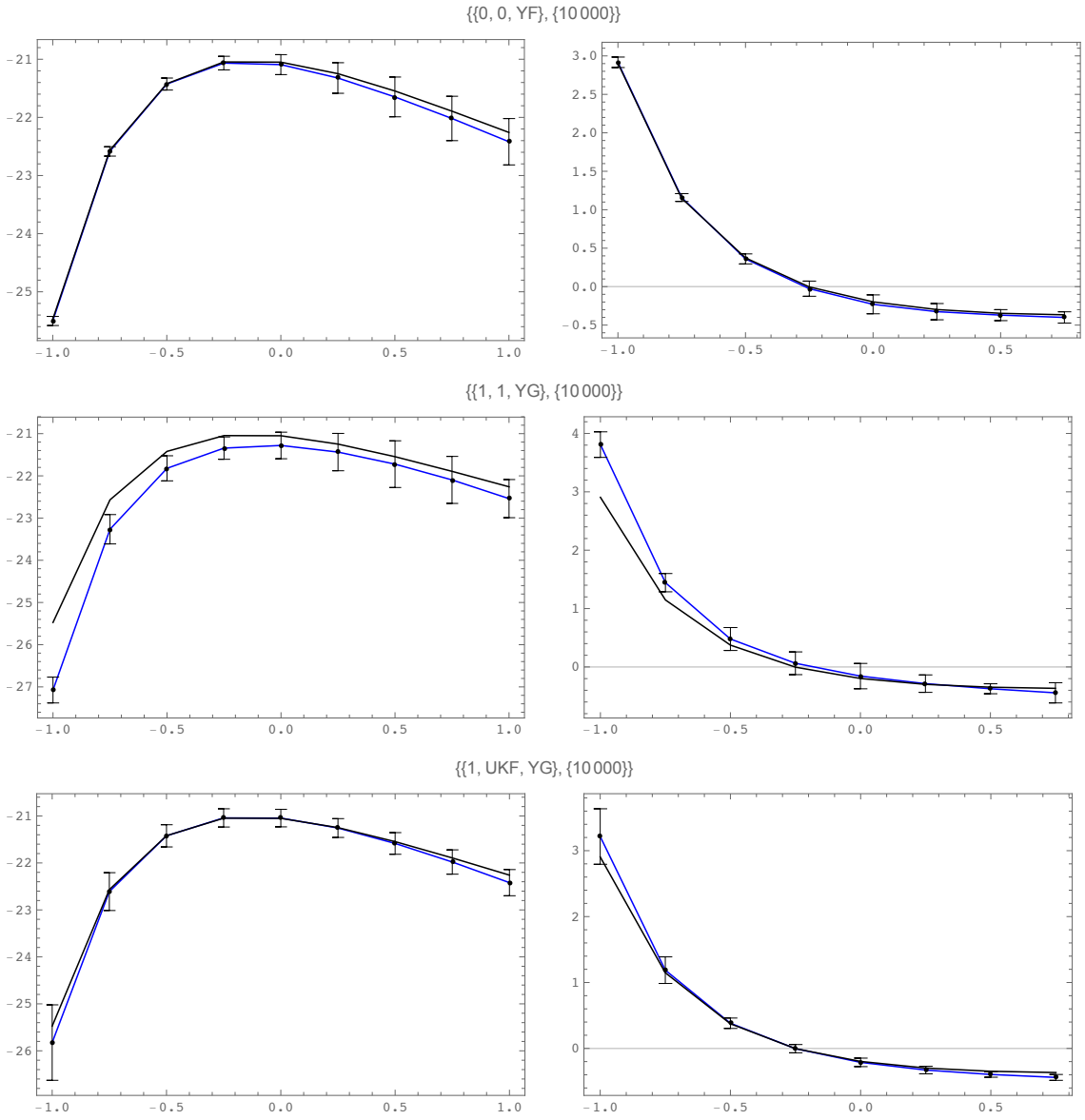

Figure 12: Ginzburg-Landau model. Extended Kalman sampling, $M = 10000$. Likelihood (left) and score (right) as function of $\sigma - \sigma_0$. Reference trajectory = smoother solution. First row: Brownian bridge. 2nd row: Extended Kalman smoother. 3rd row: Unscented Kalman smoother. Black line: transition kernel filter with Euler transition kernel.

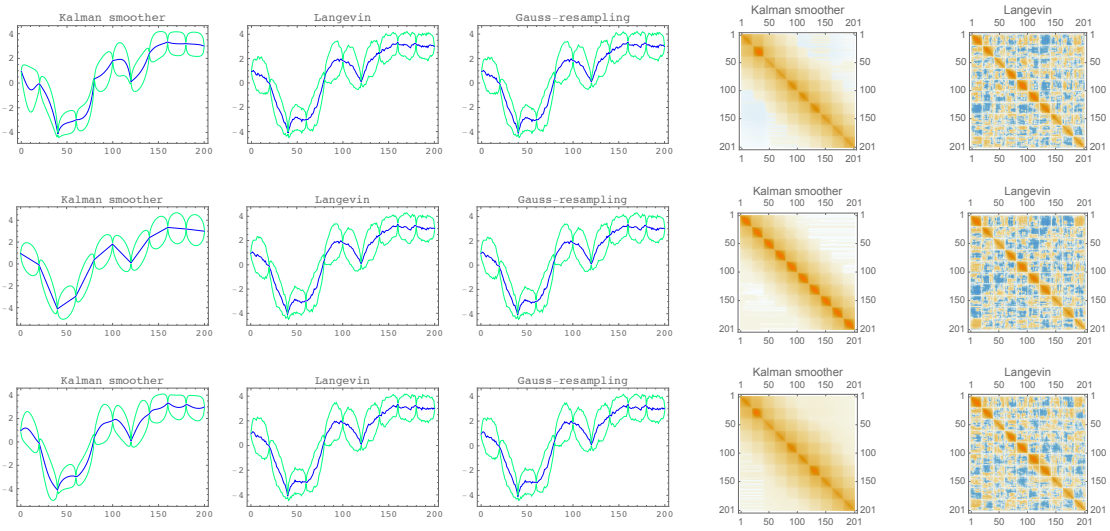

Figure 13: Ginzburg-Landau model. First line, from left: Conditional mean $\pm$ standard deviation: extended Kalman smoother, Langevin sampler, Gauss resampling, conditional covariance matrix: extended Kalman smoother, Langevin sampler. 2nd line, Brownian bridge. 3rd line, unscented Kalman smoother (see text).

$\mathrm{Cov}(\eta_j, \eta_{j'}|Z) : (J+1)p \times (J+1)p$, two variants of the linearization can be used: computation of the Jacobians $\partial f(y)/\partial y$ and $\partial h(y)/\partial y$ either along the filtered or along the smoothed trajectory. Furthermore, one can set $f \equiv 0$ in the smoother, which corresponds to a Brownian bridge (cf. Durham and Gallant; 2002). The error bars were obtained using 10 likelihood surface replications with different random seeds. Clearly, the Brownian bridge performs best, and the likelihood surface is smooth as a function of the parameter (figs. 11 and 12, first line, right). This is in contrast to methods based on particle filters (cf. Pitt; 2002; Singer; 2003). It should be noted that not only the mean over the likelihood surfaces, but also each single one is a smooth curve, which facilitates the usage in Newton type algorithms or in Langevin samplers for Bayesian estimation.

The superiority of the Brownian bridge sampler can be understood from fig. 13. The Kalman smoother (first line, left) tends to predict values in the potential wells of $\Phi(y)$ (cf. Singer; 2005, 2008), whereas the exact smoothing solution computed from the Langevin sampler (first line, second column) predicts more or less straight lines between the measurements. This behaviour is exactly produced by the Brownian bridge (second line), since the drift was set to $f = 0$. The unscented Kalman smoother also gives a conditional mean similar to the Langevin sampler (3rd line, left and middle).

Thus, the conditional mean given by the Brownian bridge and the unscented Kalman smoother is (at least in this example) nearer to the exact smoothing solution (Langevin sampler) than the Kalman smoother, producing a better approximate

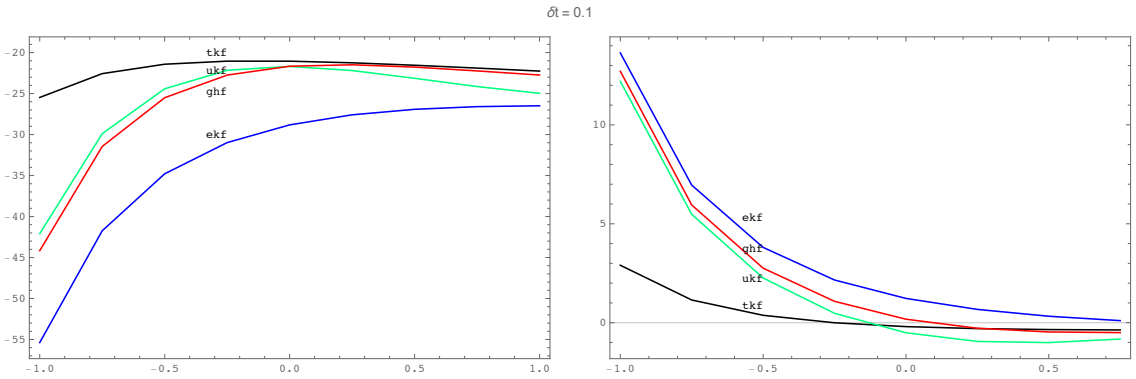

Figure 14: Ginzburg-Landau model. Likelihood of TKF, EKF, UKF and GHF as function of $\sigma - \sigma_0$.

importance density (38). This, in turn, improves the importance sampling results.

Since the extended Kalman sampler is based on an extended Kalman smoother, it is interesting to inspect the likelihood surfaces produced by the EKF and other integration based filters, such as the unscented Kalman filter (UKF) and the Gauss-Hermite filter (GHF). The comparison with the exact TKF/ETK is displayed in fig. 14. It is seen that the integration based filters are superior to the Taylor based EKF. Still, the likelihood surfaces strongly deviate from the exakt TKF solution (black line).

As noted above, the EKF moment equation $\dot{\mu} = E[f(y)] \approx f(\mu)$ gives solutions in the minima of $\Phi(y)$. This does not happen for the integration based equation $\dot{\mu} = E[f(y)] \approx \sum \alpha_l f(\eta_l)$, where μ tends to zero (cf. Singer; 2005, 2008). Therefore, an integration based smoother should give better results, as is the case.

Table 1 serves to compare the ML-estimates given by the several estimation methods with the reference solution TKF/ETK. The likelihood function was maximized using a quasi-Newton algorithm with numerical score and BFGS secant updates. The standard errors were computed using the negative inverse Hessian (observed Fisher information) after convergence. The results of the Kalman sampler with $M = 10,000$ replications are very similar to the reference solution TKF/ETK. A smaller sample size ($M = 1000$) gives good results for the Brownian bridge sampler, the unscented Kalman sampler and the Gauss-Hermite sampler. The extended Kalman smoother leads to bad results in this case. The integration based filters UKF and GHF produce somewhat biased ML estimates whereas the EKF, again, performs worst, especially for parameter α. Clearly, this findings are only preliminary.

Other estimation methods for the importance density, such as kernel density or regression methods, are very time consuming when a reasonable approximation for the likelihood surface should be achieved. These methods are presently under further study.

true	*est*	*std*	*est*	*std*	*est*	*std*	*est*	*std*
	TKF/ETK		EKF		UKF, $\kappa = 0$		GHF, $m = 3$	
-1	-0.5005	0.5407	-0.1039	0.2432	-0.349	0.7421	-0.349	0.7421
0.1	0.0609	0.0568	0.0364	0.0415	0.0406	0.0642	0.0406	0.0642
2	1.6157	0.4739	1.543	0.3801	1.7353	0.7355	1.7353	0.7355
Extended Kalman sampling, $M = 1000$								
	EKS		BBS		UKS, $\kappa = 0$		GHS, $m = 3$	
-1	-1.0676	0.3985	-0.4176	0.4457	-0.4176	0.4457	-0.4762	0.4357
0.1	0.1166	0.0439	0.0518	0.0464	0.0518	0.0464	0.0572	0.0424
2	1.73	0.3122	1.5631	0.4245	1.5631	0.4245	1.5969	0.4239
Extended Kalman sampling, $M = 10,000$								
	EKS		BBS		UKS, $\kappa = 0$		GHS, $m = 3$	
-1	-0.4864	0.4659	-0.4908	0.5061	-0.4501	0.4449	-0.4498	0.4468
0.1	0.0581	0.0437	0.06	0.0533	0.0555	0.0454	0.0554	0.0454
2	1.5882	0.4169	1.6101	0.4606	1.5899	0.4352	1.5888	0.4286

Table 1: ML estimates for several estimation methods. The transition kernel filter with Euler transition kernel (TKF/ETK) (top, left) serves as reference value. EKF: extended Kalman filter, UKF: unscented Kalman filter, GHF: Gauss-Hermite Kalman filter, EKS: extended Kalman sampler, BBS: Brownian bridge sampler ($f = 0$), UKS: unscented Kalman sampler, GHS: Gauss-Hermite sampler

6 Conclusion

We analytically computed the drift function of a Langevin sampler for the continuous-discrete state space model, including a state dependent diffusion function. In the continuum limit, a stochastic partial differential equation is obtained. From this, we can draw random vectors from the conditional distribution of the latent states, given the data. This sample can be used for the estimation of the unknown importance density and in turn to the determination of a variance reduced MC estimator of the likelihood function. Moreover, one obtains a numerical solution of the optimal smoothing problem.

The unknown importance density was estimated from the sampler data using kernel density and regression methods. Alternatively, a gaussian reference density with suboptimal properties, but known analytical form was used. Methods based on transition kernels and the Fokker-Planck equation generally gave good results, but seem to be restricted to low dimensional state spaces. In the geometrical Brownian motion model, well known from finance, all methods gave encouraging results.

However, in a strongly nonlinear system (Ginzburg-Landau model), estimation methods for the importance density based on kernel density and regression methods performed disappointingly, whereas the extended Kalman sampler EKS (using

a gaussian importance density determined by extended Kalman smoothers) gave smooth likelihood surfaces near to the true ones.

In further work, the methods will be tested with simulation studies and extended to higher dimensional models such as the 3 dimensional Lorenz model well known from chaos theory.

More generally, one can compute variance reduced estimates of functionals involving Itô processes such as the Feynman-Kac formula used in finance and quantum theory (Singer; 2016).

Appendix: Continuum limit

The expressions in the main text were obtained by using an Euler discretization of the SDE (1), so in the limit $\delta t \to 0$ one expects a convergence of η_j to the true state $y(\tau_j)$ (see Kloeden and Platen; 1999, ch. 9). Likewise, the $(J+1)p$-dimensional Langevin equation (14) for $\eta_{j\alpha}(u)$ will be an approximation of the stochastic partial differential equation (SPDE) for the random field $Y_\alpha(u,t)$ on the temporal grid $\tau_j = t_0 + j\delta t$.

A rigorous theory (assuming constant diffusion matrices) is presented in the work of Reznikoff and Vanden-Eijnden (2005); Hairer et al. (2005, 2007); Apte et al. (2007); Hairer et al. (2011). In this section it is attempted to gain the terms, obtained in this literature by functional derivatives, directly from the discretization, especially in the case of state dependent diffusions. Clearly, the finite dimensional densities w.r.t. Lebesgue measure loose their meaning in the continuum limit, but the idea is too use large, but finite J, so that the Euler densities $p(\eta_0, ..., \eta_J)$ are good approximations of the unknown finite dimensional densities $p(y_0, \tau_0; ...; y_J, \tau_J)$ of the process $Y(t)$ (cf. Stratonovich 1971, 1989, Bagchi 2001 and the references cited therein).

Constant diffusion matrix

First we consider constant and (nonsingular) diffusion matrices Ω. The Lagrangian (15) attains the formal limit (Onsager-Machlup functional)

$$S = \tfrac{1}{2}\int dy(t)'(\Omega dt)^{-1}dy(t) \tag{51}$$

$$- \int f(y)'\Omega^{-1}dy(t) + \tfrac{1}{2}\int f(y)'\Omega^{-1}f(y)dt. \tag{52}$$

If $y(t)$ is a sample function of the diffusion process $Y(t)$ in (1), the first term (51) does not exist, since the quadratic variation $dy(t)dy(t)' = \Omega dt$ is of order dt. Thus we have $dy(t)'(\Omega dt)^{-1}dy(t) = \mathrm{tr}[(\Omega dt)^{-1}\ dy(t)dy(t)'] = \mathrm{tr}[I_p] = p$. Usually, (51) is written as the formal expression $\frac{1}{2}\int \dot{y}(t)'\ \Omega^{-1}\dot{y}(t)dt$, which contains the (nonexisting) derivatives $\dot{y}(t)$. Moreover, partial integration yields

$$-\tfrac{1}{2}\int y(t)'\Omega^{-1}\ddot{y}(t)dt \tag{53}$$

so that $C^{-1}(t,s) = \Omega^{-1}(-\partial^2/\partial t^2)\delta(t-s)$ is the kernel of the inverse covariance (precision) operator of $Y(t)$ (for drift $f = 0$; i.e. a Wiener process). Indeed, since

$$\partial^2/\partial t^2 \min(t,s) = -\delta(t-s), \tag{54}$$

the covariance operator kernel $C(t,s)$ is

$$C(t,s) \;=\; \Omega(-\partial^2/\partial t^2)^{-1}\delta(t-s) = \Omega\min(t,s).$$

Thus, $p(y) \propto \exp[-\frac{1}{2}\int y(t)'\Omega^{-1}\ddot{y}(t)dt]$ is the formal density of a Gaussian process $Y(t) \sim N(0,C)$.

In contrast, the terms in (52) are well defined and yield the Radon-Nikodym derivative (cf. 17)

$$\alpha(y) \;=\; \exp\Big\{\int f(y)'\Omega^{-1}dy(t) - \tfrac{1}{2}\int f(y)'\Omega^{-1}f(y)dt\Big\}. \tag{55}$$

This expression can be obtained as the ratio of the finite dimensional density functions $p(y_J,\tau_J,...,y_1,\tau_1|y_0,\tau_0)$ for drifts f and $f=0$, respectively, in the limit $\delta t \to 0$ (cf. Wong and Hajek; 1985, ch. 6, p. 215 ff). In this limit, the (unkown) exact densities can be replaced by the Euler densities (5). Now, the terms of the Langevin equation (14) will be given. We start with the measurement term (21), $\alpha = 1,...,p$

$$\delta\log p(z|y)/\delta y_\alpha(t) \;=\; \sum_{i=0}^{T} h'_{i\bullet,\alpha}R_i^{-1}(z_i-h_i)\delta(t-t_i) \tag{56}$$

where the scaled Kronecker delta ($\delta_{jj_i}/\delta t$) was replaced by the delta function (see footnote 2). Clearly, in numerical implementations a certain term of the delta sequence $\delta_n(t)$ must be used (cf. Lighthill; 1958). Next, the term stemming from the driftless part (18) is

$$-\delta S_0/\delta y_\alpha(t) \;=\; \Omega^{-1}_{\alpha\bullet}\ddot{y}(t) = \Omega^{-1}_{\alpha\bullet}y_{tt}(t),$$

or $\Omega^{-1}y_{tt}(t)$ in matrix form, which corresponds to (53). The contributions of S_1 are (cf. 19)

$$-\delta S_1/\delta y_\alpha(t) \;=\; f(y)_{\beta,\alpha}\Omega^{-1}_{\beta\gamma}dy_\gamma(t)/dt - \Omega^{-1}_{\alpha\gamma}df_\gamma(y)/dt.$$

The first term is of Itô form. Transformation to Stratonovich calculus (Apte et al.; 2007, sects. 4, 9) yields

$$h_{\alpha\beta}dy_\beta \;=\; h_{\alpha\beta}\circ dy_\beta - \tfrac{1}{2}h_{\alpha\beta,\gamma}\Omega_{\beta\gamma}dt \tag{57}$$

$$df_\alpha \;=\; f_{\alpha,\beta}dy_\beta + \tfrac{1}{2}f_{\alpha,\beta\gamma}\Omega_{\beta\gamma}dt = f_{\alpha,\beta}\circ dy_\beta \tag{58}$$

Thus, we obtain

$$\begin{aligned}-\delta S_1/\delta y_\alpha(t) \;&=\; f(y)_{\beta,\alpha}\Omega^{-1}_{\beta\gamma}\circ dy_\gamma(t)/dt - \tfrac{1}{2}f(y)_{\beta,\alpha\beta}\\ &-\; \Omega^{-1}_{\alpha\gamma}f(y)_{\gamma,\delta}\circ dy_\gamma(t)/dt\\ &=\; (f'_y\Omega^{-1} - \Omega^{-1}f_y)\circ y_t(t) - \tfrac{1}{2}\partial_y[\partial_y\cdot f(y)]\end{aligned}$$

where $\partial_y\cdot f(y) = f_{\beta,\beta} = \mathrm{div}(f)$. Finally we have (cf. 20)

$$-\delta S_2/\delta y(t) \;=\; -f'_y\Omega^{-1}f$$

and $\delta_{y(t)} \log p(y(t_0)) = \partial_{y_0} \log p(y_0)\delta(t-t_0)$. Putting all together one finds the Langevin drift functional (in matrix form)

$$\begin{aligned} -\frac{\delta\Phi(y|z)}{\delta y(t)} &:= F(y|z) \\ &= \sum_{i=0}^{T} h'_{iy}(y) R_i^{-1}(z_i - h_i(y))\delta(t-t_i) \\ &+ \Omega^{-1} y_{tt} + (f'_y\Omega^{-1} - \Omega^{-1} f_y) \circ y_t \\ &- \tfrac{1}{2}\partial_y[\partial_y \cdot f(y)] - f'_y\Omega^{-1} f \\ &+ \partial_{y_0} \log p(y_0)\delta(t-t_0) \end{aligned}$$

and the SPDE (cf. Hairer et al.; 2007)

$$dY(u,t) = F(Y(u,t)|z))du + \sqrt{2}\, dW_t(u,t), \tag{59}$$

where $W_t(u,t) = \partial_t W(u,t)$ is a cylindrical Wiener process with $E[W_t(u,t)] = 0$, $E[W_t(u,t)$ $W_s(v,s)'] = I_p \min(u,v)\delta(t-s)$ and $W(u,t)$ is a Wiener field (Brownian sheet). See, e.g. Jetschke (1986); Da Prato and Zabczyk (1992, ch. 4.3.3). The cylindrical Wiener process may be viewed as continuum limit of $W_j(u)/\sqrt{\delta t}$, $E[W_j(u)/\sqrt{\delta t}\ W'_k(v)/\sqrt{\delta t}] = I_p \min(u,v)\delta t^{-1}\delta_{jk}$.

State dependent diffusion matrix

In this case, new terms appear. Starting with the first term in (25), one gets

$$-\Omega^{-1}_{j\alpha\beta}\delta t^{-2}(\eta_{j+1;\beta} - 2\eta_{j\beta} + \eta_{j-1;\beta}) \quad \to \quad -\Omega(y(t))^{-1} \circ \ddot{y}(t).$$

The second term in (25) contains terms of the form $h_j\ (\eta_j - \eta_{j-1})$ which appear in a backward Itô integral. Here we attempt to write them in symmetrized (Stratonovich) form. It turns out, that the Taylor expansion (24) must be carried to higher orders. Writing (for simplicity in scalar form)

$$\Omega^{-1}_{j-1}\delta\eta_{j-1} - \Omega^{-1}_j\delta\eta_j := h_{j-1}\delta\eta_{j-1} - h_j\delta\eta_j$$

and expanding around η_j

$$h_{j-1} = h_j + \sum_{k=1}^{\infty} \frac{1}{k!} h_{j,k}(\eta_{j-1} - \eta_j)^k$$

one obtains

$$h_{j-1}\delta\eta_{j-1} - h_j\delta\eta_j = h_j(\delta\eta_{j-1} - \delta\eta_j) + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} h_{j,k}\delta\eta_{j-1}^{k+1}. \tag{60}$$

To obtain a symmetric expression, $h_{j,k}$ is expanded around $\eta_{j-1/2} := \frac{1}{2}(\eta_{j-1} + \eta_j)$. Noting that $\eta_j - \eta_{j-1/2} = \frac{1}{2}\delta\eta_{j-1}$ we have

$$h_{j,k} = \sum_{l=0}^{\infty} \frac{(\frac{1}{2})^l}{l!} h_{j-1/2,k+l}\delta\eta_{j-1}^l \tag{61}$$

and together

$$h_j(\delta\eta_{j-1} - \delta\eta_j) + \sum_{k=1,l=0}^{\infty} \frac{(-1)^k(\frac{1}{2})^l}{k!\ l!} h_{j-1/2,k+l}\delta\eta_{j-1}^{k+l+1}. \tag{62}$$

Multiplying with δt^{-2} and collecting terms to order $O(\delta t^2)$ one gets the continuum limit

$$-\Omega^{-1} \circ \ddot{y} - \Omega_\eta^{-1} \circ \dot{y}^2 - \tfrac{1}{24}\Omega^{-1}_{\eta\eta\eta}\Omega^2. \tag{63}$$

The last term in (25) is absorbed in the expression (29).

The continuum limit of the first two terms in the derivative of S_1 (see (27) is

$$-f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma}dy_\gamma(t)/dt + d[\Omega(y)^{-1}_{\alpha\gamma}f_\gamma(y)]/dt.$$

Transforming to Stratonovich calculus (57–58) yields

$$-\{f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma} - [\Omega(y)^{-1}_{\alpha\beta}f_\beta(y)]_{,\gamma}\} \circ dy_\gamma(t)/dt + \tfrac{1}{2}[f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma}]_{,\delta}\Omega_{\gamma\delta}. \tag{64}$$

Equation (28) yields

$$\delta S_2/\delta y_\alpha(t) \quad = \quad f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma}f(y)_\gamma + \tfrac{1}{2}f(y)_\beta\Omega(y)^{-1}_{\beta\gamma,\alpha}f(y)_\gamma. \tag{65}$$

The last term to be discussed is (29). Formally,

$$\begin{aligned} \tfrac{1}{2}\delta t^{-2}\mathrm{tr}\{\Omega^{-1}_{,\alpha}[\Omega dt - (dy - fdt)(dy - fdt)']\} \quad &= \\ \tfrac{1}{2}\mathrm{tr}\{\Omega^{-1}_{,\alpha}[\Omega\delta t^{-1} - (\dot{y} - f)(\dot{y} - f)']\}. & \end{aligned} \tag{66}$$

From the quadratic variation formula $(dy - fdt)(dy - fdt)' = \Omega dt$ it seems that it can be dropped. But setting $\delta\eta_j - f_j\delta t = g_j z_j\sqrt{\delta t}$ (from the Euler scheme, see (3)), one gets

$$X := \tfrac{1}{2}\delta t^{-1}\mathrm{tr}\{\Omega^{-1}_{j,\alpha}\Omega_j\ (I - z_j z_j')\}$$

In scalar form, one has $X := \frac{1}{2}\delta t^{-1}\Omega^{-1}_{j,\alpha}\Omega_j\ (I - z_j^2)$ which is χ^2_1-distributed, conditionally on η_j. One has $E[1 - z^2] = 0; \mathrm{Var}(1 - z^2) = 1 - 2 + 3 = 2$, thus $E[X] = 0$ and $\mathrm{Var}[X] = \frac{1}{2}\delta t^{-2}E[\Omega^{-2}_{j,\alpha}\Omega_j^2]$.

Therefore, the drift functional in the state dependent case is

$$\begin{aligned} -\frac{\delta\Phi(y|z)}{\delta y(t)} \quad &:= \quad F(y|z) \\ &= \quad (56) - (63) - (64) - (65) + (66) + \partial_{y_0}\log p(y_0)\delta(t - t_0) \end{aligned}$$

Discussion

The second order time derivative (diffusion term w.r.t. t) $\Omega^{-1}y_{tt}$ in the SPDE (59) resulted from the first term (51) in the Lagrangian corresponding to the driftless process (random walk process). Usually this (in the continuum limit) infinite term is not considered and removed by computing a density ratio (17) which leads to a well defined Radon-Nikodym density (52). On the other hand, the term is necessary to obtain the correct SPDE. Starting from the Radon-Nikodym density (55) for the process $dY(t) = fdt + GdW(t)$ at the outset, it is not quite clear how to construct the appropriate SPDE. Setting for

simplicity $f = 0$ and dropping the initial condition and the measurement part, eqn. (59) reads

$$dY(u,t) \quad = \quad \Omega^{-1}\, Y_{tt}(u,t)du + \sqrt{2}\, dW_t(u,t).$$

This linear equation (Ornstein-Uhlenbeck process) can be solved using a stochastic convolution as ($A := \Omega^{-1}\partial_t^2$)

$$Y(u,t) \quad = \quad \exp(Au)Y(0,t) + \int_0^u \exp(A(u-s))\sqrt{2}\, dW_t(s,t).$$

(cf. Da Prato; 2004, ch. 2). It is a Gaussian process with mean $\mu(u) = \exp(Au)E[Y(0)]$ and variance $Q(u) = \exp(Au)\mathrm{Var}(Y(0))\exp(A^*u) + \int_0^u \exp(As)2\exp(A^*s)ds$ where A^* is the adjoint of A. Thus the stationary distribution ($u \to \infty$) is the Gaussian measure $N(0, Q(\infty))$ with $Q(\infty) = -A^{-1} = -\Omega \cdot [\partial_t^2]^{-1}$, since $A = A^*$. But this coincides with $C(t,s) = \Omega \min(t,s)$, the covariance function of the scaled Wiener process $G \cdot W(t)$ (see (54); $\Omega = GG'$). Thus, for large u, $Y(u,t)$ generates trajectories of $GW(t)$. More generally ($f \neq 0$), one obtains solutions of SDE (1). A related problem occurs in the state dependent case $\Omega(y)$. Again, the term $\int dy'(\Omega dt)^{-1}dy$ yields a second order derivative in the SPDE, but after transforming to symmetrized Stratonovich form, also higher order terms appear (62, 63).

Moreover, the differential of Ω^{-1} in the Lagrangian (51-52) imports a problematic term similar to (51) into the SPDE, namely $\frac{1}{2}(\dot{y} - f)'(\Omega^{-1})_y(\dot{y} - f)$, which can be combined with the derivative of the Jacobian (cf. 66). Formally, it is squared white noise where the differentials are in Itô form. A procedure similar to (61), i.e.

$$h_{j,k} \quad = \quad \sum_{l=0}^{\infty} \frac{(-\frac{1}{2})^l}{l!} h_{j+1/2,k+l}\delta\eta_j^l \tag{67}$$

can be applied to obtain Stratonovich type expressions. Because of the dubious nature of these expressions, only the quasi continuous approach based on approximate finite dimensional densities and Langevin equations will we used in this paper.

References

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

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions, *Annals of Statistics* **36, 2**: 906–937.

Apte, A., Hairer, M., Stuart, A. M. and Voss, J. (2007). Sampling the posterior: An approach to non-Gaussian data assimilation, *Physica D: Nonlinear Phenomena* **230**(1–2): 50–64.

Apte, A., Jones, C. K. R. T., Stuart, A. M. and Voss, J. (2008). Data assimilation: Mathematical and statistical perspectives, *International Journal for Numerical Methods in Fluids* **56**(8): 1033–1046.

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

Bagchi, A. (2001). Onsager-Machlup function, *Encyclopedia of Mathematics*, Springer, Berlin, Heidelberg, New York. http://www.encyclopediaofmath.org/index.php/Onsager-Machlup_function.

Basawa, I. and Prakasa Rao, B. (1980). *Statistical Inference for Stochastic Processes*, Academic Press, London.

Bergstrom, A. (1988). The history of continuous-time econometric models, *Econometric Theory* **4**: 365–383.

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.

Chang, J., Chen, S. X. et al. (2011). On the approximate maximum likelihood estimation for diffusion processes, *The Annals of Statistics* **39**(6): 2820–2851.

Da Prato, G. (2004). *Kolmogorov equations for stochastic PDEs*, Birkhäuser, Basel.

Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions*, Cambridge University Press, New York.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for simulated maximum likelihood estimation of stochastic differential equations, *Journal of Business and Economic Statistics* **20**: 297–316.

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

Feller, W. (1951). Two singular diffusion problems, *Annals of Mathematics* **54**: 173–182.

Gard, T. (1988). *Introduction to stochastic differential equations*, Dekker, New York.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(2): 123–214.

Hairer, M., Stuart, A. M. and Voss, J. (2007). Analysis of SPDEs arising in path sampling, part II: The nonlinear case, *Annals of Applied Probability* **17**(5): 1657–1706.

Hairer, M., Stuart, A. M. and Voss, J. (2009). Sampling conditioned diffusions, *Trends in Stochastic Analysis*, Vol. 353 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, pp. 159–186.

Hairer, M., Stuart, A. M. and Voss, J. (2011). Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods, *in* D. Crisan and B. Rozovsky (eds), *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press, pp. 833–873.

Hairer, M., Stuart, A. M., Voss, J. and Wiberg, P. (2005). Analysis of SPDEs arising in path sampling, part I: The Gaussian case, *Communications in Mathematical Sciences* **3**(4): 587–603.

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. and Stock, J. (1985). The estimation of higher order continuous time autoregressive models, *Econometric Theory* **1**: 97–112.

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

Jetschke, G. (1986). On the Equivalence of Different Approaches to Stochastic Partial Differential Equations, *Math. Nachr.* **128**: 315–329.

Jetschke, G. (1991). Lattice approximation of a nonlinear stochastic partial differential equation with white noise, *International Series of Numerical Mathematics 102*, Birkhäuser Verlag, Basel, pp. 107–126.

Jones, R. (1984). Fitting multivariate models to unequally spaced data, *in* E. Parzen (ed.), *Time Series Analysis of Irregularly Observed Data*, Springer, New York, pp. 158–188.

Jones, R. and Tryon, P. (1987). Continuous time series models for unequally spaced data applied to modeling atomic clocks, *SIAM J. Sci. Stat. Comput.* **8**: 71–81.

Kac, M. (1980). *Integration in function spaces and some of its applications*, Scuola normale superiore.

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

Kloeden, P. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin. corrected third printing.

Langevin, P. (1908). Sur la théorie du mouvement brownien [on the theory of brownian motion], *Comptes Rendus de l'Academie des Sciences (Paris)* **146**: 530–533.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions, *The Annals of Statistics* **41**(3): 1350–1380.

Lighthill, M. (1958). *Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press, Cambridge, U.K.

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

Louis, T. (1982). Finding the Observed Information Matrix when Using the EM Algorithm, *Journal of the Royal Statistical Association B* **44, 2**: 226–233.

Mardia, K., Kent, J. and Bibby, J. (1979). *Multivariate Analysis*, Academic Press, London.

Moler, C. and VanLoan, C. (2003). Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later, *SIAM Review* **45,1**: 1–46.

Okano, K., Schülke, L. and Zheng, B. (1993). Complex Langevin Simulation, *Progress of Theoretical Physics Supplement* **111**: 313–346.

Onsager, L. and Machlup, S. (1953). Fluctuations and Irreversible Processes, *Physical Review* **91,6**: 1505–1515.

Oud, J. and Jansen, R. (2000). Continuous Time State Space Modeling of Panel Data by Means of SEM, *Psychometrika* **65**: 199–215.

Oud, J. H. and Singer, H. (2008). Continuous time modeling of panel data: SEM versus filter techniques, *Statistica Neerlandica* **62**(1): 4–28.

Ozaki, T. (1985). Nonlinear Time Series and Dynamical Systems, *in* E. Hannan (ed.), *Handbook of Statistics*, North Holland, Amsterdam, pp. 25 – 83.

Parisi, G. and Wu, Y.-S. (1981). Perturbation theory without gauge fixing, *Scientia Sinica* **24**: 483.

Pedersen, A. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations, *Scandinavian Journal of Statistics* **22**: 55–71.

Pitt, M. (2002). Smooth Particle Filters for Likelihood Evaluation and Maximisation, *Warwick economic research papers 651*, University of Warwick. http://wrap.warwick.ac.uk/1536/.

Reznikoff, M. and Vanden-Eijnden, E. (2005). Invariant measures of stochastic partial differential equations and conditioned diffusions, *C. R. Acad. Sci. Paris* **Ser. I 340**: 305–308.

Risken, H. (1989). *The Fokker-Planck Equation*, second edn, Springer, Berlin, Heidelberg, New York.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm, *Biometrika* **88**(3): 603–621.

Roberts, G. O. and Stramer, O. (2002). Langevin Diffusions and Metropolis-Hastings Algorithms, *Methodology And Computing In Applied Probability* **4**(4): 337–357.

Rümelin, W. (1982). Numerical treatment of stochastic differential equations, *SIAM J. Numer. Anal.* **19, 3**: 604–613.

Särkkä, S. (2013). *Bayesian filtering and smoothing*, Vol. 3, Cambridge University Press.

Särkkä, S., Hartikainen, J., Mbalawata, I. and Haario, H. (2013). Posterior inference on parameters of stochastic differential equations via non-linear Gaussian filtering and adaptive MCMC, *Statistics and Computing* pp. 1–11.

Schiesser, W. E. (1991). *The Numerical Method of Lines*, Academic Press, San Diego.

Shoji, I. and Ozaki, T. (1997). Comparative Study of Estimation Methods for Continuous Time Stochastic Processes, *Journal of Time Series Analysis* **18, 5**: 485–506.

Shoji, I. and Ozaki, T. (1998). A statistical method of estimation and simulation for systems of stochastic differential equations, *Biometrika* **85, 1**: 240–243.

Silverman, B. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Singer, H. (1990). *Parameterschätzung in zeitkontinuierlichen dynamischen Systemen [Parameter estimation in continuous time dynamical systems; Ph.D. thesis, University of Konstanz, in German]*, Hartung-Gorre-Verlag, Konstanz.

Singer, H. (1993). Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components, *Journal of Time Series Analysis* **14, 5**: 527–545.

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. (1998). Continuous Panel Models with Time Dependent Parameters, *Journal of Mathematical Sociology* **23**: 77–98.

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

Singer, H. (2005). Continuous-Discrete Unscented Kalman Filtering, *Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 384*, FernUniversität in Hagen. http://www.fernuni-hagen.de/ls_statistik/publikationen/ukf_2005.shtml.

Singer, H. (2008). Nonlinear Continuous Time Modeling Approaches in Panel Research, *Statistica Neerlandica* **62, 1**: 29–57.

Singer, H. (2011). Continuous-discrete state-space modeling of panel data with nonlinear filter algorithms, *Advances in Statistical Analysis* **95**: 375–413.

Singer, H. (2012). SEM modeling with singular moment matrices. Part II: ML-Estimation of Sampled Stochastic Differential Equations., *Journal of Mathematical Sociology* **36**(1): 22–43.

Singer, H. (2014). Importance Sampling for Kolmogorov Backward Equations, *Advances in Statistical Analysis* **98**: 345–369.

Singer, H. (2016). Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling, *Diskussionsbeiträge Fakultät Wirtschaftswissenschaft 497*, FernUniversität in Hagen. http://www.fernuni-hagen.de/wirtschaftswissenschaft/forschung/beitraege.shtml.

Stramer, O., Bognar, M. and Schneider, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions, *Journal of Financial Econometrics* **8**(4): 450–480.

Stratonovich, R. (1971). On the Probability Functional of Diffusion Processes, *Selected Translations in Mathematical Statistics and Probability*, Vol. 10, pp. 273–286.

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.

Stuart, A. M., Voss, J. and Wiberg, P. (2004). Conditional path sampling of SDEs and the Langevin MCMC method, *Communications in Mathematical Sciences* **2**(4): 685–697.

Wei, G. W., Zhang, D. S., Kouric, D. J. and Hoffman, D. K. (1997). Distributed approximating functional approach to the Fokker-Planck equation: Time propagation, *J. Chem. Phys.* **107 (8)**: 3239–3246.

Wong, E. and Hajek, B. (1985). *Stochastic Processes in Engineering Systems*, Springer, New York.

Yoo, H. (2000). Semi-discretization of stochastic partial differential equations on $\mathbb{R}^1$ by a finite-difference method, *Math. Comp.* **69, 230**: 653–666.

Die Diskussionspapiere ab Nr. 183 (1992) bis heute, können Sie im Internet unter http://www.fernuni-hagen.de/wirtschaftswissenschaft/forschung/beitraege.shtml einsehen und zum Teil downloaden.
Ältere Diskussionspapiere selber erhalten Sie nur in den Bibliotheken.

Nr	Jahr	Titel	Autor/en
420	2008	Stockkeeping and controlling under game theoretic aspects	Fandel, Günter Trockel, Jan
421	2008	On Overdissipation of Rents in Contests with Endogenous Intrinsic Motivation	Schlepütz, Volker
422	2008	Maximum Entropy Inference for Mixed Continuous-Discrete Variables	Singer, Hermann
423	2008	Eine Heuristik für das mehrdimensionale Bin Packing Problem	Mack, Daniel Bortfeldt, Andreas
424	2008	Expected A Posteriori Estimation in Financial Applications	Mazzoni, Thomas
425	2008	A Genetic Algorithm for the Two-Dimensional Knapsack Problem with Rectangular Pieces	Bortfeldt, Andreas Winter, Tobias
426	2008	A Tree Search Algorithm for Solving the Container Loading Problem	Fanslau, Tobias Bortfeldt, Andreas
427	2008	Dynamic Effects of Offshoring	Stijepic, Denis Wagner, Helmut
428	2008	Der Einfluss von Kostenabweichungen auf das Nash-Gleichgewicht in einem nicht-kooperativen Disponenten-Controller-Spiel	Fandel, Günter Trockel, Jan
429	2008	Fast Analytic Option Valuation with GARCH	Mazzoni, Thomas
430	2008	Conditional Gauss-Hermite Filtering with Application to Volatility Estimation	Singer, Hermann
431	2008	Web 2.0 auf dem Prüfstand: Zur Bewertung von Internet-Unternehmen	Christian Maaß Gotthard Pietsch
432	2008	Zentralbank-Kommunikation und Finanzstabilität – Eine Bestandsaufnahme	Knütter, Rolf Mohr, Benjamin
433	2008	Globalization and Asset Prices: Which Trade-Offs Do Central Banks Face in Small Open Economies?	Knütter, Rolf Wagner, Helmut
434	2008	International Policy Coordination and Simple Monetary Policy Rules	Berger, Wolfram Wagner, Helmut
435	2009	Matchingprozesse auf beruflichen Teilarbeitsmärkten	Stops, Michael Mazzoni, Thomas
436	2009	Wayfindingprozesse in Parksituationen - eine empirische Analyse	Fließ, Sabine Tetzner, Stefan
437	2009	ENTROPY-DRIVEN PORTFOLIO SELECTION a downside and upside risk framework	Rödder, Wilhelm Gartner, Ivan Ricardo Rudolph, Sandra
438	2009	Consulting Incentives in Contests	Schlepütz, Volker

439	2009	A Genetic Algorithm for a Bi-Objective Winner-Determination Problem in a Transportation-Procurement Auction"	Buer, Tobias Pankratz, Giselher
440	2009	Parallel greedy algorithms for packing unequal spheres into a cuboidal strip or a cuboid	Kubach, Timo Bortfeldt, Andreas Tilli, Thomas Gehring, Hermann
441	2009	SEM modeling with singular moment matrices Part I: ML-Estimation of time series	Singer, Hermann
442	2009	SEM modeling with singular moment matrices Part II: ML-Estimation of sampled stochastic differential equations	Singer, Hermann
443	2009	Konsensuale Effizienzbewertung und -verbesserung – Untersuchungen mittels der Data Envelopment Analysis (DEA)	Rödder, Wilhelm Reucher, Elmar
444	2009	Legal Uncertainty – Is Hamonization of Law the Right Answer? A Short Overview	Wagner, Helmut
445	2009	Fast Continuous-Discrete DAF-Filters	Mazzoni, Thomas
446	2010	Quantitative Evaluierung von Multi-Level Marketingsystemen	Lorenz, Marina Mazzoni, Thomas
447	2010	Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density	Mazzoni, Thomas Reucher, Elmar
448	2010	Solving a Bi-Objective Winner Determination Problem in a Transportation Procurement Auction	Buer, Tobias Pankratz, Giselher
449	2010	Are Short Term Stock Asset Returns Predictable? An Extended Empirical Analysis	Mazzoni, Thomas
450	2010	Europäische Gesundheitssysteme im Vergleich – Effizienzmessungen von Akutkrankenhäusern mit DEA –	Reucher, Elmar Sartorius, Frank
451	2010	Patterns in Object-Oriented Analysis	Blaimer, Nicolas Bortfeldt, Andreas Pankratz, Giselher
452	2010	The Kuznets-Kaldor-Puzzle and Neutral Cross-Capital-Intensity Structural Change	Stijepic, Denis Wagner, Helmut
453	2010	Monetary Policy and Boom-Bust Cycles: The Role of Communication	Knütter, Rolf Wagner, Helmut
454	2010	Konsensuale Effizienzbewertung und –verbesserung mittels DEA – Output- vs. Inputorientierung –	Reucher, Elmar Rödder, Wilhelm
455	2010	Consistent Modeling of Risk Averse Behavior with Spectral Risk Measures	Wächter, Hans Peter Mazzoni, Thomas

456	2010	Der virtuelle Peer – Eine Anwendung der DEA zur konsensualen Effizienz-bewertung –	Reucher, Elmar
457	2010	A two-stage packing procedure for a Portuguese trading company	Moura, Ana Bortfeldt, Andreas
458	2010	A tree search algorithm for solving the multi-dimensional strip packing problem with guillotine cutting constraint	Bortfeldt, Andreas Jungmann, Sabine
459	2010	Equity and Efficiency in Regional Public Good Supply with Imperfect Labour Mobility – Horizontal versus Vertical Equalization	Arnold, Volker
460	2010	A hybrid algorithm for the capacitated vehicle routing problem with three-dimensional loading constraints	Bortfeldt, Andreas
461	2010	A tree search procedure for the container relocation problem	Forster, Florian Bortfeldt, Andreas
462	2011	Advanced X-Efficiencies for CCR- and BCC-Modell – Towards Peer-based DEA Controlling	Rödder, Wilhelm Reucher, Elmar
463	2011	The Effects of Central Bank Communication on Financial Stability: A Systematization of the Empirical Evidence	Knütter, Rolf Mohr, Benjamin Wagner, Helmut
464	2011	Lösungskonzepte zur Allokation von Kooperationsvorteilen in der kooperativen Transportdisposition	Strangmeier, Reinhard Fiedler, Matthias
465	2011	Grenzen einer Legitimation staatlicher Maßnahmen gegenüber Kreditinstituten zur Verhinderung von Banken- und Wirtschaftskrisen	Merbecks, Ute
466	2011	Controlling im Stadtmarketing – Eine Analyse des Hagener Schaufensterwettbewerbs 2010	Fließ, Sabine Bauer, Katharina
467	2011	A Structural Approach to Financial Stability: On the Beneficial Role of Regulatory Governance	Mohr, Benjamin Wagner, Helmut
468	2011	Data Envelopment Analysis - Skalenerträge und Kreuzskalenerträge	Wilhelm Rödder Andreas Dellnitz
469	2011	Controlling organisatorischer Entscheidungen: Konzeptionelle Überlegungen	Lindner, Florian Scherm, Ewald
470	2011	Orientierung in Dienstleistungsumgebungen – eine explorative Studie am Beispiel des Flughafen Frankfurt am Main	Fließ, Sabine Colaci, Antje Nesper, Jens

471	2011	Inequality aversion, income skewness and the theory of the welfare state	Weinreich, Daniel
472	2011	A tree search procedure for the container retrieval problem	Forster, Florian Bortfeldt, Andreas
473	2011	A Functional Approach to Pricing Complex Barrier Options	Mazzoni, Thomas
474	2011	Bologna-Prozess und neues Steuerungsmodell – auf Konfrontationskurs mit universitären Identitäten	Jost, Tobias Scherm, Ewald
475	2011	A reduction approach for solving the rectangle packing area minimization problem	Bortfeldt, Andreas
476	2011	Trade and Unemployment with Heterogeneous Firms: How Good Jobs Are Lost	Altenburg, Lutz
477	2012	Structural Change Patterns and Development: China in Comparison	Wagner, Helmut
478	2012	Demografische Risiken – Herausforderungen für das finanzwirtschaftliche Risikomanagement im Rahmen der betrieblichen Altersversorgung	Merbecks, Ute
479	2012	“It’s all in the Mix!” – Internalizing Externalities with R&D Subsidies and Environmental Liability	Endres, Alfred Friehe, Tim Rundshagen, Bianca
480	2012	Ökonomische Interpretationen der Skalenvariablen u in der DEA	Dellnitz, Andreas Kleine, Andreas Rödder, Wilhelm
481	2012	Entropiebasierte Analyse von Interaktionen in Sozialen Netzwerken	Rödder, Wilhelm Brenner, Dominic Kulmann, Friedhelm
482	2013	Central Bank Independence and Financial Stability: A Tale of Perfect Harmony?	Berger, Wolfram Kißmer, Friedrich
483	2013	Energy generation with Directed Technical Change	Kollenbach, Gilbert
484	2013	Monetary Policy and Asset Prices: When Cleaning Up Hits the Zero Lower Bound	Berger, Wolfram Kißmer, Friedrich
485	2013	Superknoten in Sozialen Netzwerken – eine entropieoptimale Analyse	Brenner, Dominic, Rödder, Wilhelm, Kulmann, Friedhelm
486	2013	Stimmigkeit von Situation, Organisation und Person: Gestaltungsüberlegungen auf Basis des Informationsverarbeitungsansatzes	Julmi, Christian Lindner, Florian Scherm, Ewald
487	2014	Incentives for Advanced Abatement Technology Under National and International Permit Trading	Endres, Alfred Rundshagen, Bianca

488	2014	Dynamische Effizienzbewertung öffentlicher Dreispartentheater mit der Data Envelopment Analysis	Kleine, Andreas Hoffmann, Steffen
489	2015	Konsensuale Peer-Wahl in der DEA -- Effizienz vs. Skalenertrag	Dellnitz, Andreas Reucher, Elmar
490	2015	Makroprudenzielle Regulierung – eine kurze Einführung und ein Überblick	Velauthapillai, Jeyakrishna
491	2015	SEM modeling with singular moment matrices Part III: GLS estimation	Singer, Hermann
492	2015	Die steuerliche Berücksichtigung von Aufwendungen für ein Studium − Eine Darstellung unter besonderer Berücksichtigung des Hörerstatus	Meyering, Stephan Portheine, Kea
493	2016	Ungewissheit versus Unsicherheit in Sozialen Netzwerken	Rödder, Wilhelm Dellnitz, Andreas Gartner, Ivan
494	2016	Investments in supplier-specific economies of scope with two different services and different supplier characters: two specialists	Fandel, Günter Trockel, Jan
495	2016	An application of the put-call-parity to variance reduced Monte-Carlo option pricing	Müller, Armin
496	2016	A joint application of the put-call-parity and importance sampling to variance reduced option pricing	Müller, Armin
497	2016	Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling	Singer, Hermann
498	2016	A Theory of Affective Communication	Julmi, Christian
499	2016	Approximations of option price elasticities for importance sampling	Müller, Armin
500	2016	Variance reduced Value at Risk Monte-Carlo simulations	Müller, Armin
501	2016	Maximum Likelihood Estimation of Continuous-Discrete State-Space Models: Langevin Path Sampling vs. Numerical Integration	Singer, Hermann
502	2016	Measuring the domain-specificity of creativity	Julmi, Christian Scherm, Ewald
503	2017	Bipartite Strukturen in Sozialen Netzen – klassische versus MaxEnt-Analysen	Rödder, Wilhelm Dellnitz, Andreas Kulmann, Friedhelm Litzinger, Sebastian Reucher, Elmar
504	2017	Langevin and Kalman Importance Sampling for Nonlinear Continuous-Discrete State Space Models	Singer, Hermann