Maximum Likelihood Estimation of Continuous-Discrete State-Space Models: Langevin Path Sampling vs. Numerical Integration

Hermann Singer

Diskussionsbeitrag Nr. 501

November 2016

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

Maximum Likelihood Estimation of Continuous-Discrete State-Space Models: Langevin Path Sampling vs. Numerical Integration

Hermann Singer
FernUniversität in Hagen *

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. We use importance sampling where the importance density is the conditional density of the latent states given the measurements. We compute analytical derivatives of the log importance density and consider a continuum limit, where the sampler is given by a stochastic partial differential equation. Since the conditional density is only known up to a normalizing constant, we use instead a kernel density estimator. Alternatively, a suboptimal, but explicitly known, importance density is employed. This may be obtained by using a certain reference parameter vector, where the drift function is zero or linear. We compare the Monte Carlo results with numerical methods based on the solution of the Fokker-Planck equation between measurements. We use the repeated multiplication of transition matrices based on a) Euler transition kernels, b) finite differences and c) discretized integral operators. The methods are illustrated for one and two-dimensional Ornstein-Uhlenbeck processes, where analytical results are known.

Key Words

Continuous-discrete state space models; Stochastic differential equations; Maximum likelihood estimation; Langevin path sampling; Nonlinear filtering and smoothing; Stochastic partial differential equations

*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 stochastic differential equations is of high practical importance since most data sets are only collected at discrete, often large sampling times. Nevertheless, exact ML estimation is possible, if one can compute the transition density between the sampling intervals. In the linear case, the approach was pioneered by Bergstrom (1988), who introduced the 'exact discrete model', an exact autoregressive scheme with parameter restrictions due to discrete sampling. 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), Monte Carlo methods (Pedersen; 1995; Elerian et al.; 2001; Singer; 2002; Beskos et al.; 2006; Stramer et al.; 2010; Särkkä et al.; 2013) and numerical solutions of the Fokker-Planck equation (Risken; 1989; Wei et al.; 1997; Mazzoni; 2012).

In this paper the likelihood function is computed by integrating out 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 (Hairer et al.; 2005, 2007; Apte et al.; 2007, 2008; Hairer et al.; 2009) combined with importance sampling.

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. Then, a continuum limit is considered. In section 3.4. we discuss the problems, when starting from a continuous time theory at the outset, especially when using probability densities w.r.t. the Wiener measure. Section 4 discusses the maximum likelihood approach and the estimation 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 computed 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 one and two-dimensional Ornstein-Uhlenbeck process are considered and the Fokker-Planck as well as Monte Carlo approach are compared.

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 and a numerically feasible approach, the SDE (1) is first replaced by the Euler-Maruyama approximation

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

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

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

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 $\tilde{\eta} := \{\eta_J, ..., \eta_1\}$]

$$
\begin{aligned}
p(\eta) &= \prod_{j=0}^{J-1} p(\eta_{j+1}|\eta_j)\; p(\eta_0) = p(\tilde{\eta}|\eta_0)p(\eta_0) && (5)\\
&:= Z^{-1}e^{-S}\; p(\eta_0) && (6)\\
S &= \tfrac{1}{2}\sum_{j=0}^{J-1}(\eta_{j+1} - \eta_j - f_j\delta t)'(\Omega_j\delta t)^{-1}(\eta_{j+1} - \eta_j - f_j\delta t) && (7)\\
Z &= \prod_{j=0}^{J-1} |2\pi\Omega_j\delta t|^{1/2}, && (8)
\end{aligned}
$$

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

$$
p(z) = \int p(z|\eta)p(\eta)d\eta. \qquad (9)
$$

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

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

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

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

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 sect. 3.3), 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 $\frac{\delta}{\delta \eta(t)}$.[1]

3.2 Drift function

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

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

must be computed analytically.[2]

---

[1] 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 sect. 3.3

[2] **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.

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[3]. Therefore, the density $p(\eta)$ can be factorized as $p(\eta) = p_0(\eta)\ \alpha(\eta)$ where

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

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 constant coefficients in multivariate models.[4] 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 ($j = 1, ..., J-1$)

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

[3]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.

[4]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).

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 derivative reads (matrix form in the second line)

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

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 $f_\eta(\eta_j)$.

3.3 Continuum limit

The expressions above 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 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).

3.3.1 Constant diffusion matrix

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

$$\begin{aligned}
S &= \tfrac{1}{2}\int dy(t)'(\Omega dt)^{-1}dy(t) && (30)\\
&- \int f(y)'\Omega^{-1}dy(t) + \tfrac{1}{2}\int f(y)'\Omega^{-1}f(y)dt. && (31)
\end{aligned}$$

If $y(t)$ is a sample function of the diffusion process $Y(t)$ in (1), the first term (30) 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, (30) 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

$$-\tfrac{1}{2}\int y(t)'\Omega^{-1}\ddot{y}(t)dt \tag{32}$$

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

$$\partial^2/\partial t^2 \min(t,s) = -\delta(t-s), \tag{33}$$

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 (31) are well defined and yield the Radon-Nikodym derivative (cf. 17)

$$\alpha(y) \;\; = \;\; \exp\Big\{\int f(y)'\Omega^{-1}dy(t) - \tfrac{1}{2}\int f(y)'\Omega^{-1}f(y)dt\Big\}. \tag{34}$$

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), $\alpha = 1,...,p$

$$\delta\log p(z|y)/\delta y_\alpha(t) \;\; = \;\; \sum_{i=0}^{T} h'_{i\bullet,\alpha}R_i^{-1}(z_i - h_i)\delta(t-t_i) \tag{35}$$

where the scaled Kronecker delta ($\delta_{jj_i}/\delta t$) was replaced by the delta function (see footnote 1). 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 (32). 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

$$h_{\alpha\beta}dy_\beta \;\; = \;\; h_{\alpha\beta}\circ dy_\beta - \tfrac{1}{2}h_{\alpha\beta,\gamma}\Omega_{\beta\gamma}dt \tag{36}$$

$$df_\alpha \;\; = \;\; f_{\alpha,\beta}dy_\beta + \tfrac{1}{2}f_{\alpha,\beta\gamma}\Omega_{\beta\gamma}dt = f_{\alpha,\beta}\circ dy_\beta \tag{37}$$

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)

$$dY(u,t) = F(Y(u,t)|z))du + \sqrt{2}\, dW_t(u,t), \tag{38}$$

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 $W_j(u)/\sqrt{\delta t}$, $E[W_j(u)/\sqrt{\delta t}\; W'_k(v)/\sqrt{\delta t}] = I_p \min(u,v)\delta t^{-1}\delta_{jk}$.

3.3.2 State dependent diffusion matrix

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

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

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

$$h_{j-1}\delta\eta_{j-1} - h_j\delta\eta_j \;=\; h_j(\delta\eta_{j-1} - \delta\eta_j) + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} h_{j,k}\delta\eta_{j-1}^{k+1}. \tag{39}$$

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

$$h_{j,k} \;=\; \sum_{l=0}^{\infty} \frac{(\frac{1}{2})^l}{l!} h_{j-1/2,k+l}\delta\eta_{j-1}^l \tag{40}$$

and together

$$h_j(\delta\eta_{j-1} - \delta\eta_j) + \sum_{k=1,l=0}^{\infty} \frac{(-1)^k(\frac{1}{2})^l}{k!\; l!} h_{j-1/2,k+l}\delta\eta_{j-1}^{k+l+1}. \tag{41}$$

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

$$-\Omega^{-1} \circ \ddot{y} - \Omega_\eta^{-1} \circ \dot{y}^2 - \tfrac{1}{24}\Omega_{\eta\eta\eta}^{-1}\Omega^2. \tag{42}$$

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 (36–37) yields

$$-\{f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma} - [\Omega(y)^{-1}_{\alpha\beta}f_\beta(y)]_{,\gamma}\} \circ dy_\gamma(t)/dt + \tfrac{1}{2}[f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma}]_{,\delta}\Omega_{\gamma\delta}. \tag{43}$$

Equation (28) yields

$$\delta S_2/\delta y_\alpha(t) \;=\; f(y)_{\beta,\alpha}\Omega(y)^{-1}_{\beta\gamma}f(y)_\gamma + \tfrac{1}{2}f(y)_\beta\Omega(y)^{-1}_{\beta\gamma,\alpha}f(y)_\gamma. \tag{44}$$

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

$$\begin{aligned} \tfrac{1}{2}\delta t^{-2}\mathrm{tr}\{\Omega_{,\alpha}^{-1}[\Omega dt - (dy - f dt)(dy - f dt)']\} \;&=\; \\ \tfrac{1}{2}\mathrm{tr}\{\Omega_{,\alpha}^{-1}[\Omega\delta t^{-1} - (\dot{y} - f)(\dot{y} - f)']\}&. \end{aligned} \tag{45}$$

From the quadratic variation formula $(dy - f dt)(dy - f dt)' = \Omega dt$ 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

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

In scalar form, one has $X := \frac{1}{2}\delta t^{-1}\Omega_{j,\alpha}^{-1}\Omega_j\ (I - z_j^2)$ which is χ_1^2-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 $\mathrm{Var}[X] = \frac{1}{2}\delta t^{-2}E[\Omega_{j,\alpha}^{-2}\Omega_j^2]$.

Therefore, the drift functional in the state dependent case is

$$\begin{aligned} -\frac{\delta\Phi(y|z)}{\delta y(t)} \;&:=\; F(y|z) \\ &=\; (35) - (42) - (43) - (44) + (45) + \partial_{y_0}\log p(y_0)\delta(t - t_0) \end{aligned}$$

3.4 Discussion

The second order time derivative (diffusion term w.r.t. t) $\Omega^{-1}y_{tt}$ in the SPDE (38) resulted from the first term (30) 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 (31). On the other hand, the term is necessary to obtain the correct SPDE. Starting from the Radon-Nikodym density (34) 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. (38) reads

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

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

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

(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 (33); $\Omega = GG'$). 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 (41, 42).

Moreover, the differential of Ω^{-1} in the Lagrangian (30-31) imports a problematic term similar to (30) 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. 45). Formally, it is squared white noise where the differentials are in Itô form. A procedure similar to (40), i.e.

$$h_{j,k} \;=\; \sum_{l=0}^{\infty} \frac{(-\frac{1}{2})^l}{l!} h_{j+1/2,k+l}\delta\eta_j^l \tag{46}$$

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 further.

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

$$p(z) \;=\; \int \frac{p(z|\eta)p(\eta)}{p(\eta|z)} p(\eta|z)d\eta := E[g(\eta,z)|z], \tag{47}$$

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

$$\hat{p}(z;U) \;=\; U^{-1}\int_0^U g(\eta(u),z)du \approx N^{-1}\sum_{l=0}^{N-1} g(\eta_l,z) := \hat{p}(z;\delta u,N) \tag{48}$$

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 (48), the importance (smoothing) density $p(\eta|z)$ is unkown, 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 was 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). Using the discretization as defined above one obtains a sequence of approximate solutions $\eta_l^{\delta u} \approx \eta_l \equiv \eta(u_l)$, which in the equilibrium (large l) generates a correlated sample from $p(\eta|z)$. Here, the error due to the numerical scheme is denoted by the superscript δu. The additional approximation error due to the Euler approximation (3) could be displayed by $\eta_l^{(\delta t,\delta u)}$. In order to eliminate effects of nonstationarity, one can drop a certain percentage of values η_l to obtain an equilibrium sample of size N.

Since only functionals of the process $\eta(u)$ must be computed (see (47)), it is sufficient to use weak approximations for the Langevin equation. One does not need strong (pathwise) approximations. Weak convergence of order β means, that the functional $E[g(\eta_l)]$ is well approximated in the limit $\delta u \to 0$

$$\left|E[g(\eta_l^{\delta u})] - E[g(\eta_l)]\right| < C(\delta u)^\beta. \tag{49}$$

Likewise, one can employ the ergodic criterion of Talay (1990), which means that $|\hat{p}(z;\delta u) - p(z)| < C(\delta u)^\beta$, where $\hat{p}(z;\delta u) = \lim_{N\to\infty}\hat{p}(z;\delta u,N)$ is the discrete approximation of the time average. This criterion may be viewed as an extension of the weak convergence criterion (49) to the infinite time horizon $l \to \infty$. Since the error term $\sqrt{2/\delta t}\; dW_j$ in (14) does not contain multiplicative noise, the Heun

scheme is of weak (and ergodic) order $\beta = 2$ (see, Talay 1990,Kloeden and Platen 1999, chs. 9.7, 15.1, 17.2, p. 542).

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

$$d\eta(u) \;=\; -K\delta_\eta \Phi(\eta|z)du + \sqrt{2}K^{1/2}\, dW(u)/\sqrt{\delta t}, \tag{50}$$

$K = K^{1/2}(K^{1/2})'$ 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.

In the following, the problem of computing the likelihood function (47) will be solved by two methods:

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

2. using a (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

$$p(z) \;=\; p_0(z) \int \frac{p(z|\eta)}{p_0(z|\eta)} \frac{p(\eta)}{p_0(\eta)} p_0(\eta|z)d\eta. \tag{51}$$

4.1.1 Density estimation approach

In the first approach, a sample η_l is generated from the Langevin equation (14) using numerical methods for SDE, e.g. Heun's method or a 4th order stochastic Runge-Kutta scheme. Then, the data η_l are used to compute a kernel density

$$\hat{p}(\eta|z) \;=\; N^{-1}\sum_{l=1}^{N} \kappa(\eta - \eta_l; H) \tag{52}$$

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.). This estimate is inserted into (47). Other estimation methods are discussed in Singer (2016). Alternatively, one can use parametric estimates, e.g. a Gauss density $\phi(\eta; \bar{\eta}, S)$.

4.1.2 Reference measure approach

A well known example of the second approach is the representation of the transition density by Dacunha-Castelle and Florens-Zmirou (1986) using a Brownian bridge. It can be obtained by considering the special case of direct measurements of the scalar SDE $dY = f(Y, \theta)dt + \sigma dW$ at two times t_0, t_1, i.e. $z_0 = y_0 = x, z_1 = y_1 = y$. Again we use the Euler approximation (3) $\eta_{j+1} = \eta_j + f_j \delta t + \sigma \delta W_j; \eta_0 = x$. Then, by inserting the auxiliary states $\tilde{\eta} = \{\eta_{J-1}, ..., \eta_1\}$, (10) takes the form (conditionally on x)

$$\begin{aligned} p(y|x) &= \int p(y|x, \tilde{\eta}) p(\tilde{\eta}|x) d\tilde{\eta} \\ &= \int \frac{p(y, \tilde{\eta}|x)}{p_0(\tilde{\eta}|y, x)} p_0(\tilde{\eta}|y, x) d\tilde{\eta} \\ &= p_0(y|x) \int \frac{p(y, \tilde{\eta}|x)}{p_0(y, \tilde{\eta}|x)} p_0(\tilde{\eta}|y, x) d\tilde{\eta}. \end{aligned}$$

Choosing a Gaussian reference measure p_0 for $\eta = \{y, \tilde{\eta}\} = \{\eta_J, ..., \eta_1\}$, $E[\eta_j] = x, \mathrm{Cov}(\eta_j, \eta_k) = \sigma^2 \min(j, k)\delta t$, corresponding to drift $f = 0$, one has $p_0(y|x) = \phi(y; x, \sigma^2(t_1 - t_0))$. Of course, the density $p(\eta|x)$ is only an (Euler) approximation of the true finite dimensional density $p(y, t_1; y_{J-1}, \tau_{J-1}; ..., y_1, \tau_1|x, t_0)$ of the process $Y(t)$ at times $t_1 = \tau_J, \tau_{J-1}, ..., \tau_1$. Thus in a continuum limit one obtains (cf. eqn. 34)

$$\begin{aligned} p(y|x) &= p_0(y|x)\, E\big[\alpha(Y)|Y(t_1) = y, Y(t_0) = x\big] \\ \alpha(y) &= \exp\left\{\sigma^{-2}\left[\int f(y)dy(t) - \tfrac{1}{2}\int f(y)^2 dt\right]\right\}, \end{aligned} \tag{53}$$

which is an exectation with respect to a Brownian bridge pinned at (x, y). The derivation may be viewed as a poor man's approach to the work cited above.

Some remarks are in order:

1. Of course, other reference densities can be used. For example, one could use the Ornstein Uhlenbeck process, derived from $dY = aYdt + \sigma dW$. In fact, the performance of the sampler is improved if the reference density is closer to the optimal choice $p_{opt} = p(\eta|z)$.

2. In the finite dimensional case, one can use arbitrary densities p_0. Usually they are derived from $\eta_{j+1} = \eta_j + f_j(\psi_0)\delta t + g_j(\psi_0)\delta W_j$ using a reference point ψ_0 in parameter space. If a continuum limit is attempted, one must require equality of the diffusion terms. Otherwise the Jacobian terms and the infinite expression (15) do not cancel when computing the density ratio (17). Only in this case the measures in function space are equivalent and the Radon-Nikodym density is given in (53). Writing the representation (51) as

$$p(z;\psi) := p_0(z;\psi_0)\; p_1(z;\psi,\psi_0), \tag{54}$$

one must consider the restriction $g(\psi) = g(\psi_0)$ in optimization algorithms. In the example above, we have $\psi = \{\theta, \sigma\}$, $\psi_0 = \{\theta_0, \sigma\}$ and $f(\psi) = f(\theta), g(\psi) = g(\psi_0) = \sigma$.

4.1.3 Score function

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

$$\begin{aligned} s(z) &= \partial/\partial\psi \log \int p(z,\eta)d\eta = p(z)^{-1} \int s(z,\eta)p(z,\eta)d\eta \\ &= E[s(z,\eta)|z], \end{aligned} \tag{55}$$

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

$$\hat{s}(z) = N^{-1} \sum_l s(z,\eta_l). \tag{56}$$

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.

4.1.4 Bayesian estimation

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

$$p(\eta,\psi|z) = p(z,\eta,\psi)/p(z) = p(z|\eta,\psi)p(\eta|\psi)p(\psi)/p(z). \tag{57}$$

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

$$d\eta(u) = \delta_\eta \log p(\eta,\psi|z)du + \sqrt{2}\; dW_1(u)/\sqrt{\delta t} \tag{58}$$

$$d\psi(u) = \partial_\psi \log p(\eta,\psi|z)du + \sqrt{2}\; dW_2(u), \tag{59}$$

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

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 also 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 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

$$\eta_{k_\alpha} = a_\alpha + k_\alpha \delta\eta_\alpha \tag{60}$$

$\eta_{K_\alpha} = b_\alpha, \alpha = 1, ..., p$, 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)

$$f(y) \approx f(y_j) + f_y(y_j)(y - y_j) + \tfrac{1}{2} f_{yy}(y_j)\Omega(y_j)(t - t_j). \tag{61}$$

The approach is known under the label local linearization (LL). Inserting (61) 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 $x(t)$ is constant in the interval $[\tau_j, \tau_{j+1})$, so $p(y_{j+1}, \tau_{j+1}|y_j, \tau_j) = \exp[F_j(y_{j+1})\delta t]\ \delta(y_{j+1} - y_j)$. 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 60) 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

$$F_{kl} = -\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\delta_{k_p l_p} f_\alpha(y_l) + \tfrac{1}{2}\delta_{k_1 l_1}...\nabla_{k_\alpha l_\alpha}...\nabla_{k_\beta l_\beta}...\delta_{k_p l_p}\Omega_{\alpha\beta}(y_l). \quad (62)$$

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

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

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. 62)

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

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 (62). 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 Ornstein-Uhlenbeck process (one dimensional)

First a univariate example is considered, which allows an exact computation of the likelihood function. The latent Ornstein-Uhlenbeck process

$$\begin{aligned} dY &= a\,Y dt + g\,dW \\ Z_i &= Y_i + \epsilon_i \end{aligned}$$

was simulated with true parameter vector $\psi^0 = \{a^0, g^0\} = \{-1, 2\}$, $\mathrm{Var}(\epsilon_i) = 0.1$ in the time interval $[t_0, t_T] = [0, 10]$ and measured with sampling interval $\Delta t = 0.5$. The discretization interval was set to $\delta t = 0.1$. Using an Euler scheme one obtains the latent trajectory $\eta = (\eta_0, ..., \eta_J), J = 10/0.1 = 100$ on the time grid $\tau_j = j\delta t$ and data points $z_i = \eta_{j_i} + \epsilon_i, i = 0, ..., 20$.

The Langevin equation (14) for the simulation of $\eta(u)$ in the fictious (simulation) time dimension u is thus an Itô equation with $J+1 = 101$ components. It can be solved approximately using numerical schemes for SDE, e.g. a 4th order Runge-Kutta scheme (see section 4.1). Using a discretization interval $\delta u, u_l = l\delta u, l =$

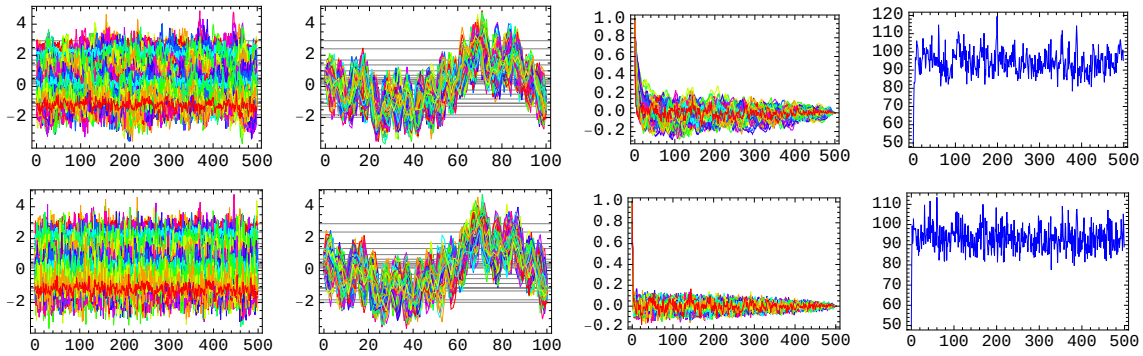

Figure 1: Langevin sampling of Ornstein-Uhlenbeck process. First row: Sampling without kernel. Second row: with kernel. First column: trajectories as a function of (fictious) time u. Second column: trajectories as a function of time t. Third column: Autocorrelation of states $\eta_j(u)$. Fourth column: potential $\Phi = -\log p(\eta(u)|z)$.

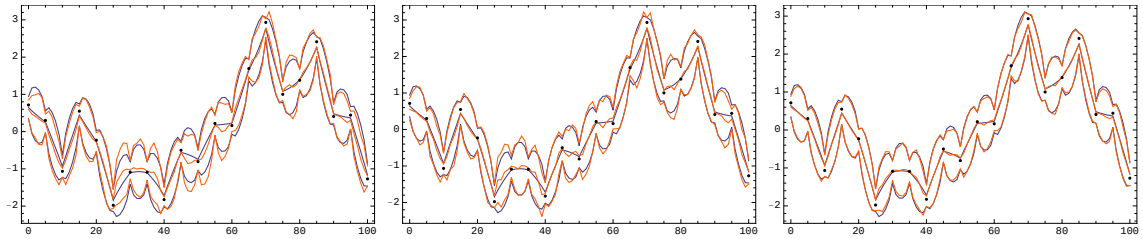

Figure 2: True (blue) and simulated (orange) a posteriori intervals $E[\eta|z] \pm \mathrm{Var}[\eta|z]^{1/2}$. Left: without kernel. Middle: with kernel, $\delta u = 1$. Right: with kernel, $\delta u = 5$.

$0, ..., N$ one obtains a sequence of approximate solutions $\eta_{lj}, l = 0, ..., N; j = 0, ..., J$, which in the equilibrium sample from $p(\eta|z)$. Fig. 1, left column, shows the sample path of η_{lj} for 500 simulation steps. For large l, one obtains conditional trajectories $\eta|z$ in the true time dimension (second column). The conditional moment $E[g(\eta)|z]$ can thus be approximated as $N^{-1} \sum_l g(\eta_l|z)$. In the following results, the first 10% of the samples were discarded to diminish effects of nonstationarity. The Langevin equation was initialized with a starting value η_{0j} obtained from a Kalman smoother (conditional mean). The third column of fig. 1 displays the sample autocorrelation function of η_{lj} as function of l. Due to the premultiplied kernel $K = -H^{-1}, H_{jj'} = \delta_{\eta_j}\delta_{\eta_{j'}} \log p(\eta|z)$ (Hessian; cf. eqn. 14 and sect. 4.1) in the second row, the autocorrelation decays faster, leading to more efficient Monte Carlo estimates (see fig. 2). The kernel was computed beforehand after a certain amount of simulation time. Finally, the right column of fig. 1 shows the potential $\Phi = -\log p(\eta(u)|z)$ as a function of simulation time. The Langevin equation may be visualized as a diffusion near the minimum of a high dimensional potential well.

Fig. 2 shows exact and estimated a posteriori intervals $E[\eta|z] \pm \mathrm{Var}[\eta|z]^{1/2}$, where the exact solution was obtained with a Kalman smoother. Clearly, the usage of a

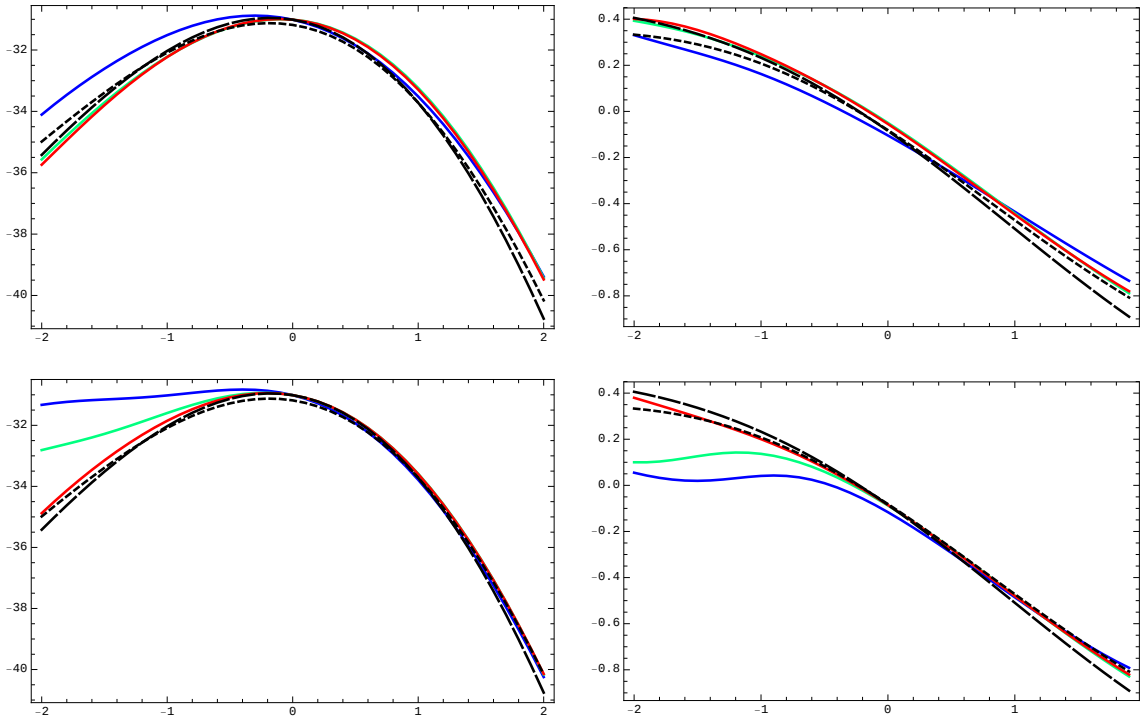

Figure 3: Likelihood surface as a function of drift a. Reference density method, $\psi_0 = \{-1, 2\}$. First row: without kernel, Second row: with kernel $-H^{-1}$. 1st column: likelihood, 2nd column: score, Shown are: Kalman filter (KF): exact (black, long dash), linearized (black, short dash), simulated: $N = 100, 500, 1000$ (blue, green, red). The abscissa is the deviation from the true value $a^0 = -1$.

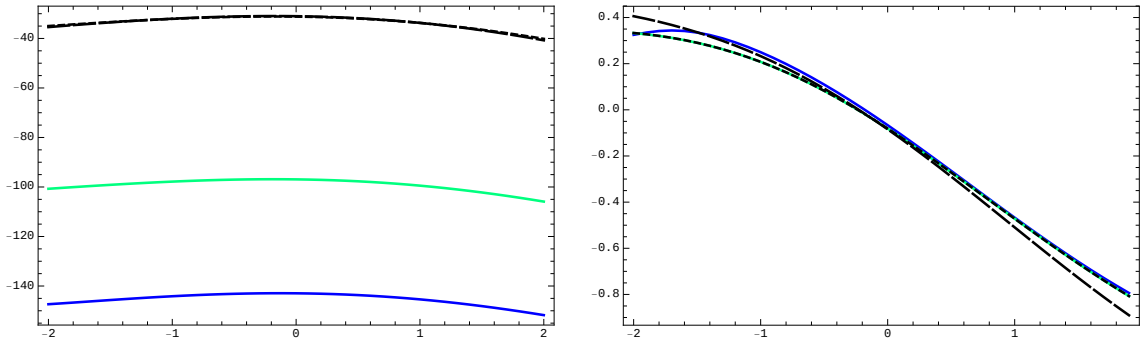

Figure 4: Likelihood surface as a function of drift a. Kernel density method. $N = 100$ (blue), $N = 200$ (green). Also shown are the exact (black, long dash) and linearized solution (black, short dash). For $N = 200$, the simulated score is nearly indistinguishable from the linearized KF solution (right picture).

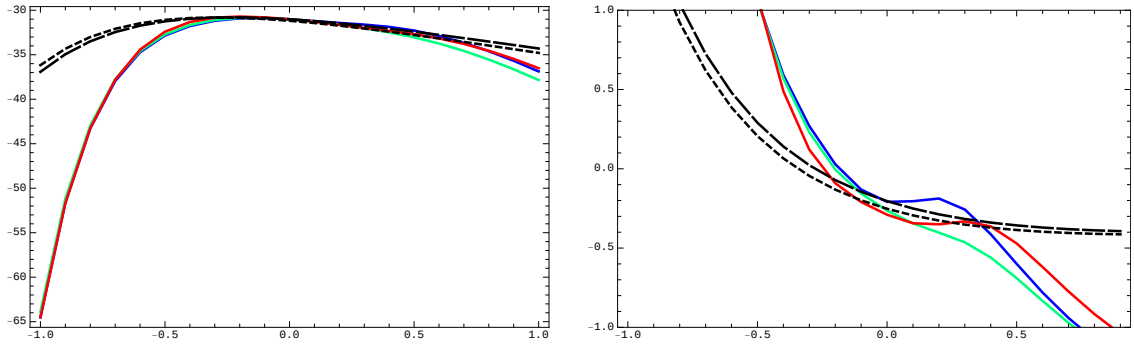

Figure 5: Likelihood surface as a function of diffusion coefficient g. Fixed reference density, $\psi_0 = \{-1, 2\}$. $N = 500, 1000, 2000$ (blue, green, red). Strong deviation from exact and linearized KF (black). Plotted on the abscissa is the deviation from the true value $g^0 = 2$.

kernel matrix improves the quality of the estimated expectation values.

Fig. 3 displays the log-likelihood surface $l(a, g = 2)$, where the function was evaluated at drift values $-3, ..., 1$ with step size 0.1. Displayed is the deviation from the true value $a^0 = -1$. The first and second rows correspond to a simulation without and with kernel $K = -H^{-1}$. The functions plotted in the columns (from left) are likelihood and gradient (score). Shown are Monte Carlo estimates obtained with the reference density method (see sect. 4.1.2), where the reference point $\psi_0 = \{-1, 2\}$ was chosen equal to the true parameter. For comparison, exact values from a Kalman filter (KF) were computed (long dashed). Also displayed are the results of a KF with linearized parameters (e.g. $\exp(a\delta t) \approx 1 + a\delta t$; KFL, short dashed).

The random vectors η_l were simulated with and without kernel, using a 4th order Runge-Kutta scheme. The results of the several integration schemes differ, but this depends on the discretization interval δu. Generally, for the Runge-Kutta method larger steps are possible. The effect of the kernel is an improvement of the likelihood simulation, depending on the sample size. For $N = 100, 500$, the results are worse, but for $N = 1000$, the likelihood is very close to the (linearized) KF solution (fig. 3, right, bottom).

For comparison, the kernel density approach (section 4.1.1) is displayed in fig. 4. It transpires that this method, using an estimate of the importance density $p(\eta|z)$, is more efficient. Already for $N = 200$, the simulated score is nearly indistinguishable from the linearized KF solution (right picture). For a correct absolute value of the likelihood one needs a larger sample size, however (left picture).

Up to now the diffusion parameter g was fixed. If g is varied (from 1,...,3 with step 0.1) and $\psi_0 = \{a_0, g_0\}$ is fixed, the density ratio in (51) does contain terms which do not cancel in the continuum limit (see sect. 4.1.2). Also numerically, the results are disappointing (see fig. 5). Therefore, one must vary the diffusion parameter both in the numerator and in the denominator of (51), i.e. $\psi = \{a, g\}$, $\psi_0 = \{a_0, g\}$.

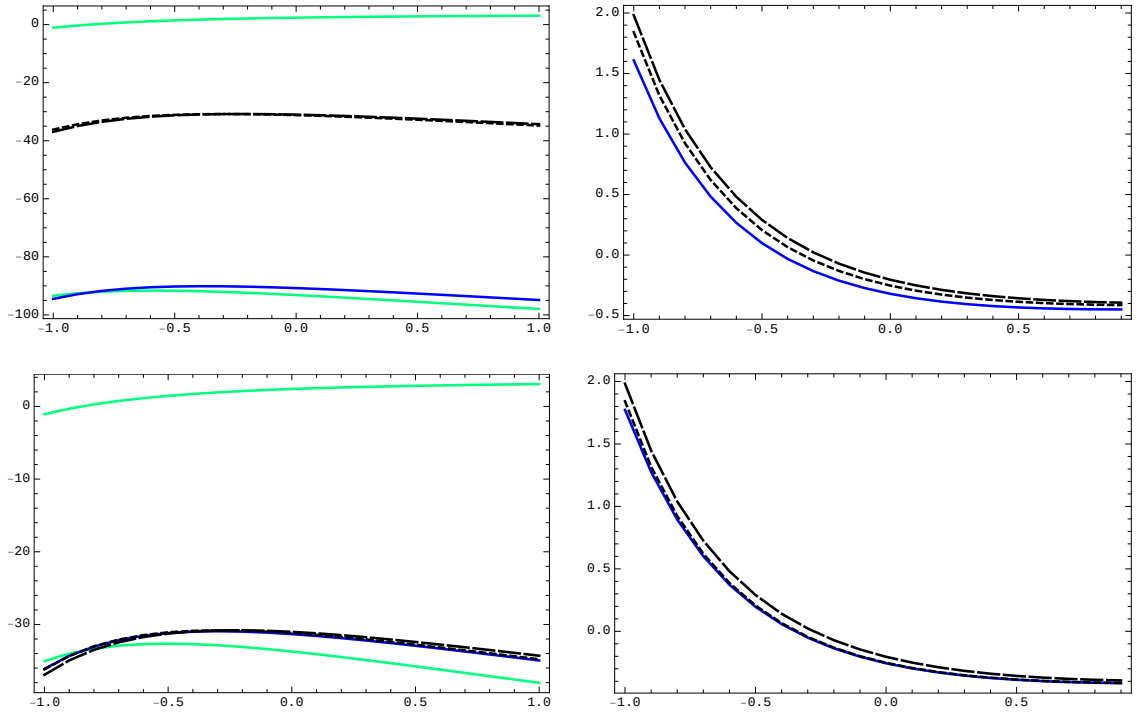

Figure 6: Likelihood surface as a function of diffusion coefficient g. Variable reference density, $\psi_0 = \{a_0, g\} = \{0, g\}$, $\psi = \{a, g\} = \{-1, g\}$. First row: simulated $p(z; \psi_0)$, $N = 100$. 2nd row: exact $p(z; \psi_0)$. Left column: $l_0 = \log p_0(z; \psi_0)$ and $l_1 = \log p_1(z; \psi, \psi_0)$ (green; see equation 54), sum of both parts (blue) and exact and linearized KF (black). Right column: score functions.

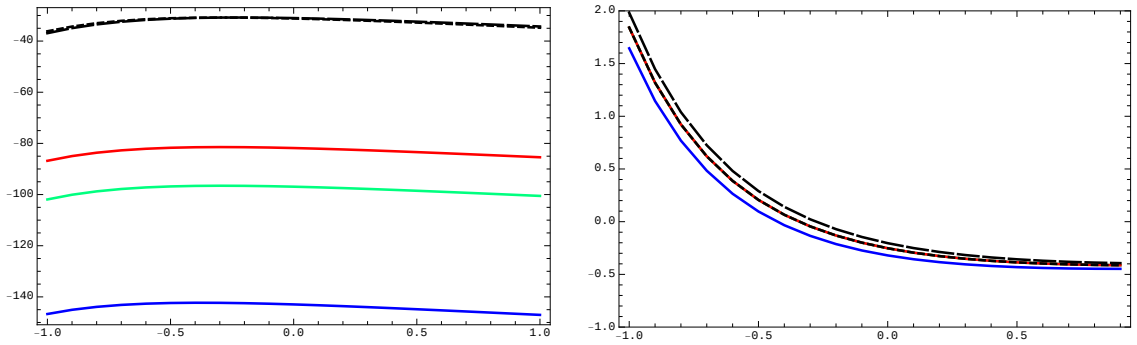

Figure 7: Likelihood surface as a function of g. Kernel density method. $N = 100$ (blue), $N = 200$ (green), $N = 500$ (red). Also shown are the exact (black, long dash) and linearized solution (black, short dash).

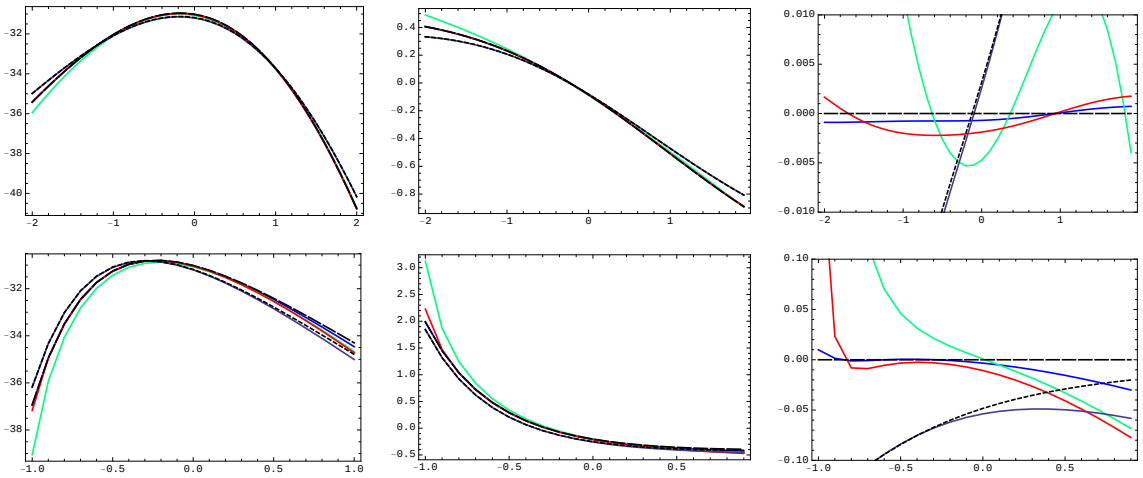

Figure 8: Likelihood surface as a function of a (first row) and g (2nd row). Numerical integration, volume element $\delta\eta = 0.5$. From left: likelihood, score function and score differences (w.r.t. the exact value = KF; black, long dashed). Displayed methods are: Euler transition kernel (ETK; dark blue), LL transition kernel (LLTK; blue), finite differences (FD; green), and integral operator (IO; red). From the right column it is seen that LLTK performs best, IO and FD are worse. The ETK is near to linearized KF (black; short dash).

Then, however, the likelihood contribution $p_0(z) = p(z; \psi_0)$ at the reference point is a function of g as well. For a practically working optimization algorithm, it must be explicitly known or simulated as well. In figure 6, the values $\psi = \{-1, g\}$, $\psi_0 = \{0, g\}$ were chosen. The left column shows the likelihood contributions $\log p_0$ (exact and simulated) and $\log p_1$ (see eqn. 54). The results are much better in comparision to the fixed reference density method.

Next, the kernel density method is displayed in fig. 7 for several sample sizes. Already for $N = 200$, using a kernel matrix, the result is nearly indistinguishable from the linearized KF. The absolute values of the likelihood contain constant simulation errors (w.r.t ψ) which attenuate with larger sample size.

For comparison, the Monte Carlo method is contrasted with numerical integration (section 4.2 and figure 8). The integration region was chosen as $[a, b] = [-4, 4]$. This choice was motivated by the stationary variance at the true parameter value which is $-(g^0)^2/(2a^0) = 2$. The volume elements were varied as $\delta\eta = 1, 0.5, 0.1, 0.05$. The likelihood surface and score for $\delta\eta = 0.5$ are already close to the exact values. Note that the mean square displacement is of order $\sqrt{\Omega\delta t} = \sqrt{4 \cdot 0.1} \approx 0.63$. To display the discrepancies in approximations, the differences of the scores in relation to the exact Kalman filter solution are displayed in the right column.

The computational costs and errors are are listed in tables 1–2. The entries correspond to figures 3, 4 and 8. The infinity norm $||d||_\infty = \max_j |d_j|, d_j = s_{approx}(\psi_j) - s_{KF}(\psi_j)$ of the difference between the approximated score and the exact

Monte Carlo integration					
sample size	100	500	1000	KFL	KF
Reference density, without kernel					
time	19.54	105.36	208.41	1.00	0.79
error/KF	0.1556	0.1018	0.1105	0.0832	0.
error/KFL	0.0724	0.06	0.0704	0.	0.0832
Reference density, with kernel					
time	22.52	108.03	228.79	1.00	0.79
error/KF	0.3517	0.3056	0.0698	0.0832	0.
error/KFL	0.288	0.2329	0.0467	0.	0.0832
Kernel density, with kernel					
sample size	100	200	500	KFL	KF
time	343.431	588.149	1536.97	1.00	0.79
error/KF	0.0964	0.0832	0.0832	0.0832	0.
error/KFL	0.0471	0.0002	0.0001	0.	0.0832

Table 1: 1-D Ornstein-Uhlenbeck process. Monte Carlo integration. Computation time and error of score functions (see text). Displayed is the maximum of absolute score differences at different parameter values a.

Numerical integration						
Method	LLTK	FD	IO	ETK	KFL	KF
$\delta\eta = 1$						
time	2.90	2.77	4.08	1.59	1.00	0.79
error/KF	0.0217	0.9269	1.1756	0.0673	0.0832	0.
error/KFL	0.0993	0.9517	1.2313	0.0175	0.	0.0832
$\delta\eta = 0.5$						
time	5.06	4.58	9.82	5.10	1.00	0.79
error/KF	0.0009	0.0846	0.0022	0.0838	0.0832	0.
error/KFL	0.0824	0.1573	0.0814	0.0009	0.	0.0832
$\delta\eta = 0.1$						
time	29.91	24.12	32.54	27.78	1.00	0.79
error/KF	0.0018	0.0058	0.0058	0.085	0.0832	0.
error/KFL	0.0814	0.0778	0.0777	0.0019	0.	0.0832

Table 2: 1-D Ornstein-Uhlenbeck process. Numerical integration. Computation time and error of score functions. Displayed is the maximum of absolute score differences at different parameter values a.

(KF) or linearized Kalman filter (KFL) score function was computed over several parameter values ψ_j, because the score is the criterion in the ML equation. Arbitrary constants in the approximated likelihoods are thus eliminated. From table 2 it is seen that the transition kernel method with locally linearized moment equation (LLTK) is best. The Euler transition kernel (ETK) yields a good approximation of the linearized KF criterion. For a very large discretization interval $\delta\eta = 1$, the finite difference (FD) and integral operator (IO) methods even give negative density approximations, thus complex log likelihoods and large errors. For a smaller interval, $\delta\eta = 0.5$, one gets the order LLTK, IO, ETK and FD, but for $\delta\eta = 0.1$, FD is equal to IO. The computation times are shortest for FD, then ETK, LLTK and IO. Generally, the Monte Carlo algorithm (table 1) takes much more computing power, at least in the one dimensional problem considered here. To get an absolute error < 0.1, one needs 228.79 sec in the MC approach (reference density, with kernel), but only 2.9 sec. for the LLTK approach.

A general observation is the remarkable smoothness of the simulated likelihood surfaces. Thus, optimization with gradient methods is possible. This is in contrast to particle filtering, where the functions can be quite rough (Pitt; 2002; Singer; 2003). In the reference density method, when a fixed reference parameter ψ_0 is used, only one stream of random vectors is drawn. Thus, $\hat{p}(z;\psi) = p(z;\psi_0)\hat{p}(z;\psi,\psi_0)$ inherits the smoothness properties of the densities $p(z|\eta;\psi)$ and $p(\eta;\psi)$. If diffusion parameters σ are estimated, $\psi_0 = \{\theta_0, \sigma\}$ is variable, thus several streams must be simulated. Nevertheless, the simulated likelihood is very smooth (fig. 6). The same applies to the kernel density method (fig. 7).

5.2 Ornstein-Uhlenbeck process (two dimensional)

In this section the linear oscillator

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = g\zeta(t)$$

is considered, with parameters γ = friction, $\omega_0 = 2\pi/T_0$ = angular frequency, T_0 = period of oscillation and g = strength of random force $\zeta = dW/dt$ (white noise). The true parameters are $\psi^0 = \{-(\omega_0^2)^0, -\gamma^0, g^0\} = \{-16, -4, 2\}$. The state $y(t)$ is assumed to be measured noisily at certain time points $t_i = i\Delta t, i = 0, \ldots, T; \Delta t = 0.5, T = 20$. The discretization interval was set to $\delta t = 0.1$ and $J = (t_T - t_0)/\delta t = (10 - 0)/0.1 = 100$.

In state space form the model reads

$$\begin{aligned} d\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix} d\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix} \\ z_i &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} y_1(t_i) \\ y_2(t_i) \end{bmatrix} + \epsilon_i, \end{aligned}$$

$\mathrm{Var}(\epsilon_i) = R = 0.01$. The model can be solved exactly using Kalman filter methods (Singer; 1993). It was chosen, because is is already demanding for the Monte Carlo

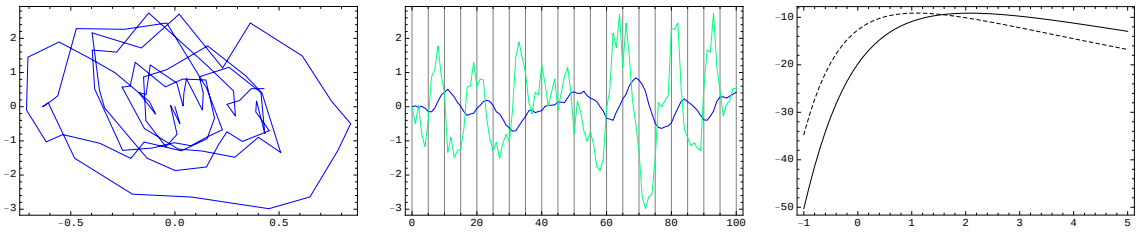

Figure 9: Stochastic oscillator. Left: trajectory in state space. Middle: Trajectory over time (index j) and measurement times (vertical lines), Right: Likelihood function as function of diffusion parameter (full: exact, dashed: linearized, abscissa: deviation from $\psi_3^0 = g^0 = 2$).

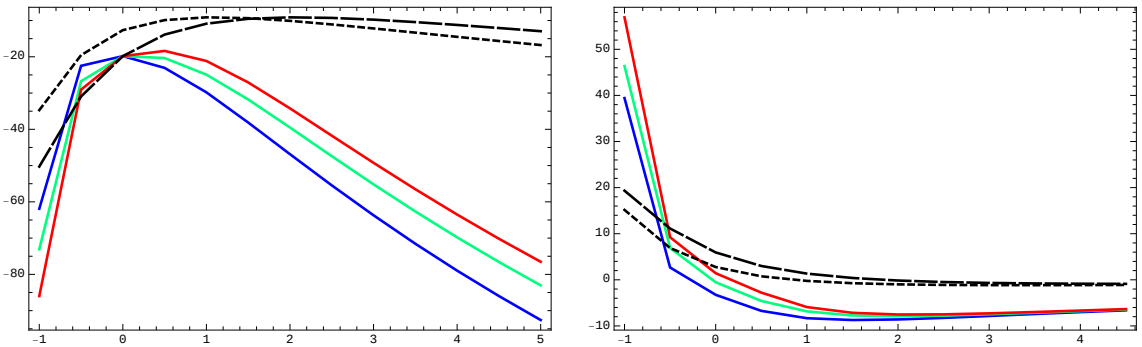

Figure 10: Stochastic oscillator. Fixed reference density, $\psi_0 = \{0, 0, 2\}$. Left: likelihood surface, right: score. Abscissa: deviation from $\psi_3^0 = g^0 = 2$. Shown are: Kalman filter (KF): exact (black, long dash), linearized (black, short dash), simulated: $N = 100, 500, 1000$ (blue, green, red).

and the numerical integration approach. Since the diffusion matrix is singular, a small parameter $\epsilon^2 = 10^{-2}$ was inserted, i.e. $\Omega = \mathrm{diag}(\epsilon^2, g^2)$. Alternatively, one could use the singular normal distribution in eqn. (5).

As in the last example, one expects problems with the variation of the diffusion parameter g. Indeed, as fig. 10 shows, a fixed reference density (using $\psi_0 = \{0, 0, 2\}$) leads to a bad approximation of the likelihood. Usage of a variable reference in the diffusion parameter, i.e. $\psi_0 = \{0, 0, g\}$, $\psi = \{-\omega_0^2, -\gamma, g\}$ gives better results (fig. 11). The density $p(z; \psi_0)$ is explicitly known in this linear example, but can be estimated by the kernel density approach (see sect. 4.1.1). Here a Gauss density $\hat{p}(\eta|z) = \phi(\eta; \bar{\eta}, S)$ was used in the denominator of (47).

The kernel density method is displayed in fig. 12. The random vectors were drawn from a scaled Langevin equation with $K = -H^{-1}$. Already a sample size of $N = 500$ gives a very accurate approximation which is close to the linearized KF solution. The absolute error of the three score curves is only $||s_n - s_{\mathrm{KFL}}|| = \{5.7915, 0.1054, 0.014\}, n = 100, 200, 500; ||f|| = \max_j |f_j|$.

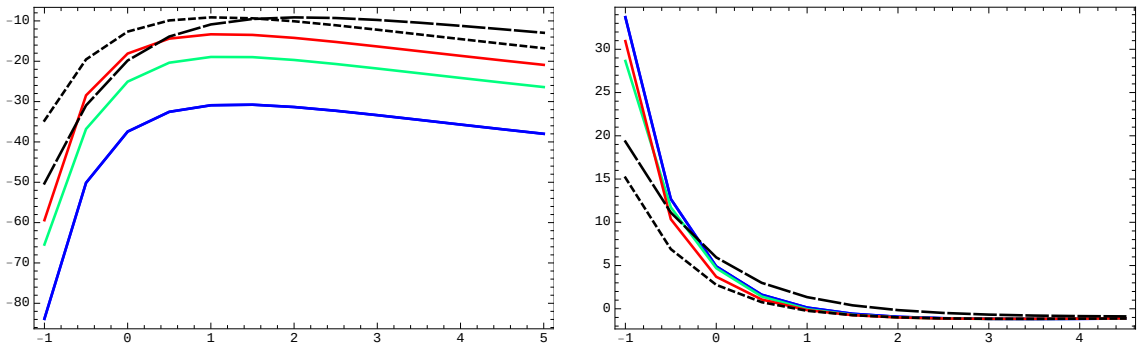

Figure 11: Stochastic oscillator. Variable reference density, $\psi_0 = \{0, 0, g\}$, $\psi = \{-16, -4, g\}$. Estimated $p(z; \psi_0)$. Left: likelihood surface, right: score. Shown are: Kalman filter (KF): exact (black, long dash), linearized (black, short dash), simulated: $N = 500, 1000, 2000$ (blue, green, red).

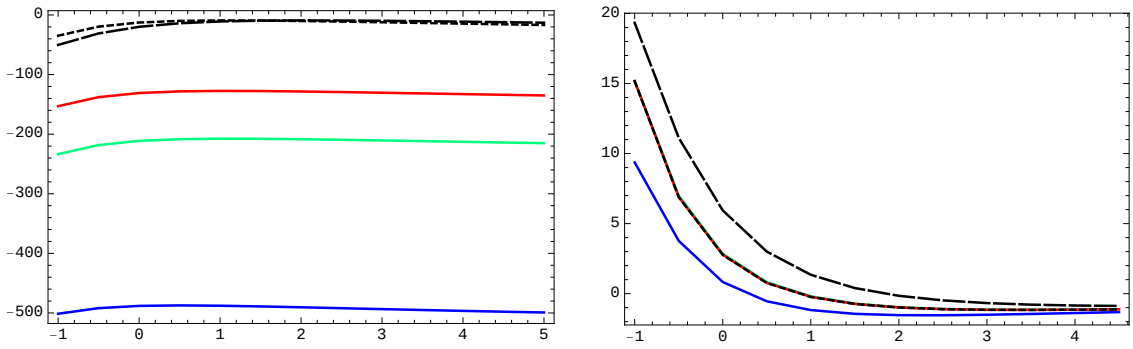

Figure 12: Stochastic oscillator. Kernel density method. Left: likelihood surface, right: score. Shown are: Kalman filter (KF): exact (black, long dash), linearized (black, short dash), simulated: $N = 100, 200, 500$ (blue, green, red). Very accurate simulation of linearized KF score (right).

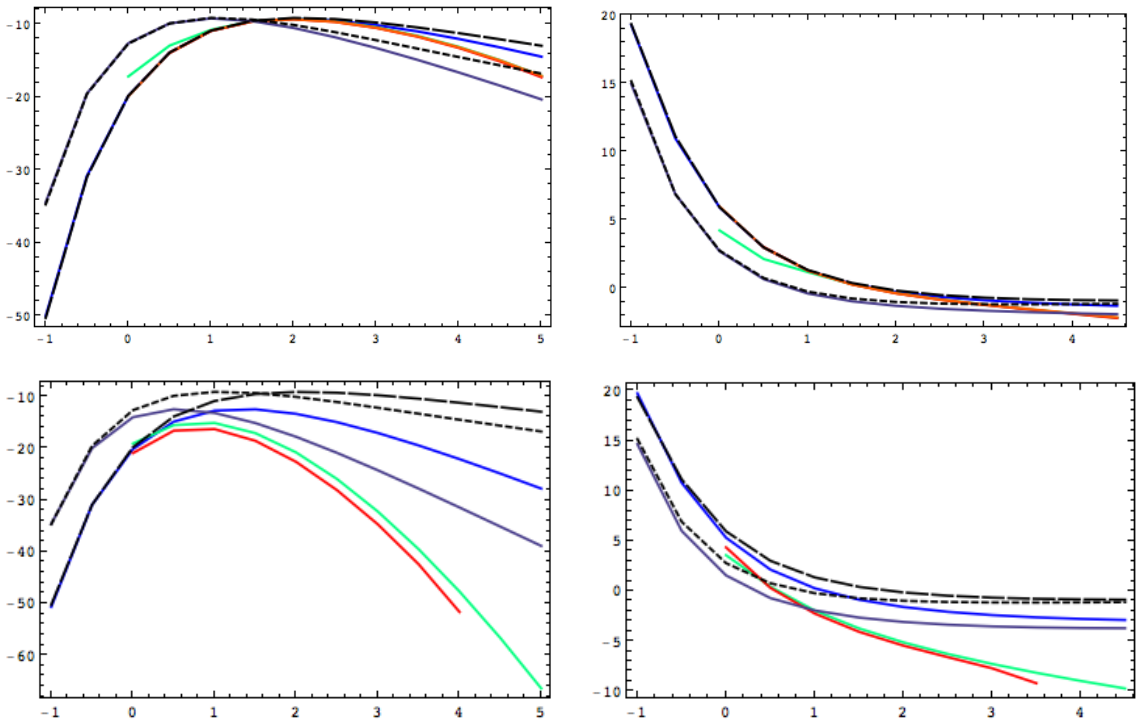

Figure 13: Stochastic oscillator. Numerical integration, spatial discretization $\delta y = \{0.1, 0.5\}$. Top: integration region $[a, b], a = \{-1, -5\}$, $b = \{1, 5\}$. Bottom: $a = \{-1, -3\}$, $b = \{1, 3\}$. Left: likelihood, right: score function. Displayed methods are: LL transition kernel (LLTK; blue), Euler transition kernel (ETK; dark blue), finite differences (FD; green), integral operator (IO; red). Missing values correspond to complex log-likelihoods (negative likelihoods).

Numerical integration							
Method	LLTK	FD	IO	IO'	ETK	KFL	KF
$a = \{-1, -5\}, b = \{1, 5\}, \delta\eta = \{0.1, 0.5\}$							
time	97.307	64.041	116.576	117.918	108.458	0.353	0.268
error/KF	0.3912	1.7004	1.287	1.2292	4.3773	4.2301	0.
error/KFL	4.1035	1.4733	2.2704	3.2272	0.7612	0.	4.2301
$a = \{-1, -3\}, b = \{1, 3\}, \delta\eta = \{0.1, 0.5\}$							
time	54.2815	33.3135	51.9409	43.86	52.6353	0.353	0.268
error/KF	1.7933	6.557	7.1006	6.6942	4.7416	4.2301	0.
error/KFL	4.0987	6.3157	6.8593	6.4528	2.304	0.	4.2301

Table 3: Stochastic oscillator. Numerical integration. Computation time and error of score functions. Displayed is the maximum of absolute score differences at different parameter values g.

Finally, the numerical integration approach is reported (fig. 13, tab. 3). In the picture, a spatial discretization of $\delta y = \{0.1, 0.5\}$ is used in rectangles $[a, b], a = \{-1, -5\}$, $b = \{1, 5\}$ (top) and $a = \{-1, -3\}$, $b = \{1, 3\}$ (bottom). This corresponds to $21*21 = 441$ and $21*13 = 273$ supporting points, respectively. Thus the transition kernel contains already $441^2 = 194481$ (top) or $273^2 = 74529$ (bottom) matrix elements. The interval $[-3, 3]$ in the second example is too small, leading to strong errors for large g (fig. 13, bottom). The approximation error of the methods is again evaluated by computing the absolute score differences (w.r.t. the exact KF solution) and taking the maximum modulus (tab. 3). The transition kernel methods (LLTK, ETK) perform best, then comes the integral operator method (IO) and finally finite differences (FD). The latter two methods in some cases lead to negative likelihood contributions. In the table, two variants of the IO method with different δ-sequences are listed. IO uses a series of Hermite functions (Wei et al.; 1997), the other (IO') uses a truncated Fourier transform $\delta_n(y-y') = \int_{-n}^{n} \exp(2\pi i(y-y')k)dk$ which seems slighty better. Moreover, it does not depend on involved numerical experiments.

The computing times from the figures are (in seconds): Numerical integration: LLTK = 97.3072, ETK = 108.458, FD = 64.041, IO = 116.576. Monte Carlo: KD = 495.065, variable reference density = 494.933. Thus, the performance of the simulation method is catching up in the 2-D system.

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 was used for the computation of a variance reduced MC estimator for 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 methods. Alternatively, a reference density with suboptimal properties, but known analytical form was used. Methods based on transition kernels and the Fokker-Planck equation gave good results, but seem to be restricted to low dimensional state spaces. In further work, improved estimation methods for the importance density will be developed. For example, one can estimate modified drift and diffusion functions and use kernel methods incorporating conditioning on the data and the conditional Markov property (Singer; 2016).

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.

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.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations, *Stochastics* **19**(4): 263–284.

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.

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.

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.

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.

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.

Mazzoni, T. (2012). Fast continuous-discrete daf-filters, *Journal of Time Series Analysis* **33**(2): 193–210.

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.

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.

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

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

Singer, H. (2016). Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling, *Diskussionsbeiträge Fakultät Wirtschaftswissenschaft*, 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.

Talay, D. (1990). Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, *Stochastics* **29, 1**: 13–36.

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