SDE
A program package for the simulation, optimal filtering and maximum likelihood estimation of nonlinear Stochastic Differential Equations

Hermann Singer
FernUniversität in Hagen, Hermann.Singer@FernUni-Hagen.de

Abstract

Continuous time models with sampled data possess several advantages over conventional time series and panel models (special issue 62:1, 2008, of Statistica Neerlandica). For example, data with unequal time intervals can be treated efficiently, since the dynamic model parameters of the system model are not affected by the measurement process. In the linear case, the nonlinear parameter restrictions of the sampled model can be implemented with specialized Kalman filter software (e.g. LSDE) or with structural equations models (SEM) allowing such nonlinear parameter restrictions. In the nonlinear case, most filtering algorithms are formulated in discrete time, but mixed continuous-discrete approaches are also scattered in the literature. The Mathematica program SDE is a collection of algorithms with consequent focus on the mixed continuous-discrete case: continuous time updates combined with discrete time measurement updates. Included are the classical methods of extended Kalman filtering and higher order nonlinear filters, but also new developments such as the unscented Kalman filter (UKF) and the Gauss-Hermite filter (GHF) using approximations of the filter density. We also use the Edgeworth-Hermite expansion of probability densities to obtain generalized Gauss filters (GGHF) utilizing higher order moments such as skewness and kurtosis.

Key Words:
Continuous-discrete state space models;
Kalman filtering;
EKF, UKF and GHF

1 Overview

1.1 Linear stochastic differential equations (LSDE)

Differential equations are the classical methods of modeling change in the natural sciences. The most simple model is the linear **growth model**

$$dY(t)/dt = AY(t) \tag{1}$$

with solution

$$Y(t) = \exp[A(t - t_0)]Y(t_0). \tag{2}$$

In applications one must model specification error using stochastic equation errors to obtain the **stochastic differential equation** (SDE)

$$dY(t)/dt = AY(t) + G\zeta(t) \tag{3}$$

where $\zeta(t)$ is a zero mean **Gaussian white noise** process with autocorrelation function $\gamma(t - s) = E[\zeta(t)\zeta(s)] = \delta(t - s)$ (Dirac delta function). We obtain the **solution**

$$Y(t) = \exp[A(t - t_0)]Y(t_0) + \int_{t_0}^{t} \exp(A(t - s))G\,\zeta(s)ds. \tag{4}$$

In symbolic notation (**Itô calculus**) one can write:

$$dY(t) = AY(t)dt + GdW(t) \tag{5}$$

$$Y(t) = \exp[A(t - t_0)]Y(t_0) + \int_{t_0}^{t} \exp[A(t - s)]GdW(s). \tag{6}$$

Using the times of measurement one obtains the **exact discrete model (EDM)** of Bergstrom (1976*b*, 1988)

$$Y_{i+1} = \exp[A(t_{i+1} - t_i)]Y_i + \int_{t_i}^{t_{i+1}} \exp[A(t_{i+1} - s)]GdW(s), \tag{7}$$

$$Y_{i+1} = \Phi(t_{i+1}, t_i)Y_i + u_i, \tag{8}$$

where Φ is the fundamental matrix of the system and $Y_i := Y(t_i)$ are the sampled measurements.
The parameters of the EDM are highly nonlinearly restricted, e.g.

$$\mathrm{Var}(u_i) = \int \Phi(t_{i+1}, s)^2 g^2 ds \tag{9}$$

in the scalar case. This is the main problem for the task of parameter estimation.
Software must be able to **implement the required nonlinear restrictions**, especially in the multivariate case where (time ordered) matrix exponentials are involved (some references are Phillips, 1976, Jones, 1984, Hamerle et al., 1991, 1993, Singer, 1998).

1.2 Advantages of differential equations

(Möbus & Nagl 1983)

1. The model specification of the system dynamics is independent of the measurement scheme and is given at the process level of the phenomenon (**micro causality** in the infinitesimal time interval dt).
2. The design of the study is specified by a measurement model, independently of the systems dynamics.
3. Changes of the variables can occur at any time, at and between the measurements. The state is defined for any time point, even if it can't be measured.
4. Extrapolation and interpolation of the data points can be obtained for arbitrary times and is not constrained to the sampling interval of the panel waves.
5. Studies with different or irregular sampling intervals can be compared since the continuous time structural parameters of the system do not depend on the measurement intervals.
6. Data sets with different sampling intervals can be analyzed together as one vector series.
7. Irregular sampling and missing data are treated in a unified framework. The parametrization is parsimonious since only the fundamental continuous time parameters must be estimated.
8. Cumulated or integrated data (flow data) can be represented explicitly.
9. Nonlinear transformations of data and variables can be handled by a differential calculus (Itô calculus).

2 Model specification and interpretation

2.1 Linear continuous/discrete state space model

(Jazwinski 1970)

$$dY(t) = [A(t,\psi)Y(t) + b(t,\psi)]dt + G(t,\psi)dW(t) \quad (10)$$

$$Z_i = H(t_i,\psi)Y(t_i) + d(t_i,\psi) + \epsilon_i \quad (11)$$

measurement times $t_i, i = 0, ..., T$.
Why use models with time-varying matrices $A(t), G(t), H(t), ...$**?**

- development psychology: children get older in the course of a longitudinal study, so the causal effects $A(t)$ are time dependent.
- factor structure $H(t)$ of a depression questionaire may be time dependent due to the psychological state of the subjects.

2.2 Exact discrete model (EDM)

From the SDE one obtains the EDM (Bergstrom 1976*b*, 1988)

$$\begin{aligned} Y(t_{i+1}) &= \Phi(t_{i+1},t_i)Y(t_i) + \\ &+ \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)b(s)ds + \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)G(s)dW(s) \end{aligned} \tag{12}$$

The EDM contains the **parameter functionals** (Arnold 1974)

$$A_i^* := \Phi(t_{i+1},t_i) \tag{13}$$

$$b_i^* := \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)b(s)ds \tag{14}$$

$$\Omega_i^* := \mathrm{Var}(u_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)G(s)G'(s)\Phi'(t_{i+1},s)ds. \tag{15}$$

with **state transition matrix**

$$\tfrac{d}{dt}\Phi(t,t_i) = A(t)\Phi(t,t_i);\ \Phi(t_i,t_i) = I. \tag{16}$$

In the time invariant and uniform sampling case one obtains the **matrix exponential function**

$$\Phi(t_{i+1},t_i) := A^* = \exp(A\Delta t) = \sum_{j=0}^{\infty}(A\Delta t)^j/j!. \tag{17}$$

It is defined as a Taylor series of the fundamental interaction matrix A. For example, the second order contribution between variables Y_k and Y_m

$$[(A\Delta t)^2]_{km} = \sum_l A_{kl}A_{lm}\Delta t^2 \tag{18}$$

involves all intermediate variables Y_l.
Alternatively, one obtains the **product representation**

$$\exp(A\Delta t) = \lim_{J\to\infty}\prod_{j=0}^{J}(I + A\Delta t/J). \tag{19}$$

2.3 Example 1: Product representation for the sampling interval $\Delta t = 2$ using $J = 2$ intermediate steps

$$A = \begin{bmatrix} -0.3 & 0 & 1 \\ 0 & -0.5 & 0.6 \\ -2 & -2 & 0 \end{bmatrix}; \Delta t = 2 \tag{20}$$

$$\lambda(A) = \{-0.18688 + 1.77645i, -0.18688 - 1.77645i, -0.42624\}$$

$$\exp[A\Delta t] = \begin{bmatrix} -0.242254 & -0.634933 & -0.131455 \\ -0.38096 & 0.0697566 & -0.116969 \\ 0.262911 & 0.389897 & -0.66265 \end{bmatrix} \tag{21}$$

There is no direct interaction between variables Y_1 and Y_2, i.e. $A_{12} = 0 = A_{21}$.

$$\exp(A\Delta t) \approx (I + A\Delta t/2)^2 = I + A\Delta t + A^2\Delta t^2/4 \tag{22}$$

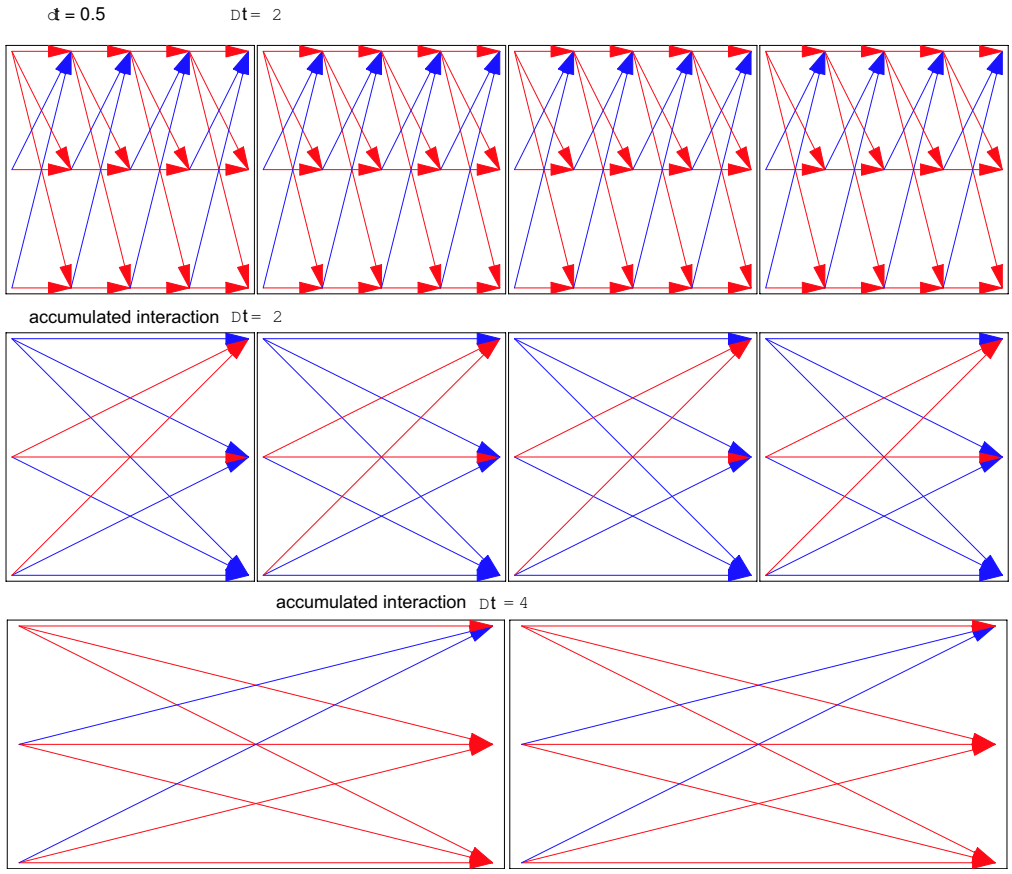

Figure 1: 3 variable model: Product representation of interactions within the measurement interval $\Delta t = 2$. Discretization interval $\delta t = 2/4 = 0.5$. Positive causal actions = red; Negative causal actions = blue.

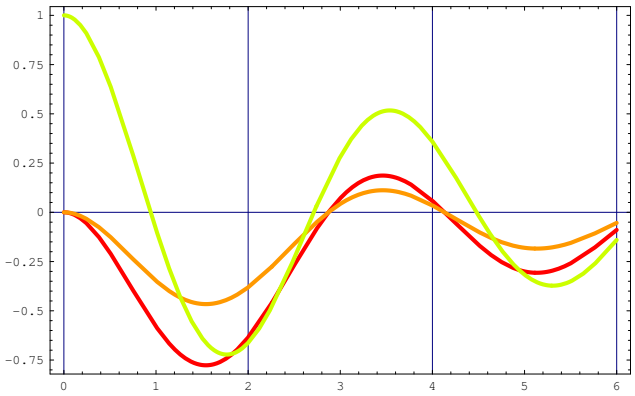

Figure 2: 3 variable model: Exact discrete matrix $A^* = \exp(A\Delta t)$ as a function of measurement interval Δt. Matrix elements A^*_{12}, A^*_{21}, A^*_{33}. Note that the discrete time coefficients change their strength and even sign.

Second order terms:

$$[\exp(A\Delta t)]_{12} \approx A_{13}A_{32}\Delta t^2/4 \tag{23}$$

- Indirect interactions are mediated through the third variable, which appear at the finite sampling interval.
- If the signs are different, one gets positive and negative contributions and the overall sign is dependent on the sampling interval.

2.4 Example 2: Linear oscillator; CAR(2)

The linear oscillator (synonyms: pendulum, swing) is the continuous time AR(2) model which can model phenomena oscillating in time. The parameters are γ = friction = 4, $\omega_0 = 2\pi/T_o = 4$ = angular frequency, T_o = period of oscillation. The second order SDE can be written in state spece form

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = bx(t) + g\zeta(t) \tag{24}$$

$$d\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} := \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}dt + \begin{bmatrix} 0 \\ b \end{bmatrix}dt + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix}d\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix} \tag{25}$$

$$z_i := \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \end{bmatrix} + \epsilon_i \tag{26}$$

and estimated with the linear state space model. The model specification in Mathematica reads

```
PAR//Clear;
(*************************************************************)
PAR[{t1_,t2_,t3_,t4_,t5_,t6_}] := Module[{OMEGA,A,B,SIGMA,MUE},
(*************************************************************)
A={{0,  1},
   {t1,t2}};
B= {{ 0},
    {t3}};
G={{.0001,0 },
   {0    ,t4}};
OMEGA=G.Transpose[G];
MUE= {t5,t6};
SIGMA={{1., 0},
       {0, 1.}};

{OMEGA,A,B,SIGMA,MUE}
]

PARD//Clear;
(*************************************************************)
PARD[{t1_,t2_,t3_,t4_,t5_,t6_}] =
(*************************************************************)
        Module[{t},
        t = {t1,t2,t3,t4,t5,t6};
        Map[DMS[#,t]&,PAR[t]]//N
]

PARM//Clear
(*************************************************************)
PARM[{phi1_,phi2_}]:= Module[{R,H,D},
```

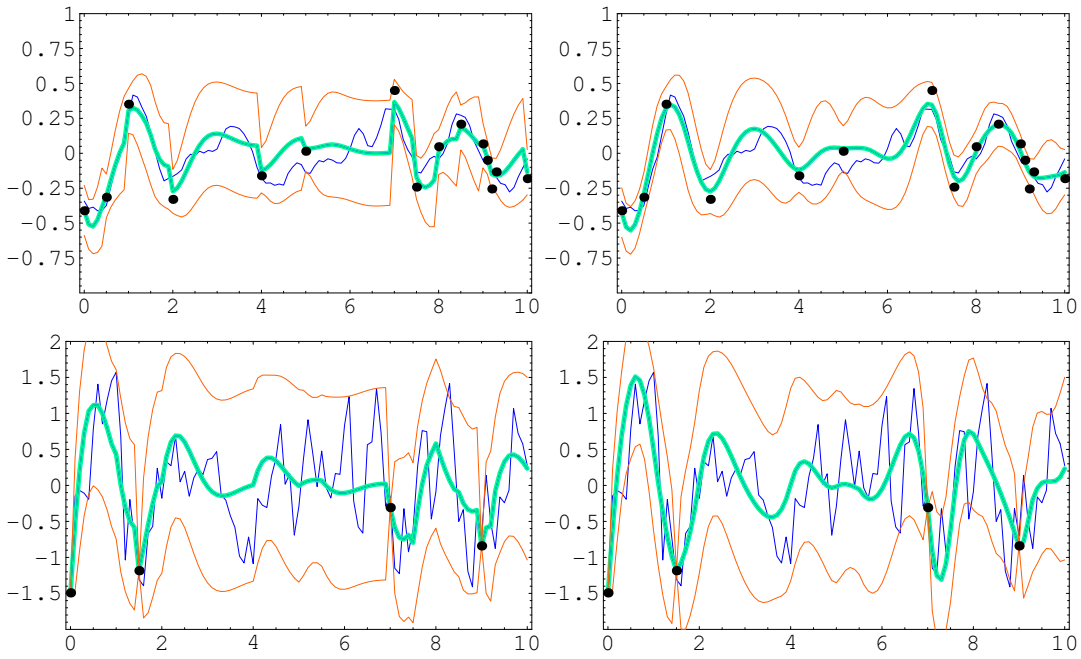

Figure 3: Linear oscillator with irregularly measured states (dots): Filtered state (left), smoothed state (right) with 95%-HPD confidence intervals. Measurements at $\tau_1 = \{0, .5, 1, 2, 4, 5, 7, 7.5, 8, 8.5, 9, 9.1, 9.2, 9.3, 10\}$ (first component; 1 st line), $\tau_2 = \{0, 1.5, 7, 9\}$ (2 nd component, 2 nd line). Discretization interval $\delta t = 0.1$. The controls $x(t)$ were measured at $\tau_3 = \{0, 1.5, 5.5, 9, 10\}$.

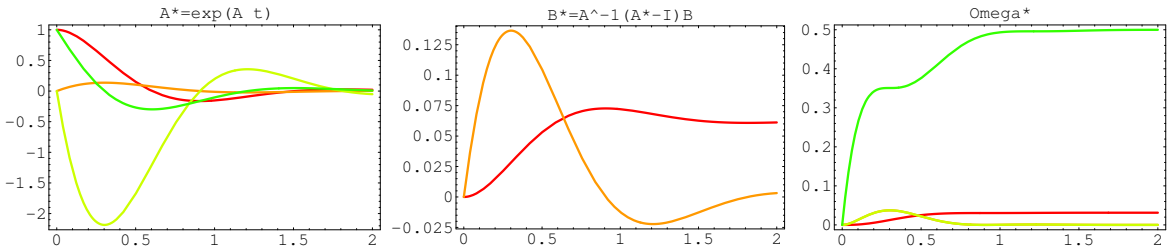

Figure 4: Linear oscillator: Exact discrete matrices $A^* = \exp(A\Delta t)$, $B^* = A^{-1}(A^* - I)B$, $\Omega^* = \int_0^{\Delta t} \exp(As)\Omega \exp(A's)ds$ as a function of measurement interval. Note that the discrete time coefficients change their strength and even sign.

```
(*************************************************************)
H={{1,0},
   {0,1}};
D= {{0},
    {0}};
R={{(phi1//Exp),     0      },
   {   0       , (phi2//Exp)}};

{R,H,D}
]

PARMD//Clear
(*************************************************************)
PARMD[{phi1_,phi2_}] =
(*************************************************************)
Module[{t},
        t = {phi1,phi2};
        Map[DMS[#,t]&,PARM[t]]
]

SeedRandom[13579];
{Z,Y} = StateSpace1[H,d,R,A,B,G,Y0,X,K,P,Q,n,T,DT,seed,UniC];

(*selection from Z*)
ZM = Array[MISS&,{T+1,K,n}];
tau1 = {0,.5,1,2,4,5,7,7.5,8,8.5,9,9.1,9.2,9.3,10}/DT + 1;
Do[ZM[[tau1[[i]],1]] = Z[[tau1[[i]],1]], {i,tau1//Length} ];
tau2 = {0,1.5,7,9}/DT + 1;
Do[ZM[[tau2[[i]],2]] = Z[[tau2[[i]],2]], {i,tau2//Length} ];

THETA = { -16, -4, 1, 2, 0, 0} //N;
PHI = {-2,-2} //N;
NONLIN = 1;
EPS = 10.^-4;
KMAX = 40;
F0 = IdentityMatrix[(THETA//Length)+(PHI//Length)];
DT = .1;
PRINT = {1,1,1,1,1,1,1,1};
Fisher = True;

K=2;Q=1;P=2;
T=100;n=10;
np = 1; (*same sampling scheme for each panel unit*)
T1=T+1;
DT=.1;

{THETA,PHI,COV,likpsi,k} =

MBFGS[THETA,PHI,F0,KMAX,EPS,ZM,Xint,K,P,Q,n,np,T,DT,1,NONLIN,MISS,PRINT];
```

2.5 Discussion

- Researchers using different sampling intervals will have a dispute over the strength and even the sign of the causal relation, but only if they use a discrete time model without the deeper structure of the continuous time approach.
- In contrast, a continuous time approach estimates the parameters related to the interval dt irrespective of the measurement intervals $\Delta t_1, \Delta t_2, ...$ of different studies or irregular intervals in one study.
- Thus sampling can be completely irregular for each panel unit and within the variables.
- One always points to the same fundamental level of the theory.

3 Estimation

3.1 General and historical remarks

- The exact discrete model is a vector autoregression with special restrictions which must be incorporated into the estimation procedure.
- Otherwise serious embeddability and identification problems arise (Phillips 1976a, Singer 1992d, Hamerle et al. 1991, 1993)
- For small sampling interval, the EDM may be linearized and time series or SEM software can be used. This was already proposed by Bergstrom in the 19sixties (rectangle or trapezium approximation; Bergstrom (1976*a*)).
- Later, there were attempts to estimate a reparametrized version of the EDM and to infer the continuous time parameters indirectly.
- This yields serious problems since the restrictions of $A, B, ..$ (see above) cannot be implemented and second, if no restrictions are imposed, embeddability and identification problems arise.
- Therefore, the likelihood function $p(Z_T, ...Z_0; \psi)$ must be expressed in terms of the EDM-matrices

$$
\begin{aligned}
A_i^* &= \exp(A\Delta t_i) && (27)\\
B_i^* &= A^{-1}(A_i^* - I)B && (28)\\
\Omega_i^* &= \int_0^{\Delta t_i} \exp(As)\Omega \exp(A's)ds && (29)
\end{aligned}
$$

and $A = A(\psi), B = B(\psi), ...$

3.2 Exact estimation methods

For ML estimation, we distinguish two methods of computing the likelihood:

1. Recursively by using the Kalman filter
2. Non-recursively by using simultaneous equations with **nonlinear** parameter restrictions (e.g. matrix exponential functions)

(for details, see Singer 2007, Oud & Singer 2008, Singer 2008)

Exact Discrete Model

(panel index $n = 1, ...N; i = 0, ...T$)

$$
\begin{aligned}
Y_{i+1,n} &= A_{in}^* Y_{in} + b_{in}^* + u_{in} && (30)\\
Z_{in} &= H_{in} Y_{in} + d_{in} + \epsilon_{in} && (31)
\end{aligned}
$$

$$A^*_{in} := \Phi_n(t_{i+1}, t_i) = \overleftarrow{T} \exp[\int_{t_i}^{t_{i+1}} A(s, x_n(s))ds] \quad (32)$$

$$b^*_{in} := \int_{t_i}^{t_{i+1}} \Phi_n(t_{i+1}, s)b(s, x_n(s))ds \quad (33)$$

$$\mathrm{Var}(u_{in}) := \Omega^*_{in} = \int_{t_i}^{t_{i+1}} \Phi_n(t_{i+1}, s)GG'(s, x_n(s))\Phi'_n(t_{i+1}, s)ds. \quad (34)$$

- Matrices are noncommutative, i.e $A(t)A(s) \neq A(s)A(t)$
- $\overleftarrow{T}\, A(t)A(s) = A(s)A(t); t < s$
  is the Wick time ordering operator (Abrikosov et al. 1963).

3.3 Kalman filter approach

likelihood (prediction error decomposition)

(panel index n is dropped)

$$l(\psi; Z) = \log p(Z_T, \ldots, Z_0; \psi) = \sum_{i=0}^{T-1} \log p(Z_{i+1}|Z^i; \psi)p(Z_0), \quad (35)$$

- $p(Z_{i+1}|Z^i; \psi) = \phi(\nu(t_{i+1}|t_i); 0, \Gamma(t_{i+1}|t_i))$
  transition densities (Gauss distributions)
- $\nu(t_{i+1}|t_i)$ is the prediction error
  (measurement minus prediction using information
  $Z^i := \{Z_i, \ldots, Z_0\}$ up to time t_i)
  Γ is the prediction error covariance matrix.
- The computation runs in a sequence of prediction and correction steps (time and measurement update).
- It was first discovered in an engineering context by (Kalman 1960) and is now well known in many fields.
- An implementation for panel data is LSDE (Singer 1991, 1993, 1995)

Kalman filter algorithm

(Liptser & Shiryayev 2001, Harvey & Stock 1985, Singer 1998)

conditional moments

- $\mu(t|t_i) = E[Y(t)|Z^i]$
- $\Sigma(t|t_i) = \mathrm{Var}[Y(t)|Z^i]$,
- $Z^i = \{Z_i, ..., Z_0\}$ are the measurements up to time t_i.

time update

$$
\begin{aligned}
(d/dt)\mu(t|t_i) &= A(t,\psi)\mu(t|t_i) + b(t,\psi) && (36)\\
(d/dt)\Sigma(t|t_i) &= A(t,\psi)\Sigma(t|t_i) + \Sigma(t|t_i)A'(t,\psi) + \Omega(t,\psi) && (37)
\end{aligned}
$$

measurement update

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + K(t_{i+1}|t_i)\nu(t_{i+1}|t_i) && (38)\\
\Sigma(t_{i+1}|t_{i+1}) &= [I - K(t_{i+1}|t_i)H(t_{i+1})]\Sigma(t_{i+1}|t_i) && (39)\\
\nu(t_{i+1}|t_i) &= Z_{i+1} - Z(t_{i+1}|t_i) && (40)\\
Z(t_{i+1}|t_i) &= H(t_{i+1})\mu(t_{i+1}|t_i) + d(t_{i+1}) && (41)\\
\Gamma(t_{i+1}|t_i) &= H(t_{i+1})\Sigma(t_{i+1}|t_i)H'(t_{i+1}) + R(t_{i+1}) && (42)\\
K(t_{i+1}|t_i) &:= \Sigma(t_{i+1}|t_i)H'(t_{i+1})\Gamma(t_{i+1}|t_i)^{-1} && (43)
\end{aligned}
$$

- $K(t_{i+1}|t_i)$ is the **Kalman gain**,
- $Z(t_{i+1}|t_i)$ is the **optimal predictor** of the measurement Z_{i+1},
- $\nu(t_{i+1}|t_i)$ is the **prediction error**
- $\Gamma(t_{i+1}|t_i)$ is the **prediction error covariance matrix**.

3.4 SEM approach

In the SEM-EDM approach (Oud & Jansen 2000, Oud et al. 1993) the exact discrete model is represented by the SEM model ($\eta_n' = [Y_{0n}', ..., Y_{Tn}']$)

$$
\begin{aligned}
\eta_n &= B\eta_n + \Gamma X_n + \zeta_n && (44)\\
Y_n &= \Lambda\eta_n + \tau X_n + \epsilon_n && (45)
\end{aligned}
$$

(stochastic ξ are absorbed in η) with structural matrices

$$
B = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ A_0^* & 0 & 0 & \dots & 0 \\ 0 & A_1^* & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & A_{T-1}^* & 0 \end{bmatrix} : (T+1)p \times (T+1)p \quad (46)
$$

$$
X_n = \begin{bmatrix} 1 \\ x_{n0} \\ x_{n1} \\ \vdots \\ x_{nT} \end{bmatrix} : (T+1)q + 1 \times 1 \quad (47)
$$

$$
\Gamma = \begin{bmatrix} \mu & 0 & 0 & \dots & 0 & 0 \\ 0 & B_0^* & 0 & \dots & 0 & 0 \\ 0 & 0 & B_1^* & \dots & 0 & 0 \\ \vdots & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & B_{T-1}^* & 0 \end{bmatrix} : (T+1)p \times (T+1)q + 1 \quad (48)
$$

$$b^*_{ni} = [\int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s)B(s, \psi)ds]x_{ni} := B^*_i x_{ni}. \quad (49)$$

Solving for η_n, one obtains the solution

$$\eta_n = (I - B)^{-1}(\Gamma X_n + \zeta_n) \quad (50)$$

$$(I - B)^{-1} = \sum_{j=0}^{\infty} B^j = \begin{bmatrix} I & 0 & 0 & \dots & 0 \\ A^*_0 & I & 0 & \dots & 0 \\ A^*_1 A^*_0 & A^*_1 & I & \dots & 0 \\ \vdots & 0 & \ddots & I & 0 \\ A^*_{T-1}...A^*_0 & A^*_{T-1}...A^*_1 & \dots & A^*_{T-1} & I \end{bmatrix}. \quad (51)$$

Thus, all $y_n(t_i)$ are expressed in terms of the initial condition $y_n(t_0) \sim N(\mu, \Sigma)$ (for details, see Singer 2007).

Likelihood function

$$l = -\tfrac{N}{2}(\log|\Sigma_y| + \mathrm{tr}[\Sigma_y^{-1}(M_y + CM_xC' - M_{yx}C' - CM_{xy})]) \quad (52)$$

$$E[Y_n] = [\Lambda(I - B)^{-1}\Gamma + \tau]X_n := CX_n \quad (53)$$

$$\Sigma_y = \mathrm{Var}(Y_n) = \Lambda(I - B)^{-1}\Sigma_\zeta(I - B)^{-T}\Lambda' + \Sigma_\epsilon. \quad (54)$$

Moment matrices

$$M_y = Y'Y : (T + 1)k \times (T + 1)k \quad (55)$$

$$M_x = X'X : (T + 1)q + 1 \times (T + 1)q + 1 \quad (56)$$

$$M_{yx} = Y'X : (T + 1)k \times (T + 1)q + 1 \quad (57)$$

$$Y' = [Y_1, ...Y_N] : (T + 1)k \times N \quad (58)$$

$$X' = [X_1, ...X_N] : (T + 1)q + 1 \times N \quad (59)$$

- $\eta'_n = [Y'_{0n}, ..., Y'_{Tn}]$ is the sampled trajectory (for panel unit n)
- ΓX_n is a deterministic intercept term.
- It is essential, that the SEM (and KF) software permits the **nonlinear parameter restrictions of the EDM** (27–29).
- The SEM (44–45) with **arbitrary nonlinear parameter restrictions** was implemented as the Mathematica program SEM (Singer 2004d)

In Mathematica, it is very simple to compute a matrix exponential (`MatrixExp`). The EDM matrices of an evenly spaced AR(2) model with sampling interval `Deltat` are given as

```
STERN[Theta_, Deltat_] :=
Module[{L, p, a, b, omega, Ip, Ip2, astern, bstern, ostern},
        a = A[Theta];
        b = B[Theta];
        omega = Omega[Theta];
        p = Length[a];
```

```
        Ip  = IdentityMatrix[p];
        Ip2 = IdentityMatrix[p^2];
        astern = MatrixExp[a  Deltat];
        bstern =  Inverse[a].(astern - Ip).b;
        L = KroneckerProduct[a, Ip] + KroneckerProduct[Ip, a];
        ostern = Inverse[L].(KroneckerProduct[astern, astern] - Ip2).Flatten[omega];
        ostern = Shape[ostern, {p, p}];
        {astern, bstern, ostern}]
```

These coefficients are inserted into the module `SystemEDMdet` specifying the structural matrices:

```
SystemEDMdet[Theta_, {Deltat_, T_}] :=
       Module[{astern, bstern, ostern, p, q, beta, gamma, sigma, sigmazeta,
                       lambdaeta, taueta, sigmaepsilon},

       {astern, bstern, ostern} = STERN[Theta, Deltat];
       {p, q} = Dimensions[bstern];
       null = Shape[0, {p, p}];
       beta = SubDiagonalMatrix[astern, {T + 1, p}];
       mu = Mue[Theta];
       gamma = Spalte[Flatten[{1, Shape[2, T]}]];
       gamma = gamma /. {1 -> Spalte[mu], 2 -> bstern};
       gamma = BlockMatrix[gamma]; (* [mu',bstern',....,bstern']' *)
       sigma = Sigma[Theta];
       sigmazeta = Flatten[{1, Shape[2, T]}];
       sigmazeta = sigmazeta /. {1 -> sigma, 2 -> ostern};
       sigmazeta = BlockDiagonalMatrix[sigmazeta]; (* Diag(sigma,ostern,...,ostern) *)
       h = H[Theta];
       lambdaeta = KroneckerProduct[IdentityMatrix[T + 1], h];
       {kp, p} = Dimensions[lambdaeta];
       {p, q} = Dimensions[gamma];
       taueta = Shape[0, {kp, q}];
       sigmaepsilon = IdentityMatrix[kp]/10.^8;

       {{beta, gamma, sigmazeta}, {lambdaeta, taueta, sigmaepsilon}}]
```

This module is plugged into the ML routine `BFGS`

```
{THETA, COV, LIKI, psii, scoreii, k, ReturnCode} =
BFGS[{{S, m, n, Deterministisch}},
     {THETA0, F0},
     {{SystemEDMdet, {Deltat, T}},
     {Lik, LISRELlik, Score, Hesse}},
     {{Sing = False, 1}, KMAX = 300, {EPS1 = .0001, EPS2 = .0001},
       OPTION = 0, PRINT = {1, 1, 1, 1, 1, 1, 1, 1, 1}}];
```

to yield the ML estimates `THETA`.

3.5 Comparision of the approaches

1. The KF computes the likelihood recursively for the data $Z = \{Z_0, ...Z_T\}$, i.e. the conditional distributions $p(Z_{t+1}|Z^t)$ are updated step by step, whereas the SEM representation utilizes the joint distribution of the vector $\{Z_0, ...Z_T\}$.

2. Therefore, the KF can work online, since new data update the conditional moments and the likelihood, whereas the SEM uses the batch of data $Z = \{Z_0, ...Z_T\}$ with dimension $(T + 1)k$. The KF only involves the data point $Z_t : k \times 1$ and one has to invert matrices of order $k \times k$ (prediction error covariance). The SEM must invert the

matrices $\mathrm{Var}(Y) : (T+1)k \times (T+1)k$ and $B : (T+1)p \times (T+1)p$ in each likelihood computation.

This will be a serious problem if long data sets $T > 100$ are analyzed, but not for short panels.

3. The KF also works in the conditionally Gaussian case, since $p(Z_{t+1}|Z^t)$ is still Gaussian, whereas the joint distribution of $Z = \{Z_0, ...Z_T\}$ is not Gaussian any more.

4. As a consequence, the KF approach can be easily generalized to nonlinear systems (extended Kalman filter EKF), since the transition probabilites are still approximately conditionally Gaussian.

5. The SEM approach is more familiar to many scientists used to work with LISREL and other programs. In the early days of SEM modeling, only linear restrictions could be implemented, but now the system (44–45) and its likelihood can be easily programmed and maximized using matrix software like Mathematica, SAS/IML etc.

6. Filtered estimates of the latent states are computed recursively by the KF (the conditional moments), and smoothed trajectories can be computed by a (fixed interval) smoother algorithm. On the other hand, in the SEM approach, it is easy to compute the conditional expectations $E[\eta|Y]$ and $\mathrm{Var}[\eta|Y]$ yielding the smoothed estimates, but again matrices of order $(T+1)k \times (T+1)k$ are involved.

7. Missing data may be treated in both cases by modifying the measurement model. The Kalman filter processes the data $z_n(t_i) : k \times 1$ for each time point and panel unit. Thus, the missing data treatment can be automatically included in the measurement update by dropping missing entries in the matrices. In the SEM approach, the so called individual likelihood approach may be utilized.

4 Nonlinear models

In this section we discuss the *nonlinear continuous-discrete state space model* (Jazwinski, 1970, ch. 6.2) for the panel units $n = 1, ..., N$ and the unit specific measurement times t_{in}

$$dY_n(t) = f(Y_n(t), x_n(t), \psi)dt + g(Y_n(t), x_n(t), \psi)dW_n(t) \quad (60)$$

$$Z_{in} = h(Y_n(t_{in}), x_n(t_{in}), \psi) + \epsilon_{ni}; i = 0, ..., T_n. \quad (61)$$

with nonlinear drift and diffusion functions $f : \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^u \to \mathbb{R}^p$ and $g : \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^u \to \mathbb{R}^p \times \mathbb{R}^r$. $\psi \in \mathbb{R}^u$ is a u-dimensional parameter vector. The state vector $Y_n(t) \in \mathbb{R}^p$ is a continuous time random process and the $x_n(t) \in \mathbb{R}^q$ are deterministic exogenous (control) variables. As usual, stochastic controls are treated by extending the state $Y_n(t) \to \{Y_n(t), X_n(t)\}$. The dependence on $x_n(t)$ includes the nonautonomous case $x_n(t) = t$. Person specific random effects π_n can be included by extending the state according to $Y_n(t) \to \{Y_n(t), \pi_n(t)\}$ and defining the trivial dynamics $d\pi_n = 0$.
In the measurement model (2), $h : \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^u \to \mathbb{R}^k$ is a measurement function mapping the latent state $y_n(t)$ onto discrete time measurements $Z_{in}, i = 0, ..., T_n, n = 1, ..., N$. The free parameters in h may be interpreted as **nonlinear factor loadings**.

In the nonlinear case it is important to interpret the SDE (60) correctly. We use the **Itô interpretation** yielding simple moment equations (for a thorough discussion of the system theoretical aspects see Arnold (1974), Van Kampen (1981), Singer (1999)).

4.1 Exact Continuous-Discrete Filter

The exact time and measurement updates of the continuous-discrete filter are given by the recursive scheme (Jazwinski, 1970) for the conditional density $p(y,t|Z^i)$:[1]

Time update:

$$\begin{aligned} \frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i)\ ; t \in [t_i, t_{i+1}] \\ p(y,t_i|Z^i) &:= p_{i|i} \end{aligned} \tag{62}$$

Measurement update:

$$p(y_{i+1}, t_{i+1}|Z^{i+1}) = \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}, t_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} := p_{i+1|i+1} \tag{63}$$

$$p(z_{i+1}|Z^i) = \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}, t_{i+1}|Z^i)dy_{i+1}, \tag{64}$$

$i = 0, \ldots, T-1$, where F in

$$\begin{aligned} Fp &= -\sum_{k=1}^{p} \frac{\partial}{\partial y_k}[f_k(y,t,\psi)p(y,t|x,s)] \\ &\quad +\tfrac{1}{2}\sum_{k,l=1}^{p} \frac{\partial^2}{\partial y_k \partial y_l}[\Omega_{kl}(y,t,\psi)p(y,t|x,s)] \end{aligned} \tag{65}$$

is the Fokker-Planck operator, $Z^i = \{z(t_j)|t_j \leq t_i\}$ are the observations up to time t_i, $y_i := y(t_i)$ and $p(z_{i+1}|Z^i)$ is the likelihood function of observation z_{i+1}. The first equation describes the time evolution of the conditional density $p(y,t|Z^i)$ given information up to the last measurement and the measurement update is a discontinuous change due to new information using the Bayes formula. The above scheme is exact, but can be solved explicitly only for the linear case where the filter density is Gaussian with conditional moments $\mu(t|t_i) = E[y(t)|Z^i]; \Sigma(t|t_i) = \mathrm{Var}[y(t)|Z^i]$ or under conditions in the Daum filter (Daum 2005).

4.2 Exact Moment Equations

Instead of solving the time update equations for the conditional density (62), the moment equations for the first, second and higher order moments are considered. The general vector case in discussed in (Singer 2006*c*). Using the Euler approximation for the SDE (60), one obtains in a short time interval δt ($\delta W(t) := W(t+\delta t) - W(t)$)

$$y(t+\delta t) = y(t) + f(y(t),t)\delta t + g(y(t),t)\delta W(t). \tag{66}$$

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

$$\mu(t+\delta t|t_i) = \mu(t|t_i) + E[f(y(t),t)|Z^i]\delta t \tag{67}$$

[1] again dropping panel index n

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

$$\dot{\mu}(t|t_i) = E[f(y(t),t)|Z^i]. \tag{68}$$

The higher order central moments

$$m_k(t|t_i) := E[(y(t)-\mu(t|t_i))^k|Z^i] := E[M_k(t|t_i)|Z^i] \tag{69}$$

fulfil (scalar notation, dropping the condition)

$$\begin{aligned} m_k(t+\delta t) &= E[(y(t)+f(y(t),t)\delta t - \mu(t+\delta t)) + g(y(t),t)\delta W(t)]^k \\ &:= E[a+bc]^k \end{aligned} \tag{70}$$

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

$$E[a+bc]^2 = E[a^2] + E[b^2]\delta t \tag{71}$$

$$\begin{aligned} &= E[y(t)+f(y(t),t)\delta t - \mu(t+\delta t)]^2 \\ &+ E[\Omega(y(t),t)^2]\delta t. \end{aligned} \tag{72}$$

Inserting the first moment (67) and setting $a := (y-E(y)) + (f-E(f))\delta t := y^* + f^*\delta t$ one obtains

$$m_2(t+\delta t) = m_2(t) + 2E[y^* f^*]\delta t + E[f^{*2}]\delta t^2 + E[\Omega]\delta t \tag{73}$$

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

$$\begin{aligned} m_k(t+\delta t) &= m_k(t) + kE[f(y,t) * (M_{k-1}(t) - m_{k-1}(t))]\delta t \\ &+ \tfrac{k(k-1)}{2} E[M_{k-2}(t)\Omega(y,t)]\delta t + O(\delta t^2). \end{aligned} \tag{74}$$

The exact moment equations (68, 74) are not differential equations, since they depend on the unknown conditional density $p(y,t|Z^i)$. Using Taylor expansions or approximations of the conditional density one obtains several filter algorithms.

4.3 Continuous-discrete filtering scheme

Using only the first and second moment equation (68,73), and the optimal linear update (normal correlation) one obtains the recursive scheme (A^- is the generalized inverse of A)

Initial condition: $t = t_0$

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

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

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

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

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

Remarks:

1. The time update for the interval $t \in [t_i, t_{i+1}]$ was written using time slices of width δt. They must be chosen small enough to yield a good approximation for the moment equations (68, 73).

2. The measurement update is written using the theorem on normal correlation (Liptser & Shiryayev 1977, 1978, ch. 13, theorem 13.1, lemma 14.1)

$$\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \ (z_{i+1} - E[z_{i+1}|Z^i]) \qquad (75) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \ \mathrm{Cov}(z_{i+1}, y_{i+1}|Z^i). \qquad (76)
\end{aligned}$$

   Inserting the measurement equation (61) one obtains the measurement update of the filter. The formula is exact for Gaussian variables and the optimal *linear* estimate for $\mu(t_{i+1}|t_{i+1}), \Sigma(t_{i+1}|t_{i+1})$ in the nongaussian case. It is natural to use, if only two moments are considered. Despite the linearity in z_{i+1}, it still contains the measurement nonlinearities in the expectations involving $h(y,t)$. Alternatively, the Bayes formula (38) can be evaluated directly. This is necessary, if strongly nonlinear measurements are taken (e.g. the threshold mechanism for ordinal data; see sect. 7)

The approximation of the expectation values containing the unknown filter density leads to several well known algorithms:

1. Taylor expansion of f, Ω and h:

   extended Kalman filter EKF, second order nonlinear filter SNF, higher order nonlinear filter HNF$(2, L)$ (Jazwinski 1970, Singer 2006*d*)). Direct linearization in the SDE (60) using the Itô formula yields the LL approach of Shoji & Ozaki (1997, 1998); cf. Singer (2002).

2. Approximation of the filter density:

   using sigma points: unscented Kalman filter UKF (Julier and Uhlmann, 1997, Julier et al., 2000), Gaussian density using Gauss-Hermite quadrature: Gauss-Hermite filter GHF (Ito & Xiong 2000), Hermite expansion of filter density: generalized Gauss-Hermite filter GGHF (Singer 2006*b*,*c*).

5 Filter Approximations based on Taylor Expansion

5.1 Extended Kalman Filter EKF

Using Taylor expansions around the conditional mean $\mu(\tau_j|t_i)$ for the nonlinear functions in the filtering scheme, one obtains

$$\begin{aligned} E[f(y(\tau_j),\tau_j)|Z^i] &\approx f(\mu(\tau_j|t_i),\tau_j) && (77)\\ \mathrm{Cov}[f(y(\tau_j),\tau_j),y(\tau_j)|Z^i] &\approx f_y(\mu(\tau_j|t_i),\tau_j)\Sigma(\tau_j|t_i) && (78)\\ E[\Omega(y(\tau_j),\tau_j)|Z^i] &\approx \Omega(\mu(\tau_j|t_i),\tau_j). && (79)\end{aligned}$$

Expanding around $\mu(t_{i+1}|t_i)$, the measurement update is approximately

$$\begin{aligned} \mathrm{Cov}[y_{i+1},h_{i+1}|Z^i] &\approx \Sigma(t_{i+1}|t_i)h'_y(\mu(t_{i+1}|t_i),t_{i+1}) && (80)\\ \mathrm{Var}[h_{i+1}|Z^i] &\approx h_y(\mu(t_{i+1}|t_i),t_{i+1})\Sigma(t_{i+1}|t_i)h'_y(\mu(t_{i+1}|t_i),t_{i+1}) && (81)\\ E[h_{i+1}|Z^i] &\approx h(\mu(t_{i+1}|t_i),t_{i+1}). && (82)\end{aligned}$$

5.2 Second Order Nonlinear Filter SNF

Expanding up to second order one obtains (using short notation and dropping third moments)

$$\begin{aligned} E[f(y(\tau_j),\tau_j)|Z^i] &\approx f(\mu(\tau_j|t_i),\tau_j)+\\ &+ \tfrac{1}{2}f_{yy}(\mu(\tau_j|t_i),\tau_j)*\Sigma(\tau_j|t_i) && (83)\\ \mathrm{Cov}[f(y(\tau_j),\tau_j),y(\tau_j)|Z^i] &\approx f_y(\mu(\tau_j|t_i),\tau_j)\Sigma(\tau_j|t_i) && (84)\\ E[\Omega(y(\tau_j),\tau_j)|Z^i] &\approx \Omega(\mu(\tau_j|t_i),\tau_j)+\\ &+ \tfrac{1}{2}\Omega_{yy}(\mu(\tau_j|t_i))*\Sigma(\tau_j|t_i). && (85)\end{aligned}$$

$$\begin{aligned} \mathrm{Cov}[y_{i+1},h_{i+1}|Z^i] &\approx \Sigma(t_{i+1}|t_i)h'_y(\mu(t_{i+1}|t_i),t_{i+1}) && (86)\\ \mathrm{Var}[h_{i+1}|Z^i] &\approx h_y(\mu(t_{i+1}|t_i),t_{i+1})\Sigma(t_{i+1}|t_i)h'_y(\mu(t_{i+1}|t_i),t_{i+1}) && (87)\\ E[h_{i+1}|Z^i] &\approx h(\mu(t_{i+1}|t_i),t_{i+1})+\\ &+ \tfrac{1}{2}h_{yy}(\mu(t_{i+1}|t_i),t_{i+1})*\Sigma(t_{i+1}|t_i), && (88)\end{aligned}$$

where $(f_{yy}*\Sigma)_i=\sum_{jk}f_{yy,ijk}\Sigma_{jk}$ etc. Expanding to higher orders in the HNF (higher order nonlinear filter) yields moments of order $k>2$ on the right hand side, which must be dropped or factorized by the Gaussian assumption $m_k=(k-1)!!m_2^{k/2}, k$ even, $m_k=0, k$ odd. For details, see Jazwinski (1970) or Singer (2006*d*).

5.3 Local linearization LL

A related algorithm occurs if the drift is expanded directly in SDE (60). Using Itô's lemma one obtains

$$\begin{aligned} &f(y(t),t)-f(y(t_i),t_i)= && (89)\\ &\int_{t_i}^t f_y(y(s),s)dy(s)+\int_{t_i}^t \tfrac{1}{2}f_{yy}(y(s),s)*\Omega(y(s),s)ds+\int_{t_i}^t f_t(y(s),s)ds.\end{aligned}$$

Freezing the coefficients at (y_i, t_i) and using a state independent diffusion coefficient $\Omega(s)$, Shoji and Ozaki (1997, 1998) obtained the linearized SDE $(t_i \leq t \leq t_{i+1})$

$$\begin{aligned} dy(t) &= [f_y(y_i,t_i)(y(t)-y_i) + f(y_i,t_i) + f_t(y_i,t_i)(t-t_i) + \\ &\quad + \tfrac{1}{2} f_{yy}(y_i,t_i) * \Omega(t_i)(t-t_i)]dt + g(t)dW(t). \end{aligned}$$

The corresponding moment equations are

$$\begin{aligned} \dot{\mu}(t|t_i) &= f_y(y_i,t_i)(y(t|t_i)-y_i) + f(y_i,t_i) + f_t(y_i,t_i))(t-t_i) + \\ &\quad + \tfrac{1}{2} f_{yy}(y_i,t_i) * \Omega(t_i)(t-t_i) \qquad (90) \\ \dot{\Sigma}(t|t_i) &= f_y(y_i,t_i)\Sigma(t|t_i) + \Sigma(t|t_i) f_y'(y_i,t_i) + \Omega(t_i). \qquad (91) \end{aligned}$$

By contrast to the EKF and SNF moment equations which is a system of nonlinear differential equations, the Jacobians are evaluated once at the measurements (y_i, t_i) and the differential equations are linear and not coupled (for details, cf. Singer, 2002).

6 Filter Approximations based on Numerical Integration

The traditional way of nonlinear filtering has been the expansion of the system functions f, Ω and h. Another approach is the approximation of the filtering density $p(y|Z^i)$.

6.1 Unscented Kalman Filtering

The idea of Julier and Uhlmann (1997) was the definition of so called sigma points with the property that the weighted mean and variance over these points coincides with the true parameters. According to Julier et al. (2000) one can take the $2p+1$ points

$$x_0 = \mu;\ x_{\pm l} = \mu \pm \sqrt{p+\kappa}\Gamma_{\cdot l}, l = 1, ..., p \qquad (92)$$

with weights

$$\alpha_0 = \kappa/(p+\kappa); \alpha_l = 1/(2(p+\kappa)) = \alpha_{-l}, \qquad (93)$$

where $\Gamma_{\cdot l}$ is the lth column of the Cholesky root of $\Sigma = \Gamma\Gamma'$, κ is a scaling factor and p is the dimension of the random vector X. For example, in the univariate case $p = 1$ one obtains the three points $\mu, \mu \pm \sqrt{1+\kappa}\,\sigma$.
The UT method may be interpreted in terms of the singular density

$$p_{UT}(x) = \sum_{l=-p}^{p} \alpha_l \delta(x - x_l). \qquad (94)$$

Then, however, only nonnegative weights α_l are admissible. Generally, the expectation $E_{UT}[X] = \int x\, p_{UT}(x)dx = \mu$ and

$$\begin{aligned} \mathrm{Var}_{UT}(X) &= \int (x-\mu)(x-\mu)' p_{UT}(x)dx \qquad (95) \\ &= \sum_{l=-p}^{p} (x_l - \mu)(x_l - \mu)'\alpha_l \qquad (96) \\ &= \sum_{l=1}^{p} \Gamma_{\cdot l}(\Gamma_{\cdot l})' = \Gamma\Gamma' = \Sigma \qquad (97) \end{aligned}$$

yields the correct first and second moment. Nonlinear expectations are easily evaluated as sums

$$E_{UT}[f(X)] = \int f(x)p_{UT}(x)dx = \sum_l f(x_l)\alpha_l \tag{98}$$

$$= \frac{\kappa}{p+\kappa}f(\mu) + \frac{1}{2(p+\kappa)}\sum_{l=-p,p\neq 0}^{p} f(x_l) \tag{99}$$

Using large κ, the EKF formula $E_{Taylor}[f(X)] = f(\mu)$ is recovered.
All expectations in the filter are evaluated using the sigma points computed from the conditional moments $\mu(\tau_j|t_i), \Sigma(\tau_j|t_i)$. To display the dependence on the moments, the notation $y_l = y_l(\mu, \Sigma)$ will be used. For example, the terms in the time update are (short notation dropping arguments)

$$E[f|Z^i] \approx \sum_l f(y_l)\alpha_l \tag{100}$$

$$\mathrm{Cov}[f, y|Z^i] = E[f(y)(y-\mu)'|Z^i] \approx \sum_l f(y_l)(y_l-\mu)'\alpha_l \tag{101}$$

$$E[\Omega|Z^i] \approx \sum_l \Omega(y_l)\alpha_l \tag{102}$$

with sigma points $y_l = y_l(\mu(\tau_j|t_i), \tau_j), \Sigma(\tau_j|t_i))$. With the new moments $\mu(\tau_{j+1}|t_i)$, $\Sigma(\tau_j|t_i)$, updated sigma points are computed.

6.2 Gauss-Hermite Filtering

For the Gaussian filter, one may assume that the true $p(x)$ is approximated by a Gaussian distribution $\phi(x; \mu, \sigma^2)$ with the same mean μ and variance σ^2. Then, the Gaussian integral

$$E_\phi[f(X)] = \int f(x)\phi(x;\mu,\sigma^2)dx = \int f(\mu+\sigma z)\phi(z;0,1)dz \tag{103}$$

$$\approx \sum_{l=1}^{m} f(\mu+\sigma\zeta_l)w_l = \sum_{l=1}^{m} f(\xi_l)w_l \tag{104}$$

may be approximated by Gauss-Hermite quadrature (cf. Ito and Xiong, 2000). Ito & Xiong (2000) The ζ_l, w_l are quadrature points and weights for the standard gaussian distribution $\phi(z; 0, 1)$. If such an approximation is used, one obtains the Gauss-Hermite filter (GHF). Generally, filters using Gaussian densities are called Gaussian filters (GF). The GHF can be interpreted in terms of the singular density $p_{GH}(x) = \sum_{l=1}^{m} w_l\delta(x-\xi_l)$ concentrated at the quadrature points ξ_l. The Gauss-Hermite quadrature rule is exact up to order $O(x^{2m-1})$. Multivariate Gaussian integrals can be computed by transforming to the standard normal distribution and p-fold application of (103).
The Gaussian filter is equivalent to an expansion of f to higher orders L

$$E[f(X)] \approx \sum_{l=0}^{L} \tfrac{1}{l!}f^l(\mu)E[X-\mu]^l = \sum_{l=0}^{L} \tfrac{1}{l!}f^l(\mu)m_l \tag{105}$$

(higher order nonlinear filter HNF$(2, L)$) and factorization of the moments according to the Gaussian assumption $m_l := E[X-\mu]^l = (l-1)!!\sigma^l$ (l even) and $m_l = 0$ (l odd). This leads

to an exact computation of (103) for $L \to \infty$. In this limit, the HNF and GF coincide. In the EKF=HNF(2,1) and SNF=HNF(2,2), the higher order corrections are neglected. Also, third and higher order moments could be used (HNF(K, L); cf. Singer (2006*d*)).
It is interesting that $\kappa = 2, p = 1$ in the UT corresponds to a Gauss-Hermite rule with $m = 3$ sample points (Ito and Xiong, 2000).

6.3 Generalized Gauss-Hermite Filtering

The Gaussian filter assumed a Gauss density $\phi(y; \mu, \sigma^2)$ for the filter distribution $p(y)$. More generally, one can use a Hermite expansion

$$p(y) \quad = \quad \phi(y; \mu, \sigma^2) \sum_{n=0}^{K} c_n H_n((y-\mu)/\sigma) = \phi(y) * H(y, K), \tag{106}$$

with Fourier coefficients $c_0 = 1, c_1 = 0, c_2 = 0$,

$$c_3 \quad := \quad (1/3!)E[Z^3] = (1/3!)\nu_3 \tag{107}$$
$$c_4 \quad := \quad (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(\nu_4 - 3), \tag{108}$$

$Z := (Y - \mu)/\sigma$ and orthogonal polynomials $H_0 = 1, H_1 = x, H_2 = y^2 - 1, H_3 = y^3 - 3y, H_4 = y^4 - 6y^2 + 3$ etc. Expectation values occuring in the update equations are again computed by Gauss-Hermite integration, including the nongaussian term

$$H(y; \{\mu, m_2, ..., m_K\}) \quad := \quad \sum_{n=0}^{K} c_n H_n((y-\mu)/\sigma). \tag{109}$$

Since $H(y; \{\mu, m_2, ..., m_K\})$ depends on higher order moments, one must use K moments equations (74). The choice $K = 2$ recovers the usual Gaussian filter, since $c_0 = 1, c_1 = 0, c_2 = 0$. The Hermite expansion can model bimodal, skewed and leptokurtic distributions. For details, see Singer (2006*d*,*b*,*c*).
Related algorithms have been developed by Srinivasan (1970), Challa et al. (2000) but we formulate the time update as integro-differential equations, solved stepwise by using Gauss-Hermite integration. Moreover, computation of the measurement update (Bayes formula) is improved. We use the normal correlation update as Gaussian weight function in the Gauss-Hermite quadrature to achieve higher numerical accuracy. The a posteriori moments are obtained directly without iterative procedure.
In contrast, Challa et al. (2000) use truncated moment equations in the time update (higher order moments are set to zero) and (iterated) EKF measurement updates as approximate means and variances for the posterior Hermite series (p. 3400). Moreover, only linear system equations are considered.
In the proposed approach, the time update equations are closed by explicit integration over the given Hermite expansion of order K. Therefore, for $K = 2$ we obtain the GHF as special case, whereas Challa et al. (p. 3399) obtain the EKF. Thus, like the GHF, higher order moments are not neglected but approximated through the approximate Hermite density.

7 Discussion

1. The density based filters UKF and GHF have the strong advantage, that no derivatives of the system functions must be computed. This is no problem for the EKF and SNF, but for higher orders in the HNF(K, L) complicated tensor expressions arise. Moreover, higher order moments must be dropped or factorized in order to obtain closed equations. In the multivariate case, the formulas for Gaussian moments are involved. The fourth moment is

$$m_{4,ijkl} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}. \tag{110}$$

   For a general formula, see (Gardiner 1996, p. 36)

2. Apart from an implementation point of view (see 1.), the low order EKF and SNF suffer from problems such as filter divergencies, especially when the sampling intervals are large. Simulation studies suggest, that the UKF and GHF are more stable and yield smaller filtering error in the mean (Singer 2006*a*).

3. The moment equations and measurement updates as derived in section (4.3) involve expectations with respect to the filter density $p(y)$, but not for the noise processes. Their statistics are already included in these updates (the terms $E[\Omega]dt = E[gdW(gdW)']$ and $R = \mathrm{Var}(\epsilon)$ stem from the noise sequences). Thus no sigma points w.r.t the noises must be computed, as suggested in the literature on the UKF (Julier et al. 2000, Sitz, Schwarz, Kurths & Voss 2002, Sitz, Schwarz & Kurths 2002). This is only necessary if the system is first modeled deterministically and afterwards extended by the noises. This is neither necessary nor efficient.

8 Parameter estimation: the Lorenz model

In most applications, the models contain unknown parameters to be estimated from the available sampled data. The likelihood function, which is computed recursively by the filter, yields maximum likelihood (ML) estimates and asymptotic standard errors by numerically maximizing the likelihood and evaluting the negative Hessian matrix (observed Fisher information). In the context of nonlinear filter theory, a more convenient approach can be taken by considering the (fixed) parameters as latent states with trivial dynamics $d\psi = 0$, i.e. constant. Thus the filter runs with the extended state $\eta(t) = \{y(t), \psi(t)\}$ and yields conditional distributions $p(y(t), \psi(t)|Z^i)$. The filters considered in this paper produce approximate conditional expectations $E(y(t), \psi(t)|Z^i)$ and variances $\mathrm{Var}(y(t), \psi(t)|Z^i)$. Thus we obtain recursive estimates $\hat{\psi}(t) = E(\psi(t)|Z^i)$ of the parameter vector which is updated at the times of measurement. Also, recursive standard errors are obtained. The Bayesian estimates may be called expected a posteriori estimates (EAP). Maximization of the marginal conditional density $p(\psi(t)|Z^i)$ with diffuse a priori distribution $p(\psi(t_0))$ would give the ML estimator.
In this example the chaotic Lorenz model (Lorenz, 1963) is considered. It is the best known simple system exhibiting chaos and has been used for many purposes. Among them are the Navier-Stokes equations in meteorology (the original application), laser theory (Haken, 1977, Graham, 1989), psychology (Singer, 1992) and sociology (Troitsch, 1990). ML estimation has

method	x	y	z	σ	r	b
EKF	34.4278	49.0568	46.4022	15.968	12.9247	8.18301
(std)	2.35341	4.8163	4.38871	6.61447	4.10521	0.979441
SNF	34.4433	49.2547	46.4675	16.0031	13.7448	8.50263
(std)	2.36142	4.81487	4.41146	6.61724	4.56789	1.1919
HNF(2,2)	34.4388	49.2483	46.4605	15.9944	13.7155	8.50246
(std)	2.35904	4.80811	4.41011	6.59141	4.48563	1.17291
UKF, $\kappa = 0$	34.4355	49.2436	46.4566	15.9841	13.6924	8.50225
(std)	2.35775	4.80203	4.411	6.57232	4.41997	1.15926
FIF, $n = 10000$	35.3055	50.9326	48.7735	22.2944	20.5525	9.43772
(std)	2.81775	5.70727	7.53299	11.3771	10.4972	2.41376

Table 1: Lorenz model: mean and standard deviation of squared filter error $A = \sum_t \nu_t^2$ in $M = 100$ samples ($\Delta t = 0.2$).

been considered by Ozaki et al. (2000) We assume that a vector of variables $Y(t) = [x, y, z](t)$ evolves according to the system of nonlinear stochastic differential equations

$$d \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} -\sigma x(t) + \sigma y(t) \\ -x(t)z(t) + rx(t) - y(t) \\ x(t)y(t) - bz(t) \end{bmatrix} dt + \begin{bmatrix} g_1 dW_1(t) \\ g_2 dW_2(t) \\ g_3 dW_3(t) \end{bmatrix} \tag{111}$$

The usual choice of parameters $\{\sigma_0, r_0, b_0\} = \{10, 28, 8/3\}$ leading to chaotic trajectories was taken. The system error matrix is $G = \mathrm{diag}(4, 4, 4)$ and the measurement error covariance was taken as $R = \mathrm{diag}(0.1, 0.1, 0.1)$. The data were simulated using a discretization interval of $\delta t = 0.01$ using an Euler-Maruyama scheme and afterwards measured at regular sampling intervals $\Delta t = 0.2, 0.5$. In order to estimate the parameters, the state was extended with $\psi(t) = \{\sigma, r, b\}$ using the dynamics $d\psi = 0$ (constant parameters). The initial state $\eta(0) = \{Y(0), \psi(0)\}$ was assumed to be distributed as $N(\mu, \Sigma)$ with $E[\eta(0)] = \{0, 0, 0, \sigma, r, b\}$ and $\mathrm{Var}(\eta(0)) = \mathrm{diag}(10, 10, 10, 1, 1, 1)$, which reflects moderate a priori knowledge of the unknown parameters. In the filter solutions (figs. 5–6), the initial value $E[\eta(0)] = \{0, 0, 0, \sigma_0 + 1, r_0 + 1, b_0 + 1\}$ was used. As is shown in the figures, the filtered parameters converge to a vicinity of the true value $\{\sigma_0, r_0, b_0\} = \{10, 28, 8/3\}$ and the approximate 95% HPD confidence intervals $E(\psi(t)|Z^i) \pm 1.96[\mathrm{Var}(\psi(t)|Z^i)]^{0.5}; t_i \leq t$ get smaller with increasing sample size. The behavior of the several algorithms is similar. The simulated filter FIF needs many trajectories to fill the 6-dimensional phase space with enough sample points. The performance can be expressed by the squared filtering error $A = \sum_t \nu_t^2$, $\nu_t = y(t) - \hat{y}(t)$. Table 1 reports the means and standard deviations of the squared filter error in a simulation study with $M = 100$ samples ($\Delta t = 0.2$). Surprisingly, the performance of the EKF is better than the UKF ($\kappa = 0$) and the SNF. The UKF and the SNF are very similar. The FIF ($n = 1000$) shows degraded performance due to simulation errors. Using $n = 10000$ trajectories leads to better results, but still a larger sample size would be necessary.

Mathematica code:

```
(* system functions *)

driftLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] :=
    {-sigma (x - y),
     -x z + r x - y,
      x y - b z     }
```

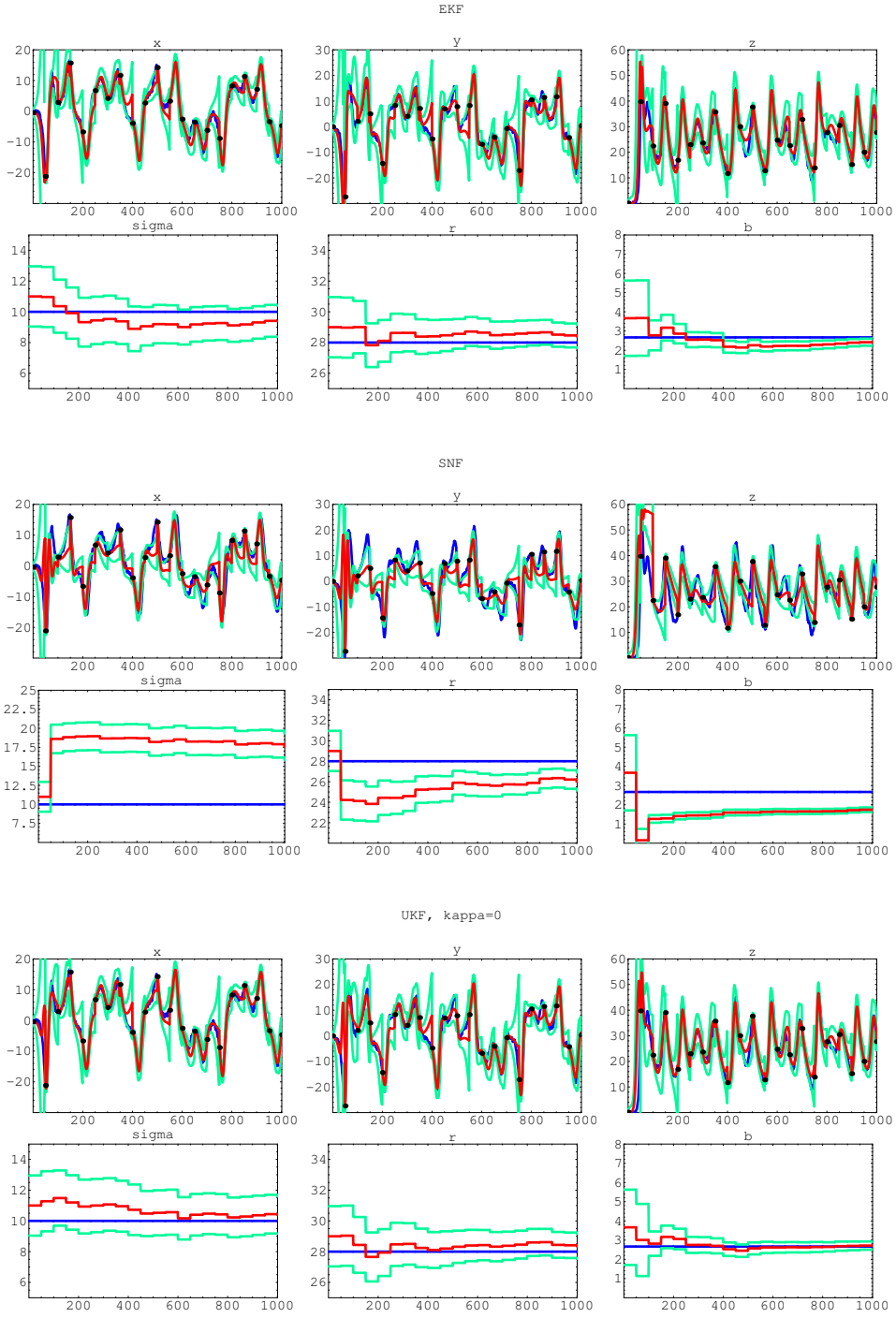

Figure 5: Filtered estimates of the Lorenz model. Comparision of several filtering algorithms. Components 4–6 are the unobserved parameters σ, r, b. Sampling interval $\Delta t = 0.5$.

Figure 6: Filtered estimates of the Lorenz model. Comparision of several filtering algorithms. Components 4–6 are the unobserved parameters σ, r, b. Sampling interval $\Delta t = 0.5$.

```
DdriftLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] =
        Outer[D,driftLorenz[{x,y,z},t,{sigma,r,b}],{x,y,z}]
diffusionLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] = DiagonalMatrix[2{1,1,1}]
hLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] := {x,y,z}
DhLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] = DiagonalMatrix[{1.,1.,1.}]
RLorenz[{x_,y_,z_},t_,{sigma_,r_,b_}] = DiagonalMatrix[{1,1,1}]
mueLorenz[t_,{sigma_,r_,b_}] = {0,0,0}
sigmaLorenz[t_,{sigma_,r_,b_}] = DiagonalMatrix[{10,10,10}]

(* extended system functions *)

driftLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_] :=
    Module[{f},
    f=driftLorenz[{x,y,z},t,{sigma,r,b}];
    {f,{0,0,0}}//Flatten]
DdriftLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_] =
    Outer[D,driftLorenzX[{x,y,z,sigma,r,b},t,Theta],{x,y,z,sigma,r,b}]
diffusionLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_]:=DiagonalMatrix[{4.,4.,4.,.0001,.0001,.0001}]
hLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_] := {x,y,z}
DhLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_] = DiagonalMatrix[{1.,1.,1.}]~ii~Shape[0,{3,3}]
RLorenzX[{x_,y_,z_,sigma_,r_,b_},t_,Theta_] = .1 DiagonalMatrix[{1,1,1}]
mueLorenzX[t_,{sigma_,r_,b_}] = {0,0,0,sigma,r,b}//N
sigmaLorenzX[t_,{sigma_,r_,b_}] =
                DiagonalMatrix[{10,10,10,1,1,1}]//N

(* simulation of data *)
k = 3;p = 6;q = 1;T = 1000; T1 = T+1;dt = 0.01;r = 6;u = 3;
Theta = {10,28,8/3}//N;
Y0 = {.01,.01,.01,Theta}//Flatten;
X = Shape[10.,{T+1,q}]//N;
SeedRandom[-999999];

{Z,Y}=NonLinStateSpace[hLorenzX, {RLorenzX, konstant}, driftLorenzX, diffusionLorenzX,
                       Y0, X, T, dt, Theta, Gauss, EulerStep];

ti={1, 21,41,81,220, 300, 301, 350, 490,500,600,610,677,800}; (* irregular sampling *)
ti=Range[1,T+1,50]; (* selection deltat= 0.5 *)

TT=Delta[ti]//Max
miss=-9.;
ZM = Array[miss&,{T+1,k}];
Map[(ZM[[#]] = Z[[#,{1,2,3}]])&,ti];

(* nonlinear filtering with EKF *)

sel=Range[T1];
{YF,PF,clik} =
        EKF[{ZM[[sel]], X[[sel]]}, Theta+1,
             {driftLorenzX, DdriftLorenzX,diffusionLorenzX, hLorenzX, DhLorenzX,
              RLorenzX, mueLorenzX,sigmaLorenzX,dt},
              {miss=-9., option=2}] ;//Timing
```

9 Stochastic Volatility

From the preceeding examples it seems, that UKF, HNF or GHF are always the best algorithms, with the FIF as the only competitor (but with high computational demand). The stochastic volatility model is interesting in this respect, since it contains a latent variable acting on the diffusion term. It is an alternative to the well known ARCH specification, where the volatility can be expressed in terms of measured states. Here, the state space model

$$dS(t) = \mu S(t)dt + \sigma(t)S(t)dW(t) \tag{112}$$

$$d\sigma(t) = \lambda[\sigma(t) - \bar{\sigma}]dt + \gamma dV(t) \tag{113}$$

$$z_i = S(t_i), \tag{114}$$

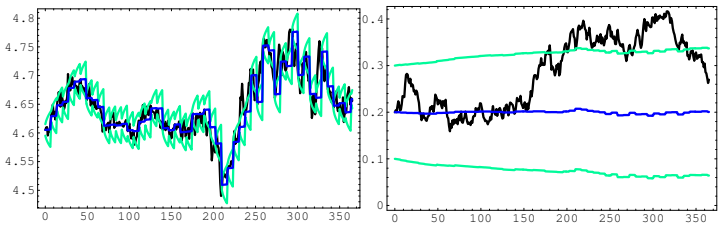

Figure 7: GGHF($K = 2$) = GHF: Stochastic volatility model. Stock price (left) and volatility (right). Similar results are obtained for the EKF, SNF and UKF.

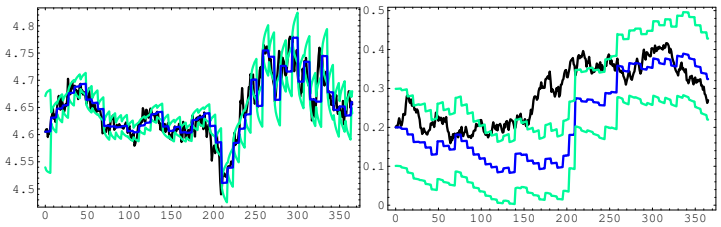

Figure 8: GGHF($K = 4$): Stochastic volatility model. Stock price (left) and volatility (right).

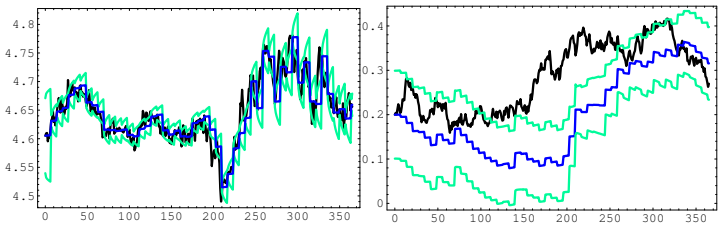

Figure 9: GGHF($K = 6$): Stochastic volatility model. Stock price (left) and volatility (right).

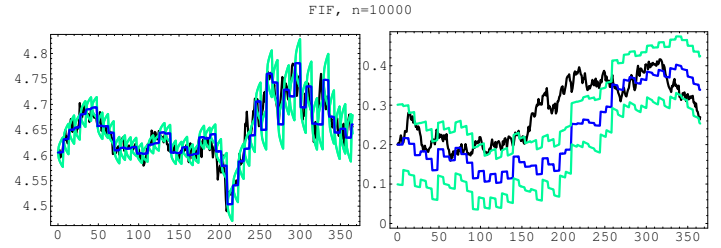

Figure 10: Monte Carlo filter ($N = 10000$ replications): Stochastic volatility model. Stock price (left) and volatility (right).

$\mathrm{Var}(dW, dV) = 0$ (Scott, 1987, Hull and White, 1987) contains an unobservable volatility process $\sigma(t)$. The data are simulated daily (365 days), but measured on a weekly sampling interval. $\{\mu, \lambda, \bar{\sigma}, \gamma\} = \{.07, -1., .2, .2\}$. The model accounts for the fact, that the returns

$$r(t) = dS/S = \mu dt + \sigma(t) dW(t) \tag{115}$$

on financial time series can exhibit a time dependent variance and leptokurtosis of the return distribution, as is found empirically (cf. Kim et al., 1998, Fridman and Harris, 1998). Using higher order moment information as in the GGHF(4), GGHF(6) yields estimates of the latent volatility trajectory (figs. 8–9) similar to Monte Carlo filtering (functional integral filter FIF, $N = 10000$ replications; fig. 10; cf. Singer, 2003).

10 Conclusion

We have presented linear (LSDE) and nonlinear filter algorithms (SDE) for sampled continuous time systems as well as SEM approaches to linear SDE estimation. In the latter case, the required nonlinear parameter restrictions can be implemented very easily in Mathematica. The nonlinear filter algorithms are based on moment equations, which can be approximated by Taylor expansions (EKF, HNF, LL) or numerical integration (UKF, GHF, GGHF). All algorithms are collected in the packages SDE and SEM.

References

Abrikosov, A., Gorkov, L. & Dzyaloshinsky, I. (1963), *Methods of Quantum Field Theory in Statistical Physics*, Dover, New York.

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

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

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

Bergstrom, A., ed. (1976*b*), *Statistical Inference in Continuous Time Models*, North Holland, Amsterdam.

Challa, S., Bar-Shalom, Y. & Krishnamurthy, V. (2000), 'Nonlinear Filtering via Generalized Edgeworth Series and Gauss-Hermite Quadrature', *IEEE Transactions On Signal Processing* **48**, 1816–1820.

Daum, F. (2005), 'Nonlinear Filters: Beyond the Kalman Filter', *IEEE A&E Systems Magazine* **20,8**, 57–69.

Fridman, M. & Harris, L. (1998), 'A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models', *Journal of the American Statistical Association* **16, 3**, 284–291.

Gardiner, C. (1996), *Handbook of Stochastic Methods*, second edn, Springer, Berlin, Heidelberg, New York.

Graham, R. (1989), Macroscopic potentials, bifurcations and noise, *in* F. Moss & P. McClintock, eds, 'Noise in nonlinear dynamical systems, vol. I', Cambridge University Press.

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

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

Hamerle, A., Nagl, W. & Singer, H. (1993), 'Identification and estimation of continuous time dynamic systems with exogenous variables using panel data', *Econometric Theory* **9**, 283–295.

Harvey, A. & Stock, J. (1985), 'The estimation of higher order continuous time autoregressive models', *Econometric Theory* **1**, 97–112.

Hull, J. & White, A. (1987), 'The Pricing of Options with Stochastic Volatilities', *Journal of Finance* **XLII,2**, 281–300.

Ito, K. & Xiong, K. (2000), 'Gaussian filters for nonlinear filtering problems', *IEEE Transactions on Automatic Control* **45, 5**, 910–927.

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

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

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

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic Volatility: Likelihood Inference and Comparision with ARCH Models', *Review of Economic Studies* **45**, 361–393.

Liptser, R. & Shiryayev, A. (1977, 1978), *Statistics of Random Processes, Volumes I and II*, Springer, New York, Heidelberg, Berlin.

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

Lorenz, E. (1963), 'Deterministic Nonperiodic Flow', *J. Atmos. Sci.* **20**, 130.

Möbus, C. & Nagl, W. (1983), Messung, Analyse und Prognose von Veränderungen (Measurement, analysis and prediction of change; in german), *in* 'Hypothesenprüfung, Band 5 der Serie Forschungsmethoden der Psychologie der Enzyklopädie der Psychologie', Hogrefe, pp. 239–470.

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

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

Oud, J., van Leeuwe, J. & Jansen, R. (1993), Kalman Filtering in discrete and continuous time based on longitudinal LISREL models, *in* J. Oud & R. van Blokland-Vogelesang, eds, 'Advances in longitudinal and multivariate analysis in the behavioral sciences', ITS, Nijmegen, Netherlands, pp. 3–26.

Ozaki, T., Jimenez, J. & Haggan-Ozaki, V. (2000), 'The Role of the Likelihood Function in the Estimation of Chaos Models', *Journal of Time Series Analysis* **21, 4**, 363–387.

Phillips, P. (1976a), The problem of identification in finite parameter continuous time models, *in* A. Bergstrom, ed., 'Statistical Inference in Continuous Time Models', North Holland, Amsterdam, pp. 123–134.

Scott, L. (1987), 'Option pricing when the variance changes randomly: Theory, estimation, and an application', *Journal of Financial and Quantitative Analysis* **22**, 419–438.

Shoji, I. & Ozaki, T. (1997), 'Comparative Study of Estimation Methods for Continuous Time Stochastic Processes', *Journal of Time Series Analysis* **18, 5**, 485–506.

Shoji, I. & Ozaki, T. (1998), 'A statistical method of estimation and simulation for systems of stochastic differential equations', *Biometrika* **85, 1**, 240–243.

Singer, H. (1991), *LSDE - A program package for the simulation, graphical display, optimal filtering and maximum likelihood estimation of Linear Stochastic Differential Equations, User's guide*, Meersburg.

Singer, H. (1992b), Dynamic structural equations in discrete and continuous time, *in* G. Haag, U. Mueller & K. Troitzsch, eds, 'Economic Evolution and Demographic Change', Springer, Berlin, Heidelberg, New York, pp. 306–320.

Singer, H. (1992d), 'The aliasing phenomenon in visual terms', *Journal of Mathematical Sociology* **17, 1**, 39–49.

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

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

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

Singer, H. (1999), *Finanzmarktökonometrie. Zeitstetige Systeme und ihre Anwendung in Ökonometrie und empirischer Kapitalmarktforschung*, Physica-Verlag, Heidelberg.

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. (2004d), SEM – Linear Structural Equations Models with Arbitrary Nonlinear Parameter Restrictions, Version 0.1, Mathematica Program, http://www.fernuni-hagen.de/STATISTIK/forschung/SEMArchive.exe, Fernuniversität in Hagen, Hagen, Germany.

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

Singer, H. (2006*b*), Generalized Gauss-Hermite Filtering, Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 391, FernUniversität in Hagen. http://www.fernuni-hagen.de/FBWIWI/forschung/beitraege/pdf/db391.pdf.

Singer, H. (2006*c*), Generalized Gauss-Hermite Filtering for Multivariate Diffusion Processes, Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 402, FernUniversität in Hagen. http://www.fernuni-hagen.de/FBWIWI/forschung/beitraege/pdf/db402.pdf.

Singer, H. (2006*d*), 'Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models', *Computational Statistics* **21,3**, 385–397.

Singer, H. (2007), Stochastic Differential Equation Models with Sampled Data, *in* K. van Montfort, H. Oud & A. Satorra, eds, 'Longitudinal Models in the Behavioral and Related Sciences', The European Association of Methodology (EAM) Methodology and Statistics series, vol. II, Lawrence Erlbaum Associates, Mahwah, London, pp. 73–106.

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

Sitz, A., Schwarz, U. & Kurths, J. (2002), 'The Unscented Kalman Filter, A Powerful Tool For Data Analysis', *International Journal of Bifurcation and Chaos* **14, 6**, 2093–2105.

Sitz, A., Schwarz, U., Kurths, J. & Voss, H. (2002), 'Estimation of parameters and unobserved components for nonlinear systems from noisy time series', *Physical Review E* **66**, 016210–1–016210–9.

Srinivasan, K. (1970), 'State Estimation by Orthogonal Expansion of Probability Distributions', *IEEE Transactions on Automatic Control* **15,1**, 3–10.

Troitzsch, K. G. (1990), *Modellbildung und Simulation in den Sozialwissenschaften*, Westdeutscher Verlag, Opladen.

Van Kampen, N. (1981), 'Itô vs. Stratonovich', *Journal of Statistical Physics* **24**, 175–187.