Computational Aspects of Continuous-Discrete Extended *Kalman*-Filtering

Thomas Mazzoni

04-23-2007

Abstract

This paper elaborates how the time update of the continuous-discrete extended *Kalman*-Filter (EKF) can be computed in the most efficient way. The specific structure of the EKF-moment differential equations leads to a new *Taylor-Heun*-approximation of the nonlinear vector field.

Furthermore, the order of consistency and stability behavior of the outlined procedure is investigated. The results are incorporated into an algorithm with adaptive controlled step size, assuring a fixed numerical precision with minimal computational effort. Additionally, the upper bound of the step size is monitored continuously in order to guarantee positive semidefiniteness of the state error covariance matrix.

1. Introduction

Recently continuous-discrete state space models associated with nonlinear filtering problems became a matter of interest not only in engineering, but in various fields of scientific research. Particularly the calculation of transition densities of nonlinear diffusion problems has become a major topic of scientific activity in physics (Sitz et al. 2002), finance (Aït-Sahalia 2002, Jensen and Poulsen 2002), sociology (Singer 2006b) and economics (Mazzoni 2007).

The most established nonlinear filter is the extended *Kalman*- or *Kalman-Schmidt*-Filter (Kalman 1960, Schmidt 1966) which is known for more than forty years. Recently, more sophisticated filters like the unscented *Kalman*-filter (Julier et al. 2000, Julier and Uhlmann 2004), DD1- and DD2-filter (Nørgaard et al. 2000) or *Gauss-Hermite*-filter (Ito and Xiong 2000) have been suggested for nonlinear problems but nevertheless, the EKF is still popular which is reflected in the numerous variants and modifications of this filter (see e.g. Tanizaki 1996, chap 3, Johnston and Krishnamurthy 2001, Lefebvre et al. 2004). Furthermore, the EKF is often used as benchmark to verify improvements of new designed filters under certain conditions. Especially in this regard it is vital to gain exact results with the EKF in order to avoid misjudgment of the new procedure.

Unlike discrete-discrete filtering problems, the continuous-discrete case requires the solution of an ordinary differential equation (ODE) for the moments. Even though there are many procedures for numerical integration, the specific structure of the involved moment equations allows the derivation of a new and

most efficient procedure for calculating prior densities. Stability features and error control of the numerical solutions are most important, because in case of maximum likelihood parameter estimation the likelihood function is calculated from the prediction error decomposition (Schweppe 1965), which is roughly spoken a kind of distance measure between prior and observation density.

In this paper it is shown step by step how the necessary elements of the corresponding algorithm can be computed most efficiently. Additionally, the related proofs concerning stability and error bounds are provided. Subsequently, the method is demonstrated for a linear *Ornstein-Uhlenbeck*-process and a highly nonlinear *Van der Pol* oscillator.

2. Moment Equations and State Vector

In the following discussion it is presumed that unique solutions of the involved ordinary and stochastic differential equations and all necessary derivatives of drift and diffusion functions exist. Sufficient conditions can be found in numerous literature (e.g. Arnold 2001, Kloeden and Platen 1992).

Starting from the stochastic *Itô*-type differential equation

$$dx = f(x)dt + G(x)dW(t) \tag{1}$$

with state vector x, nonlinear drift function $f(x)$, diffusion matrix $G(x)$ and the appropriate dimensioned *Wiener*-process $W(t)$, the coupled moment differential equations for the continuous-discrete EKF can be derived (Jazwinski 1970, chap. 6.4)

$$\dot{\mu} = f(\mu) \tag{2a}$$

$$\dot{\Sigma} = A(\mu)\Sigma + \Sigma A^T(\mu) + \Omega(\mu) \tag{2b}$$

with the *Jacobi*an $A = \frac{\partial f}{\partial x}$ and $\Omega = GG^T$. The time dependency of the variables is suppressed for notational simplicity. Equations (2a) and (2b) can be motivated in several ways. One particular simple method is to calculate the *Taylor*-linearized expectation and covariance of the *Euler-Maruyama*-approximation of the stochastic differential equation (1) in the limit $\Delta t \to 0$ (see Singer 2006a).

Because the *Jacobi*an A has to be computed anyway in equation (2b) it can be used to enhance the approximation quality of the state expectation (2a).

2.1. *Taylor-Heun*-Approximation

We start with the trapezoidal approximation (*Heun*-scheme) of equation (2a)

$$\mu_{t+\Delta t} \approx \mu_t + \frac{1}{2}\big(f(\mu_t) + f(\mu_{t+\Delta t})\big)\Delta t. \tag{3}$$

The drift vector is assumed autonomous because a time dependency can always be eliminated by extension of the state space. The vector field f at $\mu_{t+\Delta t}$ is unknown, but it can be approximated by linear *Taylor*-expansion of the drift function around μ_t. If this expansion is inserted into (3) one obtains

$$\mu_{t+\Delta t} \approx \mu_t + f(\mu_t)\Delta t + \frac{1}{2}A(\mu_t)\big(\mu_{t+\Delta t} - \mu_t\big)\Delta t$$

and after certain algebraic manipulations the *Taylor-Heun*-formula reads

$$\boxed{\mu_{t+\Delta t} \approx \mu_t + \left(I - A(\mu_t)\frac{\Delta t}{2}\right)^{-1} f(\mu_t)\Delta t} \tag{4}$$

with the identity matrix I. Unlike the original *Heun*-scheme, using an *Euler*-predictor for the function value at the upper boundary, scheme (4) uses linear *Taylor*-expansion of f.

For any practical purpose the matrix $I - A(\mu_t)\frac{\Delta t}{2}$ can be assumed to be nonsingular. Regarding computational efficiency, instead of matrix inversion, the corresponding linear system can be solved. The *Taylor-Heun*-integration is consistent and therefore convergent with order $\mathcal{O}(\Delta t^2)$ (see appendix proof A.1).

Furthermore, the suggested *Taylor-Heun*-scheme is A-stable, which can be shown with the help of the scalar test equation $\dot{y} = \lambda y$, introduced by Dahlquist[1], where λ is a complex number with negative real part, $\lambda \in \mathbb{C}^-$. For $\kappa = \lambda\Delta t$, the discrete approximation step associated with (4) is

$$y_{t+\Delta t} = y_t + \Phi_t(y_t, \Delta t)\Delta t = y_t + \frac{\lambda y_t}{1 - \lambda\frac{\Delta t}{2}}\Delta t = \frac{1 + \frac{\kappa}{2}}{1 - \frac{\kappa}{2}} y_t = H(\kappa) y_t, \tag{5}$$

where Φ is called the increment function and H is the stability function. It is easily seen that $|H(\kappa)| < 1$ for any finite $\kappa \in \mathbb{C}^-$, thus the approximation scheme is A-stable. The stability function $H(\kappa)$ in (5) is identical to the one of the implicit *Gauss-Legendre*-formula with the increment function

$$\Phi_t(y_t, \Delta t) = f\left(t + \frac{\Delta t}{2}, y_t + \Phi_t(y_t, \Delta t)\frac{\Delta t}{2}\right), \tag{6}$$

which can be shown analogously. In fact, the *Taylor-Heun*-increment offers characteristics of a *Rosenbrock-Wanner*-type approximation (Rosenbrock 1963). These methods start with a linear implicit *Runge-Kutta*-scheme and calculate the increment function using a *Newton*-step. Finally, an increment of the form

$$\Phi(t, y_t, \Delta t) = \big(I - \gamma A(y_t)\Delta t\big)^{-1} f(y_t)$$

is derived, where the parameter γ has to be determined by solving a linear equation system.

The *Taylor-Heun*-method is motivated in a completely different way, but results in a very similar increment, featuring A-stability and second order consistency. The main advantage of this method is the low computational effort in the EKF context, because the necessary *Jacobi*an has to be calculated anyway. Furthermore, unlike the *Heun*-scheme, the increment function requires no nested calculations and unlike *Rosenbrock-Wanner*-schemes, there are no additional coefficients to be determined by solving a linear system.

[1]For a review of A-stability concept and Dahlquist's theorems see Wanner (2003).

3. State Error Covariance

The state error covariance equation (2a) is a non-autonomous, linear matrix differential equation. To simplify the following algebraic operations, it can be presented in the vectorial form $\sigma_t = \text{vec}\ \Sigma_t$. The vec-operator decomposes the quadratic $(n \times n)$ matrix into their column vectors and attaches them to a single $(n^2 \times 1)$ vector (see Magnus and Neudecker 1988). For arbitrary appropriate dimensioned matrices A, B, C the operation

$$\text{vec}\ ABC = (C^T \otimes A)\,\text{vec}\ B,$$

with $\otimes$ denoting the right handed *Kronecker*-product, is true (Magnus and Neudecker 1988, chap. 2.4). In the following, vectorized matrices are symbolized by their small letter counterparts. Therefore the vectorial covariance equation reads

$$\begin{aligned} \dot{\sigma}_t &= \big(A(\mu_t) \otimes I\big)\sigma_t + \big(I \otimes A(\mu_t)\big)\sigma_t + \omega(\mu_t) \\ &= L(A_t)\sigma_t + \omega_t, \end{aligned} \tag{7}$$

with $L(A_t) = A(\mu_t) \otimes I + I \otimes A(\mu_t)$.

3.1. *Gauss-Legendre*-Approximation

The numerical integration procedure for the state error covariance ODE should provide three key features. In first place, it should be consistent with the same order as the *Taylor-Heun*-approximation of the state expectation. In second place, it should be able to process nonautonomous differential equations and finally, it should be A-stable as well.

These features are supported by the *Gauss-Legendre*-formula with the implicit increment rule (6). In particular *Gauss-Legendre* has the same stability function as *Taylor-Heun*. Because of the linear structure of (7), the increment can be calculated explicitly

$$\begin{aligned} \phi(t, \sigma_t, \Delta t) &= L(A_\tau)\sigma_t + L(A_\tau)\phi(t, \sigma_t, \Delta t)\frac{\Delta t}{2} + \omega_\tau \\ &= \left(I - L(A_\tau)\frac{\Delta t}{2}\right)^{-1}\big(L(A_\tau)\sigma_t + \omega_\tau\big), \end{aligned} \tag{8}$$

with $\tau = t + \frac{\Delta t}{2}$. Again, ϕ denotes the vectorized increment function. Unfortunately, (8) does not allow to invert the initial vec-operation explicitly, thus the appropriate matrix partition has to be compelled by $\text{vec}^{-1}\,\phi$. This repartition may not be computationally demanding, but the calculation of ϕ involves operations containing the matrix L with dimension $(n^2 \times n^2)$.

Hence the small term $A_\tau \otimes A_\tau \frac{\Delta t^2}{4}$ is added into the inverse matrix in (8), which allows the algebraic inversion of the vec-operation. One obtains the mod-

ified increment matrix

$$\begin{aligned}\tilde{\Phi}(t,\Sigma_t,\Delta t) &= \mathrm{vec}^{-1}\left[\left(I - L(A_\tau)\frac{\Delta t}{2} + A_\tau \otimes A_\tau \frac{\Delta t^2}{4}\right)^{-1}\left(L(A_\tau)\sigma_t + \omega_\tau\right)\right] \\ &= \mathrm{vec}^{-1}\left[\left(I - A_\tau\frac{\Delta t}{2}\right)^{-1} \otimes \left(I - A_\tau\frac{\Delta t}{2}\right)^{-1} \mathrm{vec}\left[\ldots\right]\right] \\ &= \left(I - A_\tau\frac{\Delta t}{2}\right)^{-1}\left(A_\tau\Sigma_t + \Sigma_t A_\tau^T + \Omega_\tau\right)\left(I - A_\tau^T\frac{\Delta t}{2}\right)^{-1}.\end{aligned} \tag{9}$$

In (9) the special property of *Kronecker*-products $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ for arbitrary nonsingular matrices A and B was used. The asymptotic error, caused by the correction term, is of the same order as the approximation error and therefore does not compromise the order of convergence (see appendix proof A.2). The approximation rule for the covariance equation (2b) therefore reads

$$\boxed{\Sigma_{t+\Delta t} \approx \Sigma_t + M_\tau\left(A_\tau\Sigma_t + \Sigma_t A_\tau^T + \Omega_\tau\right)M_\tau^T\Delta t} \tag{10a}$$

with

$$\boxed{M_\tau = \left(I - A_\tau\frac{\Delta t}{2}\right)^{-1} \quad \text{and} \quad \tau = t + \frac{\Delta t}{2}.} \tag{10b}$$

The modified approximation formula is also A-stable, which can be shown by adding a small correction term $|\varepsilon| < 1$ into the denominator of the *Gauss-Legendre*-increment function for the linear test equation $\dot{y} = \lambda y$

$$y_{t+\Delta t} = y_t + \frac{\lambda y_t}{1 - \lambda\frac{\Delta t}{2} + \varepsilon}\Delta t.$$

The resulting stability function

$$H(\kappa) = \frac{1 + \frac{\kappa}{2} + \varepsilon}{1 - \frac{\kappa}{2} + \varepsilon}$$

shows that $|H(\kappa)| < 1$ for any finite $\kappa \in \mathbb{C}^-$ still holds. Thus the modification of the integration formula does not affect stability.

3.2. Interpolation of μ_τ

The predictor μ_τ at $\tau = t + \frac{\Delta t}{2}$, used in the *Jacobi*an A and diffusion matrix Ω of (10a) and (10b) can be computed from the *Taylor-Heun*-formula (4). However, this calculation requires the solution of a linear system. Thus, it seems more convenient to interpolate μ_τ from μ_t and $\mu_{t+\Delta t}$ with a precision of at least $\mathcal{O}(\Delta t^2)$.

By means of the series expansion used in proof A.2, one is able to calculate the difference between a simple polygon interpolation and the correct predictor

$$\frac{\mu_t + \mu_{t+\Delta t}}{2} - \mu_\tau = A(\mu_t)f(\mu_t)\frac{\Delta t^2}{8} + \mathcal{O}(\Delta t^3).$$

Hence, the half-step predictor can be approximated by

$$\boxed{\mu_\tau \approx \frac{1}{2}\left(\mu_t + \mu_{t+\Delta t} - A(\mu_t) f(\mu_t)\frac{\Delta t^2}{4}\right).} \tag{11}$$

4. Adaptive Step Size

The step size can be monitored to maintain a given tolerance level $e > 0$ during every integration step. Therefore a scalar estimate $\hat{e}$ for the maximum error, considering all components of the approximation error vector ϵ, is needed. In order to get useful estimates even for small errors, the maximum total-relative error

$$\boxed{\hat{e} = \max_{i=1,\ldots,n} \frac{|\epsilon^i|}{|\mu^i_{t+\Delta t}| + 1}} \tag{12}$$

is deployed, where the upper index i denotes the i-th component of the vector. To guarantee this upper bound for the approximation error the new step is calculated by multiplication with an adjustment factor $\Delta t_{new} = \varrho \Delta t$. Because the integration procedure is of order $\mathcal{O}(\Delta t^2)$,

$$\hat{e} = c\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{13}$$

holds asymptotically with constant $c > 0$. At the same time the next step has to stay inside the error bound too, therefore

$$e = c\Delta t_{new}^2 + \mathcal{O}(\Delta t_{new}^3) = c\varrho^2 \Delta t^2 + \mathcal{O}(\Delta t^3) \tag{14}$$

holds. Dividing (14) by (13) shows

$$\frac{e}{\hat{e}} = \varrho^2\big(1 + \mathcal{O}(\Delta t)\big),$$

which leads to a simple rule for the adjustment factor

$$\boxed{\varrho = \beta\sqrt{\frac{e}{\hat{e}}}} \tag{15}$$

if the term $\mathcal{O}(\Delta t)$ is neglected. The parameter β is inserted into (15) for practical purposes, in order to avoid frequent recalculations of the step size due to operations at maximum range. Reasonably it may be set to $\beta = 0.8$.

4.1. Approximation Error

The approximation error ϵ can be calculated from the series expansion derived in proof A.1. For the coefficient c_2 in (23b) one obtains

$$\begin{aligned} c_2(t) &= \frac{\ddot{f}(\mu_t)}{3} - \frac{A^2(\mu_t) f(\mu_t)}{2} \\ &= \frac{\dot{A}(\mu_t) f(\mu_t)}{3} + \frac{A^2(\mu_t) f(\mu_t)}{3} - \frac{A^2(\mu_t) f(\mu_t)}{2}. \end{aligned} \tag{16}$$

Replacing the time derivative of the *Jacobi*an in (16) by the appropriate finite difference for small Δt, the error can be approximated by

$$\boxed{\epsilon \approx \frac{\Delta t^2}{2}\left(\frac{A(\mu_{t+\Delta t}) - A(\mu_t)}{3\Delta t} - \frac{A^2(\mu_t)}{6}\right) f(\mu_t).} \tag{17}$$

Now the scalar error prediction used in equation (15) can be computed as maximum total-relative error (12).

4.2. Positive Semidefiniteness of State Error Covariance

Providing positive semidefiniteness of the state error covariance is difficult because usually the eigenvalues of Σ have to be monitored. In doing so, the spectral decomposition has to be calculated or the *Rayleigh*-quotient has to be minimized, which is very expensive.

An alternative approach is to monitor the decay of the determinant of the state error covariance, if the initial matrix Σ_0 is positive definit. This property can always be provided in practical applications by incorporating a small measurement error, e.g. $10^{-8}I$, which should be done generally to enhance numerical stability.

If B is an arbitrary positive definite matrix with determinant $|B|$, the derivative of the determinant is given by

$$d|B_t| = |B_t| \operatorname{Tr}\left[B_t^{-1} dB_t\right],$$

where Tr[...] denotes the trace of the matrix (a proof is given in Magnus and Neudecker 1988, p. 150). The differential can be approximated by a finite difference if the increment is sufficiently small

$$|B_t + \Phi_t \Delta t| - |B_t| \approx |B_t| \operatorname{Tr}\left[B_t^{-1}\Phi_t\right] \Delta t. \tag{18}$$

Taking the approximative character of (18) into account, it seems reasonable to restrict the step size to a radius around $|B_t|$ where the approximation can be trusted. This trust region should be scaled down as $|B_t|$ becomes smaller in order to prevent the determinant from passing through zero because of approximation errors. One possible choice, providing these features is

$$|B_t + \Phi_t \Delta t| = \frac{|B_t|}{a} \quad \text{for } 1 < a < \infty, \tag{19}$$

where for example $a = 2$ corresponds to the approximated half-life of the determinant. This trust region has proven useful in practical applications. A smaller value of a may enhance the certainty of maintaining a positive determinant, but may also increase the number of calculation steps dramatically. A relaxation of the boundary, which means a greater value of a, may save additional integration steps but also increases the risk of loosing the positive definiteness of B.

Inserting (19) into (18) and dividing by its right hand side results in an adjustment factor, delivering the maximum range of the following step, if it is multiplied with the current step size. Obviously, restrictions of the step size are

necessary only if the right hand side of (18) is negative. Assigning these results to the state error covariance formula (10a) and (10b) leads to the following condition

$$\boxed{\Delta t_{max} = \begin{cases} -\frac{1}{2\,\mathrm{Tr}\left[\Sigma_t^{-1}\Psi_t\right]} & \text{for Tr } [\ldots] < 0 \\ \infty & \text{else} \end{cases}} \tag{20a}$$

with

$$\boxed{\Psi_t = M_\tau(A_\tau\Sigma_t + \Sigma_t A_\tau^T + \Omega_\tau)M_\tau^T.} \tag{20b}$$

Now all necessary components (boxed equations) for algorithmic construction are available. Figure 1 shows a flowchart of the complete time update algorithm. Notice that the step size is tuned regarding the system state only. This is done for various reasons. First of all, the state ODE is nonlinear and therefore expected to show more rapid fluctuations than the linear state error covariance differential equation. Furthermore, the *Jacobi*an of the nonlinear drift vector acts as coefficient matrix in the linear state error covariance ODE. The temporal change and the squared value of this coefficient matrix are key elements in calculating the asymptotic error (17), which rules the integration step size. Thus major elements of the state error covariance equation are included indirectly in the step size calculation. Therefore, it is convenient to spare extra computational time needed for calculating asymptotical errors of the covariance equation.

5. Applications

This section presents two applications of the derived time update procedure for the extended *Kalman*-Filter. The system state and state error covariance of two commonly known models are calculated and compared to the exact solutions. The results are pooled graphically.

5.1. *Ornstein-Uhlenbeck*-Process

The *Ornstein-Uhlenbeck*-process is a linear second order ODE, which is sometimes used as interest rate or bond pricing model[2]. The vector autoregressive form reads

$$d\begin{pmatrix} y \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix}\begin{pmatrix} y \\ \dot{y} \end{pmatrix} dt + \begin{pmatrix} 0 \\ b \end{pmatrix} dt + \begin{pmatrix} 0 \\ g \end{pmatrix} dW(t)$$

with the *Wiener*-process $W(t)$. Because of the linear structure of the state equation this problem can be solved analytically. Figure 2 shows the exact and approximated time evolution of the system state estimate (upper row) and the state error covariance matrix for the parameter set $(\omega_0^2, \gamma, b, g) = (16, 2, 8, 2)$ and the initial conditions $\mu_0 = (0, 0)^T$ and $\Sigma_0 = \mathrm{diag}\,[0, 3]$. The individual

[2] In finance such models are known as *Vasicek*-models or mean reversion processes (Vasicek 1977, Jensen and Poulsen 2002, Mamon 2004).

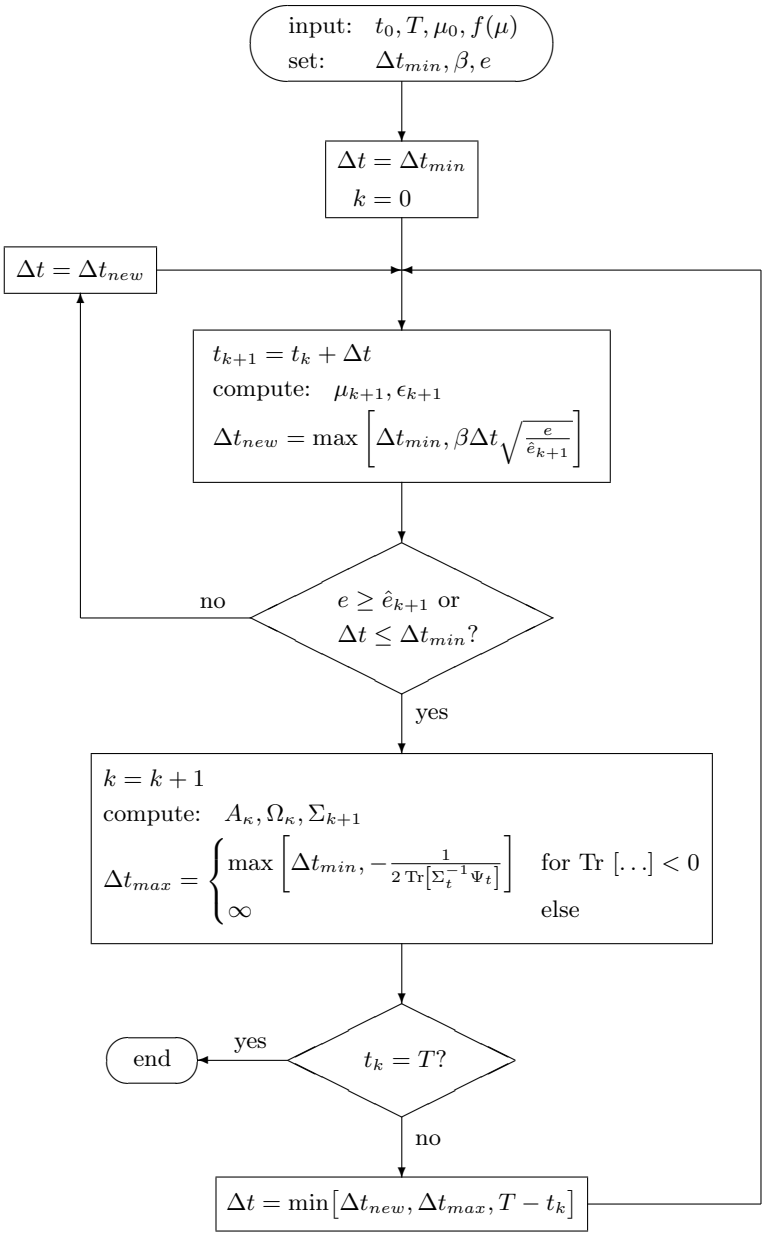

Figure 1: *Flowchart of EKF time update*

approximation steps are indicated by dots connected by a gray line, which is calculated via cubic spline interpolation. The exact solution is displayed as red line. The blue boundaries give the absolute-relative error band around the exact solution. In this example, the maximum error bound was chosen 10^{-2}, which means a relative error of one percent in the unit interval.

Obviously, the approximated solutions stay clearly inside the error bounds. A total of 59 integration steps were calculated, which accords to an average amount of 11.8 steps per unit interval.

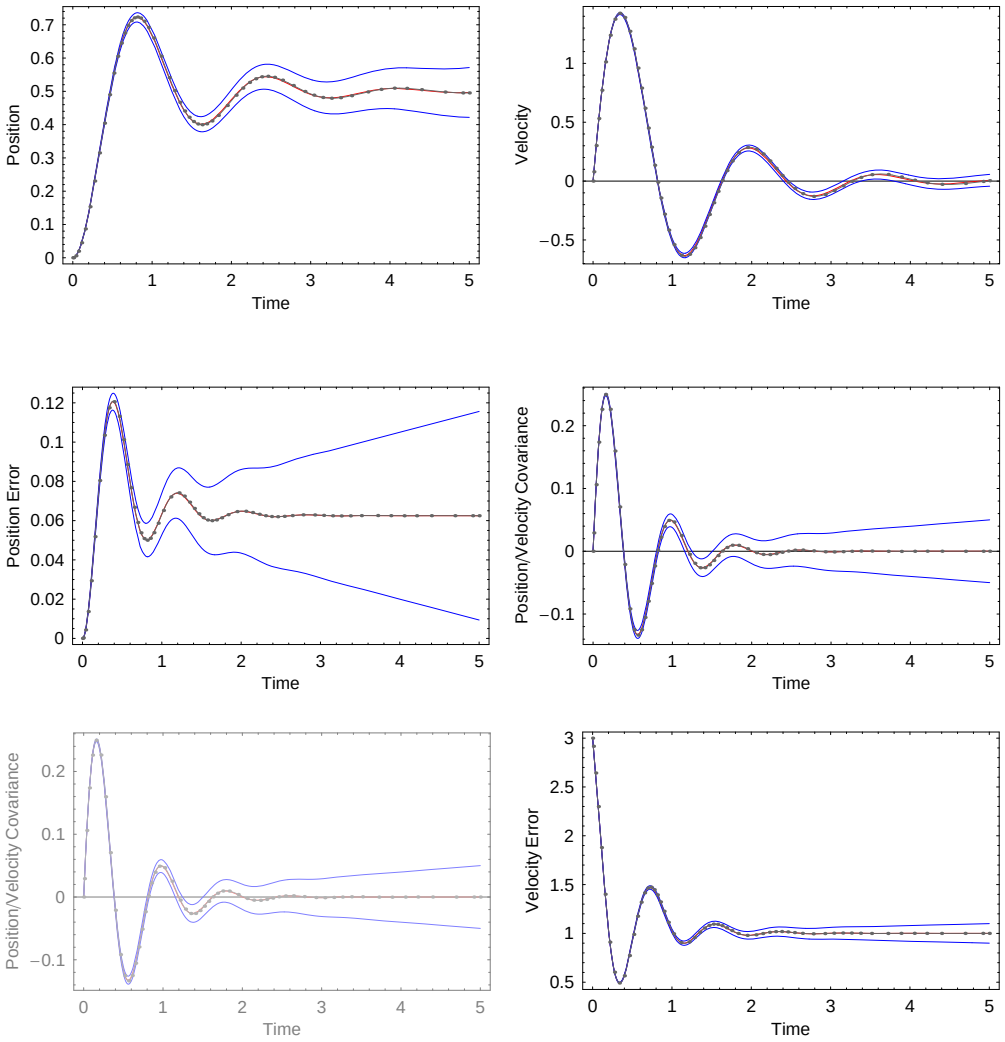

Figure 2: Ornstein-Uhlenbeck-*trajectory – state estimate and covariance matrix*

5.2. *Van der Pol* Oscillator

The *Van der Pol* oscillator is a simple model describing a stable limit cycle and therefore often used as benchmark or example for numerical computations (e.g. Sitz et al. 2002). Its vector autoregressive representation is

$$d\begin{pmatrix} y \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \dot{y} \\ \varepsilon(1-y^2)\dot{y} - y \end{pmatrix} dt + \begin{pmatrix} 0 \\ (1+y^2)g \end{pmatrix} dW(t),$$

where the stochastic influence is modeled state dependent, in order to provide full nonlinearity of the process. Because of the nonlinearity of the *Van der Pol* oscillator, solutions cannot be calculated analytically. Thus for reference, an implicit *Runge-Kutta*-scheme is used, providing numerical precision of 10^{-8}.

The process is started with parameters $(\varepsilon, g) = (1.5, 0.1)$ and initial conditions $\mu_0 = (0.5, 0.5)^T$ and $\Sigma_0 = \text{diag}\,[0, 0.1]$. The approximation again stays clearly inside the preliminary fixed error bound of 10^{-2}. Figure 3 is organized in

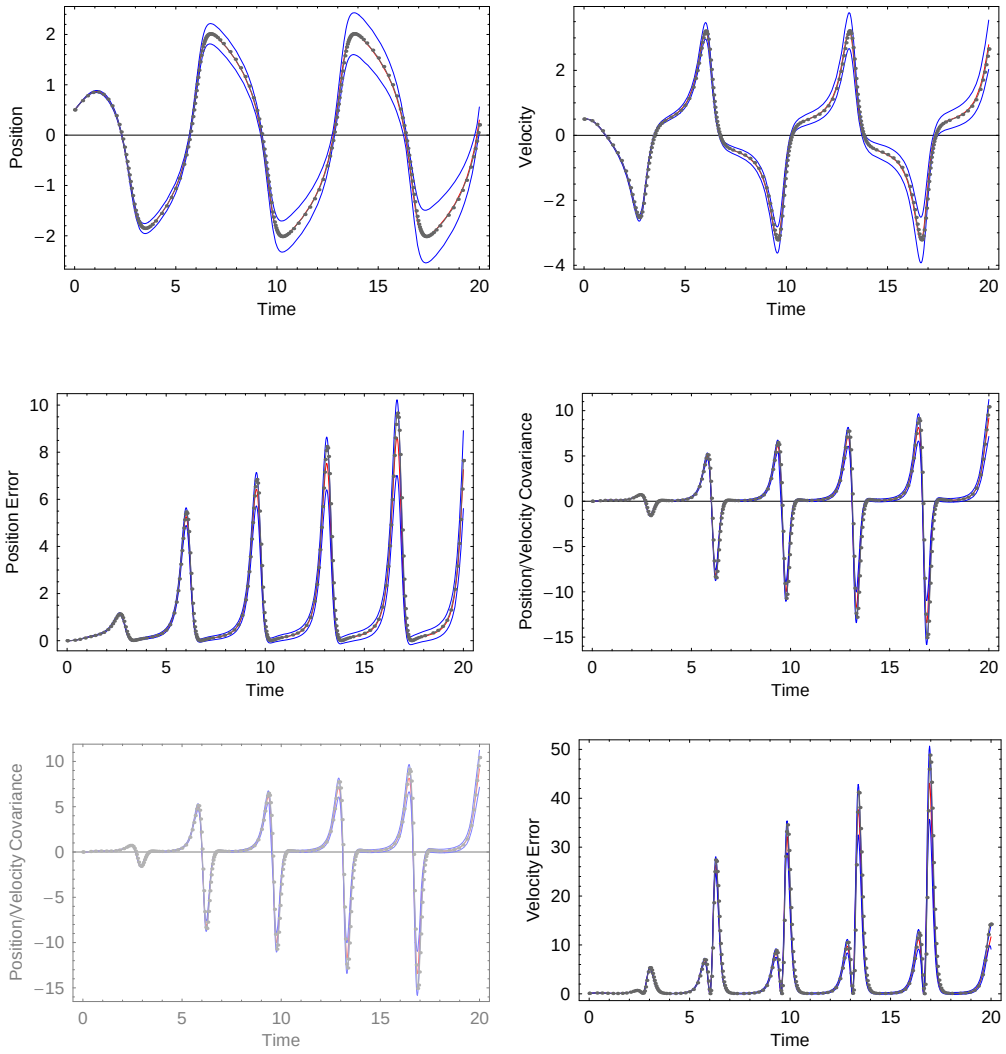

Figure 3: Van der Pol-*trajectory – state estimate and covariance matrix*

the same way as figure 2. A total of 221 integration steps were calculated, which is equivalent to an average of 11.05 steps per unit interval. This step count is a strong evidence for the efficiency of the step size calculating algorithm, because despite of significant nonlinearities the quantity of integration steps needed to provide a given precision is not larger than in the linear case.

6. Conclusions

A numerical integration scheme was developed specific to the structure of the continuous time moment equations of the extended *Kalman*-filter. This procedure features single step approximations of order $\mathcal{O}(\Delta t^2)$ for both, system state expectation and state error covariance. Additionally adaptive step size calculation is provided, assuring a preliminary fixed numerical precision with a minimum of integration steps. The derived algorithm is easy to implement, and

may serve as basis for more sophisticated integration procedures.

Two examples were presented, which confirm the stable and efficient impression of the suggested procedure. It was shown that both, linear and nonlinear problems can be treated without limitations. Further, all necessary proofs concerning stability and consistency were provided.

A. Proofs

A.1. Consistency of the *Taylor-Heun*-approximation

The general single step solution for an arbitrary integration problem reads

$$y_{t+\Delta t} = y_t + \Phi(t, y_t, \Delta t)\Delta t + \epsilon \Delta t \tag{21}$$

with the increment function Φ and the prediction error ϵ. In case of the *Taylor-Heun*-formula the increment function is

$$\Phi(t, \mu_t, \Delta t) = \left(I - A(\mu_t)\frac{\Delta t}{2} \right)^{-1} f(\mu_t). \tag{22}$$

By algebraic manipulation and *Taylor*-expansion of (21) one obtains an explicit equation for the approximation error

$$\begin{aligned}
\epsilon &= \frac{1}{\Delta t}\big(y_{t+\Delta t} - y_t\big) - \Phi(t, y_t, \Delta t) \\
&= \frac{1}{\Delta t}\left(\sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} y_t^{(k)} - y_t \right) - \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} \frac{\partial^k}{\partial \Delta t^k} \Phi(t, y_t, 0) \\
&= \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} \left(\frac{1}{k+1} y_t^{(k+1)} - \frac{\partial^k}{\partial \Delta t^k} \Phi(t, y_t, 0) \right) \\
&= \sum_{k=0}^{q-1} \frac{\Delta t^k}{k!} c_k(t) + \mathcal{O}(\Delta t^q)
\end{aligned} \tag{23a}$$

with the coefficient functions

$$c_k(t) = \frac{1}{k+1} y_t^{(k+1)} - \frac{\partial^k}{\partial \Delta t^k} \Phi(t, y_t, 0) \quad \text{for} \quad k = 0, \ldots, q-1. \tag{23b}$$

Obviously, the approximation is consistent with order $\mathcal{O}(\Delta t^q)$, if the coefficients $c_k(t)$ vanish for $k = 0, \ldots, q-1$. For the *Taylor-Heun*-increment function (22) one obtains

$$\begin{aligned}
c_0(t) &= \dot{\mu}_t - f(\mu_t) = 0 \\
c_1(t) &= \frac{\ddot{\mu}_t}{2} - \frac{A(\mu_t) f(\mu_t)}{2} = \frac{\dot{f}(\mu_t)}{2} - \frac{A(\mu_t)\dot{\mu}_t}{2} = 0 \\
c_2(t) &= \frac{\ddot{f}(\mu_t)}{3} - \frac{A^2(\mu_t) f(\mu_t)}{2} \neq 0.
\end{aligned}$$

Thus the *Taylor-Heun*-scheme is consistent with order $\mathcal{O}(\Delta t^2)$. $\square$

A.2. Asymptotic Error of the Increment Correction

In this proof the series expansion

$$(I-B)^{-1}=\sum_{k=0}^{\infty}B^k$$

for an arbitrary quadratic matrix B with $\|B\|<1$ is used. The *Jacobi*an A of the vector field f, given in equation (1) is bounded because of the *Lipschitz*-condition if the existence of a unique solution is presumed. Therefore, it is sufficient to assume Δt small enough to guarantee $\|L(A_\tau\frac{\Delta t}{2})\|<1$, which complies with the requirements of the series expansion. Thus, the inverse matrix of the original increment function can be expanded to

$$\begin{aligned}\Big(I-L\Big(A_\tau\frac{\Delta t}{2}\Big)\Big)^{-1}&=\sum_{k=0}^{\infty}\left(A_\tau\frac{\Delta t}{2}\otimes I+I\otimes A_\tau\frac{\Delta t}{2}\right)^k\\&=\sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}(A_\tau\otimes I)^{k-j}(I\otimes A_\tau)^j\left(\frac{\Delta t}{2}\right)^k\\&=\sum_{k=0}^{\infty}\sum_{j=0}^{k}\binom{k}{j}A_\tau^{k-j}\otimes A_\tau^j\left(\frac{\Delta t}{2}\right)^k.\end{aligned}\tag{24}$$

On the other hand, the modified inverse can also be expanded into

$$\begin{aligned}\left(I-A_\tau\frac{\Delta t}{2}\right)^{-1}\otimes\left(I-A_\tau\frac{\Delta t}{2}\right)^{-1}&=\sum_{k=0}^{\infty}\left(A_\tau\frac{\Delta t}{2}\right)^k\otimes\sum_{j=0}^{\infty}\left(A_\tau\frac{\Delta t}{2}\right)^j\\&=\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}A_\tau^k\otimes A_\tau^j\left(\frac{\Delta t}{2}\right)^{k+j}.\end{aligned}\tag{25}$$

Inserting (24) and (25) into the appropriate vectorized increment function and calculating the difference shows

$$\phi-\tilde{\phi}=(A_\tau\otimes A_\tau)\big(L(A_\tau)\sigma_t+\omega_\tau\big)\frac{\Delta t^2}{4}+\mathcal{O}(\Delta t^3).\tag{26}$$

Result (26) proves that the asymptotic error, caused by the correction term, is the same order as the approximation error of the *Gauss-Legendre*-formula. Therefore the consistence of the procedure is not downgraded. □

References

Arnold, V.I. (2001): *Gewöhnliche Differentialgleichungen*. Springer, Berlin, Heidelberg, New York, 2nd edn.

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

Ito, K. and K. Xiong (2000): Gaussian Filters for Nonlinear Filtering Problems. *IEEE Transactions on Automatic Control*, 45(5):910–927.

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

Jensen, B. and R. Poulsen (2002): Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques. *Journal of Derivatives*, 9(4):18–32.

Johnston, L.A. and V. Krishnamurthy (2001): Derivation of a Sawtooth Iterated Extended Kalman Smoother via the AECM Algorithm. *IEEE Transaction on Signal Processing*, 49(9):1899–1909.

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

Julier, S.; J. Uhlmann; and H.F. Durrant-White (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.E. (1960): A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82 (Series D):35–45.

Kloeden, P.E. and E. Platen (1992): *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, Heidelberg, New York, 3rd edn.

Lefebvre, T.; H. Bruyninckx; and J. de Schutter (2004): Kalman Filters for Non-Linear Systems: A Comparison of Performance. *International Journal on Control*, 77(7):639–653.

Magnus, J.R. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York, Brisbane, Toronto.

Mamon, R.S. (2004): Three Ways to Solve for Bond Prices in the Vasicek Model. *Journal of Applied Mathematics and Decision Sciences*, 8(1):1–14.

Mazzoni, T. (2007): *Stetig/diskrete Zustandsraummodelle dynamischer Wirtschaftsprozesse*. Shaker, Aachen.

Nørgaard, M.; N.K. Poulsen; and O. Ravn (2000): New Developments in State Estimation for Nonlinear Systems. *Automatica*, 36:1627–1638.

Rosenbrock, H.H. (1963): Some General Implicit Processes for the Numerical Solution of Differential Equations. *The Computer Journal*, 5(4):329–330.

Schmidt, S.F. (1966): Application of State-Space Methods to Navigation Problems. In: *Advances in Control Systems. Theory and Applications*, ed. C.T. Leondes, Academic Press, New York, San Francisco, London, vol. 3, pp. 293–340.

Schweppe, F. (1965): Evaluation of Likelihood Functions for Gaussian Signals. *IEEE Transactions on Information Theory*, 11:61–70.

Singer, H. (2006a): *Continuous-Discrete Unscented Kalman Filtering*. Tech. Rep. 384, FernUniversität in Hagen.

Singer, H. (2006b): Stochastic Differential Equation Models with Sampled Data. In: *Longitudinal Models in the Behavioral and Related Sciences*, eds. K. van Montfort; J. Oud; and A. Satorra, Lawrence Erlbaum Associates, London, chap. 4, pp. 73–106.

Sitz, A.; U. Schwarz; J. Kurths; and H.U. Voss (2002): Estimation of Parameters and Unobserved Components for Nonlinear Systems from Noisy Time Series. *Physical Review E*, 66(016210):1–9.

Tanizaki, H. (1996): *Nonlinear Filters. Estimation and Applications*. Springer, Berlin, Heidelberg, New York, 2nd edn.

Vasicek, O. (1977): An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

Wanner, G. (2003): Dahlquist's Classical Papers on Stability Theory. *BIT Numerical Mathematics*, 43(1):1–18.