Mathematical Biosciences 251 (2014) 54–62

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements

Jacob Leander [a,b,*], Torbjörn Lundh [b], Mats Jirstrand [a]

[a] Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Gothenburg, Sweden
[b] Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 27 September 2013
Received in revised form 28 February 2014
Accepted 1 March 2014
Available online 12 March 2014

Keywords:
Parameter estimation
Ordinary differential equations
Stochastic differential equations
Extended Kalman filter
Lotka–Volterra
FitzHugh–Nagumo

ABSTRACT

In this paper we consider the problem of estimating parameters in ordinary differential equations given discrete time experimental data. The impact of going from an ordinary to a stochastic differential equation setting is investigated as a tool to overcome the problem of local minima in the objective function. Using two different models, it is demonstrated that by allowing noise in the underlying model itself, the objective functions to be minimized in the parameter estimation procedures are regularized in the sense that the number of local minima is reduced and better convergence is achieved. The advantage of using stochastic differential equations is that the actual states in the model are predicted from data and this will allow the prediction to stay close to data even when the parameters in the model is incorrect. The extended Kalman filter is used as a state estimator and sensitivity equations are provided to give an accurate calculation of the gradient of the objective function. The method is illustrated using *in silico* data from the FitzHugh–Nagumo model for excitable media and the Lotka–Volterra predator–prey system. The proposed method performs well on the models considered, and is able to regularize the objective function in both models. This leads to parameter estimation problems with fewer local minima which can be solved by efficient gradient-based methods.

1. Introduction

This paper concerns the problem of estimating parameters in dynamical systems described by ordinary differential equations given discrete time measurement data. Dynamical systems and ordinary differential equations (ODEs) are applicable in a large number of areas, for example biology, medicine, aerospace, and process engineering. For an introduction to dynamical systems described by ordinary differential equations, see for example [1].

Estimation of parameters in ordinary differential equations given discrete time measurement data is a complex problem, which has been addressed by several authors in many different fields. Given a model structure and discrete time measurement data we are interested in identifying the values of the parameters in the model that best agree with observed data. The definition of deviation from the model can differ in various ways. Schittkowski [2] uses an geometric approach where the goal is to minimize the quadratic distances between model and data while Bohlin and Graebe [3], Särkää [4] and Kristensen et al. [5] use a probabilistic approach. Throughout this paper, we will utilize the probabilistic approach.

As discussed by Schittkowski [2], there are a number of possible difficulties regarding parameter estimation in dynamical systems. These include convergence to local minima, flat objective functions and non-differentiable terms in the system dynamics. Existing methods for parameter estimation in dynamical systems include least-square methods, multiple shooting methods [6–8], stochastic methods [9], and hybrid methods [10,11]. For a review on the parameter estimation problem in biological systems, see [12].

We propose a method observed to regularize the objective function used for parameter estimation in dynamical systems. The introduction of uncertainty in the underlying model can be utilized to decrease the number of local minima, which in turn leads to a less complex optimization problem that can be solved by efficient gradient-based methods. To regularize the objective function and overcome the problem of local minima, the introduction of noise in the differential equations describing the dynamics is considered. By extending the deterministic model to a stochastic model the

* Corresponding author at: Fraunhofer-Chalmers Centre, Chalmers Science Park, SE-412 88 Gothenburg, Sweden. Tel.: +46 31 7724297.
E-mail address: jacob.leander@fcc.chalmers.se (J. Leander).

http://dx.doi.org/10.1016/j.mbs.2014.03.001
0025-5564/

objective function may be better suited for a gradient-based search method. The reason for this is that when the underlying model is described by a stochastic differential equation the state and its covariance are updated at each measurement. This will in turn lead to that the model predictions will stay closer to the measurements even when the parameters values are far from being optimal.

Stochastic differential equations have received great attention in a large number of fields, including finance, pharmaceutics, and systems biology. For an introduction to the theory and numerical solution of stochastic differential equations, see [13–16]. Stochastic differential equations serve as a natural way of introducing uncertainty into a deterministic model. In contrast to the classical approach where uncertainty only exists in the measurements, stochastic differential equations can provide a more flexible framework to account for deviations in states and parameters that describe the underlying system.

Parameter estimation in stochastic differential equations is an area where several methods are available, as reviewed in [17]. For applications to pharmacokinetic and pharmacodynamic models, see [18,19]. There are several available software tools for parameter estimation in stochastic differential equations. Bohlin and Graebe [3] presented a parameter estimation scheme and an associated software named IdKit, further developed into a more sophisticated tool named MoCaVa [20]. Another estimation software tool is CTSM (Continuous Time Stochastic Modelling) [5,21], developed at DTU Compute.

However, the observation that the introduction of noise in the system dynamics provides a mean to regularize the optimization problem associated with the parameter estimation in ordinary differential equations seems to have been unnoticed. This paper illustrates this fact using two different models, the FitzHugh–Nagumo model describing excitable media [22,23] and the Lotka–Volterra predator–prey system. A gradient-based search method is proposed, using the extended Kalman filter as a state estimator. In addition, sensitivity equations for the underlying system and the filter updating equations are derived, which are used for an accurate gradient calculation.

2. Modeling of dynamical systems

Consider a continuous dynamical system described by a set of n potentially nonlinear ordinary differential equations

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, t, \boldsymbol{\theta}), \quad \mathbf{x}(t_0) = \mathbf{x}_0(\boldsymbol{\theta}), \tag{1}$$

where $t \in \mathbb{R}$ denotes time, $\mathbf{x}_t \in \mathbb{R}^n$ is the state vector, $\mathbf{u}_t \in \mathbb{R}^m$ is the input, and $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector. The function $\mathbf{f}$ is a function describing the underlying dynamics. Furthermore, the system is sampled at discrete time points $t_k,\ k = 1, \ldots, N$ under Gaussian white noise according to

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_{t_k}, \mathbf{u}_{t_k}, t_k, \boldsymbol{\theta}) + \mathbf{e}_k,\ k = 1, \ldots, N, \tag{2}$$

where $\mathbf{y}_k \in \mathbb{R}^l$ is the vector of output variables at discrete time point t_k and $\mathbf{h}$ is a function describing the measurement structure. Here $\mathbf{e}_k \in \mathbb{R}^l$ is a l-dimensional white noise process with $\mathbf{e}_k \sim N(0, \mathbf{S}(\mathbf{u}_{t_k}, t_k, \boldsymbol{\theta}))$.

In the model above the uncertainty is introduced through the measurement Eq. (2). However, this assumption may not always be sufficient. Such situations may arise if the underlying system includes stochastic parts, if the model fails to capture the true dynamics, or if some of the parameters in the model are uncertain.

By extending the ordinary differential Eq. (1) to a stochastic differential equation (SDE) the dynamical system includes a stochastic part which we refer to as the system noise (also known as diffusion). The system noise serves as a tool to account for all the unknown phenomena which are not captured by the deterministic model, for example approximations, modeling errors and oversimplifications. Hence the noise in the model is divided into two parts, measurement noise and system noise. The stochastic differential equation model written on differential form is

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\Sigma}(\mathbf{x}_t, \mathbf{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t, \quad \mathbf{x}(t_0) = \mathbf{x}_0(\boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\Sigma}(\mathbf{x}_t, \mathbf{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t$ is the system noise with $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times q}$ and $\boldsymbol{\omega} \in \mathbb{R}^q$ is a standard q-dimensional Wiener process, also known as Brownian motion. Note that $\boldsymbol{\Sigma} = \mathbf{0}$ corresponds to the initial model (1).

The solution $\mathbf{x}$ to (3) is a stochastic process, which is described by its transition probability density $\mathbf{p}(\mathbf{x}, t; \mathbf{y}, \tau)$. The transition probability density $\mathbf{p}(\mathbf{x}, t; \mathbf{y}, \tau)$ is given by the solution to the Kolmogorov-Forward partial differential equation (with dependence on $\mathbf{u}_t$ and $\boldsymbol{\theta}$ omitted)

$$\frac{\partial \mathbf{p}(\mathbf{x}, t; \mathbf{y}, \tau)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(\mathbf{p}(\mathbf{x}, t; \mathbf{y}, \tau) f_i(\mathbf{x}, t)) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left(\mathbf{p}(\mathbf{x}, t; \mathbf{y}, \tau)(\boldsymbol{\Sigma}(\mathbf{x}, t)\boldsymbol{\Sigma}^T(\mathbf{x}, t))_{ij}\right), \tag{4}$$

which has no closed form solution, except for a few special cases. The interested reader is referred to [13,14].

3. Parameter estimation in stochastic differential equations

Given a parameterized model and a set of measurements, we seek the parameter values $\boldsymbol{\theta}$ that give a model in good agreement with observed data. This is commonly referred to as the parameter estimation problem. As mentioned in the introduction, we need to define a measure of deviation from the model.

Due to the random elements in our underlying model we will utilize the probabilistic approach where we work in a maximum likelihood setting. Hence the goal is to find the parameters that maximize the likelihood function of a given sequence of measurements. To put the objective in a formal way; given a sequence of measurements $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k, \ldots, \mathbf{y}_N$, let

$$\mathcal{Y}_k \triangleq \{\mathbf{y}_k, \ldots, \mathbf{y}_1\} \tag{5}$$

denote the collection of measurements up to time point t_k. The likelihood function is defined as

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) \triangleq p(\mathcal{Y}_N|\boldsymbol{\theta}), \tag{6}$$

which is a function of the parameters $\boldsymbol{\theta}$. By repeated use of Bayes law, $P(A \cap B) = P(A|B)P(B)$, the likelihood function can be rewritten as

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = \left(\prod_{k=2}^{N} p(\mathbf{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta})\right) p(\mathbf{y}_1|\boldsymbol{\theta}). \tag{7}$$

An exact evaluation of the likelihood function is in general computationally infeasible since it requires solving the Kolmogorov-Forward partial differential Eq. (4). Since the differential equations are driven by Wiener processes and the increments of a Wiener process are Gaussian, it may be reasonable to assume that the conditional densities can be approximated reasonably well by Gaussian densities [21], which are characterized by their mean and covariance. We introduce the notation

$$\hat{\mathbf{y}}_{k|k-1} \triangleq \mathrm{E}\{\mathbf{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}\} \tag{8}$$

$$\mathbf{R}_{k|k-1} \triangleq \mathrm{Var}\{\mathbf{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}\}. \tag{9}$$

For a Gaussian random variable with mean $\hat{\mathbf{y}}_{k|k-1}$ and covariance $\mathbf{R}_{k|k-1}$ the likelihood function is

$$L(\boldsymbol{\theta}; \mathcal{Y}_N) = \left(\prod_{k=2}^{N} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_k^T \mathbf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k\right)}{\sqrt{\det(\mathbf{R}_{k|k-1})}(\sqrt{2\pi})^l}\right) p(\mathbf{y}_1|\boldsymbol{\theta}), \tag{10}$$

where $\boldsymbol{\epsilon}_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}$. Taking the negative logarithm gives the negative log likelihood

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \mathcal{Y}_N) = -\ln(L(\boldsymbol{\theta}; \mathcal{Y}_N)$$
$$= \frac{1}{2}\sum_{k=1}^{N}\Big(\ln(\det(\mathbf{R}_{k|k-1})) + \boldsymbol{\epsilon}_k^T \mathbf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon} + l\ln(2\pi)\Big), \tag{11}$$

which is to be minimized with respect to the parameter vector $\boldsymbol{\theta}$. Here $\mathbf{R}_{1|0}$ corresponds to the initial covariance. Since the objective function depends nonlinearly on the parameters $\boldsymbol{\theta}$ there may be several local minima, which can introduce severe problems in the optimization routine. The aim of this paper is to demonstrate the impact of the introduction of system noise on the presence of local minima in the objective function (11).

4. Extended Kalman filter

When the underlying model is described by a stochastic differential equation the states can change randomly due to the random fluctuations in the Wiener process. Given measurements and the underlying structure the state and covariance of the system have to be estimated in order to compute the residuals $\boldsymbol{\epsilon}_k$ and output covariance $\mathbf{R}_{k|k-1}$. To do this, we make use of the extended Kalman filter (EKF), which is a state estimator in nonlinear continuous-discrete state space models [24] of the form

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\Sigma}(\mathbf{u}_t, t, \boldsymbol{\theta})d\boldsymbol{\omega}_t, \quad \mathbf{x}(t_0) = \mathbf{x}_0(\boldsymbol{\theta}),$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, t_k, \boldsymbol{\theta}) + \mathbf{e}_k.$$

The EKF is an extension of the Kalman filter [25] to nonlinear models. Note that there is no state-dependence in the expression for the system noise $\boldsymbol{\Sigma}$. When there is state-dependent diffusion the EKF may fail and a transformation, known as the Lamperti transform, of the stochastic differential equation may be necessary. For applications of the Lamperti transform to stochastic differential equations with state-dependent diffusion, see [26].

When the dynamics is linear, the Kalman filter provides an optimal state estimator for a given parameter vector $\boldsymbol{\theta}$. For nonlinear models the EKF uses a first order linear approximation of the model. The EKF provides estimates of the conditional expectation of the state $\hat{\mathbf{x}}_{k|k} = \mathrm{E}\{\mathbf{x}_{t_k}|\mathcal{Y}_k, \boldsymbol{\theta}\}$ and its covariance $\mathbf{P}_{k|k} = \mathrm{Var}\{\mathbf{x}_{t_k}|\mathcal{Y}_k, \boldsymbol{\theta}\}$. Following the notation of [21,27] the scheme is as follows. Given initial conditions $\hat{\mathbf{x}}_{1|0} = \mathbf{x}_0$ and $\mathbf{P}_{1|0} = \mathbf{P}_0$ and linearizations

$$\mathbf{A}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{x}_t}\Big|_{\mathbf{x}_t = \hat{\mathbf{x}}_{t|k}} \tag{12}$$

$$\mathbf{C}_k = \frac{\partial \mathbf{h}}{\partial \mathbf{x}_t}\Big|_{\mathbf{x}_t = \hat{\mathbf{x}}_{k|k-1}} \tag{13}$$

the state and state covariance are predicted between two consecutive measurement time points according to

$$\frac{d\hat{\mathbf{x}}_{t|k}}{dt} = \mathbf{f}(\hat{\mathbf{x}}_{t|k}, \mathbf{u}_t, t, \boldsymbol{\theta}), \ t \in [t_k, t_{k+1}], \tag{14}$$

$$\frac{d\mathbf{P}_{t|k}}{dt} = \mathbf{A}_t\mathbf{P}_{t|k} + \mathbf{P}_{t|k}\mathbf{A}_t^T + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T, \ t \in [t_k, t_{k+1}]. \tag{15}$$

From the predicted state and state covariance we get the output prediction equations

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}, \mathbf{u}_k, t_k, \boldsymbol{\theta}), \tag{16}$$

$$\mathbf{R}_{k|k-1} = \mathbf{C}_k\mathbf{P}_{k|k-1}\mathbf{C}_k^T + \mathbf{S}. \tag{17}$$

From the state covariance $\mathbf{P}_{k|k-1}$ and measurement covariance $\mathbf{R}_{k|k-1}$ the Kalman gain is given by

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{C}_k^T\mathbf{R}_{k|k-1}^{-1}. \tag{18}$$

Finally the state and its covariance are updated according to

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\boldsymbol{\epsilon}_k, \tag{19}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{R}_{k|k-1}\mathbf{K}_k^T, \tag{20}$$

where the residual $\boldsymbol{\epsilon}_k$ is given by

$$\boldsymbol{\epsilon}_k = \mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}. \tag{21}$$

Note that the Kalman gain $\mathbf{K}_k$ is a combination of the state covariance and output covariance. If there is no system noise at all ($\boldsymbol{\Sigma} = \mathbf{0}$), we trust the model completely. If there is no measurement variance at all ($\mathbf{S} = \mathbf{0}$), we trust the measurements completely. Depending on the relation between the covariances, the updated state prediction is more or less influenced by the measurement. This will in turn lead to that the model prediction is closer to the actual measurements. This is the core and strength of the EKF and we refer the interested reader to [24]. Pseudocode for the EKF is given in Algorithm 1.

Algorithm 1. Extended Kalman Filtering

Given a parameter vector $\boldsymbol{\theta}$ and initial state $\mathbf{x}_0$ with covariance $\mathbf{P}_0$
for $k = 1$ to N **do**
  Predict state and covariance with (14) and (15)
  Predict output and output covariance with (16) and (17)
  Compute Kalman gain with (18)
  Update state and covariance with (19) and (20)
end for
Return $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ for $k = 1, \ldots, N$

5. Differentiation of the extended Kalman filter equations

We are considering the problem of finding a minimum of the nonlinear objective function $l(\boldsymbol{\theta})$. When the underlying system is described by a system of ordinary differential equations there may be problems with gradient-based methods due to the existence of local minima in the objective function [2]. However, in this paper we demonstrate how the objective function can be regularized to obtain an objective function with fewer or no local minima. In the case of no local minima the objective function can be optimized by efficient gradient-based methods to achieve global convergence.

The optimization method used in this paper is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method, which is a Quasi-Newton optimization algorithm. BFGS is a robust, well-established method for gradient-based optimization and we omit the details of the algorithm for the sake of simplicity. For information about BFGS, see for example [28].

Since we consider a gradient-based optimization method we need to calculate the gradient $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. A common choice for calculating the gradient is to approximate it by finite differences, something that is utilized in CTSM [21]. In this paper we instead derive the analytical expression of the gradient based on differentiation of the underlying system of equations.

Differentiation of the objective function (11) with respect to the ith component of the parameter vector $\boldsymbol{\theta}$ gives the ith component of the gradient according to

$$\frac{dl(\boldsymbol{\theta})}{d\theta_i} = \frac{1}{2}\sum_{k=1}^{N}\left(\frac{d\boldsymbol{\epsilon}_k^T}{d\theta_i}\mathbf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k + \boldsymbol{\epsilon}_k^T\frac{d\mathbf{R}_{k|k-1}^{-1}}{d\theta_i}\boldsymbol{\epsilon}_k + \boldsymbol{\epsilon}_k^T\mathbf{R}_{k|k-1}^{-1}\frac{d\boldsymbol{\epsilon}_k}{d\theta_i} + \frac{d\ln(\det(\mathbf{R}_{k|k-1}))}{d\theta_i}\right). \tag{22}$$

Using the fact that

$$\frac{d\mathbf{R}^{-1}}{d\theta_i} = -\mathbf{R}^{-1}\frac{d\mathbf{R}}{d\theta_i}\mathbf{R}^{-1} \tag{23}$$

together with

$$\frac{d\ln(\det(\mathbf{R}_{k|k-1}))}{d\theta_i} = \mathrm{Tr}\left(\mathbf{R}_{k|k-1}^{-1}\frac{d\mathbf{R}_{k|k-1}}{d\theta_i}\right), \tag{24}$$

the final expression for the ith component $\frac{dl(\boldsymbol{\theta})}{d\theta_i}$ of the gradient is

$$\frac{dl(\boldsymbol{\theta})}{d\theta_i} = \frac{1}{2}\sum_{k=1}^{N}\left(\frac{d\boldsymbol{\epsilon}_k^T}{d\theta_i}\mathbf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_k^T\mathbf{R}_{k|k-1}^{-1}\frac{d\mathbf{R}_{k|k-1}}{d\theta_i}\mathbf{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k + \boldsymbol{\epsilon}_k^T\mathbf{R}_{k|k-1}^{-1}\frac{d\boldsymbol{\epsilon}_k}{d\theta_i} + \mathrm{Tr}\left(\mathbf{R}_{k|k-1}^{-1}\frac{d\mathbf{R}_{k|k-1}}{d\theta_i}\right)\right). \tag{25}$$

To calculate this gradient, the partial derivatives $\frac{d\boldsymbol{\epsilon}_k}{d\theta_i}$ and $\frac{d\mathbf{R}_{k|k-1}}{d\theta_i}$ are needed. They are obtained from the sensitivity analysis of the EKF, which has been done in [4]. In this paper the model is extended to allow for parameters in the output function $\mathbf{h}$ and measurement covariance $\mathbf{S}$. This has been utilized in [29,30].

Differentiation of the state predictions Eqs. (14) and (15) yields the sensitivity equations. For convenience, the dependence on the state, input, time, and parameters has been omitted below.

$$\frac{d}{dt}\frac{d\hat{\mathbf{x}}_{t|k}}{d\theta_i} = \frac{\partial\mathbf{f}}{\partial\hat{\mathbf{x}}_{t|k}}\frac{d\hat{\mathbf{x}}_{t|k}}{d\theta_i} + \frac{\partial\mathbf{f}}{\partial\theta_i}, \quad t\in[t_k, t_{k+1}], \tag{26}$$

$$\frac{d}{dt}\frac{d\mathbf{P}_{t|k}}{d\theta_i} = \frac{d\mathbf{A}_t}{d\theta_i}\mathbf{P}_{t|k} + \mathbf{A}_t\frac{d\mathbf{P}_{t|k}}{d\theta_i} + \frac{d\mathbf{P}_{t|k}}{d\theta_i}\mathbf{A}_t^T + \mathbf{P}_{t|k}\frac{d\mathbf{A}_t^T}{d\theta_i} + \frac{d\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T}{d\theta_i}, \quad t\in[t_k, t_{k+1}]. \tag{27}$$

In the same fashion we have the derivative of the output prediction Eqs. (16) and (17)

$$\frac{d\hat{\mathbf{y}}_{k|k-1}}{d\theta_i} = \frac{\partial\mathbf{h}}{\partial\hat{\mathbf{x}}_{k|k-1}}\frac{d\hat{\mathbf{x}}_{k|k-1}}{d\theta_i} + \frac{\partial\mathbf{h}}{\partial\theta_i}, \tag{28}$$

$$\frac{d\mathbf{R}_{k|k-1}}{d\theta_i} = \frac{d\mathbf{C}_k}{d\theta_i}\mathbf{P}_{k|k-1}\mathbf{C}_k^T + \mathbf{C}_k\frac{d\mathbf{P}_{k|k-1}}{d\theta_i}\mathbf{C}_k^T + \mathbf{C}_k\mathbf{P}_{k|k-1}\frac{d\mathbf{C}_k^T}{d\theta_i} + \frac{d\mathbf{S}}{d\theta_i}. \tag{29}$$

Differentiation of the Kalman gain Eq. (18) and the residual Eq. (12) gives

$$\frac{d\mathbf{K}_k}{d\theta_i} = \frac{d\mathbf{P}_{k|k-1}}{d\theta_i}\mathbf{C}_k^T\mathbf{R}_{k|k-1}^{-1} + \mathbf{P}_{k|k-1}\frac{d\mathbf{C}_k^T}{d\theta_i}\mathbf{R}_{k|k-1}^{-1} + \mathbf{P}_{k|k-1}\mathbf{C}_k^T\frac{d\mathbf{R}_{k|k-1}^{-1}}{d\theta_i}, \tag{30}$$

$$\frac{d\boldsymbol{\epsilon}_k}{d\theta_i} = -\frac{d\hat{\mathbf{y}}_{k|k-1}}{d\theta_i}. \tag{31}$$

At last the updating Eqs. (19) and (20) are differentiated according to

$$\frac{d\hat{\mathbf{x}}_{k|k}}{d\theta_i} = \frac{d\hat{\mathbf{x}}_{k|k-1}}{d\theta_i} + \frac{d\mathbf{K}_k}{d\theta_i}\boldsymbol{\epsilon}_k + \mathbf{K}_k\frac{d\boldsymbol{\epsilon}_k}{d\theta_i}, \tag{32}$$

$$\frac{d\mathbf{P}_{k|k}}{d\theta_i} = \frac{d\mathbf{P}_{k|k-1}}{d\theta_i} - \frac{d\mathbf{K}_k}{d\theta_i}\mathbf{R}_{k|k-1}\mathbf{K}_k^T - \mathbf{K}_k\frac{d\mathbf{R}_{k|k-1}}{d\theta_i}\mathbf{K}_k^T - \mathbf{K}_k\mathbf{R}_{k|k-1}\frac{d\mathbf{K}_k^T}{d\theta_i}. \tag{33}$$

In the general case $\mathbf{A}_t = \mathbf{A}_t(\hat{\mathbf{x}}_{t|k}, \boldsymbol{\theta})$ and $\mathbf{C}_k = \mathbf{C}_k(\hat{\mathbf{x}}_{t|k}, \boldsymbol{\theta})$ which implies that

$$\frac{d\mathbf{A}_t}{d\theta_i} = \frac{\partial\mathbf{A}_t}{\partial\hat{\mathbf{x}}_{t|k}}\frac{d\hat{\mathbf{x}}_{t|k}}{d\theta_i} + \frac{\partial\mathbf{A}_t}{\partial\theta_i}, \tag{34}$$

$$\frac{d\mathbf{C}_k}{d\theta_i} = \frac{\partial\mathbf{C}_k}{\partial\hat{\mathbf{x}}_{t|k}}\frac{d\hat{\mathbf{x}}_{t|k}}{d\theta_i} + \frac{\partial\mathbf{C}_k}{\partial\theta_i}. \tag{35}$$

By applying the differentiated filter Eqs. (26)–(32), the gradient for the likelihood can be obtained together with the ordinary filter equations. This in turn leads to a more robust optimization procedure since there is no need to approximate the gradient using finite differences. Instead the step size control of the algorithm for numerical integration of the total system of ordinary differential equations is utilized to obtained necessary precision and accuracy of both filter entities and their parametric sensitivities. When using a combination of an optimization routine and integration scheme utilization of finite differences may lead to unsatisfactory results due to the adaptive step length in the integration scheme, see for example [30]. With differentiated filter equations no such problems occurs and the precision in the gradient only depends on how accurate the used integration scheme is.

6. FitzHugh–Nagumo model for excitable media

In this section we consider a nonlinear model describing the reciprocal dependencies of the voltage V across an exon membrane and a recovery variable R summarizing outward currents. The model was developed in [22,23] based on the model by Hodgkin and Huxley [31]. The model is general and is also used to model excitable media, for example heart tissue.

In [32], it is pointed out that the objective function for parameter estimation in this model has a large number of local minima, and an regularization method is proposed. The purpose of this section is to illustrate how the likelihood can be regularized by introducing system noise in the ordinary differential equation.

6.1. Model equations

The model is described by the following system of ordinary differential equations

$$\frac{dV}{dt} = \gamma\left(V - \frac{V^3}{3} + R\right), \tag{36}$$

$$\frac{dR}{dt} = -\frac{1}{\gamma}(V - \alpha + \beta R), \tag{37}$$

with parameters α, β, and γ, initial conditions $V(0) = -1$ and $R(0) = 1$ together with the measurement equation

$$y_k = V(t_k) + e_k, \tag{38}$$

where $e_k \sim N(0, S)$. The voltage $V(t)$ is assumed to be sampled between $t = 0$ and $t = 20$ at discrete time points $t_k = 0, 1, 2, \ldots, 20$ with an additive measurement variance $S = 0.1$, using parameter values $\gamma = 3$, $\alpha = 0.2$, and $\beta = 0.2$. In Fig. 1 the voltage $V(t)$ (black, solid) and recovery variable $R(t)$ (purple, dashed) are plotted as functions of time together with the sampled voltage (black dots).

The negative log likelihood function (11) is calculated as a function of the parameter values $-1 \leqslant \alpha \leqslant 2$ and $-1 \leqslant \beta \leqslant 2$, keeping

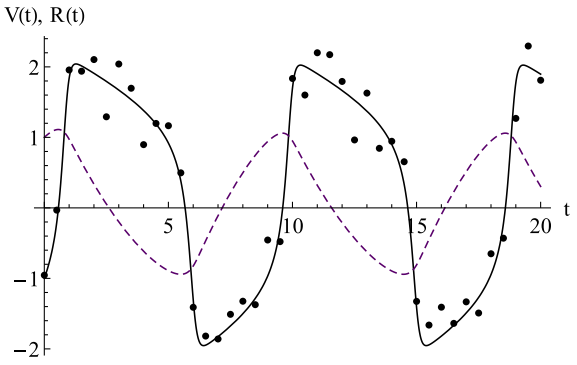


Fig. 1. Voltage V (black, solid) and recovery variable R (purple, dashed) for the FitzHugh–Nagumo model with $\alpha = 0.2$, $\beta = 0.2$, and $\gamma = 3$. The voltage is sampled at discrete time points $t_k = 0, 1, 2, \ldots, 20$, shown as black dots. The measurement variance was set to $S = 0.1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$\gamma = 3$. The reason to only consider two parameters is for easy visualization and demonstration. In Fig. 3(a) the objective function is shown as a contour plot, where darker colors implies lower function values. From Fig. 3(a) we conclude that the objective function has several local minima. Using gradient-based methods and starting at a point far away from the global minima will lead to an unsatisfactory estimate as the optimization most likely will end up in a local minima. For this data set, the optimum (the maximum likelihood estimate) was found to be $\hat{\alpha} = 0.18$ and $\hat{\beta} = 0.31$. The reason that the optimum is not in the parameter values used for simulation is that we sample with noise.

6.2. Regularization of the objective function

To regularize the objective function (11) the idea is to introduce system noise in the ordinary differential equation system used in the estimation procedure. The motivation of this is that when there is no system noise in the ordinary differential system the prediction of the model output, given by the expected value of the model output conditioned on past measurements, will follow the noise free trajectory, even when parameters values are far away from their true values. When system noise is introduced in the system the state prediction and its covariance will change at each measurement according to the updating formula in the EKF. This is due to the correlation between the variables introduced by the use of stochastic differential equations instead of ordinary differential equations. This will in turn lead to that the prediction is closer to the measurements and the corresponding piecewise trajectory will not drift away from the measured data. This concept is illustrated in Fig. 2 where the state prediction for deterministic model (black, solid) and the stochastic model (piecewise, blue) are shown for two different values of α with $\beta = 0.2$. The black dots correspond to measurements and blue crosses state updates. From Fig. 2(b) we see that even when the parameters are far from the true values the prediction with the stochastic model is still close to the measurements. The system noise level Σ was set to a diagonal matrix with diagonal elements 0.1.

Since $\mathbf{\Sigma} \in \mathbb{R}^{n\times q}$ one has to choose the dimension q. In this paper we restrict to the case $q = n$. Hence $\mathbf{\Sigma}$ is a 2×2 matrix which we will assume to be diagonal with diagonal elements σ, which implies that the system noise is of equal magnitude in both states. The corresponding stochastic differential equation system becomes

$$dV = \gamma\left(V - \frac{V^3}{3} + R\right)dt + \sigma d\omega_{1t}, \tag{39}$$

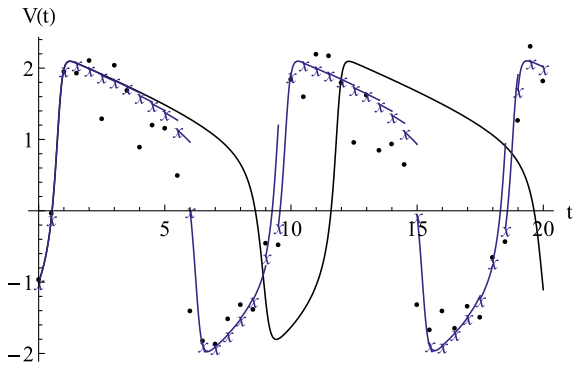
$$dR = -\frac{1}{\gamma}(V - \alpha + \beta R)dt + \sigma d\omega_{2t}, \tag{40}$$

with measurement equation

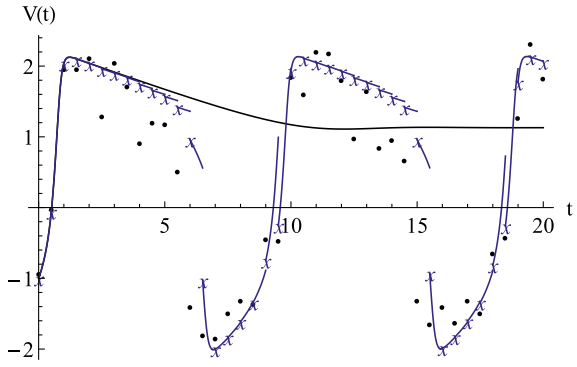
$$y_k = V(t_k) + e_k. \tag{41}$$

In the equations above, ω_{1t} and ω_{2t} correspond to the two components of the 2-dimensional Wiener process $\boldsymbol{\omega}_t$.

The objective function is given by (11) with $\mathbf{R}_{k|k-1}$ and $\boldsymbol{\epsilon}_k$ given by the EKF, described in Algorithm 1. Note that $\sigma = 0$ corresponds to the deterministic model where $\mathbf{R}_{k|k-1} = \mathbf{S}$ for all k. When $\sigma > 0$ the output covariance $\mathbf{R}_{k|k-1}$ changes in each time step according to the filter equations. This will force the state estimate towards the observed data. In Fig. 3 the corresponding contour plots of the negative log likelihood (11) for the two cases $\sigma = 0.1$ and $\sigma = 0.2$ are shown. The actual values of the likelihood is not of large importance, instead we would like to draw the reader's attention to the qualitative properties of the objective function compared with Fig. 3(a).



(a) $\alpha = 0.7$



(b) $\alpha = 1$

Fig. 2. Illustration of the extended Kalman filter. Prediction with no system noise (solid, black) and system noise (piecewise, blue) using different parameter values of α. The black dots corresponds to measurements and the blue crosses corresponds to state updates. Fig. 2(b) shows that even when the parameters are far from the true values and the dynamic is different the prediction is still close to data. The parameter β is set to 0.2 and system noise level Σ was set to a diagonal matrix with diagonal elements 0.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.3. Optimization benchmarking

To illustrate the regularizing effect on the optimization when introducing system noise we optimize the negative log likelihoods in the deterministic and stochastic setting, respectively. Using 50 random starting values of the parameters α and β between -1 and 2 the estimates from the two cases are compared to illustrate the difference in convergence. In the case when system noise is present, the noise level is set to $\sigma = 0.2$.

In Fig. 4 the estimates are shown together with the corresponding contour plots. The black circles indicate the different initial values for the optimizations and the red dots correspond to the estimated parameters. The solid lines show the convergence of the different starting values.

The plots of the attractor points reveal that local minima exist in the deterministic model and that the optimization sometimes converges to a local minima. For the deterministic model, 16 out of the 50 optimizations ended up near the global minima. For the model with system noise, all the 50 optimizations converged to the same point (using the same starting values as in the deterministic model). The mean value for the 50 optimizations in the stochastic model was $\bar{\alpha} = 0.27$ and $\bar{\beta} = 0.48$. Since we now consider a different model incorporating system noise, the optimum is not the same as in the deterministic model (with optimum $\hat{\alpha} = 0.18$, $\hat{\beta} = 0.31$). However, the values $\bar{\alpha} = 0.27$, $\bar{\beta} = 0.48$ is still in the region of attraction of the global optimum in the deterministic model. By introducing system noise in the ordinary

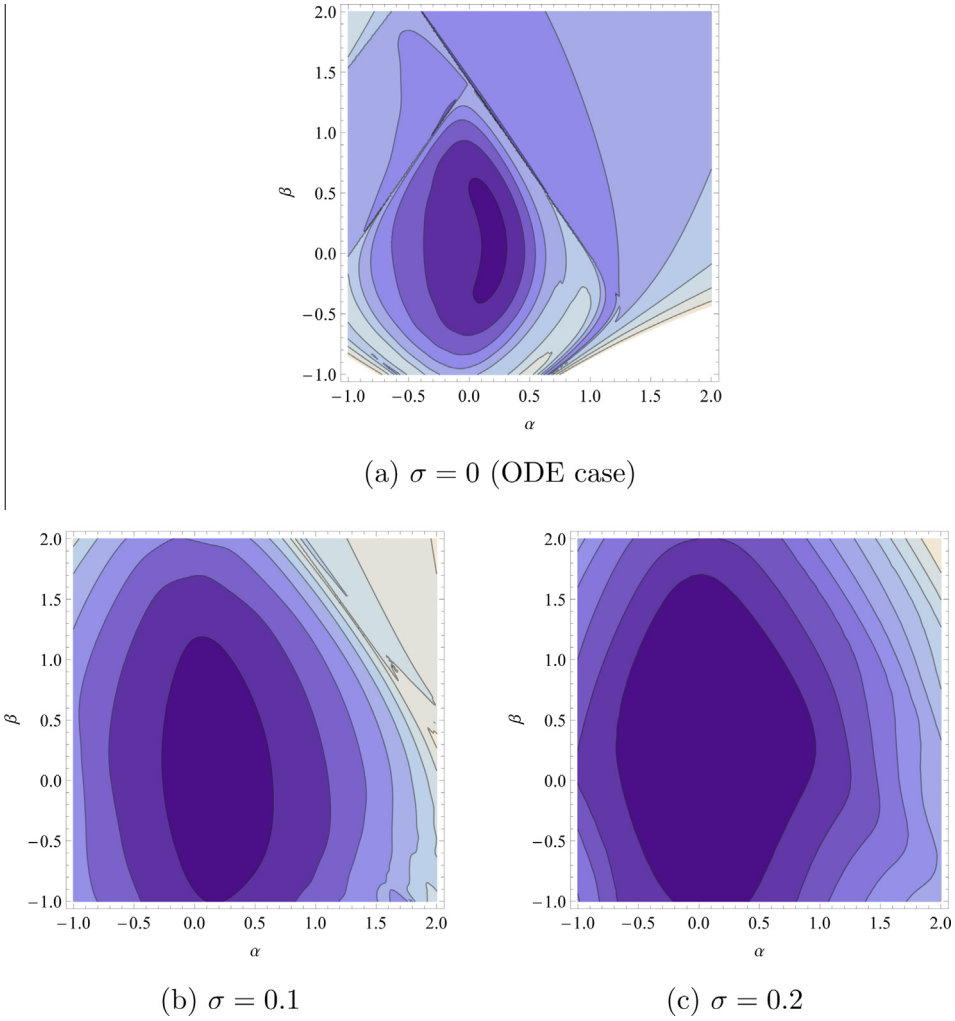


Fig. 3. Contour plots of the objective function (11) for different values of σ with $-1 \leqslant \alpha \leqslant 2$ and $-1 \leqslant \beta \leqslant 2$, keeping $\gamma = 3$, where darker colors implies lower function values. As the system noise level increases, the objective function shows a more regular behavior without local minima.

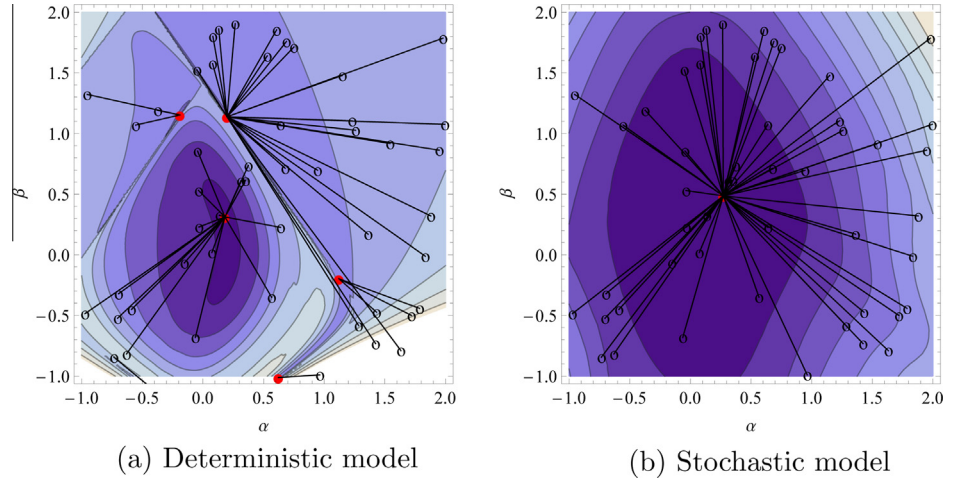


Fig. 4. Estimated parameters visualized as points in the contour plots in the FitzHugh–Nagumo model. The dots correspond to the estimated parameters and the black circles to the initial guesses. The solid lines show the convergence of the different starting values. A total of 50 random starting values was considered.

differential equation the objective function is observed to be more regular and smooth which has a large effect on the estimated parameters.

To further demonstrate the results from the FitzHugh–Nagumo model we will perform the same analysis on another model, namely the Lotka–Volterra predator–prey model.

7. Lotka–Volterra predator–prey model

As a second example of the implication of introducing noise in the model we consider the Lotka–Volterra predator–prey system. The system describes the interaction between two species (predator and prey) in an ecological system and shows an oscillatory behavior which in turn leads to a hard optimization problem when using gradient-based methods.

7.1. Model equations

The system of ordinary differential equations describing the relation between the prey $x(t)$ and predator $y(t)$ is

$$\frac{dx}{dt} = x(\alpha - \beta y), \tag{42}$$

$$\frac{dy}{dt} = -y(\gamma - \delta x), \tag{43}$$

with model parameters $\alpha,\ \beta,\ \gamma$, and δ describing the interaction between the two species. The initial conditions are $x(0) = 10$ and $y(0) = 10$. The two states are measured between $t = 0$ and $t = 24$ at discrete time points $t_k = 0, 0.5, 1, \ldots, 24$ using parameter values $\alpha = 1,\ \beta = 0.2,\ \gamma = 1$, and $\delta = 0.15$. In Fig. 5 we depict the states with added measurement noise with variance $S = 0.1$ in both states.

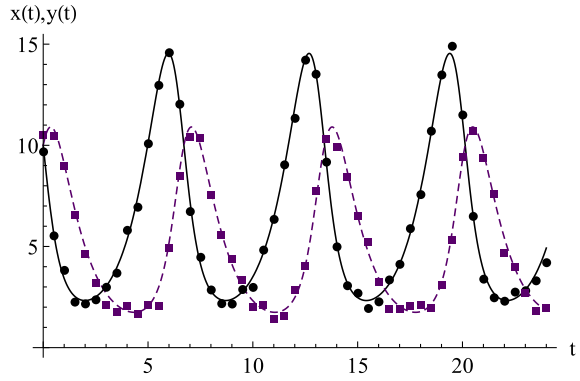


Fig. 5. Prey (black, solid) and predator (purple, dashed) in the Lotka–Volterra system considered together with sampled data. The parameter values are $\alpha = 1,\ \beta = 0.2,\ \gamma = 1$, and $\delta = 0.15$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The negative log likelihood function (11) is calculated as a function of the parameter values α and δ keeping $\beta = 0.2$ and $\gamma = 1$ fixed. In Fig. 6(a) a contour plot of the negative log likelihood function is shown for $0.4 \leqslant \alpha \leqslant 2$ and $0.1 \leqslant \delta \leqslant 0.5$. For the considered data set, the optimum was $\hat{\alpha} = 1.00,\ \hat{\delta} = 0.15$, which is the same as the parameter values used for simulation. This objective function is very problematic for local, gradient-based optimization algorithm

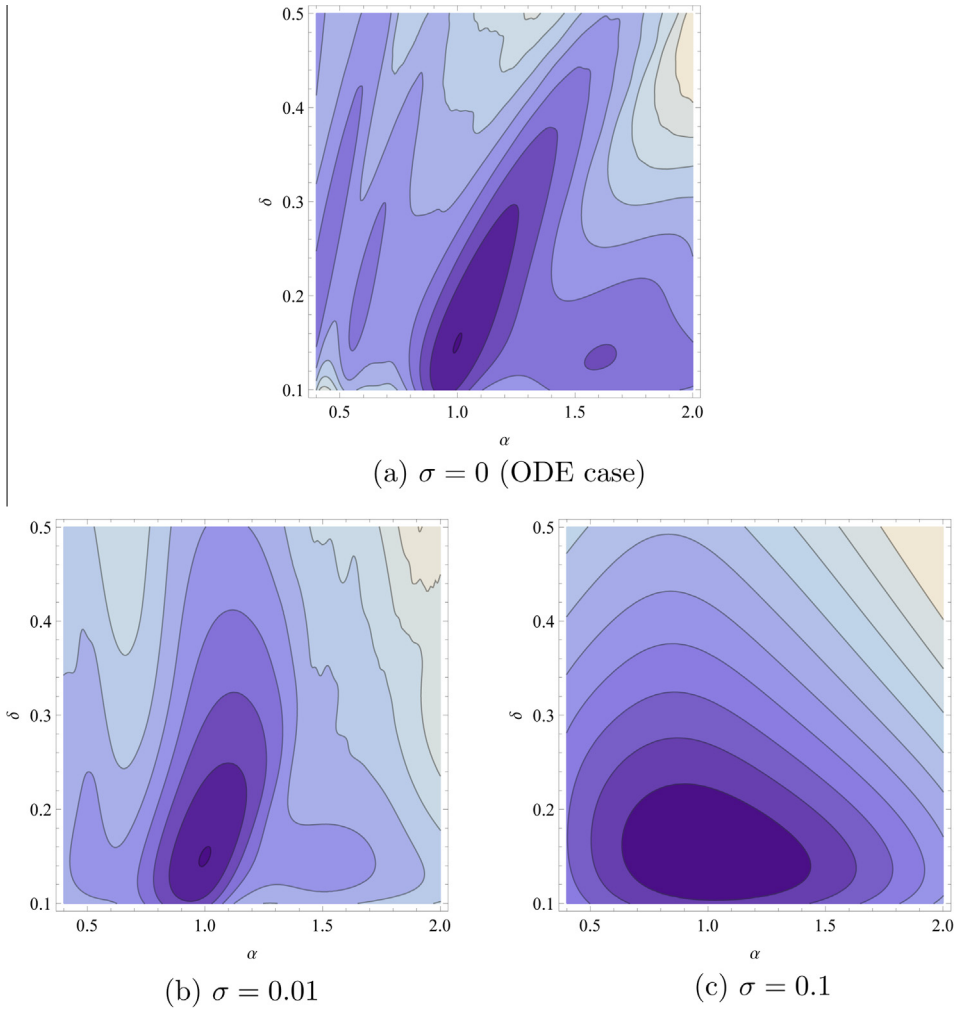


Fig. 6. Contour plots of the objective function (11) for different value σ for parameters $0.4 \leqslant \alpha \leqslant 2$ and $0.1 \leqslant \delta \leqslant 0.5$. The other parameters are fixed using $\beta = 0.2$ and $\gamma = 1$.

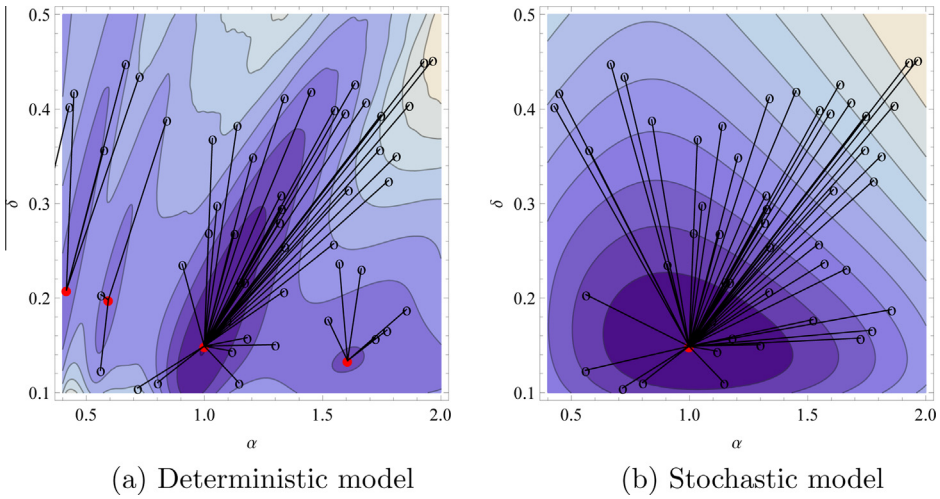


Fig. 7. Estimated parameters visualized as points in the contour plots in the Lotka–Volterra model. The dots correspond to the estimated parameters and the black circles to the initial guesses. The solid lines show the convergence of the different starting values. A total of 50 random starting values was considered.

since it contains a large number of local minima, which is often the case when considering a model describing oscillating phenomena.

7.2. Regularization of the log likelihood

Following the same approach as in the FitzHugh–Nagumo model the likelihood is regularized by introducing system noise. As before, we set $\mathbf{\Sigma}$ to a diagonal 2×2 matrix with diagonal elements σ. In Fig. 6 the contour plot of the objective function is shown for two different levels of noise using parameter values $0.4 \leqslant \alpha \leqslant 2$ and $0.1 \leqslant \delta \leqslant 0.5$.

7.3. Optimization benchmarking

Again, the objective function is minimized using 50 different starting values of the parameters α and δ. The minimization is performed using the deterministic model and the model incorporating system noise at level $\sigma = 0.1$. The estimated parameters for the 50 runs are shown in Fig. 7 with the respective contour plots. The black circles indicate the different initial values for the optimizations and the red dots correspond to the estimated parameters. The solid lines show the convergence of the different starting values.

For the deterministic model the optimization converged to the global minima 36 times out of the 50 optimizations. Note that the valleys in the contour plot serve as attractors for the optimization. For the model incorporating system noise all the 50 runs converged to a value very close (mean value $\bar{\alpha} = 1.00,\ \bar{\delta} = 0.15$) to the global minima $\hat{\alpha} = 1.00,\ \hat{\delta} = 0.15$. We conclude that the optimum point is the same up to two decimal places.

Regularizing the objective function using stochastic differential equations works well for the Lotka–Volterra model as well using a system noise level of $\sigma = 0.1$. The objective function shows no local minima where the optimization can converge to a suboptimal point. Instead, the objective function for the stochastic model has a global minima which exists at a point very close to the parameter values that was used for generating the measurements.

8. Conclusions and future work

In this paper, the extension of ordinary differential equations to stochastic differential equations was considered as a tool for regularizing the objective function for parameter estimation in dynamical systems. Using two biological different examples, the FitzHugh–Nagumo model and the Lotka–Volterra predator–prey system, the method was described and demonstrated using *in silico* data. The two examples revealed that the objective function could be regularized using an appropriate choice of the system noise. By allowing noise in the model itself, the state estimates are attracted towards the observed data and the number of local minima in the objective function are observed to be reduced.

Both the level of the system noise and the measurement covariance will have an impact on the regularization. Since the Kalman gain depends on both the state covariance and the measurement covariance, the updating depends on the relation between those two.

Moreover, it is important to note that the level of the system noise could be treated as a design parameter in the optimization and that the final estimate will depend on the selected level. However, starting with a larger level of the system noise and then decreasing it as the optimization approaches the global minimum may also be considered. This approach can be compared to the method of Simulated annealing [33], when cooling of a temperature is used during the optimization. In the beginning of the optimization the search space is increased by allowing a large uncertainty (which would correspond to the system noise). The uncertainty is then decreased during the optimization until the optimization has reached an optimum. In contrast to stochastic methods, gradient-based methods are known to be highly efficient when applicable and the structure of the model can be used in the optimization. Their applicability, though, is determined by general properties of the objective function such as smoothness and a limited number of local minima, which is what our regularization approach tries to achieve. Stochastic optimization methods, on the other hand, require a large number of function evaluations and have much slower convergence but are necessary for irregular, non-smooth, multi-modal objective functions. Using the regularizing method proposed in the paper to remove local minima together with a gradient-based optimization method is a highly efficient approach, in particular if the regularization succeeds in removing all but one minimum.

The gradient in the objective function was calculated by parametric sensitivities of the EKF for a robust calculation. This is a great benefit to the frequently used finite difference approximation. When the system of differential equations becomes larger we need to solve a large number of differential equations to obtain the required sensitivities. However, in many cases the covariance

of the states reaches a steady-state rather quickly. One possible solution to decrease the number of differential equations is to compute the steady-state solution of the matrix differential equation describing the evolution of the covariance at each and every measurement time point (a system of linear equations), provided this is as a reasonable approximation. This has not been implemented in the current version of the algorithm.

The proposed method has some similarities with the method of multiple shooting [6–8]. In this method the time domain is divided into disjoint intervals and the solution is re-initialized at the start of each interval, giving a discontinuous model solution. Slack variables to be penalized are introduced in the objective function to eventually achieve a continuous solution. However, the proposed method using stochastic differential equations guarantees that the states of the underlying system are estimated from data and not only forces the output of the system to be close to the measurements.

If some prior knowledge exists about the parameters one may consider the Bayesian approach, as illustrated in [34,35]. In these papers, the authors illustrate the benefit of using a prior when such information about the parameters are available from previous studies. The prior can then be used to make the parameters in question identifiable. However, in our work we assume no á priori information about the model parameters. In the case of flat prior the Bayesian approach reduces to the maximum likelihood approach which has been considered in this paper.

We conclude that a stochastic differential equation setup can be used as a tool to regularize a complex objective function used for parameter estimation in ordinary differential equations. The use of SDEs, which is a model class including ODEs as a special case (for $\mathbf{\Sigma} = \mathbf{0}$), provide means to regularize the estimation problem. The cost of using this more complicated model structure is well motivated by the reduced complexity of the associated optimization problem (in terms of number of local minima). In this paper we only considered objective functions depending on two parameters to allow for better visualization and understanding of the behavior of the objective function. For future work, we suggest an extended analysis to high dimensional problems that are known to be multi-modal. This has not been investigated in the present work, but will naturally be considered in a future paper.

References

[1] P. Deuflhard, F. Bornemann, Scientific Computing with Ordinary Differential Equations, Springer, 2002.
[2] K. Schittkowsky, Numerical Data Fitting in Dynamical Systems – A Practical Introduction with Applications and Software, Kluwer Academic Publishers, 2002.
[3] T. Bohlin, S.F. Graebe, Issues in nonlinear stochastic grey box identification, Int. J. Adapt. Control Signal Process. 9 (6) (1995) 465.
[4] I.S. Mbalawata, S. Särkkä, H. Haario, Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering, Comput. Stat. (2012) 1.
[5] N.R. Kristensen, H. Madsen, S.B. Jørgensen, Parameter estimation in stochastic grey-box models, Automatica 40 (2) (2004) 225.
[6] H. Bock, Recent advances in parameter identification techniques for ODE, in: P. Deuflhard, E. Hairer (Eds.), Numerical Treatment of Inverse Problems in Differential and Integral Equations, Birkhäuser, Boston, 1983, p. 95.
[7] M. Peifer, J. Timmer, Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting, IET Syst. Biol. 1 (2) (2007) 78.
[8] H.U. Voss, J. Timmer, J. Kurths, Nonlinear dynamical system identification from uncertain and indirect measurements, Int. J. Bifur. Chaos 14 (6) (2004) 1905.
[9] C.G. Moles, P. Mendes, J.R. Banga, Parameter estimation in biochemical pathways: a comparison of global optimization methods, Genome Res. 13 (11) (2003) 2467.
[10] M. Rodriguez-Fernandez, P. Mendes, J. Banga, A hybrid approach for efficient and robust parameter estimation in biochemical pathways, Biosystems 83 (2–3) (2006) 248.
[11] M. Rodriguez-Fernandez, J. Egea, J. Banga, Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems, BMC Bioinf. 7 (1) (2006) 483.
[12] I. Chou, E. Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems, Math. Biosci. 219 (2) (2009) 57.
[13] F. Klebaner, Introduction To Stochastic Calculus With Applications, Imperial College Press, 2005.
[14] B. ksendal, Differential Equations. An Introduction with Applications, Springer, 2003.
[15] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, New York, 1992.
[16] D. Higham, An algorithmic introduction to numerical simulation of stochastic differential equations, SIAM Rev. 43 (3) (2001) 526.
[17] J.N. Nielsen, H. Madsen, P.C. Young, Parameter estimation in stochastic differential equations: an overview, Ann. Rev. Control 24 (2000) 83.
[18] S. Donnet, A. Samson, A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models, Adv. Drug Delivery Rev. 65 (7) (2013) 929.
[19] N.R. Kristensen, H. Madsen, S.H. Ingwersen, Using stochastic differential equations for pk/pd model development, J. Pharmacokinetics Pharmacodynamics 32 (1) (2005) 109.
[20] T. Bohlin, Practical Grey-box Process Identification, Springer, 2006.
[21] N. Kristensen, H. Madsen, Continuous time stochastic modeling, CTSM 2.3, Math. Guide (2010).
[22] R. FitzHugh, Impulses and physiological states in theoretical models of nerve membrane, Biophys. J. 1 (6) (1961) 445.
[23] J. Nagumo, S. Arimoto, S. Yoshizawa, An active pulse transmission line simulating nerve axon, Proc. IRE 50 (10) (1962) 2061.
[24] A. Jazwinsky, Stochastic Processes and Filtering Theory, Academic Press, 1970.
[25] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. ASME – J. Basic Eng. 82 (Series D) (1960) 35–45.
[26] J. Møller, N. Carstensen, H. Madsen, Stochastic state space modelling of nonlinear systems – with application to marine ecosystems, IMM-PHD-2010, Technical University of Denmark (DTU), 2011.
[27] S. Klim, S. Mortenson, Stochastic PK/PD modeling (Master's thesis), Kongens Lyngby, 2006.
[28] J. Nocedal, S. Wright, Numerical Optimization, Springer Verlag, New York, 1999.
[29] N. Skaar, Parameter estimation methods for continuous time dynamical systems given discrete time measurements (Master's thesis), Chalmers University of Technology, 2008.
[30] J. Carlsson, C. Nordheim, A parameter estimation method for continuous time dynamical systems based on the unscented Kalman filter and maximum likelihood (Master's thesis), Chalmers University of Technology, 2011.
[31] A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, J. Physiol. 117 (4) (1952) 500.
[32] J.O. Ramsay, G. Hooker, D. Campbell, J. Cao, Parameter estimation for differential equations: a generalized smoothing approach, J. R. Stat. Soc.: Ser. B (Stat. Method.) 69 (5) (2007) 741.
[33] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671.
[34] P. Magni, R. Bellazzi, A. Nauti, C. Patrini, G. Rindi, Compartmental model identification based on an empirical bayesian approach: the case of thiamine kinetics in rats, Med. Biol. Eng. Comput. 39 (6) (2001) 700.
[35] P. Magni, G. Sparacino, R. Bellazzi, C. Cobelli, Reduced sampling schedule for the glucose minimal model: importance of bayesian estimation, Am. J. Physiol. – Endocrinol. Metab. 290 (1) (2005) E177.