THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Contributions to nonlinear mixed-effects modeling

Jacob Leander

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Division of Mathematics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden 2014

Contributions to nonlinear mixed-effects modeling

Jacob Leander

ISSN 1652-9715
NO 2014:12

Division of Mathematics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
SE–412 96 Gothenburg, Sweden
Phone: +46 (0)31–772 10 00

Printed by Reproservice
Gothenburg, Sweden, 2014

Abstract

The topic of this thesis is nonlinear mixed-effects models. A nonlinear mixed-effects model is a hierarchical regression model used to analyze measurements from several individuals simultaneously, which allows sharing of information between similar individuals. In many applications, the response is described by an ordinary differential equation together with an equation describing the measurement process. To allow for uncertainty in the underlying model as well, we consider extension of the ordinary differential equation to a stochastic differential equation. Moreover, since the parameter estimation problem requires solution of an optimization problem, we derive a novel method for calculating the gradient of the objective function in nonlinear mixed-effects models. Instead of utilizing a finite difference approximation of the gradient, we instead derive an expression for the gradient of the objective function using sensitivity equations.

In Paper I, stochastic differential equations are introduced in the single individual case. In this paper we show the ability of stochastic differential equations to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. We also introduce the concept of the extended Kalman filter, which serves as a state estimator in the stochastic models.

In Paper II we propose a combination of the nonlinear mixed-effects model and stochastic differential equations. By utilizing stochastic differential equations in the nonlinear mixed-effects model three different sources of variability in data can be modeled. In contrast to the commonly used measurement error and parameter variability we also consider uncertainty in the underlying dynamics. The stochastic mixed-effects model is applied to two different pharmacokinetic models describing drug interaction, where we show the ability of the proposed method to separate the three sources of variability in the stochastic differential mixed-effects model.

In Paper III we present the theory for efficient gradient calculation in nonlinear mixed-effects models. Instead of utilizing the common finite difference approximation of the gradient, we consider explicit derivation of the objective function using sensitivity equations. It is shown that the novel method significantly decreases the time needed for gradient calculation and at the same time increases the precision.

List of publications

This thesis is based on the work contained in the following papers:

Paper I
J. Leander, T. Lundh, M. Jirstrand, Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements, Mathematical Biosciences, Volume 251 (May 2014), 54-62

Paper II
J. Leander, J. Almquist, C. Ahlström, J. Gabrielsson, M. Jirstrand, Mixed-effects modeling using stochastic differential equations: Illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats, Submitted to Journal of Pharmacokinetics and Pharmacodynamics

Paper III
J. Almquist, J. Leander, M. Jirstrand, Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood, Manuscript in preparation

Other related publications not included in the thesis

S. Tapani, J. Almquist, J. Leander, C. Ahlström, J. Gabrielsson, M. Jirstrand, Joint feedback modeling of non-esterified fatty acids in obese Zucker rats and normal Sprague Dawley rats after different routes of administration of nicotinic acid, Journal of Pharmaceutical Sciences, Volume 103, 8 (Aug 2014), 2571-2584

Acknowledgements

As one of the first three students in the two-year licentiate program Advanced Engineering Mathematics I have had the opportunity of working at Fraunhofer-Chalmers Centre and at the same time be a graduate student in mathematics. It has been two exciting years, working collaborative with the Department of Mathematical Sciences and Fraunhofer-Chalmers Centre.

First of all, I would like to thank my two supervisors Mats Jirstrand at Fraunhofer-Chalmers Centre and Torbjörn Lundh at the Department of Mathematical Sciences. I have learned a great deal from you, and you have always had time for my questions and discussions. Without you this licentiate thesis would not be possible. Thank you!

Secondly, a huge thank you to Joachim Almquist. Without you the development and implementation of the nonlinear mixed-effect model framework wouldn't been the same. I have enjoyed all the discussions about problems and bugs, jokes, and long evenings at work. Thank you!

I would also like to thank all the people that I have met on my journey so far. Thanks for all the weird conversations, discussions, coffee breaks and cold beers. A special thank you to Staffan, Tomas, Domenico, Anton, Roland, Fredrik, Daniel, Mikael and Johan. You all made my time at Fraunhofer-Chalmers Centre pass too fast.

Finally, I would like to thank my family and girlfriend, who always stand by my side.

Gothenburg, 2014-07-24

Contents

Page

Abstract **i**

List of publications **iii**

Acknowledgments **v**

Prologue **1**
1.1 Regression modeling 1
1.2 Multivariate normal distribution 1
1.3 Maximum likelihood 2
1.4 Optimization techniques 2

Introduction **2**
2.1 Brief background 3
2.2 The nonlinear mixed-effects model 4
2.3 Inference in nonlinear mixed-effects models 6

Summary of Papers **8**
3.1 Summary of Paper I 9
3.2 Summary of Paper II 11
3.3 Summary of Paper III 13

Paper I **17**

Paper II **29**

Paper III **51**

Prologue

For convenience of the reader, we here repeat the basic mathematical tools needed throughout this thesis.

1.1 Regression modeling

Regression models play an important role in many applications. In regression analysis, the aim is to understand how the typical value of a dependent variable depends on one or more independent variables. The regression model relates the dependent variables to a function of the independent variables and model parameters. In a linear regression model we model the relationship between the independent variables $\boldsymbol{X}$ and the dependent variables $\boldsymbol{Y}$ by the linear combination

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{\beta}$ is the vector of model parameters and $\boldsymbol{\epsilon}$ is the vector of errors (which in most cases is assumed to be normal distributed). In many applications we are interested in identifying the values of the model parameters $\boldsymbol{\beta}$ which best explains the dependent variables $\boldsymbol{Y}$. In the linear case with assumption of normal distributed errors, the problem has an *ordinary least-squares* solution. There is a vast collection of literature on regression models and we refer the interested reader to [8].

1.2 Multivariate normal distribution

The normal distribution is the most common distribution to describe errors in regression models. The motivation for using the normal distribution comes from the fact that a sum of a large number of independent random variables will be approximately normally distributed. The generalization of the normal distribution to higher dimensions is given by the multivariate normal distribution.

The multivariate normal distribution of a k-dimensional random vector $\boldsymbol{x}$ is denoted

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{1.2}$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the $k \times k$ covariance matrix. The probability density function for the multivariate normal distribution is

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}^k \sqrt{\det(\boldsymbol{\Sigma})}} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big). \tag{1.3}$$

1.3 Maximum likelihood

In more complex regression models and the in stochastic models we shall encounter in this thesis ordinary least squares may not be sufficient to determine the model parameters. Parameter estimation in a more general regression models is often accomplished using the maximum likelihood approach. Assume that we have observed our dependent variable $\boldsymbol{Y}$ and are interested in identifying the model parameters $\boldsymbol{\beta}$. To estimate the model parameters we can utilize the *likelihood function* $L(\boldsymbol{\beta};\boldsymbol{Y})$. The likelihood function is defined as the probability density evaluated for the observed data and parameterized by the model parameters, that is

$$L(\boldsymbol{\beta};\boldsymbol{Y}) = p(\boldsymbol{Y};\boldsymbol{\beta}). \tag{1.4}$$

By maximizing the likelihood function we obtain the *maximum likelihood estimate* of the model parameters $\boldsymbol{\beta}$.

1.4 Optimization techniques

When faced with an optimization problem, we shall distinguish between global and local methods. The methods we will encounter throughout this thesis is local methods, which means that they converge to a local solution (which may or may not be the optimal solution). The local methods usually show a much faster convergence to a solution than the global methods which in turn aims at a global solution. Many of the local methods utilizes information about the *gradient* of the objective function.

The gradient is a generalization of the usual concept of derivative to a function of several parameters. Consider a function $L(\boldsymbol{\theta})$ depending on a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$. The gradient $\frac{dL}{d\boldsymbol{\theta}}$ of the vector valued function $L(\boldsymbol{\theta})$ is defined as the vector of partial derivatives $(\frac{\partial L}{\partial \theta_1}, \ldots, \frac{\partial L}{\partial \theta_p})$. The calculation of the derivatives needed to calculate the gradient is often approximated using finite differences. By taking a small step in parameter space and divide by the step length h the partial derivative of the parameter θ_k is approximated by

$$\frac{\partial L}{\partial \theta_k} \approx \frac{L(\boldsymbol{\theta} + h\boldsymbol{e}_k) - L(\boldsymbol{\theta})}{h}, \tag{1.5}$$

where $\boldsymbol{e}_k$ denotes the unit vector in the k:th dimension, which gives the first-order finite difference approximation. More accurate and higher order approximations may also be considered. By utilizing the gradient of the objective function the local methods iterate through the parameter space in the search of a local solution.

Introduction

In many fields, repeated measurements are collected on several subjects to gain information about some underlying phenomena of interest. The topic of this thesis is nonlinear mixed-effects models, which are hierarchial regression models often used in such situations. This introduction aims at giving a short survey on the topic of nonlinear mixed-effects models. Starting with a brief background, we continue to formally state and explain the nonlinear mixed-effects model. After the formal introduction, a few of the most common methods for parameter estimation in nonlinear mixed-effects models are presented.

This licentiate thesis is the result of the two-year long graduate programme *Advanced Engineering Mathematics*, which has resulted in four papers. Three of these papers constitute the main part of the thesis, and the papers are appended together with a summary of each paper.

When writing an introduction on a mathematical topic you always have a choice whether you should fill it with definitions, equations and proofs. I have chosen not to do so, and the mathematical part is mainly in the three articles. With that being said, we start with some brief background

2.1 Brief background

There is a vast collection of regression models, ranging from simple linear models to multi-level nonlinear models. The term *linear* in regression modeling refers to model linearity in model parameters. However, in the applications we shall encounter in this thesis, the models are *nonlinear*, which clearly complicates the inference. Moreover, our special interest in this thesis lies in modeling time-dependent systems that embodies state variables dependent of time, for example the system of drug metabolism, the interaction between predator and prey in an ecosystem, and seasonal models. In such time-dependent models, the dynamics is often assumed to be described by a system of ordinary differential equations.

In this thesis, we consider the case when repeated measurements are performed on several individuals assumed to be randomly drawn from a population of interest. To handle such cases, many statistical models incorporate so called *fixed effects*, which are parameters associated with a certain population of interest, and *random effects*, which are associated with individuals. A model with both fixed and random effects is called a *mixed-effects model*.

The mixed-effects model has gained increasing popularity, and is today used in for example pharmacometrics [23], behavioral science, image analysis [4], and forestry

[18]. It has several advantages over the so called standard two-stage approach. In the standard two-stage approach the parameters of each individual are estimated from the individual data. From the collection of estimated parameters inference about the population as a whole is done, for example by calculating the population mean and variance of each parameter. However, if the parameters for a specific individual are unreliable estimated or even unidentifiable (e.g., due to missing data) it can lead to incorrect inference about the population. The mixed-effects model instead provides a framework that allows sharing of information from similar individuals, which can be valuable when sufficient data is not available for a specific individual [6]. Moreover, it enables modeling and estimation of the variance-covariance structures for both random effects and residuals.

The mixed-effects model is a very popular tool in pharmacometrics where it is used to analyze data from several individuals to increase the knowledge in for example drug metabolism, administration, and drug effect. One of the first inclusions of mixed-effects modeling in pharmacokinetics was done by Lewis Sheiner [23], where the concept and a first method for parameter estimation in nonlinear mixed-effects model was presented. In 1977, he presented the first case study of estimation of population characteristics of pharmacokinetic parameters from clinical data. This paper was then followed by three seminal papers by Lewis Sheiner and Stuart Beal [20, 21, 22]. Their models and estimation methods are incorporated in the NONMEM program, a popular software for nonlinear mixed-effects modeling in drug development.

After the introduction in the pharmaceutical field, several statistical developments in nonlinear mixed-effects models have been proposed. The parametric approach was further analyzed and in 1990 Lindstrom & Bates [11] presented an extension of the method proposed by Sheiner. Further work includes for example nonparametric methods, Bayesian parametric approach, Laplacian approximation, and adaptive Gaussian quadrature. The last decade, several stochastic methods have been presented for maximum likelihood estimation in nonlinear mixed-effects models. For more information and methods in nonlinear mixed-effects model, see e.g. [5], [11] and [17].

2.2 The nonlinear mixed-effects model

In this section we formally state and explain the nonlinear mixed-effects model, proposed by Lindstrom & Bates [11]. In this thesis, we restrict the case to single-level of grouping, and refer the interested reader to [17] for multilevel nonlinear mixed-effects models.

Nonlinear mixed-effects models are mixed-effects models in which some, or all, of the fixed and random effects occur nonlinearly in the model function. The nonlinear mixed-effects model can be seen as a hierarchal model in two stages. The first level in the hierarchy models the underlying system and measurement process for a single individual. The individual response model can for example be an analytical expression depending on the explanatory variables in the regression model or a more complex, dynamic model.

Stage 1: Individual level

Nonlinear mixed-effects models are used to describe data of the form

$$\boldsymbol{y}_{ij}, \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i, \tag{2.1}$$

where the vector $\boldsymbol{y}_{ij}$ denotes the jth observation for the ith individual. The response $\boldsymbol{y}_{ij}$ is assumed to be described by a regression model given by an ordinary differential equation together with a measurement model. The complete model for the response is

$$d\boldsymbol{x}_i = \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{u}_i, t, \boldsymbol{\phi}_i)dt, \quad \boldsymbol{x}_i(t_0) = \boldsymbol{x}_0(\boldsymbol{\phi}_i) \tag{2.2}$$

$$\boldsymbol{y}_{ij} = \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{u}_i, t_{ij}, \boldsymbol{\phi}_i) + \boldsymbol{e}_{ij}, \tag{2.3}$$

where time is denoted by t, the state vector of the system is denoted $\boldsymbol{x}_i$ and the input to the system is denoted $\boldsymbol{u}_i$. The vector-valued function $\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{u}_i, t, \boldsymbol{\phi}_i)$ describes the dynamics of the system and $\boldsymbol{\phi}_i$ denotes the individual parameter vector for individual i. Note that the differential equation is written on differential form, which is the standard notation for stochastic differential equations which will be introduced later. Measurements of the system are assumed to be taken at discrete points in time and are characterized by the measurement function $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{u}_i, t_{ij}, \boldsymbol{\phi}_i)$ and the measurement error $\boldsymbol{e}_{ij} \sim N(0, \mathbf{R}(\boldsymbol{x}_i, \boldsymbol{u}_i, t_{ij}, \boldsymbol{\phi}_i))$.

The individual model described above can be seen as a nonlinear regression model depending on the individual parameters $\boldsymbol{\phi}_i$. The link between the individual response and the population model is given by another hierarchal relationship, namely the population model.

Stage 2: Population level

On a population level, the individual parameters $\boldsymbol{\phi}_i$ are assumed to be described by the functional relationship

$$\boldsymbol{\phi}_i = \boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{\eta}_i, \boldsymbol{Z}_i), \tag{2.4}$$

where $\boldsymbol{g}$ is a function depending on a vector of fixed effects $\boldsymbol{\theta}$, a vector of random effects $\boldsymbol{\eta}_i$, and covariates $\boldsymbol{Z}_i$. The random effects vector $\boldsymbol{\eta}_i$ is assumed to be described by a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Omega}$,

$$\boldsymbol{\eta}_i \sim N(\mathbf{0}, \boldsymbol{\Omega}). \tag{2.5}$$

The nonlinear mixed-effects model provide means to model the population by the fixed effects $\boldsymbol{\theta}$, i.e., the parameters associated with the specific population of interest. The function $\boldsymbol{g}$ models the relationship between the fixed effects and the individual parameters $\boldsymbol{\phi}_i$, which in turn determine the individual response. The link between the fixed effects and the individual parameters is given by the random parameter vector $\boldsymbol{\eta}_i$ and possible covariates $\boldsymbol{Z}_i$. The functional relationship $\boldsymbol{g}$ provides means to model specific distributions of parameters within the population. A common choice in for example pharmacometrics is to assume that the individual parameters are log-normally distributed in the population, which is modeled by the functional relationship

$$\boldsymbol{\phi}_i = \boldsymbol{\theta} \exp(\boldsymbol{\eta}_i). \tag{2.6}$$

Given a nonlinear mixed-effects model our primary interest is to make inference on the fixed effects and the covariance structure governing the distribution of parameters. In many applications we may not be interested in the random effects $\boldsymbol{\eta}_i$ *per se*, but they are often calculated in the inference procedure.

2.3 Inference in nonlinear mixed-effects models

There is a vast collection of inference methods for nonlinear mixed-effects models. As in any regression model, a popular tool for inference is maximum likelihood. We start to derive the formal likelihood function of the observed data given the model parameters. Due to the nonlinearity of the model one often have to rely on approximations. The aim is to give a brief overview of some of the most common approximations of the likelihood function.

Derivation of the population likelihood

We let $\boldsymbol{y}_i$ denote all the available data for individual i, and $\boldsymbol{y}$ the collection of all individual data. The likelihood for a single individual is defined as the probability density of the observed data given the parameters of the model, denoted $p(\boldsymbol{y}_i|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega})$. Recall that the observed data depends on the underlying model and the measurement error. The underlying model depends on the random effects $\boldsymbol{\eta}_i$. Since the random effects $\boldsymbol{\eta}_i$ are unobserved quantities, the individual likelihood can be obtained by marginalizing over the random effects

$$p(\boldsymbol{y}_i|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega}) = \int p(\boldsymbol{y}_i,\boldsymbol{\eta}_i|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega})d\boldsymbol{\eta}_i. \tag{2.7}$$

Utilizing the fact that $p(A,B) = p(A|B)p(B)$ and that $\boldsymbol{\eta}_i$ is independent of $\boldsymbol{\theta}$ and $\mathbf{R}$ we get

$$p(\boldsymbol{y}_i|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega}) = \int p(\boldsymbol{y}_i|\boldsymbol{\eta}_i,\boldsymbol{\theta},\mathbf{R})p(\boldsymbol{\eta}_i|\boldsymbol{\Omega})d\boldsymbol{\eta}_i. \tag{2.8}$$

If we assume individuals to be independent, the likelihood for the collection of individuals is simply a product of individual likelihoods

$$p(\boldsymbol{y}|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega}) = \prod_{i=1}^{N}\int p(\boldsymbol{y}_i|\boldsymbol{\eta}_i,\boldsymbol{\theta},\mathbf{R})p(\boldsymbol{\eta}_i|\boldsymbol{\Omega})d\boldsymbol{\eta}_i. \tag{2.9}$$

Since we assume that our dynamics is deterministic and that measurements are taken under independent and identically distributed normal errors the probability density $p(\boldsymbol{y}_i|\boldsymbol{\eta}_i,\boldsymbol{\theta},\mathbf{R})$ is simply a product of Gaussian densities. Moreover, $p(\boldsymbol{\eta}_i|\boldsymbol{\Omega})$ is a Gaussian density with covariance matrix $\boldsymbol{\Omega}$. Hence, the expression for the likelihood is

$$p(\boldsymbol{y}|\boldsymbol{\theta},\mathbf{R},\boldsymbol{\Omega}) = \prod_{i=1}^{N}\int \exp(l_i)d\boldsymbol{\eta}_i, \tag{2.10}$$

where

$$l_i = -\frac{1}{2}\sum_{j=1}^{n_i}\left(\boldsymbol{\epsilon}_{ij}^T\mathbf{R}_{ij}^{-1}\boldsymbol{\epsilon}_{ij} + \log|\mathbf{R}_{ij}|\right) - \frac{1}{2}\boldsymbol{\eta}_i^T\boldsymbol{\Omega}^{-1}\boldsymbol{\eta}_i - \frac{1}{2}\log|\boldsymbol{\Omega}| + C, \tag{2.11}$$

where $\boldsymbol{\epsilon}_{ij}$ denotes the residual of the jth observation of the ith individual, $\mathbf{R}_{ij}$ the covariance of the corresponding residual, and where the constant C does not depend on the model parameters. Without loss of generality we set $C = 0$. Note that the residuals $\boldsymbol{\epsilon}_{ij}$ depend on $\boldsymbol{\eta}_i$, and so may the covariances $\mathbf{R}_{ij}$.

To calculate the population likelihood, we have to calculate the integrals $\int \exp(l_i) d\boldsymbol{\eta}_i$, $i = 1, \ldots, N$. This is a computationally demanding task and there is only in rare cases that an analytical expression exists. Many approaches utilizes approximations on different levels.

Approximation of the population likelihood

In this thesis we consider approximation of the integral by utilizing Laplace's approximation [24, 25]. The Laplace approximation uses a second order Taylor expansion of l_i around a point $\boldsymbol{\eta}_{i0}$. That is

$$l_i(\boldsymbol{\eta}_i) \approx l_i(\boldsymbol{\eta}_{i0}) + \nabla l_i(\boldsymbol{\eta}_{i0})(\boldsymbol{\eta} - \boldsymbol{\eta}_{i0}) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_{i0})^T \Delta l_i(\boldsymbol{\eta}_{i0})(\boldsymbol{\eta} - \boldsymbol{\eta}_{i0}), \tag{2.12}$$

where $\nabla l_i(\boldsymbol{\eta}_{i0})$ and $\Delta l_i(\boldsymbol{\eta}_{i0})$ denote the gradient and the Hessian of the individual likelihood evaluated at $\boldsymbol{\eta}_{i0}$, respectively. There are several possible choices to choose the point $\boldsymbol{\eta}_{i0}$. One of the simplest approximations is to choose $\boldsymbol{\eta}_{i0}$ to be equal to the expectation of $\boldsymbol{\eta}_i$, that is

$$\boldsymbol{\eta}_{i0} = \mathrm{E}(\boldsymbol{\eta}_i | \boldsymbol{\Omega}) = \mathbf{0}. \tag{2.13}$$

Another approximation, is to choose $\boldsymbol{\eta}_{i0}$ to be the mode of (2.11). In this case, called *conditional estimation*, the gradient of the individual likelihood vanishes and the approximate expression for the integral $\int \exp(l_i) d\boldsymbol{\eta}_i$ depends only on the individual likelihood (2.11) and the Hessian of the individual likelihood. The approximate likelihood becomes

$$L(\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\Omega} | \boldsymbol{y}) = p(\boldsymbol{y} | \boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\Omega}) \approx \prod_{i=1}^{N} \exp(l_i(\boldsymbol{\eta}_i^*)) \left| \frac{-\Delta l_i(\boldsymbol{\eta}_i^*)}{2\pi} \right|^{-1/2}, \tag{2.14}$$

where $\boldsymbol{\eta}_i^*$ maximizes l_i given $\boldsymbol{\theta}$, $\mathbf{R}$, and $\boldsymbol{\Omega}$. By taking the logarithm the approximate population likelihood is

$$\log L(\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\Omega} | \boldsymbol{y}) \approx \sum_{i=1}^{N} l_i(\boldsymbol{\eta}_i^*) - \frac{1}{2} \log \left| \frac{-\Delta l_i(\boldsymbol{\eta}_i^*)}{2\pi} \right|. \tag{2.15}$$

Due to difficulties to compute the Hessian matrix Δl_i, a second level of approximation is to approximate the Hessian as well. Utilizing a so called *first order* approximation, only first order derivatives in the expression of the Hessian matrix are kept. In the general case the covariance matrices may depend on $\boldsymbol{\eta}_i$ which gives an approximate expression of the Hessian matrix that includes first order derivatives of the residuals and the covariances with respect to $\boldsymbol{\eta}_i$. The conditional estimation together with keeping of all first order derivatives is known as the *first order conditional estimation with interaction* (FOCEI) method. If no interaction (dependence) between the output covariance and

the random effects is assumed, the method is known as the *first order conditional estimation* (FOCE) method. For a more detailed derivation of the objective function, we refer the reader to [25].

The (approximate) maximum likelihood estimate is found by maximizing the approximate population likelihood with respect to the parameters in the model. A common approach to maximize the objective function is to utilize a gradient-based search method together with a finite difference approximation of the gradient of the objective function.

Summary of Papers

In this thesis the nonlinear mixed-effects model has been further developed in two aspects.

First, we have extended the dynamic model to include stochastic differential equations. The combination of mixed-effects models and stochastic differential equations is the topic of Paper II. A stochastic differential equation provide means to allow for a more general model specification. By combining the mixed-effects model and stochastic differential equations three sources of variability can be described. Moreover, as investigated in Paper I, a stochastic differential equation can also be used as a regularization tool in the parameter estimation problem.

Secondly, we have tackled the optimization problem of the approximate population likelihood. A very common approach in the optimization of the approximate population likelihood is to use a gradient-based optimization method. To calculate the gradient, one often relies on approximations based on finite differences. However, finite difference approximations might become an unreliable description of the gradient due to the numerical solutions of the model equations. Numerical ODE solvers using adaptive step length are known to introduce quantification errors to the objective function, making it non-smooth on small scales [3]. Instead, to overcome such problems, the gradient can be determined by formally differentiating the objective function. This approach is known as sensitivity equations and has so far, as known by the authors, not been done in a mixed-effects model setting. The use of sensitivity equations in nonlinear mixed-effects model is described in Paper III.

3.1 Summary of Paper I

In this paper we consider the single individual case, and hence we hereby drop the notation i. In many applications, the underlying dynamics are described by a system of ordinary differential equations of the general form

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})\, dt, \tag{3.1}$$

where the differential equation is written on differential form as this is the standard notation for stochastic differential equations. The data is assumed to be described by a set of measurement equations

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k, \quad k = 1, \ldots, N. \tag{3.2}$$

Estimation of model parameters in ordinary differential equations given discrete time measurement data is a complex problem, due to the nonlinearity of model parameters. As discussed by Schittkowski [19], 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.

In this paper, we propose a novel method used to regularize the objective function used for parameter estimation in dynamical systems. We consider extension of the underlying model to be described by a stochastic differential equation. Stochastic differential equations have received great attention in a large number of fields, including finance, pharmaceutics, and systems biology. Stochastic differential equations serve as a natural way of introducing uncertainty in 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. For an introduction to the theory and numerical solution of stochastic differential equations, see [14].

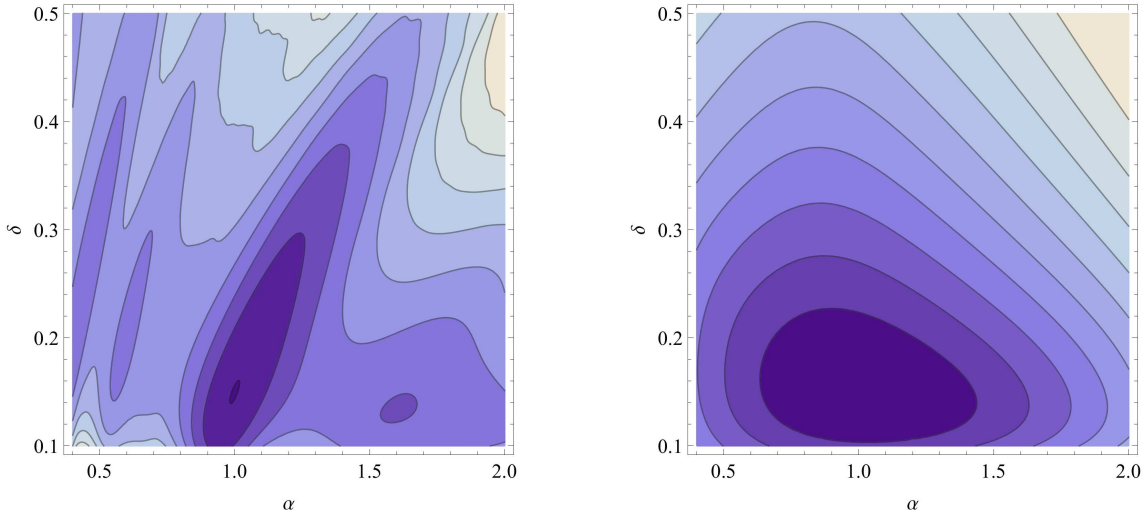
Formally, we extend the ordinary differential equation (3.1) to a stochastic differential equation on the form

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})\, dt + \boldsymbol{\Sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})\, d\boldsymbol{W}. \tag{3.3}$$

In the stochastic differential equation, $\boldsymbol{W}$ is a Wiener process and we refer to $\boldsymbol{\Sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{W}$ as the system noise. 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 SDE include the ODE case, which corresponds to $\boldsymbol{\Sigma} = \boldsymbol{0}$. The advantage of using stochastic differential equations is that the actual states in the model can be predicted from data and this will usually keep the prediction close to data even when the parameters in the model are incorrect.

We consider parameter estimation using the maximum likelihood approach. Using two different models from mathematical biology (FitzHugh-Nagumo for excitable media and the Lotka-Volterra predator-prey model) we investigate the impact of the system noise on the objective function. The approximate likelihood is derived using the extended Kalman filter, which is recursive algorithm for state estimation in nonlinear state-space models [9]. To maximize the objective function, a gradient-based search method is proposed. In contrast to the commonly used finite difference approximation of the gradient, sensitivity equations for the underlying system and the filter updating equations are utilized for an robust and accurate gradient calculation.

By considering an objective function depending on two model parameters a visualization of the complex likelihood function can be obtained. In the two models of interest, the ordinary differential equation setting gives an objective function with several local minima. By utilizing the derived sensitivity equations, a number of gradient-based optimizations is performed for each model. The two examples reveal that the objective function can 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.



(a) *ODE* (b) *SDE*

Figure 3.1: Illustration of the regularization using stochastic differential equation in the Lotka-Volterra predator-prey model. The two plots show the level curves of the objective function in the parameter estimation problem using the ordinary differential equation setting (a) and the stochastic differential equation setting (b). Note that in (a) there are several local minima to where an local optimization algorithm might converge.

The conclusion in the paper is 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. Note that in this paper, we only consider objective functions depending on two parameters. This choice was 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 exhibit problems with local minima.

3.2 Summary of Paper II

In this paper we consider the nonlinear mixed-effects model described in the introduction.

In [6], the authors stress that it is vital to recognize that published applications of nonlinear mixed-effects models are almost always predicated on correctness of $\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{u}_i, t, \boldsymbol{\phi}_i)$. Using ordinary differential equations, the underlying model is seen as true and the only error in the mixed-effects model is incorporated as measurement noise. However, in many applications there may also exist an uncertainty in the dynamics itself. In recent years, there has been several articles describing an extension of the classical mixed-effects model framework to include stochastic differential equations. As described in Paper I, the stochastic differential equations can be used 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.

In this paper, we propose a combination of the nonlinear mixed-effects model and

stochastic differential equations to allow for uncertainty in the model dynamics. As argued in [2], the random part of the model may be due to true randomness in parameters and state variables over time. Also, in systems without true randomness, SDEs can be used to model an incomplete or imperfect model structure. By combining stochastic differential equations and mixed-effects modeling three sources of variability can be distinguished; measurement error, parameter variability and model uncertainty. As in the regular nonlinear mixed-effects model we consider measurement error on the individual level and parameter variability on the population level. However, since stochastic differential equations includes a stochastic part, introduced as the system noise in Paper I, we also consider variability in the underlying dynamics.

Parameter estimation in the stochastic mixed-effects framework is a complex problem. On a population level, one have to rely on approximations of the population likelihood, as described in the introduction. Moreover, since the state of the system is uncertain when using stochastic differential equations it has to be estimated from observed data. Several approaches has been proposed, see e.g. [15, 16, 2, 7]. In [12, 10], the authors propose a combination of the first-order conditional estimation approximation of the population likelihood, together with the extended Kalman filter (see e.g. [9]) for state estimation.

Paper II serves as an introduction to pharmacokinetic modeling using stochastic differential equations. In this paper, we have extended the analysis to include the first-order conditional estimation with interaction approximation of the population likelihood together with correlation between random effects parameters. One of the major contributions to the parameter estimation problem in mixed-effects models is the utilization of sensitivity equations to calculate the gradient in the optimization problem. This was not the main objective of Paper II, and the new gradient method is instead derived in Paper III.

As mentioned in the introduction, the nonlinear mixed-effects model is frequently used for data analysis in pharmaceutical applications. Pharmacokinetics is a large field of pharmacometrics, and is used to model the administration, distribution, metabolism, and elimination of drugs. In Paper II, we consider two different pharmacokinetic models. First, we simulate a data set from a known model and estimate the model parameters from the simulated data. This serves a proof-of-concept of the derived method and we investigate whether the model parameters and the three sources of variability can be estimated from data. We also investigate the impact of assuming a deterministic model and neglecting the system noise. Secondly, we extend a previously published pharmacokinetic model of nicotinic acid (NiAc) in obese Zucker rats [1] to a stochastic pharmacokinetic model. The aim of the investigation is to compare the previously published model with the stochastic model, in terms of parameter estimates and model prediction.

To estimate the model parameters in the stochastic mixed-effects model framework we have developed the necessary functionality in Mathematica, which also features nonlinear mixed-effects modeling using ordinary differential equations. The estimation results from the simulated model show that the model parameters and the three sources of variability can be estimated from the data. Moreover, if the stochastic part is neglected the measurement variance is significantly overestimated. The results for the stochastic pharmacokinetic NiAc disposition model show that the error previously

described as pure measurement error can be divided into a reduced measurement error and a significant system noise. The significant system noise implies that the deterministic model structure may not be sufficient to be describe the observed data. The extension of the deterministic model to decrease the system noise has not been investigated in the current paper.

3.3 Summary of Paper III

To estimate the parameters in the nonlinear mixed-effect model described in the introduction, the approximate population likelihood has to be maximized with respect to the parameters in the model. Moreover, due to the conditional estimation methods (FOCE, FOCEI) the approximation of the population likelihood requires another optimization of the individual likelihoods. Further on, we will refer to the optimization of the approximate population likelihood with respect to the model parameters as the *outer* optimization problem and optimization of individual likelihoods with respect to the random effects as the *inner* optimization problem. For each step in the outer optimization problem the inner optimization problem has to be performed for each individual in the population.

In many applications, the optimization problem is solved by adopting a gradient-based method, for which a popular method is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) updating formula [13]. Using a gradient-based method, an expression for the gradient of the objective function is necessary. As noted in [16], three different solutions to the problem is of interest, namely finite difference approximation, symbolic differentiation and automatic differentiation.

To calculate the gradients needed in the optimizations (outer and inner) the most common approach is to utilize a finite difference approximation. However, due to the adaptive schemes for solving the underlying ordinary differential equations, finite difference approximations can be unreliable on small scales [3]. In Paper III, we have considered symbolic differentiation of the approximate population likelihood. By formally differentiating the objective function, expressions of the gradient can be obtained, which is achieved by repeated use of the chain rule. To our knowledge, this has not been done in mixed-effects models until now.

In Paper III, we derive the gradients of the objective functions in the inner and outer optimization problem using the FOCEI (and FOCE, which is a special case of FOCEI) approximation of the population likelihood. Derivation of the objective functions implies the need of derivation of the system of ordinary differential equations describing the dynamics, which results in the so called sensitivity equations of the original system of ordinary differential equations. Consider a system of ordinary differential equations depending on a parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}). \tag{3.4}$$

To obtain the sensitivity of a parameter θ_k both sides of the ordinary differential equation is differentiated with respect to θ_k. Hence

$$\frac{d}{d\theta_k}\frac{d\boldsymbol{x}}{dt} = \frac{d}{d\theta_k}\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}). \tag{3.5}$$

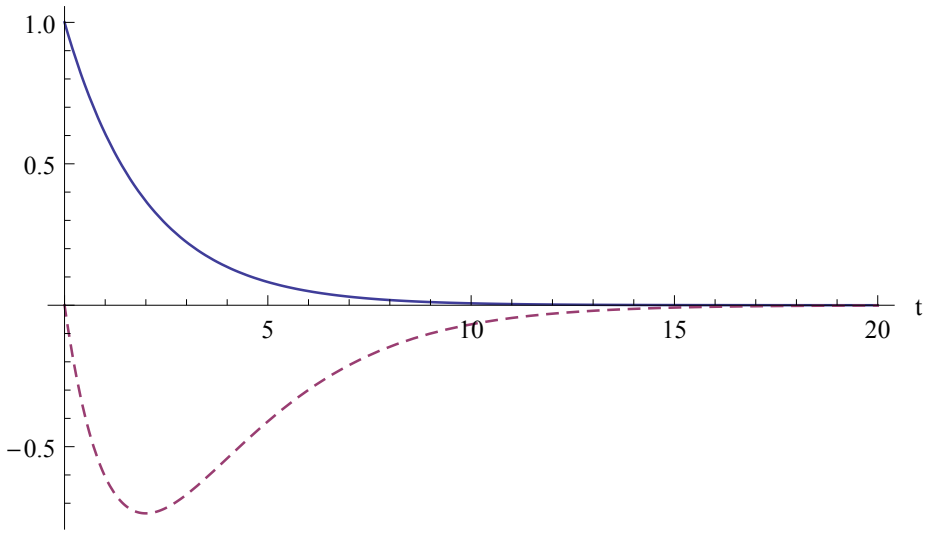


Figure 3.2: Illustration of sensitivity equations. The solution to the differential equation $\frac{dA}{dt} = -kA$ (blue, solid) and its corresponding sensitivity with respect to the parameter k (purple, dashed). The initial condition is $A(0) = 1$ and the parameter value is $k = 0.5$.

Rearranging the order of integration on the left side and expanding the right side we get

$$\frac{d}{dt}\frac{d\boldsymbol{x}}{d\theta_k} = \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{x}}\frac{d\boldsymbol{x}}{d\theta_k} + \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_k}. \tag{3.6}$$

By solving the original system of ordinary differential equations together with the sensitivity equations we are able to form the expressions of the gradients needed. This leads to a larger system of ordinary differential equations to solve. However, this system is only solved once to obtain both the value of the objective function and its gradient, in contrast to the finite difference case where the original system is solved several times to approximate the gradient using the finite difference approximation.

The proposed method is evaluated using two different pharmacokinetic models. We consider gradient evaluation using finite difference approximation and compare it with the cases where sensitivity equations are used in the inner optimization problem, the outer optimization problem, and both inner and outer, respectively. We are not interested in the running time *per se*, but merely the relative time differences between the cases.

Bibliography

[1] AHLSTRÖM, C., KROON, T., PELETIER, L. A., AND GABRIELSSON, J. Feedback modeling of non-esterified fatty acids in obese Zucker rats after nicotinic acid infusions. *J Pharmacokinet Pharmacodyn 40*, 6 (2013), 623–638.

[2] BERGLUND, M., SUNNÅKER, M., ADIELS, M., JIRSTRAND, M., AND WENNBERG, B. Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Math Med Biol 29*, 4 (2012), 361–384.

[3] BOHLIN, T. *Practical grey-box process identification: Theory and applications*. Springer-Verlag, 2006.

[4] CHEN, G., SAAD, Z. S., BRITTON, J. C., PINE, D. S., AND COX, R. W. Linear mixed-effects modeling approach to FMRI group analysis. *NeuroImage 73* (Jun 2013), 176–190.

[5] DAVIDIAN, M., AND GILTINAN, D. M. *Nonlinear Models for Repeated Measurement Data (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1995.

[6] DAVIDIAN, M., AND GILTINAN, D. M. Nonlinear models for repeated measurement data: An overview and update. *J Agric Biol Environ Stat 8*, 4 (2003), 387–419.

[7] DELATTRE, M., AND LAVIELLE, M. Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat Interface 6* (2013), 519–532.

[8] DRAPER, N. R., AND SMITH, H. *Applied Regression Analysis (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 1998.

[9] JAZWINSKY, A. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[10] KLIM, S., MORTENSEN, S. B., KRISTENSEN, N. R., STIG OVERGAARD, R. V., AND MADSEN, H. Population stochastic modelling (PSM)–an R package for mixed-effects models based on stochastic differential equations. *Comput Methods Programs Biomed 94*, 3 (Jun 2009), 279–289.

[11] LINDSTROM, M. L., AND BATES, D. M. Nonlinear mixed effects models for repeated measures data. *Biometrics 46*, 3 (Sep 1990), 673–687.

[12] MORTENSEN, S. B., KLIM, S., DAMMANN, B., KRISTENSEN, N. R., MADSEN, H., AND OVERGAARD, R. V. A matlab framework for estimation of NLME models using stochastic differential equations: applications for estimation of insulin secretion rates. *J Pharmacokinet Pharmacodyn 34*, 5 (Oct 2007), 623–642.

[13] NOCEDAL, J., AND WRIGHT, S. *Numerical Optimization*. Springer-Verlag New York, 1999.

[14] ØKSENDAL, B. *Stochastic Differential Equations. An Introduction with Applications*. Springer, 2003.

[15] OVERGAARD, R. V., JONSSON, N., TORNØE, C. W., AND MADSEN, H. Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *J Pharmacokinet Pharmacodyn 32*, 1 (2005), 85–107.

[16] PICCHINI, U., AND DITLEVSEN, S. Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput Stat Data Anal 55*, 3 (Mar 2011), 1426–1444.

[17] PINHEIRO, J. C., AND BATES, D. *Mixed-Effects Models in S and S-PLUS*. Springer, 2000.

[18] ROHNER, B., BUGMANN, H., AND BIGLER, C. Estimating the age-diameter relationship of oak species in Switzerland using nonlinear mixed-effects models. *Eur J Forest Res 132*, 5-6 (Nov 2013), 751–764.

[19] SCHITTKOWSKY, K. *Numerical Data Fitting in Dynamical Systems - A Practic Introduction with Applications and Software*. Kluwer Academic Publishers, 2002.

[20] SHEINER, L. B., AND BEAL, S. L. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm 8*, 6 (Dec 1980), 553–571.

[21] SHEINER, L. B., AND BEAL, S. L. Evaluation of methods for estimating population pharmacokinetic parameters II. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm 9*, 5 (Oct 1981), 635–651.

[22] SHEINER, L. B., AND BEAL, S. L. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm 11*, 3 (Jun 1983), 303–319.

[23] SHEINER, L. B., ROSENBERG, B., AND MARATHE, V. V. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm 5*, 5 (Oct 1977), 445–479.

[24] VONESH, E. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika 83*, 2 (Jun 1996), 447–452.

[25] WANG, Y. Derivation of various NONMEM estimation methods. *J Pharmacokinet Pharmacodyn 34*, 5 (Oct 2007), 575–593.