Least Squares Methods for Differential Equation based Models and Massive Data Sets

Josef Kallrath
BASF Aktiengesellschaft, GVCS, B009, D-67056 Ludwigshafen
e-mail: josef.kallrath@web.de

June 29, 2007

1 Introduction

Least squares problems and solution techniques to solve them have a long history briefly addressed by Björck (2001, [4]). In this article we focus on two classes of complex least squares problems. The first one is established by models involving *differential equations*. The other class is made by least squares problems involving difficult models which need to be solved for many independent observational data sets. We call this *least squares problems with massive data sets*.

1.1 A Standard Formulation for Unconstrained Least Squares Problem

The unconstrained least squares problem can be expressed by

$$\min_{\mathbf{p}} l_2(\mathbf{p}) \quad , \quad l_2(\mathbf{p}) := \|\mathbf{r}_1\left[\mathbf{x}(t_1),\ldots,\mathbf{x}(t_k),\mathbf{p}\right]\|_2^2 = \sum_{k=1}^{N}\left[r_{1k}(\mathbf{p})\right]^2 \quad , \quad \mathbf{r}_1 \in \mathbb{R}^N \quad . \tag{1}$$

The minimization of this functional, *i.e.*, the minimization of the sum of weighted quadratic residuals, under the assumption that the statistical errors follow a Gaußian distribution with variances as in (4), provides a maximum likelihood estimator (Brandt, 1976, [7] Chp.7) for the unknown parameter vector $\mathbf{p}$. This objective function dates back to Gauß (1809, [14]) and in the mathematical literature the problem is synonymously called least squares or ℓ_2 approximation problem.

The least squares structure (1) may arise either from a nonlinear overdetermined system of equations

$$r_{1k}(\mathbf{p}) = 0 \quad , \quad k = 1,...,N \quad , \quad N > n \quad , \tag{2}$$

or from a data fitting problem with N given data points $(t_k, \tilde{Y}_k)$ and variances σ_ν, a model function $\tilde{F}(t, \mathbf{p})$, and n adjustable parameters $\mathbf{p}$:

$$r_{1k} := r_{1k}(\mathbf{p}) = Y_k - F_k(\mathbf{p}) = \sqrt{w_k}\left[\tilde{Y}_k - \tilde{F}(t_k, \mathbf{p})\right] \quad . \tag{3}$$

The weights w_k are related to the variances σ_k by

$$w_k := \beta/\sigma_k^2 \quad . \tag{4}$$

Traditionally, the weights are scaled to a variance of unit weights. The factor β is chosen so as to make the weights come out in a convenient range. In short vector notation we get

$$\mathbf{r_1} := \mathbf{Y} - \mathbf{F}(\mathbf{p}) = [r_{11}(\mathbf{p}), \ldots, r_{1N}(\mathbf{p})]^{\mathrm{T}} \quad , \quad \mathbf{F}(\mathbf{p}), \mathbf{Y} \in \mathbb{R}^N \quad .$$

Our least squares problem requires us to provide the following input:

1. model,
2. data,
3. variances associated with the data,
4. measure of goodness of the fit, *e.g.*, the Euclidean norm.

In many practical applications, unfortunately, less attention is paid to the variances. It is also very important to point out that the use of the Euclidean norm requires pre-information related to the problem and statistical properties of the data.

1.2 Solution Methods

Standard methods for solving linear version of (1), *i.e.*, $\mathbf{F}(\mathbf{p}) = \mathsf{A}\mathbf{p}$, are reviewed by Björck (2001, [4]). Nonlinear methods for unconstrained least squares problems are covered in detail by Xu (2001a, 2001b, 2001c; [36], [35], [37]). In addition, we mention a popular method to solve unconstrained least squares problems: the Levenberg-Marquardt algorithm proposed independently by Levenberg (1944, [21]) and Marquardt (1963, [22]) and sometimes also called "damped least squares". It modifies the eigenvalues of the normal equation matrix and tries to reduce the influence of eigenvectors related to small eigenvalues [*cf.* Dennis and Schnabel (1983, [8])]. Damped (step-size cutting) Gauß-Newton algorithms combined with orthogonalization methods control the damping by natural level functions [Deuflhard and Apostolescu (1977, 1980, [9][10]), Bock (1987)] seem to be superior to Levenberg-Marquardt type schemes and can be more easily extended to nonlinear constrained least squares problems.

1.3 Explicit Versus Implicit Models

A common basic feature and limitation of least squares methods, but seldom explicitly noted, is that they require some *explicit* model to be fitted to the data. However, not all models are explicit. For example, some pharmaceutical applications for receptor-ligand binding studies are based on specifically coupled

mass equilibrium models. They are used, for instance, for the radioimmunological determination of Fenoterol or related substances, and lead to least squares problems in systems of nonlinear equations [31], in which the model function $\mathbf{F}(\mathbf{p})$ is replaced by $\mathbf{F}(t; \mathbf{p}, \mathbf{z})$ which, besides the parameter vector $\mathbf{p}$ and the time t, depends on a vector function $\mathbf{z} = \mathbf{z}(t; \mathbf{p})$ implictly defined as the solution of the nonlinear equations

$$\mathbf{F}_2(t; \mathbf{p}, \mathbf{z}) = \mathbf{0} \quad , \quad \mathbf{F}_2(\mathbf{p}) \in \mathbb{R}^{n_2} \quad . \tag{5}$$

This is a special case of an implicit model. There is a much broader class of implicit models. Most models in science are based on physical, chemical and biological laws or include geometry properties, and very often lead to differential equations which may, however, not be solvable in a closed analytical form. Thus, such models do not lead to explicit functions or models we want to fit to data. We rather need to fit an implicit model (represented by a system of differential equations or another implicit model). The demand for and the applications of such techniques are widespread in science, especially in the rapidly increasing fields of nonlinear dynamics in physics and astronomy, nonlinear reaction kinetics in chemistry [5], nonlinear models in material sciences (Kallrath *et al.*, 1998, [16]) and biology [2], and nonlinear systems describing ecosystems [Richter and Söndgerath (1990, [29]), Richter *et al.* (1992, [28])] in biology, or the environmental sciences. Therefore, it seems desirable to focus on least squares algorithms that use nonlinear equations and differential equations as constraints or side conditions to determine the solution implicitly.

1.4 Practical Issues of Solving Least Squares Problems

Solving least squares problems involves various difficulties among them to find an appropriate model, non-smooth models with discontinuous derivatives, data quality and checking the assumption of the underlying error distribution, and dependence on initial parameter or related questions of global convergence.

1.4.1 Models and Model Validation

A model may be defined as an appropriate abstract representation of a real system. In the natural sciences (*e.g.*, Physics, Astronomy, Chemistry and Biology) models are used to gain a deeper understanding of processes occurring in nature (an epistemological argument). The comparison of measurements and observations with the predictions of a model is used to determine the appropriateness and quality of the model. Sir Karl Popper (1980, [26]) in his famous book *Logic of Scientific Discovery* uses the expressions *falsification* and *verification* to describe tasks that the models can be used to accomplish as an aid to scientific process. Models were used in early scientific work to explain the movements of planets. Then, later, aspects and questions of accepting and improving global and fundamental models (*e.g.,* general relativity or quantum physics) formed part of the discussion of the philosophy of science. In science models are usually falsified, and, eventually, replaced by modified or completely different ones.

In industry, models have a rather local meaning. A special aspect of reality is to be mapped in detail. Pragmatic and commercial aspects are usually the motivation. The model maps most of the relevant features and neglect less important aspects. The purpose is to

- provide insight into the problem,
- allow numerical, *virtual* experimentation but avoid expensive and/or dangerous *real* experiments, or
- tune a model for later usage, *i.e.*, determine, for instance, the reaction coefficients of a chemical system – once these parameters are known the dynamics of the process can be computed.

A (mathematical) model represents a *real-world problem* in the language of mathematics, *i.e.,* by using mathematical symbols, variables (in this context: the adjustable least squares parameters), equations, inequalities, and other relations. How does one get a mathematical model for a real-world problem? To achieve that is neither easy nor unique. In some sense it is similar to solving exercises in school where problems are put in a verbal way [25]. The following points are useful to remember when trying to build a model:

- there will be no precise recipe telling the user how to build a model,
- experience and judgment are two important aspects of model building,
- there is nothing like a *correct* model,
- there is no concept of a *unique* model, as different models focusing on different aspects may be appropriate.

Industrial models are eventually validated which means that they reached a sufficient level of consensus among the community working with these models.

Statistics provide some means to discriminate models but this still is an art and does not replace the need for appropriate model validation. The basic notion is: *with a sufficient number of parameters on can fit an elefant.* This leads us to one important consequence: it seems to be necessary that one can interpret these model parameters. A reasonable model derived from the laws of science with interpretable parameters is a good candidate to become accepted. Even, if it may lead to a somewhat worse looking fits than a model with a larger number of formal parameters without interpretation.

1.4.2 Non-Smooth Models

The algorithm reviewed by Xu (2001a, 2001b, 2001c; [36], [35], [37]) for solving least squares problems usually require the continuous first derivatives of the model function with respect to the parameters. We might, however, encounter models for which the first derivatives are discontinuous. Derive-free methods such as Nelder & Mead's (1965, [23]) downhill Simplex method, or direction set

methods; *cf.* Press *et al.* (1992, [27], p.406) have been successfully used to solve least squares problems. The Simplex method provides the benefit of exploring parameter space and good starting values for derivative based methods. Powell's direction set method with appropriate conjugate directions preserve the derivative free nature of the method.

1.4.3 Global Convergence

Nonlinear least squares algorithms usually converge only if the initial parameters are close to the best fit parameters. Global convergence can be established for some algorithms, *i.e.*, they converge for all initial parameters. An essential support tool accompanying the analysis of difficult least squares problem is to visualize the data and the fits. Inappropriate or premature fits can easily be excluded. Inappropriate fits are possible because all algorithms mentioned in Sect. 1, 2, and 3 are local algorithm. Only if the least squares problem is convex, they yield the global least squares minimum. Sometimes, it is possible to identify false local minima from the residuals.

1.4.4 Data and Data Quality

Least squares analysis is concerned by fitting data to a model. The data are not exact but subject to unknown random errors ε_k. In ideal cases these errors follow a Gaussian normal distribution. One can test this assumption after the least squares fit by analyzing the distrbution of the residuals as described in Sect. 1.4.5. Another important issue is whether the data are appropriate to estimate all parameters. Experimental design is the discipline which addresses this issue.

1.4.5 Residual Distributions, Covariances and Parameter Uncertainties

Once the minimal least squares solution has been found one should at first check with the χ^2-test or Kolmogoroff-Smirnov test whether the usual assumption that the distribution really follows a Gaussian normal distribution. With the Kolmogoroff-Smirnov test [see, *e.g.*, Ostle (1963, [24])] it is possible to check as follows whether the residuals of a least-squares solution are normally distributed around the mean value 0.

1. let $M := (x_1, x_2, ..., x_n)$ be a set of observations for which a given hypothesis should be tested;
2. let $G : x \in M \rightarrow \mathbb{R}$, $x \rightarrow G(x)$, be the corresponding cumulative distribution function;
3. for each observation $x \in M$ define $S_n(x) := k/n$, where k is the number of observations less than or equal to x;
4. determine the maximum $D := \max(G(x) - S_n(x) \mid x \in M)$;

5. D_{crit} denotes the maximum deviation allowed for a given significance level and a set of n elements. D_{crit} is tabulated in the literature, *e.g.*, [24, Appendix 2, p. 560]; and

6. if $D < D_{\text{crit}}$, the hypothesis is accepted.

For the least squares problem formulated in Sect. 1.1 the hypothesis is "The residuals $\mathbf{x} := \mathbf{r_1} = \mathbf{Y} - \mathbf{F}(\mathbf{p})$ are normally distributed around the mean value 0." Therefore, the cumulative distribution function $G(x)$ takes the form

$$\sqrt{2\pi}G(x) = \int_{-\infty}^{x} g(z)\mathrm{d}z = \int_{-\infty}^{-x_0} g(z)\mathrm{d}z + \int_{-x_0}^{x} g(z)\mathrm{d}z, \quad g(z) := e^{-\frac{1}{2}z^2} \quad .$$

The value x_0 separates larger residuals; this is problem specific control parameter.

The derivative based least squares methods usually also give the covariance matrix from which the uncertainties of the parameter are derived; *cf.* (Brandt, 1976, [7], Chp.7). Least squares parameter estimations without quantifying the uncertainty of the parameters are very doubtful.

2 Parameter Estimation in ODE Models

Consider a differential equation with independent variable t for the state variable

$$\mathbf{x}'(t) = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{x}, \mathbf{p}) \quad , \quad \mathbf{x} \in \mathbb{R}^{n_d} \quad , \quad \mathbf{p} \in \mathbb{R}^{n_p} \tag{1}$$

with a right hand side depending on an unknown parameter vector $\mathbf{p}$. Additional requirements on the solution of the ODE (1) like periodicity, initial or boundary conditions or range restrictions to the parameters can be formulated in vectors $\mathbf{r}_2$ and $\mathbf{r}_3$ of (component wise) equations and inequalities

$$\mathbf{r}_2\left[\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{p}\right] = 0 \quad \text{or} \quad \mathbf{r}_3\left[\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{p}\right] \geq 0 \quad . \tag{2}$$

The multi-point boundary value problem is linked to experimental data via minimization of a least squares objective function

$$l_2(\mathbf{x}, \mathbf{p}) := \left\|\mathbf{r}_1\left[\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{p}\right]\right\|_2^2 \quad . \tag{3}$$

In a special case of (3) the components ℓ of the vector $\mathbf{r}_1 \in \mathbb{R}^L$ are "equations of condition" and have the form

$$r_{1\ell} = \sigma_{ij}^{-1}[\eta_{ij} - g_i(\mathbf{x}(t_j), \mathbf{p})] \quad , \quad \ell = 1, \ldots, L := \sum_{i=1}^{N_j} J_i \quad . \tag{4}$$

This case leads us to the least squares function

$$l_2(\mathbf{x}, \mathbf{p}) := \sum_{j=1}^{N^D} \sum_{i=1}^{N_j} \sigma_{ij}^{-2}[\eta_{ij} - g_i(\mathbf{x}(t_j), \mathbf{p})]^2 \quad . \tag{5}$$

Here, N^D denotes the number of values of the independent variable (here called time) at which observed data are available, N_j denotes the number of observables measured at time t_j and η_{ij} denotes the observed value which is compared with the value of observable i evaluated by the model where the functions $g_i(\mathbf{x}(t_j), \mathbf{p})$ relate the state variables to $\mathbf{x}$ this observable

$$\eta_{ij} = g_i(\mathbf{x}(t_j), \mathbf{p}) + \varepsilon_{ij} \quad . \tag{6}$$

The numbers ε_{ij} are the measurement errors and σ_{ij}^2 are weights that have to be adequately chosen due to statistical considerations, *e.g.* as the variances. The unknown parameter vector $\mathbf{p}$ is determined from the measurements such that the model is optimally adjusted to the measured (observed) data. If the errors ε_{ij} are independent, normally distributed with the mean value zero and have variances σ_{ij}^2 (up to a common factor β^2), then the solution of the least squares problem is a *maximum likelihood estimate*.

2.1 The Initial Value Problem Approach

An obvious approach to estimate parameters in ODE which is also implemented in many commercial packages is the initial value problem approach. The idea is to guess parameters and initial values for the trajectories, compute a solution of an initial value problem (IVP) (1) and iterate the parameters and initial values in order to improve the fit. Characteristic features and disadvantages are discussed in, *e.g.,* [6] or Kallrath *et al.* (1993, [18]). In the course of the iterative solution one has to solve a sequence of IVPs. The state variable $\mathbf{x}(t)$ is eliminated for the benefit of the unknown parameter $\mathbf{p}$ and the initial values. Note that no use is made of the measured data while solving the IVPs. They only enter in the performance criterion. Since initial guesses of the parameters may be poor, this can lead to IVPs which may be hard to solve or even have no solution at all and one can come into badly conditioned regions of the IVPs, which can lead to the loss of stability.

2.2 The Boundary Value Problem Approach

Alternatively to the IVP approach, in the "boundary value problem approach" invented by Bock (1981), the inverse problem is interpreted as an over-determined, constrained, multiple-point boundary problem. This interpretation does not depend on whether the direct problem is an initial or boundary value problem. The algorithm used here consists of an adequate combination of a multiple shooting method for the discretization of the boundary value problem side condition in combination with a generalized Gauss-Newton method for the solution of the resulting structured nonlinear constrained least squares problem [Bock (1981, 1987); [5], [6]). Depending on the vector of signs of the state and parameter dependent switching functions $\mathbf{Q}$ it is even possible to allow piecewise smooth right hand side functions f, *i.e.*, differential equations with switching conditions

$$\mathbf{x}' = f(t, \mathbf{x}, \mathbf{p}; sign(\mathbf{Q}(t, \mathbf{x}, \mathbf{p}))) \quad , \tag{7}$$

where the right side may change discontinuously if the vector of signs of the switching functions $\mathbf{Q}$ changes. Such discontinuities can occur, *e.g.* as a result of unsteady changes of physical values. The switching points are in general given by the roots of the state-dependent components of the switching functions

$$\mathbf{Q}_i(t, \mathbf{x}, \mathbf{p}) = 0 \quad . \tag{8}$$

Depending on the stability behavior of the ODE and the availability of information about the process (measured data, qualitative knowledge about the problem, etc.) a grid $\mathcal{T}_m$

$$\mathcal{T}_m : \tau_1 < \tau_2 < ... < \tau_m \quad , \quad \Delta\tau_j := \tau_{j+1} - \tau_j \quad , \quad 1 \leq j \leq m-1, \tag{9}$$

of m multiple shooting nodes τ_j ($m-1$ subintervals I_j) is chosen. The grid is adapted to the problem and data and is defined such that it includes the measuring interval ($[\tau_1, \tau_m] = [t_0, t_f]$). Usually, the grid points τ correspond to values of the independent variable t at which observations are available but additional grid points may be chosen for strongly nonlinear models. At each node τ_j an IVP

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}, \mathbf{p}) \quad , \quad \mathbf{x}(t = \tau_j) = \mathbf{s}_j \in \mathbb{R}^{n_d} \tag{10}$$

has to be integrated from τ_j to τ_{j+1}. The $m-1$ vectors of (unknown) initial values $\mathbf{s}_j$ of the partial trajectories, the vector $\mathbf{s}_m$ representing the state at the end point and the parameter vector $\mathbf{p}$ are summarized in the (unknown) vector $\mathbf{z}$

$$\mathbf{z}^T := (\mathbf{s}_1^T, ..., \mathbf{s}_m^T, \mathbf{p}^T) \quad . \tag{11}$$

For a given guess of $\mathbf{z}$ the solutions $\mathbf{x}(t; \mathbf{s}_j, \mathbf{p})$ of the $m-1$ independent initial value problems in each sub interval I_j are computed. This leads to an (at first discontinuous) representation of $\mathbf{x}(t)$. In order to replace (1) equivalently by these $m-1$ IVPs matching conditions

$$\mathbf{h}_j(\mathbf{s}_j, \mathbf{s}_{j+1}, p) := \mathbf{x}(\tau_{j+1}; \mathbf{s}_j, \mathbf{p}) - \mathbf{s}_{j+1} = 0 \quad , \quad \mathbf{h}_j : \mathbb{R}^{2n_d+n_p} \rightarrow \mathbb{R}^{n_d} \tag{12}$$

are added to the problem. (12) ensures the continuity of the final trajectory $\mathbf{x}(t)$.

Replacing $\mathbf{x}(t_i)$ and $\mathbf{p}$ in (5) by $\mathbf{z}$ the least squares problem is reformulated as a nonlinear constrained optimization problem with the structure

$$\min_{\mathbf{z}} \left\{ \frac{1}{2} \|\mathbf{F}_1(\mathbf{z})\|_2^2 \, | \mathbf{F}_2(\mathbf{z}) = 0 \in \mathbb{R}^{n_2} \quad , \quad \mathbf{F}_3(\mathbf{z}) \geq 0 \in \mathbb{R}^{n_3} \right\} \quad , \tag{13}$$

wherein n_2 denotes the number of the equality and n_3 the number of the inequality constraints. This usually large constrained structured nonlinear problem is solved by a damped generalized Gauss-Newton method [5]. If $\mathsf{J}_1(\mathbf{z}_k) := \partial_z \mathbf{F}_1(\mathbf{z}_k)$, $\mathrm{J}_2(z_k) := \partial_z \mathbf{F}_2(\mathbf{z}_k)$ vis. $\mathsf{J}_3(\mathbf{z}_k) := \partial_z \mathbf{F}_3(\mathbf{z}_k)$ denote the Jacobi matrices of $\mathbf{F}_1$, $\mathbf{F}_2$ vis. $\mathbf{F}_3$, then the iteration proceeds as

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha_k \mathbf{\Delta z}_k \tag{14}$$

with damping constant $\alpha_k, 0 < \alpha_{min} \leq \alpha_k \leq 1$, and the increment $\mathbf{\Delta z}_k$ determined as the solution of the constrained linear problem

$$\min_{\mathbf{z}} \left\{ \frac{1}{2} \left\| \mathsf{J}_1(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_1(\mathbf{z}_k) \right\|_2^2 \;\middle|\; \begin{array}{l} \mathsf{J}_2(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_2(\mathbf{z}_k) = 0 \\ \mathsf{J}_3(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_3(\mathbf{z}_k) \geq 0 \end{array} \right\} \quad . \tag{15}$$

Global convergence can be achieved if the damping strategy is properly chosen [6].

The inequality constraints that are active in a feasible point are defined by the index set

$$\mathcal{I}(\mathbf{z}_k) := \{ i | F_{3i}(\mathbf{z}_k) = 0 \quad , \quad i = 1, ..., n_3 \} \quad . \tag{16}$$

The inequalities which are defined by the index set $\mathcal{I}(\mathbf{z}_k)$ or their derivatives are denoted with $\hat{\mathbf{F}}_3$ or $\hat{\mathsf{J}}_3$ in the following. In addition to (16) we define

$$\mathbf{F}_c := \begin{pmatrix} \mathbf{F}_2 \\ \hat{\mathbf{F}}_3 \end{pmatrix} \quad , \quad \mathsf{J}_c := \begin{pmatrix} \mathsf{J}_2 \\ \hat{\mathsf{J}}_3 \end{pmatrix} \quad . \tag{17}$$

In order to derive the necessary conditions that have to be fulfilled by the solution of the problem (13) the Lagrangian

$$L(\mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := \frac{1}{2} \|\mathbf{F}_1(\mathbf{z})\|_2^2 - \boldsymbol{\lambda}^T \mathbf{F}_2(\mathbf{z}) - \boldsymbol{\mu}^T \mathbf{F}_3(\mathbf{z}) \tag{18}$$

and the reduced Lagrangian

$$\hat{L}(\mathbf{z}, \boldsymbol{\lambda}_c) := \frac{1}{2} \|\mathbf{F}_1(\mathbf{z})\|_2^2 - \boldsymbol{\lambda}_c^T \mathbf{F}_c(\mathbf{z}) \quad , \quad \boldsymbol{\lambda}_c := \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu}_c \end{pmatrix} \tag{19}$$

are defined. The Kuhn-Tucker-conditions, *i.e.* the necessary conditions of first order, are the feasibility conditions

$$\mathbf{F}_2(\mathbf{z}^*) = 0 \quad , \quad \mathbf{F}_3(\mathbf{z}^*) \geq 0 \tag{20}$$

ensuring that $\mathbf{z}^*$ is feasible, and the stationarity conditions stating that the adjoined variables λ^*, μ^* exist as solution of the stationary conditions

$$\frac{\partial L}{\partial z}(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{F}_1^T(\mathbf{z}^*) \cdot \mathsf{J}(\mathbf{z}^*) - (\boldsymbol{\lambda}^*)^T \mathsf{J}_2(\mathbf{z}^*) - (\boldsymbol{\mu}^*)^T \mathsf{J}_3(\mathbf{z}^*) = 0 \tag{21}$$

and

$$\boldsymbol{\mu}^* \geq 0 \quad , \quad i \notin \mathcal{I}(\mathbf{z}^*) \Rightarrow {\boldsymbol{\mu}_i}^* = 0 \quad . \tag{22}$$

If $(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ fulfills the conditions (20), (21) and (22), it is called a Kuhn-Tucker-point and $\mathbf{z}^*$ a stationary point. The necessary condition of second order means that for all directions

$$\mathbf{s} \in T(\mathbf{x}^*) := \left\{ \mathbf{s} \neq 0 \;\middle|\; \begin{array}{l} \mathsf{J}_2(\mathbf{z}^*)\mathbf{s} = 0 \\ \mathsf{J}_3(\mathbf{z}^*)\mathbf{s} \geq 0 \end{array} \quad , \quad \mu_i \mathsf{J}_{3i}(\mathbf{z}^*)\mathbf{s} = 0 \right\} \tag{23}$$

the Hessian $\mathsf{G}(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of the Lagrangian is positive semi-definite:

$$\mathbf{s}^T \mathsf{G}(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{s} \geq 0 \quad , \quad \mathsf{G}(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) := \frac{\partial^2}{\partial \mathbf{z}^2} L(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad . \tag{24}$$

As $\mu_i = 0$ for $i \notin \mathcal{I}(\mathbf{z}^*)$ it is sufficient to postulate the stationary condition for the reduced Lagrangian (19). For the linear problem (15) follows: $(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Kuhn-Tucker-point of the nonlinear problem (13) if and only, if $(0, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Kuhn-Tucker-point of the linear problem. The necessary conditions for the existence of a local minimum of problem (13) are:

1. $(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Kuhn-Tucker-point of the non-linear problem
2. the Hessian $\mathsf{G}(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of the Lagrangian is positive definite for all directions $\mathbf{s} \in T(\mathbf{x}^*)$, vis. $\mathbf{s}^T G(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\mathbf{s} > 0$

If the necessary conditions for the existence of the local minimum and the condition $\mu_i \neq 0$ for $i \in \mathcal{I}(\mathbf{z}^*)$ are fulfilled, two perturbation theorems [6] can be formulated. If the sufficient conditions are fulfilled it can be shown for the neighborhood of a Kuhn-Tucker-point $(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of the nonlinear problem (13) that the local convergence behavior of the inequality constrained problem corresponds to that of the equality constrained problem which represents active inequalities and equations. Under the assumption of the regularity of the Jacobians J_1 and J_c, *i.e.*

$$rank \begin{pmatrix} \mathsf{J}_1(\mathbf{z}_k) \\ \mathsf{J}_c(\mathbf{z}_k) \end{pmatrix} = n_d + n_p \quad , \quad rank(\mathsf{J}_c(\mathbf{z}_k)) = n_c \quad , \tag{25}$$

a unique solution $\Delta\mathbf{z}_k$ of the linear problem (15) exists and an unique linear mapping ${\mathsf{J}_k}^+$ can be constructed which satisfies the relation

$$\Delta\mathbf{z}_k = -{\mathsf{J}_k}^+ \mathbf{F}(\mathbf{z}_k) \quad , \quad {\mathsf{J}_k}^+ \mathsf{J}_k {\mathsf{J}_k}^+ = {\mathsf{J}_k}^+ \quad , \quad \mathsf{J}_k^T := \left[\mathsf{J}_1^T(\mathbf{z}_k), \mathsf{J}_c^T(\mathbf{z}_k)\right] \quad . \tag{26}$$

The solution $\Delta\mathbf{z}_k$ of the linear problem or formally the generalized inverse ${\mathsf{J}_k}^+$ [5] of J_k results from the Kuhn-Tucker conditions. But it should be noticed that $\mathbf{z}_k$ is not calculated from (26) because of reasons of numerical efficiency but is based on a decomposition procedure using orthogonal transformations.

By taking into consideration the special structure of the matrices J_i caused by the continuity conditions of the multiple shooting discretization (13) can be reduced by a condensation algorithm described in [Bock (1981, 1987)) to a system of lower dimension

$$\min \left\{ \frac{1}{2} \|\mathsf{A}_1 \mathbf{x}_k + \mathbf{a}_1\|_2^2 | \mathsf{A}_2 \mathbf{x}_k + \mathbf{a}_2 = 0 \quad , \quad \mathsf{A}_3 \mathbf{x}_k + \mathbf{a}_3 \geq 0 \right\} \quad , \tag{27}$$

from which $\mathbf{x}_k$ can be derived at first and at last $\Delta\mathbf{z}_k$. This is achieved by first performing a "backward recursion", the "solution of the condensed problem" and a "forward recursion" [6]. Kilian (1992, [20]) has implemented an active set

strategy following the description in [6] and [33] utilizing the special structure of J_2.

The details of the parameter estimation algorithms which are incorporated in the efficient software package `PARFIT` (a software package of stable and efficient boundary value problem methods for the identification of parameters in systems of nonlinear differential equations) are found in [6]. The damping constant α^k in the k-th iteration is computed with the help of *natural level functions* which locally approximate the distance $\|\mathbf{z}_k - \mathbf{z}^*\|$ of the solution from the Kuhn-Tucker point $\mathbf{z}^*$.

The integrator `METANB` [for the basic discretization see, for instance, Bader and Deuflhard (1981, [3])] embedded in `PARFIT` is also suitable for the integration of stiff differential equation systems. It allows the user to compute simultaneously the sensitivity matrixes G,

$$\mathsf{G}(t; t_0, \mathbf{x}_0, \mathbf{p}) := \frac{\partial}{\partial \mathbf{x}_0}\mathbf{x}(t; t_0, \mathbf{x}_0, \mathbf{p}) \in \mathcal{M}(n_d, n_d) \tag{28}$$

and H,

$$\mathsf{H}(t; t_0, \mathbf{x}_0, \mathbf{p}) := \frac{\partial}{\partial \mathbf{p}}\mathbf{x}(t; t_0, \mathbf{x}_0, \mathbf{p}) \in \mathcal{M}(n_d, n_p) \tag{29}$$

which are the most costly blocks of the Jacobians J_i via the so-called *internal numerical differentiation* as introduced by Bock (1981, [5]). This technique does not require the often cumbersome and error prone formulation of the variational differential equations

$$\mathsf{G}' = \mathbf{f}_{\mathbf{x}}(t, \mathbf{x}, \mathbf{p}) \cdot \mathsf{G} \quad , \quad \mathsf{G}(t_0; t_0, \mathbf{x}_0, \mathbf{p}) = \mathbb{1} \tag{30}$$

and

$$\mathsf{H}' = \mathbf{f}_x(t, \mathbf{x}, \mathbf{p}) \cdot \mathsf{H} + \mathbf{f}_p(t, \mathbf{x}, \mathbf{p}) \quad , \quad \mathsf{H}(t_0; t_0, \mathbf{x}_0, \mathbf{p}) = 0 \tag{31}$$

by the user.

Using the multiple shooting approach described above, differential equation systems with poor stability properties and even chaotic systems can be treated (Kallrath *et al.*, 1993, [18]).

3 Parameter Estimation in DAE Models

Another, even more complex class of problems, are parameter estimation in mechanical multibody systems, *e.g.*, in the planar slider crank mechanisms, a simple model for a cylinder in an engine. These problems lead to boundary problems for higher index differential algebraic systems (Schulz *et al.*, 1998). Singular controls and state constraints in optimal control also lead to this structure. Inherent to such problems are invariants that arise from index reduction but also additional physical invariants such as the total energy in conservative mechanical systems or the Hamiltonian in optimal control problems.

A typical class of DAEs in mechanical multibody systems is given by the equations of motion

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{v} \\ \mathsf{M}(t,\mathbf{x})\dot{\mathbf{v}} &= \mathbf{f}(t,\mathbf{x}) - \boldsymbol{\nabla}_x \mathbf{g}(t,\mathbf{x})\boldsymbol{\lambda} \quad , \\ 0 &= \mathbf{g}(t,\mathbf{x}) \end{aligned} \tag{32}$$

where $\mathbf{x} = \mathbf{x}(t)$ and $\mathbf{v} = \mathbf{v}(t)$ are the coordinates and velocities, M is the mass matrix, $\mathbf{f}$ denotes the applied forces, $\mathbf{g}$ are the holonomic constraints, and $\boldsymbol{\lambda}$ are the generalized constraint forces. Usually, M is symmetric and positive definite. A more general DAE system might have the structure

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(t,\mathbf{x},\mathbf{z};\mathbf{p}) \\ 0 &= \mathbf{g}(t,\mathbf{x},\mathbf{z};\mathbf{p}) \quad , \end{aligned}$$

where $\mathbf{p}$ denotes some parameters and $\mathbf{z} = \mathbf{z}(t)$ is a set of algebraic variables, *i.e.*, the differentials $\dot{\mathbf{z}}$ do not appear; in (32) $\boldsymbol{\lambda}$ is the algebraic variable. In addition we might have initial values $\mathbf{x}_0$ and $\mathbf{z}_0$. Obviously, some care is needed regarding the choice of $\mathbf{z}_0$ because it needs to be consistent with the constraint. In some exceptional cases (in which $\mathsf{Z} := \boldsymbol{\nabla}_z \mathbf{g}$ has full rank and can be inverted analytically) we might insert $\mathbf{z} = \mathbf{z}(t,\mathbf{x};\mathbf{p})$ into the differential equation. DAE systems with a regular matrix Z are referred to as index-1 systems. Index-1-DAEs can be transformed into equivalent ordinary differential equations by differencing the equations w.r.t. t. At first we get the implicit system of differential equations

$$\mathbf{g}_t + \mathsf{X}\dot{\mathbf{x}} + \mathsf{Z}\dot{\mathbf{z}} = \mathbf{0} \quad , \quad \mathsf{X} := \boldsymbol{\nabla}_x \mathbf{g}$$

which, according to the assumption of the regularity of Z, can be written as the explicit system

$$\dot{\mathbf{z}} = \mathsf{Z}^{-1}\left(\mathbf{g}_t + \mathsf{X}\mathbf{f}\right) \quad .$$

Many practical DAEs have index 1, *e.g.*, in some chemical engineering problems, where algebraic equations are introduced to describe, for instance, mass balances or the equation of state. However, multibody systems such as (32) have higher indices; (32) is of index 3. The reason is, that the multiplier variables, *i.e.*, the algebraic variables, do not occur in the algebraic constraints and it is therefore not possible to extract them directly without further differentiation. If Z does not have full rank the equations are differentiated successively, until the algebraic variables can be eliminated. The smallest number of differentiations required to transform the original DAE system to an ODE system is called the *index* of the DAE. The approach developed and described by Schulz *et al.* (1998, [34]) is capable to handle least squares problems without special assumption to the index.

An essential problem for the design, optimization and control of chemical systems is the estimation of parameters from time-series. These problems lead to nonlinear DAEs. The parameters estimation problem leads to a non-convex optimization problem for which several local minima exist. Esposito & Floudas

(2000, [13]) developed two global branch&bound and convex-underestimator based optimization approaches to solve this problem. In the first approach, the dynamical system is converted into an algebraic system using orthogonal collocation on finite elements. In the second approach, state profiles are computed by integration. In Esposito & Floudas (2000, [12]) a similar approach is used to solve optimal control problems.

4 Parameter Estimation in PDE Models

A very complex class of least squares problems are data fitting problems in partial differential equations based models. These include eigenvalue problems, as well as initial and boundary value problems and cover problems in atomic physics, elasticity, electromagnetic fields, fluid flow or heat transfer. Some recent problems are, for instance, in models describing the water balance and solid transport used to analyze the distributions of nutrients and pesticides (Arning, 1994, [1]), in the determination of diffusive constants in water absorption processes in hygroscopic liquids discussed in Kallrath (1999, [15]), or in multispecies reactive flows through porous media (Zieße *et al.*, 1996, [38]). Such nonlinear multispecies transport models can be used to describe the interaction between oxygen, nitrate, organic carbon and bacteria in aquifers. They may include convective transport and diffusion/dispersion processes for the mobile parts (that is the mobile pore water) of the species. The immobile biophase represents the part where reactions caused by microbial activity take place and which is coupled to transport through mobile pore water. The microorganisms are assumed to be immobile. The model leads to partial differential algebraic equations

$$\begin{aligned} \mathsf{M}\partial_t \mathbf{u} - \nabla(D\nabla\mathbf{u}) + q\nabla\mathbf{u} &= \mathbf{f}_1(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \\ \partial_t \mathbf{v} &= \mathbf{f}_2(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \\ 0 &= \mathbf{g}(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \end{aligned} \tag{33}$$

where D and q denote the hydraulic parameters of the model, $\mathbf{p}$ denotes a set of reaction parameters, $\mathbf{u}$ and $\mathbf{v}$ refer to the mobile and immobile species, and $\mathbf{z}$ is related to source and sink terms.

4.1 Methodology

To solve least squares problems based on PDE models requires sophisticated numerical techniques but also great attention with respect to the quality of data and identifiability of the parameters. To solve such problems we might use the following approaches:

1. Unstructured approach: The PDE model is, for fixed parameters $\mathbf{p}$, integrated by any appropriate method yielding estimations of the observations. The parameters are adjusted by a derivative-free optimization procedure, *e.g.*, by the Simplex method by Nelder & Mead (1965, [23]).

This approach is relatively easy to implement, it solves a sequence of direct problems, and is comparable to what in Section 2 has been called the IVP approach. Arning (1994) uses such an approach.

2. Structured approach (for initial value PDE problems): Within the PDE model spatial coordinates and time are discretized separately. Especially for models with only one spatial coordinate, it is advantageous to apply finite difference or finite element discretizations to the spatial coordinate. The PDE system is transformed into a system of (usually stiff) ordinary differential equations. This approach is known as the *method of lines* [see, for example, Schiesser (1991, [30])]. It reduces parameter estimation problems subject to time-dependent partial differential equations to parameter identification problems in systems of ordinary differential equations to be integrated w.r.t. time. Now it is possible to distinguish again between the IVP and BVP approach. Schittkowski (1997, [32]) in his software package `EASY-FIT` applies the method of lines to PDEs with one spatial coordinate and uses several explicit and implicit integration methods to solve the ODE system. The integration results are used by an SQP optimization routine or a Gauß-Newton method to estimate the parameters. Zieße *et al.* (1996) and Dieses *et al.* (1999, [11]), instead, couple the method of lines (in one and two spatial coordinates) with Bock's (1987, [6]) BVP approach, discretize time, for instance, by multiple shooting and use an extended version of `PARFIT`.

The method of lines has become one of the standard approaches for solving time-dependent PDEs with only one spatial coordinate. It is based on a partial discretization, which means that only the spatial derivative is discretized but not the time derivative. This leads to a system of N coupled ordinary differential equation, where N is the number of discretization points. Let us demonstrate the method by applying it to the diffusion equation

$$\frac{\partial}{\partial t}c(t,z) = D\frac{\partial^2}{\partial z^2}c(t,z) \quad , \quad \begin{array}{l} 0 \le t < \infty \\ 0 \le z \le L \end{array} \tag{34}$$

with constant diffusion coefficient D. We discretize the spatial coordinate z according to

$$z_i = i\Delta z \quad , \quad \Delta z := \frac{L}{N} \quad , \quad c_i = c_i(t) = c(t,z_i) \quad , \quad i = 0,\ldots,N \quad . \tag{35}$$

If we choose a finite difference approximation we get

$$\frac{\partial^2}{\partial z^2}c(t,z) \approx \frac{c(t,z-\Delta z) - 2c(t,z) + c(t,z+\Delta z)}{(\Delta z)^2} = \frac{c_{i-1} - 2c_i + c_{i+1}}{(\Delta z)^2} \quad , \tag{36}$$

which replaces the diffusion equation (34) by N ordinary differential equations

$$\dot{c}_i(t) = \frac{c_{i-1} - 2c_i + c_{i+1}}{(\Delta z)^2} \quad . \tag{37}$$

A detailed example of this method is discussed in Kallrath (1999, [15]). The water transport and absorption processes within a hygroscopic liquid are described by a model containing the diffusion equation (34) describing the water transport within the hygroscopic liquid, a mixed Dirichlet-Neumann condition representing a flux balance equation at the surface of the liquid, and an additional integral relation describing the total amount of water in the liquid. The model included three parameters to be estimated.

The available measurement data provide the total time dependent concentration $C(t)$ of water in the liquid. A further complication was that the mathematical solution of the diffusion equation is the water concentration $c(t, z)$ in the hygroscopic liquid and it is a function of time *and* location. Therefore, in order to compare the mathematical solution with the observed data one had to integrate $c(t, z)$ over the space coordinate z, *i.e.*, the depth of the fluid.

5 Least Squares Problems with Massive Data Sets

We motivate the necessity to analyze massive data sets by an example taken from astrophysics (Kallrath & Wilson, 2007, [19]). We outline the method for a huge set of millions of observed data curves in which time is the independent parameter and for each of the N, $N \simeq 10^6$, curves there is a different underlying parameter set we want to estimate by a least squares method. Note that we assume that there is a model in the sense of (1) or (5) available involving an adjustable parameter vector $\mathbf{p}$. We are further assume that we are dealing with nonlinear least squares problems which are not easy to solve. The difficulties could arise from the dependence on initial parameters, non-smoothness of the model, the number of model evalutions, or the CPU time required for one model evaluation. For each available curve we can, of course, solve this least squares problem by the techniques mentioned or discussed earlier in this article. However, the CPU time required to solve this least squares problem for several million curves is prohibitive. The archive approach described in this section is appropriate for this situation.

Examples of massive data sets subject to least squares analyses are surveys in astrophysics where millions of stars are observed over a range of time. About 50% of them are binary stars or multiple systems. The observed data could be flux of photons (just called *light* in the discipline of binary star researchers) in a certain wavelength region or radial velocity as a function of time. Thus we have to analyze millions of light and radial velocity curves. There are well validated models and methods (*cf.*, Kallrath & Milone, (1999, [17]) to compute such curves on well defined physical and geometrical parameters of the binary systems, e.g., the mass ratio, the ratio of their radii, their temperatures, inclination, semi-major axis and eccentricity to mention a few. Thus one is facing the problem how to analyze the surveys and to derive the stellar parameters $\mathcal{P}$ relevant to astrophysicists. In this eclipsing binary star example it suffices to consider the

range $[0, P]$ for the independent parameter time because the observed curves are periodic with respect to the period P. The period could be determined a priori from a frequency analysis of the observed curve. Under certain assumptions, in eclipsing binary star analyses, time can be replaced by phase.

The critical issues are speed and stability. Speed is obviously necessary to analyze large number of data, light and radial velocity curves in the example. Stability is required to automatize the procedure. Automatization enables the user to analyze large sets of eclipsing binary data produced by surveys. Stability and automatization need to overcome the problem of initial parameters usually experienced in nonlinear least squares. There is a price to be paid in terms of accuracy. But nevertheless, such an approach will produce good approximate results and may indicate interesting eclipsing binary stars for detailed follow-up analysis.

The method we propose to solve least squares problems with massive data sets is a matching approach: match one or several curves to a large test sets of pre-computed archive curves for an appropriate set of combinations of $|\mathcal{P}|$ parameters.

5.1 The Matching Approach

Let for a given binary system ℓ^o_{ic} be any observed light value for observable c, $c = 1 \ldots C$, at phase θ_i, $i = 1, \ldots, I$. Correspondingly, ℓ^c_{ick} denotes the computed light value at the same phase θ_i for the *archive* light curve k, $k = 1 \ldots K$. Note that K easily might be a large number such as 10^{10}. Each archive light curve k is computed by a certain parameter combination.

The idea of the matching approach is to pick that light curve from the archive which matches the observed curve of binary j best. The best fit solution is obtained by linear regression. The matching approach returns, for each j, the number of the archive light curve which fits best, a scaling parameter, a, and a shift parameter, b, (which might be interpreted as a constant third light) by solving the following nested minimization problem for all j, $j = 1, \ldots, N$:

$$\min_k \left\{ \min_{a_{kc}, b_{kc}} \sum_{i=1}^{I} w_i \left[\ell^o_{ic} - (a_{kc} \ell^c_{ick} + b_{kc}) \right]^2 \right\}$$

Note that the inner minimization problem requires just to solve a linear regression problem. Thus, for each k, there exists an analytic solution for the unknown parameters a_{kc} and b_{kc}. Further note that the ℓ^c_{ick} values might be obtained by interpolation. The archive light curves are generated in such a way that they have a good covering in the eclipses while a few points will do in those parts of the light curves which show only small variation with phase. Thus, there might be a non-equidistant distribution of phase grid points. A cubic interpolation will probably suffice.

Thus, the matching approach requires us to provide the following components:

1. solving linear regression problems determining a and b for all archive curves and all observed curves (the sequence of the loops is important),

2. generating the archive curves,

3. cubic interpolation in the independent time-like quantity and interpolation after the best matching solution has been found.

In the sequel we briefly comment on the last two components.

5.1.1 Generating and Storing the Archive Curves

As the number of archive curves can easily reach 10^{10} one should carefully think about storing them. That requires also appropriate looping over the parameters $p = 1, \ldots, |\mathcal{P}|$. For the eclipsing binary example the details are given in Kallrath & Wilson (2007). Among the efficiency issues is the usage of non-equidistant parameter grids exploiting the sensitivity of the parameters on the model function ℓ_{ic}^{c}.

One might think to store the archive light curves in a type of data base. However, data base techniques become very poor when talking about 10^{10} curves. Therefore, it is probably easier to use a flat storage scheme. In the simplest case, for each k we store the physical and geometric parameters, then those parameters describing observable c, and then the values of the observable. If we use the same number of phase values for each observable and each k, we have the same amount of data to be stored.

5.1.2 Exploiting Interpolation Techniques

Within the matching approach interpolation can be used at two places. The first occurrence is in the regression phase. The test curves in the archive are computed for a finite grid of the independent parameter time (phase in this example). The observed curves might be observed at time values not contained in the archive. We can interpolate from the archive values by linear or cubic interpolation to the observed time values. However, it may well pay out to have some careful thoughts on the generation of the time grid points.

The second occurrence is when it comes to determining the best fit. The linear regression returns that parameter set which matches the observed one best. Alternatively, we could exploit several archive points to obtain a better fit to the observed curve. Interpolation in an appropriately defined neighborhoods of the best archive solution can improve the fit of the observed curve.

5.1.3 Numerical Efficiency

The efficiency of a least squares method could be measured by the number of function or model evaluation per unknown parameter. If we assume that for each model parameter p we generate n_p archive curves in the archive, the archive

contains test curves $N_c = \prod_{p=1}^{|\mathcal{P}|} n_p$ and thus requires N_c model evaluation; n_p is the number of archive grid points of parameter p.

6 Conclusions

This contribution outlines how to solve ODE and PDE based least squares problems. Academic and commercial least squares solvers as well as software packages are available. Massive data sets and observations arise in data mining problems, medicine, the stock market, and surveys in astrophysics. The approach described in Sect. 5.1 has been proven efficient for surveys in astrophysics. It can also support the generation of impersonal good initial parameter estimations for further analysis. The archive approach is also suitable for parameter fitting problems with non-smooth models. Another advantage is that on the archve grid it provides the global least squares minimum.

Acknowledgements: Thanks is directed to Steffen Rebennack (University of Florida, Gainesville, FL) for a careful reading of the manuscript, and Johannes P. Schlöder (IWR, Universität Heidelberg, Germany) and Gerhard Krennrich (BASF Aktiengesellschaft, Ludwigshafen) for discussions on the subject of parameter estimation.

References

[1] M. Arning. Lösung des Inversproblems von partiellen Differentialgleichungen beim Wassertransport im Boden. Dissertation, TU Carolo-Wilhelmina zu Braunschweig, Braunschweig, Germany, 1994.

[2] E. Baake and J. P. Schlöder. Modelling the Fast Fluorescence Rise of Photosynthesis. *Bulletin of Mathematical Biology*, 54:999–1021, 1992.

[3] G. Bader and P. Deuflhard. A Semi-Implicit Mid-Point Rule for Stiff Systems of Ordinary Differential Equations. Preprint 114, Universität Heidelberg, SFB 123, Institut für Angewandte Mathematik, 69120 Heidelberg, 1981.

[4] A. Björck. Least Squares Problems. In C. A. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*, pages 160–170. Kluwer Academic Publishers, Dordrecht, Holland, 2001.

[5] H. G. Bock. Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics. In K. H. Ebert, P. Deuflhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, Series in Chemical Physics, pages 102–125. Springer, Heidelberg, 1981.

[6] H. G. Bock. Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. Preprint 142, Universität

Heidelberg, SFB 123, Institut für Angewandte Mathematik, 69120 Heidelberg, 1987.

[7] S. Brand. *Statistical and Computational Methods in Data Analysis*. North Holland, Amsterdam - New York - Oxford, 2nd edition, 1976.

[8] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimisation and Nonlinear Equations*. Prentice Hall, Englewood Cliffs, New Jersey, 1983.

[9] P. Deuflhard and V. Apostolescu. An Underrelaxed Gauss-Newton Method for Equality Constrained Nonlinear Least Squares Problems. In J. Stoer, editor, *Proc. 8th IFIP Conf Würzburg Symposium on the Theory of Computing*, number 23 in Springer Lecture Notes Control Inf. Sci., Heidelberg-Berlin-New York, 1977. Springer.

[10] P. Deuflhard and V. Apostolescu. A Study of the Gauss-Newton Method for the Solution of Nonlinear Least Squares Problems. In J. Frehse, D. Pallaschke, and U. Trottenberg, editors, *Special Topics of Applied Mathematics*, pages 129–150. North-Holland, Amsterdam, 1980.

[11] A. E. Dieses, J. P. Schlöder, H. G. Bock, and O. Richter. Parameter Estimation for Nonlinear Transport and Degradation Processes of Xenobiotica in Soil. In J. Neumann, editor, *Proceedings of the 2nd International Workshop on Scientific Computing in Chemical Engineering*, Hamburg, Germany, 1999. Technical University Hamburg-Harburg (TUUH).

[12] W. R. Esposito and C. A. Floudas. Deterministic Global Optimization in Nonlinear Optimal Control Problems. *Journal of Global Optimization*, 17:97–126, 2000.

[13] W. R. Esposito and C. A. Floudas. Global Optimization for the Parameter Estimation of Differential-Algebraic Systems. *Industrial and Engineering Chemistry Research*, 39(5):1291–1310, 2000.

[14] C. F. Gauß. *Theoria Motus Corporum Coelestium in Sectionibus Conicus Solem Ambientium*. F. Perthes and J. H. Besser, Hamburg, 1809.

[15] J. Kallrath. Least Squares Methods for Models Including Ordinary and Partial Differential Equations. In R. Dvorak, H. F. Haupt, and K. Wodnar, editors, *Modern Astrometry and Astrodynamics honouring Eichhorn Heinrich*, pages 61–75, Vienna, Austria, 1999. Austrian Academy of Sciences.

[16] J. Kallrath, V. Altstädt, J. P. Schlöder, and H. G. Bock. Analysis of Crack Fatigue Growth Behaviour in Polymers and their Composites based on Ordinary Differential Equations Parameter Estimation. *Polymer Testing,*, in print, 1998.

[17] J. Kallrath and E. F. Milone. *Eclipsing Binary Stars: Modeling and Analysis*. Springer, New York, 1999.

[18] J. Kallrath, J. Schlöder, and H. G. Bock. Parameter Fitting in Chaotic Dynamical Systems. *CMDA*, 56:353–371, 1993.

[19] J. Kallrath and R. E. Wilson. Eclipsing Binary Analysis via Light Curve Archives. *ApJ*, in preparation, 2008.

[20] C. Kilian. Numerische Behandlung von Ungleichheitsrestriktionen für Parameterschätzprobleme in Systemen gewöhnlicher Differentialgleichungen. Diploma thesis, Fachhochschule Darmstadt, Darmstadt, 1992.

[21] K. Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.

[22] D. W. Marquardt. An Algorithm for Least Squares Estimation of Nonlinear Parameters. *SIAM J. Applied Math.*, 11:431–441, 1963.

[23] J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7:308–313, 1965.

[24] B. Ostle. *Statistics in Research*. Iowa State University Press, Ames, IA, 1963.

[25] G. Polya. *Vom Lernen und Lösen mathematischer Aufgaben. Einsicht und Entdeckung. Lernen und Lehren*. Birkhäuser Verlag, Basel, 1979.

[26] K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 10th edition, 1980.

[27] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes - The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[28] O. Richter, P. Nörtersheuser, and W. Pestemer. Non-linear parameter Estimation in Pesticide Degradation. *The Science of the Total Environment*, 123/124:435–450, 1992.

[29] O. Richter and D. Söndgerath. *Parameter Estimation in Ecology*. VCH-Verlag, Weinheim, Germany, 1990.

[30] W. E. Schiesser. *The Numerical Methods of Lines*. Academic Press, San Diego, California, 1991.

[31] K. Schittkowski. Parameter Estimation in Systems of Nonlinear Equations. *Numerische Mathematik*, 68:129–142, 1994.

[32] K. Schittkowski. Parameter Estimation in One-Dimensional Time-Dependent Partial Differential Equations. *Optimization Methods and Software*, 7:165–210, 1997.

[33] J. P. Schlöder. Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung. Preprint 187, Bonner Mathematische Schriften, Institut für Angewandte Mathematik, 5300 Bonn, 1988.

[34] V. H. Schulz, H. G. Bock, and M. C. Steinbach. Exploiting Invariants in the Numerical Solution of Multipoint Boundary Value Problems for DAE. *SIAM Journal of Scientific Computing*, 19:440–467, 1998.

[35] C. Xu. Nonlinear Least Squares. In C. A. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*, pages 75–80. Kluwer Academic Publishers, Dordrecht, Holland, 2001.

[36] C. Xu. Nonlinear Least Squares: Newton-type Methods. In C. A. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*, pages 67–69. Kluwer Academic Publishers, Dordrecht, Holland, 2001.

[37] C. Xu. Nonlinear Least Squares: Trust Region Methods. In C. A. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*, pages 80–86. Kluwer Academic Publishers, Dordrecht, Holland, 2001.

[38] M. W. Zieße, H. G. Bock, J. V. Gallitzendörfer, and J. P. Schlöder. Parameter Estimation in Multispecies Transport Reaction Systems Using Parallel Algorithms. In J. Gottlieb and P. DuChateau, editors, *Parameter Identification and Inverse Problems in Hydrology, Geology and Ecology*, pages 273–282, Dordrecht, Holland, 1996. Kluwer Academic Publishers.