Least Squares Methods for Models Including Ordinary and Partial Differential Equations

Josef Kallrath
BASF-AG, ZX/ZC, C13, 67056 Ludwigshafen
e-mail: kallrath@zx.basf-ag.de

September 16, 2002

Abstract

This contribution gives a brief overview on nonlinear constrained least squares methods. The focus is on data fitting problems in which the models include ordinary or partial differential equations.

1 Introduction

Legendre (1805) first published the method of least squares, applying it to data from the 1795 French meridian arc survey as an example. In 1809, Gauß in "Theoria motus corporum coelestium" (Gauß, 1809) derived the justification for the method in terms of the normal error law, showed how to obtain the errors of the estimated parameters, and also how nonlinear problems could be linearized, so that the method could be applied to the problem of nonlinear parameter estimation. He also claimed he had been using the method since 1795, ten years before Legendre's work was published. It appears that Gauß had indeed been using the method as he claimed, but had not appreciated its wider importance until Legendre's publication. For a more detailed discussion of this priority conflict we refer the reader to Stigler (1986) or Schneider (1988).

Since the time of Gauß, numerical methods for solving several types of least squares problems have been developed and improved, and there is still much active research in that area. For a review of the methods of least squares as known and used in astronomy, especially astrometry, we refer to Eichhorn (1993).

The unconstrained least squares problem can be expressed by [1]

$$\min_{\mathbf{p}} l_2(\mathbf{p}) \quad , \quad l_2(\mathbf{p}) := \|\mathbf{r}_1\left[\mathbf{x}(t_1),\ldots,\mathbf{x}(t_k),\mathbf{p}\right]\|_2^2 = \sum_{k=1}^{N}\left[r_{1k}(\mathbf{p})\right]^2 \quad , \quad \mathbf{r}_1 \in \mathbb{R}^N \quad . \tag{1.1}$$

This structure may arise either from a nonlinear over-determined system of equations

$$r_{1k}(\mathbf{p}) = 0 \quad , \quad k = 1, ..., N \quad , \quad N > n \quad , \tag{1.2}$$

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 $\mathbf{x}$:

$$r_{1k} := r_{1k}(\mathbf{p}) = Y_k - F_k(\mathbf{p}) = \sqrt{w_k}\left[\tilde{Y}_k - \tilde{F}(t_k, \mathbf{p})\right] \quad . \tag{1.3}$$

The weights w_k are related to the variances σ_k by

$$w_k := \beta/\sigma_k^2 \quad . \tag{1.4}$$

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

$$\mathbf{r_1} := \mathbf{Y} - \mathbf{F}(\mathbf{p}) = \left[r_{11}(\mathbf{p}), \ldots, r_{1N}(\mathbf{p})\right]^{\mathrm{T}} \quad , \quad \mathbf{F}(\mathbf{p}), \mathbf{Y} \in \mathbb{R}^N \quad . \tag{1.5}$$

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.

A popular method to solve unconstrained least squares problems is the Levenberg-Marquardt algorithm proposed independently by Levenberg (1944) and Marquardt (1963) 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)]. Damped (step-size cutting) Gauß-Newton algorithms combined

[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 (1.4), provides a maximum likelihood estimator (Brandt, 1976, Chp.7) for the unknown parameter vector $\mathbf{p}$. This objective function dates back to Gauß (1809) and in the mathematical literature the problem is synonymously called least squares or ℓ_2 approximation problem.

with orthogonalization methods control the damping by natural level functions [Deuflhard and Apostolescu (1977, 1980), Bock (1987)] seem to be superior to Levenberg-Marquardt type schemes and can be more easily extended to nonlinear constrained least squares problems.

A common basic feature and limitation of least squares methods used in astronomy, but seldom explicitly noted, is that they require some explicit model to be fitted to the data. However, not all models are explicitly available. 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 (Schittkowski, 1994), 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

$$\mathbf{F}_2(t;\mathbf{p},\mathbf{z}) = \mathbf{0} \quad , \quad \mathbf{F}_2(\mathbf{p}) \in \mathbb{R}^{n_2} \quad . \tag{1.6}$$

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 (Bock, 1981), nonlinear models in material sciences (Kallrath *et al.*, 1998) and biology (Baake and Schlöder, 1992), and nonlinear systems describing ecosystems [Richter and Söndgerath (1990), Richter *et al.* (1992)] 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.

2 Parameter Estimation in ODE Models

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

$$\mathbf{x}'(t) = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(t,\mathbf{x},\mathbf{p}) \quad , \quad \mathbf{x} \in \mathbb{R}^{n_d} \quad , \quad \mathbf{p} \in \mathbb{R}^{n_p} \tag{2.1}$$

with a right hand side depending on an unknown parameter vector $\mathbf{p}$. Additional requirements on the solution of the ODE (2.1) 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

$$\mathbf{r}_2\left[\mathbf{x}(t_1),\ldots,\mathbf{x}(t_k),\mathbf{p}\right] = 0 \quad \text{or} \quad \mathbf{r}_3\left[\mathbf{x}(t_1),\ldots,\mathbf{x}(t_k),\mathbf{p}\right] \geq 0 \quad . \tag{2.2}$$

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

$$l_2(\mathbf{x}, \mathbf{p}) := \|\mathbf{r}_1 [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_k), \mathbf{p}]\|_2^2 \quad . \tag{2.3}$$

In a special case of (2.3) the components ℓ of the vector $\mathbf{r}_1 \in \mathbb{R}^L$ are "equations of condition" and have the form

$$r_{1\ell} = \sigma_{ij}^{-1} [\eta_{ij} - g_i(\mathbf{x}(t_j), \mathbf{p})] \quad , \quad \ell = 1, \ldots, L := \sum_{i=1}^{N_j} J_i \quad . \tag{2.4}$$

This case leads us to the least squares function

$$l_2(\mathbf{x}, \mathbf{p}) := \sum_{j=1}^{N^D} \sum_{i=1}^{N_j} \sigma_{ij}^{-2} [\eta_{ij} - g_i(\mathbf{x}(t_j), \mathbf{p})]^2 \quad . \tag{2.5}$$

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

$$\eta_{ij} = g_i(\mathbf{x}(t_j), \mathbf{p}) + \varepsilon_{ij} \quad . \tag{2.6}$$

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) (2.1) and iterate the parameters and initial values in order to improve the fit. Characteristic features and disadvantages are discussed in, *e.g.,* (Bock, 1987) or Kallrath *et al.* (1993). 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)). 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

$$\mathbf{x}' = f(t, \mathbf{x}, \mathbf{p}; sign(\mathbf{Q}(t, \mathbf{x}, \mathbf{p}))) \quad , \tag{2.7}$$

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

$$\mathbf{Q}_i(t, \mathbf{x}, \mathbf{p}) = 0 \quad . \tag{2.8}$$

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 $\mathcal{T}_m$

$$\mathcal{T}_m : \tau_1 < \tau_2 < ... < \tau_m \quad , \quad \Delta\tau_j := \tau_{j+1} - \tau_j \quad , \quad 1 \le j \le m-1, \tag{2.9}$$

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

$$\mathbf{x}'(t) = \mathbf{f}(t, \mathbf{x}, \mathbf{p}) \quad , \quad \mathbf{x}(t = \tau_j) = \mathbf{s}_j \in \mathbb{R}^{n_d} \tag{2.10}$$

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 $\mathbf{z}$

$$\mathbf{z}^T := (\mathbf{s}_1^T, ..., \mathbf{s}_m^T, \mathbf{p}^T) \quad . \tag{2.11}$$

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 (2.1) equivalently by these $m-1$ IVPs matching conditions

$$\mathbf{h}_j(\mathbf{s}_j, \mathbf{s}_{j+1}, p) := \mathbf{x}(\tau_{j+1}; \mathbf{s}_j, \mathbf{p}) - \mathbf{s}_{j+1} = 0 \quad , \quad \mathbf{h}_j : \mathbb{R}^{2n_d+n_p} \to \mathbb{R}^{n_d} \tag{2.12}$$

are added to the problem. (2.12) ensures the continuity of the final trajectory $\mathbf{x}(t)$.

Replacing $\mathbf{x}(t_i)$ and $\mathbf{p}$ in (2.5) by $\mathbf{z}$ the least squares problem is reformulated as a nonlinear constrained optimization problem with the structure

$$\min_{\mathbf{z}} \left\{ \frac{1}{2} \|\mathbf{F}_1(\mathbf{z})\|_2^2 \,|\mathbf{F}_2(\mathbf{z}) = 0 \in \mathbb{R}^{n_2} \quad , \quad \mathbf{F}_3(\mathbf{z}) \geq 0 \in \mathbb{R}^{n_3} \right\} \quad , \tag{2.13}$$

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 (Bock, 1981). If $\mathsf{J}_1(\mathbf{z}_k) := \partial_z \mathbf{F}_1(\mathbf{z}_k)$, $\mathsf{J}_2(z_k) := \partial_z \mathbf{F}_2(\mathbf{z}_k)$ 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

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha_k \mathbf{\Delta z}_k \tag{2.14}$$

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

$$\min_{\mathbf{z}} \left\{ \frac{1}{2} \left\| \mathsf{J}_1(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_1(\mathbf{z}_k) \right\|_2^2 \,\middle|\, \begin{array}{l} \mathsf{J}_2(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_2(\mathbf{z}_k) = 0 \\ \mathsf{J}_3(\mathbf{z}_k)\mathbf{\Delta z}_k + \mathbf{F}_3(\mathbf{z}_k) \geq 0 \end{array} \right\} \quad . \tag{2.15}$$

Global convergence can be achieved if the damping strategy is properly chosen (Bock, 1987).

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

$$\mathcal{I}(\mathbf{z}_k) := \{ i | F_{3i}(\mathbf{z}_k) = 0 \quad , \quad i = 1, ..., n_3 \} \quad . \tag{2.16}$$

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 (2.16) we define

$$\mathbf{F}_c := \begin{pmatrix} \mathbf{F}_2 \\ \hat{\mathbf{F}}_3 \end{pmatrix} \quad , \quad \mathsf{J}_c := \begin{pmatrix} \mathsf{J}_2 \\ \hat{\mathsf{J}}_3 \end{pmatrix} \quad . \tag{2.17}$$

In order to derive the necessary conditions that have to be fulfilled by the solution of the problem (2.13) the Lagrangian

$$L(\mathbf{z}, \lambda, \mu) := \frac{1}{2} \|\mathbf{F}_1(\mathbf{z})\|_2^2 - \lambda^T \mathbf{F}_2(\mathbf{z}) - \mu^T \mathbf{F}_3(\mathbf{z}) \tag{2.18}$$

and the reduced Lagrangian

$$\hat{L}(\mathbf{z}, \lambda_c) := \frac{1}{2}\|\mathbf{F}_1(\mathbf{z})\|_2^2 - \lambda_c^T \mathbf{F}_c(\mathbf{z}) \quad , \quad \lambda_c := \binom{\lambda}{\mu_c} \tag{2.19}$$

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

$$\mathbf{F}_2(\mathbf{z}^*) = 0 \quad , \quad \mathbf{F}_3(\mathbf{z}^*) \geq 0 \tag{2.20}$$

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

$$\frac{\partial L}{\partial z}(\mathbf{z}^*, \lambda^*, \mu^*) = \mathbf{F}_1^T(\mathbf{z}^*) \cdot \mathsf{J}(\mathbf{z}^*) - (\lambda^*)^T \mathsf{J}_2(\mathbf{z}^*) - (\mu^*)^T \mathsf{J}_3(\mathbf{z}^*) = 0 \tag{2.21}$$

and

$$\mu^* \geq 0 \quad , \quad i \notin \mathcal{I}(\mathbf{z}^*) \Rightarrow {\mu_i}^* = 0 \quad . \tag{2.22}$$

If $(\mathbf{z}^*, \lambda^*, \mu^*)$ fulfills the conditions (2.20), (2.21) and (2.22), it is called a Kuhn-Tucker-point and $\mathbf{z}^*$ a stationary point. The necessary condition of second order means that for all directions

$$\mathbf{s} \in T(\mathbf{x}^*) := \left\{ \mathbf{s} \neq 0 \;\middle|\; \begin{array}{l} \mathsf{J}_2(\mathbf{z}^*)\mathbf{s} = 0 \\ \mathsf{J}_3(\mathbf{z}^*)\mathbf{s} \geq 0 \end{array} \quad , \quad \mu_i \mathsf{J}_{3i}(\mathbf{z}^*)\mathbf{s} = 0 \right\} \tag{2.23}$$

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

$$\mathbf{s}^T \mathsf{G}(\mathbf{z}^*, \lambda^*, \mu^*)\mathbf{s} \geq 0 \quad , \quad \mathsf{G}(\mathbf{z}^*, \lambda^*, \mu^*) := \frac{\partial^2}{\partial \mathbf{z}^2} L(\mathbf{z}^*, \lambda^*, \mu^*) \quad . \tag{2.24}$$

As $\mu_i = 0$ for $i \notin \mathcal{I}(\mathbf{z}^*)$ it is sufficient to postulate the stationary condition for the reduced Lagrangian (2.19). For the linear problem (2.15) follows: $(\mathbf{z}^*, \lambda^*, \mu^*)$ is a Kuhn-Tucker-point of the nonlinear problem (2.13) if and only, if $(0, \lambda^*, \mu^*)$ is a Kuhn-Tucker-point of the linear problem. The necessary conditions for the existence of a local minimum of problem (2.13) are:

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

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 (Bock, 1987) can be formulated. If the sufficient conditions are fulfilled it can be shown for the neighborhood of a Kuhn-Tucker-point $(\mathbf{z}^*, \lambda^*, \mu^*)$ of the nonlinear problem

(2.13) 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.*

$$rank \begin{pmatrix} \mathsf{J}_1(\mathbf{z}_k) \\ \mathsf{J}_c(\mathbf{z}_k) \end{pmatrix} = n_d + n_p \quad , \quad rank(\mathsf{J}_c(\mathbf{z}_k)) = n_c \quad , \tag{2.25}$$

a unique solution $\Delta\mathbf{z}_k$ of the linear problem (2.15) exists and an unique linear mapping ${\mathsf{J}_k}^+$ can be constructed which satisfies the relation

$$\Delta\mathbf{z}_k = -{\mathsf{J}_k}^+\mathbf{F}(\mathbf{z}_k) \quad , \quad {\mathsf{J}_k}^+\mathsf{J}_k{\mathsf{J}_k}^+ = {\mathsf{J}_k}^+ \quad , \quad \mathsf{J}_k^T := \left[\mathsf{J}_1^T(\mathbf{z}_k), \mathsf{J}_c^T(\mathbf{z}_k)\right] \quad . \tag{2.26}$$

The solution $\Delta\mathbf{z}_k$ of the linear problem or formally the generalized inverse ${\mathsf{J}_k}^+$ (Bock, 1981) of J_k results from the Kuhn-Tucker conditions. But it should be noticed that $\mathbf{z}_k$ is not calculated from (2.26) 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 (2.13) can be reduced by a condensation algorithm described in [Bock (1981, 1987)) to a system of lower dimension

$$\min\left\{\frac{1}{2}\|\mathsf{A}_1\mathbf{x}_k + \mathbf{a}_1\|_2^2 | \mathsf{A}_2\mathbf{x}_k + \mathbf{a}_2 = 0 \quad , \quad \mathsf{A}_3\mathbf{x}_k + \mathbf{a}_3 \geq 0\right\} \quad , \tag{2.27}$$

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" (Bock, 1987). Kilian (1992) has implemented an active set strategy following the description in (Bock, 1987) and (Schlöder, 1988) 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 (Bock, 1987). 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)] 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,

$$\mathsf{G}(t; t_0, \mathbf{x}_0, \mathbf{p}) := \frac{\partial}{\partial\mathbf{x}_0}\mathbf{x}(t; t_0, \mathbf{x}_0, \mathbf{p}) \in \mathcal{M}(n_d, n_d) \tag{2.28}$$

and H,

$$\mathsf{H}(t;t_0,\mathbf{x}_0,\mathbf{p}) := \frac{\partial}{\partial \mathbf{p}}\mathbf{x}(t;t_0,\mathbf{x}_0,\mathbf{p}) \in \mathcal{M}(n_d,n_p) \tag{2.29}$$

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

$$\mathsf{G}' = \mathbf{f}_{\mathbf{X}}(t,\mathbf{x},\mathbf{p}) \cdot \mathsf{G} \quad , \quad \mathsf{G}(t_0;t_0,\mathbf{x}_0,\mathbf{p}) = \mathbb{1} \tag{2.30}$$

and

$$\mathsf{H}' = \mathbf{f}_x(t,\mathbf{x},\mathbf{p}) \cdot \mathsf{H} + \mathbf{f}_p(t,\mathbf{x},\mathbf{p}) \quad , \quad \mathsf{H}(t_0;t_0,\mathbf{x}_0,\mathbf{p}) = 0 \tag{2.31}$$

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).

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

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{v} \\ \mathsf{M}(t,\mathbf{x})\dot{\mathbf{v}} &= \mathbf{f}(t,\mathbf{x}) - \nabla_x \mathbf{g}(t,\mathbf{x})\lambda \quad , \\ 0 &= \mathbf{g}(t,\mathbf{x}) \end{aligned} \tag{3.32}$$

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 λ are the generalized constraint forces. Usually, M is symmetric and positive definite. A more general DAE system might have the structure

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(t,\mathbf{x},\mathbf{z};\mathbf{p}) \\ 0 &= \mathbf{g}(t,\mathbf{x},\mathbf{z};\mathbf{p}) \quad , \end{aligned} \tag{3.33}$$

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 (3.32) λ 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 $\mathsf{Z} := \nabla_z \mathbf{g}$ 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

$$\mathbf{g}_t + \mathsf{X}\dot{\mathbf{x}} + \mathsf{Z}\dot{\mathbf{z}} = \mathbf{0} \quad , \quad \mathsf{X} := \nabla_x \mathbf{g} \tag{3.34}$$

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

$$\dot{\mathbf{z}} = \mathsf{Z}^{-1} \left(\mathbf{g}_t + \mathsf{X}\mathbf{f} \right) \quad . \tag{3.35}$$

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 (3.32) have higher indices; (3.32) 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 only numerical approach capable to handle least squares problems without special assumption to the index seems to be the one developed and described by Schulz *et al.* (1998).

4 Parameter Estimation in PDE Models

Finally, we reach 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), in the determination of diffusive constants in water absorption processes in hygroscopic liquids discussed in Section 4.2, or in multispecies reactive flows through porous media (Zieße *et al.*, 1996). 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

$$\begin{aligned} \mathsf{M}\partial_t \mathbf{u} - \nabla(D\nabla\mathbf{u}) + q\nabla\mathbf{u} &= \mathbf{f}_1(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \\ \partial_t \mathbf{v} &= \mathbf{f}_2(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \\ 0 &= \mathbf{g}(\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{p}) \quad , \end{aligned} \tag{4.36}$$

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 and Mead (1965). 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)]. 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) in his software package `EASY-FIT`[2] applies the method of lines to PDEs with one spatial coordinate and uses several explicit and implicit integration methods to solve the

[2] `EASY-FIT` offers the method of lines to integrate the parabolic PDEs, several integration routines for ODEs, different least-squares methods for solving optimisation problems and statistical subroutines and graphic tools for evaluation and presentation of results. The program works under `WINDOWS-95/NT` and has a `MS-ACCESS` user interface which controls Fortran programs for integration and optimization. The least squares problem is either solved by `DFNLP` or by `DN2GB`. `DFNLP`: the original problem is transformed to a more general nonlinear program-

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), instead, couple the method of lines (in one and two spatial coordinates) with Bock's (1987) 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 discretisation points. Let us demonstrate the method by applying it to the diffusion equation

$$\frac{\partial}{\partial t}c(t,z) = D\frac{\partial^2}{\partial z^2}c(t,z) \quad , \quad \begin{array}{l} 0 \le t < \infty \\ 0 \le z \le L \end{array} \tag{4.37}$$

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

$$z_i = i\Delta z \quad , \quad \Delta z := \frac{L}{N} \quad , \quad c_i = c_i(t) = c(t,z_i) \quad , \quad i = 0,\ldots,N \quad . \tag{4.38}$$

If we choose a finite difference approximation we get

$$\frac{\partial^2}{\partial z^2}c(t,z) \approx \frac{c(t,z-\Delta z) - 2c(t,z) + c(t,z+\Delta z)}{(\Delta z)^2} = \frac{c_{i-1} - 2c_i + c_{i+1}}{(\Delta z)^2} \quad , \tag{4.39}$$

which replaces the diffusion equation (4.37) by N ordinary differential equations

$$\dot{c}_i(t) = \frac{c_{i-1} - 2c_i + c_{i+1}}{(\Delta z)^2} \quad . \tag{4.40}$$

4.2 An Example: Parameter Fitting in Hygroscopic Liquids

The water transport and absorption processes within a hygroscopic liquid are described by a model containing the diffusion equation (4.37) 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

ming problem but typical features of a Gauss-Newton and quasi-Newton method are retained (Schittkowski,1986). The resulting optimisation problem is solved using a standard sequential quadratic programming code [NLPQL, *cf.* Schittkowski (1983)].

`DN2GB`: this subroutine (also available in NETLIB-TOMS FORTRAN library) has been developed by Dennis *et al.* (1981), and is based on a combined Gauss-Newton and quasi-Newton approach.

additional integral relation describing the total amount of water in the liquid. The model includes three parameters to be estimated.

The available measurement data provide the total time dependent concentration $C(t)$ of water in the liquid. The mathematical solution of the diffusion equation, however, is the water concentration $c(t, z)$ in the hygroscopic liquid and it is a function of time and location. To be able to compare the mathematical solution with the observed data one has to integrate $c(t, z)$ over the space coordinate z, *i.e.*, the depth of the fluid.

4.2.1 The Description of the Diffusive Process

The PDE in the model is the diffusion equation (4.37), where the diffusion constant, D, is assumed constant across the whole vessel of depth L.

At the start ($t = 0$) of the experiments the liquid contains unavoidably already some initial and known total water concentration C_0 of water assumed to be uniformally distributed. Therefore the initial values for the partial differential equation are chosen as follows:

$$c(t, z) = C_0 \quad \text{for} \quad 0 < z < L \quad \text{and} \quad t = 0 \quad . \tag{4.41}$$

It would be possible to include other assumptions, *e.g.*, initial concentrations depending on location.

In order to be sure that no liquid leaves the vessel at the bottom ($z = L$) we apply the flux condition

$$\frac{\partial}{\partial z} c(t, z) = 0 \quad \text{for} \quad z = L \quad \text{and} \quad t \geq 0 \quad . \tag{4.42}$$

At the upper boundary ($z = 0$) separating air and liquid we have to describe a process in which absorption and diffusion act simultaneously. The surface boundary is not saturated except for a negligible time interval. If it were we could apply the upper boundary condition of the PDE

$$c(t, z) = C_\infty \quad \text{for} \quad z = 0 \quad \text{and} \quad t \geq 0 \quad . \tag{4.43}$$

Since water is absorbed at the surface on a time-scale comparable to that of the diffusion process, the surface layer is not always in a saturated status. Therefore we have to consider absorption and diffusion simultaneously, and to include a time-dependent water concentration in the surface layer. This feature is modeled as an equilibrium relation relating the absorption of water by the hygroscopic liquid to the emission of water. The absorbing flux is $k_A a$ whereas the flux vice versa is $k_B c(t, z = 0)$, k_A and k_B are additional least squares parameters, and a is the concentration of water in the air, assumed to be constant during the experiment. The function $c(t, z = 0)$ is the concentration of water at the surface of the hygroscopic liquid. The upper boundary condition of the PDE can be

formulated as a flux balance equation

$$-D\frac{\partial}{\partial z}c(t,z)\,|_{z=0} = k_A a - k_B c(t,z=0) \quad , \tag{4.44}$$

which is of the type of a mixed Dirichlet-Neumann condition.

Since the boundary condition (4.44) describes absorption and emission of water only approximately we included processes of higher order, *e.g.*,

$$\begin{aligned} -D\frac{\partial}{\partial z}c(t,z)\,|_{z=0} &= k_{A,0}\cdot a + k_{A,1}\cdot a^2 + ... \\ -k_{B,0}\cdot c(t,z &= 0) - k_{B,1}\cdot c(t,z=0)^2 - ... \quad . \end{aligned} \tag{4.45}$$

However, the inclusion of processes of higher order did not improve the least squares fits. In all least squares fits the higher order coefficients were estimated to be zero.

4.2.2 Local Concentrations, Total Concentration, and Weights

The diffusion equation (4.37) describes the concentration of water in the hygroscopic liquid. However, the result of the experiments is the time-dependent weight of liquid containing the absorbed water. Therefore, one has to convert the weight into concentration. This can be done as follows:

$$W' = W_0 + \rho_{H2O}\cdot S\cdot L\cdot C_0 \quad , \tag{4.46}$$

where W' is measured initial weight, W_0 is the real initial weight of the liquid (without contained water), ρ_{H_2O} is the specific density of water, and S is the surface area of the liquid.

Using $C(t)$ for the time-dependent total water concentration (relative concentration in percent due to the weight of liquid) and $W(t)$ for the time-depending weight of the liquid one can formulate

$$C(t) = \frac{W(t) - W_0}{W_0} \quad . \tag{4.47}$$

After substitution of the unknown terms this formula yields

$$C(t) = \frac{W(t) - W' + \rho_{H2O}\cdot S\cdot L\cdot C_0}{W' - \rho_{H2O}\cdot S\cdot L\cdot C_0} \quad , \tag{4.48}$$

which converts the weights into relative concentrations.

For the experiments discussed here it was not possible to measure $c(t,z)$ reliably. Only the total water concentration $C(t)$ can be measured. Therefore, the solution of the partial differential equation, *i.e.*, the space and time dependent concentration $c(t,z)$, have to be converted to the time-dependent

total concentration $C(t)$ in order to be able to compare model function and measurement data.

For this purpose the calculation of $W(t)$ has to be substituted by

$$W(t) = W_0 + \rho_{H_2O} \cdot S \cdot \int_0^L c(t,z)\mathrm{d}z \quad , \tag{4.49}$$

which considers the location dependent water concentration $c(t,z)$. Substituting this expression into (4.48) for $C(t)$ one gets

$$C(t) = \frac{\rho_{H2O} \cdot S}{W' - \rho_{H2O} \cdot S \cdot L \cdot C_0} \cdot \int_0^L c(t,z)\mathrm{d}z \quad . \tag{4.50}$$

Usually the concentration of water in the air (humidity) is expressed in relative values, *i.e.*, in percent of the absorption capacity of air due to the temperature. In order to keep concentrations comparable one has to convert the water concentration in air from these relative values to ppm or to relative values related to the weight. This conversion is based on linear interpolation within the table of Buchholz (1954).

4.2.3 Constants and Parameters

Besides the measurement data for the total water concentration $C(t)$ in the hygroscopic liquid one also needs the radius of the vessel (the problem is assumed to be radially symmetric), the depth of the liquid, the temperature and the relative humidity. There is an estimate for the diffusion constant which can serve as an initial value and as a clue for comparison with the result as well.

Based on theoretical considerations the following estimate $D \approx 4.8 \cdot 10^{-5}$ [cm^2/min] was found using Perry and Green (1984). To avoid being trapped in local minima several initial values for the diffusion constant D and the velocity constants k_A and k_B have been chosen.

4.2.4 Software and Solution Method

To solve the least squares problem and estimate the diffusion coefficient D and the velocity constants k_A and k_B we used the software package `EASY-FIT`. The ODEs are integrated by `RADAU5`, a program using an implicit Radau-type Runge-Kutta method of order five for stiff equations [see, for instance, Hairer and Wanner (1991)].

4.2.5 Results and Interpretation

Due to the nature of nonlinear problems it is possible that the solver finds several local minima. The problem introduced here has at least two local minima.

However, there are criteria available which allow to eliminate false ones:

· · the quality of the fitted measurements can be significantly different. We have chosen to compare the sum of squared deviations divided by the sum of squared observation values, *i.e.*,

$$\sigma_{\text{fit}} := \sum_{k=1}^{N} \left[\frac{C_k^{\text{obs}} - C(t_k)}{C_k^{\text{obs}}} \right]^2 \tag{4.51}$$

· · the user also should compare the fits optically and decide which is the better one.

· · besides the quality of the fit the solution should be also plausible and should be realistic. For example, if the value for the diffusion coefficient is much smaller or larger than one would expect from experiments or other reasoning, the related solution is not very probable.

For a set of experiments typical results are shown in the table below. There might be additional solutions, especially for the velocity constants k_A and k_B, as they cannot be calculated uniquely from the equations.

	Dim.	08_4_z	16_4_z	24_4_z
$10^4\sigma_{\text{fit}}$		2.27	1.15	0.679
$10^4 D$	cm^2/min	1.65± 0.11	2.12± 0.09	2.25± 0.07
$10^2 k_A$	cm/min	12.6± 6.2	1.88± 0.08	1.79± 0.06
$10^3 k_B$	cm/min	15.3± 7.4	2.22± 0.09	2.12± 0.07
C_∞	ppm	8400 (8420)	8980 (8860)	8970 (8960)

In the last line the calculated asymptotic water concentration C_∞ is mentioned together with its measured value (in brackets).

For testing the reliability of a solution or identifying false local minima one can consider their asymptotic properties. For the upper boundary condition

$$-D \left.\frac{\partial c}{\partial z}\right|_{z=0} = k_A \cdot a - k_B \cdot c(t, z=0) \tag{4.52}$$

the left-hand-side approaches zero for times t because the water concentration has reached its saturation value and is almost constant. Then this equation can be reformulated as:

$$\frac{k_A}{k_B} = \frac{c(t \to \infty, z=0)}{a} \quad . \tag{4.53}$$

Relation (4.53) could also be added as an equality constraint to the model. The following table shows the results [$c(t, z=0)$ and a are specified in ppm] for the single terms and the quotients mentioned above:

Dataset	k_A	k_B	$c(t, z = 0)$	a	$\frac{k_A}{k_B}$	$\frac{c(t\to\infty, z=0)}{a}$
08_4_y	10	1.3	8440	1059	7.69	7.97
24_4_y	9.3	1.1	9180	1059	8.45	8.67
08_4_z	0.12	0.015	8400	1059	8.00	7.93
16_4_z	0.0188	0.00222	8980	1059	8.47	8.48
24_4_z	0.0179	0.00212	8970	1059	8.44	8.47
16a	0.086	0.0097	7046	794	8.87	8.87
16b	0.022	0.00333	8850	1334	6.61	6.63

The ratios deviate for the two first results by less than 5% for all other results even by less than 1%, which confirms the theoretical considerations well.

Finally, the current software and techniques were used to analyze the temperature dependence of D, k_A and k_B. One of the anticipated results was that temperature can affect significantly absorption and diffusion of water.

5 Conclusions

This review demonstrates that ODE and PDE based least squares problems appear often in the applied sciences. Academic and commercial least squares solvers as well as software packages are available. The structured methods described in this paper are not yet frequently used in astronomy and astrophysics but there are examples. Kallrath *et al.* (1993) show how such methods could be applied to celestial mechanics problems and parameter fitting in chaotic dynamical systems. Wehrse and Rosenau (1997) are working on the application of the boundary value problem approach described in Section 2 to the analysis of stellar spectra of late type stars. They fit the model flux to the observed flux at several wavelength, and their least squares problem includes the radiative transfer equation describing the atmosphere, the energy equation, the hydrostatic equation, the equation of state as well as expressions for the absorption and scattering coefficients and bounds on gravity, effective temperature and abundances.

Thus, historically, we see least squares problems and techniques arising in astronomy, and now, after more than 200 years in which these methods saw many improvements, they still have great relevance in astronomy and astrophysics.

Acknowledgements: Thanks is directed to Anna Schreieck (BASF-AG) for a careful reading of the manuscript and Johannes P. Schlöder (IWR, Universität Heidelberg) for discussion on the subject of parameter estimation.

References

Arning, M.: 1994, Lösung des Inversproblems von partiellen Differential-

gleichungen beim Wassertransport im Boden, Dissertation, TU Carolo-Wilhelmina zu Braunschweig, Braunschweig, Germany

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

Bader, G. and Deuflhard, P.: 1981, *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

Bock, H. G.: 1981, Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics, in K. H. Ebert, P. Deuflhard, and W. Jäger (eds.), *Modelling of Chemical Reaction Systems*, Series in Chemical Physics, pp 102–125, Springer, Heidelberg

Bock, H. G.: 1987, *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, Preprint 142, Universität Heidelberg, SFB 123, Institut für Angewandte Mathematik, 69120 Heidelberg

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

Buchholz, E.: 1954, Das $i-x$ Diagramm von Mollier und seine Anwendung bei der Bedienung von Luftbehandlungsanlagen, *Energie* 10

Dennis, J. E., Gay, D. M., and Welsch, D. M.: 1981, Algorithm 573: NL2SOL-An Adaptive Nonlinear Least-squares Algorithm, *Transactions on Mathematical Software* **7**, 369–383

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

Deuflhard, P. and Apostolescu, V.: 1977, An Underrelaxed Gauss-Newton Method for Equality Constrained Nonlinear Least Squares Problems, in J. Stoer (ed.), *Proc. 8th IFIP Conf Würzburg Symposium on the Theory of Computing*, No. 23 in Springer Lecture Notes Control Inf. Sci., Springer, Heidelberg-Berlin-New York

Deuflhard, P. and Apostolescu, V.: 1980, A Study of the Gauss-Newton Method for the Solution of Nonlinear Least Squares Problems, in J. Frehse, D. Pallaschke, and U. Trottenberg (eds.), *Special Topics of Applied Mathematics*, pp 129–150, North-Holland, Amsterdam

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

Eichhorn, H.: 1993, Generalized Least Squares Adjustment, a Timely but Much Neglected Tool, *CMDA* **56**, 337–351

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

Kallrath, J., Altstädt, V., Schlöder, J. P., and Bock, H. G.: 1998, Analysis of

Crack Fatigue Growth Behaviour in Polymers and their Composites based on Ordinary Differential Equations Parameter Estimation, *Polymer Testing,* in print

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

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

Legendre, A. M.: 1805, *Nouvelles Methodes pour la Determination des Orbites des Comètes*, Courcier, Paris

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

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

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

Perry, R. H. and Green, D.: 1984, *Perry's Chemical Engineering Handbook*, McGraw-Hill, New York, 6th edition

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

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

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

Schittkowski, K.: 1983, On the Convergence of a Sequential Quadratic Programming Method with an Augmented Lagrangian Search Direction, *Mathematische Operationsforschung und Statistik, Serie Optimization* **14**, 197–216

Schittkowski, K.: 1986, NLPQL: A FORTRAN Subroutine Solving Constrained Nonlinear Programming Problems, *Annals of Operations Research* **5**, 485–500

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

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

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

Schneider, I.: 1988, *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*, Wissenschaftliche Buchgesellschaft, Darmstadt

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

Stigler, S. M.: 1986, *The History of Statistics*, Belknap, Harvard, Cambridge, MA

Wehrse, R. and Rosenau, P.: 1997, Parameter Identification – A New Way of Assessing Errors in Stellar Atmosphere Analysis, in T. R. B. *et al.* (ed.), *Fundamental Stellar Properties: The Interaction between Observations and Theory*, pp 235–238, IAU, The Netherlands

Zieße, M. W., Bock, H. G., Gallitzendörfer, J. V., and Schlöder, J. P.: 1996, Parameter Estimation in Multispecies Transport Reaction Systems Using Parallel Algorithms, in J. Gottlieb and P. DuChateau (eds.), *Parameter Identification and Inverse Problems in Hydrology, Geology and Ecology*, pp 273–282, Kluwer Academic Publishers, Dordrecht, Holland