LEAST SQUARES PARAMETER ESTIMATION IN CHAOTIC DIFFERENTIAL EQUATIONS

JOSEF KALLRATH*
BASF-AG, ZXT/C, Kaiser-Wilhelm-Str.52, D-6700 Ludwigshafen, Federal Republic of Germany

and

JOHANNES P. SCHLÖDER and HANS GEORG BOCK
IWR and Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 368, D-6900 Heidelberg, Federal Republic of Germany

Abstract. A recent least squares algorithm, which is designed to adapt implicit models to given sets of data, especially models given by differential equations or dynamical systems, is reviewed and used to fit the Hénon-Heiles differential equations to chaotic data sets.

This numerical approach for estimating parameters in differential equation models, called the *boundary value problem approach*, is based on discretizing the differential equations like a boundary value problem, *e.g.* by a multiple shooting or collocation method, and solving the resulting constrained least squares problem with a structure exploiting generalized Gauss-Newton-Method (Bock,1981).

Dynamical systems like the Hénon-Heiles system which can have initial values and parameters that lead to positive Lyapunov exponents or phase space filling Poincaré maps give rise to chaotic time series. Various scenarios representing ideal and noisy data generated from the Hénon-Heiles system in the chaotic region are analyzed *w.r.t.* initial conditions, parameters and Lyapunov exponents. The original initial conditions and parameters are recovered with a given accuracy. The Lyapunov spectrum is then computed directly from the identified differential equations and compared to the spectrum of the "true" dynamics.

Key words: least squares techniques – numerical parameter estimation – boundary value problem approach – dynamical systems – Hénon-Heiles system – Lyapunov spectrum

1. Introduction

The method of least squares, introduced by Gauss (1809) to determine the orbits of Ceres and Pallas, is still of great significance for the analysis of observations, in astronomy as well as in experimental sciences in general. Since the time of Gauss, numerical methods for solving several types of least squares problems have been developed and improved, and there is still active research in that area. For a review on the methods of least squares as known and used in astronomy we refer to Eichhorn (1992).

A common basic feature and limitation of least squares methods used in astronomy, which is however seldom explicitly noted, is that they require an explicit model which is fitted to the data. Most models have a well founded base in physics, and are often described in terms of differential equations, which are often not solvable in a closed analytical form. Therefore, it seems desirable to develop solution algorithms that require only the statement of the model equations as a side condition, which implicitly determines the solution. The demand for such techniques

* presently at IWR, Universität Heidelberg, Im Neuenheimer Feld 368, D-6900 Heidelberg, Germany

Celestial Mechanics and Dynamical Astronomy **56**: 353–371, 1993.

becomes even stronger when considering the rapidly increasing field of non-linear dynamics in physics and astronomy, non-linear reaction kinetics in chemistry, and non-linear systems describing ecosystems in biology or environmental sciences.

In this paper, a *boundary value problem approach* is reviewed and used to fit a four dimensional, quadratic system of Hamiltonian differential equations with four unknown initial values x_{10}, x_{20}, x_{30} and x_{40} and three unknown parameters a, b and c to chaotic time series. For $a = 1$, $b = 1$ and $c = -1$ the equations (see section 5 for details)

$$\begin{aligned} \dot{x}_1 &= x_3 \\ \dot{x}_2 &= x_4 \\ \dot{x}_3 &= -ax_1 - 2x_1x_2 \\ \dot{x}_4 &= -bx_2 - x_1^2 - cx_2^2 \end{aligned} \tag{1.1}$$

are identical to the system investigated by Hénon and Heiles (1964) describing the motion of a star within the potential of a cylindrical galaxy.

Usually, experimental data from dynamical systems in the chaotic region are analyzed with time series analysis techniques (Eckmann and Ruelle, 1985), and the goal is to identify basic characteristic quantities of the system like Lyapunov exponents λ_i, or in dissipative systems the dimension of the attractor. These procedures, however, have the disadvantage that in order to obtain the λ_i from time series one may easily need a few thousand data points according to Zeng *et al.*(1991) or Holzfuss and Parlitz (1991).

While the λ_i are certainly a useful tool to distinguish between quasi-periodic or chaotic orbits, it still seems to be a greater advantage if the underlying dynamical system of the time series could be derived from the data. Therefore, one objective of this paper is to demonstrate that it is actually possible to fit the solution of a differential equation system to a chaotic time series and estimate the parameters, and that a few hundred data points are sufficient to reach that goal. In addition, we use the differential equations and the identified parameters to compute the Lyapunov spectrum by the algorithm of Bennetin *et al.* (1980).

Since a dynamical system at first sight has the structure of an initial value problem, an obvious approach is to integrate the equations with some initial values and parameters, to insert the trajectories into the least squares function, and to correct the unknowns by an iterative procedure till a minimum is reached. However, the numerical error propagation involved in the numerical integration of dynamical systems with a strong sensitivity to changes in initial values, *e.g.* those with positive Lyapunov exponents λ_i, prohibits the link of an initial value problem solver to a least squares solver in this fashion.

On the other hand, however, one may expect that the trajectories are very sensitive to perturbations of the parameters which is another way of saying that the problem of parameter estimation itself may be especially well-posed.

The problem remains to find *adequate numerical methods* that are able to cope with parameter-dependent instabilities in differential equations and are thus capable to recover parameters from given noisy chaotic trajectories.

It will be shown that on numerical methods based on the *boundary value problem* (BVP) *approach* as described in section 4 are indeed able to solve such problems (section 5) because of their stability properties. The suitability of this class of methods for the identification of unstable and chaotic processes was explained in Bock and Schlöder (1986), a thorough investigation to chaotic data from dissipative systems is given in Baake *et al.* (1992).

2. The Structure of the Problem

Parameter estimation means to optimally adapt models to given data by determining their parameter values such that the deviation of model and data is minimized in a suitable norm. In this paper we concentrate on models given by ordinary differential equations (ODE)

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{p}; sign\boldsymbol{Q}(t, \boldsymbol{x}, \boldsymbol{p})), \quad \boldsymbol{x}\epsilon\mathbf{R}^{n_d}, \quad \boldsymbol{p}\epsilon\mathbf{R}^{n_p} \tag{2.1}$$

where t is the independent variable, $\boldsymbol{x}\epsilon\mathbf{R}^{n_d}$ are the states, and $\boldsymbol{p}\epsilon\mathbf{R}^{n_p}$ is the vector of unknown parameters. Depending on the vector of signs of the state and parameter dependent switching functions $\boldsymbol{Q}$ the right hand side $\boldsymbol{f}$ is defined piecewise smooth. At a root τ of a component of $\boldsymbol{Q}$, jumps of the states

$$\boldsymbol{x}(\tau^+) = \boldsymbol{x}(\tau^-) + \boldsymbol{s}(\tau^-, \boldsymbol{x}(\tau^-), \boldsymbol{p}) \tag{2.2}$$

are admitted, and the right hand side f may change discontinuously

$$\dot{\boldsymbol{x}} = \begin{cases} \boldsymbol{f}_i \quad (t, \boldsymbol{x}, \boldsymbol{p}), & t \leq \tau \\ \boldsymbol{f}_{i+1}(t, \boldsymbol{x}, \boldsymbol{p}), & t > \tau \end{cases} \tag{2.3}$$

Additional requirements on the solution of the ODE (2.1), like periodicity, initial or boundary conditions, range restrictions to the parameters can be formulated in a vector of (component wise) equality and inequality conditions

$$\boldsymbol{r}_2[t_1, \boldsymbol{x}(t_1), \ldots, t_k, \boldsymbol{x}(t_k), \boldsymbol{p}] = 0 \; or \geq 0 \tag{2.4}$$

All in all we arrive at a fairly general boundary value problem with jumps (2.1) and switching condition (2.2) and (2.3).

More general dynamics (implicit differential equations, differential algebraic systems, discretized partial differential equations etc.) can often be reduced in principle to the above form. However, efficient and stable algorithms have to take into account their special structures. For a treatment of the differential algebraic case see Bock *et al.* (1988).

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

$$\ell_2(\boldsymbol{x},\boldsymbol{p}) := \|\boldsymbol{r}_1[t_1, \boldsymbol{x}(t_1), ..., t_k, \boldsymbol{x}(t_k), p]\|_2^2 = \min_{\boldsymbol{x},\boldsymbol{p}} \tag{2.5}$$

This includes the probably most common case, where one component of the vector function $\boldsymbol{r}_1$ is given by

$$r_{1j}[t_1, \boldsymbol{x}(t_1), ..., t_k, \boldsymbol{x}(t_k), \boldsymbol{p}] = [\eta_{ij} - g_{ij}(t_i, \boldsymbol{x}(t_i), \boldsymbol{p})]/\sigma_{ij} \tag{2.6}$$

leading to

$$\ell_2(\boldsymbol{x},\boldsymbol{p}) := \sum_{i,j} {\sigma_{ij}}^{-2} \cdot [\eta_{ij} - g_{ij}(t_i, \boldsymbol{x}(t_i), \boldsymbol{p})]^2 \tag{2.7}$$

Here, η_{ij} is the observed value which is related to states and parameters at time t_i by

$$\eta_{ij} = g_{ij}(t_i, \boldsymbol{x}(t_i), \boldsymbol{p}) + \epsilon_{ij} \tag{2.8}$$

The numbers ϵ_{ij} are the measurement errors, ${\sigma_{ij}}^2$ are weights that have to be adequately chosen due to statistical considerations, *e.g.* as the measurement error variances. In the case of independent Gaussian $N(0, {\sigma_{ij}}^2)$ errors and known variances ${\sigma_{ij}}^2$ (up to a common factor β^2) the solution of the least squares problem is a maximum likelihood estimate.

Note, that also correlated measurements may be considered by (2.5) using an according formulation of the vector $\boldsymbol{r}_1$.

Note further, that $\boldsymbol{r}_1$ may be a nonlinear function of states and parameters and observation times which themselves may be disturbed by measurement errors. Neither all components of the states nor dense data are required (see sect.5). A procedure for the solution of such a problem should of course deliver an optimal fit that satisfies all restrictions. In addition, it is indispensable in practical problems to deliver also information about the statistical quality of the parameters and the states (*e.g.* in terms of a $variance - covariance$ matrix of these quantities).

3. The Initial Value Problem Approach

The obvious approach to estimate parameters in ODE, which is also implemented in most commercial packages, 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. However, this procedure often exhibits poor performance or even does not work at all depending on the (usually parameter dependent) stability behavior of the IVPs that may lead *e.g.* to exploding trajectories or solutions that run into steady states with extremely poor sensitivities *w.r.t.* parameter variations. In the case of dynamical systems with positive Lyapunov exponents this approach is not advisable since error propagation dominates the system.

On the other hand the parameter estimation problem itself may be very well posed. Since small perturbations in the parameters may lead to strong perturbations of the trajectories, given data of a trajectory can be expected to determine the parameters very well. The next section describes an approach that optimally exploits the inverse structure of parameter estimation problems and is thus able to cope with the disastrous error propagation in chaotic systems.

4. The Boundary Value Problem Approach

In the following we sketch a versatile realization of an alternative method — the *boundary value problem approach* introduced in Bock (1981). It consists of a multiple shooting method for the discretization of the boundary value problem side condition and a generalized Gauss-Newton-Method for the solution of the resulting structured nonlinear constrained least squares problem. A detailed description and analysis of this family of methods is included in Bock (1987).

Depending on the stability behaviour of the ODE and the availability of information about the process (measured data, qualitative knowledge about the problem, etc.) the user chooses a suitable grid $\mathcal{T}_m$ of m multiple shooting nodes τ_j ($m-1$ subintervals $\mathcal{I}_j$)

$$\mathcal{T}_m : \tau_1 < \tau_2 < \ldots < \tau_m, \quad \Delta\tau_j := \tau_{j+1} - \tau_j, \quad 1 \leq j \leq m-1 \tag{4.1}$$

covering the interval where measurements are given ($[\tau_1, \tau_m] \supseteq [t_0, t_f]$). At each grid point new variables $\{s_j := x(\tau_j)\}$ are introduced and $m-1$ initial value problems

$$\dot{x} = f(t, x, p), \quad x(\tau_j) = s_j, \quad t\epsilon[\tau_j, \tau_{j+1}] \tag{4.2}$$

are considered on the subintervals. The $m-1$ vectors of initial values s_j, the value s_m at the end point, the parameter vector p and the observation times t_i are summarized in an augmented vector

$$z^t := (s_1^t, \ldots, s_m^t, p^t, t_1, \ldots, t_k) \tag{4.3}$$

The least squares function (2.5) is then regarded as a function of this vector

$$\ell_2(z) = \|\hat{r}_1[x(t_1, z), \ldots, x(t_k, z), z] \tag{4.4}$$

or in the special case (2.7) $\ell_2(z) = \sum_j \sigma_{ij}^{-2} \cdot [\eta_{ij} - \hat{g}_{ij}(x(t_i, z), z)]^2$, and the ODE (2.1) is replaced by the matching conditions

$$x(\tau_{j+1}; s_j, p) - s_{j+1} = 0, \quad 1 \leq j \leq m-1 \tag{4.5}$$

which ensure continuity of the final trajectory. Note, however, that the initial trajectories provided with initial guesses z_i and the trajectories in the course of

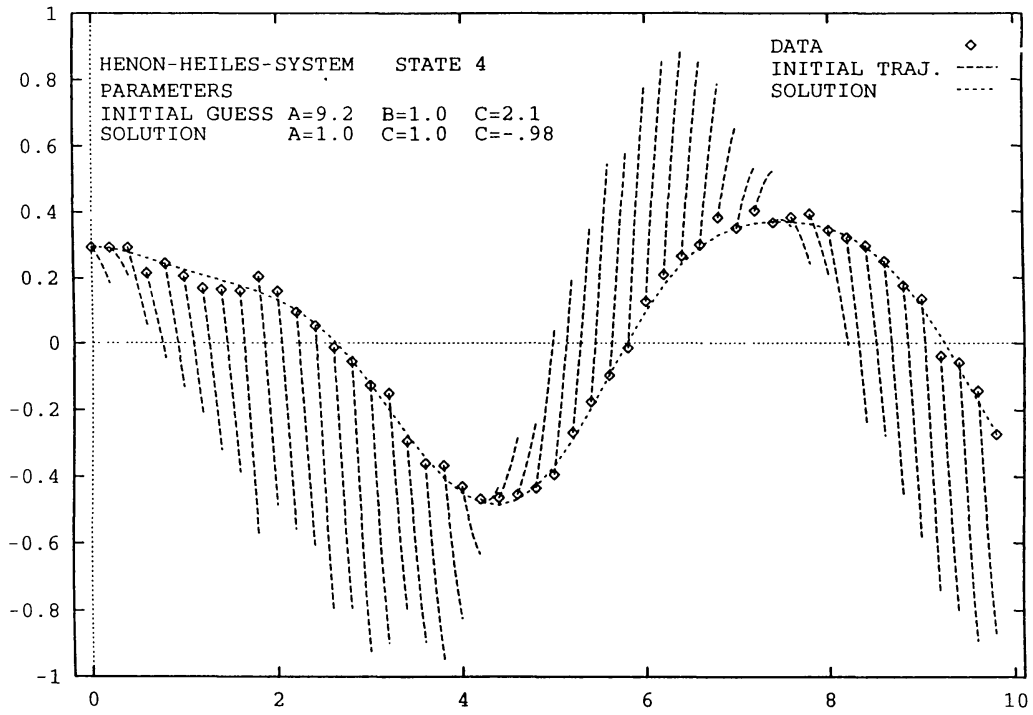

Fig. 1. Data, solution and initial trajectory for x_4 in the Hénon-Heiles system generated with initial values $x_{10} = x_{20} = 0$, $x_{30} = 0.3$ and $x_{40} = -0.41$ and parameters $(a, b, c) = (+9.2, +1, +2.1)$ as initial guesses for the fitting routine. This figure should demonstrate that the initial trajectories may be discontinuous and deviate significantly from the solution.

the iterative procedure are allowed to be discontinuous (see Fig.1). By choosing the multiple shooting intervals $\Delta\tau_j := \tau_{j+1} - \tau_j$ sufficiently small, existence of a (discontinuous) initial trajectory can be guaranteed under mild conditions, and error propagation can be controlled and limited.

Formally, the least squares problem described so far is a constrained optimization problem of the type

$$\min_{z}\{\|F_1(z)\|_2^2 \mid F_2(z) = 0 \; or \geq 0 \in \mathbf{R}^{n_c}\} \tag{4.6}$$

where n_c is the number of constraints (continuity constraints (4.5) and additional constraints (2.4)). This usually large constrained structured non-linear problem is solved by a damped generalized Gauss-Newton method, which is described here for the equality constrained case (see Bock (1987), where also the treatment of inequalities is described). Starting with an initial guess z_0 the variables are iterated via

$$z_{k+1} = z_k + \alpha_k \cdot \Delta z_k \tag{4.7}$$

with a damping constant α_k, $0 < \alpha_{min} \leq \alpha_k \leq 1$. The increment Δz_k is the solution of the problem linearized at z_k

$$\min_{\Delta z_k}\{ \| J_1(z_k)\Delta z_k + F_1(z_k)\|_2^2 | J_2(z_k)\Delta z_k + F_2(z_k) = 0, J_i(z_k) := \partial_z F_i(z_k)\} \tag{4.8}$$

Under appropriate assumptions *w.r.t.* the regularity of the Jacobians J_i there exists a unique solution Δz_k of the linear problem and a unique linear mapping $J_k{}^+$ obeying the relations

$$\Delta z_k = -J_k{}^+ F(z_k), \quad J_k{}^+ J_k J_k{}^+ = J_k{}^+, \quad J_k^t := [J_1(z_k)^t, J_2(z_k)^t] \tag{4.9}$$

The solution Δz_k of the linear problem, or formally the generalized inverse $J_k{}^+$ (Bock 1981) of J_k, follows uniquely from the Kuhn-Tucker conditions

$$J_1^t J_1 \Delta z_k - J_2^t \lambda_c + J_1^t F_1 = 0, \quad J_2 \Delta z_k + F_2 = 0 \tag{4.10}$$

where $\lambda_c \epsilon \mathrm{R}^{n_c}$ is a vector of Lagrange multipliers. For the numerical solution Δz_k of the linear constrained problem (4.8) several structure exploiting methods have been developed that compute special factorizations of J_1 and J_2 and thus implicitly, but not explicitly the generalized inverse J^+. There even exist methods (Schlöder,1988) that generate and decompose the Jacobians simultaneously, and that require an amount of computing time for the generation of the derivatives that is proportional only to the number of degrees of freedom and not to the number of variables.

The availability of the Jacobians J_1 and J_2 allows rank checks in every iteration and automatic detection of violations of the regularity assumptions. In that case automatic regularization and computation of a relaxed solution is possible.

The iteration (4.7) can be forced to converge globally to a stationary point of the problem if the damping factors α_k are chosen appropriately. In the treatment of a large number of practical problems strategies based on "natural level functions" have proven to be very successful (see Appendix A). In the region of local convergence of the full step method, the algorithm converges linearly to a solution that is stable to statistical variations in the observations. An iterate z_k is accepted as solution z^* of the nonlinear constrained problem, if a scaled norm of the increments Δz_k is below a user specified tolerance $(10^{-5}, say)$.

As the Jacobians and their decompositions are available in each iteration, covariance and correlation matrices are easily computable for the full variable vector z. In large scale problems this is usually not desirable since sparsity is destroyed. Therefore very fast algorithms have been developed that compute only the diagonal elements of the covariance matrix. For details of the algorithms, the treatment of more general classes of problems and the properties of the optimization package PARFIT we refer to Bock (1987).

The integrator, which is also used for the computations of the Lyapunov spectrum, is based on an extrapolation method by Bulirsch and Stoer (1966). In addition,

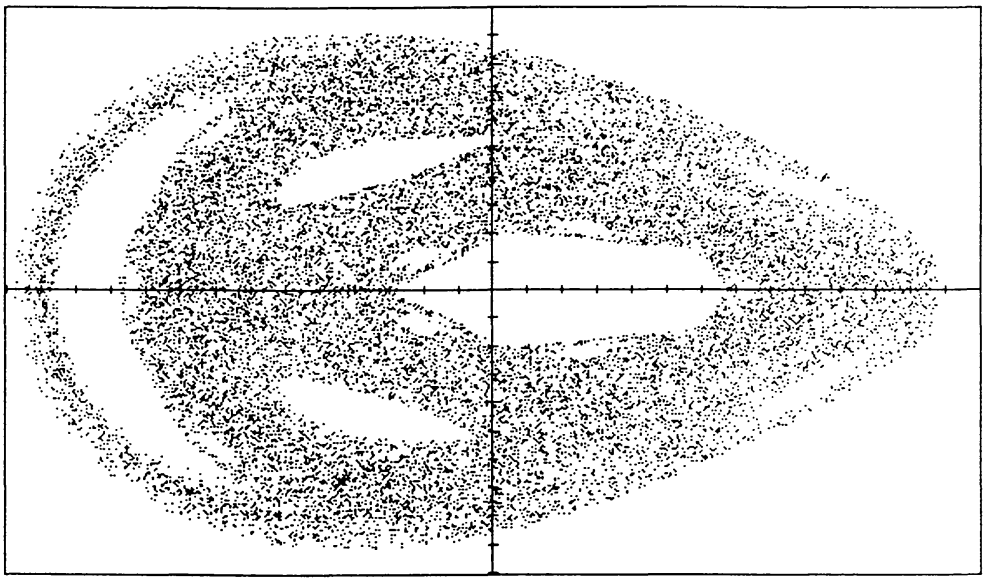

Fig. 2. Poincaré map x_4 versus x_2 in the Hénon-Heiles system generated with initial values $(x_{10}, x_{20}, x_{30}, x_{40}) = (0, 0, +0.3, -0.41)$ and parameters $(a, b, c) = (1, 1, -1)$. The plane of section is defined by $x_1 = 0$ and $x_3 \geq 0$. The ranges of the plot are $-0.5 \leq x_2 \leq +0.7$ and $-0.5 \leq x_4 \leq +0.5$.

it allows the user to compute simultaneously the sensitivity matrices $\boldsymbol{G}$ or $\boldsymbol{H}$ which are the most costly part in computing the Jacobians $\boldsymbol{J}_i$

$$\boldsymbol{G}(t; t_0, \boldsymbol{x}_o, \boldsymbol{p}) := \frac{\partial}{\partial \boldsymbol{x}_o} \boldsymbol{x}(t; t_0, \boldsymbol{x}_o, \boldsymbol{p})$$

$$\boldsymbol{H}(t; t_0, \boldsymbol{x}_o, \boldsymbol{p}) := \frac{\partial}{\partial \boldsymbol{p}} \boldsymbol{x}(t; t_0, \boldsymbol{x}_o, \boldsymbol{p}) \tag{4.11}$$

via the so-called *internal numerical differentiation* (Appendix B) as introduced by Bock (1981), which does not require the often cumbersome and error prone formulation of the variational differential equations

$$\frac{d\boldsymbol{G}}{dt} = \boldsymbol{f}_x(t, \boldsymbol{x}, \boldsymbol{p}) \cdot \boldsymbol{G}, \quad \frac{d\boldsymbol{H}}{dt} = \boldsymbol{f}_x(t, \boldsymbol{x}, \boldsymbol{p}) \cdot \boldsymbol{H} + \boldsymbol{f}_p(t, \boldsymbol{x}, \boldsymbol{p}) \tag{4.12}$$

by the user.

5. Results for the Hénon-Heiles System

In this section, we consider the Hamiltonian

$$H(x, y, \dot{x}, \dot{y}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + U(x, y) \tag{5.1}$$

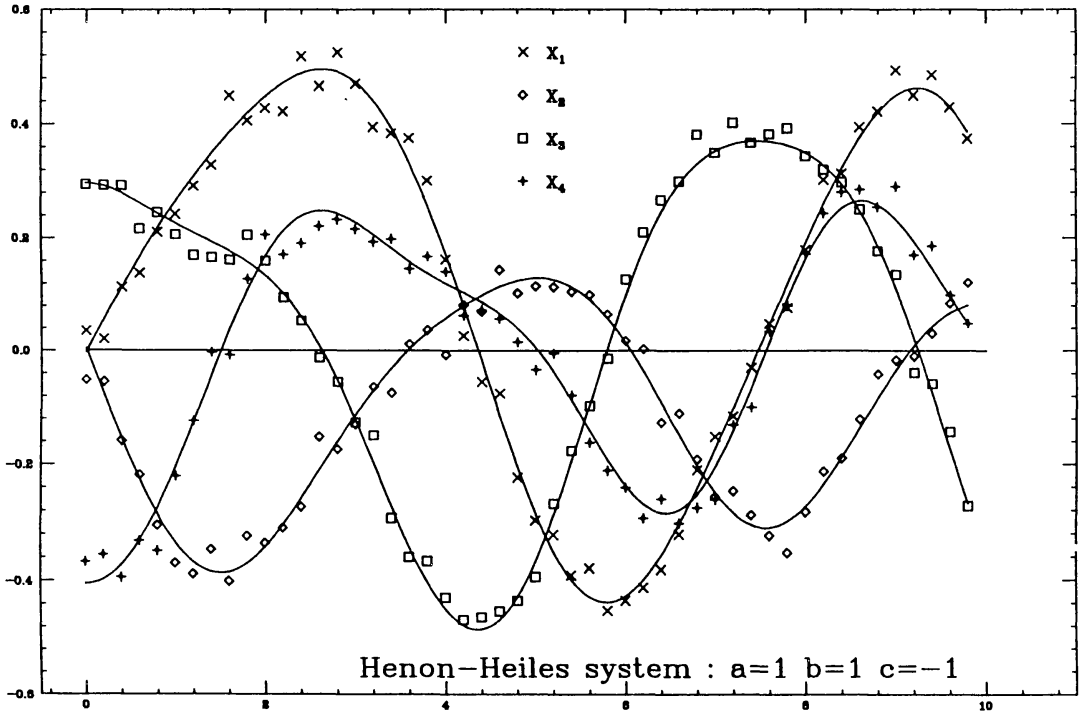

Fig. 3. Trajectories in the Hénon-Heiles system generated with initial values $x_{10} = x_{20} = 0$, $x_{30} = 0.3$ and $x_{40} = -0.41$ and parameters $(a, b, c) = (1, 1, -1)$. Gaussian random numbers with standard deviation $\sigma = 0.05$ are added to the trajectories generated by integration of the dynamical system.

and the potential function

$$U(x, y) = \frac{1}{2}(ax^2 + by^2 + 2x^2y + \frac{2}{3}cy^3) \tag{5.2}$$

From that we derive the equations of motion

$$\ddot{x} = -U_x = -ax - 2xy \tag{5.3}$$

$$\ddot{y} = -U_y = -by - x^2 - cy^2 \tag{5.4}$$

or — via ($x_1 = x$, $x_2 = y$, $x_3 = \dot{x}$, $x_4 = \dot{y}$) — the equivalent first-order system $\dot{x} = f(x)$

$$\dot{x_1} = x_3$$

$$\dot{x_2} = x_4$$

$$\dot{x_3} = -ax_1 - 2x_1x_2$$

$$\dot{x_4} = -bx_2 - {x_1}^2 - c{x_2}^2 \tag{5.5}$$

For the sake of computing the Lyapunov spectrum in a later section we also write down the linearized or variational differential equations along a given nominal trajectory $\boldsymbol{x}$

$$\dot{\xi}_1 = \xi_3$$

$$\dot{\xi}_2 = \xi_4$$

$$\dot{\xi}_3 = f_{31}\xi_1 + f_{32}\xi_2, \quad f_{31} = -(a + 2x_2), \quad f_{32} = -2x_1$$

$$\dot{\xi}_4 = f_{41}\xi_1 + f_{42}\xi_2, \quad f_{41} = -2x_1, \quad f_{42} = -(b + 2cx_2) \tag{5.6}$$

Choosing the parameters $\boldsymbol{p} = (a, b, c) = (+1, +1, -1)$ we get the system investigated by Hénon and Heiles (1964). For energies between $E^- = 1/12$ and $E^+ = 1/8$ one observes a transition from regular behavior characterized by a family of curves which almost completely fill the available space in the (x_2, x_4)-Poincaré surface of section $(x_1 = 0, \dot{x}_3 \geq 0)$ to ergodicity. As shown in Figure 2 one *ergodic* trajectory [*e.g.* initial values $\boldsymbol{x}_o = (x_{10}, x_{20}, x_{30}, x_{40}) = (0, 0, 0.3, -0.41)$] may cover large parts of the permissible area

$$U(0, x_2) + \frac{1}{2} \cdot {x_4}^2 \leq E = 0.12905 \tag{5.7}$$

in the $x_1 = 0$ plane.

In order to set up interesting scenarios for our parameter fitting algorithm we have varied the parameters (a, b, c), have inspected the Poincaré's surfaces of section and have chosen the initial conditions $\boldsymbol{x}_o$ such that they are placed in the chaotic region.

The generation of *simulated data* is performed in two steps starting with the choice of $(\boldsymbol{x},\boldsymbol{p})_s$. By integration of (5.5) we generate a simulated trajectory. Since we expect to have a strong error propagation due to the chaoticity we choose a very stringent local error bound for the integrator ($\epsilon_{int} = 10^{-13}$) and restrict the integration interval such that the global error remains significantly below the intended accuracy of the estimation problem ($\epsilon = 10^{-5}$). To simulate noisy data we add Gaussian random numbers of standard deviation σ within a 3σ strip to our fiducial trajectory. Alternatively, we might use a generalization of the boundary value problem approach to multiple experiments (Schlöder, 1988), in which the time range is split and the according data are treated as separate time series.

In the different scenarios summarized in Tables I to III we vary the parameters $\boldsymbol{p}_s$ and initial conditions $\boldsymbol{x}_o$ for the integration as well as the noise σ added to the simulated trajectories. Figure 3 shows the trajectories in the Hénon-Heiles system generated with initial values $\boldsymbol{x}_o = (0, 0, 0.3, -0.41)$, parameters $\boldsymbol{p} = (1, 1, -1)$ and Gaussian random numbers with standard deviation $\sigma = 0.05$ added to generate the *simulated* data.

TABLE I

This table defines the setup for generating trajectories with energy $E_s = 0.125$, *i.e.* initial values $x_{0s} = (0.00, 0.00, \sqrt{0.1275} \approx 0.35707, -0.35)$ and parameters $(a, b, c)_s = (1, 1, -1)$, and contains different scenarios with 400 *'exact'* data points in the first part. Given some initial guesses a_i, b_i and c_i, the original parameters are recovered with an absolute error less than $5 \cdot 10^{-5}$. The second part of the table contains *simulated* data. In both cases we used 100 multiple shooting nodes. The length of the data series is $t_{end} = (n-1) \cdot \Delta t$. The estimated parameters and their statistical errors derived from the covariance matrix are listed as well as the initial guesses. The estimated errors can be transformed to confidence intervals by multiplying them with the Fisher factor $\gamma = 3.75$. All Lyapunov exponents have been computed but we write down only $\lambda_1 = 0.044$, λ_2 and λ_3 are of opposite sign with absolute values of the order of 10^{-7} while we find $\lambda_4 = -\lambda_1$ with five digits accuracy.

		S11	S12	S13	S14	S15	S16	S17	S18	S19
Interv.:	Δt	0.4	0.4	0.4	0.2	0.2	0.2	0.2	0.2	0.2
Nodes:	n	100	100	100	100	100	100	100	100	100
St. dev:	σ							0.01	0.03	0.05
Guess:	a_i	5	5	10	10	10	20	10	10	10
	b_i	5	5	10	10	10	20	10	10	10
	c_i	-2	2	2	2	10	10	2	2	2
	E_i	E_s	E_s	E_s	E_s	E_s	E_s	0.124	0.122	0.130
Iter.:	k	6	6	6	6	6	16	6	11	14
Estim.:	x_{10}							-0.0003	-0.0017	-0.0036
	x_{20}							0.0006	0.0031	0.0067
	x_{30}							0.3574	0.3585	0.3603
	x_{40}							-0.3496	-0.3479	-0.3455
Estim.:	a							0.9999	0.9995	0.9989
	b							0.9995	0.9969	0.9926
	c							-1.0062	-1.0335	-1.0764
Estim.:	E							0.1250	0.1248	0.1246
Estim.:	Δx_{10}							7.6D−4	3.9D−3	8.4D−3
	Δx_{20}							7.6D−4	3.9D−3	8.1D−3
	Δx_{30}							4.1D−4	2.1D−3	4.5D−3
	Δx_{40}							4.6D−4	2.4D−3	5.1D−3
	Δa							3.2D−4	1.7D−3	3.6D−3
	Δb							2.5D−3	1.3D−2	2.7D−2
	Δc							1.2D−2	6.3D−2	1.4D−1

TABLE II

This table defines the setup for generating trajectories with energy $E_s = 0.12905$, *i.e.* initial values $\boldsymbol{x}_{0s} = (0, 0, 0.3, -0.41)$ and parameters $(a, b, c)_s = (1, 1, -1)$, and contains different scenarios for generating *simulated data*. The estimated parameters and their statistical errors derived from the covariance matrix (Fisher factor $\gamma = 3.75$ to convert to confidence intervals) are listed as well as the initial guesses. As in Table 1 we only write down $\lambda_1 = 0.053$.

		S21	S22	S23	S24	S25	S26
Intervals:	Δt	0.2	0.2	0.2	0.2	0.2	0.2
Nodes:	n	50	50	100	100	100	100
Range:	t_{end}	9.8	9.8	19.8	19.8	19.8	19.8
Points:	n	200	200	400	400	400	400
St. dev.:	σ	0.050	0.020	0.020	0.050	0.025	0.050
Guess:	a_i	+9.2	+1.2	+1.2	+1.2	+10.0	+10.0
	b_i	+1.0	+1.0	+1.0	+1.0	+10.0	+10.0
	c_i	+2.1	-2.0	-2.0	-2.0	+20.0	+20.0
	E_i	0.124	0.124	0.124	0.111	0.127	0.148
Iter.:	k	7	5	5	6	7	14
Estim.:	x_{10}	-0.0036	-0.0009	-0.0013	-0.0049	-0.0018	-0.0049
	x_{20}	0.0036	0.0009	0.0016	0.0062	0.0022	0.0062
	x_{30}	0.2963	0.2991	0.2997	0.2988	0.2996	0.2988
	x_{40}	-0.4060	-0.4090	-0.4081	-0.4024	-0.4073	-0.4024
Estim.:	a	0.9958	0.9989	0.9986	0.9944	0.9980	0.9944
	b	1.0047	1.0011	1.0007	1.0030	1.0010	1.0030
	c	-0.9815	-0.9957	-1.0079	-1.0317	-1.0111	-1.0317
Estim.:	E	0.1263	0.1284	0.1282	0.1256	0.1278	0.1256
Estim.:	Δx_{10}	9.5D−3	2.4D−3	1.6D−3	6.3D−3	2.3D−3	6.3D−3
	Δx_{20}	8.7D−3	2.2D−3	1.7D−3	6.6D−3	2.4D−3	6.6D−3
	Δx_{30}	3.9D−3	9.8D−4	7.6D−4	3.0D−3	1.1D−3	3.0D−3
	Δx_{40}	4.8D−3	1.2D−3	8.4D−4	3.3D−3	1.2D−3	3.3D−3
	Δa	8.4D−3	2.0D−3	8.4D−3	3.3D−3	1.2D−3	3.3D−3
	Δb	3.4D−2	9.1D−3	1.1D−3	4.5D−3	1.5D−3	4.5D−3
	Δc	1.7D−1	4.2D−2	7.5D−3	3.0D−3	1.1D−2	3.0D−3

In order to demonstrate the robustness of the algorithm in Figure 4 we present trajectories obtained with initial values $\boldsymbol{x}_0 = (0, 0, 0.3, -0.41)$ and parameters $\boldsymbol{p}_i = (9.2, 1, 2.1)$ as initial guesses for the fitting routine. As listed in Table II, the original parameters $\boldsymbol{p} = (1, 1, -1)$ are well recovered although the initial trajectories correspond to escape orbits and are far off from the solution.

Tables I to III contain three different scenarios, and in each of them we have tried to identify the parameters under different circumstances, *i.e.* different distances between the initial parameters and solution parameters, different noise and available data for fitting, and different lengths t_{end} of the time series.

The parameter estimations listed in Table I have been derived from *exact* data, *i.e.* from trajectories not disturbed by noise. These experiments demonstrate the

TABLE III

This table defines the setup for generating trajectories with energy $E_s = 0.1849$, *i.e.* initial values $x_{0s} = (0,0,0.43,0.43)$ and parameters $(a,b,c)_s = 1.3 \cdot (1,1,-1)$, and contains different scenarios for re-estimating the parameters from *simulated data*. The initial parameter guess is $p_i = (5,5,5)$ corresponding to a guess energy of $E_i = 0.2027$. The estimated parameters and their statistical errors are listed as a function of the data points considered for fitting. If a curve has weight 0 then it is not considered in the least-squares problem. As in Table I we only write down $\lambda_1 = 0.037$.

		S31	S32	S33	S34	S35	S36
Intervals:	Δt	0.2	0.2	0.2	0.2	0.4	0.4
Nodes:	n	100	100	100	100	100	100
Range:	t_{end}	19.8	19.8	19.8	19.8	39.6	39.6
St. dev.:	σ	0.05	0.05	0.05	0.05	0.05	0.07
Weight:	w	1111	1110	1100	1000	1000	1111
Iter.:	k	5	5	5	11	7	6
Estim.:	x_{10}	0.0017	0.0041	-0.0030	-0.0144	-0.0066	0.0083
	x_{20}	0.0110	0.0054	0.0042	0.0362	0.0509	-0.0080
	x_{30}	0.4023	0.4324	0.4338	0.4433	0.4322	0.4329
	x_{40}	0.4834	0.4367	0.4371	0.4272	0.4276	0.4292
Estim.:	a	1.3010	1.2965	1.3024	1.3618	1.2984	1.2997
	b	1.2905	1.3000	1.3019	1.2021	1.3150	1.2937
	c	-1.2683	-1.2848	-1.2882	-0.9491	-1.3957	-1.2669
Estim.:	E	0.1868	0.1889	0.1896	0.1900	0.1854	0.1859
Estim.:	Δx_{10}	6.0D−3	6.3D−3	9.0D−3	1.7D−2	1.2D−2	8.5D−3
	Δx_{20}	6.9D−3	9.6D−3	1.0D−2	4.0D−2	3.5D−2	1.2D−2
	Δx_{30}	3.3D−3	3.6D−3	4.8D−3	1.2D−2	6.3D−3	6.0D−3
	Δx_{40}	3.0D−3	4.2D−3	4.2D−3	1.8D−2	9.2D−3	5.5D−3
	Δa	6.0D−3	6.9D−3	8.4D−3	6.5D−2	4.3D−3	3.0D−3
	Δb	9.D−3	1.2D−2	1.2D−2	8.9D−2	1.3D−2	6.0D−3
	Δc	3.0D−2	3.9D−2	3.9D−2	3.7D−1	7.2D−2	3.1D−2
Fisher:	γ	3.75	3.75	3.75	3.90	3.90	3.75

large region of convergence of the method. Table II contains scenarios defined by time series with 200 or 400 data points perturbed with Gaussian random numbers varying between $\sigma = 0.02$ and 0.05. We choose the standard deviation σ according to

$$\sigma \approx 0.1 \cdot \max(|x_i(t)|) \tag{5.8}$$

and use only those perturbed data points which lie within a 3σ strip. The error limits given in the tables are derived from the diagonal elements of the covariance matrix. In order to get confidence intervals based on a 95% confidence level, these values have to be multiplied with the Fisher factor

$$\gamma := \sqrt{\ell_1 \cdot F(1-\alpha, \ell_1, \ell_2)}, \quad \ell_2 = n - \ell_1, \quad \alpha = 0.95 \tag{5.9}$$

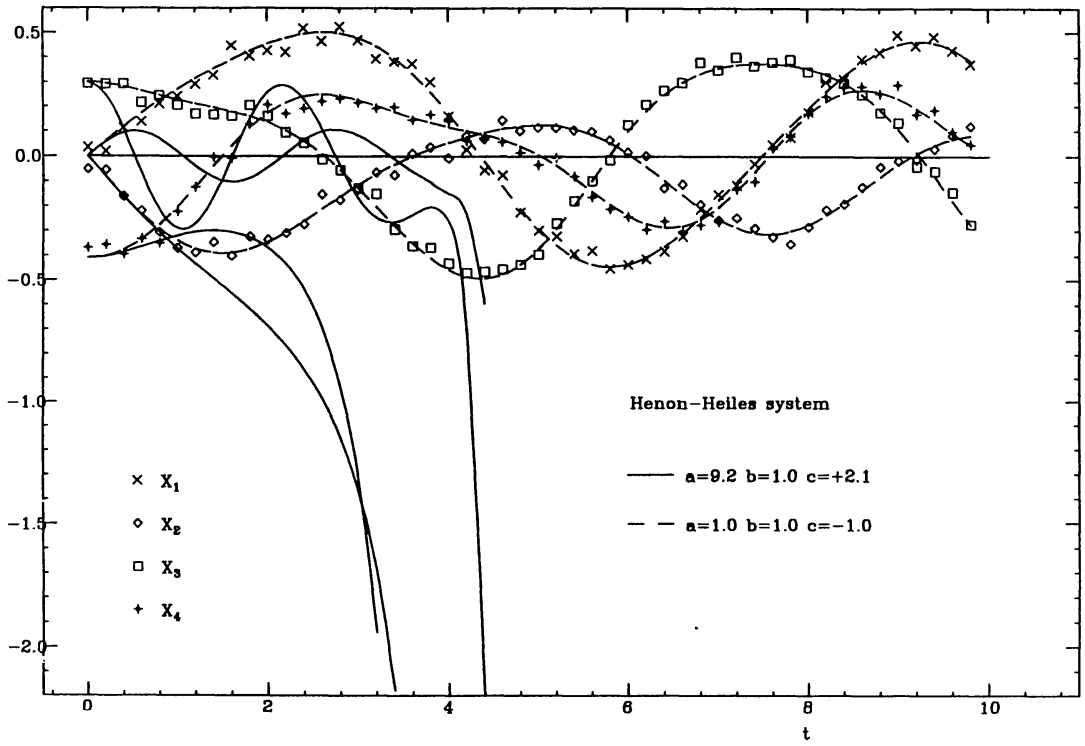

Fig. 4. Trajectories in the Hénon-Heiles system generated with initial values $x_{10} = x_{20} = 0$, $x_{30} = 0.3$ and $x_{40} = -0.41$ and parameters $(a, b, c) = (+9.2, +1, +2.1)$ as initial guesses for the fitting routine. As listed in Table I the original parameters $(a, b, c) = (1, 1, -1)$ are well recovered although the trajectories corresponding to the initial parameter guesses are far of from the solution.

where $\ell_1 = 7$ gives the degree of freedom, and $F(1-\alpha, \ell_1, \ell_2)$ is the F-distribution (Abramowitz and Stegun, 1970) for the quantil α. The Fisher factor is also listed in the tables. As expected the parameters are determined more accurately if the number n of data points or t_{end} is increased.

The scenarios in Table III belong to a different set of initial values and parameters corresponding to an energy of $E = 0.1849$ slightly below the escape energy $E_{esc} \approx 0.21$ derived from the contour lines of the potential. Starting with some moderately perturbed parameter guesses we decrease n by disregarding successively the x_4, x_3 and x_2 data. For this selective procedure we observe, of course, that the sum of squared residuals also decreases, while the errors of the parameters increase. Scenario S34 contains only the x_1 data and the error of parameter c increases to more than 10%. In scenario S35 the error is reduced again by increasing Δt to 0.4, or equivalently t_{end} to 39.6 which gives a time series of doubled length. This experiment shows that in order to identify the parameters of the Henon-Heiles system, two curves, for instance the x_1 and x_2 curves, prove to be sufficient, and even one curve is sufficient if it contains enough data points. This result agrees well with similar experiences of Baake *et al.* (1991) in the dissipative Lorenz and Rössler systems. Eventually, in S36 we contaminated all curves with Gaussian errors ($\sigma = 0.07$) which is again 10% of the amplitude of the x_i-curves.

The Lyapunov spectrum $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ also contained in Tables I to III has been computed with the code listed in Wolf *et al.* (1985). This algorithm is based on the work of Bennetin *et al.* (1980). For a continuous dynamical system in an n-dimensional phase space the long-term evolution of an infinitesimal n-sphere of initial conditions is obtained by generation of appropriate directional derivatives, *e.g.* via the integration of (5.6), or by using internal numerical differentiation techniques (sect. 4). This sphere will become a n-ellipsoid with principal axes $r_i(t)$ which will lead to the i^{th} one-dimensional Lyapunov exponent $\hat{\lambda}_i$ (Wolf *et al.*, 1985)

$$\hat{\lambda}_i := \lim_{t\to\infty} \frac{1}{t} \cdot log_2 \frac{r_i(t)}{r_i(0)}, \quad \lambda_i = \ln(2) \cdot \hat{\lambda}_i \approx 0.693 \cdot \hat{\lambda}_i \tag{5.10}$$

In order to get a numerically stable procedure Bennetin *et al.* (1980) suggested that at regular time intervals Δt_{orth} the evolved tangent vectors associated with the n-sphere mentioned above are replaced by a set of new orthonormal vectors, using the Gram-Schmidt procedure, which avoids small angles between the vectors defining the n-ellipsoid.

We would like to stress that the λ_i are ordered with λ_1 being the largest exponent, and that although they are related to the expanding or contracting nature of different directions in phase space it is not possible to assign a well-defined direction with a given exponent. Since the Hénon-Heiles differential equations are derived from a Hamiltonian system with $f = 2$ degrees of freedom we use some additional facts on the Lyapunov spectra of such systems (Froeschlé 1984) such as symmetry with respect to zero, which in our particular case lead to $\lambda_1 \geq 0$, $\lambda_2 = \lambda_3 = 0$ and $\lambda_4 = -\lambda_1$ and $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 0$ corresponding to the Liouville theorem. While the exponents may be interpreted geometrically as average exponential rates of divergence or convergence of nearby orbits in phase space or growth of an infinitesimal volume element, the log_2 function in definition (5.10) allows also to measure the rate at which the system dynamics creates or destroys information in $bits/orbit$. The complete Lyapunov spectra $\boldsymbol{\lambda}$ for the analyzed scenarios were calculated with an integration time of $t_{int} = 10^6$ leading to spectra with $\lambda_1 \approx 5 \cdot 10^{-2}$ while those exponents λ_2 and λ_3 to be expected to vanish reached values between 10^{-7} and 10^{-5} [for comparison in the case $E = 0.125$ see Bennettin et al. (1976)]. For regular orbits, identified as closed curves in the Poincaré's surfaces of section, we found that all Lyapunov exponents tended to zero as expected, *i.e.* their absolute values fell below 10^{-7}.

Concerning the interpretation of Lyapunov exponents derived from fitted ODE we would like to add the following remarks: The tables contain only the spectra $\boldsymbol{\lambda}$ for the generated trajectories and their associated parameters $\boldsymbol{p}_s$, and not those for the identified parameters $\tilde{\boldsymbol{p}}$ since for even slightly perturbed parameters $\tilde{\boldsymbol{p}}$, $\tilde{\lambda}_1$ differed by about 20 to 30% from λ_1. The reason is that slightly different initial values or parameters may lead to a significantly different energy surface. Similar to the strong dependence of the Lyapunov exponents on the distance d_0 between the

third object and the second primary in the restricted three body problem investigated by Gonczi and Froeschlé (1981), in our case we have a strong sensitivity of the Lyapunov exponents $w.r.t.$ energy E. For the Hénon-Heiles system this is in particular true when $E \approx 1/8$. We are forced to adopt the following point of view: The Lyapunov spectra derived from fitted ODE can be used to give a qualitative picture of ergodicity. They are still a quantitative tool to indicate chaos. In particular, this is important in dynamical systems with more than 2 degrees of freedom since in those cases the method of Poincaré's surfaces of section is inconvenient to use. And, finally, it is still an efficient method to determine the number of isolating integrals.

The major advantage, however, is that there is no explicit limitation as in a finite time series. Once the ODE have been identified the computation of the λ_i can be performed to any desired degree of accuracy. For some celestial mechanics problems, for instance that of the stochasticity of the orbit of Halley's comet (Froeschlé and Gonczi, 1988), there are typical integration times of 10^6 years required to reach convergence. In most cases, however, there is no a priori condition determining t_{int} available. Therefore the only chance is to integrate for a long period until convergence.

6. Conclusions

The *boundary value problem approach* recovers the parameters of the Hénon-Heiles system in a variety of scenarios differing in initial values and parameters of the simulated trajectories, noise of the Gaussian random errors and distance of the initial guesses of the parameters from the solution. From the identified dynamical systems the Lyapunov exponents are derived and are found to obey the conditions for Hamiltonian systems. In order to identify the parameters of the Hénon-Heiles system two components, $e.g.$ the x_1 and x_2 curves, prove to be sufficient, and even one component is sufficient if it contains enough data points. This is a very advantageous property, since in real experiments measurements, $e.g.$ of instance population densities in biological systems, concentrations in chemical and pharmacokinetical systems, or observations in astronomy or celestial mechanics seldom yield data for the whole phase space.

Appendix

A. Damping Strategies

Full step (Gauss-)Newton methods usually require unrealistically good initial guesses to guarantee convergence. To alleviate this problem step size strategies have been developed. According to an appropriately chosen level function the damping factor $\alpha_k \epsilon (\alpha_{min}, 1]$ is chosen in every iteration such that the value of the level function is reduced for the new iterate.

Depending on the choice of the level function which serves as a measure of the distance of the actual iterate to the solution different iterates are computed. For parameter estimation algorithms the following two level functions are important (Bock, 1981)

$$T_1(z) = \frac{1}{2} \cdot \|F_1(z)\|_2^2 + \sum_{i=1}^{n_c} \beta^i \cdot |F_2{}^i(z)| \qquad (A1)$$

where the weights β_i satisfy $\beta^i > |\lambda_c(z_k)|$ for all $(\Delta z_k, \lambda_c(z_k))$, $z_k \epsilon \mathbf{D} := \{z | T(z) \leq T(z_k)\}$. The vector $\lambda_c(z_k)$ is the Lagrange parameter of the problem linearized at z_k, Δz_k the solution of the linearized problem. The damping factor α_k is determined by

$$T_1(z_k + \alpha_k \cdot \Delta z_k) = \min_{\alpha_k} \qquad (A2)$$

It can be shown under mild conditions that then the Gauss-Newton-Method starting with an arbitrary initial guess converges to a stationary point.

Unfortunately, the practical benefits of this theoretically elegant result are limited: Even in mildly nonlinear ill-conditioned problems the direction of steepest descent of T_1 and the search direction Δz_k are nearly orthogonal. Intolerable small step sizes are computed.

As a remedy to that so called "natural level functions" (cf. Deuflhard, 1974)

$$T_n^k(z) = \|J^+(z_k)F(z)\|_2^2 \qquad (A3)$$

have been developed. These iteratively reweighted functions adapt to the local geometry of the problem. Together with special simplified line searches they allow significantly larger step sizes than by using the level function T_1, which are, in particular independent of the condition number. In the treatment of practical problems they have proven to be very effective.

B. Internal Numerical Differentiation

Consider the solution $y(t)$ of an IVP

$$\dot{y} = f(t, y, p), \quad y(t_0) = y_0 \qquad (B1)$$

as a function of the initial values and parameters $y(t) := Y(t, y_0, p)$. Provided differentiability of this function, the question of reliable and user friendly computation of $\partial Y / \partial y_0$ and $\partial Y / \partial p$ arises. In PARFIT an approximation of these quantities is generated by numerical differentiation of the approximating scheme $\hat{Y}(t, y_0, p)$ which is established by an integration procedure in the course of the computation of an approximation of $y(t) =: Y(t, y_0, p)$. $\hat{Y}$ is constituted by the chosen underlying integration method (extrapolation, Runge Kutta, multistep, etc.) and the step size and order strategies. The derivatives $\partial \hat{Y} / \partial y_0$ and $\partial \hat{Y} / \partial p$ are

computed by finite differences or by adjoint schemes. This requires no preparation of the user and is implemented in such a way that even discontinuous dynamics are tractable. Details and refined variants can be found in Bock (1987) and Schlöder (1988).

Acknowledgements

J. Kallrath gratefully acknowledges many encouraging discussions with V. Schulz and M. Steinbach (Institut für Mathematik, University of Augsburg) and their assistance concerning the application of PARFIT. Thanks are also directed to R. Dvorak (Institut für Astronomie, Vienna) for hospitality at Vienna Observatory and fruitful discussions on celestial mechanics, and to E. Baake (MPI-EB,Tübingen) for critical reading of the manuscript. We would also like to thank the careful referee (H. Eichhorn, Gainesville, University of Florida) for a couple of remarks and comments which helped to improve this paper. This work was sponsored by *Deutsche Forschungsgemeinschaft* under "Forschungsschwerpunkt Anwendungsbezogene Optimierung und Steuerung".

References

Abramowitz, M. and Stegun, I.A. : 1970, *Handbook of Mathematical Functions*, Dover Publication, 9th printing.

Baake, E., Baake, M., Bock, H.G. and Briggs, K.M. : 1992, 'Fitting Ordinary Differential Equations to Chaotic Data', *Phys.Rev.* **A45**.

Bennetin, G., Galgani, L. and Strelcyn, J.M.: 1976, 'Kolmogorov entropy and numerical experiments', *Phys.Rev.* **A14**, 2338-2345.

Bennetin, G., Galgani, L., Giorgilli, A. and Strelcyn, J.M.: 1980, 'Lyapunov Characteristic Exponents for Smooth Dynamical Systems — A Method for Computing all of them', *Meccania* (March), 9-20, 21-30.

Bock, H.G. : 1981. in: Ebert, K.H.,Deuflhard, P. & Jäger, W. (Eds.) *Modelling of Chemical Reaction Systems*, Springer Series in Chemical Physics, Springer, Heidelberg.

Bock, H.G. : 1987. 'Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen', *Bonner Mathematische Schriften* **183**, Bonn.

Bock, H.G. and Schlöder, J.P. : 1986, 'Recent progress in the development of algorithm and software for large-scale parameter estimation problems in chemical reaction systems', in: Kotobh, P.(Ed.) *Automatic Control in Petrol , Petrochemical and Desalination Industries*, IFAC Congress, Pergamon, Oxford.

Bock, H.G., Eich, E. and Schlöder, J.P. : 1988, 'Numerical Solution of Constrained Least Squares Boundary Value Problems in Differential Algebraic Equations', in: Strehmel(Ed.) *Numerical Treatment of Differential Equations*, BG Teubner, Leipzig.

Bulirsch, R. and Stoer, J. : 1966, 'Numerical Treatment of Ordinary Differential Equations by Extrapolations Methods', *Numerische Mathematik* **8**, 1–13.

Eckmann, J.-P. and Ruelle, D. : 1985, 'Ergodic Theory of chaos and strange attractors', *RevMod.Phys.* **57**, 617–657.

Eichhorn, H. : 1992, 'Generalized Least Squares Adjustment, a Timely but Much Neglected Tool', *Celestial Mechanics and Dynamical Astronomy*, this issue.

Deuflhard, P.: 1974, 'A modified Newton Methode for the Solution of Ill-conditioned Systems of Nonlinear Equations with Applications to Multiple Shooting', *Numerische Mathematik* **22**, 289–311.

Froeschlé, C. : 1984, 'The Lyapunov Characteristic Exponents — Applications to Celestial Mechanics', *Celest. Mech.* **34**, 95–115.

Froeschlé, C. and Gonczi, R. : 1988, 'On the Stochasticity of Halley like Comets', *Celest. Mech.* **43**, 325-330.

Gauss, C.F. : 1809, '*Theoria Motus Corporum Coelestium in Sectionibus Conicus Solem Ambientium*', Perthes, F. and Besser, J. H., Hamburg, [reprinted in Carl Friedrich Gauss Werke, Vol. VII, Königliche Gesellschaft der Wissenschaften zu Göttingen, Göttingen 1871.]

Gonczi, R. and Froeschlé, C. : 1981, 'The Lyapunov Characteristic Exponents as Indicators of Stochasticity in the Restricted Three-Body problem', *Celest. Mech.* **25**, 271-280.

Hénon, M. and Heiles, C. : 1964, 'The Applicability of the Third Integral of Motion: Some Numerical Experiments', *Astron.J.* **69**, 73-79.

Holzfuss, J. and Parlitz, U. : 1991, 'Lyapunov exponents from time series', in: Arnold, L., Cranel, H. and Eckmann, J.-P., *Lyapunov Exponents*, Proceedings, Oberwolfach 1990, 263–271.

Schlöder, J.P. : 1988, 'Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung', *Bonner Mathematische Schriften*, **187**, Bonn.

Wolf ,A., Swift, J.B.,Swinney, H.L. and Vastano, J.A. : 1985, 'Determining Lyapunov Exponents from a Time Series', *Physica* **16D**, 285–317.

Zeng, X., Eykholt, R. and Pielke, R.A. : 1991, 'Estimating the Lyapunov-Exponents from Short Time Series of Low Precision', *Physical Review Letters*, **66**, 3229–3232.