NONLINEAR DYNAMICAL SYSTEM IDENTIFICATION FROM UNCERTAIN AND INDIRECT MEASUREMENTS

HENNING U. VOSS and JENS TIMMER
Freiburg Center for Data Analysis and Modeling,
University of Freiburg, 79104 Freiburg, Germany

JÜRGEN KURTHS
Center for Dynamics of Complex Systems,
University of Potsdam, 14469 Potsdam, Germany

We review the problem of estimating parameters and unobserved trajectory components from noisy time series measurements of continuous nonlinear dynamical systems. It is first shown that in parameter estimation techniques that do not take the measurement errors explicitly into account, like regression approaches, noisy measurements can produce inaccurate parameter estimates. Another problem is that for chaotic systems the cost functions that have to be minimized to estimate states and parameters are so complex that common optimization routines may fail. We show that the inclusion of information about the time-continuous nature of the underlying trajectories can improve parameter estimation considerably. Two approaches, which take into account both the errors-in-variables problem and the problem of complex cost functions, are described in detail: shooting approaches and recursive estimation techniques. Both are demonstrated on numerical examples.

Keywords: System identification; multiple shooting algorithm; unscented Kalman filter; maximum likelihood.

1. Introduction

For a quantitative understanding and control of time-varying phenomena from nature and technology, it is necessary to relate the observed dynamical behavior to mathematical models. These models usually depend on a number of parameters whose values are unknown or only known imprecisely. Furthermore, often only a part of the system's dynamics can be measured. For example, in an electrophysiological experiment the voltage across a nerve cell membrane may be measured but not the ionic currents, which are not accessible but would also be needed for modeling the observed dynamics. As being part of a network, the neuron may also depend on concealed and thus unobserved inputs which need to be identified from the measured neuron dynamics. To duly apply the powerful principles of the theory of nonlinear dynamics in modeling such systems, a precedent system identification is inevitable. In the nerve cell example, not only unknown parameters have to be identified but also time-varying unobserved states. Similar problems arise in a variety of low-dimensional nonlinear dynamical systems, be they electronic circuits, mechanical devices, physiological rhythm generators, competing biological populations, nonlinear optical devices, chemical reactors, or even macroeconomic systems, among others.

We consider the problem of estimating parameters and unobserved trajectory components from noisy time series measurements of continuous

nonlinear dynamical systems. Rather than emphasize mathematical rigor, this tutorial aims at providing a comprehensible introduction to this field for the applied scientist, being from physics, biology, chemistry, or the engineering or economic sciences.

There is a vast literature on parameter estimation and system identification in general, so we restrict ourselves to a class of problems that are often encountered in applications: The *models* considered are assumed to be deterministic but in general nonlinear with continuous time dependence, such that their dynamics are described by systems of first-order ordinary differential equations. These models are assumed to be known; they result in general from first principle equations, but some or all of the parameters involved are assumed to be unknown. The *data* are assumed to be measurements on experimental setups or from the "real world", whereby it is assumed that the dynamical behavior of the system is correctly described by the model if only the correct parameter values are used. The measurements may be disturbed by a certain amount of *measurement noise.*

It is *not* assumed, however, that all system components are indeed measured; if the model consists of say, a D-dimensional system of first-order ordinary differential equations, then only less than D trajectory components may be experimentally accessible and measured.

Besides our interest in parameter estimation, we also want to estimate the unobserved trajectories at all time points of interest. To make the problem even more complicated, it is not assumed that the observed components are measured directly but we allow for smooth nonlinear *measurement functions.* For example, in terms of the previous example from electrophysiology, the measurement could be distorted by a nonlinearity in the signal amplifier.

This paper aims at solving such problems using mainly two different approaches:

We will not only describe the "standard" approach of parameter estimation in ordinary differential equations based on *shooting methods*, but also a less known approach based on *recursive estimation*, where parameters are estimated for each time point recursively proceeding through the time series. These two approaches are somewhat complementary; the former focuses on parameter estimation, whereby also the trajectories can be estimated. The latter focuses on the estimation of trajectories, whereby also parameters can be estimated. We will give a worked-out example to compare these two

approaches and will find that in this example the recursive approach can compete well with the standard approach, albeit being computationally much simpler.

The outline of this paper is as follows: In the next section the problem of parameter and state estimation is made more precise, and problems related to methods that are not based on the state space concept are highlighted. In Sec. 3 the errors-in-variables problem and its implications for parameter estimation is illustrated. The other main problem, complex cost functions, is described in Sec. 4. Then, two different approaches to solve these problems are worked out on examples: The multiple shooting approach (Sec. 5) and recursive estimation (Sec. 6). Further nonlinear filtering approaches are mentioned in Sec. 7. Finally, nonparametric modeling and recursive estimation in spatiotemporal systems is considered in Sec. 8.

Citations of methods are given in the text where they are mentioned first. Although we tried to provide an up-to-date bibliography, it is by no means complete. Nevertheless, we hope that it will provide a useful starting point for further reading.

2. The Problem of Parameter and Trajectory Estimation

Throughout the paper, only autonomous dynamical systems are considered. The parameter estimation methods further described will be straightforwardly also applicable to nonautonomous systems, but to keep notation simple, we suppress explicit time dependence. Further, if not stated otherwise, we focus on deterministic dynamics, and stochasticity enters only through the measurement errors.

We start with the description of the model system. The model is given by an autonomous system of D_x first-order differential equations with D_λ parameters,

$$\dot{x}_1(t) = f_1(x_1(t), x_2(t), \ldots, x_{D_x}(t), \lambda_1, \ldots, \lambda_{D_\lambda})$$
$$\dot{x}_2(t) = f_2(x_1(t), x_2(t), \ldots, x_{D_x}(t), \lambda_1, \ldots, \lambda_{D_\lambda})$$
$$\vdots$$
$$\dot{x}_{D_x}(t) = f_{D_x}(x_1(t), x_2(t), \ldots, x_{D_x}(t), \lambda_1, \ldots, \lambda_{D_\lambda}).$$
$$(1)$$

In vector notation, system (1) can be written as

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\lambda}). \qquad (2)$$

The parameter vector $\boldsymbol{\lambda}$ contains all parameters of both the system and the measurement function which will be introduced later.

For the rest of this paper the following situation is encountered: The states $\mathbf{x}(t)$ cannot be measured directly; what is observed is the quantity $\mathbf{y}(t)$ which is related to $\mathbf{x}(t)$ by some transformation due to a smooth measurement function plus some independent random errors $\boldsymbol{\eta}(t)$. Furthermore, the state is only sampled at discrete instants of time. For notational simplification, only uniform sampling is encountered, that is, the measurements are performed with a fixed sampling interval Δt at time $t, t + \Delta t, \ldots, t + (N-1)\Delta t$.

These assumptions cover a wide range of situations; they can be expressed in a denser form by introducing a *measurement equation*. It consists of a measurement function $\mathbf{G}: R^{D_x} \to R^{D_y}$ which maps the state vector $\mathbf{x}(t)$ to a quantity of equal or smaller dimension, and the measurement errors $\boldsymbol{\eta}_t$ which are assumed to be Gaussians with zero mean. They are independent for different times but need not have a constant covariance matrix over time. The measurement equation then becomes

$$\mathbf{y}_t = \mathbf{G}(\mathbf{x}_t, \boldsymbol{\lambda}) + \boldsymbol{\eta}_t , \qquad (3)$$

yielding a multivariate time series $\mathbf{y}_{t_i}$ $(i = 1, \ldots, N)$. From here on, subscripts denote discrete values, like sampling time.

The problem that will be encountered can be formulated as: Given a D_y-dimensional time series $\mathbf{y}_{t_i}$ $(i = 1, \ldots, N)$ and the model functions in Eqs. (2) and (3), we have to estimate the states $\mathbf{x}(t)$ for any time t of interest and the parameters $\boldsymbol{\lambda}$. This problem is referred to as a *dynamical system identification problem*.

There is one approach to solve this estimation problem which could instantly come to mind; it is based on embedding and least squares: Write system (1) as a scalar differential equation of higher order, if possible, and estimate the derivatives from the data sample. Since it is equivalent to a differential embedding of the time series [Packard *et al.*, 1980; Takens, 1981], this gives access to the unobserved system components or at least some transformations of them, which can be used in a least squares (LS) fit to estimate the parameters. There are several variants of the LS approach [Crutchfield & McNamara, 1987; Cremers & Hübler, 1987; Breeden & Hübler, 1990; Gouesbet, 1991b; Kadtke *et al.*, 1993; Aguirre & Billings, 1995; Hegger *et al.*, 1998], which have been shown to work for small amounts of noise and if the system of Eqs. (1) can be transformed into a single equation of higher order. Besides, this transformation may be quite involved and not easy to find [Gouesbet, 1991a, 1992], there are two main drawbacks:

(i)   Measurement noise prevents the accurate and even appropriate estimation of derivatives, especially of higher order. Filtering is not always a cure since filtering introduces temporal dependencies into the data and may disturb the data distribution. For example, think of a moving average window which involves also neighbored data points to smooth the data samples. These dependencies may then pretend dynamics which are not inherent in the data, leading to wrong results for the parameters [Timmer *et al.*, 2000].

(ii) The *errors-in-variables problem*. This problem stems from regression analysis and occurs if not only the dependent variables have errors but also the independent variables. If not appropriately taken into account, it leads to biased, i.e. wrong, parameter estimates. Biased parameters usually occur in connection with the naive application of a least squared estimation which is not suited to this problem. The errors-in-variables problem is unavoidable for the problem considered here, the estimation of parameters in dynamical systems with measurement noise.

3. State Space Modeling and the Errors-in-Variables Problem

The errors-in-variables problem is explained in more detail in the following, and the failure of the least squares approach is demonstrated. Also, the sometimes encountered cure by means of a total LS approach is discussed.

The errors-in-variables problem is known in statistics for a long time [Madansky, 1959]. Its significance in the context of nonlinear time series analysis has been introduced by [Kostelich, 1992]. To explain the problem by means of our system identification problem, we rewrite Eqs. (2) and (3) in discrete time [Gelb, 1974], to get the *discrete nonlinear deterministic state space model*

$$\mathbf{x}_{t+\Delta t} = \mathbf{F}(\mathbf{x}_t, \boldsymbol{\lambda}) , \qquad (4)$$

$$\mathbf{y}_{t+\Delta t} = \mathbf{G}(\mathbf{x}_{t+\Delta t}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_{t+\Delta t} . \qquad (5)$$

The time step Δt corresponds to the sampling time, and the function $\mathbf{F}$ is obtained from the function $\mathbf{f}$ in Eq. (2) via

$$\mathbf{F}(\mathbf{x}_t, \boldsymbol{\lambda}) = \mathbf{x}_t + \int_t^{t+\Delta t} \mathbf{f}(\mathbf{x}(T), \boldsymbol{\lambda})\, dT \,. \qquad (6)$$

The states described by Eqs. (4) and (6) obey the Markov condition, i.e. each state follows uniquely from its predecessor.

In this section we consider the linear state space model, but the errors-in-variables problem carries over to nonlinear models as well [Carroll *et al.*, 1995; McSharry & Smith, 1999]. A linear state space model has the form

$$\mathbf{x}_{t+\Delta t} = F(\boldsymbol{\lambda})\mathbf{x}_t + \boldsymbol{\varepsilon}_t \,, \qquad (7)$$

$$\mathbf{y}_{t+\Delta t} = G(\boldsymbol{\lambda})\mathbf{x}_{t+\Delta t} + \boldsymbol{\eta}_{t+\Delta t} \,. \qquad (8)$$

The system functions $\mathbf{F}$ and $\mathbf{G}$ are now described by $D_x \times D_x$ and $D_y \times D_x$-dimensional matrices F and G, respectively, multiplied with the states. Linear deterministic models have asymptotically, i.e. for long times, only trivial dynamics generically; the state components relax to the origin or diverge to infinity. Therefore, to establish this linear model with a nontrivial albeit asymptotically finite dynamics, the *process noise* $\boldsymbol{\varepsilon}_t$ is added in this example. It is defined as being Gaussian with zero mean and independent over time. Its covariance matrix is not obliged to be constant over time. The covariances of the process and measurement noise are denoted by Q_t and R_t, respectively, i.e.

$$Q_t := E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right] \quad \text{and} \quad R_t := E\left[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'\right] \,, \qquad (9)$$

where $E[\cdot]$ is the expectation value. Therefore, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, Q_t)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, R_t)$. The expression $\mathcal{N}(\boldsymbol{\mu}, B)$ denotes a vectorial Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix B.

Now a particularly simple case is considered [Timmer, 1998]: The state space model (7), (8) in one dimension with the trivial measurement function $Gx_t = x_t$ and constant noise variances Q and R:

$$x_{t+\Delta t} = Fx_t + \varepsilon_t \quad (|F| < 1) \qquad (10)$$

$$y_{t+\Delta t} = x_{t+\Delta t} + \eta_{t+\Delta t} \,. \qquad (11)$$

First assume that there is no measurement noise. Then the constant F can be estimated by multiply-

ing both sides of Eq. (10) with x_t. Averaging yields

$$\hat{F} = \frac{\sum\limits_t x_{t+\Delta t} x_t}{\sum\limits_t x_t^2} = \frac{\text{cov}\,(x_{t+\Delta t}, x_t)}{\text{var}\,(x_t)} \,, \qquad (12)$$

with "var" and "cov" being the variance and the covariance, respectively.

This is also the least squares estimate of F which would arise from minimizing the *cost function* $\chi^2(F)$:

$$\hat{F} = \arg\min_F \chi^2(F) \qquad (13)$$

$$= \arg\min_F \sum_t \frac{(x_{t+\Delta t} - Fx_t)^2}{Q} \,. \qquad (14)$$

It can be shown that the least squared estimator here is a *maximum likelihood estimator* (ML estimator). Maximum likelihood estimators of parameters are those which maximize the probability of the observed data given the model parameters. They are generally unbiased, i.e. they yield asymptotically the correct values for the parameters.

Now assume that there is a finite measurement noise with constant variance R, but the least squares estimator is used in a naive way. Proceeding as above, this results in

$$\hat{F}_{\text{naive}} = \frac{\text{cov}\,(y_{t+\Delta t}, y_t)}{\text{var}\,(y_t)} \,. \qquad (15)$$

Since $\text{var}\,(y_t) = \text{var}\,(x_t) + R$ and $\text{cov}\,(y_{t+\Delta t}, y_t) = F\text{var}\,(x_t)$, one gets

$$\hat{F}_{\text{naive}} = F\frac{\text{var}\,(x_t)}{\text{var}\,(x_t) + R} \,, \qquad (16)$$

from which it follows that

$$|\hat{F}_{\text{naive}}| < |F| \,. \qquad (17)$$

The larger the measurement variance R, the more the naive estimator is biased. This bias is independent of the size of the data sample and hence cannot be compensated by measuring more data. Independent of the true value of F, for large noise the naive estimate asymptotically vanishes.

How can one understand this? The parameter F was estimated using a regression approach, which solves a problem of the form: Given a *dependent* variable which can be written as a linear function of an *independent* variable plus some noise, estimate the parameters of the linear function. In this setting it is assumed that only the dependent variable is uncertain but not the independent variable. In time series analysis problems, like Eqs. (10) and
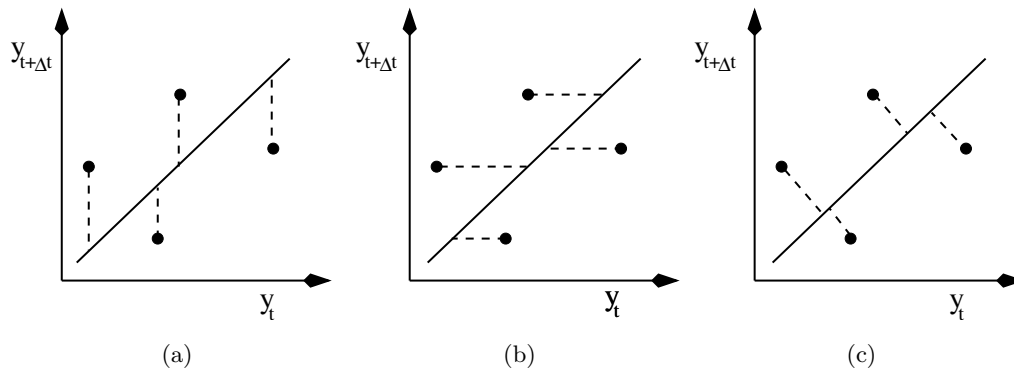
Fig. 1. The LS approach estimates the regression line by minimizing the mean squared distance between the measurements $y_{t+\Delta t}$ and y_t. The latter are wrongly taken to be given exactly (a). For uncertain y_t but certain $y_{t+\Delta t}$ the LS method works only if the meaning of y_t and $y_{t+\Delta t}$ is exchanged (b), whereas for uncertain y_t and $y_{t+\Delta t}$ the orthogonal distance method treats both variables in the same way (c).

(11), the data points are usually measurements of both dependent variables, e.g. at time $t + \Delta t$, and independent variables, e.g. at time t.

Therefore, in time series analysis the independent variables become uncertain as well, giving rise to the errors-in-variables problem. In other words, the reason for the bias is that the naive estimator is not a ML estimate any more; for a proper ML estimate the probability of the data given the model should also depend on the uncertainty of the independent variables, here x_t, rather than only on the uncertainty of the dependent variables, here $x_{t+\Delta t}$.

In this particular example, the naive estimator can still be used to correctly yield F in a ML sense; the known bias can be used to correct the wrong estimate by using Eq. (16) and the fact that in this model the variance of x_t is known to be var $Q/(1 - F^2)$. However, it stands to reason that the search for such correction formulas becomes rather cumbersome for more general models [Fuller, 1987; Seber & Wild, 1989; Jaeger & Kantz, 1996], and in our nonlinear model (4), (5) it should hardly be possible generally to estimate the system functions by correcting a least squares fit to the measurements. Note that only the observational noise causes trouble but not the system noise in Eq. (7). Therefore, the errors-in-variables problem is also significant in the determinsitic state space model (4), (5).

To summarize this result, we see that all problems have arisen because the values of x_t are not given but are uncertain themselves; a least squares approach does not take the distribution of x_t into account but takes the x_t to be given exactly. In other words, in the LS approach the regression coefficient F is determined such as to minimize the

mean squared distance between the measurements $y_{t+\Delta t}$ and y_t [Fig. 1(a)].

A tutorial of the errors-in-variables problem with a worked-out solution for the Hénon map is provided by Jaeger and Kantz [1996]. They also show how this problem affects other kinds of nonlinear estimation problems like the estimation of invariant quantities of dynamical systems.

A more general approach is the method of *total least squares* (TLS) [Jeffreys, 1980, 1988; Boggs *et al.*, 1987; van Huffel & Vandewalle, 1991]. This method treats both $y_{t+\Delta t}$ and y_t in the same way by estimating F such as to minimize the orthogonal distances between the regression line and the data [Fig. 1(c)]. This seems to be sensible, since this is the intermediate case for certain y_t and uncertain $y_{t+\Delta t}$ on the one hand and certain $y_{t+\Delta t}$ and uncertain y_t on the other hand [Fig. 1(b)]. One could also say that the LS method estimates the true values of $y_{t+\Delta t}$ via $x_{t+\Delta t} = Fy_t$, and the orthogonal distance method estimates the true values of $y_{t+\Delta t}$ and y_t. In this sense the orthogonal distance method copes with the errors-in-variables problem, but it is still not optimal, and, therefore, not a ML estimate: In its simplest form, it completely ignores information about the distributions of $y_{t+\Delta t}$ and y_t. Furthermore, the TLS approach leads to significantly larger estimation errors. Consequently, it has been shown that the TLS method is of only rather limited use for nonlinear time-discrete estimation problems [Kostelich, 2001; McSharry & Smith, 1999]. For these reasons we do not adhere to the orthogonal distance method here.

What we should keep in mind from this lesson is the following: The TLS method works by

estimating not only the unobserved values of $x_{t+\Delta t}$ from the measurements, but also the values of x_t. Albeit being not optimal, this is the right direction to proceed: The errors-in-variables problem can only be solved if the values of the independent variables x_t are also estimated from the data in addition to the values of the dependent variables $x_{t+\Delta t}$. This approach will be revived later on, when

in addition to the estimation of parameters of a dynamical system the unobserved trajectories will also be estimated (Secs. 5 and 6).

4. The Problem of Complex Cost Functions

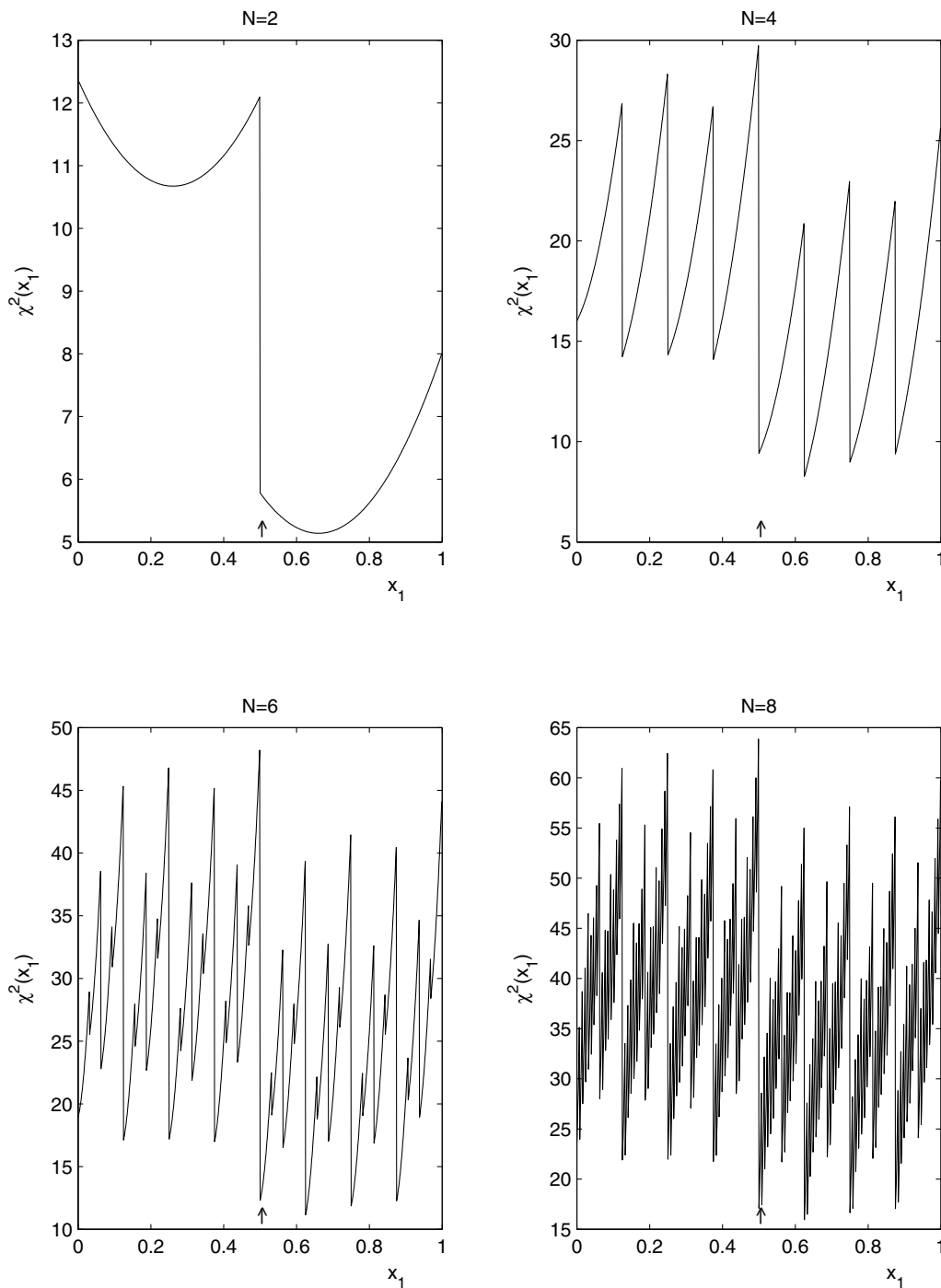The errors-in-variables problem is not the only fact

Fig. 2. The cost function of the shift map (18) with observational noise for different numbers of measurements N. The true initial value of the sequence is marked by an arrow.
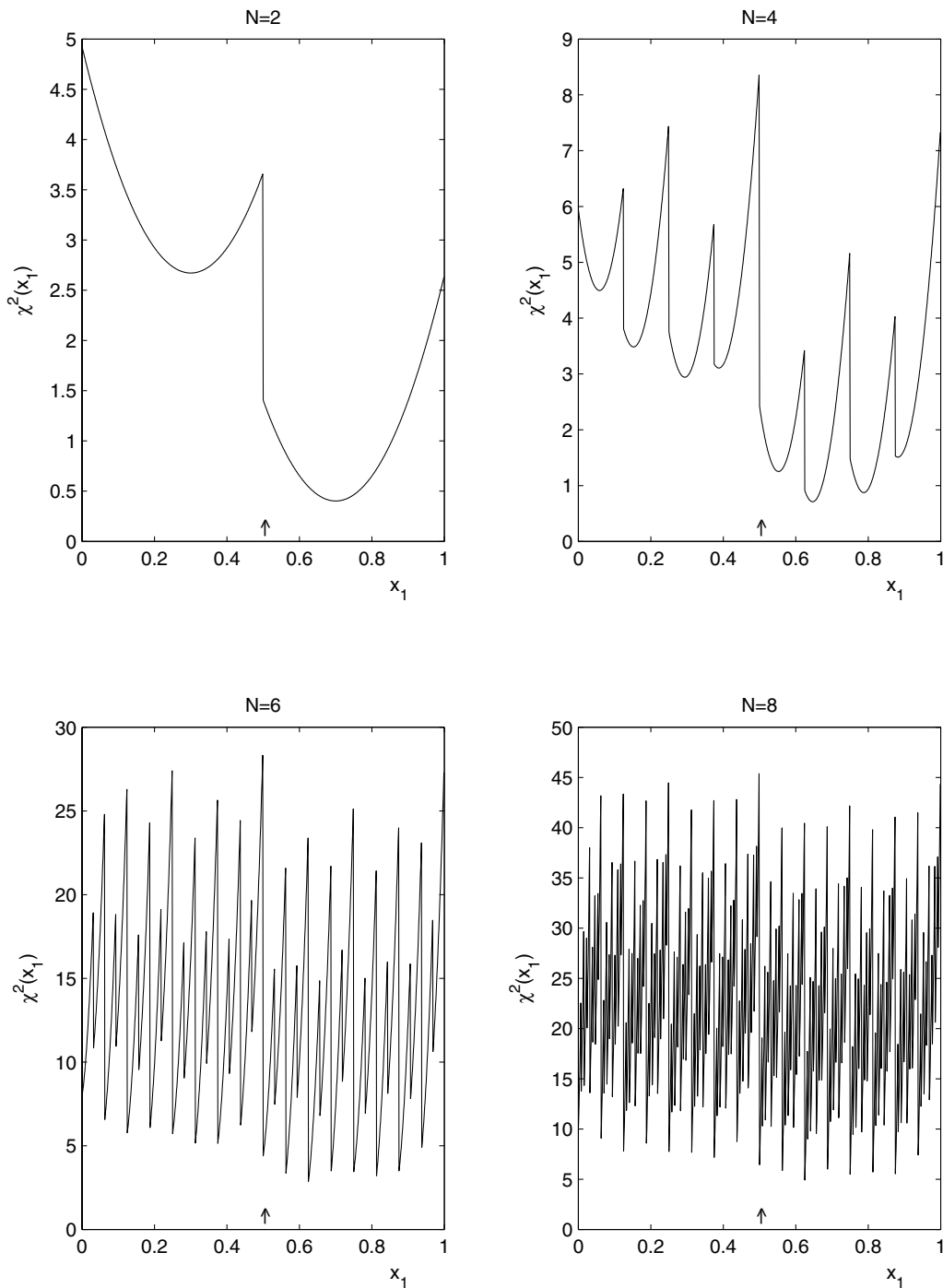
Fig. 3. The same as Fig. 2 but with a different noise realization.

that prevents an easy nonlinear dynamical system identification. As will be shown now, another problem arises, especially for chaotic dynamics: The cost function may become so complex that a numerical minimization by standard optimization methods fails. Whereas the former problem is of a genuinely statistical nature, the latter one seems to be more of technical nature. However, as it will turn out,

this is only part of the story, because the cost functions for data from chaotic dynamical systems show anomalous scaling behavior.

The problem of complex cost functions is demonstrated on a simple example. Consider the nonlinear state space model (4), (5) with the shift map

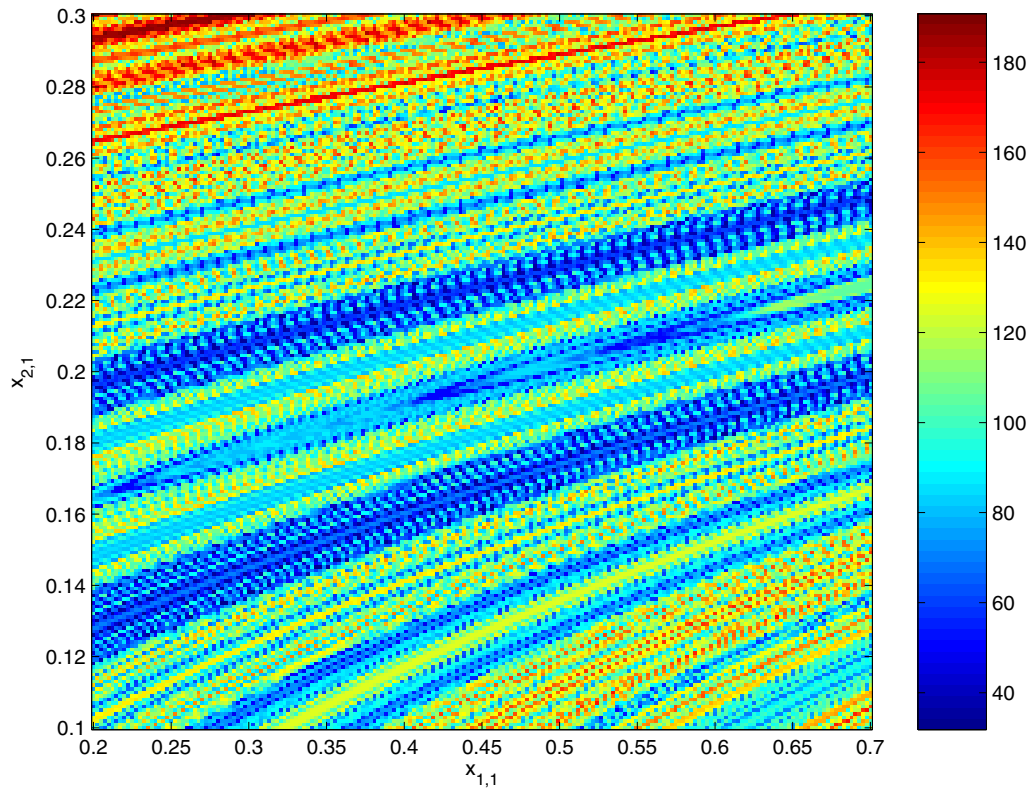$$x_{t+1} = f(x_t), \quad f(x_t) = 2x_t \bmod 1, \quad (18)$$

Fig. 4.   Clip of the cost function of the estimate of an initial state vector of the Hénon map.

which maps the unit interval onto itself. The shift map is chaotic and serves as a theoretic model to understand also dynamical aspects of the more general cases of two-dimensional maps and autonomous vector fields of the form of Eq. (2) [Wiggins, 1990].

Rather than estimating parameters, to keep the problem simple we treat the initial value x_1 as the quantity that should be estimated. The initial value x_1 can also be seen as a parameter controlling the deterministic sequence of x_t for $t > 1$. If the states x_t, including the initial state, are disturbed by Gaussian noise η_t with constant variance R and N measurements

$$y_t = x_t + \eta_t \tag{19}$$

are made, the cost function for the estimation of the initial state at time $t = 1$ is

$$\chi^2(x_1) = \sum_{t=1}^{N} \frac{(y_t - x_t)^2}{R} . \tag{20}$$

Two cost functions for different noise realizations and $x_1 = 0.501$ and $R = 0.2$ are shown in Figs. 2 and 3. One observes the following:

(i) The cost functions become rather complex rapidly. Optimization methods could fail

already for as small a number of measurements as 4, since there are many local minima with a similar value of the cost function. Similar observations have been made by [Miller *et al.*, 1994] for the Lorenz oscillator.

(ii) For the two noise realizations the cost functions are structurally rather different for as few as six data points.

(iii) Still for eight measurements the cost functions point to a wrong estimate in both cases.

(iv) One can guess from the figures that the estimation error does not decrease by a $1/\sqrt{N}$-scheme, as it holds for nonlinear regression. The reason is the strong dependency in the data caused by the deterministic origin of the dynamics. Similar anomalous scaling behavior has also been reported for dimension estimates [Theiler, 1990] and for Lyapunov exponent estimates [Theiler & Smith, 1995].

Matters become more complicated for a cost function that depends on more than one variable. One could think of the estimation of a parameter in addition to the initial state. Rather than considering this case, the problem should be illustrated by computing the cost function for the initial vector of

the two-dimensional Hénon map. This map is given by

$$x_{1,t+1} = 1 - 1.4\, x_{1,t}^2 + x_{2,t}$$
$$x_{2,t+1} = 0.3\, x_{1,t}\,. \tag{21}$$

With the measurements

$$y_{1,t} = x_{1,t} + \varepsilon_{1,t}$$
$$y_{2,t} = x_{2,t} + \varepsilon_{2,t}\,, \tag{22}$$

the cost function is defined to be

$$\chi^2(x_{1,1}, x_{2,1}) = \sum_{t=1}^{N} \left[\frac{(y_{1,t} - x_{1,t})^2}{R_1} + \frac{(y_{2,t} - x_{2,t})^2}{R_2} \right]. \tag{23}$$

The true initial value is $(x_{1,1}, x_{2,1}) = (0.51, 0.18)$, and $N = 20$ measurements are taken with a constant noise variance of $R_1 = R_2 = 0.2$. From this example another observation can be made (Fig. 4):

(v) There are long separated "valleys" where the cost function locally takes on similar minimal values. This points to the fact that the two estimates are highly dependent on each other. Therefore, it seems that the true trajectory cannot be estimated only by searching for its initial value, but in principle each of the $(x_{1,t}, x_{2,t})$ $(t = 1, \ldots, N)$ should be estimated.

5. The Multiple Shooting Approach

The problem of local minima in the cost function will be treated by means of two different approaches: multiple shooting and recursive estimation. Multiple shooting is an ML approach and will, therefore, also solve the errors-in-variables problem. For weakly nonlinear systems, recursive estimation is an ML approach as well.

If a model like Eqs. (2) and (3) is known, however, the data must not be treated as independent over time, and one should for a ML estimation also estimate $\mathbf{x}(t)$ in addition to the parameters $\boldsymbol{\lambda}$. This will be done below by utilizing the temporal dependencies as induced by the model (2). In other words, the information that the system produces a continuous trajectory is incorporated.

5.1. *The initial value approach*

The system (2) together with an initial value $\mathbf{x}_{t_1}$ and a known parameter vector $\boldsymbol{\lambda}$ constitutes an initial value problem: Find a trajectory $\mathbf{x}(t)$ that for

given $\boldsymbol{\lambda}$ solves this system for all times of interest t. This is the direct problem, in contrast to the inverse problem we are interested in; we try to calculate those trajectories and parameters that are the most likely ones, given measurements at discrete times. To keep things simple, in this chapter we restrict ourselves to scalar measurements, i.e. the measurement function (3) becomes

$$y_t = G(\mathbf{x}_t, \boldsymbol{\lambda}) + \eta_t\,. \tag{24}$$

In the *initial value approach* [Schittkowski, 1994] initial guesses for $\mathbf{x}_{t_1}$ and $\boldsymbol{\lambda}$ are chosen and the dynamical equations are solved numerically. The likelihood is given as the probability distribution ρ of the data given the states and parameters, and the ML estimator becomes

$$(\hat{\mathbf{x}}_{t_1}, \hat{\boldsymbol{\lambda}}) = \arg \max_{\mathbf{x}_{t_1}, \lambda} \rho[\{y_{t_i}\} | \mathbf{x}_{t_1}, \boldsymbol{\lambda}]\,. \tag{25}$$

It is asymptotically, i.e. for an infinite number of data samples, unbiased. In the case of independent Gaussian measurement noise, maximizing the likelihood amounts to the minimization of the cost function

$$\chi^2(\mathbf{x}_{t_1}, \boldsymbol{\lambda}) = \sum_{i=1}^{N} \frac{(y_{t_i} - G(\mathbf{x}_{t_i}(\mathbf{x}_{t_1}, \boldsymbol{\lambda}), \boldsymbol{\lambda}))^2}{R_{t_i}}\,. \tag{26}$$

It is the sum of squared residuals between the data and the model trajectory, weighted with the inverse variances of the measurement noise. The parameters are identified as those minimizing $\chi^2(\mathbf{x}_{t_1}, \boldsymbol{\lambda})$.

Often prior knowledge about the parameters can be formulated as equality or inequality constraints. For instance, "rate constants" are usually non-negative. In the context of the multiple shooting approach described later on, the continuity of the true trajectory is ensured by means of equality constraints.

The initial value approach is an iterative procedure. Thus, an initial guess is updated again and again unless some convergence criterion is met. The update step is usually computed from the gradient or the Hessian, leading to different optimization methods: the steepest descent method [Bryson & Ho, 1969], the conjugate gradient method [Fletcher & Reeves, 1964], the Levenberg–Marquardt algorithm [Stortelder, 1996], and the Gauss–Newton algorithm [Schittkowski, 1995], among others. Overviews of approaches for iterative solutions of nonlinear optimization problems are given in several textbooks [Gill *et al.*, 1981; Stoer & Bulirsch, 1993; Press *et al.*, 1997]. To make

the optimization task effective, at least first derivatives of the system equations with respect to the parameters and trajectory initial conditions, the *sensitivities*, should be provided.

Simulation studies have shown that for many types of dynamics this approach is numerically unstable by yielding a diverging trajectory or stopping in a local rather than in the global minimum [Horbelt, 2001]. The reason is that for slightly wrong parameters the estimated trajectory may deviate from the true trajectory after a short time. This is most evident in the case of chaotic dynamics, where due to the sensitivity with respect to initial conditions the estimated trajectory is expected to follow the true trajectory of the system only for a limited amount of time [Lorenz, 1963; Ott *et al.*, 1994; Wiggins, 1990; Nicolis, 1995; Alligood *et al.*, 1997]. This is related to the well-known shadowing problem of chaotic dynamics [Hammel *et al.*, 1988; Nusse & Yorke, 1989; Farmer & Sidorowich, 1990]. The divergence of the numerical and measured trajectory introduces many local minima in the cost function (26), as described above in Sec. 4.

5.2. *Multiple shooting*

A possible solution to the above mentioned problems is the multiple shooting algorithm introduced by [van Domselaar & Hemker, 1975] in the context of parameter estimation, and further developed by [Bock, 1983; Baake *et al.*, 1992]. It has been applied, for example, to plant physiology [Baake & Schlöder, 1992], chemical reaction kinetics [Bock, 1981], nonlinear optics [Horbelt *et al.*, 2001], and electronic circuits [Timmer *et al.*, 2000; Horbelt *et al.*, 2002].

The motivation to use multiple shooting rather than the initial-value approach, which is a single-shooting method, is a more efficient minimization of the cost function without getting lost too easily in local minima. In multiple shooting, initial conditions are estimated at several points along the time series, the *shooting nodes*, and thus the estimated trajectory can be forced to stay closer to the true one for a longer time. Mathematically, the task is considered as a multi-point boundary value problem [Stoer & Bulirsch, 1993]. It is based on a shooting method to solve two-point boundary value problems [Bellman & Kalaba, 1965; Bryson & Ho, 1969]. The fitting interval $[t_1, t_N]$ is partitioned into m subintervals,

$$t_1 = \tau_1 < \tau_2 < \cdots < \tau_{m+1} = t_N. \qquad (27)$$

For each subinterval $[\tau_j, \tau_{j+1}]$, *local* initial values $\mathbf{x}_j = \mathbf{x}(\tau_j)$ are introduced as additional parameters. The dynamical equation is integrated piecewise (Fig. 5) and the cost function $\chi^2(\mathbf{x}_1, \ldots, \mathbf{x}_m, \boldsymbol{\lambda})$ is evaluated and minimized as in the initial value approach. Whereas the dynamical parameters $\boldsymbol{\lambda}$ are constant over the entire interval, the local initial values are optimized separately in each subinterval.

This approach leads to an initially discontinuous trajectory, which is, however, close to the true one. The final trajectory should be continuous, i.e. the computed solution at the end of one subinterval must match the local initial values of the next one. These continuity conditions are linearized in each iteration of the procedure and are then taken into account as equality constraints [Stoer, 1971]. The particular structure of their linearized form permits the continuity constraints and the extra variables $\mathbf{x}_j$ ($j = 1, \ldots, m$) to be eliminated easily from the resulting large linear system. In this way the dimension of the system of equations to be solved in each iteration is no larger than with the initial-value approach. This procedure, called *condensation* [Bock, 1983], can be applied to many kinds of optimization strategies. Since only the linearized continuity constraints are imposed on the update step, the iteration is allowed to proceed to the final continuous solution through "forbidden ground": the iterates will generally be discontinuous trajectories. This freedom allows the method to stay close to the observed data, prevents divergence of the numerical solution and reduces the problem of local minima.

Here a generalized Gauss–Newton algorithm is used [Hanson & Haskell, 1982], and both the model nonlinearities and the continuity constraints are linearized in each iteration. Their first order Taylor expansions make up a large linear optimization problem with linear equality constraints. For the following reasons, in some cases it may become necessary to omit the continuity constraints at some shooting nodes: For time series of chaotic systems it will not be possible in general to find a trajectory that follows the true trajectory for arbitrary long times. Furthermore, if the model is not correct, the Gauss–Newton method may not converge if the whole time interval $[t_1, t_N]$ of measurement is taken as input data [Horbelt, 2001].

After convergence, confidence intervals for the parameters can be calculated from the curvature of the logarithm of the likelihood [Shumway & Stoffer, 2000]: If the model has been specified correctly,
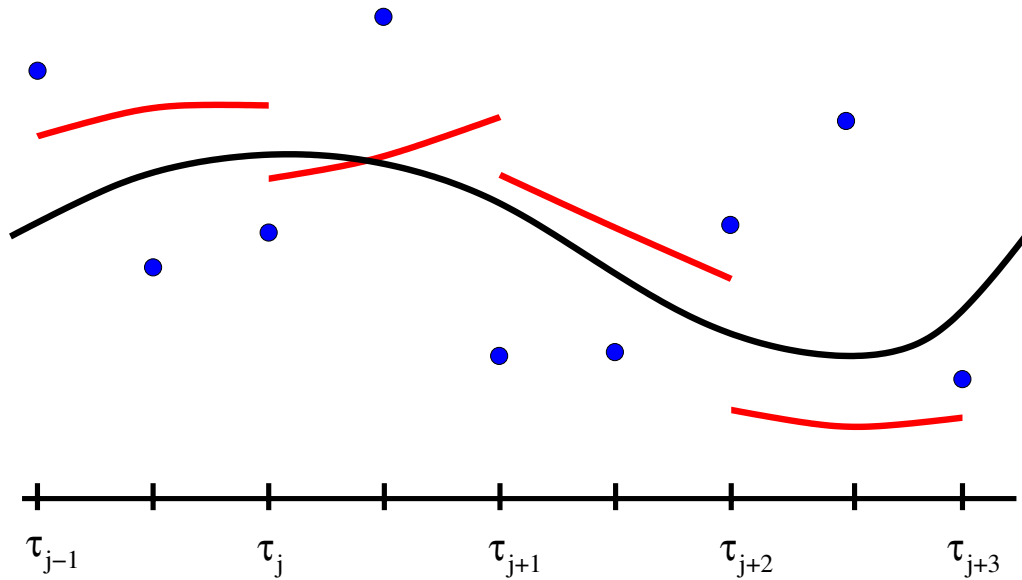
Fig. 5.   Schematic illustration of the multiple shooting approach. Black: true trajectory; blue dots: measurements; red: multiple shooting trajectories.

the global minimum of $\chi^2(\boldsymbol{\lambda})$ is found, and its quadratic approximation holds in a large enough region around the solution point, then the covariance matrix of the parameter estimate is given by the inverse Hesse matrix of the objective function:

$$(\Sigma)_{ij} = \left(\frac{1}{2} \frac{\partial^2 \chi^2(\boldsymbol{\lambda})}{\partial \lambda_i \partial \lambda_j} \right)^{-1} . \tag{28}$$

An eigenvalue analysis of Σ reveals parameters or linear combinations of parameters that are not identifiable from the given data. This happens,

for instance, when parameters or states cannot be estimated unambiguously, for example if the model is invariant under a transformation that alters the state vector and the parameters, but not the measurements.

A good correspondence between the estimated trajectory and the true trajectory has little significance if it has been achieved by adjusting a large number of free parameters. *Overfitting*, i.e. when the number of degrees of freedom is not in a reasonable relation to the amount of information
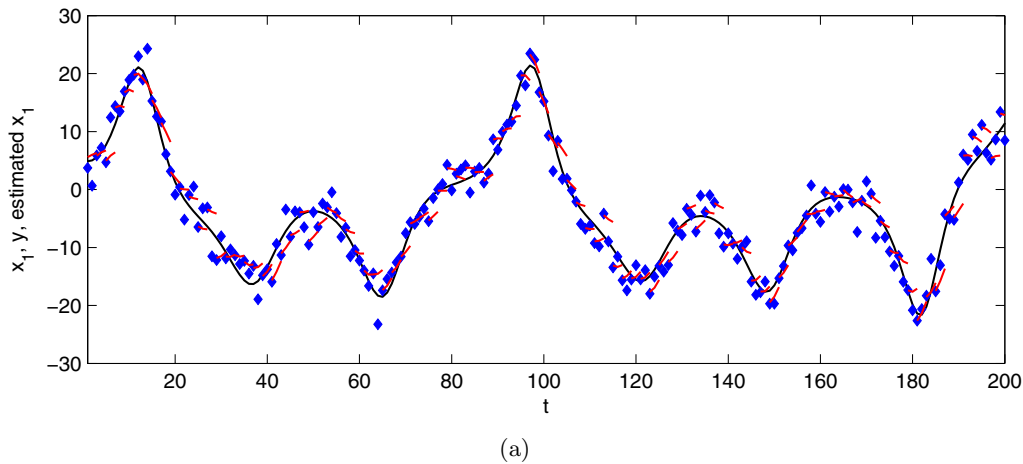

(a)

Fig. 6.   Identification of the Lorenz model (29) with the multiple shooting approach. The states and parameters are estimated by observing noisy data from the first component of $\mathbf{x}$, x_1. Measured data set (blue dots), true trajectory (black), and multiple shooting trajectories (red). For better visibility, only the first 200 data points of the 400 data points time series are shown. (a) after three iterations, (b) after ten iterations, (c) after 20 iterations. This is the final trajectory estimate.
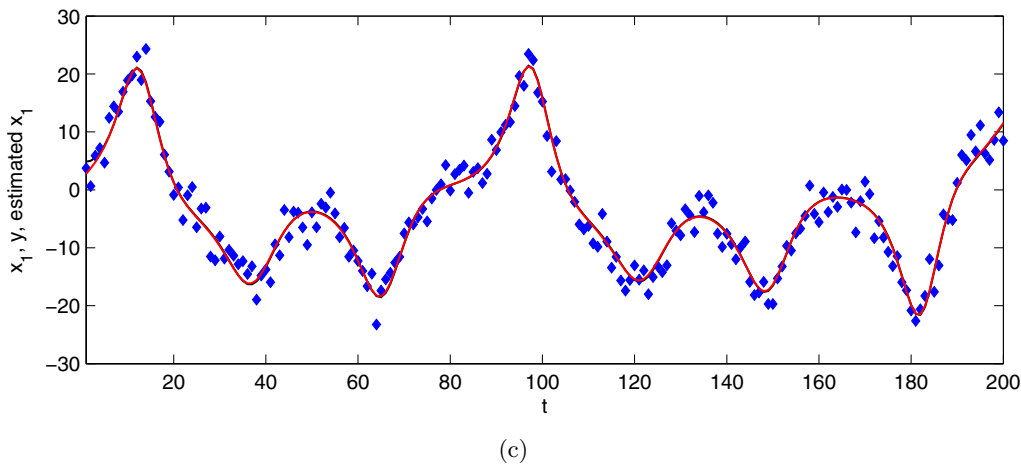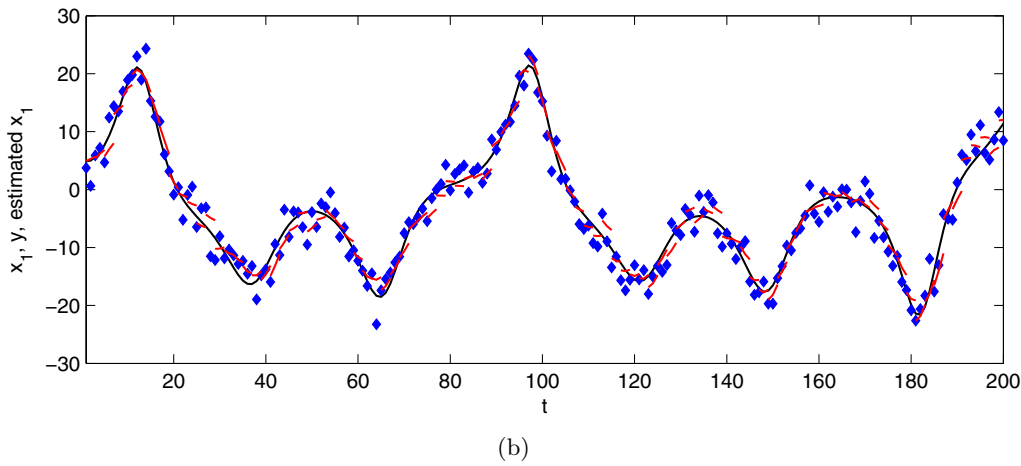
(b)

(c)

Fig. 6   (*Continued*)

contained in the data, can be detected by disproportionately large confidence intervals.

The problem of finding useful initial guesses for the parameters is not fully solved yet. Recently, a hybrid approach between parametric and nonparametric estimation has been proposed to solve this task for a rather large class of continuous time models [Timmer *et al.*, 2000]. See also Sec. 8.

5.3.   *An example*

The multiple shooting method is applied to a noisy time series of the chaotic Lorenz model [Lorenz, 1963]

$$\dot{x}_1 = -\lambda_1 x_1 + \lambda_1 x_2 \,,$$
$$\dot{x}_2 = \lambda_2 x_1 - x_2 - x_1 x_3 \,, \qquad (29)$$
$$\dot{x}_3 = -\lambda_3 x_3 + x_1 x_2 \,,$$

with the parameter vector $\boldsymbol{\lambda} = (10, 46, 2.67)$.

The measurements are constructed by corrupting the component $x_1(t)$ with Gaussian measurement noise $\eta_t \sim \mathcal{N}(0, R)$ at 400 equidistant ($\Delta t = 0.02$) sample times. The standard deviation of η, $\sqrt{R}$, is 25% of the standard deviation of the x_1-component of the system. We choose the initial parameter vector $\boldsymbol{\lambda}/2$ and the initial values of the 200 multiple shooting intervals equal to 0.9 times the true values.

The convergence of the algorithm is shown in Figs. 6(a)–6(c). It results in $\chi^2 = 353$ and a parameter vector estimate of $\hat{\boldsymbol{\lambda}} = (9.75 \pm 0.26, 47.2 \pm 1.0, 2.58 \pm 0.07)$ (1σ confidence intervals).

A recent overview of other related nonrecursive system identification approaches has been given in [Edsberg & Wedin, 1995], citing most of the relevant literature and providing a Matlab toolbox.

6.   Recursive Parameter Estimation

In the previous chapter the information contained in the temporal order of the data samples was exploited by putting a global constraint onto the estimation procedure for the trajectories. The reason

was to force the estimated trajectories to be continuous in time, since only then they are in compliance with the model. An alternative method for estimating underlying trajectories is to proceed recursively through the data sample and to obey the continuity constraint at each time step. The system trajectory is then estimated recursively by a prediction-correction method: Predict the value of the trajectory at time $t + \Delta t$ from its value at time t, and correct this estimate taking the new measurement at time $t + \Delta t$ into account. This way, one can proceed through the data sample until the last data point is reached. There is a bunch of recursive estimation methods which are, however, all based on this prediction-correction scheme, and in case of Gaussian system and measurement noise all these amount into a *recursive least squares algorithm* [Ljung & Söderström, 1983]. In this setup, parameters can be estimated simultaneously by augmenting the system's state by the parameter vector.

We first describe the well-known Kalman and extended Kalman filters for stochastic systems and will see that the covariances of the state components are essential for the calculation of the prediction-correction scheme. Therefore, after deriving the Kalman filter in its familiar form for linear stochastic systems, we will also state it in writing using *fully nonlinear* and *deterministic* system functions. In applications to nonlinear deterministic systems, the covariance matrices have to be evaluated numerically in general, as opposed to the linear case where they can be explicitly computed from the matrices in the state space model.

The application of Kalman filtering to nonlinear deterministic systems will be worked out for the same system as in the previous chapter. We will concentrate on the particular approach of "unscented Kalman filtering", which is nowadays preferred more due to its superiority to the conventional approach of extended Kalman filtering. See also [Sitz *et al.*, 2002].

A Matlab routine for parameter estimation with the unscented Kalman filter is available from the authors or via *www.fdm.uni-freiburg.de/∼hv*.

6.1. *The Kalman filter*

The *Kalman filter* (KF) [Kalman, 1960; Kalman & Bucy, 1960] is one of the most powerful signal analysis tools. The reason is that for linear systems given in a state space formulation the KF is an ML estimator and thus an optimal estimator for unobserved system trajectories. To be more specific, "optimal" means that the KF yields the most probable, but not necessarily a typical trajectory, since it has some low-pass properties. For weakly nonlinear systems the *extended Kalman filter* (EKF) [Anderson & Moore, 1979] can be used. It is based on a Taylor-series expansion of the system nonlinearities. For higher order accuracy it may consist of rather complicated expressions. Therefore, in most cases the EKF of first order is chosen, which is based on a local linearization of the system functions. In this case the Gaussianity of the state variables is maintained, which is a prerequisite for the optimal functioning of the KF. Tutorials and textbooks about Kalman filtering, including applications, are [Jazwinski, 1970; Gelb, 1974; Bar-Shalom & Fortmann, 1988; Aoki, 1990; Hamilton, 1994; Mendel, 1995; Gershenfeld, 1999; Grewal & Andrews, 2001; Honerkamp, 2002; Walker, 2002].

Some novel approaches of nonlinear Kalman filtering are not based on a Taylor expansion of the system functions any more but approximate the density of the state by Gaussians, retaining thereby the whole system nonlinearities. Since these novel approaches are easier to implement and more accurate than the EKF of first order, we will use them in the following. However, to better understand how nonlinear Kalman filtering works, it is necessary to focus on the linear case first. Already for the linear case, the KF equations will also be written in a more flexible way than usual, using nonapproximated system functions. This is the basis for the straightforward application of the novel methods to our particular system identification problem in the subsequent section.

The KF is a recursive estimator with a prediction-correction structure: We start with the nonlinear state space model (4), (5) and focus our attention on the time point $t + \Delta t$, the reference time. The state estimate at time $t + \Delta t$, given all state estimates $\dots, \hat{\mathbf{x}}_{t-\Delta t}, \hat{\mathbf{x}}_t$ and all measurements $\dots, \mathbf{y}_t, \mathbf{y}_{t+\Delta t}$ until then, is denoted by $\hat{\mathbf{x}}_{t+\Delta t|t+\Delta t}$. This is the *a posteriori*, or complete, estimate which utilizes all information taking the new measurement into account. To save indices, the *a posteriori* estimate at the reference time point $t + \Delta t$ in the following will be written as

$$\hat{\mathbf{x}} := \hat{\mathbf{x}}_{t+\Delta t|t+\Delta t} \,. \qquad (30)$$
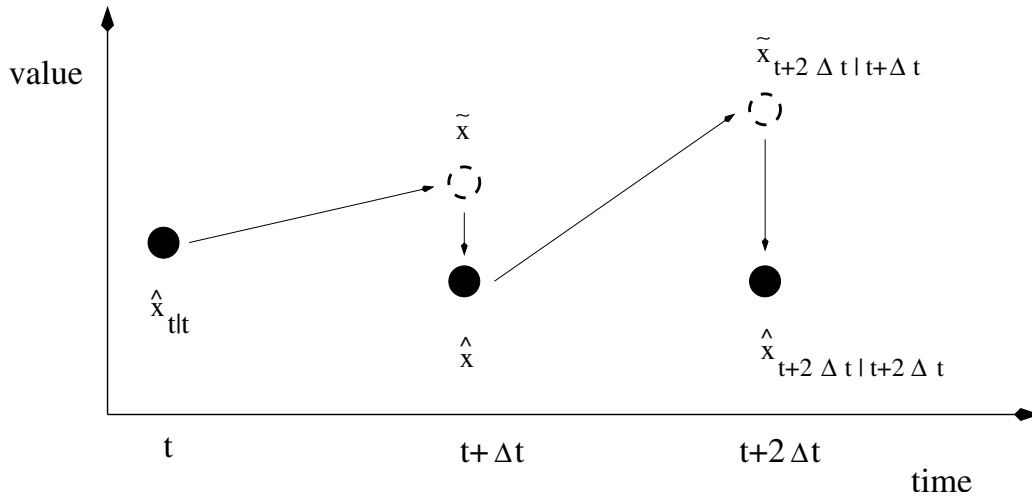
Fig. 7. Timing diagram of the Kalman filter. The indices at the reference time $t + \Delta t$ are suppressed, as in the text.

Equivalently, the *a priori*, or preliminary, state estimate at time $t + \Delta t$ again utilizes all previous state estimates but only measurements $\ldots, \mathbf{y}_{t-\Delta t}, \mathbf{y}_t$ up to one time step before the reference time $t + \Delta t$. It will be denoted by $\tilde{\mathbf{x}}_{t+\Delta t | t}$, or simply

$$\tilde{\mathbf{x}} := \tilde{\mathbf{x}}_{t+\Delta t | t} \,. \tag{31}$$

These quantities are illustrated in Fig. 7.

The KF estimation procedure of the state at time $t + \Delta t$ consists of two parts:

(i) an *a priori* estimate of the state $\tilde{\mathbf{x}}$ prior to observation of the new measurement $\mathbf{y}_{t+\Delta t}$, and

(ii) a correction after observation of the new measurement $\mathbf{y}_{t+\Delta t}$. It is proportional to the difference of the new measurement $\mathbf{y}_{t+\Delta t}$ and the predicted value of the new measurement $\tilde{\mathbf{y}}$:

$$\hat{\mathbf{x}}_{t+\Delta t | t+\Delta t} = \tilde{\mathbf{x}}_{t+\Delta t | t} + K_{t+\Delta t}(\mathbf{y}_{t+\Delta t} - \tilde{\mathbf{y}}_{t+\Delta t | t}). \tag{32}$$

The $D_x \times D_y$-dimensional *Kalman gain matrix* K is defined below. Dropping the time index of the reference time $t + \Delta t$, and using the definitions above yields simply

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} + K(\mathbf{y} - \tilde{\mathbf{y}}) \,. \tag{33}$$

Therefore, the Kalman estimate of the state is recursively given by a prediction-correction structure. The prediction uncertainty is given by the covariance matrix of the state estimate,

$$\hat{P}_{xx} = \tilde{P}_{xx} - K\tilde{P}'_{xy} \,. \tag{34}$$

The apostrophe denotes the transpose of a matrix. The Kalman gain matrix is given by

$$K = \tilde{P}_{xy}\tilde{P}_{yy}^{-1} \,. \tag{35}$$

The covariance matrices are defined as the expectation of the outer (dyadic) products of the deviation of the states from the *a priori* estimates:

$$\tilde{P}_{xx} = E[(\mathbf{x} - \tilde{\mathbf{x}})(\mathbf{x} - \tilde{\mathbf{x}})'] \,, \tag{36}$$

$$\tilde{P}_{xy} = E[(\mathbf{x} - \tilde{\mathbf{x}})(\mathbf{y} - \tilde{\mathbf{y}})'] \,, \tag{37}$$

and

$$\tilde{P}_{yy} = E\left[(\mathbf{y} - \tilde{\mathbf{y}})(\mathbf{y} - \tilde{\mathbf{y}})'\right] \,. \tag{38}$$

The vectors are understood to be column vectors, and the apostrophe denotes their transpose, i.e. making them row vectors.

In these equations it is assumed implicitly, without indexing, that the expectation value $E[\cdot]$ is conditioned on all measurements made up to the time indicated by the covariance on the l.h.s of the equation. For example, Eq. (36) exactly means $\tilde{P}_{xx,t+\Delta t | t} = E[(\mathbf{x}_{t+\Delta t} - \tilde{\mathbf{x}}_{t+\Delta t | t})(\mathbf{x}_{t+\Delta t} - \tilde{\mathbf{x}}_{t+\Delta t | t})' | \ldots, \mathbf{x}_{t-\Delta t}, \mathbf{x}_t]$. The *a priori* state estimate $\tilde{\mathbf{x}}$ is

$$\tilde{\mathbf{x}} = E[\mathbf{F}(\hat{\mathbf{x}}_{t | t})] \,. \tag{39}$$

Similarly, the *a priori* measurement estimate is

$$\tilde{\mathbf{y}} = \mathbf{G}(\tilde{\mathbf{x}}) \,. \tag{40}$$

Equations (33) to (40) will be derived in Appendix A. They yield the optimal filter for the case of normally distributed states in a linear state space model and constitute approximations for the nonlinear case.

If a state space model is explicitly given in the form of Eqs. (7) and (8), the KF can be expressed also by using the system matrices F and G:

$$K = \tilde{P}_{xx}G'(G\tilde{P}_{xx}G' + R)^{-1}, \quad (41)$$

$$\hat{P}_{xx} = (1 - KG)\tilde{P}_{xx}, \quad (42)$$

$$\tilde{P}_{xx} = F\hat{P}_{xx,t|t}F' + Q_t, \quad (43)$$

$$\tilde{\mathbf{x}} = F\hat{\mathbf{x}}_{t|t}, \quad (44)$$

$$\tilde{\mathbf{y}} = G\tilde{\mathbf{x}}. \quad (45)$$

These equations will also be derived in Appendix A. The first three equations can be computed offline.

Note that the process noise $\boldsymbol{\varepsilon}$ may vanish without causing problems in the expressions for the KF [Gelb, 1974], but not the measurement noise $\boldsymbol{\eta}$. Therefore, the matrix Q may be omitted for deterministic dynamics, but the matrix R should always be retained; otherwise the KF does not make much sense.

For the limit of infinitesimally small sampling time steps, the time discretization cannot be performed so straightforwardly as done here; rather, one should start with the Riccati equations which constitute the KF equations for the time-continuous case [Sinha & Rao, 1991]. Since we are dealing with data with a finite sampling time step, using the model (4), (5), we adhere to this writing for the rest of the paper.

6.2. *Nonlinear Kalman filtering*

The KF (41) to (45) yields the optimum recursive ML estimates for the states in the linear state space model (7), (8). The reason is that Gaussian random variables remain Gaussians for all times, and due to the Markov property the ML estimator amounts to a recursive least squares estimator. To return to our main problem, the estimation of states and parameters in the nonlinear system (4), (5), we need to relax the assumption of model linearity. We discuss two different approaches to achieve an approximation for weakly nonlinear functions: The method of *extended Kalman filtering* approximates the *system function* by locally linear functions, and the method of *unscented Kalman filtering* approximates the *density of the states* by a Gaussian.

In both cases, the problem of recursive least squares estimation boils down to the following: Given a nonlinear relationship

$$\mathbf{x}_{t+\Delta t} = \mathbf{F}(\mathbf{x}_t) \quad (46)$$

and

$$\bar{\mathbf{x}}_t := E[\mathbf{x}_t], \quad P_{xx,t} = E[(\mathbf{x}_t - \bar{\mathbf{x}}_t)(\mathbf{x}_t - \bar{\mathbf{x}}_t)'], \quad (47)$$

how can mean and covariance be best approximated in the next time step? That is, find $\bar{\mathbf{x}}_{t+\Delta t}$ and $P_{xx,t+\Delta t}$.

(i) The extended Kalman filter: In the EKF, the nonlinearity in Eq. (46) is Taylor expanded up to a given order, and then expectations are taken to compute the updated mean and covariance; for the EKF of first order, these so obtained mean and covariance are accurate up to first and second orders, respectively:

$$\bar{\mathbf{x}}_{t+\Delta t} \approx \mathbf{F}(\mathbf{x}_t), \quad (48)$$

$$P_{xx,t+\Delta t} \approx \nabla\mathbf{F}(\mathbf{x}_t)\,P_{xx,t}(\nabla\mathbf{F}(\mathbf{x}_t))'. \quad (49)$$

This approximation is equivalent to a state space model with time-varying system matrix F_t; it is allowed to do this because the derivation of the Kalman equations (41) to (45) would go through unchanged with a time-dependent matrix F [Honerkamp, 2002].

(ii) The unscented Kalman filter: There are alternative approaches to the EKF of which some have significant advantages with respect to implementation or accuracy. The *unscented Kalman filter* (UKF) retains the exact nonlinearity $\mathbf{F}$ but approximates the *a posteriori* probability density of the state $\mathbf{x}_{t+\Delta t}$ by a Gaussian. It is motivated by the fact that it is easier to approximate a distribution by a Gaussian than to approximate an arbitrary nonlinear function by linearization [Julier *et al.*, 1995; Julier *et al.*, 2000]. The approximation of the *a posteriori* density is accomplished by using a set of characteristic points, the so-called *sigma points*. They will be be used to adjust the probability density of the state after transformation by the system function $\mathbf{F}$ to a Gaussian again:

A normally distributed random variable $\mathbf{x}$ of dimension D_x is completely described by its mean and covariance matrix. This information can be stored, with some redundancy, in $2D_x$ sigma points $\mathcal{X}_i$ of dimension D_x each:

$$\mathcal{X}_i := \bar{\mathbf{x}} + (\sqrt{D_x P})_i \quad (i = 1, \ldots, D_x), \quad (50)$$

$$\mathcal{X}_{i+D_x} := \bar{\mathbf{x}} - (\sqrt{D_x P})_i \quad (i = 1, \ldots, D_x). \quad (51)$$

The square root is any matrix square root of choice [Strang, 1988], and $(\sqrt{\cdot})_i$ denotes its ith row or column. Considered as a data sample, this set of sigma points has the same mean and covariance as $\mathbf{x}$.

A simple example may shed some light on this: If the random variable $\mathbf{x}$ has zero mean and its covariance matrix is

$$P_{xx} = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}, \qquad (52)$$

then the set of sigma points is

$$\sqrt{2}\left\{\begin{pmatrix} \sqrt{R_1} \\ 0 \end{pmatrix}, -\begin{pmatrix} \sqrt{R_1} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \sqrt{R_2} \end{pmatrix}, -\begin{pmatrix} 0 \\ \sqrt{R_2} \end{pmatrix}\right\}, \qquad (53)$$

and it follows that

$$P_{\mathcal{X}\mathcal{X}} = \begin{pmatrix} \frac{1}{4}(2R_1 + 2R_1) & 0 \\ \\ 0 & \frac{1}{4}(2R_2 + 2R_2) \end{pmatrix} = P_{xx}. \qquad (54)$$

It can be shown that the transformation of sigma points,

$$\mathcal{X}_{i,t+\Delta t} = \mathbf{F}(\mathcal{X}_{i,t}), \qquad (55)$$

$$\bar{\mathbf{x}}_{t+\Delta t} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} \mathcal{X}_{i,t+\Delta t}, \qquad (56)$$

$$P_{xx,t+\Delta t} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} (\mathcal{X}_{i,t+\Delta t} - \bar{\mathbf{x}}_{t+\Delta t})$$
$$\times (\mathcal{X}_{i,t+\Delta t} - \bar{\mathbf{x}}_{t+\Delta t})', \qquad (57)$$

preserves statistics up to second order in a Taylor series expansion [Julier & Uhlmann, 1996]:

$$\bar{\mathbf{x}}_{t+\Delta t} \approx \mathbf{F}(\mathbf{x}_t) + \frac{\nabla' P_{xx,t} \nabla}{2} \mathbf{F}(\mathbf{x}_t), \qquad (58)$$

$$P_{xx,t+\Delta t} \approx \nabla \mathbf{F}(\mathbf{x}_t) P_{xx,t} (\nabla \mathbf{F}(\mathbf{x}_t))'. \qquad (59)$$

Compared with the EKF of first order, Eq. (48), (49), it is more accurate for nonlinear systems, and much simpler to computationally implement than the extended KF. In the case of linear state space models, both methods yield exactly the same estimates.

The application of the unscented transformation (55) to (59) to Kalman filtering in the nonlinear deterministic state space model (4), (5) is straightforward: Use the KF, Eqs. (33) to (40), with the following mean values and covariances:

$$\tilde{\mathbf{x}} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} \tilde{\mathcal{X}}_i \quad (\tilde{\mathcal{X}}_i = \mathbf{F}(\mathcal{X}_{i,t})), \qquad (60)$$

$$\tilde{\mathbf{y}} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} \tilde{\mathcal{Y}}_i \quad (\tilde{\mathcal{Y}}_i = \mathbf{G}(\tilde{\mathcal{X}}_i)), \qquad (61)$$

$$\tilde{P}_{xx} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} (\tilde{\mathcal{X}}_i - \tilde{\mathbf{x}})(\tilde{\mathcal{X}}_i - \tilde{\mathbf{x}})', \qquad (62)$$

$$\tilde{P}_{xy} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} (\tilde{\mathcal{X}}_i - \tilde{\mathbf{x}})(\tilde{\mathcal{Y}}_i - \tilde{\mathbf{y}})', \qquad (63)$$

$$\tilde{P}_{yy} = \frac{1}{2D_x} \sum_{i=1}^{2D_x} (\tilde{\mathcal{Y}}_i - \tilde{\mathbf{y}})(\tilde{\mathcal{Y}}_i - \tilde{\mathbf{y}})' + R_{t+\Delta t}. \qquad (64)$$

Equations (33) to (35) together with (60) to (64) provide the complete UKF. Its computational implementation is rather easy, since the system function $\mathbf{F}$ is used as it is; for example, if the numerical solution of the system requires a particular integration scheme, this routine can be used unmodified for estimation also; the only difference is that rather than computing a single trajectory, now $2D_x$ trajectories have to be integrated, one for each sigma point.

The approximations imposed by the UKF are to neglect cumulants of order higher than two in the densities of states. If higher order cumulants become significant, e.g. for heavy-tailed or multimodal distributions, the estimates can become biased. However, it often turns out that the UKF gives excellent results for even higher order nonlinearities. The reason is that the nonlinearities that arise from the temporal discretization of the dynamics of a time continuous system are only weakly nonlinear, unless the time steps are too large; each numerical integration scheme is based on a decomposition of the dynamics into an identity mapping to which a comparably small correction is added.

One of the main advantages of this approach is that there is no need for a computation of derivatives with respect to the state. This allows the straightforward use of state space models that contain nondifferentiable terms or models where the Jacobian cannot be computed easily. This is often the case for high-dimensional systems occurring with partial differential equations.

If it is not necessary to work recursively or in real-time, the estimates can be improved by using

the *Kalman smoother* rather than the Kalman filter. In the Kalman smoother, after a forward sweep over the time series, a backward sweep is performed, such that the estimate at time t depends not only on past but also on future times. The combination of forward and backward filter yields, at least in the linear case, the optimal smoother [Fraser & Potter, 1969]. Equivalent concepts for the UKF are not known so far to our knowledge.

6.3. *Parameter estimation with the nonlinear Kalman filter*

There are different possibilities to estimate parameters using the nonlinear Kalman filter based on Eqs. (33) to (35) and (60) to (64). We will concentrate on the so-called *augmented state vector* approach, since it can be implemented without hardly any modification of the filtering equations as they stand now. Other approaches are mentioned below.

In the augmented state vector approach, the dynamical system (4), (5) is rewritten in the following way:

$$
\begin{aligned}
\boldsymbol{\lambda}_{t+\Delta t} &= \boldsymbol{\lambda}_t \,, \\
\mathbf{x}_{t+\Delta t} &= \mathbf{F}(\mathbf{x}_t, \boldsymbol{\lambda}) \,, \\
\mathbf{y}_{t+\Delta t} &= \mathbf{G}(\mathbf{x}_{t+\Delta t}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_{t+\Delta t} \,,
\end{aligned}
\tag{65}
$$

i.e. the state vector is augmented by the constant parameter vector. Also without a stochastic term in the parameter equation, the *estimate* of the parameter can still vary in time due to the Kalman update (33), which influences also the estimate of $\boldsymbol{\lambda}_{t+\Delta t}$. But if the dynamics is stationary, the covariance of $\hat{\boldsymbol{\lambda}}_t$ should decrease steadily whilst recursively proceeding through the time series and approaching the true value. This behavior is illustrated in the following example.

For a further discussion of this way of parameter estimation, including the problem of estimation biases due to the approximate nature of recursive estimation and possible solutions, see e.g. [Bar-Shalom *et al.*, 2001].

6.4. *An example*

We again consider the Lorenz system (29), where only disturbed values of the x_1 component are measured and all three parameters are unknown. The data used are the same as in Sec. 5.3.

For parameter estimation, the augmented state vector approach is used, i.e. the three parameters of the Lorenz system are extended to the state of

the system to $D_x = 6$ dimensions. The complete nonlinear state space model is

$$
\begin{aligned}
\lambda_{i,t+\Delta t} &= \lambda_{i,t} \quad (i = 1,\, 2,\, 3)\,, \\
x_{i,t+\Delta t} &= x_{i,t} + I_i \quad (i = 1,\, 2,\, 3)\,, \\
y_t &= x_{1,t} + \eta_t
\end{aligned}
\tag{66}
$$

with

$$
\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \int_t^{t+\Delta t} \begin{pmatrix} -\lambda_{1,T} x_{1,T} + \lambda_{1,T} x_{2,T} \\ \lambda_{2,T} x_{1,T} - x_{2,T} - x_{1,T} x_{3,T} \\ -\lambda_{3,T} x_{3,T} + x_{1,T} x_{2,T} \end{pmatrix} dT \,.
\tag{67}
$$

The application of the UKF is shown in Figs. 8(a) to 8(c). The matrix square root used for calculating the sigma points is taken as one factor of the Cholesky decomposition of the covariance matrix [Press *et al.*, 1997]. The whole procedure is used in an iterative way, i.e. in addition to the recursive structure of the algorithm, it is applied iteratively to obtain several "sweeps" over the data set. This method can significantly reduce that part of estimation error which is caused by the nonlinearities of the system [Gelb, 1974; Ljung & Söderström, 1983]. In the first sweep over the data [Fig. 8(a)] the parameter estimates start with rather large uncertainties, which where specified initially to be large. In this case the trajectory adjusts to the true trajectory quickly, thus has low bias, but for the price of a large variance. The initial guesses for the second sweep are set in the following way: The initial unobserved state components are retained from the first sweep, and the initial parameter estimates are set to their final values at $t = 400$ from the first sweep. The covariance matrix P_{xx} is taken as the final value of the first sweep, but with entries related to a cross-covariance between states and parameters removed.

This procedure results in $\chi^2 = 321$ and a parameter vector estimate of $\hat{\boldsymbol{\lambda}} = (9.97 \pm 0.11,\ 46.47 \pm 0.31,\ 2.68 \pm 0.02)$ (1σ confidence intervals). The parameter vector estimate is obtained by averaging the estimates over all times; its error is obtained by taking the standard deviation of these parameter estimates.

The comparison of the recursive estimation result [Fig. 8(c)] with the multiple shooting result [Fig. 6(c)] reveals the following:

(i) The recursively estimated trajectory is not as smooth as the one estimated with multiple shooting. On the other hand, its χ^2 value is

smaller, namely 321 versus 353, respectively. The reason may be that there are no continuity constraints in recursive estimation, and the trajectory can follow the noisy measurements more closely due to the prediction-correction structure of the estimator. Both χ^2 are statistically compatible, and one cannot say that one method outperforms the other.

(ii) In this example, the parameter estimates are less biased and more accurate for recursive estimation. Although we believe that this result should not be generalized to other cases too early, the reason may be that the recursive parameter estimate is taken as the mean value over all estimates over time. The price one has to pay for the better parameter estimate is the diminished accuracy of the state estimate.

Which of these two approaches should be preferred in applications depends on many factors and cannot be said across the board. Up to now there are no studies about a performance comparison of recursive and shooting approaches for deterministic nonlinear systems.

Some other recent numerical examples and applications of these approaches are given in e.g. [Sitz *et al.*, 2002; Sitz *et al.*, 2003; Horbelt *et al.*, 2000; Horbelt *et al.*, 2001; Timmer *et al.*, 2000; Horbelt *et al.*, 2002; Müller & Timmer, 2002].

6.5. *Parameter tracking with the nonlinear Kalman filter*

If the parameter vector is not constant but *slowly* varying in time, instead of a parameter estimation problem one has the problem of *parameter tracking*. The difference to the state estimation problem is that there is no model needed to describe the variation of the parameter over time; rather, as long as the parameter varies slowly as compared to the state dynamics, it can behave in an arbitrary way. This problem can be tackled using Kalman filters in the following way:

In the augmented state vector approach, the dynamical system (4), (5) is rewritten as

$$\boldsymbol{\lambda}_{t+\Delta t} = \boldsymbol{\lambda}_t + \boldsymbol{\varepsilon}_t \,,$$
$$\mathbf{x}_{t+\Delta t} = \mathbf{F}(\mathbf{x}_t, \boldsymbol{\lambda}) \,, \qquad (68)$$
$$\mathbf{y}_{t+\Delta t} = \mathbf{G}(\mathbf{x}_{t+\Delta t}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_{t+\Delta t} \,,$$

i.e. the state vector is augmented by a stochastically varying parameter vector. There are now different ways to incorporate the information about

this variability, given by the covariance matrix Q_t of $\boldsymbol{\varepsilon}_t$, into the estimation procedure [Gelb, 1974; Bar-Shalom *et al.*, 2001]. Note that the estimation procedure still always yields a unique result; the term $\boldsymbol{\varepsilon}_t$ is just an auxiliary term, and only Q_t and not $\boldsymbol{\varepsilon}_t$ enters the estimation procedure. This will become clear in the following example.

6.6. *An example*

As an example, we consider the FitzHugh–Nagumo system [FitzHugh, 1961] which is in widespread use as a simplified model of the Hodgkin–Huxley system [Hodgkin & Huxley, 1952] for the excitable dynamics in certain nerve cells. It is given by the two-dimensional system

$$\dot{x}_1 = c\left(x_2 + x_1 - \frac{x_1^3}{3} + z(t)\right)$$
$$\dot{x}_2 = -\frac{x_1 - a + bx_2}{c} \,. \qquad (69)$$

The parameters are $a = 0.7$, $b = 0.8$, and $c = 3$. The variable x_1 is a membrane voltage which can be measured, and x_2 is a usually unobserved lumped variable that describes the combined effect of different kinds of ionic currents. The external voltage $z(t)$ influences the dynamics. Here we assume that $z(t)$ is varying slowly in comparison with the other two variables; it is treated here as the parameter λ_t to be tracked. Similar setups have been used to describe bursting in this system [Honerkamp *et al.*, 1985].

Our aim is to track the variation of $z(t)$ over time from the measurement of only x_1, disturbed by a significant amount of measurement noise as it is usually the case in experiments of this kind. In this example no parameters are estimated.

The measurements are constructed by corrupting the component $x_{1,t}$ with Gaussian noise $\boldsymbol{\eta}_t \sim \mathcal{N}(0, R)$ at 800 equidistant ($\Delta t = 0.2$) sample times. The standard deviation of $\boldsymbol{\eta}$, $\sqrt{R}$, is 20% of the standard deviation of the x_1-component of the system. The external variable $z(t)$ is constructed as a cosine-halfwave, up to an additive constant, as shown in Fig. 9(b). The initial guesses of z and x_2 are set to zero, the initial guess of x_1 is set to the measured value. For parameter tracking, the augmented state vector approach (68) is used with a constant covariance matrix $Q = 0.015$. In each estimation step, the component of the covariance matrix P_{xx} which corresponds to the variance of z
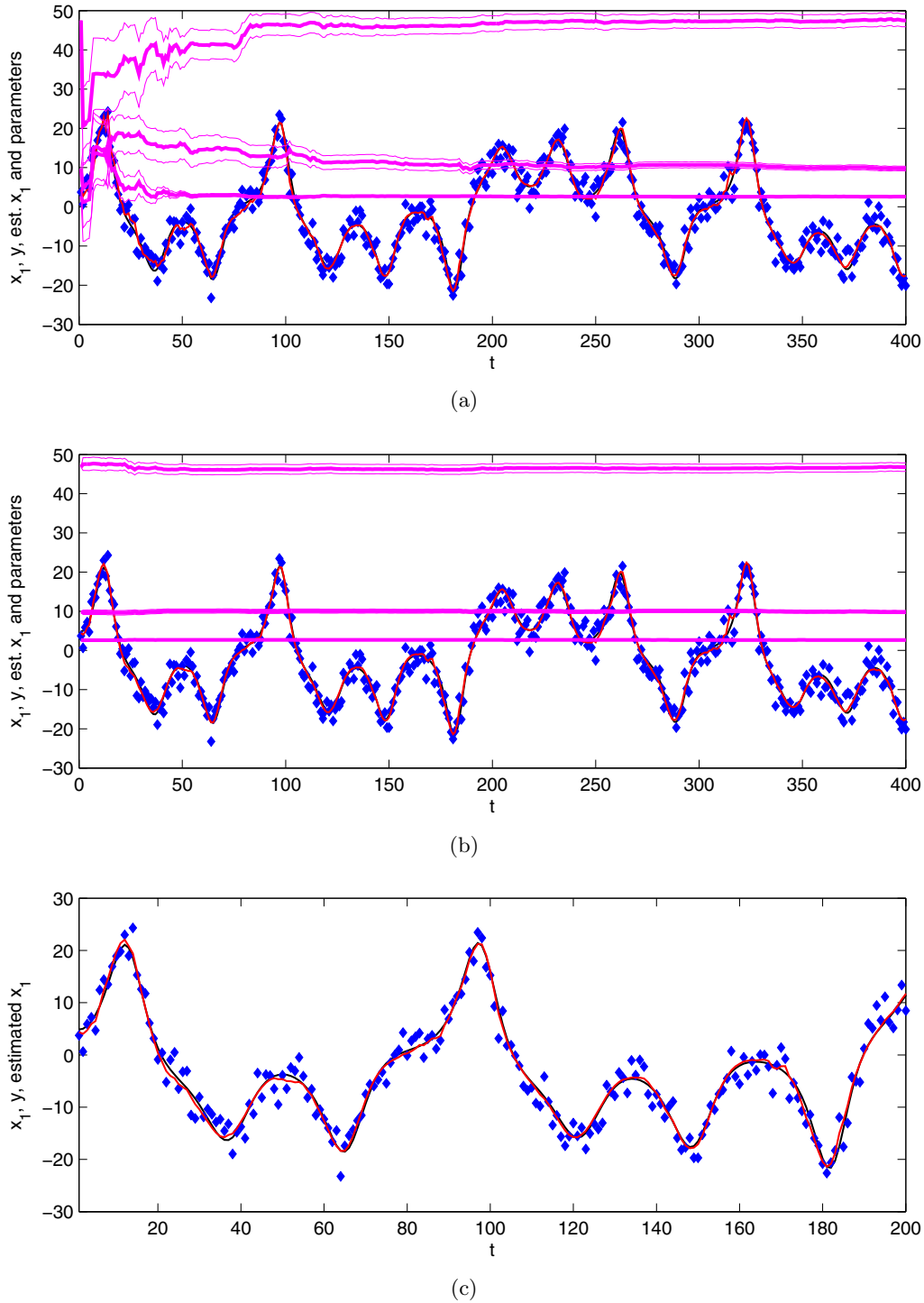
(a)

(b)

(c)

Fig. 8. Recursive parameter estimation in the Lorenz model (29) with the UKF approach. Like in Fig. 6, the states and parameters are estimated by observing noisy data from the first component of **x**. Measured data set (blue dots), true trajectory (black), estimated trajectory (red), and estimated parameters with confidence intervals (pink). (a) The result after the first iteration of the process. (b) After the second iteration of the process. (c) The first 200 data points, and the true and estimated trajectories of (b) to facilitate a comparison with the results of multiple shooting parameter estimation in Fig. 6(c).
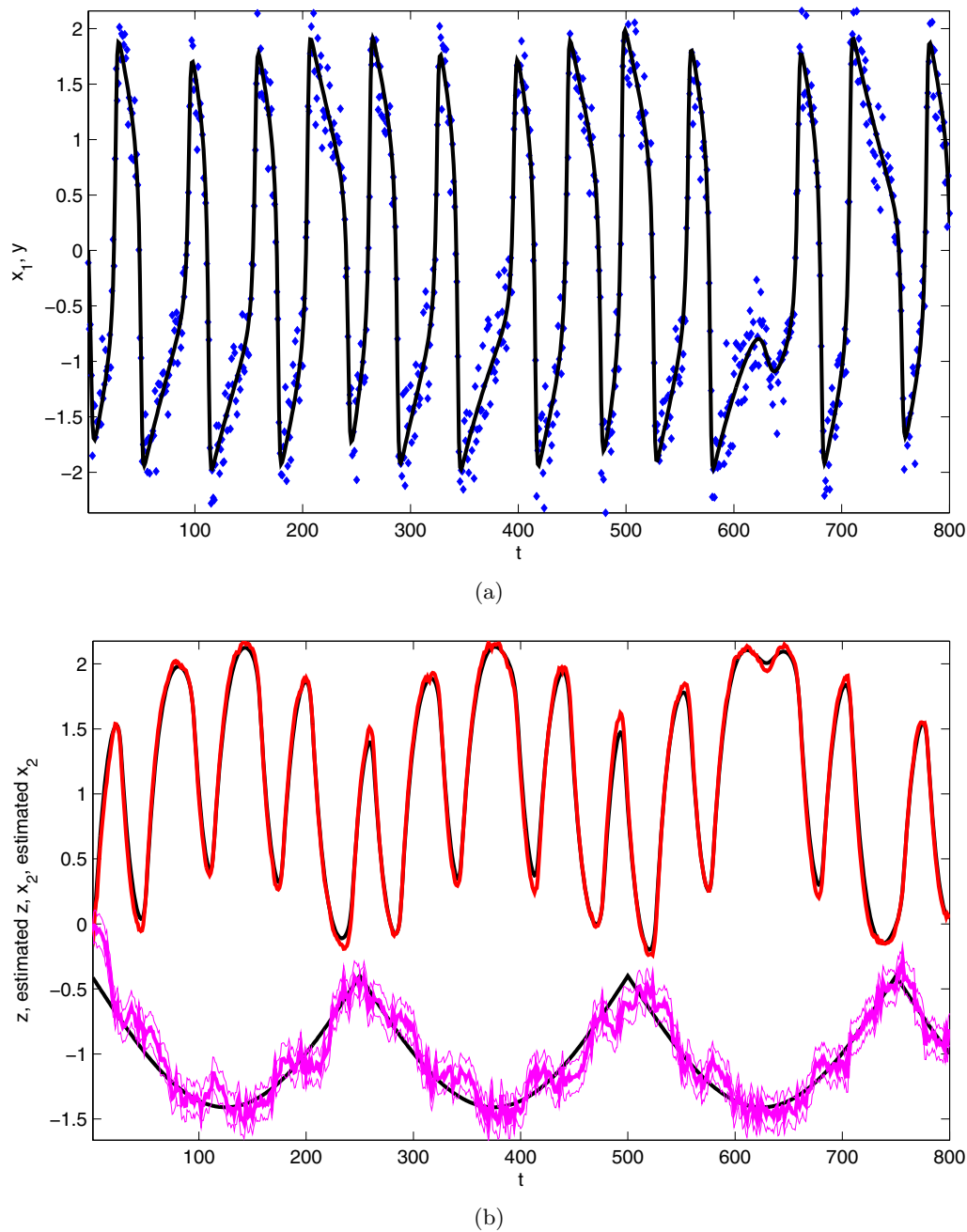
Fig. 9.   Recursive parameter tracking in the FitzHugh–Nagumo model (69) with the UKF approach. Like in the previous examples, the states and the external parameter are estimated by observing noisy data from the first component of $\mathbf{x}$. (a) Measured data set (blue dots) and the true trajectory of x_1 (black). (b) The unobserved external variable z (black), its tracked estimate (pink, shown with 1σ confidence intervals), the true trajectory of the unobserved component x_2 (black), and its estimate (red).

is set to the constant value Q to mimic the uncertainty of the variable to be tracked.

The application of the UKF is shown in Fig. 9. Since the parameters of the model were given, the variable z, corresponding to $\boldsymbol{\lambda}$ in Eq. (68), is tracked quite accurately, along with a very precise esti-

mate of the unobserved component of the state vector. Note that this example constitutes an application which cannot be treated with nonrecursive approaches like multiple shooting.

A Matlab routine reproducing this example is available via *www.fdm.uni-freiburg.de/~hv*.

7. Further Nonlinear Recursive Approaches

We have seen that the basis of all recursive non-linear filters is to reproduce as best as it gets the temporal evolution of the probability densities to maximize the likelihood. Here only the first two moments, mean and covariance, were taken into account. Recently, there evolved several other methods related to this task, some of them being more general. However, this is not an exhausting overview of all these methods, but can be used as a starting point for further reading, reflecting many of the actual main developments:

Methods to treat estimation in deterministic state space models recursively rather than with direct nonlinear optimization as in the initial value approach or multiple shooting go back to [Pearson, 1967]. In these methods, sometimes denoted by *nonlinear least squares methods* [Gelb, 1974], the probability density of the state $\mathbf{x}$ is unknown, like in our case.

The ideas of unscented Kalman filtering go back to the more general idea of *stochastic approximation* [Robbins & Munroe, 1951; Phaneuf, 1968; Sunahara, 1969; Mahalanabis & Farooq, 1971] and are put forward in [Norgaard *et al.*, 2000].

An alternative approach to the augmented state vector for parameter estimation with the KF is *dual estimation*. In dual estimation, the parameter system is treated as a system of its own, and the estimation procedure alternates between the parameter state and the system state [Wan *et al.*, 2000].

Related to dual estimation again is the *Expectation Maximization* (EM) algorithm [Dempster *et al.*, 1977; Shumway & Stoffer, 1982; Honerkamp, 2002] applied to nonlinear dynamical system identification [Ghahramani & Roweis, 1999]. In the EM algorithm, the parameter vector is not treated as evolving in time but estimated by the following iterative algorithm: In each iteration, the expectation of the state given the current parameter estimate and the data is calculated (E-step), then the parameter is updated by maximizing the conditional likelihood (M-step). The E-step is computed again with a KF, and the M-step usually with a gradient based algorithm, unless an analytic form as in linear models can be found. For hidden Markov models, which are based on a discrete state space, there exists also a recursive estimator based on the EM algorithm [Radons & Stiller, 1999].

Closely related to the estimation of unobserved states are model based nonlinear noise reduction methods which take the errors-in-variables problem into account [Farmer & Sidorowich, 1990; Hammel, 1990; Davies, 1992]. The problem of convergence of constrained minimization approaches for nonlinear noise reduction, which was not covered in this tutorial, has been investigated recently by [Bröcker & Parlitz, 2001].

The case of varying parameters in recursive estimation, i.e. nonstationary problems, has been investigated for a long time now by Young [Young, 1970, 1981, 1984]. Newer results allow even for recursive nonparametric model estimation [Young, 2000].

Filtering in uncertain models, where the explicit form of the state space model is known only up to some disturbances, has been recently studied in [Sayed & Subramanian, 2002]. This is related to the framework of *adaptive filtering* [Xie *et al.*, 1994; Petersen & Savkin, 1999].

The prediction-correction structure of the KF reminds one on coupled synchronizing dynamical systems and the *observer problem* in engineering science [Nijmeijer, 2001]. Indeed, recent methods appeared which use the principle of synchronization in coupled dynamical systems to estimate parameters [Stojanovski *et al.*, 1996; Maybhate & Amritkar, 1999; Sakaguchi, 2002].

In the UKF approach, only mean and covariance of the posterior probability are calculated exactly, hence it is valid only approximately if this assumption is violated. For strong nonlinearities, i.e. of more than second order if given by a polynomial, this approximation may cause a bias in the estimates. A filter that can cope with arbitrary nonlinearities in an exact way, however, has to be infinite dimensional, because it cannot be given in closed form [Hazewinkel *et al.*, 1983; Arulampalam *et al.*, 2002]. This is true at least up to some exceptions especially designed to yield a finite dimensional filter [Beněs, 1981; Daum, 1986; Schmidt, 1993]. The UKF used an optimally chosen set of sampling points, the sigma points, to represent the posterior density under the assumption of Gaussianity.

There are other novel methods based on Monte Carlo sampling [Müller, 1991] which allow to loosen this assumption by using more general densities and (*sequential*) *importance sampling* [Tanizaki, 1993], *bootstrapping* [Gordon *et al.*, 1993; Kitagawa, 1996],

particle filters [Arulampalam *et al.*, 2002], or *condensation* (from "conditional density estimation") [MacCormick & Blake, 1999]. The general idea of Monte Carlo sampling is to represent the densities by finite samples and to use them in the Kalman recursion. Thereby the samples are drawn in a random manner, unlike the sigma points which are computed uniquely. It is especially useful if the probability densities deviate strongly from a Gaussian, e.g. by being multimodal. Of course, this may be frequently the case for chaotic systems, but we are not aware of an application of particle filters to chaotic systems.

Importance sampling belongs to the *independent sampling family*. If the sample at time $t + \Delta t$ is configured such as to depend on the sample at time t, the samples are not independent any longer. This can be exploited to design highly efficient sampling strategies, known as Markov chain Monte Carlo estimations [Carlin *et al.*, 1992; Davies, 1998; Bremer & Kaplan, 2001], with the burden of a much higher computational load and difficult problems of slow convergence [Frühwirth-Schnatter, 1994; Shephard, 1994] and convergence diagnosis. However, these approaches are promising for distinguishing between process and measurement noise, a problem not encountered here at all, because in nonlinear system estimation this problem is in its general form unsolved yet.

The latter approaches are all rather new and research is proceeding rapidly.

8. Nonparametric Estimation and Spatiotemporal Systems

Nonparametric model estimation is a suitable tool for modeling in the case that explicit model equations are unknown; for sufficiently clean data, often a model or at least an initial guess of a model can be derived nonparametrically from time series of dynamical systems [Voss & Kurths, 1997; Timmer *et al.*, 2000; Peifer *et al.*, 2003] or from spatiotemporal data [Voss *et al.*, 1998; Voss *et al.*, 1999a; Andersen *et al.*, 2002]. The statistical measure of *maximal correlation* [Hirschfeld, 1935; Gebelein, 1941; Rényi, 1959; Csaki & Fischer, 1963] serves as a quantitative criterion for the goodness of the nonparametric model fits [Voss *et al.*, 1999b]. The herein often used ACE-algorithm [Breiman & Friedman, 1985; Frank & Lanteri, 1988; Buja, 1990; Härdle, 1990; Schimek, 2000] is accessible via *www.fdm.uni-freiburg.de/∼hv*. For nonparamet-

ric modeling, see also [Abel, 2004; Wessel *et al.*, 2004].

At an even more fundamental level, e.g. if one is not interested primarily in the model equations but in the dimension of the dynamical system, methods based on time-delay embeddings have been developed and applied to many physical and physiological systems [Kantz & Schreiber, 1995, 1997; Hegger *et al.*, 1999]. A method that gives more reliable results of the phase space dimension for noisy data than a precise estimation of correlation dimensions [Grassberger & Procaccia, 1983a, 1983b; Grassberger *et al.*, 1991; Albano *et al.*, 1988] is based on the false nearest-neighbor approach [Kennel *et al.*, 1992; Abarbanel *et al.*, 1993; Abarbanel, 1996; Hegger & Kantz, 1999]. This may be used as a tool to explore if one has to cope with a low-dimensional system or a higher dimensional one.

The latter case is true for spatially extended systems whose phase space dimension is infinite. Fortunately, many systems have only a limited range of scales for the spatial components of the dynamics. In these systems numerical simulations can still be performed in an approximate sense. Therefore, it should be possible also to perform some kind of parameter estimation for such systems. The generalization of methods mentioned in Secs. 5 and 6 to these systems is not straightforward though, because of the continued high-dimensionality which causes computational difficulties that cannot be compensated by large computer power alone; as an example, consider a two-dimensional system with measurements on a 100×100 spatial grid, and assume that the number of state components is 5. The dynamics of this system can then be imagined as a $5 \cdot 10^4$-dimensional state vector evolving in time, leading to covariance matrices with $25 \cdot 10^8$ entries in the UKF. This problem is only tractable if approximated somehow [Sitz *et al.*, 2003; Müller & Timmer, 2002]. See also [Busch & Hütt, 2004; Mandelj *et al.*, 2004].

9. Summary and Discussion

The problem of identifying parameters and unobserved state components from samples of continuous-time nonlinear dynamical systems faces two main difficulties: The errors-in-variables problem and the problem of complex cost functions. Whereas the former one can be overcome by taking the smooth dependence of the states properly into

account, the latter can be managed by splitting the optimization procedure to minimize the cost function into several parts. This methodology results in two somehow orthogonal approaches: multiple shooting methods and recursive filtering methods. Both turn out to be useful especially for the analysis of chaotic processes. Although the recursive methods can only yield approximate results, it has been shown that they may compete well with the multiple shooting approach. An advantage is that the implementation is much simpler than for methods based on global, nonrecursive, optimization.

Some words are in order why we consider only deterministic systems, although e.g. recursive estimation was developed historically in stochastic systems. The reason is twofold: Parameter estimation in nonlinear stochastic differential equations, as opposed to time-discrete models, suffers from the fact that the integration time steps must usually be orders of magnitude smaller than the sampling time steps, unless the data are highly sampled [Dacunha-Castelle & Florens-Zmirou, 1986; Florens-Zmirou, 1989], and there must be no measurement noise. These situations do hardly occur in real applications. Whether the proposed quasi-maximum-likelihood approaches of [Bibby & Sørensen, 1995; Timmer, 2000] can be extended to data with measurement noise still needs to be investigated. The other reason is that estimation approaches based on the smoothness of trajectories cannot rely on this assumption any more, because on small scales the trajectories cannot be approximated by smooth ones [Kloeden & Platen, 1995]. The problem of identification of stochastic dynamical systems provides much room for future research. See also [Lotosky & Rozovskii, 1998; Siegert *et al.*, 1998] and [Siefert & Peinke, 2004].

Before being applied to experimental data, both multiple shooting and recursive estimation should be validated first on numerical simulations for the following reason: We did not encounter the problem of identifiability of system parameters and trajectories. For example, even if the state space formulation of the system is linear, it may happen that a solution cannot be determined in an unambiguous way. In other words, it may be impossible to uniquely estimate the complete state from the indirect measurements. This is not a question of the amount of noise but of the system's structure. For linear systems, criteria can be formulated that allow to test for so-called *observ-ability* [Gelb, 1974; Conti, 1976; Barnett, 1983; Anderson & Deistler, 1984; Chui & Chen, 1989; Söderström *et al.*, 2002], but there is no concise theory up to now for nonlinear systems.

In recent years nonlinear system identification has made much progress, but developed along two different branches. Both branches already found a lot of applications; the recursive estimation approach predominantly in the engineering sciences, and shooting approaches in the natural sciences. We expect that both could profit from each other if combined, and many novel solutions for the still open challenges could be found.

Acknowledgments

References

Abarbanel, H. D. I. *et al.* [1993] "The analysis of observed chaotic data in physical systems," *Rev. Mod. Phys.* **65**, 1331–1392.

Abarbanel, H. D. I. [1996] *Analysis of Observed Chaotic Data* (Springer, NY).

Aguirre, L. A. & Billings, S. A. [1995] "Retrieving dynamical invariants from chaotic data using NARMAX models," *Int. J. Bifurcation and Chaos* **5**, 449–474.

Albano, A. M. *et al.* [1988] "Singular-value decomposition and the Grassberger–Procaccia algorithm," *Phys. Rev.* **A38**, 3017–3026.

Alligood, K., Sauer, T. D. & Yorke, J. A. [1997] *Chaos: An Introduction to Dynamical Systems* (Springer, Heidelberg).

Anderson, B. D. O. & Moore, J. B. [1979] *Optimal Filtering* (Prentice-Hall).

Anderson, B. D. O. & Deistler, M. [1984] "Identifiability of dynamic errors-in-variables models," *J. Time Series Anal.* **5**, 1–13.

Andersen, K. H. *et al.* [2002] "Pattern dynamics of vortex ripples in sand: Nonlinear modeling and experimental validation," *Phys. Rev. Lett.* **88**, 234302.

Aoki, M. [1990] *State Space Modeling of Time Series* (Springer, NY).

Arulampalam, S., Maskell, S. R., Gordon, N. J. & Clapp, T. [2002] "A tutorial on particle filters for

on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Sign. Process*, to appear.

Baake, E., Baake, M., Bock, H. G. & Briggs, K. M. [1992] "Fitting ordinary differential equations to chaotic data," *Phys. Rev.* **A45**, 5524–5529.

Baake, E. & Schlöder, J. P. [1992] "Modeling the fast fluorescence rise of photosynthesis," *Bull. Math. Biol.* **54**, 999–1021.

Bar-Shalom, Y. & Fortmann, T. E. [1988] *Tracking and Data Association* (Academic Press, Orlando, Florida).

Bar-Shalom, Y., Li, X. R. & Kirubarajan, T. [2001] *Estimation with Applications to Tracking and Navigation* (Wiley, NY).

Barnett, S. [1983] *Polynomials and Linear Control Systems* (Marcel Dekker, NY).

Bellman, R. E. & Kalaba, R. [1965] *Quasiliniearization and Boundary Value Problems* (American Elsevier, NY).

Beněs, V. E. [1981] "Exact finite-dimensional filters for certain diffusions with nonlinear drift," *Stochastics* **5**, 65–92.

Bibby, B. M. & Sørensen, M. [1995] "Martingale estimation functions for discretely observed diffusion processes," *Bernoulli* **1**, 17–39.

Bock, H. G. [1981] "Numerical treatment of inverse problems in chemical reaction kinetics," *Modeling of Chemical Reaction Systems*, eds. Ebert, K. H., Deuflhard, P. & Jäger, W., Vol. 18 (Springer, NY), pp. 102–125.

Bock, H. G. [1983] "Recent advances in parameter identification for ordinary differential equations," *Progress in Scientific Computing*, Vol. 2, eds. Deuflhard, P. & Hairer, E. (Birkhäuser, Boston), pp. 95–121.

Boggs, P. T., Byrd, R. H. & Schnabel, R. B. [1987] "A stable and efficient algorithm for nonlinear orthogonal distance regression," *SIAM J. Scient. Stat. Comput.* **8**, 1052–1078.

Breeden, J. L. & Hübler, A. [1990] "Reconstructing equations of motion from experimental data with unobserved variables," *Phys. Rev.* **A42**, 5817–5826.

Breiman, L. & Friedman, J. H. [1985] "Estimating optimal transformations for multiple regression and correlation," *J. Am. Stat. Assoc.* **80**, 580–598.

Bremer, C. L. & Kaplan, D. T. [2001] "Markov chain Monte Carlo estimation of nonlinear dynamics from time series," *Physica* **D160**, 116–126.

Bröcker, J. & Parlitz, U. [2001] "Efficient noncausal noise reduction for deterministic time series," *Chaos* **11**, 319–326.

Bryson, A. E. & Ho, Y. C. [1969] *Applied Optimal Control* (Waltham, Ginn).

Buja, A. [1990] "Remarks on functional canonical variates, alternating least squares methods and ACE," *Ann. Stat.* **18**, 1032–1069.

Carlin, B. P., Polson, N. G. & Stoffer, D. S. [1992]

"A Monte Carlo approach to nonnormal and nonlinear state space modeling," *J. Amer. Stat. Assoc.* **87**, 493–500.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. [1995] *Measurement Error in Nonlinear Models* (Chapman and Hall, London).

Chui, C. K. & Chen, G. [1989] *Linear Systems and Optimal Control* (Springer, Berlin).

Conti, R. [1976] *Linear Differential Equations and Control* (Academic Press, London).

Cremers, J. & Hübler, A. [1987] "Construction of differential equations from experimental data," *Z. Naturforsch.* **42a**, 797–802.

Crutchfield, J. P. & McNamara, B. S. [1987] "Equations of motion from data series," *Compl. Syst.* **1**, 417–451.

Csaki, P. & Fischer, J. [1963] "On the general notion of maximal correlation," *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **8**, 27–51.

Dacunha-Castelle, D. & Florens-Zmirou, D. [1986] "Estimation of the coefficients of a diffusion from discrete observations," *Stochastics* **19**, 263–284.

Daum, F. E. [1986] "Exact finite-dimensional nonlinear filters," *IEEE Trans. Automat. Contr.* **31**, 616–622.

Davies, M. [1992] "Noise reduction by gradient descent," *Int. J. Bifurcation and Chaos* **3**, 113–118.

Davies, M. E. [1998] "Nonlinear noise reduction through Monte Carlo sampling," *Chaos* **8**, 775–781.

Dempster, A. P., Laird, N. M. & Rubin, D. B. [1977] "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Stat. Soc.* **B39**, 1–38.

Edsberg, L. & Wedin, P. [1995] "Numerical tools for parameter estimation in ODE-systems," *Opt. Meth. Softw.* **6**, 193–217.

Farmer, J. D. & Sidorowich, J. J. [1990] "Optimal shadowing and noise reduction," *Physica* **D47**, 373–392.

FitzHugh, R. [1961] "Impulses and physiological states in theoretical models of nerve membrane," *Biophys. J.* **1**, 445–466.

Fletcher, R. & Reeves, C. M. [1964] "Function minimization by conjugate gradients," *Computer J.* **7**, 149–154.

Florens-Zmirou, D. [1989] "Approximate discretetime schemes for statistics of diffusion processes," *Statistics* **4**, 547–557.

Frank, I. E. & Lanteri, S. [1988] "ACE: Nonlinear regression model," *Chemom. Intell. Lab. Syst.* **3**, 301–313.

Fraser, D. C. & Potter, J. E. [1969] "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Automat. Contr.* **7**, 387–390.

Frühwirth-Schnatter, S. [1994] "Data augmentation and dynamic linear models," *J. Time Series Anal.* **15**, 183–202.

Fuller, W. A. [1987] *Measurement Error Models* (John Wiley, NY).

Gebelein, H. [1941] "Das statistische Problem der

Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung," *Zeitschrift für angew. Math. und Mech.* **21**, 364–379.

Gelb, A. [1974] *Applied Optimal Estimation* (The M.I.T. Press, Cambridge, MA).

Gershenfeld, N. [1999] *The Nature of Mathematical Modeling* (Cambridge University Press, Cambridge).

Ghahramani, Z. & Roweis, S. T. [1999] "Learning nonlinear dynamical systems using an EM algorithm," *Advances in Neural Information Processing Systems*, Vol. 11, eds. Kearns, M. J., Solla, S. A. & Cohn, D. A. (MIT Press).

Gill, P. E., Murray, W. & Wright, M. H. [1981] *Practical Optimization* (Academic Press, London).

Gordon, N. J., Salmond, D. J. & Smith, A. F. M. [1993] "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Comm. Radar Sign. Process.* **140**, 107–113.

Gouesbet, G. [1991a] "Reconstruction of standard and inverse vector fields equivalent to a Rössler system," *Phys. Rev.* **A44**, 6264–6280.

Gouesbet, G. [1991b] "Reconstruction of the vector fields of continuous dynamical systems from scalar time series," *Phys. Rev.* **A43**, 5321–5331.

Gouesbet, G. [1992] "Reconstruction of vector fields: The case of the Lorenz system," *Phys. Rev.* **A46**, 1784–1796.

Grassberger, P. & Procaccia, I. [1983a] "Characterization of strange attractors," *Phys. Rev. Lett.* **50**, 346–349.

Grassberger, P. & Procaccia, I. [1983b] "Measuring the strangeness of strange attractors," *Physica* **D9**, 189–208.

Grassberger, P., Schreiber, T. & Schaffrath, C. [1991] "Nonlinear time sequence analysis," *Int. J. Bifurcation and Chaos* **1**, 521–547.

Grewal, M. S. & Andrews, A. P. [2001] *Kalman Filtering: Theory and Practice Using MATLAB* (John Wiley, NY).

Hamilton, J. D. [1994] *Time Series Analysis* (Princeton University Press, Princeton).

Hammel, S., Grebogi, C. & Yorke, J. A. [1988] "Numerical orbits of chaotic processes represent true orbits," *Bull. Amer. Math. Soc.* **19**, 465–470.

Hammel, S. M. [1990] "A noise reduction method for chaotic systems," *Phys. Lett.* **A148**, 421–428.

Hanson, R. J. & Haskell, K. H. [1982] "Two algorithms for the linearly constrained least squares problem," *ACM Trans. Math. Softw.* **8**, 323–333.

Härdle, W. [1990] *Applied Nonparametric Regression* (Cambridge University Press, Cambridge).

Hazewinkel, M., Marcus, S. I. & Sussmann, H. J. [1983] "Nonexistence of finite dimensional nonlinear filters for conditional statistics of the cubic sensor problem," *Syst. Cont. Lett.* **3**, 331–340.

Hegger, R. *et al.* [1998] "Dynamical properties of a ferroelectric capacitor observed through nonlinear time series analysis," *Chaos* **8**, 727–736.

Hegger, R. & Kantz, H. [1999] "Improved false nearest neighbor method to detect determinism in time series," *Phys. Rev.* **E60**, 4970–4973.

Hegger, R., Kantz, H. & Schreiber, T. [1999] "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos* **9**, 413–435.

Hirschfeld, H. O. [1935] "A connection between correlation and contingency," *Proc. Camb. Phil. Soc.* **31**, 520–524.

Hodgkin, A. L. & Huxley, A. F. [1952] "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol* (London) **117**, 500–544.

Honerkamp, J., Mutschler, G. & Seitz, R. [1985] "Coupling of a slow and a fast oscillator can generate bursting," *Bull. Math. Biol.* **47**, 1–21.

Honerkamp, J. [2002] *Statistical Physics* (Springer, Berlin).

Horbelt, W., Müller, T., Timmer, J., Melzer, W. & Winkler, K. [2000] "Analysis of nonlinear differential equations: parameter estimation and model selection," *Medical Data Analysis*, Lecture Notes in Mathematics, Vol. 1933, eds. Brause, R. & Hanisch, E. (Springer, Berlin), pp. 152–158.

Horbelt, W. [2001] *Maximum Likelihood Estimation in Dynamical Systems*, PhD thesis, University of Freiburg.

Horbelt, W., Timmer, J., Bünner, M., Meucci, R. & Ciofini, M. [2001] "Identifying physical properties of a CO_2 laser by dynamical modeling of measured time series," *Phys. Rev.* **E64**, 016222.

Horbelt, W., Timmer, J. & Voss, H. U. [2002] "Parameter estimation in nonlinear delayed feedback systems from noisy data," *Phys. Lett.* **A299**, 513–521.

Jaeger, L. & Kantz, H. [1996] "Unbiased reconstruction of the dynamics underlying a noisy chaotic time series," *Chaos* **6**, 440–450.

Jazwinski, A. H. [1970] *Stochastic Processes and Filtering Theory* (Academic Press, NY).

Jeffreys, W. H. [1980] "On the method of least squares," *Astron. J.* **85**, 177–181.

Jeffreys, W. H. [1988] "Erratum: On the method of least squares," *Astron. J.* **95**, 1299.

Julier, S. J., Uhlmann, J. K. & Durrant-Whyte, H. F. [1995] "A new approach for filtering nonlinear systems," in *Proc. 1995 American Control Conf.*, Seattle, pp. 1628–1632.

Julier, S. J. & Uhlmann, J. K. [1996] "A general method for approximating nonlinear transformations of probability distributions," Technical Report, Dept. of Engineering Science, University of Oxford.

Julier, S., Uhlmann, J. & Durrant-Whyte, H. F. [2000] "A new method for the nonlinear transformation of

means and covariances in filters and estimators," *IEEE Trans. Automat. Contr.* **45**, 477–482.

Kadtke, J., Brush, J. & Holzfuss, J. [1993] "Global dynamical equations and Lyapunov exponents from noisy chaotic time series," *Int. J. Bifurcation and Chaos* **3**, 607–616.

Kalman, R. [1960] "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic. Eng. Series* **82/D**, 35–46.

Kalman, R. E. & Bucy, R. S. [1960] "New results in linear filtering and prediction theory," *Trans. ASME J. Basic. Eng. Series* **83/D**, 95–108.

Kantz, H. & Schreiber, T. [1995] "Dimension estimates and physiological data," *Chaos* **5**, 143–153.

Kantz, H. & Schreiber, T. [1997] *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge).

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. [1992] "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev.* **A45**, 3403–3411.

Kitagawa, G. [1996] "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Comput. Graph. Stat.* **5**, 1–25.

Kloeden, P. & Platen, E. [1995] *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin).

Kostelich, E. J. [1992] "Problems in estimating dynamics from data," *Physica* **D58**, 138–152.

Kostelich, E. J. [2001] "Bootstrap estimates of chaotic dynamics," *Phys. Rev.* **E64**, 016213.

Ljung, L. & Söderström, T. [1983] *Theory and Practice of Recursive Identification* (MIT Press, Cambridge, MA).

Lorenz, E. N. [1963] "Deterministic non-periodic flow," *J. Atmos. Sci.* **20**, 130–141.

Lotosky, S. V. & Rozovskii, B. L. [1998] "Recursive nonlinear filter for a continuous-discrete time model: Separation of parameters and observables," *IEEE Trans. Autom. Contr.* **43**, 1154–1158.

MacCormick, J. & Blake, A. [1999] *A Probabilistic Exclusion Principle for Tracking Multiple Objects* (IEEE Press).

Madansky, A. [1959] "The fitting of straight lines when both variables are subject to error," *J. Amer. Stat. Assoc.* **54**, 173–205.

Mahalanabis, A. K. & Farooq, M. [1971] "A second-order method for state estimation of nonlinear dynamical systems," *Int. J. Contr.* **14**, 631–639.

Maybhate, A. & Amritkar, R. [1999] "Use of synchronization and adaptive control in parameter estimation from time series," *Phys. Rev.* **E59**, 284–293.

McSharry, P. E. & Smith, L. A. [1999] "Better nonlinear models from noisy data: Attractors with maximum likelihood," *Phys. Rev. Lett.* **83**, 4285–4288.

Meinhold, R. J. & Singpuwalla, N. D. [1983] "Understanding the Kalman filter," *Amer. Statist.* **37**, 123–127.

Mendel, J. M. [1995] *Lessons in Estimation Theory for Signal Processing, Communications and Control* (Prentice Hall, NJ).

Miller, R. N., Ghil, M. & Gauthiez, F. [1994] "Advanced data assimilation in strongly nonlinear dynamical systems," *J. Atm. Sci.* **51**, 1037–1056.

Müller, P. [1991] "Monte Carlo integration in general dynamic models." *Statistical Multiple Integration*, eds. Flournoy, N. & Tsutakawa, R. K. (American Mathematical Society, Providence), pp. 145–163.

Müller, T. G. & Timmer, J. [2002] "Fitting parameters in partial differential equations from partially observed noisy data," *Physica* **D171**, 1–7.

Nicolis, G. [1995] *Introduction to Nonlinear Science* (Cambridge University Press, Cambridge).

Nijmeijer, H. [2001] "A dynamical control view on synchronization," *Physica* **D154**, 219–228.

Norgaard, M., Poulsen, N. K. & Ravn, O. [2000] "New developments in state estimation for nonlinear systems," *Automatica* **36**, 1627–1638.

Nusse, H. E. & Yorke, J. A. [1989] "Is every approximate trajectory of some process near an exact trajectory of a nearby process?" *Commun. Math. Phys.* **114**, 363–379.

Ott, E., Sauer, T. & Yorke, J. A. [1994] *Coping with Chaos*, Wiley Series in Nonlinear Science (John Wiley, NY).

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Pearson, J. B. [1967] "On nonlinear least squares filtering," *Automatica* **4**, 97–105.

Peifer, M., Timmer, J. & Voss, H. U. [2003] "Nonparametric identification of nonlinear oscillating systems," *J. Sound Vibr.*, to appear.

Petersen, I. R. & Savkin, A. V. [1999] *Robust Kalman Filtering for Signals and Systems with Large Uncertainties* (Birkhäuser, Boston).

Phaneuf, R. J. [1968] *Approximate Nonlinear Estimation*, PhD thesis, MIT.

Press, W. H. *et al.* [1997] *Numerical Recipes in C* (Cambridge University Press, Cambridge).

Radons, G. & Stiller, J. C. [1999] "Online estimation of hidden Markov models," *IEEE Sign. Process. Lett.* **6**, 213–215.

Rényi, A. [1959] "On measures of dependence," *Acta. Math. Acad. Sci. Hungar.* **10**, 441–451.

Robbins, H. & Munroe, S. [1951] "A stochastic approximation method A," *Math. Statist.* **22**, 400–407.

Sakaguchi, H. [2002] "Parameter evaluation from time sequences using chaos synchronization," *Phys. Rev.* **E65**, 1–4.

Sayed, A. H. & Subramanian, A. [2002] "State-space estimation with uncertain models," *Total Least Squares and Errors-in-Variables Modeling, III: Analysis,*

Algorithms and Applications, eds. van Huffel, S. & Lemmerling, P. (Kluwer, NY).

Schimek, M. G. (ed.) [2000] *Smoothing and Regression. Approaches, Computation and Application* (Wiley & Sons, NY).

Schittkowski, K. [1994] "Parameter estimation in systems of nonlinear equations," *Numer. Math.* **68**, 129–142.

Schittkowski, K. [1995] "Parameter estimation in differential equations," *Recent Trends in Optimization Theory and Applications*, ed. Agarwal, R. (World Scientific, Singapore), pp. 353–370.

Schmidt, G. C. [1993] "Designing nonlinear filters based on Daum's theory," *J. Guidance Contr. Dyn.* **16**, 371–376.

Seber, G. A. F. & Wild, C. J. [1989] *Nonlinear Regression* (John Wiley, NY).

Shephard, N. [1994] "Partial non-Gaussian state space," *Biometrika* **81**, 115–131.

Shumway, R. H. & Stoffer, D. S. [1982] "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Ser. Anal.* **3**, 253–264.

Shumway, R. H. & Stoffer, D. S. [2000] *Time Series Analysis and its Applications* (Springer, NY).

Siegert, S., Friedrich, R. & Peinke, J. [1998] "Analysis of data sets of stochastic systems," *Phys. Lett.* **A243**, 275–280.

Sinha, N. K. & Rao, G. P. (eds.) [1991] *Identification of Continuous-Time Systems* (Kluwer, Dordrecht).

Sitz, A., Schwarz, U., Kurths, J. & Voss, H. U. [2002] "Estimation of parameters and unobserved components for chaotic systems from noisy time series," *Phys. Rev.* **E66**, 016210.

Sitz, A., Kurths, J. & Voss, H. U. [2003] "Identification of nonlinear spatiotemporal systems via partitioned filtering," *Phys. Rev.* **E68**, 016202.

Söderström, T., Soverini, U. & Mahata, K. [2002] "Perspectives on errors-in-variables estimation for dynamic systems," *Sign. Process.* **82**, 1139–1154.

Stoer, J. [1971] "On the numerical solution of constrained leastsquares problems," *SIAM J. Numer. Anal.* **8**, 382–411.

Stoer, J. & Bulirsch, R. [1993] *Introduction to Numerical Analysis* (Springer, Heidelberg).

Stojanovski, T., Kocarev, L. & Parlitz, U. [1996] "A simple method to reveal the parameters of the Lorenz system," *Int. J. Bifurcation and Chaos* **6**, 2645–2652.

Stortelder, W. J. H. [1996] "Parameter estimation in chemical engineering, a case study for resin production," *Scientific Computing in Chemical Engineering*, eds. Keil, F., Mackens, W., Voß, H. & Werther, J. (Springer, Heidelberg), pp. 226–238.

Strang, G. [1988] *Linear Algebra and Its Applications* (Harcourt Brace Jovanovich, San Diego).

Sunahara, Y. [1969] "An approximate method of state estimation for nonlinear dynamical systems," in *Joint Automatic Control Conf.*, University of Colorado.

Takens, F. [1981] "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics, Vol. 898, eds. Rand, D. & Young, L. (Springer, Berlin), pp. 366–381.

Tanizaki, H. (ed.) [1993] *Nonlinear Filters; Estimation and Applications* (Springer, Berlin).

Theiler, J. [1990] "Statistical precision of dimension estimators," *Phys. Rev.* **A41**, 3038–3051.

Theiler, J. & Smith, L. A. [1995] "Anomalous convergence of Lyapunov exponent estimates," *Phys. Rev.* **E51**, 3738–3741.

Timmer, J. [1998] "Modeling noisy time series: Physiological tremor," *Int. J. Bifurcation and Chaos* **8**, 1505–1516.

Timmer, J. [2000] "Estimating parameters in nonlinear stochastic differential equations," *Chaos Solit. Fract.* **11**, 2571–2578.

Timmer, J., Rust, H., Horbelt, W. & Voss, H. U. [2000] "Parametric, nonparametric and parametric modelling of a chaotic circuit time series," *Phys. Lett.* **A274**, 123–134.

van Domselaar, B. & Hemker, P. [1975] "Nonlinear parameter estimation in initial value problems," Technical Report NW 18/75, Mathematical Centre Amsterdam.

van Huffel, S. & Vandewalle, J. [1991] *The Total Least Squares Problem: Computational Aspects and Analysis* (SIAM, Philadelphia).

Voss, H. & Kurths, J. [1997] "Reconstruction of nonlinear time delay models from data by the use of optimal transformations," *Phys. Lett.* **A234**, 336–344.

Voss, H., Bünner, M. J. & Abel, M. [1998] "Identification of continuous, spatiotemporal systems," *Phys. Rev.* **E57**, 2820–2823.

Voss, H. U., Kolodner, P., Abel, M. & Kurths, J. [1999a] "Amplitude equations from spatiotemporal binary-fluid convection data," *Phys. Rev. Lett.* **83**, 3422–3425.

Voss, H. U., Schwache, A., Kurths, J. & Mitschke, F. [1999b] "Equations of motion from chaotic data: A driven optical fiber ring resonator," *Phys. Lett.* **A256**, 47–54.

Walker, D. M. [2002] "Kalman filtering of time series data," *Modelling and Forecasting Financial Data-Techniques of Nonlinear Dynamics*, eds. Soofi, A. & Cao, L. (Kluwer Academic Publishers, Boston), pp. 137–157.

Wan, E. A., Merwe, R. & Nelson, A. T. [2000] "Dual estimation and the unscented transformation," *Advances in Neural Information Processing Systems*, Vol. 12, eds. Solla, S., Leen, T. & Müller, K.-R. (The MIT Press), pp. 666–672.

Wiggins, S. [1990] *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Springer, NY).

Xie, L., Soh, Y. C. & de Souza, C. E. [1994] "Robust Kalman filtering for uncertain discrete-time systems," *IEEE Trans. Autom. Contr.* **39**, 1310–1340.

Young, P. [1981] "Parameter estimation for continuous-time models — a survey," *Automatica* **17**, 23–39.

Young, P. [2000] "Stochastic, dynamic modelling and signal processing: Time variable and state dependent parameter estimation," *Nonlinear and Non-stationary Signal Processing*, eds. Fitzgerald, W. J. *et al.* (Cambridge University Press, Cambridge), pp. 74–114.

Young, P. C. [1970] "An instrumental variable method for real-time identification of a noisy process," *Automatica* **6**, 271–287.

Young, P. C. [1984] *Recursive Estimation and Time Series Analysis* (Springer, Berlin).

Appendix A
Derivation of the Filtering Equations (33) to (40) and (41) to (45)

We derive the KF putting more emphasis on a heuristic understanding rather than a more fundamental derivation based on the theorem of Bayes. For the latter approach, see for example [Honerkamp, 2002; Gershenfeld, 1999]. Another introduction into Kalman filtering from a more statistical viewpoint is given in [Meinhold & Singpuwalla, 1983].

First consider the linear state space model (7), (8) with parameter dependence neglected. To estimate the state $\mathbf{x}_t$ from a time series of measurements $\mathbf{y}_t$ recursively, we first note that

$$\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{t+\Delta t|t} = E\left[\mathbf{x}_{t+\Delta t|t}\right] = E[F\mathbf{x}_{t|t} + \boldsymbol{\varepsilon}_t] = F\hat{\mathbf{x}}_{t|t}\,. \tag{A.1}$$

Thereby it was taken that $E[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ and that the previous state $\mathbf{x}_t$ has to be estimated again from the states up to time $t - \Delta t$ and the measurements up to time t. Similarly,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_{t+\Delta t|t} = G\tilde{\mathbf{x}}_{t+\Delta t|t} = G\tilde{\mathbf{x}}\,. \tag{A.2}$$

The Kalman gain is now derived as follows: We try to estimate $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{t+\Delta t|t+\Delta t}$ in Eq. (32) as accurately as possible, i.e. with the lowest estimation error. Any random variable that results from a linear transformation of a Gaussian random variable again is a Gaussian random variable. Therefore, $\mathbf{x}_{t+\Delta t}$ is, like $\boldsymbol{\varepsilon}_t$, normally distributed, and completely described by its mean and covariance. The estimation error is thus given by the covariance matrix

$$\hat{P}_{xx,t+\Delta t|t+\Delta t} = E[(\mathbf{x}_{t+\Delta t} - \hat{\mathbf{x}}_{t+\Delta t|t+\Delta t}) \times (\mathbf{x}_{t+\Delta t} - \hat{\mathbf{x}}_{t+\Delta t|t+\Delta t})']\,, \tag{A.3}$$

or in our shorthand notation by

$$\hat{P}_{xx} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})']\,. \tag{A.4}$$

The Kalman gain K should be chosen such that the trace of this matrix is minimized, because the trace is the sum over the squared errors of the components of $\hat{\mathbf{x}}$. Here it is assumed that variances are additive. It is not necessary to think about a minimization of the cross-diagonal elements of $\hat{P}_{xx}$, since $\hat{P}_{xx}$ is symmetric and can thus be diagonalized; the trace is invariant under this transformation. To perform this minimization explicitly, we rewrite Eq. (A.4): From

$$\mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \tilde{\mathbf{x}} - K(\mathbf{y} - \tilde{\mathbf{y}}) \tag{A.5}$$

and

$$\mathbf{y} - \tilde{\mathbf{y}} = G\mathbf{x} + \boldsymbol{\eta} - G\tilde{\mathbf{x}} \tag{A.6}$$

one gets

$$\mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \tilde{\mathbf{x}} - KG(\mathbf{x} - \tilde{\mathbf{x}}) - K\boldsymbol{\eta}$$
$$= (1 - KG)(\mathbf{x} - \tilde{\mathbf{x}}) - K\boldsymbol{\eta}\,. \tag{A.7}$$

Therefore, using definition (9), we obtain

$$\hat{P}_{xx} = (1 - KG)E[(\mathbf{x} - \tilde{\mathbf{x}})(\mathbf{x} - \tilde{\mathbf{x}})']$$
$$\times (1 - KG)' + KRK'$$
$$= (1 - KG)\tilde{P}_{xx}(1 - KG)' + KRK'$$
$$= \tilde{P}_{xx} - KG\tilde{P}_{xx} - \tilde{P}_{xx}G'K'$$
$$+ K(G\tilde{P}_{xx}G' + R)K'\,. \tag{A.8}$$

Now choose K such that

$$\operatorname{tr}\hat{P}_{xx} \overset{!}{=} \min\,, \tag{A.9}$$

using the relations

$$\frac{d}{dA}\operatorname{tr}(AB) = B' \quad (AB \text{ square})\,, \tag{A.10}$$

$$\frac{d}{dA}\operatorname{tr}(ACA') = 2AC \quad (C \text{ symmetric})\,, \tag{A.11}$$

$$\left(\frac{df}{dA}\right)_{ij} := \frac{df}{dA_{ij}}\,, \tag{A.12}$$

and

$$\operatorname{tr}A = \operatorname{tr}A' : \tag{A.13}$$

$$\frac{d\operatorname{tr}\hat{P}_{xx}}{dK} = -2\tilde{P}_{xx}G' + 2K(G\tilde{P}_{xx}G' + R) = 0\,. \tag{A.14}$$

Solving this for the Kalman gain yields a minimum of (A.9) at

$$K = \tilde{P}_{xx}G'(G\tilde{P}_{xx}G' + R)^{-1}. \qquad (A.15)$$

It can be checked that this is indeed a minimum. Inserting this into Eq. (A.8) one obtains

$$\hat{P}_{xx} = \tilde{P}_{xx} - \tilde{P}_{xx}G'(G\tilde{P}_{xx}G' + R)G\tilde{P}_{xx}$$
$$= (1 - KG)\tilde{P}_{xx}. \qquad (A.16)$$

It remains to compute $\tilde{P}_{xx}$:

$$\tilde{P}_{xx} = E[(\mathbf{x}_{t+\Delta t} - \tilde{\mathbf{x}})(\mathbf{x}_{t+\Delta t} - \tilde{\mathbf{x}})']$$
$$= E[(F\mathbf{x}_t + \boldsymbol{\varepsilon}_t - F\hat{\mathbf{x}}_{t|t})(F\mathbf{x}_t + \boldsymbol{\varepsilon}_t - F\hat{\mathbf{x}}_{t|t})']$$
$$= E[(F(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}) + \boldsymbol{\varepsilon}_t)(F(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t}) + \boldsymbol{\varepsilon}_t)']$$
$$= E[F(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})'F'] + Q_t$$
$$= F\hat{P}_{xx,t|t}F' + Q_t. \qquad (A.17)$$

Therefore, $\tilde{P}_{xx}$ can be computed recursively from $\hat{P}_{xx}$ one time step before. This proves Eqs. (41) to (45).

Finally, we come to the proof of the filtering equations (33) to (40), which hold also for deterministic but nonlinear state space models, our main goal: A Gaussian random variable $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, P_{xx})$ transforms into the Gaussian $\mathbf{y} = G\mathbf{x} + \boldsymbol{\eta}$ via

$$\mathbf{y} \sim \mathcal{N}(G\bar{\mathbf{x}}, GP_{xx}G' + R) \quad (R = E[\boldsymbol{\eta}\boldsymbol{\eta}']). \quad (A.18)$$

Therefore,

$$\tilde{P}_{yy} = G\tilde{P}_{xx}G' + R \qquad (A.19)$$

and

$$\tilde{P}_{xy} = \tilde{P}_{xx}G'. \qquad (A.20)$$

Inserting Eqs. (A.19) and (A.20) into Eq. (A.15) proves Eq. (35). Equation (34) follows immediately.