Commun Nonlinear Sci Numer Simulat 18 (2013) 2733–2742

Contents lists available at SciVerse ScienceDirect

Commun Nonlinear Sci Numer Simulat

journal homepage: www.elsevier.com/locate/cnsns

Nonlinear system identification employing automatic differentiation

Jan Schumann-Bischoff, Stefan Luther, Ulrich Parlitz *

Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany
Institute for Nonlinear Dynamics, Georg-August-Universität Göttingen, Am Faßberg 17, 37077 Göttingen, Germany
DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, and Heart Research Center Göttingen, D-37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 1 November 2012
Received in revised form 26 February 2013
Accepted 28 February 2013
Available online 15 March 2013

Keywords:
Nonlinear modelling
Parameter estimation
Delay differential equations
Data assimilation

ABSTRACT

An optimization based state and parameter estimation method is presented where the required Jacobian matrix of the cost function is computed via automatic differentiation. Automatic differentiation evaluates the programming code of the cost function and provides exact values of the derivatives. In contrast to numerical differentiation it is not suffering from approximation errors and compared to symbolic differentiation it is more convenient to use, because no closed analytic expressions are required. Furthermore, we demonstrate how to generalize the parameter estimation scheme to delay differential equations, where estimating the delay time requires attention.

1. Introduction

For many processes in physics or other fields of science mathematical models exist (in terms of differential equations, for example), but not all state variables are easily accessible (measurable) and proper values of model parameters may be (partly) unknown. In particular, detailed biological cell models (e.g., cardiac myocytes [1]) may include many variables which are difficult to access experimentally and, in addition, depend on up to hundreds of physical parameters whose values have to be determined. To estimate unobserved variables (as a function of time) and model parameters different identification methods have been devised [2–14,16–18,20,19]. These methods have in common that an attempt is made to adjust the model output (in general a function of the state variables) to some (experimentally) observed time series. To achieve agreement, unobserved variables and unknown model parameters are suitably adjusted such that the model reproduces and follows the observed time series. In geosciences and meteorology (e.g., whether forecasting) this procedure is often called *data assimilation* and describes the process of incorporating new (incoming) data into a computer model of the real system.

A general framework for state estimation provides, for example, the path integral formalism including a saddle point approximation [15,16]. This formalism can be used to state the estimation problem as an optimization problem [19,12–14,18]. If an optimization method is employed that is based on gradient descent (such as the well-known Levenberg–Marquard method [21,22]), in general the Jacobian matrix of the cost function has to be provided, whose derivation may be quite cumbersome (and error-prone), depending on the structure of the cost function and the underlying mathematical model of the dynamical system. To estimate the Jacobian matrix one may approximate it by numerical derivatives (often spoiled by unacceptably large truncation errors) or use symbolic mathematics, which requires, however, that the function to be derived has to be given in closed form.

* Corresponding author at: Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany. Tel.: +49 551 5176 369.
E-mail addresses: jan.schumann-bischoff@ds.mpg.de (J. Schumann-Bischoff), stefan.luther@ds.mpg.de (S. Luther), ulrich.parlitz@ds.mpg.de (U. Parlitz).

1007-5704/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cnsns.2013.02.017

A convenient alternative to both of these methods is *automatic differentiation* [24] where exact numerical values of the required derivatives are computed by analyzing a given source code implementation of the cost function. As will be shown here automatic differentiation leads in this context not only to a very flexible and efficient algorithm for computing the required Jacobian but also provides the sparsity pattern of the Jacobian which is exploited by suitable optimization methods. In Section 2 we will give a formal description of the optimization problem to be solved for state and parameter estimation. Then we briefly present in Section 3 the concept of automatic differentiation in the form used here. As an illustrative example we show in Section 4 how to estimate the model parameters of the Lorenz-96 model. In Section 5 we discuss how to estimate the delay time in delay differential equations and provide in Section 6 an example (Mackey–Glass model).

2. State and parameter estimation method

The method used here to adapt a model to a time series is based on minimizing a cost function and was introduced in Ref. [19]. For completeness we present in the following an extended version covering also delay differential equations (DDEs).

We assume that a multivariate R-dimensional time series $\{\boldsymbol{\eta}(n)\}$ is given consisting of $N+1$ samples $\boldsymbol{\eta}(n)\hat{=}\boldsymbol{\eta}(t_n)\in\mathbb{R}^R$ measured at times $\mathcal{T}=\{t_n=n\cdot\Delta t|n=0,1,\ldots,N\}$. For simplicity the observation times t_n are equally spaced (with a fixed time step Δt) and start at $t_0=0$. The estimation method can easily be extended to nonuniformly sampled observations (see Ref. [19]). Here we consider the general case of a model given by a set of coupled delay differential equations (DDEs)

$$\frac{\mathrm{d}\boldsymbol{y}(t)}{\mathrm{d}t}=\boldsymbol{F}(\boldsymbol{y}(t),\boldsymbol{y}_\tau(t),\boldsymbol{p},t), \tag{1}$$

with $\boldsymbol{y}_\tau(t)=\boldsymbol{y}(t-\tau)$. The state vector (s) $\boldsymbol{y}(t)=(y_1(t),\ldots,y_D(t))^{\mathrm{T}}$, the delay parameter $\tau\in\mathbb{R}$ and the U model parameters $\boldsymbol{p}=(p_1,\ldots,p_U)^{\mathrm{T}}$ are unknown and have to be estimated from the time series $\{\boldsymbol{\eta}(n)\}$. Estimating τ can not be conducted as estimating $\boldsymbol{p}$, because $\boldsymbol{F}(\boldsymbol{y}(t),\boldsymbol{y}_\tau(t),\boldsymbol{p},t)$ does *not* explicitly depend on τ. In fact $\boldsymbol{F}(\boldsymbol{y}(t),\boldsymbol{y}_\tau(t),\boldsymbol{p},t)$ depends on $\boldsymbol{y}_\tau(t)$ which is a function of τ. We shall later come back to this topic.

Note that (1) also describes (as a special case) models given by coupled ordinary differential equations (ODEs). In this case the right-hand side of (1) is independent of $\boldsymbol{y}_\tau(t)$ and thus can be replaced by $\boldsymbol{F}(\boldsymbol{y}(t),\boldsymbol{p},t)$ (see Ref. [19] for details).

To estimate the unknown quantities a measurement function

$$\boldsymbol{z}(t)=\boldsymbol{h}(\boldsymbol{y}(t),\boldsymbol{q},t), \tag{2}$$

is required to represent the relation between model states $\boldsymbol{y}(t)$ and the $\boldsymbol{z}(t)$ corresponding to the observations $\{\boldsymbol{\eta}(n)\}$. This measurement function may contain V additional unknown parameters $\boldsymbol{q}=(q_1,\ldots,q_V)^{\mathrm{T}}$ that also have to be estimated using information from the given time series $\{\boldsymbol{\eta}(n)\}$.

2.1. Cost function

The goal of the estimation process is to find a set of values for all unknown quantities such that the model equations provide via measurement function (2) a model times series $\{\boldsymbol{z}(t_n)\}$ that matches the experimental time series $\{\boldsymbol{\eta}(t_n)\}$. In other words, the average difference between $\boldsymbol{\eta}(t_n)$ and $\boldsymbol{z}(t_n)$ should be small. Furthermore, the model equations should be fulfilled as well as possible. This means that modeling errors $\boldsymbol{u}(t)$ are allowed, but should be small. Therefore, model (1) is extended to

$$\frac{\mathrm{d}\boldsymbol{y}(t)}{\mathrm{d}t}=\boldsymbol{F}(\boldsymbol{y}(t),\boldsymbol{y}_\tau(t),\boldsymbol{p},t)+\boldsymbol{u}(t). \tag{3}$$

The smaller $\boldsymbol{u}(t)$ is the better the model Eq. (1) are fulfilled. Next, for simplicity, $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$ will be discretized at the times in $\mathcal{T}$. This means that $\boldsymbol{y}(t)$ will be sampled at the same time when data are observed. With $\boldsymbol{y}(n)\hat{=}\boldsymbol{y}(n\cdot\Delta t)\hat{=}\boldsymbol{y}(t_n)$ and $\mathcal{Y}(a,b)=\{\boldsymbol{y}(n)|n=a,a+1,\ldots,b\}$ the set of values of the discretized model variables can be written as $\mathcal{Y}(0,N)$. The quantities in $\mathcal{Y}(0,N)$ have to be estimated in addition to $\boldsymbol{p}$ and $\boldsymbol{q}$. With the same discretization we have $\{\boldsymbol{u}(n)\}=\{\boldsymbol{u}(t_n)\}$. At this point we assume a fixed (not to be estimated) delay $\tau=k\cdot\Delta t$ with $k\in\mathbb{N}$ which is not necessarily equal to the delay parameter of the physical process underlying the data. This simplifies the discretization of the delayed variable to $\boldsymbol{y}_\tau(t)=\boldsymbol{y}(n\cdot\Delta t-k\cdot\Delta t)=\boldsymbol{y}((n-k)\cdot\Delta t)=\boldsymbol{y}(n-k)=\boldsymbol{y}_k(n)$. The set of the discretized delayed variable is then $\mathcal{Y}_k(0,N)=\mathcal{Y}(-k,N-k)$. Note that $\mathcal{Y}(-k,N-k)=\mathcal{Y}(-k,-1)\cup\mathcal{Y}(0,N-k)$. Since $\mathcal{Y}(0,N-k)\subset\mathcal{Y}(0,N)$, $\mathcal{Y}(0,N-k)$ contains no additional quantities to be determined. Only the variables in $\mathcal{Y}(-k,-1)$ are additional quantities which have to be estimated. Typically the delay time is much shorter than the length of the time series $N\cdot\Delta t$ and hence the number of elements in $\mathcal{Y}(-k,-1)$ is much smaller than in $\mathcal{Y}(0,N)$. Therefore the number of quantities to be estimated does not increase much compared to a model given by ODEs (with similar D and N) where $\mathcal{Y}(-k,-1)$ has not to be estimated.

The discretization of (3) is then given by

$$\boldsymbol{u}(n)\approx\left.\frac{\Delta\boldsymbol{y}}{\Delta t}\right|_{t_n}-\boldsymbol{F}(\boldsymbol{y}(n),\boldsymbol{y}_k(n),\boldsymbol{p},t_n), \tag{4}$$