Application of one-step method to parameter estimation in ODE models

Itai Dattner

Department of Statistics, University of Haifa, 199 Aba Khoushy Ave., Mount Carmel, Haifa 3498838, Israel

Shota Gugushvili*

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

In this paper, we study application of Le Cam's one-step method to parameter estimation in ordinary differential equation models. This computationally simple technique can serve as an alternative to numerical evaluation of the popular non-linear least squares estimator, which typically requires the use of a multistep iterative algorithm and repetitive numerical integration of the ordinary differential equation system. The one-step method starts from a preliminary $\sqrt{n}$-consistent estimator of the parameter of interest and next turns it into an asymptotic (as the sample size $n \to \infty$) equivalent of the least squares estimator through a numerically straightforward procedure. We demonstrate performance of the one-step estimator via extensive simulations and real data examples. The method enables the researcher to obtain both point and interval estimates. The preliminary $\sqrt{n}$-consistent estimator that we use depends on non-parametric smoothing, and we provide a data-driven methodology for choosing its tuning parameter and support it by theory. An easy implementation scheme of the one-step method for practical use is pointed out.

Keywords and Phrases: non-linear least squares, ordinary differential equations, smooth and match estimator, integral estimator, Levenberg–Marquardt algorithm, one-step estimator.

AMS 2000 classifications: Primary: 62F12, Secondary: 62G08, 62G20

1 Introduction

Systems of ordinary differential equations (ODEs in short) are commonly used for the mathematical modelling of the rate of change of dynamic processes (e.g. in mathemat-

---

*shota.gugushvili@math.leidenuniv.nl

ical biology, see Edelstein-Keshet, 2005; in the theory of chemical reaction networks, see Feinberg, 1979 and Sontag, 2001; and in biochemistry, see Voit, 2000). Statistical inference for ODEs is not a trivial task, because numerical evaluation of standard estimators, like the maximum likelihood or the least squares estimators, may be difficult or computationally costly. Therefore, over the last few decades, first in the numerical analysis and mathematical biology literature and lately also in the statistical literature, various alternative, primarily non-parametric smoothing-based methods have been proposed to tackle the problem, see, e.g. Bellman and Roth (1971), Varah (1982), Voit and Savageau (1982), Ramsay *et al.* (2007), Hooker (2009), Hooker *et al.* (2011), Gugushvili and Klaassen (2012), Campbell and Lele (2014), Vujačić *et al.* (2015), Dattner (2015), Dattner and Klaassen (2015), among others. These techniques typically share the property of being computationally simpler, but often also statistically less efficient than the maximum likelihood or the least squares methods.

The ODE systems that we have in mind take the form

$$\begin{cases} x'(t) = F(x(t), \theta), & t \in [0, 1], \\ x(0) = \xi \end{cases} \tag{1}$$

where $x(t) = (x_1(t), \ldots, x_d(t))^{tr}$ is a d-dimensional state variable, $\theta = (\theta_1, \ldots, \theta_p)^{tr}$ denotes a p-dimensional parameter, while the column d-vector $x(0) = \xi$ defines the initial condition. We define $\eta := (\xi, \theta)$ and denote the solution to Equation 1 corresponding to the parameter η by

$$x(\eta, t) := (x_1(\eta, t), \ldots, x_d(\eta, t))^{tr}.$$

Knowledge regarding the system parameters ξ and θ is of vital importance for the study of a process that Equation 1 models. Indeed, these parameters affect the qualitative properties of the system, and their knowledge allows one to predict the system behaviour. However, in practice, the parameter θ and possibly also the initial condition ξ are unknown to the researcher. Typically, they cannot be measured directly but have to be inferred from noisy measurements of the process under study.

Let $\eta_0 = (\xi_0, \theta_0)$ be the 'true' parameter value that governs the underlying process. The common statistical model considered for the noisy measurements of the process at time instances $t_1, \ldots, t_n$ (not necessarily equally spaced) is the additive measurement error model,

$$Y_{ij} = x_i(\eta_0, t_j) + \epsilon_{ij}, \quad i = 1, \ldots, d, j = 1, \ldots, n, \tag{2}$$

where the random variables ϵ_{ij} are independent measurement errors (not necessarily Gaussian). Based on observation pairs (t_j, Y_{ij}), $i = 1, \ldots, d, j = 1, \ldots, n$, the goal is to estimate the parameter η_0.

A classical approach to parameter estimation for ODEs is the non-linear least squares (NLS) method. Its use is based on the observation that the problem at hand in its essence is a non-linear regression problem, where the regression function

$x(\eta, \cdot)$ is defined implicitly as the solution to Equation 1. The least squares estimator $\tilde{\eta}_n = (\tilde{\xi}_n, \tilde{\theta}_n)$ of η_0 is defined as a minimizer of the least squares criterion function $R_n(\cdot)$,

$$\tilde{\eta}_n = (\tilde{\xi}_n, \tilde{\theta}_n) = \operatorname{argmin}_\eta \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\eta, t_j))^2 \tag{3}$$

$$=: \operatorname{argmin}_\eta R_n(\eta).$$

The strongest justification for the use of the least squares estimator lies in its attractive asymptotic properties; see, e.g. Jennrich (1969) and Wu (1981). In most practical applications, the solution $x(\eta, \cdot)$ to Equation 1 is non-linear in the parameter η, and therefore, some iterative procedure has to be used to compute the NLS estimator. Such procedures require an initial guess for a minimizer $\tilde{\eta}_n$ and, then, proceed by constructing successive approximations to the least squares estimator (in a direction guided by the gradient of the criterion function, when a gradient-based optimization method, e.g. the Levenberg–Marquardt method, is used). However, the noisy and non-linear character of the optimization problem may lead for the procedure to end up in a local minimum of the least squares criterion function, especially when good initial guesses of the parameter values are not available. Furthermore, in most of the interesting applications, the system (Equation 1) is non-linear and does not have a closed-form solution. In that case, at every step of the iterative procedure, one has to numerically integrate Equation 1 (as well as the system of the associated sensitivity equations in order to compute the gradient of the criterion function, in case a gradient-based optimization method is used). Because the number of iterations made until convergence of the algorithm can be ascertained is usually large, in most cases, this leads to a computational bottleneck. This is the case especially in mathematical biology and biochemistry, where a highly non-linear character of dependence of the solution $x(\eta, \cdot)$ on the parameter η leads to 'stiff' integration problems. For a penetrating discussion of these points, see, e.g. Ramsay *et al.* (2007) and Voit and Almeida (2004).

Although NLS algorithms and ODE integration routines are constantly improving, and so is the available computational power, admittedly much time and effort can be saved with alternative, less computationally intense approaches, see Voit and Almeida (2004). In this paper, we explore application of Le Cam's one-step estimator (see, e.g. van der Vaart, 1998) to parameter estimation for systems of ODEs. Some examples of similar studies in different areas are Bickel (1975), Simpson *et al.* (1992), Field and Wiens (1994), Cai *et al.* (2000), Delecroix *et al.* (2003) and Rieder (2012). In particular, our main goal is to show that the one-step method is at least comparable with NLS, first asymptotically and second in finite samples. We would like to stress the fact that the one-step method is not simply a numerical approximation to an algorithm used for numerical evaluation of NLS: It is an estimation method on its own.

The main contributions of our paper are: (i) Smoothing-based parameter estimation methods for ODE systems can be upgraded to have statistical efficiency of NLS through a computationally simple one-step method. (ii) If one wants to avoid using NLS (as is often the case in the applied literature, see, e.g. Stein *et al.*, 2013 and Bucci

et al., 2016), one can still do this, while not losing statistical efficiency of NLS and computational properties of smoothing-based methods. (iii) We show how to perform smoothing in a data-driven manner and provide theory supporting our data-driven algorithm. (iv) We point out a very simple scheme for implementing the one-step estimator, which is readily available in any software that implements Newton-type optimization algorithms, such as R, see R Core Team (2017), and Matlab, see The Mathworks, Inc. (2017).

Pertaining to point (i) mentioned earlier, we highlight the extent of loss of efficiency of smoothing-based methods compared with the NLS and the one-step method, which in some simulation setups is of alarming degree. With high throughput, dense-in-time data, that is becoming increasingly available in practice, specifically in molecular biology (see Voit and Almeida, 2004; Goel *et al.*, 2008), and that would allow an in-depth study of underlying biological processes, such a statistical efficiency loss is clearly undesirable. On the other hand, current ODE inference algorithms must also meet challenges with massive amounts of data and complex models awaiting in the near future. Pertaining to point (ii), as noted in Chou and Voit (2009), that far no parameter estimation technique for ODEs has arisen as a clear winner in terms of efficiency, robustness and reliability in realistic data scenarios. In this sense, addition of the one-step method (that shares some of the better properties of both the smoothing-based methods and NLS) to a practitioner's toolbox appears a sensible option. Concerning (iii), we note that much of the literature dealing with smoothing-based inference methods for ODEs in practice does smoothing either in a theoretically suboptimal or even an ad hoc way. A distinct advantage of our proposed approach is providing theoretical guarantees for data-depending smoothing that our procedure employs as an intermediate step. Finally, concerning our contribution (iv), we point out an important relation between the one-step estimator and the Levenberg–Marquardt algorithm, which leads to a very practical and straightforward implementation of the method: When computational time is an issue, our simulations and theory justify the use of the Levenberg–Marquardt method with *one iteration*, provided it is initialized at an appropriate smoothing-based parameter estimator, because this reduces to the one-step estimation framework.

The rest of the paper is organized as follows: in Section 2, we describe the one-step estimator in the context of ODEs. In Section 3, we provide theoretical results for it. Section 4 presents a detailed simulation study illustrating the performance of the one-step method, with further examples in Section 5, while Section 6 contains numerical results based on real data examples. Section 7 summarizes our contribution and outlines potential future research directions. Finally, Appendices give a proof of our theoretical result and some further implementational details on the methods in the main text of the paper.

2 One-step estimate for ordinary differential equations

When one adopts an asymptotic point of view on statistics, all the estimators with the same asymptotic variance can be considered as equivalent. We now demonstrate how once a preliminary $\sqrt{n}$-consistent estimator $\hat{\eta}_n$ of the parameter η is available (see

succeeding text for our choice), one can obtain an asymptotically equivalent estimator to the least squares estimator in just one extra step, referred to as the one-step method in the statistical literature, see, e.g. section 5.7 in van der Vaart (1998) for the motivation behind it and a detailed exposition.

Introduce the function

$$\Psi_n(\eta) = \sum_{j=1}^{n} \psi_\eta(t_j, Y_j), \tag{4}$$

where

$$\psi_\eta(t, y) = (x'_\eta(\eta, t))^{tr}(y - x(\eta, t)), \tag{5}$$

with $x'_\eta(\eta, t)$ denoting the derivative of $x(\eta, t)$ with respect to η. Specifically, the ith row of $x'_\eta(\eta, t)$ is the gradient of $x_i(\eta, t)$ with respect to η.

The one-step estimator $\bar{\eta}_n$ of η_0 is defined as a solution in η of the equation

$$\Psi_n(\hat{\eta}_n) + \frac{d}{d\eta}\Psi_n(\hat{\eta}_n)(\eta - \hat{\eta}_n) = 0.$$

If $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$ is invertible, the estimator $\bar{\eta}_n$ can be expressed as

$$\bar{\eta}_n = \hat{\eta}_n - \left(\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)\right)^{-1}\Psi_n(\hat{\eta}_n). \tag{6}$$

In order to implement the estimator just defined, the two essential steps that have to be done are (i) evaluation of a preliminary estimator $\hat{\eta}_n$, and (ii) evaluation of $\Psi_n(\hat{\eta}_n)$ and the derivative matrix $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$. The computational cost for that is very modest. Indeed, as mentioned in Section 1, step (i) is very fast, when a smoothing-based estimator is used, see examples in the succeeding text. Furthermore, step (ii) reduces to requiring just one numerical integration of the sensitivity and variational equations associated with the system (Equation 1), as we will now explain. This material is standard in the numerical analysis and ODE literature (cf. Schittkowski, 2002; Ramsay and Hooker, 2017) but perhaps less familiar to statisticians, hence our decision to provide full details. It is helpful to think of F in Equation 1 as a function of η rather than only θ. Thus, we write the right-hand side F of Equation 1 as $F(x(\eta, t), \eta)$. Differentiating both sides of Equation 1 with respect to η and interchanging the order of a t-derivative with an η-derivative, we get

$$\begin{cases} \frac{d}{dt}\frac{\partial}{\partial\eta}x(\eta, t) = F'_x(x(\eta, t), \eta)\frac{\partial}{\partial\eta}x(\eta, t) + F'_\eta(x(\eta, t), \eta), \\ \frac{\partial}{\partial\eta}x(\eta, 0) = (1, 0)^{tr}, \end{cases} \tag{7}$$

where 1 and 0 in the initial conditions here and in Equations 8–9 in the succeeding text should be understood as vectors of 1's and 0's of the appropriate dimensions. The system (Equation 7) is a matrix differential equation and is usually referred to in the

literature as a system of sensitivity equations. By replacing η with $\hat{\eta}_n$, we arrive at the system

$$\begin{cases} \frac{d}{dt}s(t) = F'_x(x(\hat{\eta}_n,t),\hat{\eta}_n)s(t) \ + \ F'_\eta(x(\hat{\eta}_n,t),\hat{\eta}_n), \\ s(0) \ = \ (1,0)^{tr}, \end{cases} \tag{8}$$

where we have defined $s(t) := \frac{d}{d\eta}x(\hat{\eta}_n,t)$. Observe that $x(\hat{\eta}_n,\cdot)$ is a known function, because it can be found by integrating Equation 1 for parameter values $\hat{\xi}_n$ and $\hat{\theta}_n$. Consequently, the system of sensitivity equations is a linear system with time-dependent coefficients and, hence, is relatively straightforward to integrate.

By differentiating Equation 7 one more time with respect to η and replacing η with $\hat{\eta}_n$, we arrive at the following set of variational equations (sometimes called second-order sensitivity equations):

$$\begin{cases} \frac{d}{dt}z(t) \ = \ F''_{\eta\eta}(x(\hat{\eta}_n,t),\hat{\eta}_n)) \ + \ F''_{\eta x}(x(\hat{\eta}_n,t),\hat{\eta}_n)s(t) \\ + \ \left\{ F''_{x\eta}(x(\hat{\eta}_n,t),\hat{\eta}_n) \ + \ F''_{xx}(x(\hat{\eta}_n,t),\hat{\eta}_n)s(t) \right\} s(t) \\ + \ F'_x(x(\hat{\eta}_n,t),\hat{\eta}_n)z(t), \\ z(0) \ = \ 0, \end{cases} \tag{9}$$

where we have set $z(t) := \frac{\partial^2}{\partial\eta^2}x(\eta,t)$. For each $z_i,\ i \ = \ 1,\dots,d$, the system (Equation 9) is a matrix differential equation and again is a linear system with time-varying coefficients. Here also, we can treat x and s as known, for they can be obtained through numerical integration of Equations 1 and 8. The process of obtaining variational equations can be made automatic through a software implementation.

Integration of Equations 1, 8 and 9 for the parameter value $\hat{\eta}_n$ allows us to compute $\Psi_n(\hat{\eta}_n)$ and $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$ and, consequently, the one-step estimator $\bar{\eta}_n$. Note that numerical integration of the variational equations (or at least the sensitivity equations) is usually required when computing the least squares estimator via gradient-based optimization methods (unless the gradient is available analytically). However, in our approach, we need to do this only *once*.

REMARK 1. A seemingly more general non-autonomous system than the autonomous system (Equation 1),

$$\begin{cases} \tilde{x}'(t) = F(\tilde{x}(t),t,\theta), \ t \in [0,1], \\ \tilde{x}(0) = \tilde{\xi}, \end{cases}$$

may and will be reduced to Equation 1 by a simple substitution $x(t) \ = \ (\tilde{x}^{tr}(t),t)^{tr}, \ t \in [0,1]$, and $\xi \ = \ (\tilde{\xi}^{tr},0)^{tr}$.

3 Theory for the one-step method

The one-step estimation methodology described in the previous section requires the user to first obtain a preliminary $\sqrt{n}$-consistent estimator of parameter of interest.

Obviously, one would like such an estimator to be cheap in computational cost. In the context of ODEs, such preliminary estimators were suggested in Bellman and Roth (1971) and Varah (1982), who use non-parametric smoothing techniques to bypass numerical integration of the ODEs required in evaluation of the maximum likelihood or the least squares estimators. This approach was studied rigorously from the theoretical point of view in Gugushvili and Klaassen (2012) (other relevant references are, e.g. Brunel, 2008; Dattner and Klaassen, 2015; Vujačić *et al.*, 2015). As mentioned, such methods use non-parametric smoothing techniques, and therefore, their good performance crucially depends on an appropriate choice of a 'tuning parameter', such as the bandwidth in the case of kernel smoothing, or the number of basis functions in the case of splines. Moreover, this dependence on the bandwidth choice propagates to performance of the one-step estimator. In this section, we describe one of the possible preliminary estimators, provide a data-driven scheme for the choice of the tuning parameter and derive the relevant theory for the one-step method.

The preliminary estimation works as follows. The observations are first smoothed, which results in an estimator $\hat{x}_n(\cdot)$ for the solution $x(\eta_0, \cdot)$ of the system, and by differentiation, in an estimator $\hat{x}'_n(\cdot)$ for $x'(\eta_0, \cdot)$. Then the estimator for θ_0 is defined as the minimizer $\hat{\theta}_n$ over θ of the function

$$\int_0^1 \| \hat{x}'_n(t) - F(\hat{x}_n(t); \theta) \|^2 \, w(t) \, \mathrm{d}t, \tag{10}$$

where w is an appropriate weight function and $\| \cdot \|$ denotes the standard Euclidean norm. Hence, this approach bypasses the need to integrate the system numerically, and as a result, the parameter estimates can be computed extremely quickly, especially when F in Equation 1 is linear in θ. Under regularity conditions, Gugushvili and Klaassen (2012) show that this *smooth and match estimator* (SME) $\hat{\theta}_n$ has the $\sqrt{n}$-rate of convergence to θ. By the general statistical theory, the $\sqrt{n}$-rate of convergence is in fact the best rate one can expect in the present context. This result thus puts the smooth and match method on a solid theoretical ground.

Note that execution of this method does not require the knowledge of the initial values in Equation 1. However, it cannot be used to estimate them. If estimation of initial values is of interest, then once the estimator $\hat{\theta}_n$ is at hand, one may obtain an estimator $\hat{\xi}_n$ by minimizing with respect to ξ the criterion

$$\int_0^1 \| \hat{x}_n(t) - \xi - \int_0^t F(\hat{x}_n(s); \hat{\theta}_n) \mathrm{d}s \|^2 \, \mathrm{d}t.$$

Notice that this is a linear least squares optimization problem and hence is easy to execute.

Actually, approaches as mentioned earlier are criticised for not being *statistically efficient*. In informal terms, this means that the resulting estimators do not squeeze as much information out of the data as the least squares estimator does. In more formal terms, their asymptotic variance is larger than that of the least squares estimator. Hence, sometimes it is suggested (see, e.g. Swartz and Bremermann, 1975, for an early

reference) to use this method only for generating preliminary estimates that should be used later as initial guesses for more accurate methods. Thus, the SME described earlier is a natural candidate for serving as a preliminary estimator to be used by the one-step method. Now, we describe our data-driven methodology for choosing the tuning parameter.

Let $\hat{\eta}_{\rho_n}$ denote an estimator of the ODE parameter η_0 that depends on smoothing parameter ρ_n (we make the dependence on the sample size n explicit in our notation). As one specific example, $\hat{\eta}_{\rho_n}$ may be a smooth and match or an integral estimator (see Appendix B), in which case ρ_n is the bandwidth h_n. Alternatively, ρ_n may also stand for the number of basis functions. Now, consider two sequences of positive numbers $\underline{R}_n \leq \bar{R}_n$ that for every n define an interval $\mathcal{R}_n = [\underline{R}_n, \bar{R}_n]$. This will be an interval in which a user selects his or her smoothing parameter (in a data-dependent way), when the sample size is equal to n. More specifically, let N be an arbitrary fixed positive integer. For every n, consider a grid of size N of smoothing parameters in $\mathcal{R}_n$:

$$R_n = \{\rho_n(k) \in \mathcal{R}_n, k = 1, \dots, N\}.$$

Here, k indexes smoothing parameter values contained in the candidate set R_n of smoothing parameter values available to a user.

Now, a data-driven one-step estimator can be defined through the following procedure:

1. Compute N preliminary estimators $\hat{\eta}_{\rho_n(k)}$ for $\rho_n(k) \in R_n$.
2. Compute N one-step estimators $\bar{\eta}_n = \bar{\eta}_n(\hat{\eta}_{\rho_n(k)})$.
3. Set

$$\bar{\eta}_n^* = \operatorname{argmin}_{\bar{\eta}_n(\hat{\eta}_{\rho_n(k)})} \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\bar{\eta}(\hat{\eta}_{\rho_n(k)}), t_j))^2. \tag{11}$$

In Section 4, we demonstrate that this procedure results in an excellent practical performance of the estimator $\bar{\eta}_n^*$. In the theorem below, we show that it has a sound theoretical basis as well.

THEOREM 1. Assume that the following conditions hold true:

(i) Observation times $t_1, \dots, t_n$ are i.i.d. with a distribution function F_T supported on the interval $[0, T]$.
(ii) Measurement errors ϵ_{ij}'s are i.i.d. with mean zero and variance $\sigma^2 > 0$, and are also independent of observation times t_j's.
(iii) The parameter set H is a compact subset of $\mathbb{R}^{d+p}$.
(iv) For all $\eta \in H$ and $t \in [0, T]$, the third partial derivatives $x'''_{\eta_j \eta_k \eta_l}(\eta, t)$ of the ODE solution $x(\eta, t)$ exist and are continuous functions of η and t.
(v) The matrix

$$I(\eta) = \frac{1}{\sigma^2} \sum_{i=1}^{d} \int_0^T \left(\frac{d}{d\eta} x_i(\eta, t) \right)^{tr} \left(\frac{d}{d\eta} x_i(\eta, t) \right) \mathrm{d}\, F_T(t) \tag{12}$$

is non-degenerate.

(vi) For every choice of a deterministic sequence of smoothing parameters $\rho_n \in \mathcal{R}_n$, the resulting estimator $\hat{\eta}_{\rho_n}$ is $\sqrt{n}$-consistent.

Then

$$\sqrt{n}(\bar{\eta}_n^* - \eta_0) \xrightarrow{d} \mathcal{N}\left(0, I(\eta_0)^{-1}\right), \tag{13}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

REMARK 2. Under conditions of Theorem 1, the limit covariance matrix in Equation 13 coincides with the limit covariance matrix of the least squares estimator; compare with Example 5.27 in van der Vaart (1998). In our examples, the distribution function F_T will typically be uniform on $[0, T]$.

REMARK 3. For a smooth and match or an integral estimator, $\sqrt{n}$-consistency for any deterministic choice of the bandwidth $h_n \in R_n$ can be achieved, e.g. by taking $\underline{R}_n = \underline{c}n^{-\underline{r}}, \bar{R}_n = \bar{c}n^{-\bar{r}}$, for suitably chosen constants $\underline{c}, \bar{c}, \bar{r}, \underline{r} > 0$. Certain freedom in their choice is in fact allowed. As a specific example, the theoretical analysis of Dattner and Klaassen (2015) shows that in order to have the $\sqrt{n}$-rate for the integral estimator, one should take a bandwidth $b = O(n^{-1/3})$. Thus, in our practical implementation in subsequent sections, we set $B = n^{-1/3} \times (c_1, \ldots, c_N)$, where the c_j's depend on the grid of points on which we evaluate the kernel estimator.

REMARK 4. The one-step method as described in Section 2 requires evaluation of the second derivative $x''_{\eta\eta}(\eta, t)$ of the ODE solution $x(\eta, t)$ as part of evaluation of the matrix $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$. A standard argument, compare with pp. 71–72 in van der Vaart (1998), shows, however, that Theorem 1 still holds true if in the definition of the one-step estimator $\bar{\eta}_n$ in Formula 6, the matrix $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$ is replaced by the matrix

$$-\sum_{j=1}^{n} (x'_{\eta}(\hat{\eta}_n, t_j))^{tr} x'_{\eta}(\hat{\eta}_n, t_j). \tag{14}$$

This version of the one-step method is useful when large numerical errors or numerical instability are expected when evaluating $x''_{\eta\eta}(\eta, t)$. A further refinement is to employ damping and to replace the derivative matrix $\frac{d}{d\eta}\Psi_n(\hat{\eta}_n)$ with

$$-\sum_{j=1}^{n} (x'_{\eta}(\hat{\eta}_n, t_j))^{tr} x'_{\eta}(\hat{\eta}_n, t_j) - \lambda_n I,$$

where $\lambda_n > 0$ is a damping parameter and I is an identity matrix of appropriate dimensions. The assumption for the asymptotic theory to go through is that $\lambda_n/n \to 0$ as $n \to \infty$. The idea of this version of the one-step method is that it numerically robustifies the one-step procedure in case the matrix (Equation 14) is nearly singular

(which is not uncommon in practice). We use this version of the one-step method in our simulation example in Section 5.

3.1 Confidence intervals

Clearly, confidence intervals for parameter η_0 can be generated using Equations 12 and 13. However, the Fisher information matrix in Equation 12 depends on the true values of the parameters, initial values and σ^2, which are not known in practice. Fully data-driven confidence intervals can be constructed by estimating the Fisher information matrix. To that end, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{d(n-1)} \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\bar{\eta}_n^*, t_j))^2,$$

where $x(\bar{\eta}_n^*, \cdot)$ stands for the solution of the system (Equation 16) using the estimated parameters and initial values obtained from the one-step method. Then, an estimate for the asymptotic variance of the estimator of the parameter η_j is given by $I_{jj}^{-1}(\bar{\eta}_n^*)/n$, where $I_{jj}^{-1}(\bar{\eta}_n^*)$ stands for the jth diagonal element of the inverse Fisher information matrix evaluated in point $\bar{\eta}_n^*$. When $s(\cdot)$ has no closed form, the integral in Equation 12 is evaluated using numerical integration (in our examples, we will use the trapezoidal rule). Specifically, an approximate $1 - \alpha$ level confidence interval for η_{0j} is given by

$$[\bar{\eta}_{j,n}^* - z_{1-\alpha/2} I_{jj}^{-1/2}(\bar{\eta}_{j,n}^*)/\sqrt{n}, \bar{\eta}_n^* + z_{1-\alpha/2} I_{jj}^{-1/2}(\bar{\eta}_{j,n}^*)/\sqrt{n}], \tag{15}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

4 Simulation study

In this section, we present the results of an extensive simulation study comparing the one-step method with the classical NLS approach. The models that we use are standard test examples for parameter inference in ODEs, as indicated in the references that we will supply in the relevant places. Our goal is to exhibit that the one-step algorithm provides statistical accuracy comparable with the NLS method in practical scenarios.

All computations in the present section were carried out using Matlab (the code will be sent by the first author upon request). The algorithms that we used for computing the NLS and one-step estimators are 'default', in the sense that we did not attempt to tweak them to fit better in specific problems. Specifically, the NLS estimator was computed using the Levenberg–Marquardt (1963) algorithm of Matlab. The variant of SME $\hat{\eta}_n$ that we used in the present and next sections to compute the one-step estimator $\bar{\eta}_n$ is detailed in Appendix B. The local polynomial estimator in some of our examples was based on the implementation from Cao (2008). Further software and hardware details are Windows 8.1 Pro, Intel ® Core™ i7-4550U CPU @ 1.50 GHz (Santa Clara, California, USA).

4.1 Linear ordinary differential equation

We start with illustrating the performance of the one-step estimator when used to estimate the parameter and initial value of a one-dimensional linear ordinary differential equation

$$\begin{cases} x'(t) = \theta_0 x(t), \ t \in [0, T], \\ x(0) = \xi_0. \end{cases} \tag{16}$$

This is a toy example, but it allows us to explore the practical performance of the one-step method in great detail and to compare it with the theoretically expected results. Advanced examples will be considered later on.

The solution of the initial value problem (Equation 16) is $x(t) = x_0 \exp(\theta_0 t)$. We generate (pseudo) random observations from the model

$$Y_j = \xi_0 \exp(\theta_0 t_j) + \epsilon_j,$$

where $t_j \in \{0(0.1)10\}$ ($n = 101$), and $\epsilon_j \sim N(0, 0.05^2)$, $j = 1 \dots, n$. We consider

$$\theta_0 \in \{-1, -0.8, -0.6, -0.4, -0.2, 0.1, 0.3, 0.5, 0.7, 0.9\}$$

and $\xi_0 \in \{0.5, 1\}$. For each pair (ξ_0, θ_0), we run a Monte Carlo study of 500 samples of $Y_1 \dots, Y_{101}$, where in each sample, we apply both the one-step method and the NLS method. This simulation study enables us to estimate the asymptotic variance of the least squares and the one-step methods. We then compare the results with the true asymptotic variance. The true and estimated asymptotic variances can be obtained for each set of parameters and initial values by inverting the Fisher information matrix (Section 3.1). The optimal bandwidth b used to compute SME was chosen in the set

$$n^{-1/3} \times (0.02, 0.3511, 0.6822, 1.0134, 1.3445, 1.6756, 2.0067, 2.3378, 2.6689, 3),$$

using the procedure outlined in Remark 3; compare with Theorem 1. We also note that in order not to overload the paper with reporting various tuning constants that depend on specific experimental setups, we will not indicate $c_1, \dots, c_N$ from Remark 3 in our subsequent examples but will supply them to the reader by email, should he or she want to know them.

A direct computation gives that in model (16), the asymptotic variance of $\bar{\bar{\xi}}_n$ depends on θ but is independent of the values of ξ itself. In Figure 1, we plot the estimated variance of the one-step estimators (plus signs) and that of the NLS (circles), for estimating ξ_0 based on 500 simulation runs. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The left plot is for $\xi_0 = 0.5$, and the right one is for $\xi_0 = 1$. As the theory suggests, independently of the values of ξ, the true asymptotic variance is the same. Note that in this specific numerical example, the estimated variances of the one-step and NLS estimators are the same. This is not surprising, because in order to apply the NLS, we used as the initial point in the parameter space the SME (resulted from using the bandwidth $3 \times n^{-1/3}$; this choice was arbitrary). The estimated
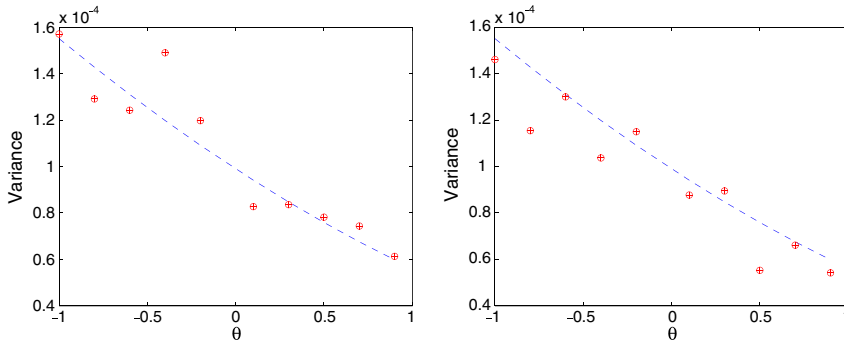
Fig. 1. The estimated variance of the one-step (plus signs) and non-linear least squares (circles) estimators $\bar{\xi}_n$ and $\tilde{\xi}_n$, respectively, based on 500 simulations with $n = 101$ and $\epsilon_j \sim N(0, 0.05^2)$, $j = 1\dots,n$. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The left plot is for $\xi_0 = 0.5$, and the right one is for $\xi_0 = 1$.

Fig. 2. The estimated variance of the one-step (plus signs) and non-linear least squares (circles) estimators $\bar{\theta}_n$ and $\tilde{\theta}_n$, respectively, based on 500 simulations with $n = 101$ and $\epsilon_j \sim N(0, 0.05^2)$, $j = 1\dots,n$. The estimates are superimposed on the theoretical asymptotic variance (dashed line). The left plot is for $\xi_0 = 0.5$, and the right one is for $\xi_0 = 1$.
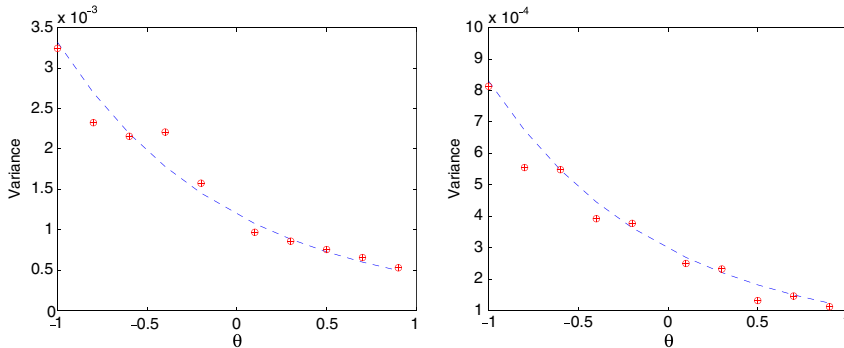
variances agree with the asymptotic one. We note that the grid of θ_0 does not include 0, where the asymptotic variance equals zero.

In Figure 2, we see similar plots corresponding to estimating the asymptotic variances of $\bar{\theta}_n$. Here, the variance has different order, depending on the value of ξ_0. Again, the estimated variances of the one-step (plus signs) and NLS (circles) estimators are the same, and both agree with the asymptotic one (dashed line). Similar plots were obtained when considering other values for σ^2, and therefore, we do not present them here.

In Table 1, we present the empirical coverage of various confidence intervals based on a Monte Carlo study with 500 simulations for different experimental setups. The results should be compared with the nominal coverage of 95%. We consider four setups denoted by A, B, C and D according to $(\xi_0 = 1/2, \theta_0 = -1), (\xi_0 = 1/2, \theta_0 = 1), (\xi_0 = 1, \theta_0 = -1), (\xi_0 = 1, \theta_0 = 1)$, respectively. Each

Table 1. Means of point estimates and actual coverage of interval estimates for the parameters of model (Equation 16) according to four different experimental setups. The results are based on 500 simulation runs. The observations are generated according to $Y_j = \xi_0 \exp(\theta_0 t_j) + \epsilon_j$, where $t_j \in \{0(0.5)10\}$ ($n = 21$), or $t_j \in \{0(0.2)10\}$ ($n = 51$) and $\epsilon_j \sim N(0, 0.05^2)$, $j = 1 \ldots, n$. The point estimates are given by Equation 6; the interval estimates are defined in Equation 15

				One-step		NLS	
Setup				Mean	Coverage	Mean	Coverage
n = 21	A	ξ_0	0.500	0.501	0.942	0.501	0.942
		θ_0	−1.000	−1.002	0.946	−1.002	0.946
	B	ξ_0	0.500	0.500	0.928	0.500	0.928
		θ_0	1.000	1.000	0.938	1.000	0.938
	C	ξ_0	1.000	0.999	0.932	0.999	0.932
		θ_0	−1.000	−0.997	0.940	−0.997	0.940
	D	ξ_0	1.000	1.000	0.944	1.000	0.944
		θ_0	1.000	1.000	0.948	1.000	0.948
n = 51	A	ξ_0	0.500	0.500	0.944	0.500	0.944
		θ_0	−1.000	−0.998	0.944	−0.998	0.944
	B	ξ_0	0.500	0.500	0.946	0.500	0.946
		θ_0	1.000	0.999	0.958	0.999	0.958
	C	ξ_0	1.000	0.999	0.932	0.999	0.932
		θ_0	−1.000	−0.999	0.938	−1.000	0.938
	D	ξ_0	1.000	1.000	0.948	1.000	0.948
		θ_0	1.000	1.001	0.952	1.001	0.952

Note: NLS, non-linear least squares.

scenario is tested for $n = 21$ and $n = 51$. Table 1 presents the point and interval estimates for the parameters of each scenario. We see that the coverage of the confidence intervals is satisfying across the different experimental scenarios.

4.2 Lotka–Volterra system

The Lotka–Volterra system of ODEs (Edelstein-Keshet, 2005) is a population dynamics model that describes evolution over time of the populations of two species, predators and their preys. The system takes the form

$$\begin{cases} x_1'(t) = \theta_1 x_1(t) - \theta_2 x_1(t)x_2(t), \\ x_2'(t) = -\theta_3 x_2(t) + \theta_4 x_1(t)x_2(t). \end{cases} \tag{17}$$

Here, x_1 represents the size of the prey population and x_2 of the predator population. In Table 2, we see the empirical coverage of the 95% confidence intervals based on a Monte Carlo study consisting of 500 simulation runs for different sample sizes.

The experimental setup is as follows: The observed time points are equidistant on $[0, 10]$; the errors are normal with zero mean and standard deviation $\sigma = 0.05$; the initial values are $\xi_0 = (1, 1/2)^{tr}$, and the parameters are $\theta_0 = (1/2, 1/2, 1/2, 1/2)^{tr}$. The point estimates are given by Equation 6, while the interval estimates are defined in Equation 15. As expected, the coverage is much better when the sample size is larger. The performance of the one-step and NLS methods is similar.

In Table 3, we present the square root of the average of the estimates of the asymptotic variance over the 500 simulations (denoted by 'ASYM'). Next to that, we present

Table 2. Means of point estimates and actual coverage of interval estimates for the parameters of model (17), where the initial values are $\xi_0 = (1, 1/2)^{tr}$ and the rate parameters are $\theta_0 = (1/2, 1/2, 1/2, 1/2)^{tr}$. The results are based on running 500 simulations. The observed time points are equidistant on $[0, 10]$, and the errors are normal with zero expectation and $\sigma = 0.05$. The one-step point estimates are given by Equation 6; the interval estimates are defined in Equation 15

Setup			One-step		NLS	
			Mean	Coverage	Mean	Coverage
n = 21	ξ_1	1.000	1.000	0.932	0.999	0.928
	ξ_2	0.500	0.500	0.936	0.500	0.934
	θ_1	0.500	0.502	0.942	0.501	0.942
	θ_2	0.500	0.502	0.932	0.501	0.938
	θ_3	0.500	0.500	0.910	0.501	0.916
	θ_4	0.500	0.500	0.918	0.501	0.922
n = 51	ξ_1	1.000	1.000	0.958	1.000	0.966
	ξ_2	0.500	0.500	0.954	0.500	0.948
	θ_1	0.500	0.502	0.964	0.500	0.968
	θ_2	0.500	0.501	0.968	0.500	0.964
	θ_3	0.500	0.500	0.958	0.500	0.958
	θ_4	0.500	0.500	0.952	0.500	0.958

Note: NLS, non-linear least squares.

Table 3. Standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). Square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM'). The experimental setup is as in Table 2

Setup		SME	One-step		NLS	
		STE	STE	ASYM	STE	ASYM
n = 21	ξ_1	0.033	0.025	0.023	0.025	0.023
	ξ_2	0.022	0.020	0.019	0.020	0.019
	θ_1	0.030	0.027	0.026	0.027	0.026
	θ_2	0.024	0.022	0.021	0.022	0.021
	θ_3	0.025	0.022	0.020	0.022	0.020
	θ_4	0.021	0.020	0.018	0.020	0.018
n = 51	ξ_1	0.021	0.015	0.016	0.014	0.016
	ξ_2	0.014	0.013	0.013	0.013	0.013
	θ_1	0.019	0.016	0.017	0.016	0.017
	θ_2	0.015	0.013	0.014	0.013	0.014
	θ_3	0.014	0.013	0.013	0.013	0.014
	θ_4	0.013	0.012	0.012	0.012	0.012

Note: NLS, non-linear least squares; SME, smooth and match estimator.

standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). The results for both the NLS and one-step methods agree with each other. Note also the first column of this table, where we report the standard errors of the SME, which are larger than those of the one-step, as expected. In this experimental setup, the loss of statistical efficiency of SME in comparison with the one-step method and NLS is relatively small, given moderate sample sizes ($n = 21$ and $n = 51$). See, however, the next subsection.

Table 4.   The rows in the table correspond (respectively) to the means of point estimates, Monte Carlo empirical standard deviation, means of estimated asymptotic standard deviation, true asymptotic standard deviation and actual coverage of interval estimates (using the estimated asymptotic standard deviation) for the parameter $\theta_0 = 1$ in the linear ordinary differential equation case (Equation 16) (initial value $\xi_0 = 1$ was estimated as well). The results are based on 1000 Monte Carlo simulations. The observed time points are equidistant on [0, 10], and the errors are normal with zero expectation and σ as in the table. The one-step point estimates are given by Equation 6; the interval estimates are defined in Equation 15

| n = 250 | | | n = 500 | | | n = 1000 | | |
$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$
1.0010	1.0010	0.9990	1.0000	1.0000	1.0010	1.0000	1.0000	1.0000
0.0130	0.0260	0.0400	0.0090	0.0190	0.0280	0.0070	0.0130	0.0200
0.0130	0.0280	0.0390	0.0090	0.0180	0.0290	0.0070	0.0130	0.0200
0.0130	0.0270	0.0400	0.0100	0.0190	0.0290	0.0070	0.0130	0.0200
0.9460	0.9540	0.9610	0.9540	0.9520	0.9530	0.9430	0.9550	0.9470

4.3 Comparison with other methods

The main theme of this paper is not to compare various parameter estimation methods for ODEs, but to show how a non-efficient estimation method such as SME can be improved statistically, to an efficient one, and to test its practical performance. Indeed, this point was demonstrated earlier by comparing the variance of the one-step estimator with that of the least squares, which is not considered as a competitor but serves as a 'gold standard' for efficient estimation. For completeness, however, we report results of a small scale comparison that can shed some additional light on the statistical effects of the one-step correction on SME. In Table 4, we present the results of a simulation study for several experimental setups of the linear ODE case (cf. Equation 16). The results should be compared with table 1 of Hall and Ma (2014), where a different variant of SME is studied. The one-step estimator is uniformly (over all experimental setups) better than the method developed in the aforementioned paper, even though unlike that work we estimate both the initial value and the parameter and, hence, have to deal with greater uncertainty. The reduction in standard error achieved by the one-step estimator over the SME is in the range of 30–50% in this example. Such an improvement of an efficient parameter estimation method over SME is not an isolated instance: Hall and Ma (2014) report results of a Monte Carlo comparison between their version of SME and the generalized smoothing (or profiling) approach of Ramsay *et al.* (2007) and find out that the latter produces twice as small standard errors for parameter estimates in a specific experimental setup in the FitzHugh–Nagumo model; this despite the fact that the SME in Hall and Ma (2014) relies on a fully observed FitzHugh–Nagumo model, whereas Ramsay *et al.* (2007) assume only one state variable out of two is measured. A lesson to be drawn from this discussion from the statistical efficiency point of view is that one should be very careful when using SME so as to fully utilize precious information contained in observations.

4.4 Computational times

We close this section by reporting one more comparison. Namely, we compare 'default' implementations of one-step and NLS with respect to computational time. Voit and Almeida (2004) consider a test example that was introduced in Robertson (1966) and point out that it is now frequently used as a benchmark for the efficiency of stiff solvers. The system is given by

$$
\begin{aligned}
x_1'(t) &= \theta_1 x_2(t) x_3(t) - \theta_2 x_1(t), \\
x_2'(t) &= \theta_2 x_1(t) - \theta_1 x_2(t) x_3(t) - \theta_3 (x_2(t))^2, \\
x_3'(t) &= \theta_3 (x_2(t))^2,
\end{aligned}
\tag{18}
$$

with initial values $\xi_0 = (1, 0, 0)^{tr}$ and parameters $\theta_0 = (10^4, 0.04, 3 \times 10^7)^{tr}$. We take the observational time interval to be (in seconds) $[0(0.5)10]$, implying that we have $n = 21$ equispaced observations at our disposal. The variance of the noise is set to be 0.01 times the mean values of the (true) solutions corresponding to the system just defined. The actual coverage of the confidence intervals for the parameters $(\theta_1, \theta_2, \theta_3)^{tr}$ for a nominal level of 95% and using the one-step and NLS estimator based on 100 Monte Carlo simulations was $(1, 0.97, 1)^{tr} \times 100\%$. The widths of the confidence intervals for one-step and NLS were comparable. A single evaluation of the one-step estimator took about 26 s on average, while that of the NLS took about 78 s.

However, one should keep in mind that a completely objective comparison of computational costs for various ODE inference techniques is hardly possible, as this depends on factors like software and hardware used, as well as the skill of the user in tailoring the methods to specific applications. Also, one cannot expect that a single best method (as far as the computational cost is concerned) will emerge across all possible experimental setups (different ODE systems, sample sizes, time scales and resolutions, noise levels).

5 Further comparison

In this section, we additionally study a notoriously difficult test example in parameter inference for ODEs. In particular, we illustrate the fact why it might be advantageous to use the one-step method instead of a 'default' implementation of NLS, such as the Levenberg–Marquardt algorithm in Matlab. Our take-home message is that overreliance on 'default' implementations of NLS estimation routines for ODEs is perhaps a strategy to be critically reconsidered. We also point out a very simple practical scheme for implementing the one-step method.

5.1 Goodwin's oscillator

Goodwin's oscillator, see Goodwin (1963), Goodwin (1965) and Griffith (1968), is a simple ODE system for modelling feedback control in gene regulatory mechanisms. Various versions of this model have been used as test examples for Markov Chain
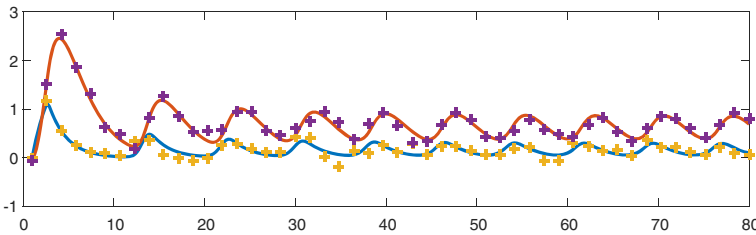
Fig. 3.   Components of the solution x_1 and x_2 of the system (19) (red and blue solid lines) with a typical realization of noisy observations (purple and yellow crosses).

Monte Carlo (MCMC) samplers in the Bayesian approach to inference in ODE models, see, e.g. Girolami (2008), Calderhead and Girolami (2009), Oates, Niederer, Lee, Briol, and Girolami (2017) and Oates, Papamarkou, and Girolami (2016). Standard Metropolis-Hastings samplers encounter severe difficulties in this setting because of a highly complex shape of the likelihood the Goodwin oscillator typically produces, with Markov chains getting trapped in local maxima of the likelihood surface. Not surprisingly, similar behaviour can be observed also in the case of default implementations of the least squares routines, as we will now demonstrate.

The following version of Goodwin's model is described, e.g. in Murray (2002), while the experimental setup mimics the one in Oates *et al.* (2016). The ODE system we consider is

$$
\begin{cases}
x_1'(t) &= \dfrac{\theta_1}{1 + \theta_2 x_3(t)^{10}} - \theta_5 x_1(t), \\
x_2'(t) &= \theta_3 x_1(t) - \theta_5 x_2(t), \\
x_3'(t) &= \theta_4 x_2(t) - \theta_5 x_3(t).
\end{cases}
\tag{19}
$$

We used the following parameter values,

$$\theta_1 = 1, \theta_2 = 3, \theta_3 = 2, \theta_4 = 1, \theta_5 = 0.5,$$

and zero initial conditions. Initial conditions and all the parameters except θ_1 and θ_5 were assumed to be known in the estimation problem. We compare the performance of the NLS and the one-step method through 100 Monte Carlo simulations for estimating the parameter $\theta = (\theta_1, \theta_5)^{tr}$. We consider the case when Equation 19 is observed only partially, with observations on x_3 not available; observed are the variables x_1, x_2 subject to additive Gaussian errors, with $n = 50$ noisy observations spread uniformly over the time interval [0, 80]. The solution to Equation 19 shows a characteristic oscillatory behaviour, and we plot it in Figure 3 together with corresponding observations in one simulation run.

We consider three scenarios corresponding to three noise levels $\sigma = 0.01, 0.15, 0.25$, respectively. It turned out that in this specific example, the version of the one-step method that we described in Section 3 in Remark 4 produced better results than the core one-step method from Section 2, so that we decided to perform a comparison of this version to a default implementation of the Levenberg–Marquardt method in

Matlab. Numerically, the one-step method in this case reduces to one iteration of the Levenberg–Marquardt algorithm, but with a difference that it is initialized at the $\sqrt{n}$-consistent preliminary parameter estimator and not an arbitrary initial guess. The default (starting) value for the damping parameter λ of the Levenberg–Marquardt algorithm in Matlab is $\lambda = 0.01$, which is also the one we used for the one-step method. Matlab successively increases the damping parameter until a proposed parameter move of the Levenberg–Marquardt method results in a decrease of the criterion function (the total number of proposals in one optimization run can be controlled by setting the maximal number of function evaluations for the algorithm). This then constitutes one iteration of the Levenberg–Marquardt method in Matlab.

We let the optimization for NLS to start from a random initial guess generated from a gamma distribution. Specifically, the initial guess for θ_1 is generated from a gamma
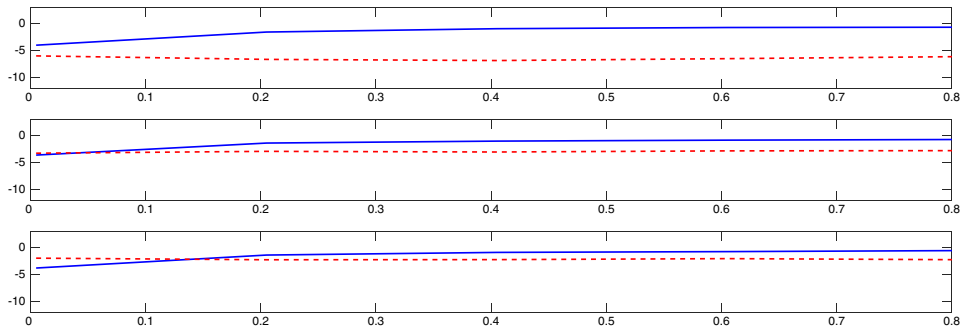

Fig. 4. Simulation results for Goodwin's oscillator in Section 5. The *y*-axis gives the logarithm of the sum of mean square errors of parameter estimates (non-linear least squares results plotted with a solid line, the one-step method ones with a dashed line). The noise level is $\sigma = 0.01, 0.15, 0.25$ in the upper, middle and bottom plots, respectively. The *x*-axis is the scale parameter of the gamma distribution used to generate initial guesses for non-linear least squares, with large values corresponding to initial guesses farther away from the true parameters values.
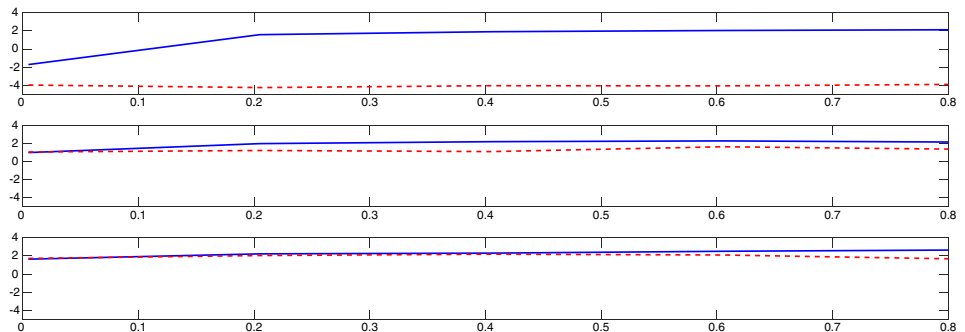

Fig. 5. Simulation results for Goodwin's oscillator in Section 5. The *y*-axis gives the logarithm of the sum of squares of model fits (non-linear least squares results plotted with a solid line, the one-step method ones with a dashed line). The noise level is $\sigma = 0.01, 0.15, 0.25$ in the upper, middle and bottom plots, respectively. The *x*-axis is the scale parameter of the gamma distribution used to generate initial guesses for non-linear least squares, with large values corresponding to initial guesses farther away from the true parameters values.

distribution with shape parameter $\theta_1/scale$, where the scale parameter is according to the x-axis of Figures 4–5, and similarly for θ_5, the shape will be $\theta_5/scale$. The one-step method, on the other hand, employs the $\sqrt{n}$-consistent estimator, namely, the direct integral estimator (although the system (19) we consider is partially observed, the direct integral approach still applies, as we explain in Appendix C). In Figure 4, we plot on y-axis the logarithm of the sum of mean square errors of parameter estimates (over 100 Monte Carlo simulation runs): NLS with a solid line, the one-step estimator with a dashed line. The noise level is $\sigma = 0.01, 0.15, 0.25$ in the upper, middle and bottom plots, respectively. The x-axis gives the scale parameter of the gamma distribution used to generate initial guesses for NLS; large values of the scale parameter correspond to a diffuse prior information on the true parameters, with initial guesses likely to be farther away from the true parameter values. In Figure 5, we show a similar setup, where now the y-axis gives the logarithm of the sum of squares of model fits (averaged over 100 Monte Carlo simulation runs).

We can see that the mean square error and the sum of squares of NLS grow together with the distance of the initial guess from the true parameter. For initial guesses close to the true parameter values, the NLS does better than the one-step method, but starts to deteriorate very quickly. Because in practice, informative prior information on true parameters is rarely available, we conclude that the one-step method is in general better in terms of both the mean square error of parameter estimates and the sum of squares of model fits than the NLS initialized at a random initial guess. This despite the fact that we allowed the Levenberg–Marquardt implementation of NLS to run for 100 iterations, while for the one-step method, we used only one iteration (as its name actually suggests). From the plots, we also see that larger the measurement error is, the more similar the two methods are in terms of the mean square error and the sum of squares. This is not surprising, because for large noise level, the direct integral estimator used as an initial input for the one-step estimator will be farther away from the true parameter (as any other estimator), and hence, the numerical performance of the one-step method will start to resemble that of the NLS initialized at a guess that is far from the true parameter.

We finally remark that the pattern observed in this low-dimensional simulation example (three-dimensional system with two unknown parameters) will readily extend to the case of more complex and realistic ODE models (depending on a particular experimental setup, in an even more pronounced form).

6 Real data examples

In this section, we study several real data examples. To check the limits of applicability of the one-step method, our emphasis is on examples with small and moderate sample sizes.

Table 5. Parameter estimates for model (Equation 20) obtained in the literature

Paper	Estimate of θ_1	Estimate of θ_2
Bellman *et al.* (1967)	0.4577×10^{-5}	0.2797×10^{-3}
Van Domselaar and Hemker (1975)	0.45×10^{-5}	0.27×10^{-3}
Esposito and Floudas (2000)	0.4593×10^{-5}	0.28285×10^{-3}
Kim and Sheng (2010)	0.46×10^{-5}	0.28×10^{-3}
Tjoa and Biegler (1991)	0.4604×10^{-5}	0.2847×10^{-3}
Varah (1982)	0.46×10^{-5}	0.27×10^{-3}

6.1 Nitrogene oxide reaction

The system

$$\begin{cases} x'(t) = \theta_1(126.2 - x(t))(91.9 - x(t))^2 - \theta_2(x(t))^2, \\ x(0) = 0 \end{cases} \tag{20}$$

describes the reversible homogeneous gas phase reaction of nitrogene oxide,

$$2NO + O_2 \rightleftharpoons 2NO_2.$$
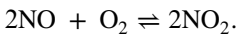
For additional chemical background, see Bodenstein (1922). Based on the experimental data from Table 39 in Bodenstein (1922), parameters of Equation 20 were estimated via different methods in Bellman *et al.* (1967); Van Domselaar and Hemker (1975), see pp. 18–19; Esposito and Floudas (2000), Section 7.4; Kim and Sheng (2010); Section 3.1; Tjoa and Biegler (1991), Problem 6 on p. 381; and Varah (1982), see pp. 37–38. The results obtained in these papers are summarized in Table 5.[†] We also remark that this problem is one of the six test problems in parameter estimation for ODEs that were included in Floudas *et al.* (1999).

Our interest in this example first went in the following direction: We used the realistic estimated parameter values from the literature, generated an artificial set of data from Equation 20 and checked how well the one-step estimator performs in this case. We also present the estimation results using the NLS estimator. Accordingly, we took the parameter estimates $\theta_1 = 0.4577 \times 10^{-5}$ and $\theta_2 = 0.2797 \times 10^{-3}$ from Bellman *et al.* (1967) together with the initial condition $\xi = x(0) = 0$, thus $\eta_0 = (\xi, \theta_1, \theta_2)^{tr}$. Then we generated observations uniformly over $t_j \in \{0(2)40\}$, ($n = 21$), according to Equation 2, where the i.i.d. measurement errors ϵ_j were generated from the normal distribution $N(0, \sigma^2)$ with mean zero and variance $\sigma^2 = 0.25$.

This setup was chosen to mimic the real data scenario related to this model, as described later on. The fact that θ_1 and θ_2 are small numbers, combined with the fact

---

[†]Note that Varah (1982) gives five different parameter estimates corresponding to different values of the smoothing parameter used in his method. Of these estimates, we report only the first pair and refer to Table 4 in Varah (1982) for the remaining ones. Note also that Esposito and Floudas (2000) use two approaches (collocation method and integration method in their terminology) and with the second of them identify another local solution to the problem, namely, $\theta_1 = 0.1306 \times 10^{-2}, \theta_2 = 0.90393$ (see Table 11 in Esposito and Floudas, 2000), which we did not report in Table 5.

Table 6. Means of point estimates and actual coverage of interval estimates for the parameters of model (Equation 20), where the initial value ξ is zero and the parameters are $\theta_1 = 0.4577 \times 10^{-5}$ and $\theta_2 = 0.2797 \times 10^{-3}$. The results are based on 500 simulation runs. There are 21 observations given on a uniform grid on $[0, 40]$, and the errors are normal with zero expectation and $\sigma^2 = 0.25$. The one-step point estimates are given by Equation 6, while the interval estimates are defined in Equation 15

Setup			One-step		NLS	
			Mean	Coverage	Mean	Coverage
n = 21	ξ_1	0	1.491e-02	0.938	7.960e-03	0.942
	θ_1	4.577e-06	4.576e-06	0.954	4.577e-06	0.952
	θ_2	2.797e-04	2.788e-04	0.932	2.798e-04	0.930

Note: NLS, non-linear least squares.

Table 7. Point estimates for the parameters of model (Equation 20) based on the real data of table 39 in Bodenstein (1922). We consider the initial value to be zero. The one-step point estimates are given by Equation 6; the confidence intervals were generated according to Equation 15. The left and right interval points are denoted by CI(L) and CI(R), respectively

	One-step			NLS		
	Point	CI(L)	CI(R)	Point	CI(L)	CI(R)
θ_1	4.579e-06	4.255e-06	4.903e-06	4.577e-06	4.253e-06	4.901e-06
θ_2	2.791e-04	1.923e-04	3.658e-04	2.796e-04	1.928e-04	3.665e-04

Note: NLS, non-linear least squares.

that their magnitudes are rather different, renders their estimation a difficult task, compare with p. 1303 in Esposito and Floudas (2000). In Table 6, we see the empirical average of point estimates and the empirical coverage of interval estimates based on Monte Carlo study consisting of 500 runs. The point estimates are given by Equation 6, while the interval estimates are defined in Equation 15.

We note that when estimating $\theta = (\theta_1, \theta_2)$, unlike Bellman *et al.* (1967), Van Domselaar and Hemker (1975), Tjoa and Biegler (1991) and Varah (1982), we did not assume that the initial condition $x(0) = 0$ was known, but estimated it as well. Notice also that our method exploits linearity in the parameters, and therefore, it is not required to supply an initial guess in the parameter space (in Bellman *et al.*,1967, and other related papers the initial guesses $\theta_1 = 10^{-6}$ and $\theta_2 = 10^{-4}$ were used). We see that even with a small sample as 21 observations, the point and interval estimates are satisfying, and again, we do not observe a substantial difference between the one-step and NLS methods.

We next tested our approach on the real data for the model (Equation 20) given in Table 39 in Bodenstein (1922) and reproduced in Table I in Bellman *et al.* (1967). There are in total 14 observations available on the interval $[0, 39]$, excluding the initial condition $x(0) = 0$.[‡] This time, we did not estimate the initial condition and considered

---

[‡]Note that in Table 39 in Bodenstein (1922) and in Table I in Bellman *et al.* (1967), the observation 48.8 corresponding to the time instance $t = 19$ appears to contain a typo: We tentatively corrected it to 38.8. The same correction was applied in Table 24 in Esposito and Floudas (2000) and in Table 1 in Kim and Sheng (2010).
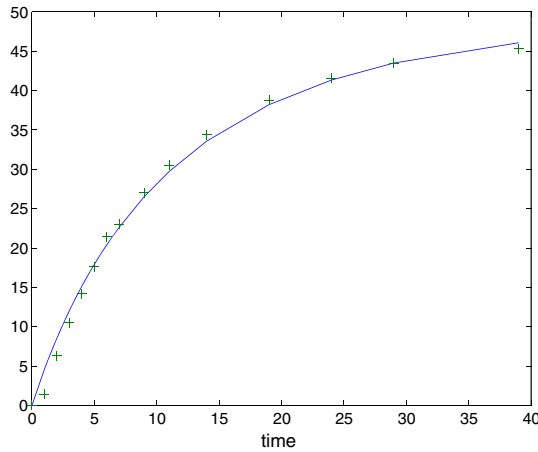
Fig. 6. The solution to Equation 20 (given by the solid line) and the observations (indicated by pluses). The parameters were estimated using the real data from Bellman *et al.* (1967). The initial value is considered to be known and equals zero.

it to be zero, which agrees with the physical phenomenon that the model describes. The estimation results are displayed in Table 7. Both point and interval estimates obtained from the one-step and NLS methods are presented.

A comparison with the results given in Table 5 shows that this is essentially the same result as already reported in the literature using the least squares estimator: This illustrates the fact that one-step is an asymptotically equivalent estimator to the least squares estimator, provided a preliminary estimator it uses is already within the $n^{-1/2}$ range of the true parameter. In Figure 6, we plot the data from Bellman *et al.* (1967) and the solution to Equation 20 evaluated with one-step fitted values of θ_1 and θ_2. The fit appears to be satisfactory given a simplistic character of the model (Equation 20).

6.2 α-Pinene problem

We now consider 'problem 8' of Tjoa and Biegler (1991). The system is given by

$$
\begin{aligned}
x_1'(t) &= -(\theta_1 + \theta_2)x_1(t), \\
x_2'(t) &= \theta_1 x_1(t), \\
x_3'(t) &= \theta_2 x_1(t) - (\theta_3 + \theta_4)x_3(t) + \theta_5 x_5(t), \\
x_4'(t) &= \theta_3 x_3(t), \\
x_5'(t) &= \theta_4 x_3(t) - \theta_5 x_5(t).
\end{aligned}
\tag{21}
$$

This system characterizes a reaction that describes the thermal isomerization of α-pinene x_1 to dipentene x_2 and alloocimene x_3, which in turn yields α-pyronene and β-pyronene x_4 and a dimer x_5. The data that we use are taken from Table 2 in Box *et al.* (1973). For each state of the system, the data include only eight observations in time. This is a challenging problem to deal with, a point raised also in Tjoa and Biegler (1991), Rodriguez-Fernandez *et al.* (2006Nov) and Brunel and Clairon (2015). In Table

Table 8.   Point estimates for the parameters of model (Equation 21) based
on the real data from Box *et al.* (1973). We consider the initial values to
be known. The one-step point estimates are given by Equation 6; the con-
fidence intervals were generated according to Equation 15. The left and
right interval points are denoted by CI(L) and CI(R), respectively

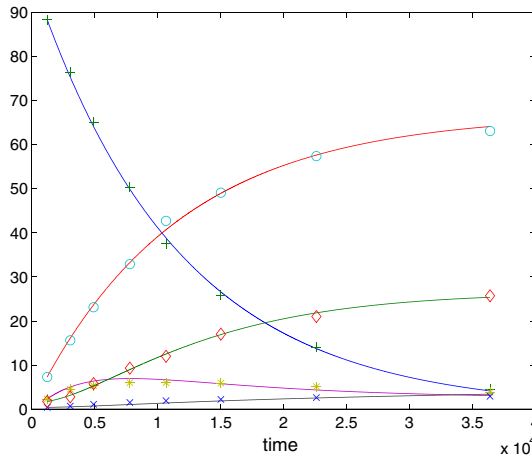|	Point	CI(L)	CI(R)	Tjoa and Biegler (1991)
θ_1	5.869e-05	5.771e-05	5.967e-05	5.926e-05
θ_2	2.830e-05	2.740e-05	2.920e-05	2.963e-05
θ_3	1.745e-05	1.305e-05	2.186e-05	2.047e-05
θ_4	2.132e-04	1.770e-04	2.494e-04	2.744e-04
θ_5	2.137e-05	1.037e-05	3.236e-05	3.997e-05

Fig. 7.   The solution to Equation 21 based on the one-step estimate; the observations are indicated by
different symbols, corresponding to the system state that they represent. The parameters were estimated
using the real data from Box *et al.* (1973).

8, we see the resulting point and interval estimates based on the real data, using the
one-step method. We do not present the results of the Monte Carlo study for the NLS
method, because it could not be completed in a reasonable amount of time using the
Levenberg–Marquardt method (as we did in all examples in our paper). In the last col-
umn of Table 8, we present the estimation result from Tjoa and Biegler (1991). The
solution of the system (Equation 21) corresponding to the one-step estimate is dis-
played in Figure 7. Unlike Tjoa and Biegler (1991), our approach does not require
to provide an initial guess in the parameter space. The parameter estimates that we
obtained are similar to those in Tjoa and Biegler (1991), except for parameters θ_4, θ_5:
The estimates computed in Tjoa and Biegler (1991) are not contained in our confidence
intervals. As explained in detail in Brunel and Clairon (2015), these two parameters
are the most difficult to estimate, and those authors also raise a question whether the
values obtained in Tjoa and Biegler (1991) are reliable and speculate the estimates in
their own work could be in fact more accurate. Without offering a resolution of this
difficult question, here we simply remark that alternative estimates computed in Brunel
and Clairon (2015) are contained in our confidence intervals.

Table 9. Means of point estimates and actual coverage of interval estimates for the parameters of model (Equation 21), where the initial value ξ is considered as known. The results are based on 500 simulation runs; see the experimental setup in the text. The one-step point estimates are given by Equation 6; the interval estimates are defined in Equation 15. Standard errors of the point estimates as calculated based on the 500 simulations (denoted by 'STE'). Square root of the average of the estimates of the asymptotic variance, over the 500 simulations (denoted by 'ASYM')

Setup		True	Mean	Coverage	STE	ASYM
$\sigma = 0.02$	θ_1	5.926e-05	5.920e-05	0.758	6.539e-07	3.913e-07
	θ_2	2.963e-05	2.958e-05	0.806	5.246e-07	3.615e-07
	θ_3	2.047e-05	2.042e-05	1.000	5.789e-07	1.815e-06
	θ_4	2.744e-04	2.709e-04	1.000	7.847e-06	2.099e-05
	θ_5	3.997e-05	3.878e-05	0.998	2.793e-06	6.060e-06
$\sigma = 0.1$	θ_1	5.926e-05	5.910e-05	0.768	3.265e-06	3.026e-02
	θ_2	2.963e-05	2.945e-05	0.820	2.669e-06	2.717e-03
	θ_3	2.047e-05	1.993e-05	0.998	2.755e-06	9.746e-06
	θ_4	2.744e-04	2.452e-04	0.946	8.382e-05	1.406e-04
	θ_5	3.997e-05	3.103e-05	0.940	2.569e-05	9.688e-05

Next, we conducted two simulation studies, corresponding to two different measurement error variances. Specifically, we generated observations according to Equations 2 and 21 under the following experimental setup: The time grid is the same as in the real data, namely,

$$t_j \in \{1230, 3060, 4920, 7800, 10\,680, 15\,030, 22\,620, 36\,420\},$$

resulting in a total of 8 observation points. Initial values are set to the observations at the first time point,

$$\xi = \{88.35, 7.3, 2.3, 0.4, 1.75\}.$$

The errors are normal with expectation zero and standard deviations

$$\sigma = a \times \{44.6833, 36.4111, 4.9570, 1.6339, 12.4147\},$$

corresponding to σ_i, $i = 1, \ldots, 5$. Here, the value a is multiplied by the mean value of each state, as calculated from the solutions based on the real data example. In the first study, we set $a = 0.02$, while in the second, we take $a = 0.1$. We note that the variance σ^2 that corresponds to $a = 0.02$ is the order of the variance that we observed in the real data example. For each scenario, we repeat the experimental setup 500 times and calculate the average of point estimates and actual coverage of the confidence intervals. We also provide the standard error of the one-step estimator as calculated based on 500 simulations ('STE'), as well as the square root of the average of estimates of the asymptotic variance ('ASYM'). The results are presented in Table 9. We see that the actual coverage is not too poor but, nevertheless, deviates noticeably from the nominal level of 95%. Further, we see a considerable difference between estimates of the asymptotic variance and the actual finite sample variance as calculated based on 500 simulations. All these results are not surprising, if we recall that we have at hand only eight observations on each system state, so that asymptotic approximations are not accurate enough yet.

7 Conclusions

Parameter estimation for ODEs is a challenging problem. In this paper, we have explored performance of Le Cam's one-step method in the ODE context both from applied and theoretical sides. Using real and simulated data examples, we have demonstrated that execution of a one-step correction on a preliminary smoothing-based estimator leads to rather satisfactory estimation results, that are comparable with those in the 'gold standard' least squares estimation. In particular, we can argue that already for small and moderate sample sizes, the one-step method yields results comparable with the NLS estimation in terms of the statistical accuracy, as suggested by the asymptotic statistical theory. The empirical coverage of the confidence intervals that we provide is good even for samples as small as $n = 21$ in the examples that we considered. On the other hand, for very small sample sizes, the NLS method appears to perform better than the one-step method, although the latter remains reasonable. Furthermore, we note that the one-step approach discussed in this work was applied for both fully and partially observed ODE systems (Section 5).

The relation between the one-step method and the Levenberg–Marquardt method that we pointed out in Section 5 leads to a very simple practical implementation: When computational time is an issue, our simulations and theory justify the use of the Levenberg–Marquardt algorithm with only one iteration, if its starting point is SME or the integral estimator. In particular, as evidenced by the results presented in Section 5, the performance of the one-step estimator is as good as or even better than that of the NLS starting from a random initial guess and using 100 iterations. This is a useful practical observation: Tuning the number of iterations is possible in software implementations of optimization algorithms, such as the one in Matlab, and hence, the one-step correction on the SME or the integral estimator is straightforward to implement.

Appendix A: Proof of Theorem 1

Note that $\bar{\eta}_n^* = \bar{\eta}_n(\hat{\eta}_{\bar{\rho}_n})$ for some data-dependent (random) smoothing parameter $\bar{\rho}_n$ taking values in the set $\mathcal{R}_n$; more formally,

$$\bar{\rho}_n = \operatorname{argmin}_{\rho_n(k) \in R_n} \sum_{i=1}^{d} \sum_{j=1}^{n} (Y_{ij} - x_i(\bar{\eta}(\hat{\eta}_{\rho_n(k)}), t_j))^2.$$

Observe that the estimator $\hat{\eta}_{\bar{\rho}_n}$ is $\sqrt{n}$-consistent. This claim appears to be self-evident, but nevertheless, we still provide its proof. Thus, for every fixed $\varepsilon > 0$, we have to show existence of a constant K_ε, such that

$$P(\sqrt{n}|\hat{\eta}_{\bar{\rho}_n} - \eta_0| \geq K_\varepsilon) \leq \varepsilon$$

for all $n \geq n_\varepsilon$, where n_ε is some integer, possibly depending on ε and K_ε. We have

$$P(\sqrt{n}|\hat{\eta}_{\bar{\rho}_n} - \eta_0| \geq K_\varepsilon) \leq P\left(\sqrt{n} \sum_{i=1}^{N} |\hat{\eta}_{\rho_n(i)} - \eta_0| \geq K_\varepsilon \right)$$

$$\leq \sum_{i=1}^{N} P\left(\sqrt{n}|\hat{\eta}_{\rho_n(i)} - \eta_0| \geq \frac{K_\varepsilon}{N} \right).$$

$\sqrt{n}$-consistency of $\hat{\eta}_{\bar{\rho}_n}$ now easily follows from the above inequality and $\sqrt{n}$-consistency of each $\hat{\eta}_{\rho_n(k)}$, $k = 1, \ldots, N$.

Now that we know the estimator $\hat{\eta}_{\bar{\rho}_n}$ is $\sqrt{n}$-consistent, the proof of our theorem consists in application of Theorem 5.45 and Addendum 5.46 in van der Vaart (1998), which in turn can be reduced to verification of conditions of Theorem 5.41 there. This amounts to verification of the following conditions:

(i) It must hold that $\sqrt{n}\Psi_n(\eta_0)$ converges in distribution. Here, Ψ_n is as in Formula 4.

(ii) It must hold that for every fixed (t, y), the function $\psi_\eta(t, y)$ is twice continuously differentiable with respect to η. Here, $\psi_\eta(t, y)$ is as in Equation 5.

(iii) It must hold that $\mathbb{E}[\psi_{\eta_0}(T_1, Y_1)] = 0$, $\mathbb{E}[|\psi_{\eta_0}(T_1, Y_1)|^2] < \infty$, and the matrix

$$\mathbb{E}\left[\frac{d}{d\eta} \psi_\eta(T_1, Y_1)|_{\eta = \eta_0} \right]$$

must be non-singular. Here, Y_1 is a shorthand notation for the vector $(Y_{11}, Y_{21}, \ldots, Y_{d1})$.

(iv) It must hold that the second-order partial derivatives of the function ψ_η with respect to η_j, η_k are dominated by an integrable (with respect to its distribution) function of (T_1, Y_1).

Arguments for verification of these conditions are quite standard and follow from the regularity assumptions in the statement of our theorem. The limit covariance matrix in Equation 13 is obtained in the process of verification of (i)–(iv) mentioned earlier.

Appendix B: Integral estimator

Given observations Y_{ij}'s, the one-step method requires first to have at hand a $\sqrt{n}$-consistent estimator of θ_0 and ξ_0. As mentioned in the previous sections, the SME provides us with such an estimator. However, this method is based on estimating the derivative x', which is hard to do accurately in practice for small or moderate sample sizes. In the case where the symbol F of the system of ODEs is linear in functions of the parameter θ, one can avoid estimation of derivatives and use an integral SME. Indeed, in such cases, one can use some version of the so called 'integral approach'

(Himmelblau *et al.*, 1967) as was studied in Dattner and Klaassen (2015). The idea works as follows: Note that for systems whose symbols are linear in parameters, $F(x(t); \theta) = g(x(t))\theta$ holds, where the measurable function $g : \mathbb{R}^d \to \mathbb{R}^{d \times p}$ maps the d-dimensional column vector x into a $d \times p$ matrix. Let $\hat{x}_n(\cdot)$ be an estimator of $x(\eta_0, \cdot)$, and denote $\hat{G}_n(t) = int_0^t g(\hat{x}_n(s), s) \, \mathrm{d}s$, $\hat{A}_n = \int_0^T \hat{G}_n(t) \, \mathrm{d}t$, $\hat{B}_n = \int_0^T \hat{G}_n^T(t) \hat{G}_n(t) \, \mathrm{d}t$, and let I_d be the $d \times d$ identity matrix. Then Dattner and Klaassen (2015) show that the direct estimators,

$$\hat{\xi}_n = \left(I_d - \hat{A}_n \hat{B}_n^{-1} \hat{A}_n^T \right)^{-1} \int_0^T \left(I_d - \hat{A}_n \hat{B}_n^{-1} \hat{G}_n^T(t) \right) \hat{x}_n(t) \, \mathrm{d}t, \qquad (\text{B.1})$$

$$\hat{\theta}_n = \hat{B}_n^{-1} \int_0^T \hat{G}_n^T(t) \left(\hat{x}_n(t) - \hat{\xi}_n \right) \mathrm{d}t, \qquad (\text{B.2})$$

are $\sqrt{n}$-consistent. In case the initial value ξ_0 is known, Appendix B.2 may be used with $\hat{\xi}_n$ replaced by ξ_0. Besides the required statistical properties, the extensive simulation study presented in the aforementioned paper suggests that this approach is much more accurate in finite samples compared to the derivative-based SME. Thus, we use the integral SME (Equations B.1– B.2) whenever applicable and the derivative-based SME otherwise.

We choose to estimate the solution x using local polynomial estimators, which are consistent and 'automatically' correct for the boundaries. Under the assumption that x are C^α-functions for some real $\alpha \geq 1$, we will approximate them by polynomials of degree $\ell = \lfloor \alpha \rfloor$ as follows (Tsybakov, 2009, Section 1.6): let

$$U(u) = \left(1, u, u^2/(2!), \ldots, u^\ell/(\ell!) \right)^T, \quad u \in \mathbb{R},$$
$$v(t) = \left(x(t), x'(t)b, x''(t)b^2, \ldots, x^{(\ell)}(t)b^\ell \right), \quad t \in \mathbb{R},$$

where $b = b_n > 0$ is a bandwidth, the $(\ell + 1)$-vector $U(u)$ is a column vector and $v(t)$ is a $d \times (\ell + 1)$-matrix. Let $K(\cdot)$ be some appropriate kernel function and define

$$\hat{v}_n(t) = \arg\min_{v \in \mathbb{R}^{d \times (\ell+1)}} \sum_{i=1}^n \left\{ Y(t_i) - v U\left(\frac{t_i - t}{b} \right) \right\}^T$$
$$\times \left\{ Y(t_i) - v U\left(\frac{t_i - t}{b} \right) \right\} K\left(\frac{t_i - t}{b} \right).$$

The local polynomial estimator of order ℓ of $x(t)$ is the first column of the $d \times (\ell + 1)$-matrix $\hat{v}_n(t)$, i.e. $\hat{x}_n(t) = \hat{v}_n(t) U(0)$.

We applied the estimation procedure described earlier to a set of bandwidths $B := \{b_{\min}, \ldots, b_{\max}\}$, and for a given $b \in B$, we denote the resulting one-step parameter estimator by $\bar{\eta}_{n,b}$. We then select $\bar{\eta}_n = \bar{\eta}_{n,\bar{b}}$ for some $\bar{b} \in B$, the choice of which is discussed in Remark 3 of the main text. Last, we use local estimators polynomials of order 1, with $K(t) = 3/4(1 - t^2)\mathbf{1}\{|t| \leq 1\}$ (cf. Dattner and Klaassen (2015), where $\mathbf{1}\{\cdot\}$ stands for the indicator function. Other kernels are also possible.

Appendix C: Goodwin's oscillator

In Section 5, we applied the direct integral method on a partially observed Goodwin's oscillator,

$$\begin{cases} x_1'(t) & = \dfrac{\theta_1}{1 + 2x_3(t)^{10}} - \theta_5 x_1(t), \\ x_2'(t) & = 2x_1(t) - \theta_5 x_2(t), \\ x_3'(t) & = x_2(t) - \theta_5 x_3(t). \end{cases} \tag{C.1}$$

The integral estimation approach works as follows in this case: First, apply the integral estimation method from Appendix B on the second equation of Equation C.1 and obtain a $\sqrt{n}$-consistent estimator of θ_5 (this is possible, because the state variables x_1, x_2 are observed in the setting of Section 5). Next, integrate the equation

$$\begin{cases} x_3'(t) & = \hat{x}_2(t) - \hat{\theta}_5 x_3(t), \\ x_3(0) & = 0, \end{cases}$$

to get an estimator $\hat{x}_3$ of the component x_3 of the solution to Equation C.1. Finally, apply the integral estimation method on the first equation of Equations C.1 to get a $\sqrt{n}$-consistent estimator of θ_1 (this is possible, because estimators $\hat{x}_1$ and $\hat{x}_3$ of x_1 and x_3 are available).

Acknowledgements

References

BELLMAN, R., J. JACQUEZ, R. KALABA and S. SCHWIMMER (1967), Quasilinearization and the estimation of chemical rate constants from raw kinetic data, *Mathematical Biosciences* **1**, 71–76.

BELLMAN, R. and R. S. ROTH (1971), The use of splines with unknown end points in the identification of systems, *Journal of Mathematical Analysis and Applications* **34**, 26–33.

BICKEL, P. J. (1975), One-step Huber estimates in the linear model, *Journal of the American Statistical Association* **70**, 428–434.

BODENSTEIN, M. (1922), Bildung und Zersetzung der Höheren Stickoxyde, *Zeitschrift fur Physikalische Chemie* **100**, 68–123.

BOX, G. E. P., W. G. HUNTER, J. F. MACGREGOR and J. ERJAVEC (1973), Some problems associated with the analysis of multiresponse data, *Technometrics* **15**, 33–51.

BRUNEL, N. J. B. (2008), Parameter estimation of ODE's via nonparametric estimators, *Electronic Journal of Statistics* **2**, 1242–1267.

BRUNEL, N. J. B. and Q. CLAIRON (2015), A tracking approach to parameter estimation in linear ordinary differential equations, *Electronic Journal Statistics* **9**, 2903–2949.

BUCCI, V., B. TZEN, N. LI, M. SIMMONS, T. TANOUE, E. BOGART, L. DENG, V. YELISEYEV, M. L. DELANEY, Q. LIU, B. OLLE, R. R. STEIN, K. HONDA, L. BRY and G. K. GERBER (2016), Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses, *Genome Biology* **17**, 121.

CAI, Z., J. FAN and R. LI (2000), Efficient estimation and inferences for varying-coefficient models, *Journal of the American Statistical Association* **95**, 888–902.

CALDERHEAD, B. and M. GIROLAMI (2009), Estimating Bayes factors via thermodynamic integration and population MCMC, *Computational Statistics & Data Analysis* **53**, 4028–4045.

CAMPBELL, D. and S. LELE (2014), An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems, *Computational Statistics & Data Analysis* **70**, 257–267.

CAO, Y. (2008), Local linear kernel regression. MATLAB Central File Exchange (http://www.mathworks.com/matlabcentral/fileexchange/19564-local-linear-kernel-regression), Retrieved March 18, 2015.

CHOU, I. C. and E. O. VOIT (2009), Recent developments in parameter estimation and structure identification of biochemical and genomic systems, *Mathematical biosciences* **219**, 57.

DATTNER, I. (2015), A model-based initial guess for estimating parameters in systems of ordinary differential equations, *Biometrics* **71**, 1176–1184.

DATTNER, I. and C. A. J. KLAASSEN (2015), Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters, *Electronic Journal Statistics* **9**, 1939–1973.

DELECROIX, M., W. HÄRDLE and M. HRISTACHE (2003), Efficient estimation in conditional single-index regression, *Journal of Multivariate Analysis* **86**, 213–226.

EDELSTEIN-KESHET, L. (2005), Mathematical models in biology. Classics in applied mathematics, Society for Industrial and Applied Mathematics, vol. 46, Philadelphia.

ESPOSITO, W. R. and C. A. FLOUDAS (2000), Global optimization for the parameter estimation of differential-algebraic systems, *Industrial & Engineering Chemistry Research* **39**, 1291–1310.

FEINBERG, M. (1979), Lectures on chemical reaction networks, Notes of lectures given at the Mathematics Research Center, University of Wisconsin. https://crnt.osu.edu/LecturesOnReactionNetworks.

FIELD, C. A. and D. P. WIENS (1994), One-step M-estimators in the linear model, with dependent errors, *Canadian Journal of Statistics* **22**, 219–231.

FLOUDAS, C. A., P. M. PARDALOS, C. S. ADJIMAN, W. R. ESPOSITO, Z. H. GÜMÜŞ, S. T. HARDING, J. L. KLEPEIS, C. A. MEYER and C. A. SCHWEIGER (1999), *Handbook of test problems in local and global optimization*, Nonconvex optimization and its applications, vol. 33, Kluwer Academic Publishers, Dordrecht.

GIROLAMI, M. (2008), Bayesian inference for differential equations, *Theoretical Computer Science* **408**, 4–16.

GOEL, G., I. C. CHOU and E. O. VOIT (2008), System estimation from metabolic time-series data, *Bioinformatics* **24**, 2505–2511.

GOODWIN, B. C. (1963), *Temporal organization in cells. a dynamic theory of cellular control processes*, Academic Press, London.

GOODWIN, B. C. (1965), Oscillatory behavior in enzymatic control processes, *Advances in Enzyme Regulation* **3**, 425–437.

GRIFFITH, J. S. (1968), Mathematics of cellular control processes I. Negative feedback to one gene, *Journal of Theoretical Biology* **20**, 202–208.

GUGUSHVILI, S. and C. A. J. KLAASSEN (2012), $\sqrt{n}$-Consistent parameter estimation for systems of ordinary differential equations: bypassing numerical integration via smoothing, *Bernoulli* **18**, 1061–1098.

HALL, P. and Y. MA (2014), Quick and easy one-step parameter estimation in differential equations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 735–748.

HIMMELBLAU, D. M., C. R. JONES and K. B. BISCHOFF (1967), Determination of rate constants for complex kinetics models, *Industrial & Engineering Chemistry Fundamentals* **6**, 539–543.

HOOKER, G. (2009), Forcing function diagnostics for nonlinear dynamics, *Biometrics* **65**, 928–936.

HOOKER, G., S. P. ELLNER, L. D. V. RODITI and J. D. EARN (2011), Parameterizing state–space models for infectious disease dynamics by generalized profiling: measles in Ontario, *Journal of The Royal Society Interface* **8**, 961–974.

JENNRICH, R. I. (1969), Asymptotic properties of non-linear least squares estimators, *Annals of Mathematical Statistics* **40**, 633–643.

KIM, T. and Y. P. SHENG (2010), Estimation of water quality model parameters, *KSCE Journal of Civil Engineering* **14**, 421–437.

MARQUARDT, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial & Applied Mathematics* **11**, 431–441.

THE MATHWORKS, INC. (2017), *Matlab version 9.3 (r2017b)*, Mathworks, Inc., Natick, Massachusetts.

MURRAY, J. D. (2002), *Mathematical biology. I* Third ed., Interdisciplinary Applied Mathematics, vol. 17, Springer-Verlag, New York, NY.

OATES, C., S. NIEDERER, A. LEE, F. X. BRIOL and M. GIROLAMI (2017), Probabilistic Models for Integration Error in the Assessment of Functional Cardiac Models, in: GUYON, I., U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN and GARNETT R. (eds), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 109–117. http://papers.nips.cc/paper/6616-probabilistic-models-for-integration-error-in-the-assessment-of-functional-cardiac-models.pdf.

OATES, C. J., T. PAPAMARKOU and M. GIROLAMI (2016), The controlled thermodynamic integral for Bayesian model evidence evaluation, *Journal of the American Statistical Association* **111**, 634–645.

R CORE TEAM (2017), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

RAMSAY, J. and G. HOOKER (2017), *Dynamic data analysis*, Springer Series in Statistics, Springer, New York.

RAMSAY, J. O, G. HOOKER, D. CAMPBELL and J. CAO (2007), Parameter estimation for differential equations: a generalized smoothing approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 741–796.

RIEDER, S. (2012), Robust parameter estimation for the Ornstein–Uhlenbeck process, *Statistical Methods & Applications* **21**, 411–436.

ROBERTSON, H. H. (1966), The solution of a set of reaction rate equations, in: WALSH, J. (ed.), Numerical analysis: an introduction, London, New York, Academic Press, 178–182.

RODRIGUEZ-FERNANDEZ, M., J. A. EGEA and J. R. BANGA (2006Nov), Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems, *BMC Bioinformatics* **7**, 483.

SCHITTKOWSKI, K. (2002), *Numerical data fitting in dynamical systems*, Applied Optimization, vol. 77, Kluwer Academic Publishers, Dordrecht.

SIMPSON, D. G., D. RUPPERT and R. J. CARROLL (1992), On one-step GM estimates and stability of inferences in linear regression, *Journal of the American Statistical Association* **87**, 439–450.

SONTAG, E. D. (2001), Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction, *IEEE Transactions on Automatic Control* **46**, 1028–1047.

STEIN, R. R., V. BUCCI, N. C. TOUSSAINT, C. G. BUFFIE, G. RÄTSCH, E. G. PAMER, C. SANDER and J. B. XAVIER (2013), Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota, *PLoS Computational Biology* **9**, e1003388.

SWARTZ, J. and H. BREMERMANN (1975), Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of the estimates, *Journal of Mathematical Biology* **1**, 241–257.

TJOA, I. B. and L. T. BIEGLER (1991), Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems, *Industrial & Engineering Chemistry Research* **30**, 376–385.

VAN DER VAART, A. W. (1998), *Asymptotic statistics*, Cambridge University Press, Cambridge.

Van Domselaar, B. and P. W. Hemker, (1975), Nonlinear parameter estimation in initial value problems, *Technical report, SIS-76-1121*.

Varah, J. M. (1982), A spline least squares method for numerical parameter estimation in differential equations, *SIAM Journal on Scientific and Statistical Computing* **3**, 28–46.

Voit, E. O. (2000), *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*, Cambridge University Press.

Voit, E. O. and J. Almeida (2004), Decoupling dynamical systems for pathway identification from metabolic profiles, *Bioinformatics* **20**, 1670–1681.

Voit, E. O. and M. A. Savageau (1982), Power-law approach to modeling biological systems, III. methods of analysis, *Journal of Fermentation Technology* **60**, 233–241.

Vujačić, I., I. Dattner, J. González and E. Wit (2015), Time-course window estimator for ordinary differential equations linear in the parameters, *Statistics and Computing* **25**, 1057–1070.

Wu, C. F. (1981), Asymptotic theory of nonlinear least squares estimation, *Annals of Statistics* **9**, 501–513.