On estimation algorithms for ordinary differential equations

M. R. Osborne[1]

(Received 22 July 2008; revised 10 October 2008)

Abstract

This article addresses the problem of estimating the parameters of a system of ordinary differential equations given data derived from noisy observations on the state variables. This problem is important in a range of applications in areas such as adaptive, real time control. There are two main classes of method for attacking this problem, and their equivalence and effectiveness (consistency) are discussed. Recent rate of convergence results for the major implementation techniques are summarized, and some matters requiring further consideration indicated.

Contents

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1363 gives this article, © Austral. Mathematical Soc. 2008. Published October 27, 2008. ISSN 1446-8735. (Print two pages per sheet of paper.)

1 Introduction

The problem under consideration is that of estimating the vector of parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ in the system of ordinary differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}, \boldsymbol{\beta}) , \tag{1}$$

where $\mathbf{x}, \mathbf{f} \in \mathbb{R}^m$, given observed data:

$$\mathbf{y}_i = \mathcal{O}\mathbf{x}^*(t_i, \boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}_i , \quad i = 1, 2, \ldots, n , \tag{2}$$

$$\mathcal{O} \in \mathbb{R}^m \to \mathbb{R}^k, \quad \mathbf{y}_i \in \mathbb{R}^k, \quad k \leq m , \quad t_i \in [0, 1] ,$$

$$\boldsymbol{\varepsilon}_i \in \mathbb{R}^m, \quad \boldsymbol{\varepsilon}_i \sim N\left(0, \sigma^2 I_k\right) \text{ and independent.} \tag{3}$$

The assumption of normal errors is standard enough, but it does have some implications (Section 5). It means that the maximum likelihood estimators of the 'true' parameter vector $\boldsymbol{\beta}^*$ and corresponding state variable values $\mathbf{x}^*(t, \boldsymbol{\beta}^*)$ are found by solving the nonlinear least squares problem

$$\hat{\boldsymbol{\beta}}_n = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathcal{O}\mathbf{x}_i\|^2 \tag{4}$$

where the allowed values of $\mathbf{x}_i = \mathbf{x}(t_i, \boldsymbol{\beta})$ are constrained to satisfy the differential equation. Typically the approximation is made that a discretized version of the differential equation is used to constrain trial values in (4). Here this is the trapezoidal rule

$$\mathbf{c}(\mathbf{x}_c)_i = \mathbf{x}_{i+1} - \mathbf{x}_i - \frac{\Delta t}{2}\left[\mathbf{f}(t_{i+1}, \mathbf{x}_{i+1}, \boldsymbol{\beta}) + \mathbf{f}(t_i, \mathbf{x}_i, \boldsymbol{\beta})\right] = 0\,, \tag{5}$$

where $(\mathbf{x}_c)_i = \mathbf{x}_i$, $i = 1, 2, \ldots, n$, $\mathbf{x}_c \in \mathbb{R}^{n\times m}$. It produces solution values that are usually sufficiently accurate when working with noise contaminated signals (see Section 4). It has the important sparsity structure

$$\mathbf{c}(\mathbf{x}_c)_i = \mathbf{c}_{ii}(\mathbf{x}_i) + \mathbf{c}_{i(i+1)}(\mathbf{x}_{i+1})\,, \quad i = 1, 2, \ldots, n-1\,. \tag{6}$$

This structure is not possessed by higher order discretizations.

Methods for solving the optimization problem (4) fall into two general classes called here *embedding* [3], and *simultaneous* [1, 6]. The embedding method provides a formal link between the problem and the closely related regression problem, but this connection involves some arbitrary choices that affects performance. However, it does allow the method to connect with an important body of theory that enables validation of the procedure. The simultaneous class avoids this specification uncertainty and should be capable of being implemented more efficiently in many cases. However, justification is technically a more difficult problem. These points are illustrated by considering in Section 3 the problem of the equivalence of the methods and in Section 4 the question of asymptotic consistency—do the methods produce the true parameter vector in the limit as the number of observations grows without bound? Completely satisfactory answers to all questions of interest are not yet available .

2 Methods

2.1 Estimation via embedding

The embedding approach leads to an unconstrained optimization problem that typically is solved by such standard methods as the Gauss–Newton algorithm. It removes the differential equation constraint on the state variable $\mathbf{x}(t, \boldsymbol{\beta})$ by embedding the differential equation into a parametrised family of boundary value problems that is solved explicitly at each step in order to generate trial solution values. This requires boundary conditions

$$B_1\mathbf{x}(0) + B_2\mathbf{x}(1) = \mathbf{b}\,,$$

where $B_1, B_2 \in \mathbb{R}^m \to \mathbb{R}^m$ are assumed known while $\mathbf{b}$ is a vector of additional parameters that must be determined as part of the estimation process. The key requirement is that the resulting system has a numerically well determined solution $\mathbf{x}(t, \boldsymbol{\beta}, \mathbf{b})$ for all $(\boldsymbol{\beta}, \mathbf{b})$ in a large enough neighborhood of $(\boldsymbol{\beta}^*, \mathbf{b}^*)$ where $\mathbf{b}^*$ is determined by the true state variable values

$$\mathbf{b}^* = B_1\mathbf{x}^*(0, \boldsymbol{\beta}^*) + B_2\mathbf{x}^*(1, \boldsymbol{\beta}^*)\,.$$

So far this leaves open the selection of appropriate B_1 and B_2. A hint is provided by the need to calculate $\mathbf{x}(t, \boldsymbol{\beta}, \mathbf{b})$ and $\nabla_\beta\mathbf{x}$, $\nabla_b\mathbf{x}$ at each Gauss–Newton step. Common to these computations is the solution of a sequence of linear problems obtained by linearising the differential equation about the current solution estimates. For example, $\nabla_b\mathbf{x}$ satisfies the linear system

$$\begin{aligned}\frac{d}{dt}\nabla_b\mathbf{x} - \nabla_x\mathbf{f}\nabla_b\mathbf{x} &= 0\,,\\ B_1\nabla_b\mathbf{x}(0) + B_2\nabla_b\mathbf{x}(1) &= I\,.\end{aligned}$$

Computation of the trapezoidal rule approximations (5) to these quantities requires the inversion of the matrix

$$F = \begin{bmatrix} C_{11} & C_{12} & & & & \\ & C_{22} & C_{23} & & & \\ & & & \ddots & & \\ & & & & C_{(n-1)(n-1)} & C_{(n-1)n} \\ B_1 & & & & & B_2 \end{bmatrix}$$

where $C_{ij} = \nabla_x \mathbf{c}_{ij}$.

Idea *Choose* B_1, B_2 *so this matrix is well conditioned at* $\mathbf{x}^*(t, \boldsymbol{\beta}^*)$.

Computation Begin by permuting the first block column of F to the last position. A transformation of the first $n-1$ block rows of the permuted matrix to block upper triangular form by orthogonal S using Householder transformations yields

$$R = S^T F P = \begin{bmatrix} R_{11} & R_{12} & 0 & \cdots & 0 & R_{1n} \\ & R_{22} & R_{23} & \cdots & 0 & R_{2n} \\ & & & \ddots & \vdots & \vdots \\ & & & & R_{(n-1)(n-1)} & R_{(n-1)n} \\ & & & & B_2 & B_1 \end{bmatrix} . \tag{7}$$

This orthogonal factorization affects quantities that depend on the differential equation only. The first and last block components of the solutions to the linearised equations with matrix F are determined by the last two block rows of R and so depend directly on B_1 and B_2. To compute suitable values make a second orthogonal factorization

$$\begin{bmatrix} R_{(n-1)(n-1)} & R_{(n-1)n} \end{bmatrix} = \begin{bmatrix} U^T & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} .$$

It follows that $\begin{bmatrix} B_2 & B_1 \end{bmatrix} = Q_2^T$ provides an appropriate choice.

The embedding method has advantages.

- An estimate of the boundary conditions can be computed by the above procedure given suitable initial $\mathbf{x}_c^0$. Previous work illustrated the stability advantages [3].
- It is readily adapted to make use of standard Gauss–Newton nonlinear least squares solvers and differential equation boundary value software.
- The availability of good boundary value software is important if the differential equation is difficult.

It has disadvantages.

- What happens if $1 - \left\| Q_2^T \left(\mathbf{x}_c^0\right)^T Q_2 \left(\mathbf{x}_c^*\right) \right\|$ is close to 1? Good initial conditions are important!
- The economics of solving a nonlinear boundary value problem for every function evaluation needs attention.
- The extra parameters $\mathbf{b}$ are not directly relevant to the problem formulation.

2.2 Simultaneous estimation

Let

$$\mathbf{r}_i = \mathbf{y}_i - \mathcal{O}\mathbf{x}_i \quad \text{and} \quad \Phi\left(\mathbf{x}_c\right) = \frac{1}{2n} \sum_{i=1}^{n} \left\|\mathbf{r}_i\right\|^2 .$$

Then the simultaneous method formulates the estimation problem as a constrained nonlinear least squares problem for the $nm + p$ unknowns $(\mathbf{x}_c, \boldsymbol{\beta})$:

$$\begin{bmatrix} \hat{\mathbf{x}}_c \\ \hat{\boldsymbol{\beta}}_n \end{bmatrix} = \arg\min_{\mathbf{x}_c, \boldsymbol{\beta}} \Phi(\mathbf{x}_c); \quad \mathbf{c}(\mathbf{x}_c, \boldsymbol{\beta}) = 0. \tag{8}$$

Solution of (8) falls within the scope of standard methods of sequential quadratic programming [2]. However, note that the number of constraints increases as the discretization of the differential equation is refined. This provides a context in which it is necessary to exploit the sparsity structure of the problem formulation.

To summarise the solution process, introduce the problem Lagrangian

$$\mathcal{L}(\mathbf{x}_c, \boldsymbol{\beta}, \boldsymbol{\lambda}_c) = \Phi(\mathbf{x}_c) + \sum_{i=1}^{n-1} \boldsymbol{\lambda}_i^T \mathbf{c}_i(\mathbf{x}_c, \boldsymbol{\beta}) . \tag{9}$$

The necessary conditions for a stationary point give

$$\nabla_{((x,\beta),\lambda)} \mathcal{L}(\mathbf{x}_c, \boldsymbol{\beta}, \boldsymbol{\lambda}_c) = 0. \tag{10}$$

The corresponding Newton iteration is

$$\begin{bmatrix} \nabla^2_{(x,\beta)} \mathcal{L} & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \triangle \mathbf{x} \\ \triangle \boldsymbol{\beta} \\ \triangle \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla_{(x,\beta)} \mathcal{L}^T \\ \mathbf{c}_c \end{bmatrix}, \tag{11}$$

where $C = \nabla_{(x,\beta)} \mathbf{c}_c \in \mathbb{R}^{nm+p} \to \mathbb{R}^{(n-1)m+p}$. The simultaneous method has advantages.

- It is completely specified given initial estimates of $\mathbf{x}_c$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}_c$. I have shown [5] that the choice $\boldsymbol{\lambda}_c = 0$ is typically suitable in large samples.
- Economy—the simultaneous method avoids the work required to solve nonlinear boundary value problems at each step of the embedding method.

It has disadvantages.

- The number of constraints grows without bound as the discretization is refined.
- This means that so does the number of constraint second derivatives that must be computed.
- Solution strategies such as mesh refinement are more difficult to formulate as exact state variable values are known only at the solution.

3 Equivalence

Superficially the embedding and simultaneous methods look rather different. This is not misleading. The relatively arbitrary component in the embedding method has been noted, while the simultaneous method has a surprising depth of structure. Perhaps the most obvious feature in common is that they address the same problem! However, some progress is possible on the question of equivalence. Specifically, an isolated local minimum of the sums of squares of residuals for one method is also an isolated local minimum of the sum of squares of residuals of the other.

Let $S_S(\mathbf{x})$ be the sum of squares of residuals in the simultaneous method corresponding to feasible $\mathbf{x}$, and let $S_E(\mathbf{x}, \mathbf{b})$ be the sum of squares of residuals in the embedding method corresponding to given boundary vector $\mathbf{b}$. Let $\mathbf{x}_S$ be an isolated local minimum of the simultaneous method in a ball $R(\mathbf{x}_S, \rho)$ of radius ρ for some $\rho > 0$. Then direct substitution gives

$$B_1\mathbf{x}_S(0) + B_2\mathbf{x}_S(1) = \mathbf{b}_S\,.$$

Because $\mathbf{x}_S$ satisfies (5) the corresponding sum of squares is defined for the embedding method and $S_E(\mathbf{x}_S, \mathbf{b}_S) = S_S(\mathbf{x}_S)$. Assume $(\mathbf{x}_S, \mathbf{b}_S)$ is not a

corresponding local minimum of $S_E(\mathbf{x}, \mathbf{b})$. Then there exists $\mathbf{x} = \mathbf{x}_P \in R(\mathbf{x}_S, \rho)$, and $\mathbf{b} = \mathbf{b}_P$ such that

$$S_E(\mathbf{x}_P, \mathbf{b}_P) < S_E(\mathbf{x}_S, \mathbf{b}_S) \ .$$

However, $\mathbf{x}_P$ is feasible for the simultaneous method. Thus

$$S_S(\mathbf{x}_P) = S_E(\mathbf{x}_P, \mathbf{b}_P) < S_S(\mathbf{x}_S) \ .$$

This is a contradiction. It follows that $(\mathbf{x}_S, \mathbf{b}_S)$ provides a local minimum for both methods. The argument can be reversed to show that if $(\mathbf{x}_E, \mathbf{b}_E)$ is a local minimum of the embedding method then it is a local minimum of the simultaneous method also.

This is a non–constructive argument. A more interesting result would be one that addressed more of the structure of the methods. For example, it would be interesting to show that satisfaction of necessary conditions for either the embedding or simultaneous methods could be deduced from satisfaction of the other. This would be particularly interesting for the discussion of consistency in the next section as a direct proof of consistency for the simultaneous method appears to be lacking.

4 Consistency

Differential equation estimation by the embedding method becomes a conventional maximum likelihood estimation problem if the boundary value problems are solved exactly. Thus methods for showing consistency of maximum likelihood can be applied to the embedding method also. The following argument [4] has the advantage that it avoids the usual assumption of knowledge of the global maximum of the likelihood function. It has the further advantage that it extends without difficulty when the likelihood is only evaluated

approximately. The maximum likelihood problem has the form

$$\hat{\boldsymbol{\beta}}_n = \arg\max_{\beta} \mathbf{L}_n(\mathbf{y}, \boldsymbol{\beta}) = \arg\max_{\beta} \sum_{i=1}^{n} L(\mathbf{y}_i, t_i, \boldsymbol{\beta}) \,,$$

where L corresponds to the log of the relevant probability density. Assume the t_i equispaced, then

$$\frac{1}{n}\nabla_\beta \mathbf{L}(\mathbf{y}, \boldsymbol{\beta}) \underset{n\to\infty}{\overset{a.s.}{\to}} \int_0^1 \mathcal{E}^*\{\nabla_\beta L(\mathbf{y}, t, \boldsymbol{\beta})\}\, dt \,, \tag{12}$$

where the expectation is evaluated using the true density. This gives a limiting form of the necessary conditions

$$\nabla_\beta \mathbf{L}_n(\mathbf{y}, \boldsymbol{\beta}) = 0 \,. \tag{13}$$

It follows from a standard identity that $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ is the limiting solution. The Kantorovich form of Newton's method is used to show that $\hat{\boldsymbol{\beta}}_n \underset{n\to\infty}{\overset{a.s.}{\to}} \boldsymbol{\beta}^*$. The idea is to apply this to solve the necessary conditions starting from $\boldsymbol{\beta}^*$. The result (12) can be used to show that the exact limiting solution leads to small residuals in (13). Then the Kantorovitch result is used to show that $\hat{\boldsymbol{\beta}}_n$ is close to $\boldsymbol{\beta}^*$ almost surely.

The Kantorovich Theorem required has the following statement. Let $\mathcal{J}_n = \frac{1}{n}\nabla^2_{\beta\beta}\mathbf{L}$, and $S_\rho = \{\boldsymbol{\beta} \mid \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \rho\}$. If the following four conditions are satisfied

1. $\|\mathcal{J}_n(\mathbf{u}) - \mathcal{J}_n(\mathbf{v})\| \le K_1 \|\mathbf{u} - \mathbf{v}\|$, for all $\mathbf{u}, \mathbf{v} \in S_\rho$,
2. $\left\|\mathcal{J}_n(\boldsymbol{\beta}_0)^{-1}\right\| = K_2$,
3. $\left\|\mathcal{J}_n(\boldsymbol{\beta}_0)^{-1}\, \frac{1}{n}\nabla_x \mathbf{L}_n(\mathbf{y}; \boldsymbol{\beta}_0)^T\right\| = K_3$, and
4. $\xi = K_1 K_2 K_3 < \frac{1}{2}$,

then the Newton iteration started at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ converges to a point $\widehat{\boldsymbol{\beta}} \in S_\rho$ satisfying the estimating equation (13), and $\widehat{\boldsymbol{\beta}}$ is the only root in S_ρ. The step to the solution $\widehat{\boldsymbol{\beta}}$ is bounded by

$$\left\| \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right\|_2 < 2K_3 < \rho \, .$$

The consistency result for the embedding method that assumes exact integration requires modification to take account of discretization error that causes the objective function to differ from the true likelihood for all finite n. The embedding consistency result extends to two important cases:

1. when each differential equation discretization grid $\mathbf{K}_n$ corresponds to the observation grid $\mathbf{T}_n$; and
2. when the discretization is made on a fixed grid $t_j \in \mathbf{K}$ independent of $\mathbf{T}_n$, $n \to \infty$.

The maximum mesh spacing $\Delta t \to 0$, $n \to \infty$ in the first case so the solution of the discretized problem tends to that of the differential equation at a satisfactory rate. For the trapezoidal rule this is $O\left(\Delta t^2\right)$. It is significantly faster than any relevant stochastic rate (typically $O\left(\Delta t^{1/2}\right)$) . In the second case Δt is fixed and finite. This means that truncation error effects persist in the solution of the discretized problem as the size of the observation set $|T_n| \to \infty$.

Typical results are the following.

1. If $\Delta t \to 0$ then consistency follows using an argument similar to that in the exact integration case. The idea is to start the iteration for each n at the exact integration solution $(\widehat{\boldsymbol{\beta}}_n, \widehat{\mathbf{b}}_n)$ and use knowledge of the discretization error to show that $K_3 = O\left(\Delta t\right)^2$ so this start is close to the finite grid solution $(\boldsymbol{\beta}_\Delta^n, \mathbf{b}_\Delta^n)$. Consistency of the finite grid solution now follows from the consistency for exact integration

2. If Δt fixed, small enough, then the best result possible is

$$\begin{bmatrix} \beta_{\Delta}^{n} \\ b_{\Delta}^{n} \end{bmatrix} \subset S\left(\begin{bmatrix} \beta^{*} \\ b^{*} \end{bmatrix}, O\left(\Delta t^{2}\right) \right), \; n \to \infty .$$

It uses $K_3 = O\left(\Delta t\right)^2$ for all $n = |T_n|$ large enough.

5 Convergence rate results

The Gauss–Newton method for nonlinear least squares minimization is typically the method of choice in the embedding method [3]. It has the key feature that the evaluation of second derivatives is avoided in the approximate Hessian. This has the consequential advantages of strong positive definiteness properties and excellent scale invariance. A key result in this case is that the convergence rate approaches second order asymptotically if the discretization error tends to zero as $|T_n| \to \infty$ [4]. However, if Δt fixed, small enough, so that discretization error effects persist, then the convergence rate is reduced to a fast first order rate if needed function values are found by linear interpolation.

The Bock iteration [1] is the method of choice in the simultaneous method. Here the Newton iteration is modified by setting the constraint second derivatives to 0 in (11). The condition needed for ignoring these curvature terms is that the associated Lagrange multipliers be small as $\Delta t \to 0$. This requirement is satisfied if the error terms are normally distributed. In this case the multipliers are $O\left(\Delta t^{1/2}\right)$ so that the Bock iteration has a similar convergence rate to Gauss–Newton [5]. This is a stronger condition on the error structure than that required for the convergence rate estimates for the embedding methods. The assumptions needed for the Gauss-Newton results require only the weaker conditions that the errors are independent and have bounded variance [4].

6 In conclusion

The two main approaches to the differential equation estimation problem have been considered from the point of view of their equivalence and consistency. Consistency of the embedding method follows relatively easily from standard results in regression analysis. Thus consistency of the simultaneous method follows from results establishing the equivalence of the two approaches. While a relatively simple argument serves to establish a form of equivalence, deeper results would have significance. One important step would be an independent proof of consistency for the simultaneous method. Both the Gauss–Newton and Bock algorithms make use of a strategy of ignoring certain second order partial derivatives. However, rather different assumptions on measurement error distributions contrast the convergence rate results obtained for the two methods.

References

[1] H. G. Bock. Recent advances in parameter identification techniques in ODE. In P. Deuflhard and E. Hairer, editors, *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, pages 95–121. Birkhäuser, 1983. C109, C118

[2] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer–Verlag, 1999. C113

[3] M. R. Osborne. Numerical questions in ODE boundary value problems. In Wayne Read, Jay W. Larson, and A. J. Roberts, editors, *Proceedings of the 13th Biennial Computational Techniques and Applications Conference, CTAC-2006*, volume 48 of *ANZIAM J.*, pages C899–C926, February 2008. http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/79 [February 11, 2008]. C109, C112, C118

[4] M. R. Osborne. Fisher's method of scoring. *Int. Stat. Rev.*, 86:271–286, 1992. C115, C118

[5] M. R. Osborne. The Bock iteration for the ODE estimation problem. 2008. in preparation. C113, C118

[6] I. Tjoa and L. T. Biegler. Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic systems. *Ind. Eng. Chem. Res.*, 30:376–385, 1991. C109

Author address

1. **M. R. Osborne**, Mathematical Sciences Institute, Australian National University, ACT 0200, Australia.
mailto:Mike.Osborne@maths.anu.edu.au