1

Stability Problems in ODE Estimation

Michael R.Osborne

Mathematical Sciences Institute, Australian National University,
Canberra 0200, Australia
Mike.Osborne@maths.anu.edu.au

Summary. The main question addressed is how does the stability of the underlying differential equation system impact on the computational performance of the two major estimation methods, the embedding and simultaneous algorithms. It is shown there is a natural choice of boundary conditions in the embedding method, but the applicability of the method is still restricted by the requirement that this optimal formulation as a boundary value problem be stable. The most attractive implementation of the simultaneous method would appear to be the null space method. Numerical evidence is presented that this is at least as stable as methods that depend on stability of the boundary value formulation.

1.1 Introduction

The description of the estimation problem begins with a system of differential equations depending explicitly on a fixed vector of parameters together with data obtained by sampling solution trajectories typically in the presence of noise. The system of differential equations is written:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(t, \mathbf{x}, \boldsymbol{\beta}\right), \tag{1.1}$$

where the state vector $\mathbf{x} \in R^m$, the parameter vector $\boldsymbol{\beta} \in R^p$, and it is assumed that $\mathbf{f} \in R \times R^m \times R^p \to R^m$ is smooth enough. The data is assumed to have the form:

$$\mathbf{y}_i = H\mathbf{x}(t_i, \boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, n, \tag{1.2}$$

where $H : R^m \to R^k$, and the observational error $\boldsymbol{\varepsilon}_i \sim N\left(0, \sigma^2 I\right)$. The problem is to estimate $\boldsymbol{\beta}$ by making use of the given data and the structural information contained in the differential equation statement. An alternative formulation of the estimation problem as a smoothing problem by incorporating the parameter vector into the state vector is also useful in certain circumstances. This approach expands the system of differential equations by making the substitutions:

$$\mathbf{x} \leftarrow \begin{bmatrix} \mathbf{x}(t) \\ \boldsymbol{\beta} \end{bmatrix}, \mathbf{f} \leftarrow \begin{bmatrix} \mathbf{f}(t, \mathbf{x}) \\ 0 \end{bmatrix} \tag{1.3}$$

The standard estimation methods of least squares and maximum likelihood are equivalent in this problem context. The basic idea is that $\boldsymbol{\beta}$ is to be estimated by minimizing the objective:

$$F\left(\mathbf{x}_c, \boldsymbol{\beta}\right) = \sum_{i=1}^{n} \left\|\mathbf{y}_i - H\mathbf{x}\left(t_i, \boldsymbol{\beta}\right)\right\|^2 \tag{1.4}$$

over all allowable values of the state variables $\mathbf{x}(t_i, \boldsymbol{\beta}),\ i = 1, 2, \cdots, n$. Methods differ in the manner of generating these comparison function values. Two well defined classes are considered here.

1. *Embedding method*: The differential equation solutions are restricted to the class of boundary value problems satisfying the conditions:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}, \boldsymbol{\beta}), \quad B_0\mathbf{x}(0) + B_1\mathbf{x}(1) = \mathbf{b}. \tag{1.5}$$

   Here the boundary matrices $B_0,\ B_1$ are imposed and $\mathbf{b}$ becomes an extra vector of parameters to be determined. The boundary matrices must be chosen in such a way that the boundary value problem has a well determined solution for the range of parameter values of interest. These methods require that the boundary value problem be solved explicitly each time a new value of the state variable is required.
2. *Simultaneous method*: The idea here is that differential equation discretization information is incorporated as explicit constraints on the state variables leading to a constrained optimization problem. In the case of the trapezoidal rule this gives

$$\mathbf{c}_i(\mathbf{x}_c) = \mathbf{x}_{i+1} - \mathbf{x}_i - \frac{h}{2}\left(\mathbf{f}_{i+1} + \mathbf{f}_i\right) = 0, \quad i = 1, 2, \cdots, n-1, \tag{1.6}$$

   with $\mathbf{x}_i = \mathbf{x}(t_i, \boldsymbol{\beta})$, $\mathbf{x}_c$ the composite vector with sub-vector components $\mathbf{x}_i$, and h the discretization mesh spacing. A feature of these methods is that the state and parameter vectors are corrected simultaneously.

Mesh selection for integrating the ODE system or defining the constraint equations would typically take the data points $\{t_i,\ i = 1, 2, \cdots, n\}$ as a starting configuration. These could be expected to be required to cluster in regions where the solution trajectory is changing rapidly. Their choice is further conditioned by two important considerations:

- The asymptotic analysis of the effects of noisy data on the parameter estimates shows that this gets small typically no faster than $O\left(n^{-1/2}\right)$.

- It is not difficult to obtain differential equation discretizations that give errors at most $O\left(n^{-2}\right)$.

This suggests that selection of the data points is a more serious consideration than reducing discretization error. Consequences include:

- That the trapezoidal rule provides an adequate integration method.
- As linear interpolation has an accuracy comparable with the trapezoidal rule it should be easily possible to integrate the differential equation on a mesh coarser than that provided by the observation points .

The basic assumption made is that the estimation problem has a well determined solution for n, the number of observations, large enough. This requirement takes slightly different forms for the two problem approaches. It becomes a stability requirement for the boundary formulation in the embedding method. This is discussed in the next section where it is shown that an "optimal" choice of boundary matrices is possible. However, the connection between stability and dichotomy suggest possible limitations to the embedding method. In the third section it is shown that the simultaneous method is capable of a number of implementations and that these can give rise to different stability considerations. It is concluded that there is likely a preferred implementation

1.2 ODE stability

The basic idea is that a system is stable if small changes to its inputs leads to small changes in its outputs. Computational considerations enter through the requirement that the discretized scheme mimic the structural properties of the original. Also this requirement could hold for all discretization scales or only for those scales small enough. These cases could be summarized as types of structurally stable discretization. In addition, in suitably controlled circumstances, it may be possible to obtain useful information by applying computational schemes to follow bounded solutions in unstable situations. Control is needed because even if the desired solution could be followed precisely in exact arithmetic it is likely unstable modes will be introduced by rounding errors and eventually swamp the computation. This is an example of numerical instability.

Initial value stability (IVS)

Here the problem considered is:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(t, \mathbf{x}\right), \quad \mathbf{x}(0) = \mathbf{b}.$$

The classical stability requirement is that solutions with close initial conditions $\mathbf{x}_1(0)$, $\mathbf{x}_2(0)$ remain close in an appropriate sense. For example:

- $\|\mathbf{x}_1(t) - \mathbf{x}_2(t)\| \to 0,\ t \to \infty$. Strong IVS.
- $\|\mathbf{x}_1(t) - \mathbf{x}_2(t)\|$ remains bounded as $t \to \infty$. Weak IVS.

In this context structurally stable discretizations which place only weak conditions on the discretization scale are described as stiffly stable. Numerical instability is an important consideration in multiple shooting [6]. Control must be exercised to ensure reasonably accurate fundamental matrices can be computed over short enough time intervals.

Example 1. Constant coefficient case: Here

$$\mathbf{f}\left(t, \mathbf{x}\right) = A\mathbf{x} - \mathbf{q}$$

If the constant matrix A is non-defective then weak IVS requires that the eigenvalues $\lambda_i(A)$ satisfy $Re\lambda_i \leq 0$, while this inequality must be strict for strong IVS.
A one-step discretization of the ODE (ignoring the q contribution) can be written

$$\mathbf{x}_{i+1} = T_h\left(A\right)\mathbf{x}_i.$$

where $T_h(A)$ is the amplification matrix. Here a stiff discretization requires the stability inequalities to map into the condition $|\lambda_i\left(T_h\right)| \leq 1$.
For the trapezoidal rule

$$\begin{aligned} |\lambda_i\left(T_h\right)| &= \left|\frac{1 + h\lambda_i(A)/2}{1 - h\lambda_i(A)/2}\right|, \\ &\leq 1 \text{ if } Re\left\{\lambda_i\left(A\right)\right\} \leq 0. \end{aligned}$$

Boundary value stability (BVS)

Here the problem is

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(t, \mathbf{x}\right), \quad B\left(\mathbf{x}\right) = B_0\mathbf{x}(0) + B_1\mathbf{x}(1) = \mathbf{b}.$$

Behaviour of perturbations about a solution trajectory $\mathbf{x}^*(t)$ is governed to first order by the linearized equation

$$L\left(\mathbf{z}\right) = \frac{d\mathbf{z}}{dt} - \nabla_x\mathbf{f}\left(t, \mathbf{x}^*(t)\right)\mathbf{z} = 0. \tag{1.7}$$

Here stability is closely related to the existence of a modest bound for the Green's matrix:

$$\begin{aligned} G\left(t, s\right) &= Z(t)\left[B_0Z(0) + B_1Z(1)\right]^{-1}B_0Z(0)Z^{-1}(s), \quad t > s, \\ &= -Z(t)\left[B_0Z(0) + B_1Z(1)\right]^{-1}B_1Z(1)Z^{-1}(s), \quad t < s. \end{aligned}$$

Where $Z(t)$ is a fundamental matrix for the linearised equation (1.7). Let α be a bound for $|G(t, s)|$. The dependence of this stability bound on the behaviour of the possible solutions $Z\mathbf{d}$ of (1.7) is explained by the idea of dichotomy:

Definition 1. *Dichotomy (weak form): $\exists$ projection P depending on the choice of Z such that, given*

$$S_1 \leftarrow \{ZP\mathbf{w},\ \mathbf{w} \in R^m\}, \quad S_2 \leftarrow \{Z\,(I-P)\,\mathbf{w},\ \mathbf{w} \in R^m\},$$

it follows that

$$\phi \in S_1 \Rightarrow \frac{|\phi(t)|}{|\phi(s)|} \le \kappa, \quad t \ge s,$$
$$\phi \in S_2 \Rightarrow \frac{|\phi(t)|}{|\phi(s)|} \le \kappa, \quad t \le s.$$

These conditions can always be satisfied if $t, s \in [0,1]$. The computational context requires modest κ. If Z satisfies $B_0 Z(0) + B_1 Z(1) = I$ then $P = B_0 Z(0)$ is a suitable projection in the sense that for separated boundary conditions an allowable setting is $\kappa = \alpha$. There is a basic equivalence between stability and dichotomy. The key paper is [2].

BVS has implications for the structural stability of possible discretizations.

- The dichotomy projection separates increasing and decreasing solutions. *Compatible* boundary conditions pin down decreasing solutions at 0, growing solutions at 1.
- Discretization needs similar property so that the given boundary conditions exercise the same control.
- This requires solutions of (1.7) which are increasing (decreasing) in magnitude to be mapped into solutions of the discretization which are increasing (decreasing) in magnitude.

This property is called di-stability in [3]. They note that the trapezoidal rule is di-stable in the constant coefficient case.

$$\lambda(A) > 0 \Rightarrow \left| \frac{1 + h\lambda(A)/2}{1 - h\lambda(A)/2} \right| > 1.$$

Example 2. The importance of compatible boundary conditions is well illustrated by the following differential equation [1].

$$A(t) = \begin{bmatrix} 1 - 19\cos 2t & 0 & 1 + 19\sin 2t \\ 0 & 19 & 0 \\ -1 + 19\sin 2t & 0 & 1 + 19\cos 2t \end{bmatrix}, \tag{1.8}$$

$$\mathbf{q}(t) = \begin{bmatrix} e^t\,(-1 + 19\,(\cos 2t - \sin 2t)) \\ -18e^t \\ e^t\,(1 - 19\,(\cos 2t + \sin 2t)) \end{bmatrix}. \tag{1.9}$$

Here the right hand side is chosen so that $\mathbf{z}(t) = e^t\mathbf{e}$ satisfies the differential equation. The fundamental matrix displays the fast and slow solutions:

$$Z(t,0) = \begin{bmatrix} e^{-18t}\cos t & 0 & e^{20t}\sin t \\ 0 & e^{19t} & 0 \\ -e^{-18t}\sin t & 0 & e^{20t}\cos t \end{bmatrix}.$$

For boundary data with two terminal conditions and one initial condition :

$$B_0 = \begin{bmatrix} 0\,0\,0 \\ 0\,0\,0 \\ 1\,0\,0 \end{bmatrix},\ B_1 = \begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,0 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} e \\ e \\ 1 \end{bmatrix},$$

the trapezoidal rule discretization scheme gives the following results. These computations are apparently satisfactory.

Table 1.1. Boundary point values – stable computation

	$\Delta t = .1$			$\Delta t = .01$		
$\mathbf{x}(0)$	1.0000	.9999	.9999	1.0000	1.0000	1.0000
$\mathbf{x}(1)$	2.7183	2.7183	2.7183	2.7183	2.7183	2.7183

In contrast, for two initial and one terminal condition:

$$B_0 = \begin{bmatrix} 0\,0\,1 \\ 0\,0\,0 \\ 1\,0\,0 \end{bmatrix},\ B_1 = \begin{bmatrix} 0\,0\,0 \\ 0\,1\,0 \\ 0\,0\,0 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 1 \\ e \\ 1 \end{bmatrix}.$$

The results are given in following Table. The effects of instability are seen clearly in the first and third solution components.

Table 1.2. Boundary point values – unstable computation

	$\Delta t = .1$			$\Delta t = .01$		
$\mathbf{x}(0)$	1.0000	.9999	1.0000	1.0000	1.0000	1.0000
$\mathbf{x}(1)$	-7.9+11	2.7183	-4.7+11	2.03+2	2.7183	1.31+2

Nonlinear stability

There are well known examples of forms of stability associated with systems of differential equations which cannot be classified as BVS. Any realization of (1.7) in which the labelling of solutions as fast or slow cannot be done unambiguously over the interval of interest, and which clearly has a local stability property provides a counterexample. One fruitful source corresponds to systems with stable limit cycles.

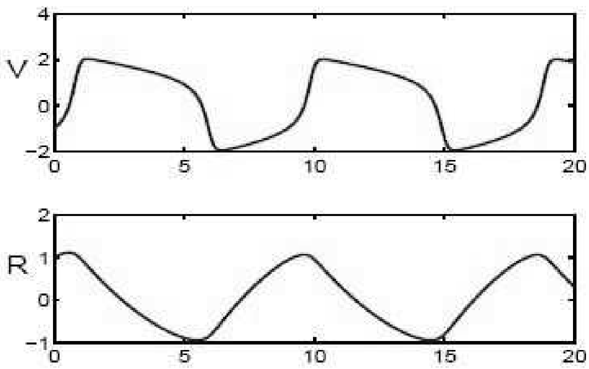
Example 3. The FitzHugh-Nagumo equations

$$\frac{dV}{dt} = \gamma\left(V - \frac{V^3}{3} + R\right), \tag{1.10}$$

$$\frac{dR}{dt} = -\frac{1}{\gamma}\left(V - \alpha - \beta R\right). \tag{1.11}$$

The limit cycle is exemplified in the case $\alpha = .2$, $\beta = .2$, $\gamma = 1.$ in figure 1.1. Figure 1.2 gives the sum of squares of discrepancies between this solution and the solution for perturbed values of the α and β parameters. It shows that the minimum is well determined in a neighbourhood of the target values, but it also shows that there are definite restrictions on the size of this neighbourhood, and that changes in solution structure would render global searching very difficult. These figures are taken from [8].

Fig. 1.1. Limit cycle trajectory.

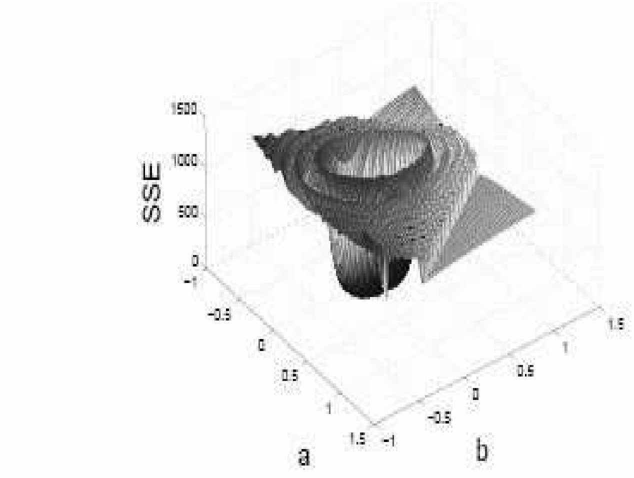


This example can be solved numerically as a boundary value problem by transforming the range of a complete cycle to $[0, 1]$, introducing the unknown range as an extra variable as in the smoothing approach, imposing periodic boundary conditions, and using (1.11) to impose a zero derivative condition at one boundary to fix the extra unknown. Thus it does not show a severe instability.

1.3 The embedding method

First problem is to set suitable boundary conditions. Expect good boundary conditions should lead to a relatively well conditioned linear system. Assume

Fig. 1.2. Objective function as function of α and β.



the ODE discretization is

$$\mathbf{c}_i\left(\mathbf{x}_i, \mathbf{x}_{i+1}\right) = \mathbf{c}_{ii}(\mathbf{x}_i) + \mathbf{c}_{i(i+1)}(\mathbf{x}_{i+1}).$$

Consider the factorization of the difference equation (gradient) matrix $C = \nabla_{\mathbf{x}} \mathbf{c}_c$ with first column permuted to end:

$$\left[\begin{array}{cccc|c} C_{12} & & & & C_{11} \\ C_{21} & C_{22} & & & \\ & & & & \\ \hline & & C_{(n-1)(n-1)} & C_{(n-1)n} & 0 \end{array}\right] \rightarrow Q \left[\begin{array}{c|c} U & V \\ \hline 0 \cdots H & G \end{array}\right] \tag{1.12}$$

This step is independent of the boundary conditions. Inserting the boundary conditions gives the system with matrix $\begin{bmatrix} H & G \\ B_1 & B_0 \end{bmatrix}$ to solve for $\mathbf{x}_1$, $\mathbf{x}_n$. Orthogonal factorization again provides a useful strategy.

$$\begin{bmatrix} H \ G \end{bmatrix} = \begin{bmatrix} L \ 0 \end{bmatrix} \begin{bmatrix} S_1^T \\ S_2^T \end{bmatrix}$$

It follows that the system determining $\mathbf{x}_1$, $\mathbf{x}_n$ is best conditioned by choosing

$$\begin{bmatrix} B_1 \ B_0 \end{bmatrix} = S_2^T. \tag{1.13}$$

These conditions depend only on the differential equation. For the Mattheij example (1.8) the "optimal" boundary matrices for $h = .1$ are given in Table 1.3. These confirm the importance of weighting the boundary data to reflect

Table 1.3. Optimal boundary matrices when $h = .1$

B_1			B_2		
.99955	0.0000	.02126	-.01819	0.0000	-.01102
0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
.02126	0.0000	.00045	.85517	0.0000	.51791

the stability requirements of a mix of fast and slow solutions. The solution does not differ from that obtained when the split into fast and slow was correctly anticipated.

Example 4. Solution of the embedding problem would typically use the Gauss-Newton method [7]. Consider the modification of the Mattheij problem (1.8) with parameters $\beta_1^* = \gamma$, and $\beta_2^* = 2$ corresponding to the solution $\mathbf{x}(t, \boldsymbol{\beta}^*) = e^t \mathbf{e}$:

$$A(t) = \begin{bmatrix} 1 - \beta_1 \cos \beta_2 t & 0 & 1 + \beta_1 \sin \beta_2 t \\ 0 & \beta_1 & 0 \\ -1 + \beta_1 \sin \beta_2 t & 0 & 1 + \beta_1 \cos \beta_2 t \end{bmatrix},$$

$$\mathbf{q}(t) = \begin{bmatrix} e^t \left(-1 + \gamma \left(\cos 2t - \sin 2t\right)\right) \\ -(\gamma - 1) e^t \\ e^t \left(1 - \gamma \left(\cos 2t + \sin 2t\right)\right) \end{bmatrix}.$$

In the numerical experiments optimal boundary conditions are set at the first iteration. The aim is to recover estimates of $\boldsymbol{\beta}^*, \mathbf{b}^*$ from simulated data $e^{t_i} H\mathbf{e} + \boldsymbol{\varepsilon}_i$, $\boldsymbol{\varepsilon}_i \sim N(0, .01I)$ using Gauss-Newton, stopping when $\nabla F \mathbf{h} < 10^{-8}$. Results are given in Table 1.4. There is relatively little change observed in the optimum boundary conditions (1.13) as $\| \begin{bmatrix} B_1 & B_2 \end{bmatrix}_1 \begin{bmatrix} B_1 & B_2 \end{bmatrix}_k^T - I \|_F < 10^{-3}$, $k > 1$. Thus no updating was deemed to be necessary.

Table 1.4. Embedding method: Gauss-Newton results for the Mattheij problem

$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$	$H = \begin{bmatrix} .5 & 0 & .5 \\ 0 & 1 & 0 \end{bmatrix}$
$n = 51,\ \gamma = 10,\ \sigma = .1$	$n = 51,\ \gamma = 10,\ \sigma = .1$
14 iterations	5 iterations
$n = 51,\ \gamma = 20,\ \sigma = .1$	$n = 51,\ \gamma = 20,\ \sigma = .1$
11 iterations	9 iterations
$n = 251,\ \gamma = 10,\ \sigma = .1$	$n = 251,\ \gamma = 10,\ \sigma = .1$
9 iterations	4 iterations
$n = 251,\ \gamma = 20,\ \sigma = .1$	$n = 251,\ \gamma = 20,\ \sigma = .1$
8 iterations	5 iterations

1.4 The simultaneous method

Associated with the equality constrained problem is the Lagrangian

$$\mathcal{L} = F(\mathbf{x}_c) + \sum_{i=1}^{n-1} \boldsymbol{\lambda}_i^T \mathbf{c}_i. \tag{1.14}$$

The necessary conditions for a stationary point give:

$$\nabla_{\mathbf{x}_i} \mathcal{L} = 0,\ i = 1, 2, \cdots, n, \quad \mathbf{c}(\mathbf{x}_c) = 0.$$

The Newton equations determining corrections $\mathbf{dx}_c, \mathbf{d}\boldsymbol{\lambda}_c$ are:

$$\nabla^2_{\mathbf{xx}} \mathcal{L} \mathbf{dx}_c + \nabla^2_{\mathbf{x}\boldsymbol{\lambda}} \mathcal{L} \mathbf{d}\boldsymbol{\lambda}_c = -\nabla_{\mathbf{x}} \mathcal{L}^T, \tag{1.15}$$

$$\nabla_{\mathbf{x}} \mathbf{c}(\mathbf{x}_c) \mathbf{dx}_c = C \mathbf{dx}_c = -\mathbf{c}(\mathbf{x}_c), \tag{1.16}$$

Note sparsity! $\nabla^2_{\mathbf{xx}} \mathcal{L}$ is block diagonal, $\nabla^2_{\mathbf{x}\boldsymbol{\lambda}} \mathcal{L} = C^T$ is block bidiagonal.

The Newton equations also correspond to necessary conditions for the quadratic program:

$$\min_{\mathbf{dx}} \nabla_{\mathbf{x}} F \mathbf{dx}_c + \frac{1}{2} \mathbf{dx}_c^T M \mathbf{dx}_c; \quad \mathbf{c} + C \mathbf{dx}_c = 0,$$

in case $M = \nabla^2_{\mathbf{xx}} \mathcal{L}$, $\boldsymbol{\lambda}^u = \boldsymbol{\lambda}_c + \mathbf{d}\boldsymbol{\lambda}_c$ [5]. A standard approach is to use the constraint equations to eliminate variables (see [4] and references given there). This can use the factorization (1.12) to give

$$\mathbf{dx}_i = \mathbf{v}_i + V_i \mathbf{dx}_1 + W_i \mathbf{dx}_n, \quad i = 2, 3, \cdots, n-1.$$

The reduced constraint equation is

$$G \mathbf{dx}_1 + H \mathbf{dx}_n = \mathbf{w}.$$

This variable elimination would appear to be restricted by BVS considerations; but there is an alternative approach called the null space method in [5]. Let $C^T = \begin{bmatrix} Q_1 \ Q_2 \end{bmatrix} \begin{bmatrix} U \\ 0 \end{bmatrix}$ then the Newton equations (1.15), (1.16) can be written

$$\begin{bmatrix} Q^T \nabla^2_{\mathbf{xx}} \mathcal{L} Q & \begin{bmatrix} U \\ 0 \end{bmatrix} \\ \begin{bmatrix} U^T \ 0 \end{bmatrix} & 0 \end{bmatrix} \begin{bmatrix} Q^T \mathbf{dx}_c \\ \boldsymbol{\lambda}^u \end{bmatrix} = - \begin{bmatrix} Q^T \nabla_{\mathbf{x}} F^T \\ \mathbf{c} \end{bmatrix}.$$

These can be solved in sequence

$$\begin{aligned} U^T Q_1^T \mathbf{dx}_c &= -\mathbf{c}, \\ Q_2^T \nabla^2_{\mathbf{xx}} \mathcal{L} Q_2 Q_2^T \mathbf{dx}_c &= -Q_2^T \nabla^2_{\mathbf{xx}} \mathcal{L} Q_1 Q_1^T \mathbf{dx}_c - Q_2^T \nabla_{\mathbf{x}} F^T, \\ U \boldsymbol{\lambda}^u &= -Q_1^T \nabla^2_{\mathbf{xx}} \mathcal{L} \mathbf{dx}_c - Q_1^T \nabla_{\mathbf{x}} F^T. \end{aligned}$$

A direct stability test is possible using the Mattheij problem data (1.8) as $Q_1^T \mathbf{dx}_c$ estimates $Q_1^T \mathrm{vec}\{\exp t_i\}$ when $\mathbf{x}_c = 0$. Computed and exact results are compared in Table 1.5.

Table 1.5. Stability test: comparison of exact and computed values

test results $n = 11$	particular integral $Q_1^T x$
.87665 -.97130 -1.0001	.87660 -.97134 -1.0001
.74089 -1.0987 -1.3432	.74083 -1.0988 -1.3432
.47327 -1.2149 -1.6230	.47321 -1.2150 -1.6231
.11498 -1.3427 -1.8611	.11491 -1.3428 -1.8612
-.32987 -1.4839 -2.0366	-.32994 -1.4840 -2.0367
-.85368 -1.6400 -2.1250	-.85376 -1.6401 -2.1250
-1.4428 -1.8125 -2.1018	-1.4429 -1.8125 -2.1019
-2.0773 -2.0031 -1.9444	-2.0774 -2.0032 -1.9444
-2.7309 -2.2137 -1.6330	-2.7310 -2.2138 -1.6331
-3.3719 -2.4466 -1.1526	-3.3720 -2.4467 -1.1527

1.5 In conclusion

- Embedding makes use of carefully constructed, explicit boundary conditions. Thus BVS restrictions must apply.
- The variable eliminations form of the simultaneous method partitions variables into sets $\{\mathbf{x}_1, \ \mathbf{x}_n\}$, and $\{\mathbf{x}_2, \cdots, \mathbf{x}_{n-1}\}$ which are found in a sequential order corresponding to a fixed pivoting sequence. This approach relies implicitly on a form of BVS .
- The null space variant partitions the variables into the sets $\left\{Q_1^T \mathbf{x}_c\right\}$, $\left\{Q_2^T \mathbf{x}_c\right\}$. It appears at least as stable as the variable elimination procedure. Sparsity preserving implementation is straightforward.

Acknowledgement

I am indebted to Giles Hooker and Jim Ramsay for their preprint [8] which gives a rather different approach to the estimation problem, and for permission to use the insightful FitzHugh-Nagumo figures taken from this paper.

References

1. U.M. Ascher, R.M.M. Mattheij, and R.D. Russell, *Numerical solution of boundary value problems for ordinary differential equations*, SIAM, Philadelphia, 1995.
2. F.R. de Hoog and R.M.M. Mattheij, *On dichotomy and well-conditioning in bvp*, SIAM J. Numer. Anal. **24** (1987), 89–105.
3. R. England and R.M.M. Mattheij, *Boundary value problems and dichotomic stability*, SIAM J. Numer. Anal. **25** (1988), 1037–1054.
4. Z. Li, M.R. Osborne, and T. Prvan, *Parameter estimation of ordinary differential equations*, IMA J. Numer. Anal. **25** (2005), 264–285.
5. J. Nocedal and S.J. Wright, *Numerical optimization*, Springer Verlag, 1999.

6. M. R. Osborne, *On shooting methods for boundary value problems*, J. Math. Analysis and Applic. **27** (1969), 417–433.
7. M.R. Osborne, *An approach to parameter estimation and model selection in differential equations*, Modeling, Simulation and Optimization of Complex Processes (H.G. Bock, Ekaterina Kostina, Hoang Xuan Phu, and R. Rannacher, eds.), Springer, 2005, pp. 393–408.
8. J.O. Ramsay, G. Hooker, C. Cao, and C. Campbell, *Estimating differential equations*, preprint, Department of Psychology, McGill University, Montreal, Canada, 2005, p. 40.