Make your way to CTAC08, ANU, Canberra
http://wwwmaths.anu.edu.au/events/ctac08/

ODE estimation – statistical properties and numerical problems

M.R. Osborne

Mathematical Sciences Institute
Australian National University

Statistical Methods for Modelling Dynamic Systems
Montreal, July 9–15

Outline

Explicit parameters

Start with signal measured in the presence of noise giving independent event outcomes $\mathbf{y}_t \in R^q$ and associated pdf $g\left(\mathbf{y}_t; \boldsymbol{\theta}_t, \mathbf{t}\right)$ indexed by "points" $\mathbf{t} \in T_n \subset R^l$, and structural information provided by a known parametric model

$$\boldsymbol{\theta}_t = \boldsymbol{\eta}\left(\mathbf{x}, \mathbf{t}\right), \ \mathcal{E}\left\{\mathbf{y}_t\right\} = \boldsymbol{\eta}\left(\mathbf{x}, \mathbf{t}\right)$$

where $\boldsymbol{\theta} \in R^q$, and $\mathbf{x} \in R^p$. Given the event outcomes $\mathbf{y}_t$ it is required to estimate the actual parameter values $\mathbf{x}^*$.
A priori information is the condition for a planned experiment. This is needed for asymptotics. Let $T_n \subset S(T), \ |T_n| = n$. Require

$$\frac{1}{n} \sum_{\mathbf{t} \in T_n} f(\mathbf{t}) \rightarrow \int_{S(T)} f(\mathbf{t}) \rho(\mathbf{t}) d\mathbf{t}$$

go ICTE

Setting the objective

Likelihood: $\mathcal{G}_n(\mathbf{y}; \mathbf{x}, T_n) = \prod_{\mathbf{t} \in T_n} g(\mathbf{y}_t; \boldsymbol{\theta}_t, \mathbf{t})$
Estimation principle: $\widehat{\mathbf{x}}_n = \arg\max_{\mathbf{x}} \mathcal{G}_n(\mathbf{y}; \mathbf{x}, T_n)$.
Target objective function is *log likelihood*:

$$\mathcal{F}_n(\mathbf{y}; \mathbf{x}, T_n) = \sum_{\mathbf{t} \in T_n} \log g(\mathbf{y}_t; \boldsymbol{\theta}_t, \boldsymbol{t}) = \sum_{\mathbf{t} \in T_n} F(\mathbf{y}_t; \boldsymbol{\theta}_t, \mathbf{t})$$

Assume:

- $\exists$ true model $\boldsymbol{\eta}$, parameter vector $\mathbf{x}^*$;
- $\mathbf{x}^*$ properly in interior of region in which $\mathcal{F}_n$ is well behaved;
- boundedness of integrals (computing expectations etc), adequate smoothness.

Necessary conditions

The necessary conditions for a maximum plus an application of the law of large numbers lead to a limiting equation satisfied by $\mathbf{x}^*$.

$$0 = \frac{1}{n} \sum_{\mathbf{t} \in T_n} \nabla_x F\left(\mathbf{y}_t; \eta\left(\mathbf{x}, \mathbf{t}\right), \mathbf{t}\right),$$

go EXPP $\mathcal{E}^*$ corresponds to expectation computed with $\mathbf{x} = \mathbf{x}^*$.

Necessary conditions

The necessary conditions for a maximum plus an application of the law of large numbers lead to a limiting equation satisfied by $\mathbf{x}^*$.

$$0 = \frac{1}{n} \sum_{\mathbf{t} \in T_n} \nabla_x F\left(\mathbf{y}_t ; \eta\left(\mathbf{x}, \mathbf{t}\right), \mathbf{t}\right),$$

$$= \frac{1}{n} \sum_{\mathbf{t} \in T_n} \nabla_x F\left(\mathbf{y}_t ; \eta\left(\mathbf{x}, \mathbf{t}\right), \mathbf{t}\right) - \mathcal{E}^* \left\{ \frac{1}{n} \sum_{\mathbf{t} \in T_n} \nabla_x F\left(\mathbf{y}_t ; \eta\left(\mathbf{x}, \mathbf{t}\right), \mathbf{t}\right) \right\}$$

$$+ \mathcal{E}^* \left\{ \frac{1}{n} \sum_{\mathbf{t} \in T_n} \nabla_x F\left(\mathbf{y}_t ; \eta\left(\mathbf{x}, \mathbf{t}\right), \mathbf{t}\right) \right\},$$

go EXPP $\mathcal{E}^*$ corresponds to expectation computed with $\mathbf{x} = \mathbf{x}^*$.

Necessary conditions

The necessary conditions for a maximum plus an application of the law of large numbers lead to a limiting equation satisfied by $\mathbf{x}^*$.

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{\mathbf{t}\in T_n} \nabla_x F\left(\mathbf{y}_t;\boldsymbol{\eta}\left(\mathbf{x},\mathbf{t}\right),\mathbf{t}\right),\\
&= \frac{1}{n}\sum_{\mathbf{t}\in T_n} \nabla_x F\left(\mathbf{y}_t;\boldsymbol{\eta}\left(\mathbf{x},\mathbf{t}\right),\mathbf{t}\right) - \mathcal{E}^*\left\{\frac{1}{n}\sum_{\mathbf{t}\in T_n} \nabla_x F\left(\mathbf{y}_t;\boldsymbol{\eta}\left(\mathbf{x},\mathbf{t}\right),\mathbf{t}\right)\right\}\\
&+ \mathcal{E}^*\left\{\frac{1}{n}\sum_{\mathbf{t}\in T_n} \nabla_x F\left(\mathbf{y}_t;\boldsymbol{\eta}\left(\mathbf{x},\mathbf{t}\right),\mathbf{t}\right)\right\},\\
&\overset{a.s.}{\to} \int_{S(T)} \mathcal{E}^*\left\{\nabla_x F\left(\mathbf{y};\boldsymbol{\eta}\left(\mathbf{x},\mathbf{t}\right),\mathbf{t}\right)\right\}\rho\left(\mathbf{t}\right)d\mathbf{t},\ n\to\infty.
\end{aligned}
$$

go EXPP $\mathcal{E}^*$ corresponds to expectation computed with $\mathbf{x}=\mathbf{x}^*$.

Consistency, limiting distribution

To prove $\widehat{\mathbf{x}}_n \xrightarrow{a.s.} \mathbf{x}^*$ can apply Newton's method to the necessary conditions

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \mathcal{J}_n\left(\mathbf{x}_i\right)^{-1} \frac{1}{n} \nabla_x \mathcal{F}_n\left(\mathbf{x}_i\right)^T,$$

with starting value $\mathbf{x}^*$ to give a small residual for n large enough and use the Kantorovich theorem.
The limiting distribution of the parameter estimates is obtained by expanding the necessary conditions about $\mathbf{x}^*$. This gives

$$\sqrt{n}\left(\widehat{\mathbf{x}} - \mathbf{x}^*\right) \sim N\left(0, \mathcal{I}\left(\mathbf{x}^*\right)^{-1}\right).$$

This is a very slow rate of convergence. If the actual parameter values are needed then so are lots of data.

Scoring/Gauss-Newton

This is a modified Newton iteration with the basic form:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathcal{I}_n(\mathbf{x}_i)^{-1} \frac{1}{n} \nabla_x \mathcal{F}_n(\mathbf{x}_i)^T .$$

The logic in using the expected Hessian, which is independent of the observed data, is as follows:

$$\begin{array}{ccccc} -\mathcal{J}_n(\mathbf{x}^*) & \stackrel{a.s.}{\to} & \mathcal{I}(\mathbf{x}^*) & & \\ & & \approx & & \|\mathbf{x}^* - \mathbf{x}\| \text{ small} \\ \mathcal{I}_n(\mathbf{x}) & \to & \mathcal{I}(\mathbf{x}) & & \end{array}$$

Table: Scoring diagram

Here is the relationship between these terms:

$$\mathcal{J}_n(\mathbf{x}^*) = \frac{1}{n}\sum_{t\in T_n} \nabla_x^2 F(\mathbf{x}^*) \xrightarrow{a.s.} \int_{S(T)} \mathcal{E}^*\left\{\nabla_x^2 F(\mathbf{x}^*)\right\}\rho(\mathbf{t})\,d\mathbf{t}$$

$$= -\int_{S(T)} \mathcal{E}^*\left\{\nabla_x F(\mathbf{x}^*)^T \nabla_x F(\mathbf{x}^*)\right\}\rho(\mathbf{t})\,d\mathbf{t} = -\mathcal{I}(\mathbf{x}^*),$$

$$\mathcal{I}_n(\mathbf{x}) = \frac{1}{n}\mathcal{E}\left\{\sum_{t\in T_n} \nabla_x F_t(\mathbf{x})^T \nabla_x F_t(\mathbf{x})\right\}, \text{ Fisher information,}$$

$$\to \int_{S(T)} \mathcal{E}\left\{\nabla_x F(\mathbf{x})^T \nabla_x F(\mathbf{x})\right\}\rho(\mathbf{t})\,d\mathbf{t} = \mathcal{I}(\mathbf{x})$$

Iteration properties

Advantages in using $\mathcal{I}_n$ include:
1.Avoids calculation of second derivatives.
2. Provides a generically positive definite replacement for the Hessian $\mathcal{J}_n$. This suggests enhanced convergence properties.
3. Possesses excellent transformation invariance properties.
4. Each iteration can be reduced to the solution of a linear least squares problem by orthogonal transformation techniques.

Disadvantage is the generic first order convergence rate. Can be serious except in cases:
1. accurate measurements (small σ),
2. large data sets (n large),
when asymptotic properties are good.

Rate of convergence 1

Consider the unit step scoring iteration in fixed point form:

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

where

$$Q_n\left(\mathbf{x}\right) = \mathbf{x} + \mathcal{I}_n\left(\mathbf{x}\right)^{-1}\frac{1}{n}\nabla_{\mathbf{x}}\mathcal{F}_n\left(\mathbf{x}\right)^T.$$

The condition for convergence is

$$\varpi\left(Q_n'\left(\widehat{\mathbf{x}}_n\right)\right) < 1,$$

where $\varpi\left(Q_n'\left(\widehat{\mathbf{x}}_n\right)\right)$ is the spectral radius of the variation $Q_n' = \nabla_x Q_n$.

$\varpi\left(Q_n'\left(\widehat{\mathbf{x}}_n\right)\right)$ is an invariant of the likelihood surface, is a measure of the quality of the modelling, and can be estimated by a modification of the power method.

Rate of convergence 2

To calculate $\varpi\left(Q_n'\left(\widehat{\mathbf{x}}_n\right)\right)$ note that $\nabla_{\mathbf{x}}\mathcal{F}_n\left(\widehat{\mathbf{x}}_n\right)=0$. Thus

$$\begin{aligned} Q_n'\left(\widehat{\mathbf{x}}_n\right) &= I+\mathcal{I}_n\left(\widehat{\mathbf{x}}_n\right)^{-1}\frac{1}{n}\nabla^2_{\mathbf{x}}\mathcal{F}_n\left(\widehat{\mathbf{x}}_n\right), \\ &= \mathcal{I}_n\left(\widehat{\mathbf{x}}_n\right)^{-1}\left(\mathcal{I}_n\left(\widehat{\mathbf{x}}_n\right)+\frac{1}{n}\nabla^2_{\mathbf{x}}\mathcal{F}_n\left(\widehat{\mathbf{x}}_n\right)\right). \end{aligned}$$

If the right hand side were evaluated at $\mathbf{x}^*$ then the result $\varpi\left(Q_n'\left(\mathbf{x}^*\right)\right)\overset{a.s.}{\to}0, n\to\infty$ would follow from the strong law of large numbers which shows that the matrix gets small (hence ϖ gets small) almost surely as $n\to\infty$. But, by consistency of the estimates, we have

$$\varpi\left(Q_n'\left(\widehat{\mathbf{x}}_n\right)\right)=\varpi\left(Q_n'\left(\mathbf{x}^*\right)\right)+O\left(\left\|\widehat{\mathbf{x}}_n-\mathbf{x}^*\right\|\right),$$

and the desired result follows.

The differential equation

Consider the differential equation:

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

where $\mathbf{x}, \mathbf{f} \in R^m$, $\beta \in R^p, t \in [0, 1]$. The general solution of this equation has m implicit degrees of freedom that must be fixed in any particular solution in addition to the p associated with the explicit vector of parameters β . Thus the solution manifold relevant to the parameter estimation problem has $m + p$ degrees of freedom. The implicit degrees of freedom are fixed typically by satisfying explicit additional conditions. For example, boundary conditions

$$B_0\mathbf{x}\left(0\right) + B_1\mathbf{x}\left(1\right) = \mathbf{b},$$

where $B_0, B_1 : R^m \to R^m$, $\mathbf{b} \in R^m$.

Approximating the ODE

go GNM A second problem is that the ODE solution manifold can only be approximated. However, this is minor. The procedure for integrating the ODE system is conditioned by two important considerations:

- The asymptotic analysis of the effects of noisy data on the parameters shows that this gets small no faster than $O\left(n^{-1/2}\right)$ under planned experiment conditions.
- It is not difficult to obtain ODE discretizations that give solution errors at most $O\left(n^{-2}\right)$.

Approximating the ODE

go GNM A second problem is that the ODE solution manifold can only be approximated. However, this is minor. The procedure for integrating the ODE system is conditioned by two important considerations:

- The asymptotic analysis of the effects of noisy data on the parameters shows that this gets small no faster than $O\left(n^{-1/2}\right)$ under planned experiment conditions.
- It is not difficult to obtain ODE discretizations that give solution errors at most $O\left(n^{-2}\right)$.

This suggests that the trapezoidal rule provides an adequate integration method. It is known to be endowed with attractive properties. Let $\mathbf{x}_c$ be the composite vector with components $\mathbf{x}_i,\ i = 1, 2, \cdots, n$.

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

Linear case

$$\mathbf{f}(t,\mathbf{x}) = A(t)\mathbf{x} + \mathbf{q}(t).$$

Let fundamental matrix $X(t,\xi)$ satisfy the IVP

$$\frac{dX}{dt} = A(t)X, \quad X(\xi,\xi) = I$$

then BVP has a solution provided $(B_0 + B_1 X(1,0))$ has a bounded inverse. The Green's matrix is

$$\begin{aligned} G(t,s) &= X(t)\left[B_0 X(0) + B_1 X(1)\right]^{-1} B_0 X(0) X^{-1}(s), t > s, \\ &= -X(t)\left[B_0 X(0) + B_1 X(1)\right]^{-1} B_1 X(1) X^{-1}(s), t < s. \end{aligned}$$

Note G does not depend on the initial condition on X. The magnitude of G is an indicator of problem stability. Set *stability constant* $\alpha = \max_{t,s} \|G(t,s)\|_2$.

Dichotomy: Key paper is de Hoog and Mattheij

This is the structural property that connects linear BVP stability with the detailed behaviour of the range of possible solutions. Weak form: $\exists$ projection P depending on choice of X such that, given

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

$$\phi \in S_1 \Rightarrow \frac{\|\phi(t)\|_2}{\|\phi(s)\|_2} \leq \kappa, \quad t \geq s,$$
$$\phi \in S_2 \Rightarrow \frac{\|\phi(t)\|_2}{\|\phi(s)\|_2} \leq \kappa, \quad t \leq s.$$

Computational context happy with modest κ for $t, s \in [0,1]$. If X satisfies $B_0X(0) + B_1X(1) = I$ then $P = B_0X(0)$ is a suitable projection in sense that for separated boundary conditions can take $\kappa = \alpha$. Dichotomy is sufficient for BVP stability.

BVS restricts possible discretizations

- Sense in which dichotomy projection separates increasing and decreasing solutions. *dichotomy compatible* BC's pin down decreasing solutions at 0, growing solutions at 1.

BVS restricts possible discretizations

- Sense in which dichotomy projection separates increasing and decreasing solutions. *dichotomy compatible* BC's pin down decreasing solutions at 0, growing solutions at 1.
- Discretization needs similar property so given BC's exercise same control.

BVS restricts possible discretizations

- Sense in which dichotomy projection separates increasing and decreasing solutions. *dichotomy compatible* BC's pin down decreasing solutions at 0, growing solutions at 1.
- Discretization needs similar property so given BC's exercise same control.
- This requires solutions of ODE which are increasing (decreasing) in magnitude to be mapped into solutions of discretization which are increasing (decreasing) in magnitude.

BVS restricts possible discretizations

- Sense in which dichotomy projection separates increasing and decreasing solutions. *dichotomy compatible* BC's pin down decreasing solutions at 0, growing solutions at 1.
- Discretization needs similar property so given BC's exercise same control.
- This requires solutions of ODE which are increasing (decreasing) in magnitude to be mapped into solutions of discretization which are increasing (decreasing) in magnitude.

This property called di-stability by England and Mattheij who showed the TR is di-stable in constant coefficient case. go MatEx

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

Bob Mattheij's example 1

Consider the differential system defined by

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

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

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:

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

Bob Mattheij's example 2

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.

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

Table: Boundary point values - stable computation

These computations are apparently satisfactory.

Bob Mattheij's example 3

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.

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

Table: Boundary point values - unstable computation

The effects of instability are seen clearly in the first and third solution components. go EMDS

Nonlinear stability

The IVP/BVP stability requirements are restrictive in sense that the classification into increasing/decreasing solutions is emphasised.

Nonlinear stability

The IVP/BVP stability requirements are restrictive in sense that the classification into increasing/decreasing solutions is emphasised.

Important conflicting examples occur in dynamical systems. These

- can have a stable character - for example, limiting trajectories which attract neighboring orbits;
- clearly cannot satisfy the IVP/BVP stability requirements.

Nonlinear stability

The IVP/BVP stability requirements are restrictive in sense that the classification into increasing/decreasing solutions is emphasised.
Important conflicting examples occur in dynamical systems. These

- can have a stable character - for example, limiting trajectories which attract neighboring orbits;
- clearly cannot satisfy the IVP/BVP stability requirements.

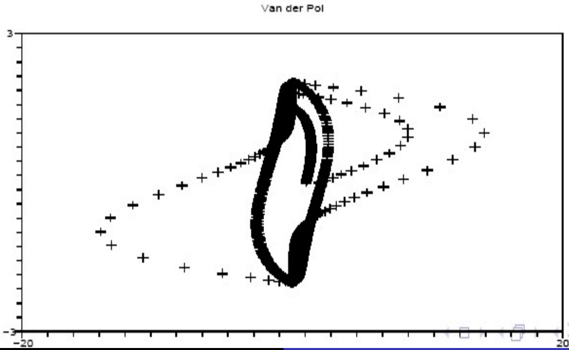
Limit cycle behavior provides a familiar example that is of this type.

Example 1 - Van der Pol equation

$$\frac{d^2x}{dt^2} - \lambda\left(1 - x^2\right)\frac{dx}{dt} + x = 0.$$

Reliable,"difficult" ODE example with difficulty increasing with λ.

scilab plot shows convergence to limit cycle for $\lambda = 1, 10$.



Example 1 - BVP formulation 1

Transformation $s = 4t/T$ puts $1/2$ period onto $[0, 2]$. Set $x_3 = T/4$. The ODE becomes

$$\frac{dx_1}{ds} = x_2, \quad \frac{dx_3}{ds} = 0$$
$$\frac{dx_2}{ds} = \lambda\left(1 - x_1^2\right)x_2x_3 - x_1x_3^2.$$

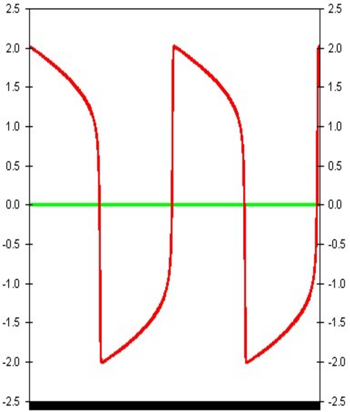
Boundary data is

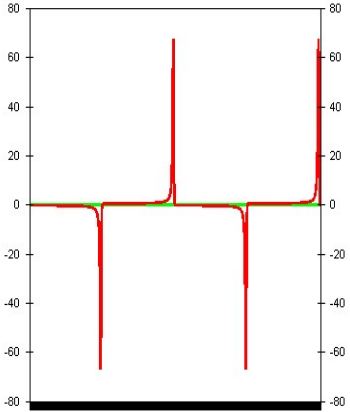
$$B_0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{b} = 0.$$

Solution for $\lambda = 0$ provides initial estimate for $\lambda = 1$. Continuation with $\Delta\lambda = 1$ used for higher values. $n = 1001$. DE discretized at shifted Chebyshev extrema.

Example 1 - BVP formulation 2

BVP results for $\lambda = 10$. Extra values by reflection.





Iteration details

Newton iteration, tolerance $= 1.e^{-10}$, line search based on
$\left\{\sum \|\mathbf{c}_i\|^2 / (t_{i+1} - t_i) + \|B_0\mathbf{x}_1 + B_1\mathbf{x}_n - \mathbf{b}\|^2\right\}^{1/2}$.

λ	(LS)/NI	(Approx. Cnd.) $* 10^{-2}$	$T/4$
1	(1)/5	0.2199	1.6658
2	(1,2)/5	0.1986	1.9075
3	(2,3)/6	0.3106	2.2148
4	(2,3)/6	0.4622	2.5509
5	(2,3)/6	0.6264	2.9030
6	(2,3)/6	0.7969	3.2654
7	(2,3)/6	0.9677	3.6349
8	(2,3)/6	1.1407	4.0095
9	(1,2)/5	1.3142	4.3881
10	(1,2)/5	1.4879	4.7697

Stability consequences

The ODE stability conditions provide sharp distinctions - in part because they are specifying global properties. Computational requirements force compromise.

In the IVP this is provided by various control devices: for example, automatic step length control.

Stability consequences

The ODE stability conditions provide sharp distinctions - in part because they are specifying global properties. Computational requirements force compromise.

In the IVP this is provided by various control devices: for example, automatic step length control.

In BVP fudge dichotomy considerations to finite interval and ask for "moderate" κ. There is an exact discretization (multiple shooting). Can write down the inverse of this matrix as $h \rightarrow 0$. It is limit of corresponding inverses of discretization matrices. Components in this limit can be interpreted using the Green's matrix and bounded by the stability constant. In practice a more unstable BVP is associated with larger bounds and a more sensitive Newton iteration. Available tools include:

- adaptive mesh control;
- continuation.

The objective

Estimation principles (least squares, (-) maximum likelihood) consider the objective:

$$\mathcal{F}_n\left(\mathbf{x}_c,\beta\right) = \frac{1}{2}\sum_{t\in T_n}\left\|\mathbf{y}_t - H\mathbf{x}\left(t,\beta\right)\right\|_2^2 = \frac{1}{2}\sum_{t\in T_n}\left\|\mathbf{r}_t\right\|_2^2 .$$

Here the observations are assumed to have the form

$$\mathbf{y}_t = H\mathbf{x}_t^* + \varepsilon_t,\ t \in [0,1]\,,$$

where $H : R^m \to R^q$, and $\varepsilon_t \sim N\left(0,\sigma^2 I_q\right)$.
For simplicity of presentation it is assumed that the points at which the observations are made coincide with the points at which the ODE is discretized.
Methods for estimating β differ in the way in which comparison function values $\mathbf{x}(t_i,\beta),\ i = 1,2,\cdots,n$ are generated in the minimization problem.

Embedding

The embedding method introduces boundary matrices B_0, B_1 and extra parameters $\mathbf{b} \in R^m$ so that $\beta, \mathbf{b}$ parametrise the solution manifold. Comparison values $\mathbf{x}(t_i, \beta, \mathbf{b})$ satisfy BVP

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

The resulting estimation problem has some advantages:

- It can adapt standard BVP software which can provide adaptive meshing and continuation facilities.

The cost involved is that the BVP must be solved for each function value required.

Embedding

The embedding method introduces boundary matrices $B_0, \ B_1$ and extra parameters $\mathbf{b} \in R^m$ so that $\beta, \mathbf{b}$ parametrise the solution manifold. Comparison values $\mathbf{x}(t_i, \beta, \mathbf{b})$ satisfy BVP

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

The resulting estimation problem has some advantages:

- It can adapt standard BVP software which can provide adaptive meshing and continuation facilities.
- Similarly some modification may be needed to use a standard nonlinear least squares program.

The cost involved is that the BVP must be solved for each function value required.

System factorization

First problem is to set suitable boundary conditions $B_0, \; B_1$. Expect good choice of boundary conditions should lead to a relatively well conditioned linear system for the Newton iteration. Write the trapezoidal rule discretization as go SVE

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

Consider the orthogonal factorization of the difference equation (gradient) matrix with first column permuted to end:

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

This step is independent of the boundary conditions.

Optimal boundary conditions

The boundary conditions can be inserted at this point. This gives the system with matrix $\begin{bmatrix} H & G \\ B_1 & B_0 \end{bmatrix}$ to solve for $\mathbf{x}_n, \mathbf{x}_1$. 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 $\mathbf{x}_n, \mathbf{x}_1$ is best conditioned by choosing

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

These boundary conditions depend only on the ODE, and S_2 is well defined as $n \to \infty$.

BC's for Mattheij example

The “optimal” boundary matrices corresponding to $h = .1$ are given in the Table. These confirm the importance of weighting the boundary data to reflect 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.

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

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

Gauss-Newton details

Let $\nabla_{(\beta,b)}\mathbf{x} = \left[\frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}, \frac{\partial \mathbf{x}}{\partial \mathbf{b}}\right]$, $\mathbf{r}_i = \mathbf{y}_i - H\mathbf{x}\left(t_i, \boldsymbol{\beta}, \mathbf{b}\right)$ then the gradient of $\mathcal{F}_n$ is

$$\nabla_{(\beta,b)}\mathcal{F}_n = -\sum_{i=1}^{n} \mathbf{r}_i^T H \nabla_{(\beta,b)}\mathbf{x}_i.$$

The gradient terms wrt $\boldsymbol{\beta}$ are found by solving the BVP's

$$B_0 \frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}(0) + B_1 \frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}(1) = 0,$$
$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}} = \nabla_x \mathbf{f} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}} + \nabla_\beta \mathbf{f},$$

Gauss-Newton details

Let $\nabla_{(\beta,b)}\mathbf{x} = \left[\frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}, \frac{\partial \mathbf{x}}{\partial \mathbf{b}}\right]$, $\mathbf{r}_i = \mathbf{y}_i - H\mathbf{x}(t_i, \boldsymbol{\beta}, \mathbf{b})$ then the gradient of $\mathcal{F}_n$ is

$$\nabla_{(\beta,b)}\mathcal{F}_n = -\sum_{i=1}^{n} \mathbf{r}_i^T H \nabla_{(\beta,b)}\mathbf{x}_i.$$

while the gradient terms wrt $\mathbf{b}$ satisfy the BVP's

$$B_0 \frac{\partial \mathbf{x}}{\partial \mathbf{b}}(0) + B_1 \frac{\partial \mathbf{x}}{\partial \mathbf{b}}(1) = I,$$

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \mathbf{b}} = \nabla_x \mathbf{f} \frac{\partial \mathbf{x}}{\partial \mathbf{b}}.$$

Embedding: Again the Mattheij example

Consider the modification of the Mattheij problem with parameters $\beta_1^* = \gamma$, and $\beta_2^* = 2$ corresponding to the solution $\mathbf{x}(t, \beta^*) = e^t \mathbf{e}$:

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

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

In the numerical experiments optimal boundary conditions are set at the first iteration. The aim is to recover estimates of $\beta^*, \mathbf{b}^*$ from simulated data $e^{t_i} H\mathbf{e} + \varepsilon_i$, $\varepsilon_i \sim N(0, .01I)$ using Gauss-Newton, stopping when $\nabla \mathcal{F}_n \mathbf{h} < 10^{-8}$. go NSMM

Embedding: Again the Mattheij example

go NSMM

$$H = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$H = \begin{bmatrix} .5 & 0 & .5 \\ 0 & 1 & 0 \end{bmatrix}$$

$n = 51,\ \gamma = 10,\ \sigma = .1$
14 iterations
$n = 51,\ \gamma = 20,\ \sigma = .1$
11 iterations
$n = 251,\ \gamma = 10,\ \sigma = .1$
9 iterations
$n = 251,\ \gamma = 20,\ \sigma = .1$
8 iterations

$n = 51,\ \gamma = 10,\ \sigma = .1$
5 iterations
$n = 51,\ \gamma = 20,\ \sigma = .1$
9 iterations
$n = 251,\ \gamma = 10,\ \sigma = .1$
4 iterations
$n = 251,\ \gamma = 20,\ \sigma = .1$
5 iterations

Here $\left\| \begin{bmatrix} B_1 & B_2 \end{bmatrix}_1 \begin{bmatrix} B_1 & B_2 \end{bmatrix}_k^T - I \right\|_F < 10^{-3}$, $k > 1$.

The constrained problem

For purposes of presentation only note $\frac{d\beta}{dt} = 0$. We introduce the parameters as extra solution variables $\{\mathbf{x}_i\}_{m+1}, \cdots, \{\mathbf{x}_i\}_{m+p}$, $i = 1, 2, \cdots, n$, and set $m \leftarrow m + p$.
The simultaneous method treats the discretized ODE as a set of constraints so the estimation problem becomes

$$\min_{\mathbf{x}_c} \frac{1}{n} \mathcal{F}_n(\mathbf{x}_c); \ \mathbf{c}_i(\mathbf{x}_c) = 0, \ i = 1, 2, \cdots, n-1.$$

The problem Lagrangian is

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

where the $\boldsymbol{\lambda}_i$ are the Lagrange multipliers. Must solve:

$$\nabla_{\mathbf{x}_i} \mathcal{L} = 0, \ i = 1, 2 \cdots, n; \ \mathbf{c}_i = 0, \ i = 1, 2, \cdots, n-1.$$

Solving the necessary conditions

Here the gradient of the Lagrangian gives the equations

$$-\frac{1}{n}\mathbf{r}_1^T H + \boldsymbol{\lambda}_1^T \nabla_{\mathbf{x}_1}\mathbf{c}_{11} = 0,$$

$$-\frac{1}{n}\mathbf{r}_i^T H + \boldsymbol{\lambda}_{i-1}^T \nabla_{\mathbf{x}_i}\mathbf{c}_{(i-1)i} + \boldsymbol{\lambda}_i^T \nabla_{\mathbf{x}_i}\mathbf{c}_{ii} = 0, \quad i = 2, 3, \cdots, n-1,$$

$$-\frac{1}{n}\mathbf{r}_n^T H + \boldsymbol{\lambda}_{n-1}^T \nabla_{\mathbf{x}_n}\mathbf{c}_{(n-1)n} = 0, .$$

The Newton equations determining corrections $\mathbf{dx}_c, \mathbf{d}\boldsymbol{\lambda}_c$ to current estimates of state and multiplier vector solutions of these equations are:

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

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

Details

Setting $\mathbf{s}\left(\boldsymbol{\lambda}_c\right)_i = \boldsymbol{\lambda}_{i-1} + \boldsymbol{\lambda}_i,\ \boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_n = 0,\ i = 1, 2, \cdots, n$, and making use of the block separability of the Lagrangian:

$$\nabla^2_{\mathbf{x}}\mathcal{L} = \mathrm{diag}\left\{\frac{1}{n}H^T H - \frac{h}{2}\nabla^2_{\mathbf{x}_i}\left(\mathbf{s}\left(\boldsymbol{\lambda}_c\right)_i^T \mathbf{f}\left(t_i, \mathbf{x}_i\right)\right),\ i = 1, 2, \cdots, n\right\},$$

$$\nabla^2_{\boldsymbol{\lambda}\mathbf{x}}\mathcal{L} = C^T,$$

$$C_{ii} = -I - \frac{h}{2}\nabla_{\mathbf{x}_i}\mathbf{f}\left(t_i, \mathbf{x}_i\right),$$

$$C_{i(i+1)} = I - \frac{h}{2}\nabla_{\mathbf{x}_{i+1}}\mathbf{f}\left(t_{i+1}, \mathbf{x}_{i+1}\right).$$

Note that the choice of the trapezoidal rule makes $\nabla^2_{\mathbf{x}}\mathcal{L}$ block diagonal, and that the constraint matrix $C : R^{nm} \rightarrow R^{(n-1)m}$ is block bidiagonal.

There is some structure in $\boldsymbol{\lambda}$

Grouping terms in the necessary conditions gives

$$-\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{i+1} + \frac{h}{2}\nabla_{\mathbf{x}_i}\mathbf{f}_{i+1}^T\left(\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_{i+1}\right) = -\frac{1}{n}H^T\mathbf{r}_i.$$

For simplicity consider the case where r_i is a scalar and the observation structure is based on a vector representer $H = \mathbf{o}^T$. Then

$$\begin{aligned} r_i H^T &= \left\{\varepsilon_i + \mathbf{o}^T\left(\mathbf{x}_i^* - \mathbf{x}_i\right)\right\}\mathbf{o}, \\ &= \sqrt{n}\left\{\frac{\varepsilon_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\mathbf{o}^T\left(\mathbf{x}_i^* - \mathbf{x}_i\right)\right\}\mathbf{o}. \end{aligned}$$

Let $\mathbf{w}_i = \sqrt{n}\boldsymbol{\lambda}_i,\ i = 1, 2, \cdots, n-1$, then

$$-\mathbf{w}_i + \mathbf{w}_{i+1} + \frac{h}{2}\nabla_{\mathbf{x}_i}\mathbf{f}_{i+1}^T\left(\mathbf{w}_i + \mathbf{w}_{i+1}\right) = -\frac{r_i}{\sqrt{n}}\mathbf{o}.$$

Multiplier estimate

This equation is important!

$$-\mathbf{w}_i + \mathbf{w}_{i+1} + \frac{h}{2}\nabla_{\mathbf{x}_i}\mathbf{f}_{i+1}^T\left(\mathbf{w}_i + \mathbf{w}_{i+1}\right) = -\frac{r_i}{\sqrt{n}}\mathbf{o}.$$

In this rescaled form the variance of the stochastic forcing term is $\left(\sigma^2/n\right)\mathbf{oo}^T$, and the remaining right hand side term is essentially deterministic with scale $O\{1/n\}$ when the generic $O\{n^{-1/2}\}$ rate of convergence of the estimation procedure is taken into account. This permits identification with a discretization of the adjoint to the linearised constraint differential equation system subject to a forcing term which contains a stochastic component. go Stoch The significant feature of this comparison is that it indicates that the multipliers $\lambda_i \to 0,\ i = 1, 2, \cdots, n-1$, on a scale which is $O\left(n^{-1/2}\right)$ as $n \to \infty$.

Example of multiplier behaviour

The effect of the random walk term can be isolated in the smoothing problem with data:

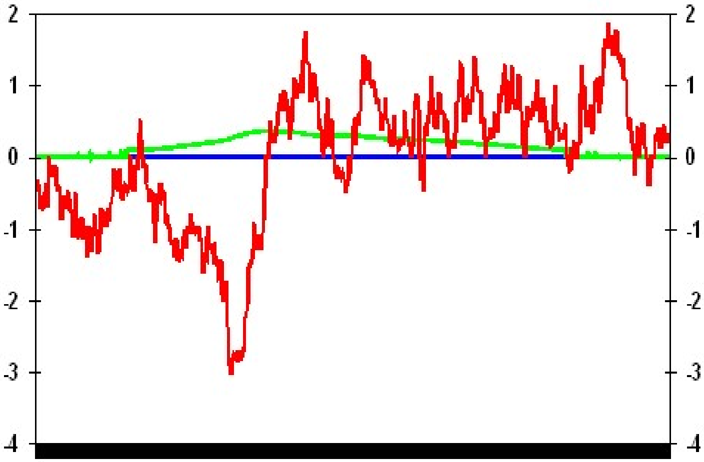
$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x},$$

$$y_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_i + \varepsilon_i = 1 + \varepsilon_i, \; \varepsilon_i \sim N(0, 1),$$

$$t_i = \frac{(i-1)}{(n-1)}, \; i = 1, 2, \cdots, n.$$

The trapezoidal rule is exact for this differential equation. The scaled solution $\mathbf{w}_i, \; i = 1, 2, \cdots, n-1$ obtained for a particular realisation of the ε_i for $n = 501, \; \sigma = 5$ is plotted below. The relation between the scale of the standard deviation σ and that of $\mathbf{w}$ seems typical. This provides a good illustration that the $n^{-1/2}$ scaling leads to an $O(1)$ result.

Scaled Lagrange multiplier plot



The null space method

Let $C^T = S \begin{bmatrix} U \\ 0 \end{bmatrix}$, where S is orthogonal and $U : R^{(n-1)m} \rightarrow R^{(n-1)m}$ is upper triangular, $S = \begin{bmatrix} S_1 & S_2 \end{bmatrix}$, $S_1 : R^{(n-1)m} \rightarrow R^{nm}$, $S_2 : R^m \rightarrow R^{nm}$. Then the Newton equations can be written

$$\begin{bmatrix} S^T \nabla_{\mathbf{x}}^2 \mathcal{L} S & \begin{bmatrix} U \\ 0 \end{bmatrix} \\ \begin{bmatrix} U^T & 0 \end{bmatrix} & 0 \end{bmatrix} \begin{bmatrix} S^T \mathbf{dx}_c \\ \mathbf{d}\boldsymbol{\lambda}_c \end{bmatrix} = \begin{bmatrix} -S^T \nabla_{\mathbf{x}} \mathcal{L}^T \\ -\mathbf{c} \end{bmatrix}.$$

The solution of this system can be found by solving in sequence: go ID2P

$$U^T \left(S_1^T \mathbf{dx}_c \right) = -\mathbf{c},$$

$$S_2^T \nabla_{\mathbf{x}}^2 \mathcal{L} S_2 \left(S_2^T \mathbf{dx}_c \right) = -S_2^T \left(\nabla_{\mathbf{x}}^2 \mathcal{L} S_1 \left(S_1^T \mathbf{dx}_c \right) + \nabla_{\mathbf{x}} \mathcal{L}^T \right),$$

$$U \mathbf{d}\boldsymbol{\lambda}_c = -S_1^T \left(\nabla_{\mathbf{x}}^2 \mathcal{L} \mathbf{dx}_c + \nabla_{\mathbf{x}} \mathcal{L}^T \right).$$

Mattheij NSM example

Figure go MEER shows state variable and multiplier plots for a Newton's method implementation of the null space approach. These results complement the embedding results presented in Example go EMMP. The data for the estimation problem is based on the observation functional representer $H = \begin{bmatrix} .5 & 0 & .5 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \end{bmatrix}$ with the true signal values being perturbed by random normal values having standard deviation $\sigma = .5$. The number of observations generated is $n = 501$. The initial values of the state variables are perturbed from their true values by up to 10%, and the initial multipliers are set to 0. The initial parameter values correspond to the true values $10, \; 2$ perturbed also by up to 10%. Very rapid convergence (4 iterations) is obtained.

Mattheij NSM results

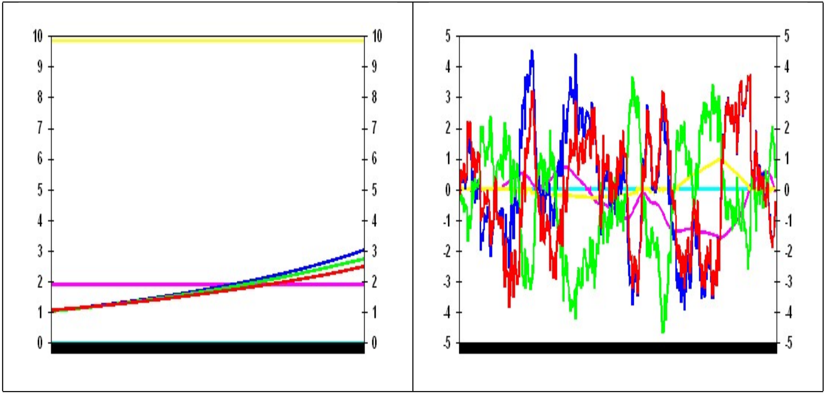


Figure: State variables $\mathbf{x}_c$ and multipliers $n\lambda_c$ for Mattheij Problem

go NSMM

A scoring related algorithm

The Newton iteration works with the augmented matrix appropriate to the problem. This is necessarily indefinite even if $\nabla_x^2 \mathcal{L}$ is positive definite. It follows that not all advantages of the Gauss-Newton iteration extend. However, the second derivative terms arising from the constraints are $O\left(1/n\right)$ through the factor h. Thus their contribution is smaller than that of the terms arising from the objective function when the $O\left(1/n^{1/2}\right)$ scale appropriate for the Lagrange multipliers is taken into account. Also, it is required that the initial Hessian (augmented) matrix be nonsingular if $\lambda_c = 0$ is an acceptable initial estimate. This suggests that ignoring the strict second derivative contribution from the constraints should lead to an iteration with asymptotic convergence properties similar to Gauss-Newton. This behaviour has been observed by Bock (first-1983) and others.

Sketch of justification

This time it is not sufficient to show that the elements of Q', the fixed point iteration variational matrix, are $O\left(n^{-1/2}\right)$. This is true, but $Q' \in R^{2nm-m} \to R^{2nm-m}$. Structure is everything!

go NSME Here $W = \begin{bmatrix} S^T & 0 \\ 0 & I \end{bmatrix} Q' \begin{bmatrix} S & 0 \\ 0 & I \end{bmatrix}$ has the form

$$W = \begin{bmatrix} X & X & X \\ X & X & 0 \\ X & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} X & X & 0 \\ X & X & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ X & Z & 0 \\ X & X & 0 \end{bmatrix},$$

$$Z = \left\{ \frac{1}{n} S_2^T \operatorname{diag}\{H^T H\} S_2 \right\}^{-1} \left\{ h S_2^T \nabla_x^2 \left(\mathbf{s}\left(\boldsymbol{\lambda}_c\right)^T \mathbf{f}_c \right) S_2 \right\} \in R^{m \times m}.$$

The key result is:

$$\varpi\left\{ Q'\left(\begin{bmatrix} \widehat{\mathbf{x}}_n \\ \widehat{\boldsymbol{\lambda}}_n \end{bmatrix} \right) \right\} = \varpi\left\{ Z\left(\begin{bmatrix} \widehat{\mathbf{x}}_n \\ \widehat{\boldsymbol{\lambda}}_n \end{bmatrix} \right) \right\} \overset{a.s.}{\to} 0, \; n \to \infty.$$

Loose ends

- The embedding and simultaneous algorithms are equivalent. Readily proved modulo some reasonable assumptions by assuming the contrary and deriving a contradiction.
- Consistency for the estimation problem follows most easily from the embedding algorithm. Set $[B_1 \; B_0] = \lim_{n\to\infty} S_2\left(\mathbf{x}^*\right)^T$ and treat result as an explicit parameter estimation problem.
- Simultaneous method avoids explicit ODE solution steps. How can adaptive meshing be introduced?

Stochastic ODE

Consider the linear stochastic differential equation

$$d\mathbf{x} = M\mathbf{x}dt + \sigma\mathbf{b}dz$$

where z is a unit Wiener process. Variation of parameters gives the discrete dynamics equation

$$\mathbf{x}_{i+1} = X(t_{i+1}, t_i)\,\mathbf{x}_i + \sigma\mathbf{u}_i,$$

where

$$\mathbf{u}_i = \int_{t_i}^{t_{i+1}} X(t_{i+1}, s)\,\mathbf{b}\frac{dz}{ds}ds.$$

From this it follows that

$$\mathbf{u}_i \backsim N\left(0, \sigma^2 R(t_{i+1}, t_i)\right),$$

where go SDES

$$R(t_{i+1}, t_i) = \int_{t_i}^{t_{i+1}} X(t_{i+1}, s)\,\mathbf{b}\mathbf{b}^T X(t_{i+1}, s)^T\,ds = O\left(\frac{1}{n}\right).$$