
10

The Bock iteration for the ODE estimation problem

M.R. Osborne

Mathematical Sciences Institute
Australian National University

MOSOCOP 08
Heidelberg, July 21–25

Outline

Gauss-Newton algorithm

Seek

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in R^p} \|\mathbf{f}(\mathbf{x})\|_2^2, \ \mathbf{f} \in R^n, \ n > p,$$

by solving the sequence of problems

$$\mathbf{h}_i = \arg\min_{\mathbf{h}\in R^P} \|\mathbf{f}(\mathbf{x}_i) + \nabla_x \mathbf{f}(\mathbf{x}_i)\,\mathbf{h}\|_2^2,$$
$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{h}_i.$$

This is a fixed point iteration of the form $\mathbf{x}_{i+1} = \mathbf{F}_n(\mathbf{x}_i)$ and $\widehat{\mathbf{x}}$ is a point of attraction provided ($\varpi(\cdot)$ denotes spectral radius)

$$\varpi\left(\mathbf{F}_n'\left(\widehat{\mathbf{x}}\right)\right) = \varpi\left(\left(\nabla_x \mathbf{f}^T \nabla_x \mathbf{f}\right)^{-1}\left(\sum_{i=1}^n f_i \nabla_x^2 f_i\right)\right) < 1.$$

Second order convergence if $\mathbf{f}\left(\widehat{\mathbf{x}}\right) = 0$. Fast convergence in large sample, data analytic problems.

Generalised Gauss-Newton algorithm

Basic iteration is formally similar:

$$\mathbf{h}_i = \arg\min_{\mathbf{h}\in R^P} \left\| \mathbf{f}(\mathbf{x}_i) + \nabla_x \mathbf{f}(\mathbf{x}_i)\,\mathbf{h} \right\|_s ,$$
$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{h}_i .$$

where we work with norms $\|\cdot\|_s$ on R^n and $\|\cdot\|_t$ on R^p. Important application to l_1 and max norms. Nicest results have to do with cases where second order convergence is possible. Local strong uniqueness is an elegant sufficient condition (Ludwig Cromme, Num. Math. 29, 179-94, 1978).

$$\exists\, \gamma > 0 \;\Rightarrow \left\| \mathbf{f}\left(\widehat{\mathbf{x}} + \mathbf{v}\right)\right\|_s \geq \left\| \mathbf{f}\left(\widehat{\mathbf{x}}\right)\right\|_s + \gamma \left\|\mathbf{v}\right\|_t ,\; \forall\; \left\|\mathbf{v}\right\|_t \text{ small enough.}$$

Implies $\|\mathbf{f}\|_s$ not smooth at $\widehat{\mathbf{x}}$.

Constrained problems

$$\min_{\mathbf{x}} \Phi(\mathbf{x}), \ \mathbf{c}(\mathbf{x}) = 0, \ \mathbf{c} \in R^m, \ m < p.$$

Introduce *Lagrangian*: $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \Phi(\mathbf{x}) + \Sigma_{i=1}^{m} \lambda_i c_i(\mathbf{x})$.
Necessary conditions are:

$$\nabla_x \mathcal{L} = 0, \ \nabla_\lambda \mathcal{L} = \mathbf{c}^T = 0.$$

Newton correction $(\triangle \mathbf{x}, \ \triangle \boldsymbol{\lambda})$ is given by

$$\begin{aligned} \nabla_{xx}^2 \mathcal{L} \triangle \mathbf{x} + \nabla_x \mathbf{c}^T \triangle \boldsymbol{\lambda} &= -\nabla_x \mathcal{L}^T, \\ \nabla_x \mathbf{c} \triangle \mathbf{x} &= -\mathbf{c}. \end{aligned}$$

Let

$$C^T = \nabla_x \mathbf{c}^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} U \\ 0 \end{bmatrix}.$$

Then a sufficient condition for local convergence at $\widehat{\mathbf{x}}, \widehat{\boldsymbol{\lambda}}$ is

$$U \text{ full rank}, \ Q_2^T \nabla_{xx}^2 \mathcal{L} Q_2 \succ 0$$

Introducing the Bock iteration

In the case of interest the values of both p and m are determined by discretization of an ODE system. Thus they are potentially large. Two consequences are immediate:

1. Sparsity needs to be respected.
2. Calculation of $\sum_{i=1}^{m} \lambda_i \nabla_{xx}^2 c_i$ is potentially a pain.

The iteration considered here sets $\nabla_{xx}^2 c_i, \; i = 1, 2 \cdots, m \rightarrow 0$. The key to success in cases of non trivial constraint curvature is the size of $\boldsymbol{\lambda}$.

Iterations formulated

Set $\nabla^2_{xx}\Phi = A$, $\sum_{i=1}^{m} \lambda_i \nabla^2_{xx} c_i = B$, $\nabla_x \mathbf{c} = C$. Then basic iteration calculations solve

$$\text{Newton} \quad \begin{bmatrix} A+B & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \triangle \mathbf{x} \\ \triangle \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L} \\ -\mathbf{c} \end{bmatrix}.$$

$$\text{Bock} \quad \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \triangle \mathbf{x} \\ \triangle \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L} \\ -\mathbf{c} \end{bmatrix}.$$

Write Bock as a fixed point iteration

$$\begin{bmatrix} \mathbf{x}_{i+1} \\ \boldsymbol{\lambda}_{i+1} \end{bmatrix} = \mathbf{F}\left(\begin{bmatrix} \mathbf{x}_i \\ \boldsymbol{\lambda}_i \end{bmatrix} \right),$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_x \mathcal{L} \\ \mathbf{c} \end{bmatrix}.$$

Condition for an attractive fixed point is $\varpi\left(\mathbf{F}'\right) < 1$.

Structure of $\mathbf{F}'$ in Bock iteration

At $\begin{bmatrix} \widehat{\mathbf{x}} \\ \widehat{\lambda} \end{bmatrix}$ necessary conditions give

$$\mathbf{F}' = I - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} A+B & C^T \\ C & 0 \end{bmatrix},$$

$$= - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}.$$

Orthogonal similarity using Q is helpful. Let

$$Q^TAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \; Q^TBQ = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Then

$$\begin{bmatrix} Q^T & 0 \\ 0 & I \end{bmatrix} \mathbf{F}' \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & U \\ A_{21} & A_{22} & 0 \\ U^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} B_{11} & B_{12} & 0 \\ B_{21} & B_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

More algebra

Let $G = \begin{bmatrix} A & B & C \\ D & E & 0 \\ F & 0 & 0 \end{bmatrix}$, then

$$G^{-1} = \begin{bmatrix} 0 & 0 & F^{-1} \\ 0 & E^{-1} & -E^{-1}DF^{-1} \\ C^{-1} & -C^{-1}BE^{-1} & C^{-1}BE^{-1}DF^{-1} - AF^{-1} \end{bmatrix}.$$

Let $W = \begin{bmatrix} R & S & 0 \\ T & Z & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then

$$G^{-1}W = \begin{bmatrix} 0 & 0 & 0 \\ E^{-1}T & E^{-1}Z & 0 \\ C^{-1}R - C^{-1}BE^{-1}T & C^{-1}S - C^{-1}BE^{-1}Z & 0 \end{bmatrix}.$$

Key result

$$\varpi\left(G^{-1}W\right)=\varpi\left(E^{-1}Z\right).$$

Here p is the number of variables, m is the number of constraints. In the ODE application these $\uparrow\infty$ as the discretization is refined, but $p-m$ is fixed and finite (equal to the order of the ODE system). The significance is:

$$p-m=\dim E^{-1}Z.$$

Thus the rate of convergence question can be reduced to the question of estimating the eigenvalues of a matrix of fixed and finite dimension. go BCP

Smoothing problem data

Estimate state variables $\mathbf{x}(t_i)$ given differential equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}), \ \mathbf{x}, \mathbf{f} \in R^m,$$

plus observations in presence of independent, $N(0, \sigma^2 I)$ errors

$$\mathbf{y}_i = \mathcal{O}\mathbf{x}^*(t_i) + \varepsilon_i, \ i = 1, 2, \cdots, n,$$

$$\mathbf{y}_i \in R^k, \ \mathcal{O} \in R^m \rightarrow R^k, \ t_i \in [0, 1].$$

Trapezoidal rule discretization gives constraints

$$\mathbf{c}_i = \mathbf{x}_{i+1} - \mathbf{x}_i - \frac{\triangle t}{2}(\mathbf{f}(t_i, \mathbf{x}_i) + \mathbf{f}(t_{i+1}, \mathbf{x}_{i+1})) = 0.$$

Note

1. dimension change $p = n * m, \ m \leftarrow (n-1)m$.
2. introduce $\mathbf{x}_c$ composite vector with block components $\mathbf{x}_i, \ i = 1, 2, \cdots, n$.
3. sparsity $\mathbf{c}_i(\mathbf{x}_c) = \mathbf{c}_{ii}(\mathbf{x}_i) + \mathbf{c}_{i(i+1)}(\mathbf{x}_{i+1})$.

Parameter estimation as smoothing

The problem of estimating auxiliary parameters in the differential equation can be reduced to smoothing problem form by introducing additional state variables and augmenting the differential equation. Not necessarily the way to compute the solution.

$$\begin{bmatrix} \mathbf{x} \\ \beta \end{bmatrix} \rightarrow \mathbf{x},$$

$$\begin{bmatrix} \mathbf{f}(t, \mathbf{x}, \beta) \\ 0 \end{bmatrix} \rightarrow \mathbf{f}(t, \mathbf{x}).$$

Consider only the smoothing form here.

Smoothing problem formulation

Set $\mathbf{r}_i = \mathbf{y}_i - \mathcal{O}\mathbf{x}_i, \ \mathbf{r}_i \in R^k$, and let

$$\Phi_n\left(\mathbf{x}_c\right) = \frac{1}{2n}\sum_{i=1}^{n}\|\mathbf{r}_i\|_2^2 .$$

Smoothing problem in simultaneous form

$$\min_{\mathbf{x}_c} \Phi_n\left(\mathbf{x}_c\right); \ \mathbf{c}\left(\mathbf{x}_c\right) = 0.$$

Lagrangian is

$$\mathcal{L}_n = \Phi_n\left(\mathbf{x}_c\right) + \sum_{i=1}^{n-1} \boldsymbol{\lambda}_i^T \mathbf{c}_i,$$

$$= \Phi_n\left(\mathbf{x}_c\right) + \boldsymbol{\lambda}_1^T \mathbf{c}_{11} + \sum_{i=2}^{n-1}\left\{\boldsymbol{\lambda}_{i-1}^T \mathbf{c}_{(i-1)i} + \boldsymbol{\lambda}_i^T \mathbf{c}_{ii}\right\} + \boldsymbol{\lambda}_{n-1}^T \mathbf{c}_{(n-1)n}.$$

Lagrangian derivatives

Set $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_n = 0$ and define $\mathbf{s}\left(\boldsymbol{\lambda}_c\right)_i = \boldsymbol{\lambda}_{i-1} + \boldsymbol{\lambda}_i$ then

$$\nabla^2_{xx}\mathcal{L} = \operatorname{diag}\left\{\frac{1}{n}\mathcal{O}^T\mathcal{O} + \frac{\triangle t}{2}\nabla^2_{xx}\left(\mathbf{s}_i^T\mathbf{f}\left(t_i, \mathbf{x}_i\right)\right),\ i = 1, 2, \cdots, n\right\},$$

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

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

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

Here $\nabla^2_{xx}\mathcal{L}$ is $m \times m$ block diagonal, and C is $m \times m$ block (upper) bidiagonal.

Solving the necessary conditions

The gradient of the Lagrangian gives the equations

$$
\begin{aligned}
-\frac{1}{n}\mathbf{r}_1^T\mathcal{O}+\boldsymbol{\lambda}_1^T\nabla_x\mathbf{c}_{11}&=0,\\
-\frac{1}{n}\mathbf{r}_i^T\mathcal{O}+\boldsymbol{\lambda}_{i-1}^T\nabla_x\mathbf{c}_{(i-1)i}+\boldsymbol{\lambda}_i^T\nabla_x\mathbf{c}_{ii}&=0,\quad i=2,3,\cdots,n-1,\\
-\frac{1}{n}\mathbf{r}_n^T\mathcal{O}+\boldsymbol{\lambda}_{n-1}^T\nabla_x\mathbf{c}_{(n-1)n}&=0,.
\end{aligned}
$$

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:

$$
\begin{aligned}
\nabla_{xx}^2\mathcal{L}\mathbf{dx}_c+\nabla_{x\lambda}^2\mathcal{L}\mathbf{d}\boldsymbol{\lambda}_c&=-\nabla_x\mathcal{L}^T,\\
\nabla_x\mathbf{c}\left(\mathbf{x}_c\right)\mathbf{dx}_c=C\mathbf{dx}_c&=-\mathbf{c}\left(\mathbf{x}_c\right),
\end{aligned}
$$

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

Grouping terms in the necessary conditions gives

$$-\boldsymbol{\lambda}_{i-1}+\boldsymbol{\lambda}_i+\frac{\triangle t}{2}\nabla_x\mathbf{f}\left(t_i,\mathbf{x}_i\right)^T\left(\boldsymbol{\lambda}_{i-1}+\boldsymbol{\lambda}_i\right)=-\frac{1}{n}\mathcal{O}^T\mathbf{r}_i.$$

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

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

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

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

Multiplier estimate for normal errors

This equation is important!

$$-\mathbf{w}_{i-1}+\mathbf{w}_i+\frac{\triangle t}{2}\nabla_x \mathbf{f}\left(t_i,\mathbf{x}_i\right)^T\left(\mathbf{w}_{i-1}+\mathbf{w}_i\right)=-\frac{r_i}{\sqrt{n}}\mathbf{o}. \qquad (1)$$

The variance of the stochastic forcing term in this rescaled form is $\left(\sigma^2/n\right)\mathbf{o}\mathbf{o}^T$, 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 $\boldsymbol{\lambda}_i \rightarrow 0,\ i=1,2,\cdots,n-1$, on a scale which is $O\left(n^{-1/2}\right)$ as $n\rightarrow\infty$ when the errors are normally distributed.

Example of multiplier behaviour

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

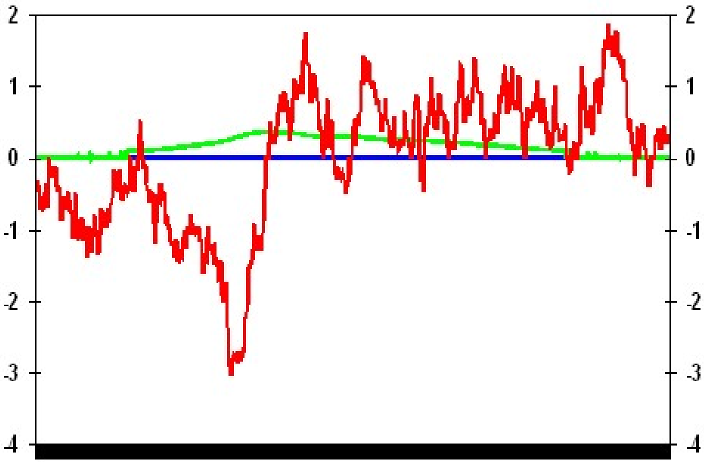
$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x},$$

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

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

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



go P

Bob Mattheij's example

Consider the differential system defined by $\mathbf{f}(t,\mathbf{x}) = A(t,\mathbf{x})\,\mathbf{x} + \mathbf{q}(t)$ with

$$A(t) = \begin{bmatrix} 1 - x_4 \cos x_5 t & 0 & 1 + x_4 \sin x_5 t & 0 & 0 \\ 0 & x_4 & 0 & 0 & 0 \\ -1 + x_4 \sin x_5 t & 0 & 1 + x_4 \cos x_5 t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{q}(t) = \begin{bmatrix} e^t\left(-1 + 10\left(\cos 2t - \sin 2t\right)\right) \\ -9e^t \\ e^t\left(1 - 10\left(\cos 2t + \sin 2t\right)\right) \\ 0 \\ 0 \end{bmatrix}.$$

Here the right hand side is chosen so that $\mathbf{x}(t)^T = \begin{bmatrix} e^t & e^t & e^t & 10 & 2 \end{bmatrix}$ satisfies the differential equation.

Mattheij example

Following Figure shows state variable and multiplier plots for a Newton's method implementation. The data for the estimation problem is based on the observation functional representer $\mathcal{O} = \begin{bmatrix} .5 & 0 & .5 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \end{bmatrix}$ 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

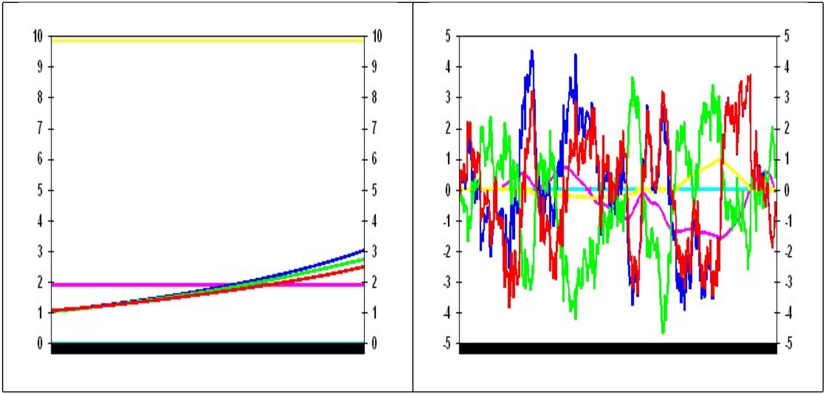


Figure: State variables $\mathbf{x}_c$ and multipliers $n\mathbf{w}_c$ for Mattheij Problem

The approximate 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 and ties simultaneous methods to the class of SQP algorithms. This means more complicated behaviour when compared to minimising sums of squares. However, the second derivative terms arising from the constraints are $O\left(1/n\right)$ through the factor $\triangle t$. 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. This suggests that ignoring the strict second derivative contribution from the constraints should lead to an iteration with asymptotic convergence rate similar to Gauss-Newton. This behaviour has been observed by Bock (first-1983) and others.

Convergence rate analysis

Theorem

Assume

$$E = \left[Q_2^T \operatorname{diag}\left\{ \mathcal{O}^T \mathcal{O},\ i = 1, 2, \cdots, n \right\} Q_2 \right]$$

has a bounded inverse for n large. Then

$$\varpi \left\{ F_n' \left(\left[\begin{array}{c} \widehat{\mathbf{x}}_c \\ \widehat{\boldsymbol{\lambda}}_c \end{array} \right] \right) \right\} \underset{n \to \infty}{\overset{a.s.}{\longrightarrow}} 0.$$

Here $\widehat{\mathbf{x}}_c,\ \widehat{\boldsymbol{\lambda}}_c$ indicate optimal values for the current n.
Note: The condition on E is an identifiability condition on the estimation problem. If $\mathcal{O} = I$ then $E = I$.

Outline of proof

go SBI Let

$$B = \operatorname{diag}\left\{ \nabla^2_{xx}\mathbf{s}\left(\widehat{\boldsymbol{\lambda}}_c\right)_i^T \mathbf{f}\left(t_i, \widehat{\mathbf{x}}_i\right),\ i = 1, 2, \cdots, n \right\}.$$

The critical quantity is

$$\varpi\left\{ \left(Q_2^T \operatorname{diag}\left\{ \frac{1}{n}\mathcal{O}^T\mathcal{O},\ i = 1, \cdots, n \right\} Q_2 \right)^{-1} Q_2^T \frac{\triangle t}{2} B Q_2 \right\}$$

As $n\triangle t = O(1)$ it follows that it is sufficient to show that

$$\left\| Q_2^T B Q_2 \right\| \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0.$$

Here the spectral norm is dominated by the spectral radius of the symmetric, $m \times m$ block, block diagonal matrix B. The desired result now follows because the diagonal blocks of this matrix all tend to 0 with $\widehat{\boldsymbol{\lambda}}_c,\ n \to \infty$.

Properties

- The figures suggest that there is scope for cancellation in summations involving the computed multipliers. However, there appears to be little scope for exploiting this apparently random behaviour as sums in B are over fixed panels of length m. go M1
- The assumption that the observational errors are normal is required here explicitly. This is in contrast to Gauss-Newton where independence (+ bounded variance) is the key to fast convergence in large samples. go Stoch

Performance

Computations with the Bock iteration on the Mattheij example make for an interesting comparison with the Newton results. It proves distinctly less satisfactory without a line search in case $\sigma = .5$, and failure to converge was noted for a high percentage of seed values for the random number generator. However, when $\sigma = .1$ the behaviour of the two iterations is essentially identical.

Stochastic ODE

go SDES go P 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

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

where

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

From this it follows that

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

where

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