The Bock iteration for the ODE estimation problem

M.R.Osborne *

Contents

Abstract

The Bock iteration [2] is a method for the minimization of an objective function in the form of a sum of squares subject to a set of equality constraints. It was formulated originally as a method for solving the problem of estimating parameters in a system of ordinary differential equations (ODE) posed in its simultaneous form, and this remains its principal application. The simultaneous method requires that the discretized ODE be imposed as equality constraints on the objective function of the data fitting problem. However, not only is the data fitting problem explicitly constrained, but the constraint system possesses the key additional feature that it becomes unbounded in size as the scale of the discretization $\Delta t \to 0$, $n \to \infty$, where n is the number of independent observations in the estimation problem. Constraints introduce Lagrange multipliers into the necessary conditions for the simultaneous method, and it is required to estimate these in

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

This paper is dedicated to Hans Georg Bock on the occasion of his 60'th birthday.

order to carry out an asymptotic convergence rate analysis of the Bock iteration for large n. It is shown here that the Lagrange multipliers are $O\left(\sigma n^{-1/2}\right),\ n \to \infty$, where σ is the standard deviation of the measurement errors in the observations, and that a similar estimate is valid for the multiplier characterizing the first order convergence rate of the iteration. This shows that the Bock iteration has excellent convergence rate properties in large sample problems. An interesting feature of these estimates is that their derivation requires that the observational errors are normally distributed. The Lagrange multiplier estimates are obtained by interpreting the necessary conditions for an optimum of the estimation problem as a discretization of a stochastic ODE system. The convergence rate analysis must take into account the unbounded size of the constraint system. This is done by reducing the calculation of the multiplier characterizing the first order convergence rate of the Bock iteration to that of estimating the spectral radius of a matrix of fixed dimension independent of n.

1 Introduction

The first point to make is that the Bock iteration differs significantly from the Gauss-Newton method which proves ubiquitous in minimizing a sum of squares objective function in unconstrained, nonlinear estimation problems. This is despite the fact that an argument based on analogy with Gauss-Newton proved important in the original derivation of the method. For this reason a brief summary of Gauss-Newton and generalized Newton methods is given.

The basic Gauss-Newton algorithm [5] seeks

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in R^p} \left\|\boldsymbol{\phi}\left(\mathbf{x}\right)\right\|_2^2,\ \boldsymbol{\phi} \in R^n,\ n > p, \tag{1}$$

by solving the sequence of problems

$$\begin{aligned} \mathbf{h}_i &= \arg\min_{\mathbf{h}\in R^p} \left\|\boldsymbol{\phi}\left(\mathbf{x}_i\right) + \nabla_x \boldsymbol{\phi}\left(\mathbf{x}_i\right)\mathbf{h}\right\|_2^2, \\ \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{h}_i. \end{aligned} \tag{2}$$

This is a fixed point iteration of the form

$$\mathbf{x}_{i+1} = \mathbf{F}_n\left(\mathbf{x}_i\right), \tag{3}$$

and $\widehat{\mathbf{x}}$ is a point of attraction provided

$$\varpi\left(\mathbf{F}_n'\left(\widehat{\mathbf{x}}\right)\right) = \varpi\left(\left(\nabla_x \boldsymbol{\phi}^T \nabla_x \boldsymbol{\phi}\right)^{-1}\left(\sum_{i=1}^{n} \phi_i \nabla_{xx}^2 \phi_i\right)\right) < 1. \tag{4}$$

where $\varpi(\cdot)$ denotes the spectral radius of the indicated matrix. If convergence is obtained then the iteration rate is generically first order with multiplier determined by this spectral radius. Second order convergence requires that $\varpi\left(\mathbf{F}_n'\left(\widehat{\mathbf{x}}\right)\right) = 0$. This occurs if $\boldsymbol{\phi}\left(\widehat{\mathbf{x}}\right) = 0$. However, small residuals are not necessary for fast convergence in the Gauss-Newton iteration. It is known that fast convergence can occur in large sample, data analytic problems provided the measurement errors are independent and have bounded variance [7]. The development of similar large sample convergence results for the Bock iteration applied to the simultaneous method is the main purpose of this paper. The Bock results turn out to be somewhat more restrictive in that they require the measurement errors to be not only independent but also normally distributed.

The generalized Gauss-Newton iteration extends the basic Gauss-Newton scheme to a wider class of norms than the Euclidean norm $\|\cdot\|_2$. The underlying idea is that of minimizing the norm of a local linearization to provide an iterative correction. This generates the iterative procedure:

$$\begin{aligned} \mathbf{h}_i &= \arg\min_{\mathbf{h}\in R^P} \left\|\boldsymbol{\phi}\left(\mathbf{x}_i\right) + \nabla_x \boldsymbol{\phi}\left(\mathbf{x}_i\right)\mathbf{h}\right\|_s, \\ \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{h}_i. \end{aligned} \tag{5}$$

Here the norms are $\|\cdot\|_s$ on R^n and $\|\cdot\|_t$ on R^p. This iteration has important application when the norms in R^n are the l_1 and max norms, and the nicest results have to do with cases where second order convergence is possible. Local strong uniqueness provides an elegant sufficient condition [3]

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

Note that it implies that $\|\boldsymbol{\phi}\|_s$ is not smooth at $\widehat{\mathbf{x}}$. Necessary conditions are considered in [4].

Consider now the constrained problem.

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

The *Lagrangian* for this problem is

$$\mathcal{L}\left(\mathbf{x}, \boldsymbol{\lambda}\right) = \Phi(\mathbf{x}) + \Sigma_{i=1}^{q} \lambda_i c_i\left(\mathbf{x}\right),$$

where the $\lambda_i,\; i = 1, 2, \cdots, q$ are the Lagrange multipliers. The corresponding necessary conditions for an optimum are

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

These can be solved by Newton's method which gives corrections $\Delta\mathbf{x}$, $\Delta\boldsymbol{\lambda}$ determined by the linear system

$$\nabla^2_{xx}\mathcal{L}\Delta\mathbf{x} + \nabla_x\mathbf{c}^T\Delta\boldsymbol{\lambda} = -\nabla_x\mathcal{L}^T, \tag{6}$$

$$\nabla_x\mathbf{c}\Delta\mathbf{x} = -\mathbf{c}. \tag{7}$$

It proves convenient to make an orthogonal Q times upper triangular U factorization of $C^T = \nabla_x\mathbf{c}^T$.

$$C^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} U \\ 0 \end{bmatrix}. \tag{8}$$

Here $Q \in R^p \to R^p$, $Q_1 \in R^q \to R^p$, and $U \in R^q \to R^q$. Then a sufficient condition for local convergence at $\widehat{\mathbf{x}}$, $\widehat{\boldsymbol{\lambda}}$ is

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

These conditions are a form of second order sufficiency. They ensure that the Jacobian of the Newton iteration is nonsingular at the stationary point.

The plan of the paper is as follows. The Bock iteration in a conventional mathematical programming setting is introduced in the next section. The principal result obtained relates the first order convergence rate multiplier to the spectral radius of a $(p-q)\times(p-q)$ matrix. The ODE estimation problem is introduced as a smoothing problem in Section 3. Here the unknowns are the state variable estimates at the points at which the data is provided. The exact constraint is the condition that the state variables satisfy the ODE system. This system is approximated by the finite system which is obtained by using the trapezoidal rule to discretize the ODE. The advantages of this discretization in the estimation context, including advantages in sparsity in formulating the Lagrangian, are summarised. The necessary conditions lead to a recurrence for the Lagrange multipliers, and this, when suitably scaled, is identified as a discretization of a stochastic differential equation with almost sure $O(1)$ solution. This permits the asymptotic scale of the Lagrange multipliers to be deduced. This scale is checked by two numerical computations. The first makes use of an ODE which is integrated exactly by the trapezoidal rule. The second makes use of the Mattheij differential equation [1] to provide a more serious test. These results are pulled together in the final section to show that the convergence rate multiplier for the Bock iteration tends to zero in large samples in much the same way as the corresponding multiplier for the Gauss-Newton method provided the more restrictive assumption of normally distributed measurement errors is made. This result underlines the effectiveness of the Bock iteration in large sample data analysis problems.

2 Introducing the Bock iteration

The Bock iteration modifies the basic Newton iteration by setting $\nabla^2_{xx} c_i \to 0,\ i = 1, 2 \cdots, q$ in the linear system (6), (7). This is correct if the constraints are linear functions of $\mathbf{x}$. It is analogous to deriving (2) by ignoring second derivatives in the full Newton formulation of the nonlinear least squares problem as these derivatives vanish in the Newton formulation if the $\phi_i(\mathbf{x})$ are linear. However, what corresponds to small residuals in the Gauss-Newton method are small values of $\boldsymbol{\lambda}$, the Lagrange multipliers, in the Bock iteration. This modification is very attractive in the simultaneous formulation of the ODE estimation problem because here the values of both p, the number of variables, and q, the number of constraints, are determined by Δt^{-1} where Δt is the grid scale in the discretization of the ODE system. Thus they are potentially large. Two consequences in the ODE application are immediate:

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

To formulate the Bock iteration set

$$\nabla^2_{xx}\Phi = A,\ \sum_{i=1}^{q} \lambda_i \nabla^2_{xx} c_i = B,\ \nabla_x \mathbf{c} = C.$$

Then the basic iteration calculations solve the following systems of equations.

Newton

$$\begin{bmatrix} A+B & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}^T \\ -\mathbf{c} \end{bmatrix}. \tag{9}$$

Bock

$$\begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}^T \\ -\mathbf{c} \end{bmatrix}. \tag{10}$$

The Bock iteration can be written in fixed point form

$$\begin{bmatrix} \mathbf{x}_{i+1} \\ \boldsymbol{\lambda}_{i+1} \end{bmatrix} = \mathbf{F}_n \left(\begin{bmatrix} \mathbf{x}_i \\ \boldsymbol{\lambda}_i \end{bmatrix} \right), \tag{11}$$

where

$$\mathbf{F}_n = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_x \mathcal{L}^T \\ \mathbf{c} \end{bmatrix}. \tag{12}$$

Let $\mathbf{F}'_n = \nabla_{(x,\lambda)} \mathbf{F}_n$. Then the condition for an attractive fixed point is

$$\varpi \left(\mathbf{F}'_n \left(\begin{bmatrix} \widehat{\mathbf{x}} \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix} \right) \right) < 1.$$

The form of $\mathbf{F}'_n$ in the Bock iteration at $\begin{bmatrix} \widehat{\mathbf{x}} \\ \widehat{\boldsymbol{\lambda}} \end{bmatrix}$ can make use of the necessary conditions $\nabla_x \mathcal{L} = 0,\ \mathbf{c} = 0$ to remove matrix derivatives. This gives

$$\begin{aligned} \mathbf{F}'_n &= I - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} A+B & C^T \\ C & 0 \end{bmatrix}, \\ &= - \begin{bmatrix} A & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned} \tag{13}$$

Orthogonal similarity using Q defined in (8) does not change $\varpi(\mathbf{F}'_n)$. This transformation is helpful in simplifying (13). Let

$$Q^T A Q = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad Q^T B Q = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Then

$$\begin{bmatrix} Q^T & 0 \\ 0 & I \end{bmatrix} \mathbf{F}'_n \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & U \\ A_{21} & A_{22} & 0 \\ U^T & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} B_{11} & B_{12} & 0 \\ B_{21} & B_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{14}$$

In these expressions $A_{11}, B_{11} \in R^q \to R^q$ and $A_{22}, B_{22} \in R^{p-q} \to R^{p-q}$.

It is possible to exploit this zero pattern. As a preliminary two elementary algebraic results are quoted.

Lemma 1 *Let*

$$G = \begin{bmatrix} H & J & K \\ L & M & 0 \\ N & 0 & 0 \end{bmatrix},$$

where the backward diagonal matrices K, M, N are all nonsingular. Then G is invertible, and

$$G^{-1} = \begin{bmatrix} 0 & 0 & N^{-1} \\ 0 & M^{-1} & -M^{-1}LN^{-1} \\ K^{-1} & -K^{-1}JM^{-1} & K^{-1}JM^{-1}LN^{-1} - HN^{-1} \end{bmatrix}. \tag{15}$$

Lemma 2 *Let*

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

then

$$G^{-1}W = \begin{bmatrix} 0 & 0 & 0 \\ M^{-1}T & M^{-1}Z & 0 \\ K^{-1}R - K^{-1}JM^{-1}T & K^{-1}S - K^{-1}JM^{-1}Z & 0 \end{bmatrix}. \tag{16}$$

An immediate consequence of Lemma 2 is:

Theorem 3

$$\varpi\left(G^{-1}W\right) = \varpi\left(M^{-1}Z\right). \tag{17}$$

Direct application to the Bock iteration gives the results:

Corollary 4

$$A_{22}^{-1}B_{22} \in R^{p-q} \to R^{p-q}, \tag{18}$$

$$\varpi\left(\mathbf{F}'_n\right) = \varpi\left(A_{22}^{-1}B_{22}\right). \tag{19}$$

Proof. The necessary identifications are made using (14). The result now follows from (17). ■

Remark 5 *This result is given in terms of p, the number of variables, and q, the number of constraints. In the ODE application these* $\uparrow \infty$ *as the discretization is refined, but* $p - q$ *is fixed and finite (equal to the order of the ODE system). Thus the rate of convergence question can be reduced to the question of estimating the eigenvalues of a matrix of fixed and finite dimension.*

3 The ODE estimation problem

The ODE estimation problem sets out to estimate parameters $\boldsymbol{\beta} \in R^s$ in the differential equation model

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}\left(t, \mathbf{x}, \boldsymbol{\beta}\right), \quad \mathbf{x}, \mathbf{f} \in R^m, \tag{20}$$

given observation data

$$\mathbf{y}_i = \mathcal{O}\mathbf{x}^*\left(t_i\right) + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \cdots, n, \tag{21}$$

where the "$*$" superscript is used to denote exact values, and

$$\mathbf{y}_i \in R^k, \quad \mathcal{O} \in R^m \to R^k, \quad t_i = (i-1)\Delta t, \quad \Delta t = 1/(n-1).$$

Equation (21) allows for vectors of data to be observed at each time point t_i. It is not required that different components of the same observation vector are independent. Statements regarding independence are to be interpreted as statements of independence between the results of measurements at distinct time points.

The estimation problem is

$$\min_{x,\beta} \Phi\left(\mathbf{x}\left(t_i\right),\ i=1,2,\cdots,n,\ \boldsymbol{\beta}\right) = \frac{1}{2n}\sum_{i=1}^{n} \left\|\mathbf{r}_i\right\|_2^2, \tag{22}$$

where

$$\mathbf{r}_i = \mathbf{y}_i - \mathcal{O}\mathbf{x}\left(t_i\right).$$

The dependence on $\boldsymbol{\beta}$ comes about because $\mathbf{x}\left(t\right)$ is constrained to satisfy (20). The estimation problem can be reformulated as a smoothing problem in which the parameters $\boldsymbol{\beta}$ are included among the state variables by setting

$$\mathbf{x} \leftarrow \left[\begin{array}{c} \mathbf{x} \\ \boldsymbol{\beta} \end{array}\right], \quad \mathbf{f}\left(t,\mathbf{x}\right) \leftarrow \left[\begin{array}{c} \mathbf{f}\left(t,\mathbf{x},\boldsymbol{\beta}\right) \\ 0 \end{array}\right],$$

in (20). This step makes use of the property $\frac{d\boldsymbol{\beta}}{dt} = 0$. The smoothing formulation has some advantages from the point of view of analysis and for this reason is employed here. It has the disadvantage for implementation that it increases the dimension of the state vector $\mathbf{x}(t)$ and hence the number of unknowns as $m \leftarrow m + s$.

It is necessary to discretize the differential equation constraint explicitly in order to complete the specification of the simultaneous method . Here the trapezoidal rule is used for this purpose. It leads to the constraint system

$$\mathbf{c}_i = \mathbf{x}_{i+1} - \mathbf{x}_i - \frac{\Delta t}{2}\left(\mathbf{f}\left(t_i,\mathbf{x}_i\right) + \mathbf{f}\left(t_{i+1},\mathbf{x}_{i+1}\right)\right) = 0, \quad i=1,2,\cdots,n-1, \tag{23}$$

where $\mathbf{x}_i$ is written for the unknown in the discrete system which provides an estimate of $\mathbf{x}(t_i)$.
This choice has the consequences:

1. The number of constraints in this formulation is $q = m(n-1)$ where m is now the order of the augmented system.

2. It is convenient to introduce $\mathbf{x}_c$ to denote the composite vector of unknowns with block components $\mathbf{x}_i,\ i=1,2,\cdots,n$. Thus the number of unknowns is $p = nm$, and $p - q = m$.

3. The trapezoidal rule has the strong sparsity property

$$\mathbf{c}_i\left(\mathbf{x}_c\right) = \mathbf{c}_{ii}\left(\mathbf{x}_i\right) + \mathbf{c}_{i(i+1)}\left(\mathbf{x}_{i+1}\right). \tag{24}$$

4. The relatively large error estimate for the trapezoidal rule of $O\left(\Delta t^2\right)$ is typically not a problem in this application. Here the influence of measurement errors tends to dominate when $\sigma > 0$. It can be shown that the standard deviation of the error in the solution estimates is $O\left(\sigma\Delta t^{1/2}\right)$. This follows from the consistency of the estimates [6].

In summary, the smoothing problem in simultaneous form is

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

The corresponding Lagrangian is

$$\begin{aligned}\mathcal{L}_n &= \Phi\left(\mathbf{x}_c\right) + \sum_{i=1}^{n-1} \boldsymbol{\lambda}_i^T \mathbf{c}_i, \\ &= \Phi\left(\mathbf{x}_c\right) + \boldsymbol{\lambda}_1^T \mathbf{c}_{11} + \sum_{i=2}^{n-1} \left\{\boldsymbol{\lambda}_{i-1}^T \mathbf{c}_{(i-1)i} + \boldsymbol{\lambda}_i^T \mathbf{c}_{ii}\right\} + \boldsymbol{\lambda}_{n-1}^T \mathbf{c}_{(n-1)n}.\end{aligned}$$

The Lagrangian derivatives are given by

$$\begin{aligned}\nabla_{xx}^2 \mathcal{L} &= \operatorname{diag}\left\{\frac{1}{n}\mathcal{O}^T\mathcal{O} + \mathbf{s}_i^T \frac{\Delta t}{2} \nabla_{xx}^2 \mathbf{f}\left(t_i, \mathbf{x}_i\right),\ i = 1, 2, \cdots, n\right\}, \\ \nabla_{x\lambda}^2 \mathcal{L}_n &= C^T, \\ C_{ii} &= -I - \frac{\Delta t}{2} \nabla_x \mathbf{f}\left(t_i, \mathbf{x}_i\right), \\ C_{i(i+1)} &= I - \frac{\Delta t}{2} \nabla_x \mathbf{f}\left(t_{i+1}, \mathbf{x}_{i+1}\right),\end{aligned}$$

where the notation $\mathbf{s}\left(\boldsymbol{\lambda}_c\right)_i = \boldsymbol{\lambda}_{i-1} + \boldsymbol{\lambda}_i$ is used with end values fixed by setting $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_n = 0$. Here $\nabla_{xx}^2 \mathcal{L}$ is block diagonal with $m \times m$ diagonal blocks, and C is block bi-diagonal also with $m \times m$ blocks.

The necessary conditions give equations that are actually a recurrence satisfied by the Lagrange multipliers.

$$-\frac{1}{n}\mathbf{r}_1^T \mathcal{O} + \boldsymbol{\lambda}_1^T \nabla_x \mathbf{c}_{11} = 0, \tag{25}$$

$$-\frac{1}{n}\mathbf{r}_i^T \mathcal{O} + \boldsymbol{\lambda}_{i-1}^T \nabla_x \mathbf{c}_{(i-1)i} + \boldsymbol{\lambda}_i^T \nabla_x \mathbf{c}_{ii} = 0, \quad i = 2, 3, \cdots, n-1, \tag{26}$$

$$-\frac{1}{n}\mathbf{r}_n^T \mathcal{O} + \boldsymbol{\lambda}_{n-1}^T \nabla_x \mathbf{c}_{(n-1)n} = 0, . \tag{27}$$

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:

$$\nabla_{xx}^2 \mathcal{L} \mathbf{dx}_c + \nabla_{x\lambda}^2 \mathcal{L} \mathbf{d}\boldsymbol{\lambda}_c = -\nabla_x \mathcal{L}^T, \tag{28}$$

$$\nabla_x \mathbf{c}\left(\mathbf{x}_c\right) \mathbf{dx}_c = C \mathbf{dx}_c = -\mathbf{c}\left(\mathbf{x}_c\right), \tag{29}$$

Remark 6 *The equations satisfied by $\widehat{\boldsymbol{\lambda}}_c$ are of particular interest. Equations (25) and (27) essentially provide boundary conditions for the recurrence. These equate the initial and final values of $\widehat{\boldsymbol{\lambda}}_i$ to quantities which are*

$O(1/n)$. *Actually the boundary conditions are over-determined. Thus part provides explicit conditions on (26) which can be recognized as an adjoint of the ODE discretization (23) – see equation (30) below. Optimality is achieved when the complementary set is satisfied at the appropriate solution of (23). Selection of an appropriate set of boundary conditions can be made provided (20) can be posed stably as a two-point boundary value problem. For example, if equation (20) is stable as an initial value problem then trial values of* $\mathbf{x}(0)$ *can be chosen as the vector of initial values, and the values of* $\mathbf{r}_i$ *estimated by a forward computation using equation (23). This permits* $\boldsymbol{\lambda}_{n-1}$ *to be determined from (27) and the adjoint recurrence (26), which can be expected to be stable for the backward computation, solved for the trial multiplier estimates. The necessary conditions are satisfied when the computed* $\boldsymbol{\lambda}_1$*, defined as a function of* $\mathbf{x}_1$ *by the above process, satisfies equation (28).*

Grouping terms in the necessary conditions gives

$$-\widehat{\boldsymbol{\lambda}}_{i-1} + \widehat{\boldsymbol{\lambda}}_i + \frac{\Delta t}{2}\nabla_x \mathbf{f}\left(t_i, \widehat{\mathbf{x}}_i\right)^T \left(\widehat{\boldsymbol{\lambda}}_{i-1} + \widehat{\boldsymbol{\lambda}}_i\right) = -\frac{1}{n}\mathcal{O}^T \mathbf{r}_i. \tag{30}$$

This equation provides the basis for the required estimate of the Lagrange multiplier scale.

Theorem 7 *For simplicity consider the case where* r_i *is a scalar and the observation structure is based on a vector representer* $\mathcal{O} = \mathbf{o}^T$*. Also, assume the* $\boldsymbol{\varepsilon}_i$ *are independent and normally distributed with standard deviation* σ*. Then*

$$\widehat{\boldsymbol{\lambda}}_i = O\left(\sigma n^{-1/2}\right), \quad i = 1, 2, \cdots, n-1. \tag{31}$$

Proof. Consider first the recurrence forcing term in (30) evaluated at $\widehat{\mathbf{x}}_i,\ i = 1, 2, \cdots, n$.

$$\begin{aligned} r_i \mathcal{O}^T &= \left\{\varepsilon_i + \mathbf{o}^T\left(\mathbf{x}_i^* - \widehat{\mathbf{x}}_i\right)\right\}\mathbf{o}, \\ &= \sqrt{n}\left\{\frac{\varepsilon_i}{\sqrt{n}} + \frac{1}{\sqrt{n}}\mathbf{o}^T\left(\mathbf{x}_i^* - \widehat{\mathbf{x}}_i\right)\right\}\mathbf{o}. \end{aligned} \tag{32}$$

The key step is a rescaling to get the right dependence on n in this term. Let $\mathbf{w}_i = \sqrt{n}\widehat{\boldsymbol{\lambda}}_i,\ i = 1, 2, \cdots, n-1$, then

$$-\mathbf{w}_{i-1} + \mathbf{w}_i + \frac{\Delta t}{2}\nabla_x \mathbf{f}\left(t_i, \widehat{\mathbf{x}}_i\right)^T \left(\mathbf{w}_{i-1} + \mathbf{w}_i\right) = -\frac{r_i}{\sqrt{n}}\mathbf{o}. \tag{33}$$

This equation is important! The variance of the stochastic forcing term in this rescaled form is $(\sigma^2/n)\,\mathbf{o}\mathbf{o}^T$, 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 results that follow from the consistency of the estimation procedure [6] are taken into account. These are the correct orders of magnitude needed to produce results that are $O(1)$ almost surely given appropriate initial/boundary conditions (Remark 6 above). This permits identification with a discretization of the adjoint to the linearized constraint differential equation system subject to a forcing term which contains a stochastic component (see the Appendix). It is at this point that it is essential that the measurement errors are normally distributed. The significant feature of this identification is that it indicates that the $\mathbf{w}_i$ are almost surely bounded from which it follows that the multipliers $\widehat{\boldsymbol{\lambda}}_i \to 0,\ i = 1, 2, \cdots, n-1$, on a scale which is $O\left(\sigma n^{-1/2}\right)$ as $n \to \infty$. ■

Example 8 *The effect of the random walk term in equation (33) can be isolated in the smoothing problem with data:*

$$\frac{d\mathbf{x}}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x}, \tag{34}$$

$$y_i = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_i + \varepsilon_i = 1 + \varepsilon_i,\ \varepsilon_i \sim N(0,1), \tag{35}$$

$$t_i = \frac{(i-1)}{(n-1)},\ i = 1, 2, \cdots, n. \tag{36}$$

The trapezoidal rule is exact for equation (34). The scaled solution $\mathbf{w}_i,\ i = 1, 2, \cdots, n-1$ *obtained for a particular realization of the* ε_i *for* $n = 501,\ \sigma = 5$ *is plotted in figure 1. 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.*

Example 9 *Consider the Mattheij differential system [1] defined by*

$$\mathbf{f}(t, \mathbf{x}) = A(t, \mathbf{x})\,\mathbf{x} + \mathbf{q}(t), \tag{37}$$

with

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

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

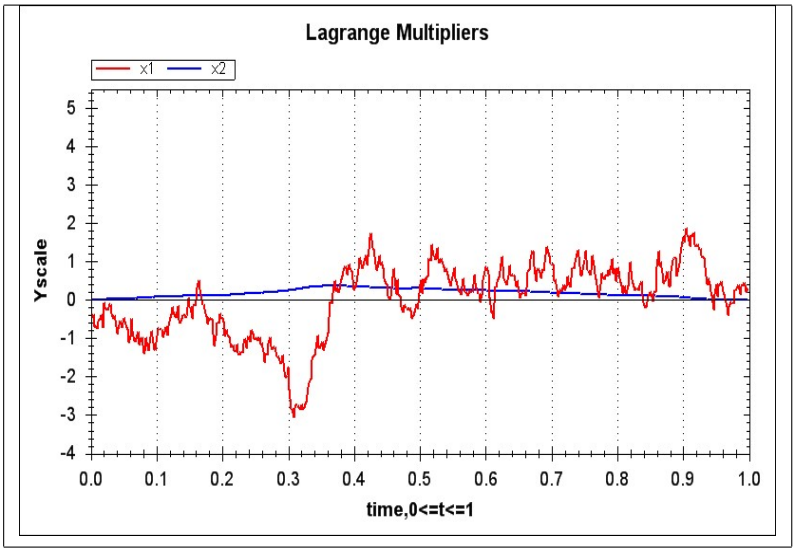

Figure 1: Plot of multipliers $\mathbf{w}$ for test problem

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. Figure 2 shows state variable and multiplier plots for the solution computed by 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 $\sigma = .1$*. 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 (typically 4 iterations) is obtained.*

4 The Bock iteration for the smoothing problem

The Newton iteration works with the Jacobian of the necessary conditions in the Lagrangian formulation. This corresponds to the augmented matrix appropriate to the problem. This is necessarily indefinite even if $\nabla_x^2 \mathcal{L}$ is positive definite. It ties the simultaneous methods to the class of sequential

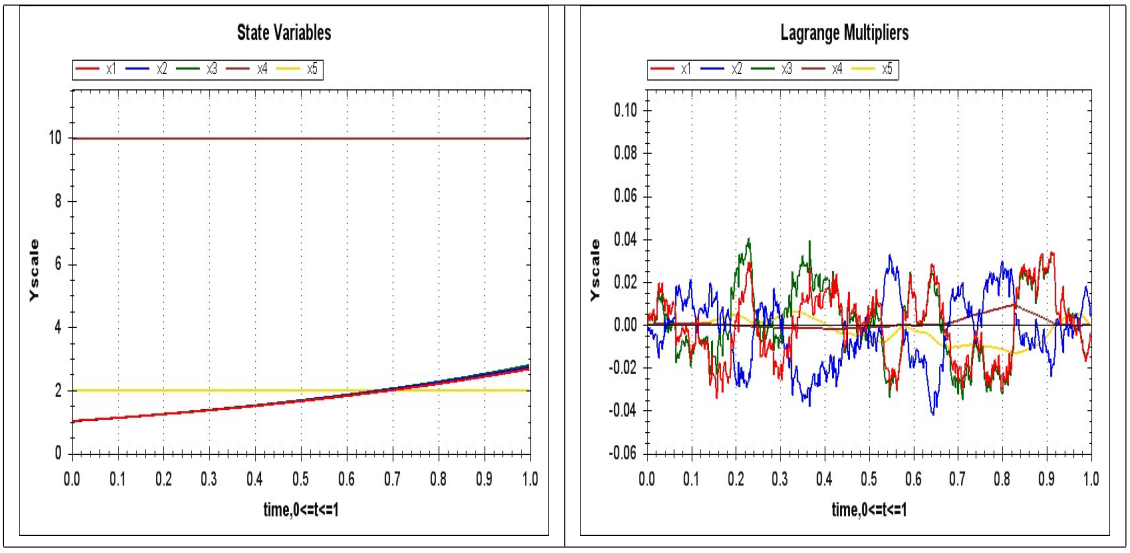

Figure 2: State variables $\mathbf{x}_c$ and multipliers $\mathbf{w}_c$ for Mattheij Problem

quadratic programming algorithms [5]. This means more complicated behaviour than that which occurs when minimizing sums of squares using the Gauss-Newton algorithm. However, the second derivative terms arising from the constraints are $O(1/n)$ in the augmented matrix through the factor Δt. Thus their contribution is relatively smaller than those of the terms arising from the objective function Φ when the $O\left(\sigma/n^{1/2}\right)$ scale appropriate to the Lagrange multipliers is taken into account. This suggests that ignoring the second derivative contributions from the constraints would result in an iteration with asymptotic convergence rate similar to Gauss-Newton where it is known that the first order convergence rate multiplier in large samples has a similar $O\left(\sigma/n^{1/2}\right)$ dependence [7]. This behaviour has been observed by Bock (first-1983) [2] and others. The next result provides a justification.

Theorem 10 *Assume*

$$E = \left[Q_2^T \operatorname{diag}\left\{\mathcal{O}^T\mathcal{O},\ i = 1, 2, \cdots, n\right\} Q_2\right] \tag{38}$$

has a bounded inverse for n large. Then

$$\varpi\left\{F_n'\left(\left[\begin{array}{c}\widehat{\mathbf{x}}_c\\ \widehat{\boldsymbol{\lambda}}_c\end{array}\right]\right)\right\} = O\left(\frac{\sigma}{n^{1/2}}\right) \underset{n\to\infty}{\overset{a.s.}{\to}} 0.$$

Proof. Let

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

By Corollary 4 the critical quantity is

$$\varpi\left\{\left(Q_2^T \operatorname{diag}\left\{\frac{1}{n}\mathcal{O}^T\mathcal{O},\ i=1,\cdots,n\right\}Q_2\right)^{-1} Q_2^T\frac{\triangle t}{2}BQ_2\right\}$$

As $n\Delta t = O(1)$ it follows as a consequence of the Theorem assumption that it is sufficient to show that

$$\left\|Q_2^T BQ_2\right\|_2 \overset{a.s.}{\underset{n\to\infty}{\to}} 0$$

at the appropriate rate. Here the spectral norm is dominated by the spectral radius of the symmetric block diagonal matrix B with $m\times m$ diagonal blocks. The desired result now follows because the elements in the i'th diagonal block of B are linear in the components of $\widehat{\mathbf{s}}_i$ and so are $O\left(\frac{\sigma}{n^{1/2}}\right)$ and tend to 0, $n\to\infty$ for $i=1,2,\cdots,n$. Thus it follows by the Gerschgorin Theorem, that the spectral radius of B tends to zero with $\frac{\sigma}{n^{1/2}}$. But this dominates $\varpi\left\{Q_2^T BQ_2\right\}$. ■

Remark 11 *The condition on E is an identifiability condition on the estimation problem. It expresses the condition that the smoothing problem with $\mathbf{f}=0$ be identifiable. The scale can be checked by noting that if $\mathcal{O}=I$ then $E=I$. This case corresponds to all state variable values being specified at each observation point.*

Computations with the Bock iteration on the Mattheij example (37) make for an interesting comparison with the Newton results. The Bock computations prove distinctly less satisfactory without a line search in case $\sigma=.5$, and failure to converge was noted for a high percentage of seed values tried for the random number generator. However, when $\sigma=.1$ the behaviour of the two iterations is essentially identical.

5 Appendix

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

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

From this it follows that

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

where

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

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] 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.

[3] Ludwig Cromme. Strong uniqueness. a far-reaching criterion for the convergence analysis of iterative procedures. *Numer. Math.*, 29:179–193, 1978.

[4] Krisorn Jittorntrum and M.R. Osborne. Strong uniqueness and second order convergence in nonlinear discrete approximation. *Numer. Math.*, 34:439–455, 1980.

[5] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Verlag, 1999.

[6] M. R. Osborne. On estimation algorithms for ordinary differential equations. In Geoffry N. Mercer and A. J. Roberts, editors, *Proceedings of the 14th Biennial Computational Techniques and Applications Conference, CTAC-2008*, volume 50 of *ANZIAM J.*, pages C107–C120, October 2008. `http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1363` [October 27, 2008].

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