IMPACT PROJECT

A Commonwealth Government inter-agency project in co-operation with the University of Melbourne, to facilitate the analysis of the impact of economic demographic and social changes on the structure of the Australian economy

Paper presented to
Economics Society of Australia and New Zealand
TENTH CONFERENCE OF ECONOMISTS
Australian National University
Canberra
24-28 August, 1981

ELIMINATION OF LINEARIZATION ERRORS WHICH ARISE IN THE SOLUTION OF A GENERAL EQUILIBRIUM MODEL USING JOHANSEN'S APPROXIMATION

by

Russell J. Rimmer

IMPACT Research Centre

Preliminary Working Paper No. OP-33 Melbourne August 1981

The views expressed in this paper do not necessarily reflect the opinions of the participating agencies, nor of the Commonwealth government.

IMPACT PROJECT RESEARCH CENTRE 153 Barry Street, Carlton 3053
Postal Address: Impact Centre, University of Melbourne, Parkville, Vic., 3052, Australia
Phones: (03) 345 1844 extensions 7417 & 7418
After hours (03) 341 7417 or 341 7418.

CONTENTS

Bibliography

1. Dieudonne, J. _Foundations of Mathematical Analysis_, Academic Press, New York, 1960.

2. Dixon, P.B., B.R. Parmenter, J. Sutton and D.P. Vincent. _ORANI, A Multisectoral Model of the Australian Economy_, to be published by North-Holland Publishing Company, Amsterdam.

3. Gray, A. "Differentiation of Composites with Respect to a Parameter", _J. Aust. Math. Soc._, Vol. 19, 1975, pp.121-128.

4. Isaacson, E. and H. Keller. _Analysis of Numerical Methods_, John Wiley and Sons, New York, 1966.

5. Johansen, L. _A Multisectoral study of Economic Growth_, North-Holland Publishing Company, Amsterdam, 1974.

6. Lang, S. _Analysis 1_, Addison-Wesley Publishing Co., Reading, Mass., 1968.

7. Maurin, K. _Analysis, part 1 : Elements_, D. Reidel Publishing Co., Dodrecht, Holland, 1976.

8. Smale, S. "The Fundamental Theorem of Algebra and Complexity Theory", _Bull. Am. Math. Soc._, vol. 4, 1981, pp.1-36.

are thrice so. The argument given by Dieudonne [1, p.263] to prove his result (10.23) now proves that B has three continuous derivatives, and so that g' is thrice continuously differentiable. Therefore, g has four continuous derivatives on U_o.

(c) Proof of (VI) : B <u>satisfies a Lipschitz condition in</u> V_o

B is continuously differentiable on $U_o \times V_o$. Therefore, the mean value theorem (see Lang [6, p.314]) ensures that

$$\|B(x, y^1) - B(x, y^2)\| \leq \sup_{0\leq t\leq 1} \|D_2B(x, y^1 + t[y^2 - y^1])\| \; \|y^1 - y^2\| \qquad (5.2)$$

for each pair of points y^1, y^2 such that $y^1 + t[y^2 - y^1]$ is an element of V_o and where t satisfies $0 \leq t \leq 1$. (V_o can be assumed to satisfy this property for we may take a small convex region within V_o and rename it as V_o, if necessary.) So, if K is chosen as

$$K = \sup_{0\leq t\leq 1} \|D_2B(x, y^1 + t[y^2 - y^1])\|$$

the result follows.

(d) Proof of (VII) :

By reducing $U_o \times V_o$ if required, the bounds for B and its first two derivatives, and for g and its first three derivatives, may be obtained using the continuity of these maps which is guaranteed by (V).

ELIMINATION OF LINEARIZATION ERRORS WHICH ARISE IN THE SOLUTION OF A GENERAL EQUILIBRIUM MODEL USING JOHANSEN'S APPROXIMATION

by

Russell J. Rimmer[1]

IMPACT Research Centre

1. <u>Introduction</u>

The m equations which describe a static model of an economy involving n prices and quantities will be denoted here by a function F which maps a subset W of $\mathbb{R}^n$ into $\mathbb{R}^m$, where W should be thought of as the set of allowable values of prices and quantities. In the case of the Johansen[2] type general equilibrium models, for example, the components of F can be classified into five groups :

(i) Equations which describe household and other final demands for each commodity. Households, for example, might be assumed to behave as if they maximize a single utility function subject to a budget constraint. This constrained maximization problem generates household commodity demands as functions of aggregate household expenditure and commodity prices.

(ii) Equations which describe demand by industries for inputs, and equations which specify the supply of commodities by each industry.

---

1. The author is indebted to Peter Dixon who carefully read a previous draft of this paper and made many helpful suggestions.

2. Named in recognition of the contributions made by L. Johansen [5].

(iii) Pricing equations setting prices equal to costs.

(iv) Market clearing equations for commodities and primary factors.

(v) Equations which define aggregated quantities such as gross domestic product, the consumer price index and aggregate employment. These equations build up a macroeconomic summary of all the micro level changes occurring in the model.

An equilibrium solution of the model is a vector of prices and quantities Z which satisfies

$$F(Z) = \underset{\sim}{0} \quad . \tag{1.1}$$

Johansen [5] used the following local approach to locate a new equilibrium Z "near" a known solution Z_I :

If F is differentiable then

$$F(Z) = F(Z_I) + F'(Z_I)(Z - Z_I) + o(Z - Z_I) \quad ,$$

where $F'(Z_I)$ denotes the Jacobian matrix of F at Z_I and $o(Z)$ is a function with the property that

$$\lim_{Z \to \underset{\sim}{0}} \frac{o(Z)}{\|Z\|} = 0 \quad . \tag{1.2}$$

Provided then that Z is a zero of F, or very nearly satisfies $F(Z) = 0$,

$$F'(Z_I)(Z - Z_I) = o(Z) \quad . \tag{1.3}$$

Intuitively, (1.2) and (1.3) show us that if Z is an equilibrium solution of the model and if Z is close to Z_I, then the left

(V) B <u>is thrice continuously differentiable on</u> $U_o \times V_o$, <u>and so</u> g <u>is four times continuously differentiable on</u> U_o ;

(VI) <u>there is</u> $K > 0$ <u>such that, for each</u> x <u>in</u> U_o, <u>and each pair</u> y^1, y^2 <u>in</u> V_o

$$\|B(x, y^1) - B(x, y^2)\| \leq K \|y^1 - y^2\| \quad ;$$

(VII) B <u>and its first two derivatives are bounded on</u> $U_o \times V_o$, <u>and</u> g <u>and its first three derivatives are bounded on</u> U_o.

PROOF

Conclusions I, II and III follow directly from the implicit function theorem (see Dieudonne [1, pp.265-267] and Lang [6, pp.362-364]).

(a) Proof of (IV) : <u>The map</u> $A_2 : (x, y) \to D_2F(x, y)$ <u>is invertible on</u> $U_o \times V_o$.

$\mathrm{Det}[D_2F(x_I, y_I)] \neq 0$ and also D_2F is continuous on $U_o \times V_o$. Therefore, continuity of the determinant function ensures that there is a neighbourhood of (x_I, y_I) on which the determinant of A_2 is nonzero. By taking $U_o \times V_o$ to be sufficiently small then we have the required property.

The derivation of the differential equation for g is given at the beginning of this section.

(b) Proof of (V) : B <u>is thrice continuously differentiable on</u> $U_o \times V_o$ <u>and</u> g <u>has four continuous derivatives on</u> U_o .

As F is four times continuously differentiable then D_1F and D_2F

equation may be rewritten as

$$\underset{\sim}{0} = D_1F(x, g(x)) + D_2F(x, g(x)) \circ Dg(x) \quad ,$$

and so

$$Dg(x) = - [D_2F(x, g(x))]^{-1} \circ D_1F(x, g(x)) \quad .$$

This equation together with the initial condition $y_I = g(x_I)$ comprises an initial value problem of the type discussed in the previous sections.

PROPOSITION 3 Suppose that $F : U \times V \to \mathbb{R}^m$ is four times continuously differentiable, and suppose that F and its four derivatives are bounded on $U \times V$. If there is (x_I, y_I) in $U \times V$ such that $F(x_I, y_I) = \underset{\sim}{0}$ and $D_2F(x_I, y_I)$ is invertible then there exist connected neighbourhoods U_o of x_I in U, V_o of y_I in V, and a function $g : U_o \to V$ such that :

(I) $g(x_I) = y_I$;

(II) for each x in U_o ,

$$F(x, g(x)) = \underset{\sim}{0} \ ;$$

(III) $\{(x, g(x)) : x \in U_o\}$ is the set of all points (x, y) in $U_o \times V_o$ with $F(x, g(x)) = \underset{\sim}{0}$;

(IV) for (x, y) in $U_o \times V_o$ the map $(x, y) \to D_2F(x, y)$ is invertible and

$$g'(x) = B(x, g(x))$$

where

$$B(x, y) = - [D_2F(x, y)]^{-1} \circ D_1F(x, y) \ ;$$

hand side of (1.3) is small. Therefore, a new equilibrium solution of the model may be approximated by a vector Z which solves the system of linear equations

$$F'(Z_I)(Z - Z_I) = \underset{\sim}{0} \quad , \qquad (1.4)$$

where, it should be noted, that $F'(Z_I)$ is an $n \times m$ matrix. Basically, then, Johansen advocated replacement of the problem of finding zeroes of F with the easier problem of solving the linear system of equations (1.4).

In general $n >> m$. That is, the number of variables (n) distinguished by the model greatly exceeds the number of linear equations (m) involving them. So, to obtain an approximation to a new equilibrium solution of the model $n - m$ prices and quantities must be pre-assigned values. This selection of $n - m$ exogenous variables partitions Z and Z_I into (X, Y) and (X_I, Y_I), respectively, where X denotes the vector of exogenous prices and quantities and Y is used to denote the remaining m endogenous variables. Equation (1.4) may now be re-written as

$$F_1'(X_I, Y_I)(X - X_I) + F_2'(X_I, Y_I)(Y - Y_I) = \underset{\sim}{0} \qquad (1.5)$$

where $F_1'(X_I, Y_I)$ is the $(n - m) \times m$ matrix of partial derivatives of F with respect to the exogenous entities, and $F_2'(X_I, Y_I)$ is the $m \times m$ matrix of partial derivatives of F with respect to the endogenous prices and quantities. These matrices are obtained by partitioning $F'(X_I, Y_I)$ into two matrices whose columns correspond to partial derivatives of F with respect to an exogenous entity (for $F_1'(X_I, Y_I)$) or an endogenous variable (for $F_2'(X_I, Y_I)$).

Provided that $F_2'(X_I, Y_I)$ is invertible (1.5) may be solved for Y.

In replacing equation (1.3) by the linearized equation (1.4) we have constructed an approximation to (1.1), and so, in general, the vector (X, Y) obtained from (1.5) will not be an equilibrium solution of the original model, but hopefully is a good estimate of such a solution. The estimation error is called the "linearization error" here. Certainly, if (X, Y) is near to (X_I, Y_I) then the linearization error will be small, but then two problems immediately arise : Being "near to (X_I, Y_I)" may be very restrictive, and, even if X is chosen to be acceptably close to X_I, this does not guarantee that Y is "near" to Y_I. Despite these problems there are very good reasons for using Johansen's linearization method to estimate equilibrium solutions of a model :

(i) Solution of systems of linear equations on modern computers is not expensive nor are these solution methods likely to be severely restricted by the computational speed and size of the machine, even though several million equations may be involved. The same is not true for solution methods employed for systems of nonlinear equations (see, for example, the comments of Smale [8, p.1]).

(ii) Exogenous variables may easily be converted to endogenous variables and vice versa.[1] This flexibility is greatly

1. Dixon, Parmenter, Sutton and Vincent, hereafter referred to as DPSV, [2, pp.49-51] discuss the importance of this feature.

5. Conditions the Equations of a General Equilibrium Model Must Satisfy if Euler's Method and Richardson's Extrapolation Rule are to be Used

The proposition proved in this section imposes conditions on the equations of a general equilibrium model (that is, the function $F : U \times V \to \mathbb{R}^m$) which ensure that application of Euler's method and Richardson's extrapolation will improve the accuracy of a Johansen style solution. These conditions are simple enough and should cause no problems, in that, if a model is built using a modelling technique which yields a function F that is differentiable, then it is unlikely to be a significant difficulty to construct F so that the hypotheses of Proposition 3 are satisfied. The other thing to notice about these conditions is that they will generally be easy to check for a particular function F.

Proposition 3 is essentially an application of the implicit function theorem. That is, if F satisfies the hypotheses of Proposition 3 then the implicit function theorem ensures the existence of a function g such that

$$F(x, g(x)) = \underset{\sim}{0} \qquad (5.1)$$

for x near x_I, and

$$g(x_I) = y_I .$$

Now, applying the chain rule to (5.1), we obtain

$$\underset{\sim}{0} = DF(x, g(x)) \circ (\iota, Dg(x))$$

where $\iota : U \to U$ denotes the identity function on U. This and

$$\|\delta_{j+1}\| \leq \left[\left\{1 + \beta\,\|h\|\right\}^{j+1} - 1\right] \|O\,(h^{(3)})\|/\beta\,\|h\|$$

$$\leq \left\{1 + \beta\,\|h\|\right\}^{j+1} \|O\,(h^{(2)})\|/\beta\ ,$$

which proves then that $e_j^s - \xi^s(j)h$ is $O(h^{(2)})$ for each $j=0,1,2,\ldots,s$ and any positive integer s.

reduced in models where nonlinear solution methods are used. With Johansen's approach switching a variable from being exogenous to endogenous, and therefore making another which is endogenous into an exogenous variable, is accomplished by swapping a column of $F_1'(X_I, Y_I)$ with a column of $F_2'(X_I, Y_I)$.

(iii) Even if large parts of the model are modified, only the coefficients in $F'(X_I, Y_I)$ will be changed. This is easily accomplished without changing programmes already written for the solution algorithm. A preliminary programme which computes the coefficients of $F'(X_I, Y_I)$ given (X_I, Y_I) is required, and when the model is changed only this programme will require modification. By contrast, a nonlinear solution method may be intimately connected with the model, so that modifications of the model could lead to major revision of the solutions routines (see DPSV [2, pp. 70-77]).

With these advantages in mind it is reasonable to ask : Is it possible to improve the Johansen approximation using a numerical method which preserves the basic structure of a linear model, and which is inexpensive to run on a modern computer?

This question was answered affirmatively by DPSV [2, section 31.4] who solved an initial value problem, of which (1.5) is a special case, using Euler's method. The initial value problem studied by DPSV may be obtained by appealing to the implicit function theorem (see Dieudonné [1, pp.265-267]) which states that, if $F_2'(X_I, Y_I)$ is invertible, then there is a function g defined for X near X_I such that

$$\begin{aligned} F(X, g(X)) &= \underset{\sim}{0} \\ g'(X) &= G(X, g(X)) \\ g(X_I) &= Y_I \end{aligned} \qquad (1.6)$$

and

$$G(X, g(X)) = -\left[F_2'(X, g(X))\right]^{-1} F_1'(X, g(X)) \quad .$$

This initial value problem is defined on a neighbourhood of X_I in $\mathbb{R}^n$ where $n >> 1$, and so the classical results on convergence of Euler's method must be generalized to apply in this situation (see section 3). Now, at each step, solution of (1.6) using Euler's algorithm involves :

(I) update of (X_I, Y_I) to, say, (X_N, Y_N) to account for small shifts in the values of the exogenous variables away from X_I,

(II) evaluation of $F'(X_N, Y_N)$,

and (III) solution of the linear system (1.5) where the partial derivatives there are now evaluated at (X_N, Y_N).

It is a simple matter to write code for step (I), and as pointed out in (iii) above, a programme which computes step (II) should already exist as part of a flexible package for the solution of (1.5).

While Euler's method exploits the structure of the linearized equation (1.5), it converges only linearly to one of the

to obtain

$$\xi^s(j+1) = \xi^s(j) + D\xi^s(j)1 + \tfrac{1}{2}\int_0^1 (1-r)D^2\xi^s(j+r)dr(1, 1)$$

where, it should be noted, the integral curves of (4.2) are defined on $\mathbb{R}$ - see [7, pp.299-313] . If $\delta_j^s = e_j^s - \xi^s(j)h$ for $j=0,1,2,\ldots,s-1$ then

$$\begin{aligned} \delta_{j+1}^s = {} & \delta_j^s + D_2B(x_j^s, g(x_j^s))(\delta_j^s)h - \tfrac{1}{2}D^s g(x_j^s)h^{(2)} \\ & - D\xi^s(j).(h) - R_2(j, 1; \xi^s)(1, 1)h \\ & + R_2((x_j^s, g(x_j^s)), e_j; B)\, e_j^{s(2)}h - R_3(x_j^s, h; g)h^{(3)} \; . \end{aligned}$$

Now, from Proposition 1 we know that e_j^s is $O(h)$, and, by applying the theorem proved in Gray [3, pp.123-125], we see that $R_2(j, 1; \xi^s)(1, 1)$ is $O(h^{(2)})$. Therefore,

$$\|\delta_{j+1}^s\| \le \|\delta_j^s\| + \|D_2B(x_j^s, g(x_j^s))\| \, \|\delta_j^s\| \, \|h\| + \|O(h^{(3)})\| \; ,$$

and an argument by induction establishes that

$$\begin{aligned} \|\delta_{j+1}^s\| \le {} & \{1 + \|D_2B(x_j^s, g(x_j^s))\| \, \|h\|\}^j \, \|\delta_o^s\| \\ & + \sum_{i=0}^{j} \{1 + \|D_2B(x_j^s, g(x_j^s))\| \, \|h\|\}^i \, \|O(h^{(3)})\| \; . \end{aligned}$$

But $e_o = \underset{\sim}{0}$ and so $\delta_o = 0$. Now if β is the bound for D_2B on $U_o \times V_o$ we get

$$\|\delta_{j+1}\| \le \sum_{i=0}^{j} \left\{1 + \beta \, \|h\|\right\}^i \, \|O(h^{(3)})\| \quad ,$$

linear differential equation,

$$D\xi^s(t) = D_2B(x_t^s, g(x_t^s))\,(\xi^s(t)h) - \tfrac{1}{2}D^s g(x_t^s)h$$

$$\xi^s(0) = 0 ,$$

and

$$x_t^s = x_I + th .$$

PROOF

As B is thrice continuously differentiable on $U_o \times V_o$ the solution g of (3.4) is four times continuously differentiable on U_o, and so we may apply Taylor's theorem to obtain

$$g(x_{j+1}^s) = g(x_j^s) + Dg(x_j^s)h + \tfrac{1}{2}D^s g(x_j^s)\,h^{(2)}$$
$$+ \tfrac{1}{2}\int_0^1 (1 - r)^2\, D^3g(x_j^s + rh)dr\, h^{(3)} ,$$

for $j=0,1,2,\ldots,s-1$. Hence

$$e_{j+1}^s = e_j^s + \left[B(x_j^s, u_j^s) - B(x_j^s, g(x_j^s))\right]h$$
$$- \tfrac{1}{2}D^2g(x_j^s)\,h^{(2)} - R_3(x_j^s, h; g)\,h^{(3)} ,$$

and as B is twice differentiable, another application of Taylor's theorem gives

$$e_{j+1}^s = e_j^s + D_2B(x_j^s, g(x_j^s))\,(e_j^s)h - \tfrac{1}{2}D^2g(x_j^s)\,h^{(2)}$$
$$+ R_2((x_j^s, g(x_j^s)), e_j^s; B)\,e_j^{s\,(2)}h - R_3(x_j^s, h; g)\,h^{(3)} .$$

Further, as ξ satisfies (4.2) then ξ is twice continuously differentiable and so, we may apply Taylor's theorem once again

zeroes $(X, g(X))$ of F given in (1.6) (this is proved in section 3). Such a slow rate of convergence could mean that a large number of Euler steps may be required, at considerable cost, to eliminate most of the linearization error. Hence some means of hastening the rate of convergence will generally be required. A suggestion of this type is made by DPSV [2, p.88] when they observed that, to a high order of accuracy, their results satisfied

$$R(2s) - R(s) = 2\{R(4s) - R(2s)\} , \qquad (1.7)$$

where $R(q)$ denotes the result obtained from a q-step Euler simulation, and where DPSV set

$$s = 1, 2, 8, \text{ and } 16.$$

Equation (1.7) is just Richardson's extrapolation formula (see Isaacson and Keller [4, pp.372-374]), which is shown, in section 4, to produce a sequence of approximations that converge quadratically to a zero $(X, g(X))$ of F.

In summary, sections 3 and 4 of this paper prove that the generalizations of Euler's method and Richardson's extrapolation converge linearly and quadratically, respectively, to zeroes of F if the functions g and G of (1.6) satisfy certain smoothness and boundedness conditions near (X_I, Y_I). But the important practical question is : What properties must the equations of the model (that is, the function F) possess to ensure that use of Euler's method and Richardson's extrapolation rule will produce a sequence of approximations which converge quadratically to an equilibrium of the model? That is, when can Euler's method and Richardson's extrapolation rule be used to improve the Johansen style solution? The answer is that the function F must only possess some easily checked smoothness and boundedness properties near (X_I, Y_I). This is proved in section 5.

The next section explains a number of mathematical concepts and results which should be consulted as required while reading sections 3, 4 and 5.

convergent in the case of initial value problems defined on some open interval. (For example, see Isaacson and Keller [4, pp.373-374].) Here we extend these results to give a proof which is valid for initial value problems defined on an open set in $\mathbb{R}^{n-m}$.

As will be seen in the following proposition on Richardson's extrapolation the function ξ^s of (4.1) is defined as the solution of the first order linear differential equation

$$D\xi^s(t) \;=\; D_2B(x_t^s,\, g(x_t^s))\;(\xi^s(t)h) - \tfrac{1}{2}D^s g(x_t^s)h \qquad (4.2)$$

for which

$$\xi^s(0) \;=\; 0 \;,$$

where $x_t^s = x_I + th$. Under the hypotheses of the proposition below the maps $t \to D_sB(x_t^s, g(x_t^s))$ and $t \to D^s g(x_t^s)$ are continuous for t near 0, and so this differential equation is known to have a unique solution ξ^s defined on $\mathbb{R}$ (see Maurin [7, pp.299-306]).

PROPOSITION 2 Suppose that the function B of (3.4) has three continuous derivatives on the neighbourhood $U_o \times V_o$ of Proposition 1. Further, suppose that D_2B is bounded on $U_o \times V_o$ and that D^3g is bounded on U_o. If $e_o = \underset{\sim}{0}$ then for any positive integer s

$$e_j^s \;=\; \xi^s(j)h \;+\; O(h^{(2)}) \;, \qquad j=0,2,\ldots,s,$$

where $h = H/s$ for H in U_o, ξ^s is a solution of the first order

4. Richardson's Extrapolation Rule

For the initial value problem (3.4) defined on $U \subset \mathbb{R}^{n-m}$ we now show how a sequence of estimates μ_j^s may be generated from the Euler estimates u_j^s of $g(x_j^s)$ which converge quadratically to $g(x_j^s)$. This is done using Richardson's extrapolation rule which depends, for its successful improvement of the convergence rate, on the assumption that the error $e_j^s = u_j^s - g(x_j^s)$ satisfies

$$e_j^s = \xi^s(j)h + 0(h^{(2)}) , \qquad (4.1)$$

where ξ is defined later. Now suppose that Euler's method is performed with spacings h and $h/2$, in which case $x_j^s = x_{2j}^{2s}$ and, if (4.1) is true,

$$u_j^s = g(x_j^s) + \xi^s(j)h + 0(h^{(2)})$$

and

$$u_{2j}^{2s} = g(x_j^s) + \tfrac{1}{2}\xi^s(j)h + 0(h^{(2)}) .$$

Therefore,

$$2u_{2j}^{2s} - u_j^s = g(x_j^s) + 0(h^{(2)}) .$$

Thus the accuracy of Euler's method is improved by an order of magnitude in h if the approximation

$$\mu_j^s = 2u_{2j}^{2s} - u_j^s$$

to $g(x_j^s)$ is used in place of u_j^s. This technique has been applied by DPSV [2, pp.88-93 and Section 47] in their work on the ORANI model of the Australian economy.

Many authors have shown that Richardson's extrapolation rule yields a sequence of approximations which are quadratically

2. Mathematical Preliminaries

A complete account of the material discussed here is given in Lang [6, Chapter 16].

Let $f : U \to \mathbb{R}^q$ and let $z \in U$ where $U \subset \mathbb{R}^p$ is open, and $z = (z_1, z_2, \ldots, z_n)$. Also let $f = (f_1, f_2 \ldots, f_q)$, where $f_i : U \to \mathbb{R}$ for $i = 1, 2, \ldots, q$ are the co-ordinate functions of f.

(i) A map $L : U \to \mathbb{R}^2$ is <u>linear</u> if and only if

$$L(x + y) = L(x) + L(y) , \qquad (2.1)$$

and

$$L(\lambda x) = \lambda L(x) ,$$

for each x, y in U and each $\lambda \in \mathbb{R}$. Now the $p \times q$ matrix denoted by $[L]$ which satisfies,

$$L(x) = [L]\, x^t \quad \text{for each} \quad x \in R^p$$

is uniquely determined for given co-ordinates on $\mathbb{R}^q$ and $\mathbb{R}^p$. In line with this identification drop the parentheses and write $L(x)$ as Lx.

Example :

Let $L(x, y) = (-y, -x)$. Then

$$[L] = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

Rather than preserve the cumbersome "[L]" notation, use L to also denote the matrix of the linear map L. The entity intended will be clear from the context.

(ii) $\|\chi\|$ denotes the <u>norm</u> of the quantity χ, where χ may be a linear map, a matrix or a vector. Hence the classical inequality involving the norm of a linear map is written here as

$$\|L(x)\| \leq \|L\| \; \|x\| \quad \text{for each } x, \tag{2.2}$$

where

$$\|L\| = \mathit{sup}\, \{\|L(x)\| : \|x\| = 1\} \quad . \tag{2.3}$$

<u>Examples</u> :

For a vector $x = (x_1, x_2, \ldots, x_p)$ in $\mathbb{R}^p$, $\|x\|$ might be defined by

$$\|x\| = \left(x_1^2 + x_2^2 + x_3^2 + \ldots + x_p^2\right)^{\frac{1}{2}} \quad .$$

Let L be defined by

$$L(x, y) = (x_1 + x_2, x_1 - x_2)$$

then

$$\|L\| = \mathit{sup}\, \{\|(x_1 + x_2, x_1 - x_2)\| : \|(x_1, x_2)\| = 1\}$$

$$= \mathit{sup}\left\{\sqrt{(x_1 + x_2)^2 + (x_1 - x_2)^2} : \|(x_1, x_2)\| = 1\right\}$$

$$= \mathit{sup}\left\{\sqrt{2(x_1^2 + x_2^2)} : \|(x_1, x_2)\| = 1\right\}$$

$$= \mathit{sup}\left\{\sqrt{2}\|(x_1, x_2)\| : \|(x_1, x_2)\| = 1\right\}$$

$$= \sqrt{2} \; .$$

(iii) Let

$$\mathcal{A}(U, \mathbb{R}^q) = \{L : U \to \mathbb{R}^q \mid L \text{ is linear}\} \quad .$$

Equipped with the norm defined in (2.3) $\mathcal{A}(U, \mathbb{R}^q)$ is a normed vector space which is similar to $\mathbb{R}^q$, and so the idea of derivative

principle of mathematical induction to the inequality above for e^s_{j+1}, we obtain

$$\|e^s_{j+1}\| \leq (1 + K\,\|h\|)^{j+1}\,\|e^s_o\| + \frac{1}{K}\left\{(1 + K\,\|h\|)^{j+1} - 1\right\}\tau^s . \tag{3.8}$$

Recall now that $u^s_o = y_I$ and so $e^s_o = \underset{\sim}{0}$. Hence, the first term on the right in (3.8) may be neglected. Also for $\beta \in \mathbb{R}$ with $\beta + 1 > 0$

$$(1 + K\,\|h\|)^{\beta+1} \leq \exp\,\{(\beta + 1)K\,\|h\|\} \quad .$$

Hence the last inequality for e^s_{j+1} may be simplified to

$$\|e^s_{j+1}\| \leq \tau^s \exp\,\{(j + 1)K\,\|h\|\}/K\,\|h\| \quad . \tag{3.9}$$

Finally to obtain the desired result τ^s must be estimated using (3.6). With the bounds α, γ and δ for B, DB_1 and DB_2, respectively, on $U_1 \times V_1$ given in hypotheses (1) and (3) of the proposition, we obtain

$$\tau^s \leq (\gamma + \alpha\delta)\,\|h\|^2 \quad .$$

follows from hypothesis (2) of the proposition and the existence - uniqueness theorem for solutions of ordinary differential equations (see Lang [6, pp.374-5]). Taylor's theorem may then be applied to g to obtain

$$g(x^s_{j+1}) = g(x^s_j) + B(x^s_j, g(x^s_j))\, h + R_2(x^s_j, h; g)\, h^{(2)} \tag{3.5}$$

where

$$R_2(x^s_j, h; g) = \int_0^1 (1 - t) D^2 g(x^s_j + th)\,dt \quad . \tag{3.6}$$

Subtraction of (3.5) from (3.3) yields

$$e^s_{j+1} = e^s_j - \left[B(x^s_j, g(x^s_j)) - B(x^s_j, u^s_j)\right] h - R_s(x^s_j, h; g)\, h^{(2)} . \tag{3.7}$$

Now by hypothesis (2) there is a real number $K > 0$ such that

$$\|B(x, y^1) - B(x, y^2)\| \leq K \|y^1 - y^2\|$$

where (x, y^1) and (x, y^2) lie in $U_o \times g(U_o)$. Therefore (3.7) implies that

$$\begin{aligned} \|e^s_{j+1}\| &\leq (1 + K\|h\|)\, \|e^s_j\| + \|R_2(x^s_j, h; g)\, h^{(2)}\| \\ &\leq (1 + K\|h\|)\, \|e^s_j\| + \|\tau^s\| \end{aligned}$$

where $\tau^s = \max\{\|R_2(x^s_j, h; g)\, h^{(2)}\| : j = 0, 1, \ldots, s-1\}$.

For each j the mapping of h into $R_2(x^s_j, h; g)(h, h)$ is continuous. (As may be seen from (2.9), $R_2(x^s_j, h; g)$ involves the second derivative of g which is continuous on U_o.) So, by shrinking U_o if necessary, we are assured that for all h and j, each $R_2(x^s_j, h; g)(h, h)$ is bounded. Now, by applying the

discussed below for functions $f : U \to \mathbb{R}^q$ may easily be extended to functions $f : V \to \mathcal{A}(U, \mathbb{R}^q)$ for an open set V in $\mathbb{R}^s$, say. As will be seen below the derivative of such a function will be an element of the vector space $\mathcal{A}(V, \mathcal{A}(U, \mathbb{R}^q))$. The norm defined on this space is given by

$$\|L\| = \sup\{\|L(M)\| : M \in \mathcal{A}(U, \mathbb{R}^q) \text{ and } \|M\| = 1\} \ .$$

(iv) f is <u>differentiable</u> at z if and only if there is a linear map $L_z : \mathbb{R}^p \to \mathbb{R}^q$ such that

$$f(z + h) = f(z) + L_z(h) + o(h) \ ,$$

where the function o has the property

$$\lim_{h \to \underset{\sim}{0}} \frac{o(h)}{\|h\|} = \underset{\sim}{0} \ .$$

For h close to $\underset{\sim}{0}$ the definition embodies the idea that the value of $f(z + h)$ may be approximated by the value of a linear map at h plus a constant. If f is differentiable then L_z is uniquely determined, and it is written as $Df(z)$. That is,

$$Df(z)(h) = L_z(h) \quad . \tag{2.4}$$

<u>Examples</u> :

(a) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(z_1, z_2) = (z_2, z_1) \text{ for } (z_1, z_2) \in \mathbb{R}^2 \ .$$

Now for (h_1, h_2) in $\mathbb{R}^2$

$$f((z_1, z_2) + (h_1, h_2)) = f(z_1 + h_1, z_2 + h_2)$$
$$= (z_2 + h_2, z_1 + h_1)$$
$$= (z_2, z_1) + (h_2, h_1)$$
$$= f(z_1, z_2) + (h_2, h_1) .$$

So, here the function $o(h_1, h_2) = 0$, and

$$L(z_1, z_2)(h_1, h_2) = (h_2, h_1) .$$

Hence the derivative $Df(z_1, z_2)$ is given by

$$Df(z_1, z_2)(h_1, h_2) = (h_2, h_1) ,$$

and the matrix of this linear map is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

(b) Let $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$g(t) = t^2 ,$$

and let $t_o \in \mathbb{R}$. For $h \in \mathbb{R}$

$$g(t_o + h) = (t + h)^2$$
$$= t_o^2 + 2ht_o + h^2$$
$$= g(t_o) + 2ht_o + o(h) ,$$

where $o(h) = h^2$. So, a suitable linear function L_{t_o} is given by

$$L_{t_o}(h) = 2t_o h .$$

We now prove the convergence result for this generalized version of Euler's method. Another proof of this result is given in DPSV [2, Chapter 35].

PROPOSITION 1 Suppose that the function $B : U \times V \to \mathcal{A}(\mathbb{R}^{n-m}, \mathbb{R}^m)$ in the initial value problem (3.4) satisfies the following conditions on some neighbourhood $U_1 \times V_1$ of (x_I, y_I) in $U \times V$:

(1) B is bounded with

$$\|B(x, y)\| \le \alpha \text{ for } (x, y) \text{ in } U_1 \times V_1 ;$$

(2) B is continuously differentiable and there exists $K > 0$ with the property that

$$\|B(x, y_1) - B(x, y_2)\| \le K\|y_1 - y_2\|$$

for x in U_1 and y_1, y_2 in V_1 ;

(3) The partial derivatives of B satisfy

$$\|D_1 B(x, y)\| \le \alpha \text{ and } \|D_2 B(x, y)\| \le \delta$$

for some α and δ in $\mathbb{R}^+$ and any (x, y) in $U_1 \times V_1$.

Then there is a neighbourbood U_o of x_I in U_1 such that for $H \in U_o$ and any integer s

$$\|e_j^s\| \le \|h\| (\gamma + \alpha\delta) \exp\{(j+1)K \|h\|\}/K$$

where $h = H/s$.

PROOF

First notice that there is a twice differentiable function g satisfying (3.4), and that this function is uniquely determined on a suitably small region $U_o \times V_o$ of (x_I, y_I) in $U_1 \times V_1$. This interval, the error in the estimate of g at x_j^s given by $\|g(x_j^s) - u_j^s\|$, is of order $h = H/s$. Further, these errors converge uniformly to zero as h approaches 0 - that is, as s, the number of steps, increases. In this section we outline how this convergence result may be extended to deal with initial value problems defined on an open set U in $\mathbb{R}^{n-m}$ by

$$Dg(x) = B(x, g(x))$$
$$g(x_I) = y_I \quad , \qquad (3.4)$$

where now $g : U \to V \subset \mathbb{R}^m$ for $n - m \geq 1$ and $m \geq 1$, $Dg : U \to \mathcal{A}(\mathbb{R}^{n-m}, \mathbb{R}^m)$ and $B : U \times V \to \mathcal{A}(\mathbb{R}^{n-m}, \mathbb{R}^m)$. As in (3.2) $s + 1$ vectors x_j^s are chosen but, to ensure that each x_j^s lies in U, it is assumed that $x_I + tH$ is in U for each t in $[0, 1]$. Again, Euler approximations u_j^s to the points $g(x_j^s)$ are obtained from (3.2) and (3.3), where (3.3) now involves evaluation of the linear map $B(x_j^s, u_j^s)$ at $h \in \mathbb{R}^{n-m}$ rather than simple multiplication of the real numbers $B(x_j^s, u_j^s)$ and h, as was the case for problem (3.1). Of course the evaluation of the linear map at h may be thought of as the product of a matrix and a vector.

Define the error e_j^s in the Euler approximation u_j^s to $g(x_j^s)$ by

$$e_j^s = u_j^s - g(x_j^s) \quad , \qquad j=0,1,2,\ldots,s,$$

where

$$e_o = u_o - g(x_I) = \underset{\sim}{0} \quad .$$

That is,

$$Dg(t_o)(h) = 2t_o h \quad .$$

At this point it should be remembered that $Dg(t_o)$ is a linear map which is used to approximate g near t_o. The slope of the graph of g at t_o is $2t_o$, which traditionally has been written as

$$g'(t_o) = 2t_o \quad .$$

In general, we now use $f'(z)$ to denote the matrix of the linear map $Df(z)$. In fact, $f'(z)$ is the matrix of first order partial derivatives of f at z, and is called the <u>Jacobian</u> of f at z.

(v) Let $f : U \to \mathbb{R}^q$ be given by $f = (f_1, f_2, \ldots, f_q)$ where for $j = 1, 2, \ldots, q$ $f_j : U \to \mathbb{R}$. Then for z in U

$$Df(z) = (Df_1(z), Df_2(z), \ldots, Df_q(z)) \quad .$$

(vi) The partial derivatives of f_j of order k are written as $D_{i_1} \ldots {}_{i_k} f_j(z)$ where $j = 1, 2, \ldots, q$ and $i_1 = 1, 2, \ldots, p$.

(vii) If we partition $\mathbb{R}^p$ into $\mathbb{R}^s \times \mathbb{R}^{p-s}$ then, for x in $\mathbb{R}^s$ and y in $\mathbb{R}^{p-s}$ with (x, y) in U, we define D_1f and D_2f by

$$D_1f(x, y)\, h_1 = Df(x, y)(h_1, \underset{\sim}{0}_1)$$

and $\qquad (2.5)$

$$D_2f(x, y)\, h_2 = Df(x, y)(\underset{\sim}{0}_2, h_1)$$

where $\underset{\sim}{0}_1 \in \mathbb{R}^{p-s}$ and $\underset{\sim}{0}_2 \in \mathbb{R}^s$ are the zero vectors. It now follows that

$$Df(x, y)(h_1, h_2) = D_1f(x, y)h_1 + D_2f(x, y)h_2 \quad . \qquad (2.6)$$

Examples :

(a) Let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x, y) = xy \quad ,$$

and let $(a, b) \in \mathbb{R} \times \mathbb{R}$. Then

$$f'(a, b) = (b \ a)$$

and for $(h_1, h_2) \in \mathbb{R}^2$

$$Df(a, b)(h_1, h_2) = (b \ a)\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = hb_1 + ah_2$$

and

$$\begin{aligned} D_1f(a, b)h_1 &= Df(a, b)(h_1, \underset{\sim}{0}) \\ &= (b \ a)\begin{pmatrix} h_1 \\ \underset{\sim}{0} \end{pmatrix} \\ &= b\,h_1 \quad . \end{aligned}$$

(b) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$f(x, y) = (xy, x^2)$$

then

$$f'(a, b) = \begin{pmatrix} b & a \\ 2a & 0 \end{pmatrix}$$

$$Df(a, b)(h_1, h_2) = (bh_1 + ah_2, 2ah_1) \quad ,$$

3. Euler's Method

Consider the initial problem defined on some interval in $\mathbb{R}$ by

$$\begin{aligned} g(x) &= B(x, g(x)) \\ g(x_I) &= y_I \quad , \end{aligned} \qquad (3.1)$$

where g maps I into some interval J and $B : I \times J \to \mathbb{R}$. The simplest, and probably the best known way of generating approximations to a solution g of (3.1) at points in the interval $[x_I, x_I + H]$ is Euler's method, which proceeds as follows :

(i) First distinguish $s + 1$ points in the interval $[x_I, x_I + H]$ by

$$x_j^s = x_I + jh \ , \quad h = H/s \qquad j=0,1,2,\ldots,s \qquad (3.2)$$

(where $x_o = x_I$ and $x_s^s = x_I + H$).

(ii) Now let $u_o = g(x_I)$ and define approximations u_j^s to the values of g at each point x_j^s by

$$u_{j+1}^s = u_j^s + B(x_j^s, u_j^s)h. \qquad j=0,1,2,\ldots,s-1 \qquad (3.3)$$

$B(x_j^s, u_j^s)$ is an approximation to the slope of the line tangent to g at x_j^s, and so the function $f(z) = u_j^s + B(x_j^s, u_j^s)z$ is an approximation to the straight line tangent to g at x_j^s.

It is well known (see Isaacson and Keller [4, pp.367-369]) for initial value problems which, like (3.1), are defined on some where $h^{(j)}$ is the j-tuple of vectors $(h, h, \ldots, h)$ and

$$R_\ell(z, h; f) = \int_0^1 \frac{(1-t)^{\ell-1}}{(\ell-1)!} D^\ell f(z + th)dt . \qquad (2.9)$$

We will use the inequality

$$\|R_\ell(z, h; f)\, h^{(\ell)}\| \le \|R_\ell(z, h; f)\|\, \|h\|^\ell \qquad (2.10)$$

quite often in this paper. This is obtained by ℓ applications of the inequality for the norm of a linear map.

and

$$D_1 f(a, b)h_1 = \begin{pmatrix} b & a \\ 2a & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ 0 \end{pmatrix}$$

$$= (bh_1 \;\; 2ah_1) .$$

(viii) Let $g : V \to U$ and $f : U \to \mathbb{R}^q$ then the function $f \circ g : V \to \mathbb{R}^2$ is defined by

$$f \circ g(z) = f(g(z))$$

for each z in V. If g is differentiable at $z \in V$ and if f is differentiable at $g(z)$ (which must lie in U) then $f \circ g$ is differentiable and

$$D(f \circ g)(z) = Df(g(z)) \circ Dg(z)$$

or

$$(f \circ g)'(z) = f'(g(z)) \,.\, g'(z) .$$

<u>Example</u> :

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$g(P, Q) = PQ$$

and

$$f(Z) = \ell og(Z) .$$

Then

$$(f \circ g)(P, Q) = \ell og(PQ) ,$$

and

$$(f \circ g)'(P, Q) = \ell og'(g(P, Q)) \,.\, g'(P, Q)$$

$$= \frac{1}{g(P, Q)} (Q \;\; P) .$$

Hence

$$(f \circ g)'(P, Q)(\Delta P, \Delta Q) = \frac{1}{g(P, Q)} (Q \ P) \begin{pmatrix} \Delta P \\ \Delta Q \end{pmatrix}$$

$$= \frac{1}{g(P, Q)} (Q\Delta P + P\Delta Q)$$

$$= \frac{\Delta P}{P} + \frac{\Delta Q}{Q} .$$

(ix) f is twice differentiable at z if the function $Df : z \to Df(z)$ is differentiable at z. That is, there is a map M_z such that

$$Df(z + k)(h) = Df(z)(h) + M_z(k)(h) + o_h(k)$$

where

$$\lim_{k \to 0} \frac{o_h(k)}{\|k\|} = 0 .$$

We may regard M_z as a bilinear map (i.e., linear in each of its arguments separately) via the identification.

$$M_z(h, k) = M_z(h)(k) \quad \text{for} \quad h, k \in \mathbb{R}^p .$$

Thus, the second derivative of f at z, if it exists, is a symmetric bilinear map, and is written as $D^2f(z)$. We may apply the norm inequality for linear maps twice to obtain

$$\begin{aligned} \|D^2f(z)\ (h, k)\| &= \|D^2f(z)\ (h)\ (k)\| \\ &\le \|D^2f(z)\ (h)\|\ \|k\| \\ &\le \|D^2f(z)\|\ \|h\|\ \|k\| . \end{aligned} \qquad (2.7)$$

Example :

Lef $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$g(x, y) = xy$$

and so,

$$D^2f(x, y)\ ((h_1, h_2), (k_1, k_2)) = (k_1\ k_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

$$= k_1h_2 + k_2h_1 .$$

The matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is the matrix of second order partial derivatives of g. For functions f whose range is $\mathbb{R}$ this is the "Hessian of f at z".

(x) Let $D^\ell f(z) = D(D^{\ell-1}f)(z)$ for $\ell \ge 1$ an integer. f is <u>ℓ times continuously differentiable</u> on U if $D^\ell f(z)$ exists for each z in U, and if the map $D^\ell f : z \to D^\ell f(z)$ is continuous on U. This is equivalent to the statement that all partial derivatives of f of order less than or equal to ℓ exist and are continuous on U.

For each ℓ $D^\ell f(z)$ is a multilinear map defined on $\mathbb{R}^{p\ell}$. When $\ell = 1$, $Df(z)$ is linear, and when $\ell = 2$, $D^2f(z)$ is bilinear.

(xi) <u>Taylor's Theorem</u>

Let $f : U \to \mathbb{R}^q$ be ℓ times continuously differentiable on U. Let $h \in \mathbb{R}^p$ with $z + th$ in U for each t satisfying $0 \le t \le 1$. Then

$$f(z + h) = f(z) + Df(z)\ h + \tfrac{1}{2}D^2f(z)\ h^{(2)} + \dots$$

$$+ \frac{1}{(\ell - 1)!} D^{\ell-1}f(z)\ h^{(\ell-1)} + R_\ell(z, h; f)\ h^{(\ell)} \qquad (2.8)$$