REVUE D'ANALYSE NUMÉRIQUE ET DE THÉORIE DE L'APPROXIMATION
Rev. Anal. Numér. Théor. Approx., vol. 32 (2003) no. 2, pp. 171–175
ictp.acad.ro/jnaat

ON SOME ONE-STEP IMPLICIT METHODS AS DYNAMICAL SYSTEMS

CĂLIN-IOAN GHEORGHIU*

Abstract. The one-step implicit methods, the backward Euler being the most known, require the solution of a nonlinear equation at each step. To avoid this, these methods can be approximated by making use of a one step of a Newton method. Thus the methods are transformed into some explicit ones. We will obtain these transformed methods, find conditions under which they generate continuous dynamical systems and show their order of convergence. Some results on the stability of these explicit schemes, as well as on the shadowing phenomenon are also carried out. Concluding remarks and some open problems end the paper.

MSC 2000. 65L05, 65L20, 65L70, 65P40.

Keywords. One-step implicit methods, Newton method, continuous dynamical systems, convergence, stability, shadowing.

1. INTRODUCTION

We consider the initial value autonomous problem (Cauchy problem):

$$\frac{\mathrm{d}u}{\mathrm{d}t} = f\,(u)\,, \quad u(0) = U \in \mathbb{R}^p, \quad p \geq 1, p \in \mathbb{N}, \tag{1}$$

where $u\,(t) \in \mathbb{R}^p$ denotes a vector valued function of $t \in \mathbb{R}$, and $f \in C\,(\mathbb{R}^p, \mathbb{R}^p)$. Further regularity assumptions on f will be described as required.

The aim of this paper is to solve the above problem by a method which avoids the solution of a nonlinear equation at each step. We will work in the spirit of dynamical systems and will use the notations from the monograph of Stuart and Humphries [9]. In fact, we use one step of a Newton method in order to transform a one-step implicit method into an explicit one.

We will consider the following two classes of implicit schemes:

A/ the *one-stage theta method*

$$U_0 = U, \quad U_{n+1} = U_n + (\Delta t)\, f\big(\,(1-\theta)\,U_n + \theta U_{n+1}\big), \quad n = 1,2,3,\ldots,\ \theta \in [0,1]\,,$$

and

B/ the *two-stage theta method*

$$U_0 = U, \quad U_{n+1} = U_n + (\Delta t)\left[\,(1-\theta)\,f(U_n) + \theta f\,(U_{n+1})\,\right],$$
$$n = 1,2,3,\ldots,\ \theta \in [0,1]\,.$$

*"T. Popoviciu" Institute of Numerical Analysis, P.O. Box 68-1, 3400 Cluj-Napoca, Romania, e-mail: ghcalin@ictp.acad.ro.

Note that both schemes **A/** and **B/** reduce to the forward Euler method when $\theta = 0$ and to the backward Euler method when $\theta = 1$. However, for $\theta \in (0,1)$ the methods differ. For example, if $\theta = \frac{1}{2}$ in **A/** we obtain the implicit midpoint rule while if $\theta = \frac{1}{2}$ in **B/** we get the trapezoidal rule. More important than that, whenever the schemes **A/** and **B/** are implicit they require at each step of time n the solution of a nonlinear system. To avoid this, the method is frequently approximated by applying one step of Newton iteration, with U_n as an initial guess. Consequently, we transform both, **A/** and **B/** into explicit schemes and find out conditions under which they are continuous dynamical systems. This is the subject of the second section. In the third section we analyse them with respect to the order of convergence. In the fourth section we consider the shadowing phenomenon and the numerical stability.

2. THE NEWTON METHOD FOR THE ONE-STAGE AND TWO-STAGE THETA METHODS

It is quite surprising that when the classical Newton method is used in order to solve the implicit equations involved by both **A/** and **B/**, with U_n as initial guess, we are lead up to the same explicit scheme, namely

$$\mathbf{AB}/ \quad U_{n+1} = U_n + h\Phi\left(U_n; h\right)$$

where $h := \Delta t$, and the *increment function* is defined by

$$\Phi\left(U_n; h\right) = \left(I_p - h\theta \mathrm{d}f\left(U_n\right)\right)^{-1} f(U_n),$$

with $\mathrm{d}f\left(\cdot\right)$ denoting the Jacobian of $f\left(\cdot\right)$.

It is a matter of evidence (see for example [2] and [7]) that for

$$h\theta \leq \frac{q}{\left\|\mathrm{d}f\left(U_n\right)\right\|}, \quad 0 < q < 1, \tag{2}$$

the matrix $I_p - h\theta \mathrm{d}f\left(U_n\right)$, has a continuous inverse and the Neumann series

$$\left(I_p - h\theta \mathrm{d}f\left(U_n\right)\right)^{-1} = I_p + h\theta \mathrm{d}f\left(U_n\right) + \cdots + h^m\theta^m\left(\mathrm{d}f\left(U_n\right)\right)^m + \cdots$$

is convergent.

Thus under the condition (2) for $h\theta$, **AB/** defines a dynamical system on $\mathbb{R}^p$ with the discrete semigroup of evolution

$$S^1_{\Delta t} U = U + h\left(I_p - h\theta \mathrm{d}f\left(U\right)\right)^{-1} f(U). \tag{3}$$

Moreover, if $f\left(\cdot\right)$ is continuously differentiable, the above dynamical system is itself continuous.

3. THE ORDER OF CONVERGENCE OF THE TRANSFORMED EXPLICIT METHOD

The truncation error for the scheme $\mathbf{AB/}$, as a numerical approximation of the problem (1), in a point U of $\mathbb{R}^p$ is defined by

$$T\left(U;\Delta t\right):=S\left(\Delta t\right)U-S^1_{\Delta t}U,$$

where $S\left(\cdot\right):\mathbb{R}^p\to\mathbb{R}^p$ is the evolution semigroup associated with (1), (see for example [9, p. 106] and [10]). Taylor expansion of $S\left(\Delta t\right)U$ shows that there exists $\Delta t_c=\Delta t_c\left(U\right)>0$ and $K=K\left(U\right)$ such that

$$\left|T\left(U;\Delta t\right)\right|\leq K\left(\Delta t\right)^2, \quad \forall\Delta t\in\left(0,\Delta t_c\right)$$

whenever $\theta=\frac{1}{2}$. We have to observe that K is the upper bound for the principal error function

$$F\left(U\right)=\tfrac{1}{6}\left(\mathrm{d}^2f\left(U\right)\cdot f^2\left(U\right)+\left(\mathrm{d}f\left(U\right)\right)^2\cdot f\right)-\theta^2f\cdot\left(\mathrm{d}f\left(U\right)\right)^2.$$

We also note that, due to the terms $\mathrm{d}^2f(\cdot)$, f^2 in d^3u, no choice of the parameter θ would imply a method of the third order of accuracy.

For $\theta\neq\frac{1}{2}$ in $\mathbf{AB/}$ the truncation error satisfies an inferior condition, namely

$$\left|T\left(U;\Delta t\right)\right|\leq K_1(\Delta t), \quad \forall\Delta t\in\left(0,\Delta t_c\right)$$

where $K_1=K_1\left(U\right)$ is an upper bound for the principal error function

$$F_1(U)=\left(\tfrac{1}{2}\mathrm{d}f\left(U\right)\cdot f\left(U\right)-\theta\mathrm{d}f\left(U\right)\cdot f\left(U\right)\right).$$

Consequently, the implicit midpoint rule and the trapezoidal rule are of order of accuracy 2, while any other methods are of order 1.

4. ON THE SHADOWING PHENOMENON AND THE NUMERICAL STABILITY OF THE EXPLICIT SCHEME

The idea of approximating a solution of (1) by a numerical scheme with different initial data is sometimes referred to as shadowing.

Thus, for $h\theta$ satisfying (2), let use $\mathbf{AB/}$ in order to solve (1) with two different initial data U and V, $U\neq V$. We have respectively

$$U_{n+1}=U_n+hf\left(U_n\right)+h^2\theta^2f\left(U_n\right)\mathrm{d}f\left(U_n\right)+\mathcal{O}\left(h^3\right), \tag{4}$$

$$V_{n+1}=V_n+hf\left(V_n\right)+h^2\theta^2f\left(V_n\right)\mathrm{d}f\left(V_n\right)+\mathcal{O}(h^3), \quad n=0,1,2,\ldots. \tag{5}$$

Subtracting these two equations, we get

$$U_{n+1}-V_{n+1}=U_n-V_n+h\left(f\left(U_n\right)-f\left(V_n\right)\right)+\mathcal{O}\left(h^2\right)$$

Whenever $f\left(\cdot\right)$ satisfies a Lipschitz condition with the Lipschitz constant denoted by L_f, the last equation implies successively

$$\|U_{n+1} - V_{n+1}\| \leq \|U_n - V_n\| \left(1 + hL_f\right),$$
$$\|U_{n+1} - V_{n+1}\| \leq \|U_1 - V_1\| \left(1 + hL_f\right)^n . \tag{6}$$

Consequently, assuming that $f(\cdot)$ is Lipschitz and (1) generates a dynamical system on $\mathbb{R}^p$, the inequality (6) ensure the existence of the local phase portrait of (1) near a critical point (see [9, Ch. 6: Convergence of Invariant Sets]).

A quite exhaustive analysis of the linear stability in the numerical solution of initial value problems is available in [3]. When the dynamical system is nonlinear the analysis becomes much more complicated.

With respect to the stability of the scheme **AB/** we have obtained the following estimation

$$\|U_{n+1} - U_n\| \leq hN\|f(U_n)\|, \quad n = 0, 1, 2, \cdots \tag{7}$$

where N is the sum of the Neumann series

$$N = \|I + h\theta df(U_n) + \cdots + h^k\theta^k \left(\mathrm{d}f(U_n)\right)^k + \cdots\|$$

whenever $h\theta$ satisfies (2). The estimation (7) could be quite useful for vector fields $f(\cdot)$ with some boundedness properties.

5. CONCLUDING REMARKS AND OPEN PROBLEMS

As it is apparent from the above analysis the most important results obtained—the existence of the dynamical system, the stability and the shadowing results—depend essentially on the inequality (2). Thus, whenever the Jacobian $\mathrm{d}f(U_n)$, $n = 1, 2, 3, \ldots$, is a normal matrix, its eigenvalues can be used in order to verify (2). Unfortunately, in the most interesting, and at the same time most important applications, $\mathrm{d}f(\cdot)$ is not a normal operator (matrix). In this situation, the spectrum of this operator could be misleading (see for example [1], [4], [11] or [12]). The pseudospectrum and some scalar measure of non-normality of square matrices, such the Henrici number [5], could have some relevance. However, the two parameters h and θ involved in the inequality (2) can be adjusted, in principle, such that this inequality is satisfied at each and every step.

A final remark about the terminology used in this paper. Finite difference schemes for initial value problems (1) are considered in the classical language in the well known monographs of Henrici [6], Richtmyer and Morton [8] or in the textbook of Süli and Meyers [10]. In contrast with this language is that of dynamical systems based essentially on semigroups of evolution. We have chosen the latter due to a multitude of advantages.

At the end of this paper we have to mention an important open problem. This refers to the specialization of the scheme **AB/**, and the subsequent results, for some specific vector fields $f(\cdot)$ such as gradient, dissipative and contractive vector fields.

REFERENCES

[1] Arveson, W., *A Short Course in Spectral Theory*, Springer, 2002.

[2] Conway, J. B., *A Course in Functional Analysis*, Springer, 1985.

[3] van Dorsselaer, J. L. M., Kraaijvanger, J. F. B. M. and Spijker, M. N., *Linear stability analysis in numerical solutions of initial value problems*, Acta Numerica, pp. 199–237, 1993.

[4] Embree, M. and Trefethen, L. N., *Pseudospectra Gateway,* http://www.comlab.ox.ac.uk/projects/pseudospectra2000.

[5] Gheorghiu, C. I., *On the measure of non-normality of matrices*, General Mathematics, ULB Sibiu, 2003.

[6] Henrici, P., *Discrete Variable Methods in Ordinary Differential Equations*, John Wiley & Sons, 1962.

[7] Kantorovich, L. V. and Akilov, G. P., *Functional Analysis*, Scientific Publishing House, Bucharest, Romania (translation from Russian).

[8] Richtmayer, R. D. and Morton, K. W., *Difference Methods for Initial-Value Problems*, John Wiley, New York, 1967.

[9] Stuart, A. M. and Humphries, A. R., *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, 1996.

[10] Süli, E. and Mayers, D., *Numerical Computation I and II*, Oxford University Computing Laboratory, Oxford, 1995.

[11] Trefethen, L. N., *Computation of Psedudospectra*, Acta Numerica, pp. 247–295, 1999.

[12] Trefethen, L. N. and Bau III, D., *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

Received by the editors: March 5, 2003.