stichting
mathematisch
centrum

MC

AFDELING NUMERIEKE WISKUNDE
(DEPARTMENT OF NUMERICAL MATHEMATICS)

NW 101/81 MAART

J.G. VERWER

ON THE PRACTICAL VALUE OF THE NOTION OF BN-STABILITY

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: 65L05

ACM Computer Review Categories: 5.17

On the practical value of the notion of BN-stability*)

by

J.G. Verwer

ABSTRACT

The oldest concept of unconditional stability of numerical integration methods for ordinary differential systems is that of A-stability. This concept is related to linear systems having constant coefficients and has been introduced by Dahlquist in 1963. More recently, since another contribution of Dahlquist in 1975, there has been much interest in unconditional stability properties of numerical integration methods when applied to non-linear dissipative systems (G-stability, BN-stability, A-contractivity). Various classes of implicit Runge-Kutta methods have already been shown to be BN-stable. However, contrary to the property of A-stability, when implementing such a method for practical use this unconditional stability property may be lost. The present note clarifies this for a class of diagonally implicit methods and shows at the same time that Rosenbrock's method is not BN-stable.

KEY WORDS & PHRASES: *Numerical analysis, Implicit Runge-Kutta methods, Stiff problems, Nonlinear stability*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

The s-stage implicit Runge-Kutta (IRK) method for numerically solving the initial value problem for systems of ordinary differential equations

$$y' = f(t,y), \quad f: \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N, \tag{1.1}$$

is given by

$$\begin{aligned} k_i &= \tau f(t_n + c_i\tau, y_n + \sum_{j=1}^{s} a_{ij}k_j), \quad i = 1,\ldots,s, \\ y_{n+1} &= y_n + \sum_{i=1}^{s} b_i k_i. \end{aligned} \tag{1.2}$$

The vector y_n denotes the numerical approximation to the exact solution $y = y(t)$ at $t = t_n$ and $\tau = t_{n+1} - t_n$ denotes the stepsize. We shall consider systems (1.1) which satisfy the dissipativity condition

$$\langle f(t,u) - f(t,v), u - v \rangle \le 0, \quad \text{all } t \in \mathbb{R}, \text{ all } u,v \in \mathbb{R}^N, \tag{1.3}$$

where $\langle \cdot,\cdot \rangle$ denotes an inner product on $\mathbb{R}^N$. A typical consequence of this inequality is that any two solutions, say y and $\tilde{y}$, behave contractive, i.e., $\|y(t_2) - \tilde{y}(t_2)\| \le \|y(t_1) - \tilde{y}(t_1)\|$ for all $t_2 \ge t_1$. Here $\|\cdot\|$ denotes the inner product norm. In what follows it is convenient to introduce C as the class of all (nonlinear) problems (1.1) satisfying (1.3).

Burrage and Butcher [3] now call the IRK method BN-stable if for all members from C it holds that

$$\|y_{n+1} - \tilde{y}_{n+1}\| \le \|y_n - \tilde{y}_n\|, \quad \text{all } t_n \in \mathbb{R}, \text{ all } y_n, \tilde{y}_n \in \mathbb{R}^N, \text{ all } \tau > 0. \tag{1.4}$$

Thus BN-stability means *unconditional* numerical contractivity on C. Since the first papers of Dahlquist and Butcher on this topic [2,5], contractivity has received a lot of attention in the numerical literature (see e.g. [4,6,8,9,11,16] and the references cited therein). In these papers one usually concentrates on the implicit formulation (1.2), that is, in the investigations one assumes that the approximations $\{y_n\}$ satisfy (1.2)

exactly. However, in actual computation (1.2) has to be combined with some sort of Newton process in order to solve the implicit relations. Because BN-stability was introduced for *nonlinear* stiff problems it makes sense to question whether this property can be preserved when *implementing* a BN-stable method for practical use. A closely related question is of course: given some BN-stable IRK method, is it then possible to solve the implicit relations numerically on the whole class C without restrictions on τ? It is to be feared, and for many certainly not surprising, that the answer to this question will be negative.

In the author's opinion these aspects deserve more attention in the literature in order to find out under what circumstances unconditional nonlinear stability can be realized in practice. To support this view the present note discusses a standard implementation of diagonally implicit (DIRK) methods having equal diagonal elements a_{ii}. Such methods are more easier to implement than fully implicit ones (see e.g. [1,12]). We show that this implementation is not unconditionally stable on a subclass of C consisting of *linear problems only*, i.e., the implementation is not BN-stable. An immediate consequence of this result is that Rosenbrock's method is not BN-stable. This fact has been pointed out earlier in [16]. A negative result on nonlinear stability of the most simple Rosenbrock scheme has already been given in 1968 by Sandberg and Shichman [15].

2. DIRK METHODS

The class of DIRK methods we consider is defined by

$$\begin{aligned} k_i &= \tau f\Big(t_n + c_i\tau, y_n + \sum_{j=1}^{i-1} a_{ij}k_j + ak_i\Big), \quad a > 0,\ i = 1,\ldots,s, \\ y_{n+1} &= y_n + \sum_{i=1}^{s} b_i k_i. \end{aligned} \tag{2.1}$$

Two examples of such BN-stable DIRK methods have been independently given in [3] and [4]; see also [7] and [16]:

(2.2) $$\begin{array}{c|c} c & A \\ \hline & b \end{array} = \begin{array}{c|cc} \lambda & \lambda & \\ 1-\lambda & 1-2\lambda & \lambda \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$ order is equal to 3, $\lambda = (3+\sqrt{3})/6$.

(2.3) $$\begin{array}{c|ccc} (1+\lambda)/2 & (1+\lambda)/2 & & \\ \frac{1}{2} & -\lambda/2 & (1+\lambda)/2 & \\ (1-\lambda)/2 & 1+\lambda & -1-2\lambda & (1+\lambda)/2 \\ \hline & 1/(6\lambda^2) & 1-1/(3\lambda^2) & 1/(6\lambda^2) \end{array}$$ order is equal to 4, $\lambda = \frac{2}{\sqrt{3}} \cos \frac{\pi}{18}$.

Because $a_{ij} = 0$ for $j > i$, the increment vectors k_i can be computed one after another. In actual computation this is done by means of the modified Newton process. Exploiting the equality $a_{ii} = a$ for all i, the usual approach in this solution process is to spend *at most* one $\partial f/\partial y$-evaluation, plus corresponding LU-decomposition, per integration step (see [1,12] for details). In practice one often integrates with a fixed Jacobian matrix over several steps. We now assume that we perform *precisely one evaluation of* $J(t,y) = \partial f(t,y)/\partial y$ *per step* at the point $(t,y) = (t_n+c_1\tau, y_n)$, and that per stage s_i modified Newton iterations are performed. We thus proceed with the *implemented DIRK method*

$$\begin{aligned} &(I - \tau a J_n)k_i^{(m)} = -\tau a J_n k_i^{(m-1)} + \\ &\tau f(t_n+c_i\tau, y_n + \sum_{j=1}^{i-1} a_{ij}k_j^{(s_j)} + a k_i^{(m-1)}), \quad i = 1,\ldots,s \text{ and } m = 1,\ldots,s_i, \\ &y_{n+1} = y_n + \sum_{i=1}^{s} b_i k_i^{(s_i)}, \end{aligned} \tag{2.4}$$

where $J_n = J(t_n+c_1\tau, y_n)$ and the starting vectors $k_i^{(0)}$ are still free to choose. In the next section we shall show that (2.4) is not unconditionally contractive on a subclass of C consisting of linear problems only, i.e., implementation (2.4) is not BN-stable.

In this connection it is worth noting that for the linear problem

$$y' = Jy + r(t), \quad J \text{ constant}, \tag{2.5}$$

schemes (2.1) and (2.4) are identical. This implies that implementation

(2.4) is A-stable if the implicit scheme itself is A-stable.

Observe that when we have to deal with unequal diagonal elements a_{ii}, s say, we still have to perform s matrix factorizations. This is the reason why in practice the equality $a_{ii} = a$ is preferred. Finally, for the sake of completeness, note that in a computer implementation the matrix expression $(1-z)^{-1}z$ is replaced by $(1-z)^{-1}-1$. Herewith one saves a matrix vector operation.

3. THE CLASS *SLC*

We define *SLC* as the class of all linear problems

$$y' = \varepsilon^{-1}F(t,\varepsilon)y, \quad \varepsilon \in (0,\varepsilon_0], \ \varepsilon_0 \text{ some constant}, \tag{3.1}$$

where, for all $t \in \mathbb{R}$ and $\varepsilon \in (0,\varepsilon_0]$, $F(t,\varepsilon)$ is a symmetric nonpositive definite N×N matrix. Furthermore, it is assumed that F is continuous and bounded on $\mathbb{R} \times (0,\varepsilon_0]$. Then it is immediate that, given a point $(t_0,y_0) \in \mathbb{R} \times \mathbb{R}^N$, for all $\varepsilon \in (0,\varepsilon_0]$ equation (3.1) has a unique solution $y(t,\varepsilon)$ on $\mathbb{R}$ such that $y(t_0,\varepsilon) = y_0$.

Because $F(t,\varepsilon)$ is symmetric nonpositive definite, any two solutions behave contractive, i.e., *SLC* $\subset$ *C*. By counterexample we shall show that the implemented DIRK method (2.4) is not unconditionally contractive on *SLC*, i.e., not BN-stable. Observe that class *SLC* bears a close resemblance to the problem class studied by Van Veldhuizen [17] (see also [18]). Also note that on *SLC* the increment vectors k_i defined by (2.1) always exist ($a > 0$).

When applied to (3.1), method (2.4) yields

$$\begin{aligned} k_i^{(m)} &= (P+Q_i)k_i^{(m-1)} + a^{-1}Q_i\Big(y_n + \sum_{j=1}^{i-1} a_{ij}k_j^{(s_j)}\Big), \quad i = 1,\ldots,s \text{ and} \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad m = 1,\ldots,s_i, \\ y_{n+1} &= y_n + \sum_{i=1}^{s} b_i k_i^{(s_i)}, \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} P &= -(I-\tau a\varepsilon^{-1}F(t_n+c_1\tau,\varepsilon))^{-1}\tau a\varepsilon^{-1}F(t_n+c_1\tau,\varepsilon), \\ Q_i &= (I-\tau a\varepsilon^{-1}F(t_n+c_1\tau,\varepsilon))^{-1}\tau a\varepsilon^{-1}F(t_n+c_i\tau,\varepsilon). \end{aligned} \tag{3.3}$$

Note that $P = -Q_1$. This implies that $k_1^{(m)} = -a^{-1}Py_n$ for all m, i.e., the vector k_1 defined by (2.1) is obtained after precisely one Newton iteration, which is not true for the remaining increment vectors unless F is constant.

Equation (3.2) can be abbreviated to the form

$$y_{n+1} = G(t_n,\tau,\varepsilon)y_n, \tag{3.4}$$

G being a complicated rational matrix expression. The idea is now to construct examples from class SLC for which (cf. [17,18])

$$\|G(t_n,\tau,\varepsilon)\| \to \infty \quad \text{as} \quad \varepsilon \to 0 \quad \text{for any fixed } \tau. \tag{3.5}$$

If this occurs, the contractivity condition (1.4) is violated. Using the terminology of Van Veldhuizen, we can also say that method (2.4) is not $D(SLC)$-stable. As shown below it is not difficult to find appropriate examples:

COUNTEREXAMPLE 1. Definition (3.1) does not exclude matrices $F(t,\varepsilon)$ which may become zero for some finite number of t-values. Now suppose that this is the case for $F(t_n+c_1\tau,\varepsilon)$, all $\varepsilon \in (0,\varepsilon_0]$. Further suppose that $s \geq 2$ and that for at least one integer $i \in \{2,\ldots,s\}$ we have that $F(t_n+c_i\tau,\varepsilon) \neq F(t_n+c_1\tau,\varepsilon)$, all $\varepsilon \in (0,\varepsilon_0]$. For such a stage the modified Newton process then degenerates to the simple Jacobi iteration process. This implies that $\|k_i^{(j)}\| \to \infty$ if $\varepsilon \to 0$, irrespective the number of iterations, i.e., it is always possible to let $\|G(t_n,\tau,\varepsilon)\| \to \infty$ as $\varepsilon \to 0$. The possibility of zero partial derivatives has also been pointed out by Vanselow [16] in an investigation on Rosenbrock methods. ■

COUNTEREXAMPLE 2. The previous example suggests to remove all members from C and SLC whose partial derivatives $\partial f/\partial y$ take zero values. However, it is then still possible to find a counterexample to BN-stability of implementation (2.4). Let us again consider scheme (3.2). It is immediate that, for every $\tau > 0$, P is bounded in $\varepsilon \in (0,\varepsilon_0]$ on the whole class SLC. Unfortunately, this need not to be true for the matrices Q_i, even if we exclude zero partial derivatives. To see this consider the problem

$$y' = E(t)\,\mathrm{diag}(d_1(t),\varepsilon^{-1}d_2(t))E^T(t)y, \tag{3.6}$$

where

$$E(t) = \begin{bmatrix} \cos \nu\tau & -\sin \nu t \\ \sin \nu t & \cos \nu t \end{bmatrix}, \ \nu \text{ constant}, \tag{3.7}$$

and where $d_1,d_2\colon \mathbb{R} \to \mathbb{R}$ are strictly negative, bounded and continuous on $\mathbb{R}$. This example was taken from Kreiss [10] and has also been studied in [14, 18,19]. It certainly belongs to *SLC*. We now select a stage for which $c_i \neq c_1$ and take, for example, $t_n+c_1\tau = 0$. The elements of the 2×2 matrix Q_i are then given by

$$\begin{aligned}
(Q_i)_{1,1} &= \tau a[d_1(t)\cos^2(\nu t)+\varepsilon^{-1}d_2(t)\sin^2(\nu t)]/(1-\tau a d_1(0)),\\
(Q_i)_{1,2} &= \tau a[(d_1(t)-\varepsilon^{-1}d_2(t))\sin(\nu t)\cos(\nu t)]/(1-\tau a d_1(0)),\\
(Q_i)_{2,1} &= \tau a[(d_1(t)-\varepsilon^{-1}d_2(t))\sin(\nu t)\cos(\nu t)]/(1-\tau a\varepsilon^{-1} d_2(0)),\\
(Q_i)_{2,2} &= \tau a[d_1(t)\sin^2(\nu t)+\varepsilon^{-1}d_2(t)\cos^2(\nu t)]/(1-\tau a\varepsilon^{-1} d_2(0)),
\end{aligned} \tag{3.8}$$

where $t = t_n+c_i\tau$. We see that, by an appropriate choice of ν, the first-row elements are bot bounded in ε. If we take s_i, the number of modified Newton iterations, finite, the increment vector $k_i^{(s_i)}(\varepsilon)$ is then also unbounded in ε. This means that it is possible to let $\|G(t_n,\tau,\varepsilon)\| \to \infty$ as $\varepsilon \to 0$ for any fixed τ. Now suppose that we do not prescribe the number of modified Newton iterations beforehand; then we have to show that it is possible to let $\|k_i^{(m)}\| \to \infty$ as $m \to \infty$ for at least one value of ε and τ, i.e., divergence to infinity. Divergence to infinity appears if the spectral radius of $P + Q_i$ is larger than 1. Because the first-row elements of $P + Q_i$ are unbounded in $\varepsilon \in (0,\varepsilon_0]$, contrary to the second-row elements, we can make this spectral radius as large as we wish. This observation completes our second counter-example. ■

If we set $s_i = 1$ and $k_i^{(0)} = 0$ (an appropriate initial vector) in (2.4), there results

$$
\begin{aligned}
&(I-\tau a J_n)k_i^{(1)} = \tau f(t_n+c_i\tau, y_n+\sum_{j=1}^{i-1} a_{ij}k_j^{(1)}), \quad i = 1,\ldots,s,\\
&y_{n+1} = y_n + \sum_{i=1}^{s} b_i k_i^{(1)}.
\end{aligned}
\tag{3.9}
$$

In literature this scheme is called a Rosenbrock scheme. It differs from the original one [13] in two ways. That scheme has been defined for autonomous problems and allows, in principle, a new Jacobian evaluation at each stage while also the parameter a may depend on the index i (for reasons of computational overhead this possibility is ruled out in practice).
A counterexample to B-stability of such a scheme can be constructed by means of a one dimensional non-increasing function f(y) [16]. Such a function is dissipative. The idea is to select s y-values such that f_y becomes zero. The scheme then degenerates to an explicit RK scheme (see also [15] for another instructive example).

4. FINAL REMARKS

Nonlinear stability has become an important subject in the literature on numerical methods for stiff nonlinear problems. Various classes of IRK methods have already been shown to be BN-stable, e.g. the DIRK-methods (2.2) and (2.3). A difficulty arises when implementing a BN-stable method on a computer. Contrary to the property of A-stability, the property of BN-stability can in general not be proven for an implemented IRK method. This makes it difficult to decide whether in practice a BN-stable scheme should be implemented instead of an A-stable one. In order to exploit the nice theoretical results on unconditional nonlinear stability for practical use, these aspects should therefore be given more attention.

In the present paper we explained the situation for an interesting class of DIRK methods. We still wish to emphasize that, for a given BN-stable DIRK method, the implementation (2.4) can be modified to become unconditionally stable on *SLC* by allowing a new Jacobian evaluation per stage. On *SLC* the implementation then can be made identical to the corresponding implicit scheme. However, in view of the large computational overhead, we do consider such a modification as being not realistic. Therefore it makes sense to study implementation (2.4) when applied to class *SLC*. Finally, if one still wishes to consider implementations allowing a new

Jacobian per stage, or even a new Jacobian per Newton iteration, the one-dimensional dissipative functions discussed in [15,16] can probably be used to construct counterexamples to BN-stability.

ACKNOWLEDGEMENT. The author wishes to acknowledge Paul Wolkenfelt for his comments and remarks on a first draft of the paper.

REFERENCES

[1] ALEXANDER, R., *Diagonally implicit Runge-Kutta methods for stiff ODEs*, SIAM J. Numer. Anal. 14 (1977), 1006-1021.

[2] BUTCHER, J.C., *A stability property of implicit Runge-Kutta methods*, BIT 15 (1975), 358-361.

[3] BURRAGE, K. & J.C. BUTCHER, *Stability criteria for implicit Runge-Kutta methods*, SIAM J. Numer. Anal. 16 (1979), 46-57.

[4] CROUZEIX, M., *Sur la B-stabilité des méthodes de Runge-Kutta*, Numer. Math. 32 (1979), 75-82.

[5] DAHLQUIST, G., *Error analysis for a class of methods for stiff nonlinear initial value problems*, Dundee Conference on Numerical Analysis 1975, in: Lecture Notes in Math. 506 (1976), 60-74.

[6] DAHLQUIST, G. & R. JELTSCH, *Generalized disks of contractivity for explicit and implicit Runge-Kutta methods*, Report TRITA-NA-7906, Royal Institute of Technology, Stockholm, 1979.

[7] HAIRER, E., *Highest possible order of algebraically stable diagonally implicit Runge-Kutta methods*, BIT 20 (1980), 254-256.

[8] HAIRER, E. & G. WANNER, *Algebraically stable and implementable Runge-Kutta methods of high order*, SIAM J. Numer. Anal., to appear.

[9] HUNDSDORFER, W.H. & M.N. SPIJKER, *A note on B-stability of Runge-Kutta methods*, Numer. Math., to appear.

[10] KREISS, H.O., *Difference methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. 15 (1978), 21-58.

[11] NEVANLINNA, O. & W. LINIGER, *Contractive methods for stiff differential equations*, BIT 18 (1978), 457-474 and BIT 19 (1979), 53-72.

[12] NØRSETT, S.P., *Semi explicit Runge Kutta methods*, Mathematics and Computation No 6/74, Dept. of Mathematics, University of Trondheim, 1974.

[13] ROSENBROCK, H.H., *Some general implicit processes for the numerical solution of differential equations*, Computer J. 5 (1963), 329-330.

[14] SAND, J., *A note on a differential system constructed by H.O. Kreiss*, Report TRITA-NA-8004, Royal Institute of Technology, Stockholm, 1980.

[15] SANDBERG, I.W. & H. SHICHMAN, *Numerical integration of systems of stiff nonlinear differential equations*, The Bell System Technical Journal 47 (1968), 511-527.

[16] VANSELOW, R., *Stabilität und Fehleruntersuchungen bei numerischen Verfahren zur Lösung steifer nichtlinearer Anfangswertprobleme*, Diplomarbeit Sektion Mathematik TU-Dresden, april 1979.

[17] VELDHUIZEN, M. VAN, *D-stability*, SIAM J. Numer. Anal., to appear.

[18] VERWER, J.G., *An analysis of Rosenbrock methods for nonlinear stiff initial value problems*, SIAM J. Numer. Anal., to appear.

[19] VERWER, J.G., *Instructive experiments with some Runge-Kutta-Rosenbrock methods*, Report NW 100/81, Mathematical Centre, Amsterdam (prepublication) 1981.