

# Computational Ordinary Differential Equations

Based on the proceedings of a conference on Computational Ordinary Differential Equations, organized by The Institute of Mathematics and its Applications and held at Imperial College of Science and Technology, University of London, in July 1989

Edited by

J. R. CASH

*Department of Mathematics*

*Imperial College of Science, Technology, and Medicine, London, UK*

and

I. GLADWELL

*Department of Mathematics*

*Southern Methodist University, Dallas, USA*

CLARENDON PRESS · OXFORD · 1992

im Norm

onicity of Polynomials Oc-  
al Value Problems, *Numer.*

J., Spijker, M.N., Stepsize  
al Solution of Ordinary and  
*ut. Appl. Math.*, **20**, 67-81,

Algebraic Stability and Error  
*Appl. Numer. Math.*, **5**, 71-

y of Runge-Kutta Methods,

and *Differential Equations in*  
ork, 1976.

inite *Difference Method in Par-*  
sons, Chichester, 1980.

rie für Allgemeine Operatoren  
*., 4*, 258-281, 1951.

se-Kutta Methods, In: *Numer-*  
*l Value Problems, Proceedings,*  
uist, G., Jeltsch, R., editors, In-  
Mathematik der RWTH Aachen,

umerical Solution of Initial Value  
90, 1983.

for Stability of One-Step Methods  
d Value Problems, *Math. Comp.*,

# Stability Criteria in the Numerical Solution of Initial Value Problems

M.N. Spijker

University of Leiden

## 1 Introduction

1.1. Consider the system of ordinary differential equations

$$U'(t) = f(t, U(t)) \quad (1.1)$$

with an initial condition  $U(0) = u_0$ . Here  $u_0$  is a given vector in the  $s$ -dimensional real vector space  $\mathbb{R}^s$  while  $U(t) \in \mathbb{R}^s$  is unknown for  $t > 0$ . In all of the following  $f(t, \xi)$  is assumed to be a given, continuous function from  $\mathbb{R} \times \mathbb{R}^s$  to  $\mathbb{R}^s$  with  $s \geq 1$ .

We will be concerned with numerical processes which yield approximations  $u_n \in \mathbb{R}^s$  to the true solution  $U(t_n)$  at the *grid points*  $t_n = nh$ . Here  $h > 0$  denotes the *step size* of the process. We assume  $u_n$  to be computed from  $u_{n-1}$ , without using the former approximations  $u_{n-2}, u_{n-3}, \dots$

We will deal with the *stability* analysis of the numerical process. Here the term *stability* is used to indicate that *any numerical errors introduced at some stage of the calculations are propagated in a mild fashion* in the subsequent steps of the process.

1.2. Much of the stability analysis in the numerical solution of equation (1.1) consists of studying the numerical method when applied to a very simple, linear, scalar *test equation*. In justification of focussing on such a simple test equation one can appeal to *two principles*.

The *first principle* underlies the replacement of (1.1) by the linear system

$$U'(t) = AU(t). \quad (1.2)$$

Here  $f$  is assumed to be differentiable, and  $A$  stands for the Jacobian matrix of  $f$  "frozen" at any point  $t^*$  belonging to the time-interval under consideration, i.e.

$$A = \frac{\partial}{\partial \xi} f(t^*, U(t^*)).$$

The first principle asserts that stability in the numerical solution of all linear equations (1.2) under consideration guarantees a stable behaviour of the original numerical process for (1.1).

The *second principle* is applied when one replaces (1.2) by the still simpler equation

$$U'(t) = \lambda U(t). \quad (1.3)$$

Here  $\lambda$  stands for any eigenvalue belonging to the spectrum  $\sigma[A]$  of  $A$ . The second principle asserts that stability of the method in the solution of all relevant equations (1.3) guarantees stability in the solution of (1.2).

Clearly, a combination of these two principles provides a justification for basing the stability analysis of numerical methods for (1.1) on the test equation (1.3). See e.g. [1], [7], [8].

In this paper we will review some situations where the above principles break down, and other principles have to be used.

**1.3.** Suppose that the numerical calculations for approximating the solution to (1.1) are performed using a slightly perturbed starting vector  $\tilde{u}_0$  say, instead of  $u_0$ . We then would obtain new approximations, denoted by  $\tilde{u}_n$ .

In a stability analysis of the numerical process for (1.1) the crucial question is whether the difference  $\tilde{u}_n - u_n$  can be bounded suitably in terms of the perturbation  $\tilde{u}_0 - u_0$ . Therefore, one is looking for estimates of the type

$$|\tilde{u}_n - u_n| \leq \gamma n^q |\tilde{u}_0 - u_0| \quad (\text{whenever } n \geq 1, \text{ and } \{u_n\}, \{\tilde{u}_n\} \text{ are generated by the same numerical process}). \quad (1.4)$$

Here  $|\cdot|$  is a given norm in  $\mathbb{R}^s$  and  $\gamma \geq 0$ ,  $q \geq 0$  are independent of  $n$ ,  $\{u_n\}$ ,  $\{\tilde{u}_n\}$ . Clearly, the estimate (1.4) amounts to *stability* if both  $\gamma$  and  $q$  are of moderate size. Property (1.4) with  $q = 0$ ,  $\gamma = 1$  is called *contractivity*.

## 2 Failing of the first principle: Runge-Kutta methods

**2.1.** As an illustration consider the method

$$u_n = u_{n-1} + \frac{h}{2}[f(t_{n-1}, u_{n-1}) + f(t_n, u_n)], \quad n = 1, 2, 3, \dots$$

in the numerical solution of the problem

$$U'(t) = -1 + \exp[10^5(1 - 3U(t))], \quad U(0) = 1/3.$$

The true solution is  $U(t) \equiv 1/3$ .

We apply the method with  $h = 1/3$ . But, starting with  $u_1 = 1/3$ . But, starting

Clearly, (1.4) does not hold. The derivative with respect to  $t$  is

$$f(t, \xi)$$

satisfies

Further, the method under consideration is not contractive in the solution of the first principle of section 1.3. A related example was given in [1].

**2.2.** The above example shows that the stability theory relevant to numerical methods is not such a theory for general  $R$ .

$$u_n = u_{n-1} + \dots$$

where  $x_j \in \mathbb{R}^s$  are computed by

$$x_i = hf(t_{n-1} + c_i h, u_{n-1}) \quad \text{for } i = 1, 2, \dots, m.$$

The coefficients  $a_{ij}$ ,  $b_i$  are given by  $a_{i1} + a_{i2} + \dots + a_{im} = 1$ .

The following definition is due to Crouzeix [4]: Method (2.1) is called *contractive* if

$$\begin{cases} \text{all } b_i \geq 0, \text{ and} \\ Q_{ij} = b_i a_{ij} + \dots \end{cases}$$

Similarly to [3], [4], we restrict the Euclidean norm

$$|\xi| = |\xi|_2$$

and to differential equation

$$\begin{cases} |\tilde{U}(t) - U(t)| \leq \dots \\ \text{whenever } \dots \end{cases}$$

We apply the method with  $h = 0.1$ . Starting with  $u_0 = 1/3$  the method yields  $u_1 = 1/3$ . But, starting from  $\tilde{u}_0 = 0.333$  the method yields

$$\tilde{u}_1 \simeq 10^{42}.$$

Clearly, (1.4) does not hold here with any  $\gamma$  of moderate size.

The derivative with respect to  $\xi$  of the function

$$f(t, \xi) = -1 + \exp[10^5(1 - 3\xi)]$$

satisfies

$$\frac{\partial}{\partial \xi} f(t, \xi) \leq 0.$$

Further, the method under consideration is  $A$ -stable, which implies contractivity in the solution of (1.2) and (1.3) with  $A = \lambda \leq 0$ . Therefore, the first principle of section 1.2 is misleading for the situation at hand. A related example was given by Wanner [20].

**2.2.** The above example shows that there is some need for a reliable stability theory relevant to nonlinear differential equations. We now review such a theory for general *Runge-Kutta methods* of the following type.

$$u_n = u_{n-1} + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \tag{2.1a}$$

where  $x_j \in \mathbb{R}^s$  are computed from  $u_{n-1}$  in such a way that

$$\begin{aligned} x_i &= hf(t_{n-1} + c_i h, u_{n-1} + a_{i1} x_1 + a_{i2} x_2 + \dots + a_{im} x_m) \\ &\text{for } i = 1, 2, \dots, m. \end{aligned} \tag{2.1b}$$

The coefficients  $a_{ij}, b_i$  are real parameters defining the method, and  $c_i = a_{i1} + a_{i2} + \dots + a_{im}$ .

The following definition originated with Butcher, Burrage [3] and Crouzeix [4]: Method (2.1) is called *algebraically stable* if

$$\begin{cases} \text{all } b_i \geq 0, \text{ and the } m \times m \text{ matrix } Q = (Q_{ij}) \text{ with} \\ Q_{ij} = b_i a_{ij} + b_j a_{ji} - b_i b_j \text{ is positive semi-definite.} \end{cases}$$

Similarly to [3], [4], we restrict ourselves in the sequel, up to section 3, to the Euclidean norm

$$|\xi| = |\xi|_2 = \sqrt{\xi^T \xi} \quad \text{for } \xi \in \mathbb{R}^s, \tag{2.2a}$$

and to differential equations (1.1) such that

$$\begin{cases} |\tilde{U}(t) - U(t)| \leq |\tilde{U}(t-h) - U(t-h)| \\ \text{whenever } h > 0 \text{ and } \tilde{U}, U \text{ satisfy (1.1).} \end{cases} \tag{2.2b}$$

Consider the property

There is contractivity whenever the method is applied with  
any  $h > 0$  to any equation (1.1) satisfying (2.2). (2.3)

**Theorem 1.** Algebraic stability is a necessary and sufficient condition on the method to guarantee property (2.3).

This theorem is valid provided method (2.1) is irreducible (cf. [1], [9], [17]).

**2.3.** Clearly, algebraic stability is a favourable property. However, within the class of Runge-Kutta methods that fail to be algebraically stable, there are methods with other favourable properties, e.g. requiring only little computational labour for computing  $u_n$  from  $u_{n-1}$ . Therefore, it seems worth looking closer at the methods which are  $A$ -stable without being algebraically stable. In particular, consider the following question: Are there methods failing to be algebraically stable and still stable [(1.4), with moderately sized  $\gamma, q$ ] for all  $h > 0$  and for all equations (1.1) satisfying (2.2)?

The method in the example of Section 2.1 corresponds to a Runge-Kutta method with  $m = 2$ ,  $a_{11} = a_{12} = 0$ ,  $a_{21} = a_{22} = b_1 = b_2 = 1/2$ . This method fails to be algebraically stable. The differential equation in Section 2.1 belongs to the class of equations under consideration, defined by requirement (2.2). Therefore, the disastrous error propagation in Section 2.1 suggests that the answer to the above question is negative.

However, the following counterexample, due to Kraaijevanger [11], shows that the answer is positive. Let  $m = 2$ , and consider method (2.1) with

$$\begin{aligned} a_{11} &= 1/2, & a_{12} &= 0 \\ a_{21} &= -1/2, & a_{22} &= 2 \\ b_1 &= -1/2, & b_2 &= 3/2. \end{aligned}$$

Since  $b_1 < 0$ , this method is *not algebraically stable*. Still,

(1.4) holds with  $\gamma = 2$ ,  $q = 0$  whenever this method is applied with any  $h > 0$  to any equation (1.1) satisfying (2.2).

This example thus suggests that algebraic stability is an unnecessarily strong demand upon a method if one is not looking for contractivity but only for stability in the sense (1.4) with moderate  $\gamma, q$ .

**2.4.** Up to now we have been dealing with a constant stepsize  $h > 0$ . In actual calculations one will often use *variable stepsizes*  $h_n > 0$  and non-equidistant gridpoints  $t_n = t_{n-1} + h_n$ . In this situation it is natural to consider the following stability property of a Runge-Kutta method:

There are fixed  $\gamma, q$  such that (1.4) holds whenever the method is applied with any variable  $h_n \in (0, 1]$  to any equation (1.1) satisfying (2.2). (2.4)

**Theorem 2.** Algebraic stability of the method to guarantee the

Thus, after all, algebraic stability upon a Runge-Kutta method stability property (2.4). For

**2.5.** The stability estimates of algebraic stability, are quite highly nonlinear differential equation 1 may break down. However, rather strong and not satisfactory e.g. [1]. Therefore it is interesting to see if stability estimates exist as well for some mildly nonlinear differential equation 1 applies. The estimates in equations (1.1) satisfying (2.4)  $A(\theta)$ -stable methods ( $0 < \theta < 1$ ) the (classical) numerical stability wedge in the complex plane. The actual scope of the present

### 3 Failure of the second principle

**3.1.** In the above we dealt with algebraic stability for a constant stepsize  $h_n$ . From now on we consider the problem (1.2) with a constant stepsize  $h$ . We apply the Runge-Kutta method (2.1) to equation

$$u_n = \varphi(h)$$

Here  $\varphi(\zeta) = \frac{P(\zeta)}{Q(\zeta)}$  is a rational function. We use the notation

$$|\varphi(h)|$$

Note that Rosenbrock method applied to (1.2), also reduce to the second principle of

The second principle of

$$|\varphi(h)|$$

is sufficient to guarantee a stability region

$$S = \{\zeta \in \mathbb{C} : |\varphi(\zeta)| \leq 1\}$$

**Theorem 2.** Algebraic stability is a necessary and sufficient condition on the method to guarantee that (2.4) holds.

Thus, after all, algebraic stability turns out to be the appropriate demand upon a Runge-Kutta method — if one is looking for the variable stepsize stability property (2.4). For the proof of Theorem 2 we refer to [11].

**2.5.** The stability estimates in (2.3), (2.4), present under the assumption of algebraic stability, are quite satisfactory in that they are still valid for highly nonlinear differential equations, for which the first principle of Section 1 may break down. However, the condition of algebraic stability is rather strong and not satisfied by various methods of practical interest (cf. e.g. [1]). Therefore it is important to note that rigorous stability estimates exist as well for some methods which fail to be algebraically stable. Such estimates were derived e.g. by Hundsdorfer [8] and Schmitt [16] for mildly nonlinear differential equations, where the first principle of Section 1 applies. The estimates in [8] cover some  $A$ -stable methods applied to equations (1.1) satisfying (2.2). The estimates in [16] are relevant to some  $A(\theta)$ -stable methods ( $0 < \theta < \pi/2$ ) applied to differential equations where the (classical) numerical range of the Jacobian matrix lies within an appropriate wedge in the complex plane. Since these results lie outside the actual scope of the present paper we refer to [16] for more details.

### 3 Failure of the second principle: general methods

**3.1.** In the above we dealt with highly nonlinear equations (1.1) and variable  $h_n$ . From now on we consider the numerical solution of the linear problem (1.2) with a constant stepsize  $h > 0$ . An application of the Runge-Kutta method (2.1) to equation (1.2) yields a numerical process of the type

$$u_n = \varphi(hA)u_{n-1}, \quad n = 1, 2, 3, \dots \tag{3.1}$$

Here  $\varphi(\zeta) = \frac{P(\zeta)}{Q(\zeta)}$  is a rational function with  $\varphi(0) = \varphi'(0) = 1$ , and we use the notation

$$\varphi(hA) = P(hA)[Q(hA)]^{-1}.$$

Note that Rosenbrock methods and higher derivative methods, when applied to (1.2), also reduce to numerical processes of type (3.1).

The second principle of Section 2.1 asserts that the condition

$$|\varphi(h\lambda)| < 1, \quad \text{for all } \lambda \in \sigma[A]$$

is sufficient to guarantee a stable behaviour of process (3.1). Defining the *stability region*

$$S = \{\zeta : \zeta \in \mathbb{C} \text{ with } |\varphi(\zeta)| < 1\}$$



(cf. [12]).

**3.2.** The unreliable condition (3.2) can be converted into a reliable one by replacing  $\sigma[A]$  by some appropriate larger set  $\tau[A] \subset \mathbb{C}$ , i.e.

$$h\tau[A] \subset S. \tag{3.2^*}$$

Satisfactory stability estimates, with various choices for  $\tau[A]$ , have been derived in [2], [5], [10], [12], [13], [15], [18], [19].

We now review one of the stability results in [12]. Let the Gerschgorin disk  $D_j$  of  $A = (\alpha_{jk})$  be defined by

$$D_j = \left\{ \zeta : |\zeta - \alpha_{jj}| \leq \sum_{k \neq j} |\alpha_{jk}| \right\}.$$

Define  $\tau[A]$  to be the convex hull of the union of all Gerschgorin disks, i.e.

$$\tau[A] = \text{conv}(D_1 \cup D_2 \cup \dots \cup D_s).$$

**Theorem 3.** With the above  $\tau[A]$ , condition (3.2\*) implies estimate (1.4) with  $q = 1$ , and  $\gamma$  depending on  $\varphi$  and  $h\tau[A]$  only.

Here  $\gamma$  does *not* depend in any transparent manner on the eigenvectors of  $A$ . In fact  $\gamma = \Gamma(h\tau[A])$  with a fixed function  $\Gamma$  satisfying

$$\Gamma(V_1) \leq \Gamma(V_2) \text{ for } V_1 \subset V_2 \subset S.$$

Moreover,  $\max\{\text{Re } z : z \in \tau[A]\}$  is precisely the logarithmic norm of  $A$  with respect to the maximum norm.

We illustrate the theorem with the same  $A$  as above and with  $h$  satisfying (3.2\*). A straightforward calculation shows that  $\tau[A]$  is contained in the union  $T_1 \cup T_2$  of the triangle

$$T_1 = \left\{ \zeta : |\pi - \arg \zeta| \leq \frac{\pi}{6} \text{ and } -\frac{3(2s+1)}{4} \leq \text{Re } \zeta < 0 \right\}$$

and the disk

$$T_2 = \{ \zeta : |\zeta + 2s + 1| \leq s + 1/2 \}.$$

Therefore (3.2\*) holds when  $h$  is so small that

$$h \cdot (T_1 \cup T_2) \subset S.$$

The latter inclusion is valid for all  $h$  with

$$h \leq h_0 = \frac{9}{4(2s+1)} \simeq 0.02777.$$

With the choice  $h = 0.027$  the theorem thus predicts stability in the sense (1.4) with  $q = 1$ , and  $\gamma$  only depending on  $\varphi$  and  $h \cdot (T_1 + T_2)$ .

Straightforward numerical calculations show that, with this  $h$ , one actually has

$$|\tilde{u}_n - u_n| \leq 1.2 |\tilde{u}_0 - u_0| \quad (\text{for all } n \geq 1; u_0, \tilde{u}_0 \in \mathbb{R}^s).$$

Hence, there is now a very mild error propagation – even better than predicted by theorem 3. We refer to [12] for more details.

## References

- [1] Butcher, J.C.: "The numerical analysis of ordinary differential equations, Runge-Kutta and general linear methods". Chichester: John Wiley (1987).
- [2] Brenner, P., Thomée, V.: On rational approximations of semigroups, *SIAM J. Numer. Anal.* **16**, 683-694 (1979).
- [3] Burrage, K., Butcher, J.C.: Stability criteria for implicit Runge-Kutta methods, *SIAM J. Numer. Anal.* **16**, 46-57(1979).
- [4] Crouzeix, M.: Sur la B-stabilité des méthodes de Runge-Kutta, *Numer. Math.* **32**, 75-82(1979).
- [5] Crouzeix, M.: On multistep approximation of semigroups in Banach spaces, Université de Rennes, IRISA, Publ.Interne no. 310 (1986).
- [6] Griffiths, D.F., Christie I., Mitchell, A.R.: Analysis of error growth for explicit difference schemes in conduction-convection problems, *Int. J. Numer. Meth. Engin.* **15**, 1075-1081 (1980).
- [7] Houwen, P.J. van der: "Construction of integration formulas for initial value problems". Amsterdam, New York, Oxford: North-Holland Publ. Comp. (1977).
- [8] Hundsdorfer, W.H.: "The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods". CWI Tract 12. Amsterdam: Centre for Mathematics and Computer Science (1985).
- [9] Hundsdorfer, W.H., Spijker, M.N.: A note on B-stability of Runge-Kutta methods, *Numer. Math.* **36**, 319-331 (1981).
- [10] Kraaijevanger, J.F.B.M., Lenferink, H.W.J., Spijker, M.N.: Stepsize restrictions for stability in the numerical solution of ordinary and partial differential equations, *J. Comp. Appl. Math.* **20**, 67-81 (1987).
- [11] Kraaijevanger, J.F.B.M. propagation in Runge-71-87(1989).
- [12] Lenferink, H.W.J., Spijker, M.N.: numerical analysis of ordinary differential equations, *J. Comp. Appl. Math.* **237** (1991).
- [13] Nevanlinna, O.: Remanent groups, Helsinki Univ. A225 (1984).
- [14] Parter, S.V.: Stability of difference equations for stiff problems, *J. Comp. Appl. Math.* **4**, 277-292 (1962).
- [15] Roux, M.N. le: Semidiscrete stability, *J. Comp. Appl. Math.* **33**, 919-931 (1991).
- [16] Schmitt, B.A.: Stability of linear stiff differential equations, *J. Comp. Appl. Math.* **4**, 277-292 (1962).
- [17] Spijker, M.N.: Feasibility of Runge-Kutta methods, *J. Comp. Appl. Math.* **4**, 277-292 (1962).
- [18] Spijker, M.N.: Stepsize restrictions for stability in the numerical solution of stiff ordinary differential equations, *J. Comp. Appl. Math.* **20**, 67-81 (1987).
- [19] Trefethen, L.N.: Lax stability, Department of Mathematics, University of Toronto, Numer. Anal. Report 1987-1.
- [20] Wanner, G.: A short proof of the stability of Runge-Kutta methods, *J. Comp. Appl. Math.* **20**, 67-81 (1987).

dicts stability in the sense  
 $\|h \cdot (T_1 + T_2)\|$   
 that, with this  $h$ , one ac-

$\geq 1; u_0, \tilde{u}_0 \in \mathbb{R}^s$ ).

on – even better than pre-  
 details.

"of ordinary differential equa-  
 tions". Chichester: John

proximations of semigroups,  
 1979).

ia for implicit Runge–Kutta  
 methods (1979).

odes de Runge–Kutta, *Nu-*

ion of semigroups in Banach  
 publ. Interne no. 310 (1986).

: Analysis of error growth for  
 convection problems, *Int.*  
 (1980).

ntegration formulas for initial  
 value problems, Oxford: North-Holland

solution of nonlinear stiff initial  
 value problems". CWI Tract 12.  
 and Computer Science (1985).

note on B-stability of Runge–  
 Kutta methods, 9-331 (1981).

H.W.J., Spijker, M.N.: Stepsize  
 control for the numerical  
 solution of ordinary and par-  
 tial differential equations, *Appl. Math.* 20, 67-81 (1987).

- [11] Kraaijevanger, J.F.B.M., Spijker, M.N.: Algebraic stability and error propagation in Runge–Kutta methods, *Applied Numer. Mathem.* 5, 71-87 (1989).
- [12] Lenferink, H.W.J., Spijker, M.N.: The use of stability regions in the numerical analysis of initial value problems, *Math. Comp.* 57, 221-237 (1991).
- [13] Nevanlinna, O.: Remarks on time discretization of contraction semigroups, Helsinki Univ. Techn., Inst. Math., Report HTPKK-MAT-A225 (1984).
- [14] Parter, S.V.: Stability, convergence and pseudo-stability of finite-difference equations for an overdetermined problem, *Numer. Math.* 4, 277-292 (1962).
- [15] Roux, M.N. le: Semidiscretization time for parabolic problems, *Math. Comp.* 33, 919-931 (1979).
- [16] Schmitt, B.A.: Stability of implicit Runge–Kutta methods for nonlinear stiff differential equations, *BIT* 28, 884-897 (1988).
- [17] Spijker, M.N.: Feasibility and contractivity in implicit Runge–Kutta methods, *J. Comp. Appl. Math.* 12 & 13, 563-578 (1985).
- [18] Spijker, M.N.: Stepsize restrictions for stability of one-step methods in the numerical solution of initial value problems, *Math. Comp.* 45, 377-392 (1985).
- [19] Trefethen, L.N.: Lax-stability vs. eigenvalue stability of spectral methods, Departm. Math., MIT, Cambridge, Massachusetts 02139, Numer. Anal. Report 88-4 (1988).
- [20] Wanner, G.: A short proof on nonlinear A-stability, *BIT* 16, 226-227 (1976).