

Centrum voor Wiskunde en Informatica Centre for Mathematics and Computer Science

W.H. Hundsdorfer

A note on monotonicity of a Rosenbrock method

Department of Numerical Mathematics

Report NM-R8606

February

Bibliotheet: Sentrum voor Wishumke on Information Amsterdam

A Note on Monotonicity of a Rosenbrock Method

W.H. Hundsdorfer

Centre for Mathematics and Computer Science P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

For a dissipative differential equation with stationary solution u^* , the difference between any solution U(t)and u* is nonincreasing with t. In this note we present necessary and sufficient conditions in order for a similar monotonicity property to hold for numerical approximations computed from a Rosenbrock method. Our results also provide global convergence results for some modifications of Newton's method.

1980 Mathematics Subject Classification: 65L20, 65H10 Key Words & Phrases: stiff differential equations, monotonicity, Rosenbrock methods, nonlinear algebraic equations, modified Newton methods.

1. INTRODUCTION

Consider an initial value problem in \mathbb{R}^m

.....

......

$$U'(t) = f(U(t)) \ (t \ge 0),$$
 (1.1a)

$$U(0) = u_0$$
 (1.1b)

whose solution U(t) tends to a stationary solution $u^* \in \mathbb{R}^m$. For the numerical solution of (1.1) we consider the Rosenbrock method

$$u_{n+1} = u_n + (I - h\theta f'(u_n))^{-1} h f(u_n)$$
(1.2)

where θ is a positive parameter, h > 0 is the stepsize and the vectors $u_n \in \mathbb{R}^m$ approximate $U(t_n)$, $t_n = nh \ (n = 0, 1, 2, ...).$

Assume the function f is dissipative with respect to an inner product $<\cdot, \cdot >$ on \mathbb{R}^m $(\langle f(\tilde{u})-f(u),\tilde{u}-u\rangle \leq 0$ for all $\tilde{u},\tilde{u} \in \mathbb{R}^m$ and let $||x|| = \langle x,x \rangle^{1/2}$ (for $x \in \mathbb{R}^m$). This assumption implies that the difference $\|\tilde{U}(t) - U(t)\|$ of any two solutions of the differential equation (1.1a) is nonincreasing with t. The corresponding property for the numerical approximations, $\|\tilde{u}_{n+1}-u_{n+1}\| \leq \|\tilde{u}_n-u_n\|$, only holds under additional, rather restrictive conditions on f (see e.g. [3]). In this note we look at the less exacting monotonicity property

$$\|u_{n+1} - u^*\| \le \|u_n - u^*\|, \tag{1.3}$$

and we shall present conditions on f which are necessary and sufficient for (1.3) to hold with arbitrary stepsize h. Under somewhat stronger conditions on f the convergence of u_n to u^* can be guaranteed. These results are relevant to stiff ordinary differential equations and to partial differential equations (via the method of lines) since neither the Lipschitz constant of f nor the dimension m are involved.

The monotonicity property (1.3) is of particular interest if scheme (1.2) is regarded as a timemarching procedure for finding stationary solutions. The scheme has been used for this purpose in [4] with $\theta = 1$ (and with an approximation to the Jacobian matrix $f'(u_n)$; cf.(2.1)). We note that in such a situation (1.2) can be considered as a modified Newton procedure for solving f(u) = 0. By introducing $\omega = 1/\theta$ and $\lambda = 1/h\theta$ we can rewrite (1.2) as

$$u_{n+1} = u_n - \omega (f'(u_n) - \lambda I)^{-1} f(u_n),$$

in which $\omega > 0$ can be viewed as a relaxation parameter and $\lambda > 0$ ensures that $f'(u_n) - \lambda I$ is

Report NM-R8606

Centre for Mathematics and Computer Science P.O. Box 4079, 1009 AB Amsterdam, The Netherlands 1

nonsingular whenever f is dissipative (see [5; sect.5.4,7.1]).

2. MONOTONICITY FOR NUMERICAL APPROXIMATIONS

Besides the Rosenbrock method (1.2) we also consider the more general linearly implicit scheme

$$u_{n+1} = u_n + (I - h\theta J(u_n))^{-1} h f(u_n)$$
(2.1)

where $J(u_n)$ is an $m \times m$ matrix. Further we shall use the following notation. By $L(\mathbb{R}^m)$ we denote the space of linear operators on \mathbb{R}^m . If $\|\cdot\|$ is a norm on \mathbb{R}^m , the corresponding operator norm on $L(\mathbb{R}^m)$ is also denoted by $\|\cdot\|$, and $\mu[\cdot]$ will stand for the logarithmaic norm (cf.[2]).

Consider the following set of assumptions (2.2)-(2.6), with $\delta > 0$ and $\epsilon \ge 0$, which will be denoted by (A_1) .

 $m \in \mathbb{N}$ and $\|\cdot\|$ is a norm on \mathbb{R}^m generated by an inner product $\langle \cdot, \cdot \rangle$; (2.2)

$$f: \mathbb{R}^m \to \mathbb{R}^m, u^* \in \mathbb{R}^m \text{ is a zero of } f, \text{ and } J: \mathbb{R}^m \to L(\mathbb{R}^m);$$
 (2.3)

$$\int D = \{u : u \in \mathbb{R}^m, ||u - u^*|| < \delta\}, f \text{ is continuously differentiable on } D$$
(2.4)

and J is continuous on D;

for any
$$u \in D$$
 we have $\mu[f'(u)] \le 0$ (2.5)

and J(u) does not have positive real eigenvalues;

for all
$$u, v \in D$$
 there is a $E(u, v) \in L(\mathbb{R}^m)$ such that (2.6)

$$|f'(v) = J(u)(I + E(u, v)), ||E(u, v)|| \leq \epsilon.$$

Further (A_2) will stand for these assumptions (2.2)-(2.6) together with

$$J(u) = f'(u) \text{ (for all } u \in D).$$

$$(2.1)$$

In (2.4) continuously differentiable means that that the matrix of partial derivatives $f'(u) = (\partial f_i(u) / \partial u_j)$ exists and depends continuously on u. The condition $\mu[f'(u)] \leq 0$ on D is equivalent to requiring that f is dissipative on D (see e.g.[5, sect.5.4]). The condition in (2.6) states that the relative difference between f'(v) and J(u) is bounded by ϵ ; in case J(u) is regular it reads $||J(u)^{-1}(f'(v)-J(u))|| \leq \epsilon$. Thus, in a relative sense, the variation of f' on D may not be too large and J(u) has to approximate f'(u) accurately enough.

In order to formulate our main results we define the real functions $\psi_k(k = 1,2)$ on the interval $[1/2,\infty)$ by

$$\psi_1(\theta) = \min\{2\theta - 1, 1\},$$
 (2.8)

$$\psi_2(\theta) = \min\{2\theta - 1, \sqrt{(2\theta - 1)/\theta}\}.$$
(2.9)

THEOREM 2.1. Let h and δ be positive, and k equal to 1 or 2. We have $||u_{n+1} - u^*|| \le ||u_n - u^*||$ (whenever $u_n \in D$ and (A_k) is valid) iff $\theta \ge 1/2$ and $\epsilon \le \psi_k(\theta)$.

This theorem is an extension of a result by M.N. SPIJKER and the present author, see [7; sect. 4]. The proof will be given in the next section. The restriction $\theta \ge 1/2$ in this theorem is not surprising since the methods with $\theta < 1/2$ are not A-stable. For $\theta = 1/2$ we see that the monotonicity property only holds for linear problems ($\epsilon = 0$).

Under slightly stronger conditions on f it can be shown that $||u_{n+1}-u^*|| < ||u_n-u^*||$ (for $u_n \in D, u_n \neq u^*$). This leads to the following result which will also be proved in section 3.

THEOREM 2.2. Let h and δ be positive, and k equal to 1 or 2. Assume $u_0 \in D, \theta \ge 1/2, \epsilon \le \psi_k(\theta)$ and

(A_k). Assume in addition that either $\epsilon < \psi_k(\theta)$ and J(u) is regular (for all $u \in D$) or $\mu[f'(u)] < 0$ (for all $u \in D$). Then u^* is the unique zero of f in D and $\lim_{u \to 0} u_n = u^*$.

3. PROOF OF THE MONOTONICITY RESULTS

3.1. Preliminaries

In order to prove the theorems of section 2 we first derive some technical results. Consider arbitrary $A, B \in L(\mathbb{R}^m)$ with $m \in \mathbb{N}$, and suppose $\|\cdot\|$ is a norm on \mathbb{R}^m generated by an inner product $\langle \cdot, \cdot \rangle$. For any $C \in L(\mathbb{R}^m)$ we denote by C^* its adjoint with respect to this inner product $\langle \cdot, \cdot \rangle$. $\langle \cdot Cx, y \rangle = \langle x, C^* y \rangle$ for all $x, y \in \mathbb{R}^m$.

The relation $||Bx|| \leq \gamma ||Ax||$ (for all $x \in \mathbb{R}^m$), with $\gamma > 0$ given, implies the existence of a $C \in L(\mathbb{R}^m)$ such that B = CA, $||C|| \leq \gamma$; if A is regular we can take $C = BA^{-1}$ and for singular A the inverse A^{-1} can be replaced by the generalized inverse of A (see e.g. [1; ch.8]). Since $||C^*|| = ||C||$ for any $C \in L(\mathbb{R}^m)$ one easily arrives at the following result.

LEMMA 3.1. Let $\gamma > 0$. We have $||B^*x|| \leq \gamma ||A^*x||$ (for all $x \in \mathbb{R}^m$) iff B = AC for some $C \in L(\mathbb{R}^m)$ with $||C|| \leq \gamma$.

Consider the following statements, with $\theta \ge \frac{1}{2}$ and $\epsilon \ge 0$,

 $B = A(I + E_1) \text{ for some } E_1 \in L(\mathbb{R}^m) \text{ with } ||E_1|| \le \epsilon,$ (3.1a)

$$A = B(I + E_2) \text{ for some } E_2 \in L(\mathbb{R}^m) \text{ with } ||E_2|| \le \epsilon,$$
(3.1b)

and

$$B = \theta A (I+F) \text{ for some } F \in L(\mathbb{R}^m) \text{ with } ||F|| \le 1.$$
(3.2)

LEMMA 3.2. (3.1a) implies (3.2) iff $\epsilon \leq \psi_1(\theta)$.

PROOF. Assuming (3.1a) and $\epsilon \leq \psi_1(\theta)$ we set $F = \theta^{-1}[E_1 + (1-\theta)I]$, in which case $B = \theta A (I+F)$ and

$$||F|| \leq \theta^{-1}(\epsilon + |1-\theta|) \leq \theta^{-1}(\psi_1(\theta) + |1-\theta|) = 1.$$

To construct a counterexample in case $\epsilon > \psi_1(\theta)$ we first consider the simple scalar (complex) example A = a, B = b with $a, b \in \mathbb{C}$. The condition in (3.1a) corresponds to

$$|b-a| \le \epsilon |a| \tag{3.3}$$

and (3.2) corresponds to

$$b - \theta a | \leq \theta |a|. \tag{3.4}$$

By simple geometrical arguments it follows that for $\epsilon > \psi_1(\theta)$ there exist $a, b \in \mathbb{C}$ satisfying (3.3) but violating (3.4).

These considerations on C lead to a counterexample with $A = A_1$ and $B = B_1 \in L(\mathbb{R}^2)$,

$$A_1 = \begin{pmatrix} \operatorname{Re} a & -\operatorname{Im} a \\ \operatorname{Im} a & \operatorname{Re} a \end{pmatrix}, B_1 = \begin{pmatrix} \operatorname{Re} b & -\operatorname{Im} b \\ \operatorname{Im} b & \operatorname{Re} b \end{pmatrix},$$

and with $\|\cdot\|$ the Euclidian norm on \mathbb{R}^2 (and the corresponding spectral norm on $L(\mathbb{R}^2)$).

LEMMA 3.3. (3.1a) & (3.1b) implies (3.2) iff $\epsilon \leq \psi_2(\theta)$.

PROOF. Assume (3.1a), (3.1b) and $\epsilon \leq \psi_2(\theta)$. To show that (3.2) holds it is, in view of lemma 3.1, sufficient to consider the remaining case $1 < \epsilon^2 \leq (2\theta - 1)/\theta$. From lemma 3.1 it follows that for any $x \in \mathbb{R}^m$

$$||B^*x||^2 - 2 < A^*x, B^*x > + ||A^*x||^2 \le \epsilon^2 ||A^*x||^2,$$

$$||B^*x||^2 - 2 < A^*x, B^*x > + ||A^*x||^2 \le \epsilon^2 ||B^*x||^2.$$

Combining these inequalities we obtain

$$\langle A^*x, B^*x \rangle \geq (1-\frac{1}{2}\epsilon^2) \|B^*x\|^2.$$

From our assumption on ϵ it follows that

$$||B^*x||^2 \leq 2\theta < A^*x, B^*x >,$$

and hence

$$||(B^* - \theta A^*)x|| \leq \theta ||A^*x||.$$

Statement (3.2) now follows by again applying lemma 3.1.

Now assume $\epsilon > \psi_2(\theta)$ and $1/2 \le \theta \le 1$. Then we obtain a (scalar real) counterexample by taking $m = 1, A = -1, B = -1 - \epsilon$.

Finally assume $\epsilon > \psi_2(\theta)$ and $\theta > 1$. Let $\xi \in ((2\theta-1)/\theta, 2)$ such that $\xi < \epsilon^2$, and take $a, b \in \mathbb{C}$ such that b/a equals $(1-\xi/2)+i\sqrt{\xi(1-\xi/4)}$. Then $|b-a| \le \epsilon |a|$ and $|b-a| \le \epsilon |b|$ but $|b-\theta a| > \theta |a|$. As in the proof of lemma 3.2 such $a, b \in \mathbb{C}$ lead to $A_2, B_2 \in L(\mathbb{R}^2)$ such that for $A = A_2, B = B_2$ the statements (3.1a), (3.1b) hold whereas (3.2) is violated. \Box

We note that in the above counterexamples which prove the necessity of $\epsilon \leq \psi_k(\theta)$ we can choose the $a, b \in \mathbb{C}$ such that $\operatorname{Re} a \leq 0$, $\operatorname{Re} b \leq 0$. This leads to $A_k, B_k \in L(\mathbb{R}^2)$ satisfying $\mu[A_k] \leq 0, \mu[B_k] \leq 0$ (for k = 1, 2).

The following lemma is a slight generalization of results in [6] and [7; lemma 4.3].

LEMMA 3.5. Assume $I - \lambda \theta A$ is regular for all $\lambda > 0$. We have $||I + (I - \lambda \theta A)^{-1} \lambda B|| \le 1$ (for all $\lambda > 0$) iff $\mu[B] \le 0$ and (3.2) holds.

PROOF. Let $C = B - \theta A$. Then $I + (I - \lambda \theta A)^{-1} \lambda B = (I - \lambda \theta A)^{-1} (I + \lambda C)$, and it follows that $||I + (I - \lambda \theta A)^{-1} \lambda B|| \le 1$ iff

 $\|(I+\lambda C^*)x\| \leq \|(I-\lambda\theta A^*)x\| \text{ (for all } x\in\mathbb{R}^m).$

The latter inequality can be written as

$$2\lambda < B^* x, x > +\lambda^2 \|C^* x\|^2 \leq \lambda^2 \|\theta A^*\|^2 \quad \text{(for all } x \in \mathbb{R}^m\text{)}.$$

Clearly this holds for all $\lambda > 0$ iff

$$\langle Bx, x \rangle \leq 0$$
 and $\|C^*x\| \leq \|\theta A^*x\|$ (for all $x \in \mathbb{R}^m$).

Application of lemma 3.1 completes the proof. \Box

3.2. The proof of theorem 2.1

For $u \in D$ we define

$$\sigma(u) = \int_{0}^{1} \|I + (I - h\theta J(u))^{-1} h f'(u^* + \tau(u - u^*))\| d\tau.$$
(3.5)

Since for any $u_n \in D$

$$\|u_{n+1} - u^*\| = \|u_n - u^* + (I - h\theta J(u_n))^{-1}h(f(u_n) - f(u^*))\|$$

it follows by the mean-value theorem that

$$||u_{n+1} - u^*|| \le \sigma(u_n) ||u_n - u^*||.$$
(3.6)

Application of the lemmas 3.2, 3.3 and 3.5 with $A = hJ(u_n)$ and $B = hf'(u^* + \tau(u_n - u^*))$ shows the sufficiency of $\epsilon \leq \psi_k(\theta)$ for having $||u_{n+1} - u^*|| \leq ||u_n - u^*||$ in case (A_k) holds, k = 1, 2. The necessity will be proved by some counterexamples.

A counterexample in case (A_1) holds, $\epsilon > \psi_1(\theta)$ is given by $hJ(u) = \lambda A_1, hf(u) = \lambda B_1 u$ (for $u \in \mathbb{R}^2$) with A_1, B_1 as in the proof of lemma 3.2, $\lambda > 0$ and $\|\cdot\|$ the Euclidean norm on \mathbb{R}^2 . With $u^* = 0, u_n \in \mathbb{R}^2$, we obtain

$$||u_{n+1} - u^*|| = ||(I + (I - \lambda \theta A_1)^{-1} \lambda B_1)(u_n - u^*)|| =$$

= ||I + (I - \lambda \theta A_1)^{-1} \lambda B_1 ||||u_n - u^*|| > ||u_n - u^*||

$$- 11 (1 1021) 10 11110n u 11 > 110n u$$

provided $\lambda > 0$ is suitably chosen (see lemma 3.5).

Next we give a scalar (real) example for $\epsilon > \psi_2(\theta), 1/2 \le \theta \le 1$ in case (A_2) is valid. This counterexample is similar to one given by SANDBERG and SHICHMAN [6].

Take, for convenience, $h = 1, \delta > 1$ and $u^* = 0$, $u_0 = 1$. Let $\eta \in (2\theta - 1, \epsilon)$ and $f(u) = \lambda g(u)$ (for $u \in \mathbb{R}$) with $\lambda > 0$ and $g: \mathbb{R} \to \mathbb{R}$ a continuously differentiable function such that

$$g(0) = 0, g'(u) \in [-1-\epsilon, -1] \text{ (for all } u \in \mathbb{R}),$$

$$g(u) = -u + \eta \text{ (for } u \leq 1), g(u) = -u - \eta \text{ (for } u \geq 1).$$

Such an f meets the conditions imposed in (A_2) . Further we have

$$u_1 = (1+\lambda\theta)^{-1}(1+\lambda(\theta-1-\eta))$$

and thus $|u_1 - u^*|$ tends to $\theta^{-1}(\eta + 1 - \theta) > 1 = |u_0 - u^*|$ for $\lambda \to \infty$.

Finally we assume $\epsilon > \psi_2(\theta)$, $\theta > 1$. For this we construct a complex, scalar counterexample, which can as before be converted to a real one by identifying \mathbb{C} with \mathbb{R}^2 in the usual way. Suppose $(2\theta-1)/\theta < \xi < \min\{2,\epsilon^2\}$ and let $a,b \in \mathbb{C}$ be such that $\operatorname{Re} a < 0$, $\operatorname{Re} b < 0$ and $b/a = (1-\xi/2)+i\sqrt{\xi(1-\xi/4)}$ (as in the proof of lemma 3.3). Then $|b-a| < \epsilon |a|, |b-a| < \epsilon |b|$ but $|b-\theta a| > \theta |a|$, and thus for $\lambda > 0$ suitably chosen $|1+(1-\lambda\theta a)^{-1}\lambda b| > 1$ (see lemma 3.5). We put $\alpha = \lambda a$ and $\beta = \lambda b$.

Let D be the unit disk in $\mathbb{C}, h = 1, u^* = 0$, and define

$$f(u) = \Phi(0)(\alpha - \beta) + \alpha u + \Phi(u)(\beta - \alpha),$$

$$\Phi(u) = -\frac{2k}{k+1} \left(\frac{1}{2} - \frac{1}{2}u\right)^{1+1/k}$$

for $u \in \mathbb{C}$, where $k \in \mathbb{N}$ is to be specified later. Then f(0) = 0 and

$$f'(u) = \alpha + \phi(u)(\beta - \alpha),$$

$$\phi(u) = (\frac{1}{2} - \frac{1}{2}u)^{1/k}.$$

The image of D under ϕ tends to the interval (0,1) on the real axis if $k \to \infty$. By using this property it can be shown that, for k sufficiently large, the conditions on f in (A₂) are satisfied. Moreover, since $f'(1) = \alpha$ and f(1) tends to β for $k \to \infty$,

$$|1+(1-\theta f'(1))^{-1}f(1)| > 1$$

provided k is sufficiently large. It follows that, for such k and u_0 close to 1,

$$|u_1-u^*| > |u_0-u^*|.$$

3.3. The proof of theorem 2.2

First we show that under the assumptions of theorem 2.2 the function σ , defined by (3.5), satisfies $\sigma(u) < 1$ (for all $u \in D$). Examination of the proof of lemma 3.5 shows that for $A, B \in L(\mathbb{R}^m)$ satisfying (3.2) and $\mu[B] \leq 0$ we have

$$||I + (I - \lambda \theta A)^{-1} \lambda B|| < 1$$
 (for all $\lambda > 0$)

provided we assume in addition either

$$\mu[B] < 0$$

or

A is regular and
$$B = \theta A (I+F), ||F|| < 1$$
.

Further it is easily seen, by regarding the proofs of lemma 3.2 and lemma 3.3, that if A is regular and we have (3.1a) with $\epsilon < \psi_1(\theta)$ or (3.1a), (3.1b) with $\epsilon < \psi_2(\theta)$ then there is an $F \in L(\mathbb{R}^m)$ such that $B = \theta A (I+F), ||F|| < 1$. By setting $A = hJ(u), B = hf'(u^* + \tau(u - u^*))$ it follows that the assumptions of theorem 2.2 imply $\sigma(u) < 1$ on D.

The function σ is continuous on D. Therefore we obtain for arbitrary $u_0 \in D$

$$||u_n - u^*|| \leq s_0^n ||u_0 - u^*||$$

with $s_0 = \max\{\sigma(u): u \in D, ||u-u^*|| \le ||u_0-u^*||\} < 1$. From this it is clear that u^* is the unique zero of f in D and that the u_n converge to u^* for $n \to \infty$.

Acknowledgement

The author is grateful to professor M.N. Spijker for the valuable and stimulating discussions.

References

- [1] A. BEN-ISRAEL, T.N.E. GREVILLE Generalized inverses: theory and applications, Wiley-Interscience, New York, 1974.
- [2] K. DEKKER, J.G. VERWER, Stability of Runge-Kutta methods for stiff nonlinear differential equations, North-Holland, Amsterdam, 1984.
- [3] W.H. HUNDSDORFER, The numerical solution of nonlinear stiff initial value problems, CWI Tract 12, Centre for Math. and Comp. Sc., Amsterdam, 1985.
- [4] B. VAN LEER, W.A. MULDER, Relaxation methods for hyperbolic equations, Report 84-20, Dept. of Math.and Informatics, Delft Univ. of Techn., Delft, 1984.
- [5] J.M. ORTEGA, W.C. RHEINBOLDT, Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.
- [6] I.W. SANDBERG, H. SHICHMAN, Numerical Integration of systems of stiff nonlinear initial value problems, B.S.T.J. 47, 511-527 (1968).
- [7] M.N. SPIJKER, A note on monotonicity and contractivity in the numerical solution of stiff differential equations, Report 1985-25, Inst. of Appl. Math. and Comp. Sc., Univ. of Leiden, 1985.