

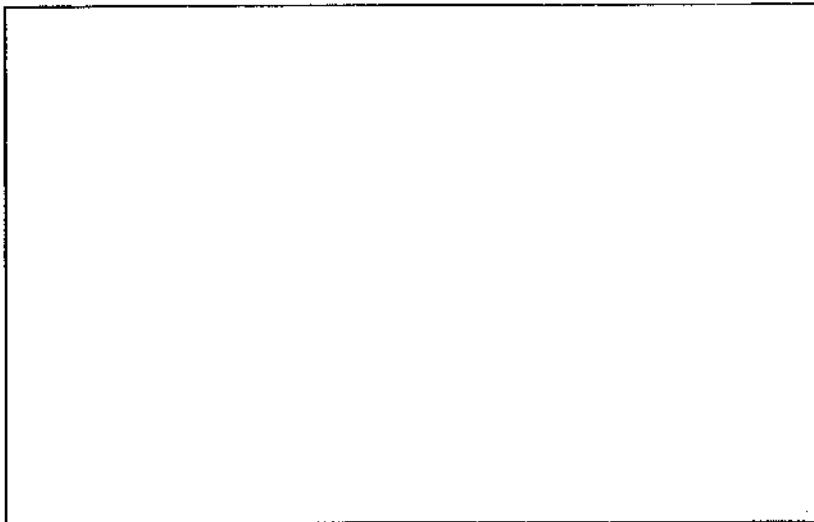
COPY NOT TO BE REMOVED FROM THE LIBRARY

TRITA-NA 7613  
9



THE ROYAL INSTITUTE OF TECHNOLOGY  
STOCKHOLM  
SWEDEN

DEPARTMENT OF  
INFORMATION PROCESSING  
COMPUTER SCIENCE



CERN LIBRARIES, GENEVA



CM-P00069420

TRITA-NA-7613

Inst för Informationsbehandling  
KTH  
100 44 STOCKHOLM 70

Dept of Information Processing  
The Royal Institute of Technology  
S-100 44 STOCKHOLM 70, Sweden

ON THE NUMERICAL INTEGRATION  
OF NONLINEAR INITIAL VALUE PROBLEMS  
BY LINEAR MULTISTEP METHODS

by

Olavi Nevanlinna \*

Report TRITA-NA-7613

\* Institute of Mathematics  
Helsinki University of Technology  
SF-02150 Espoo 15, Finland

A B S T R A C T

We study the numerical solution of the nonlinear initial value problem

$$\begin{cases} \frac{du(t)}{dt} + Au(t) = f(t), & t > 0 \\ u(0) = c, \end{cases}$$

where  $A$  is a nonlinear operator in a real Hilbert space. The problem is discretized using linear multistep methods, and we assume that their stability regions have nonempty interiors. We give sharp bounds for the global error by relating the stability region of the method to the monotonicity properties of  $A$ . In particular we study the case where  $Au$  is the gradient of a convex functional  $\varphi(u)$ .

## 1. INTRODUCTION

The classical error bound for numerical integration of nonlinear initial value problems is basically of the form

$$|u_n - u(nh)| \leq C_1 e^{C_2 M T} h \sum_{j=0}^{T/h} |q_j|, \quad 0 \leq n \leq T/h, \quad (1.1)$$

where  $M$  denotes the Lipschitz-constant of the system and  $\{q_j\}$  a sequence of local errors. The bound (1.1) holds for linear multistep methods whose stability region contains the origin. Using one-sided Lipschitz-conditions sharper bounds, for Lipschitz-continuous systems and some special methods, was obtained by Uhlmann 1957 [13], Lošinskij 1958 [8] and Dahlquist 1959 [3]. Recently, in the works of Dahlquist and Nevanlinna [4,9,10,5,11] and Odeh and Liniger [12] there has been progress in deriving error bounds which take directly use of the stability region of the methods. In this paper we unify and extend these results and present the following error bound

$$|u_n - u(nh)| \leq C h \sum_{j=0}^n \theta^{n-j} |q_j|. \quad (1.2)$$

Here  $C, \theta$  are independent on  $n$  and  $\theta$  is less than (equal to) or greater than one depending on whether the nonlinear operator lies, in a circle condition sense, strictly inside (inside) the stability region or overlaps the instability region. The treatment covers all linear multistep methods whose stability region has a nonempty interior, and the properties required on the nonlinear operator are formulated using one-sided Lipschitz-conditions. Gradients of convex functionals form an important class of such (monotone) operators. In that case the solutions of the initial value problem have some special nonoscillation properties, and we study those multistep methods which carry these properties to the difference equation.

The author spent the year 1975-76 at Institut Mittag-Leffler, The Royal Swedish Academy of Sciences. He is very grateful for that year, especially, since it made possible an inspiring collaboration with professor G. Dahlquist.

## 2. EXISTENCE OF SOLUTIONS

Let  $H$  be a real Hilbert space and  $A$  a *nonlinear* singlevalued operator  $D(A) \subset H \rightarrow H$ . We consider the numerical solution of the evolution equation

$$\begin{cases} \frac{du(t)}{dt} + Au(t) = f(t), & t > 0 \\ u(0) = c \in D(A) \end{cases} \quad (2.1)$$

using the difference analogue

$$\begin{cases} h^{-1}\rho(E)u_n + \sigma(E)Au_n = f_{n+k}, & n \geq 0 \\ u_j = c_j \in D(A), & 0 \leq j < k \end{cases} \quad (2.2)$$

In (2.2)  $E$  denotes the translation operator  $Ey_n = y_{n+1}$  and  $\rho, \sigma$  are real polynomials

$$\rho(\zeta) = \sum_{j=0}^k \alpha_j \zeta^j, \quad \sigma(\zeta) = \sum_{j=0}^k \beta_j \zeta^j \quad (2.3)$$

with no common divisor. Further, we assume that  $\alpha_k > 0$ ,  $\rho(1) = 0$ ,  $\rho'(1) = \sigma(1)$ .

REMARK 2.1. With some additional notation we could consider cases where  $A$  were multivalued and time dependent and  $f$  Lipschitz-continuous in  $u$ .

DEFINITION 2.1. An operator  $B: D(B) \subset H \rightarrow H$  is called monotone if

$$\langle u - v, Bu - Bv \rangle \geq 0 \quad \text{for all } u, v \in D(B). \quad (2.4)$$

It is called maximally monotone if additionally  $R(I+B) = H$ .  $\square$

On monotone operators see [1].

Assume that

$$(I + ahA)^{-1} \text{ is Lipschitz-continuous, with constant } L, \quad (2.5)$$

and

$$A(I + ahA)^{-1} - \alpha I \text{ is maximally monotone} \quad (2.6)$$

with some fixed  $a, \alpha \in \mathbb{R}$ ,  $h \in \mathbb{R}_+$ .

THEOREM 2.1. Let  $\{f_{n+k}\} \subset H$ ,  $\{c_j\}_{j=0}^{k-1} \subset D(A)$  and assume that (2.5), (2.6) hold. If

$$\beta_k \neq 0, \beta_k - a\alpha_k \geq 0 \text{ and } ah(\beta_k - a\alpha_k)/\alpha_k > -1 \quad (2.7)$$

then there exists a unique  $\{u_n\} \subset D(A)$  such that (2.2) holds.  $\square$

PROOF. Assuming that  $u_m$  are given for  $m < n+k$  the equation (2.2) can be written as

$$u + (h\beta_k/\alpha_k)Au = d \quad (2.8)$$

with  $d$  a given and  $u = u_{n+k}$  the unknown vector. Put  $u + ahAu = w$ , then (2.8) implies

$$w + \lambda(1+\alpha\lambda)^{-1}(A(I+ahA)^{-1} - \alpha I)w = (1+\alpha\lambda)^{-1}d, \quad (2.9)$$

with  $\lambda = h(\beta_k/\alpha_k - a) \geq 0$ . By (2.6) the equation (2.9) has a unique solution (see Prop. 2.2 in [1]). But then (2.5) implies the existence of  $u \in H$  such that (2.8) holds. Since  $\beta_k \neq 0$  we have  $u \in D(A)$ .  $\square$

REMARK 2.2. Draw a circle with real center passing through  $a^{-1}$  and  $-ah/(1-ah)$ . Applied to the operator  $Au = -\lambda u$  with  $\lambda \in \mathbb{C}$  (2.6) means, that  $\lambda h$  lies inside that circle if  $1-ah > 0$  and outside if  $1-ah < 0$ . (In complex spaces one considers the real part of the inner product in (2.4)). For later reference call the corresponding disc  $D(a, ah)$ .

### 3. MAIN RESULT

The main tool in deriving our error bound is to associate with  $\rho$  and  $\sigma$  a convolution operator and relate its Fourier transform to the stability region of the method. This idea was developed in [10] and [5], and the technique used in [12] is very similar.

DEFINITION 3.1. The stability region  $S$  of a method  $(\rho, \sigma)$  consists of those  $q \in \mathbb{C}$  for which the characteristic polynomial  $\rho(\zeta) - q\sigma(\zeta)$  satisfies the root condition, i.e. its roots satisfy  $|\zeta_j| \leq 1$  and those of modulus one are simple.  $\square$

Fix a number  $a \in \mathbb{R}$  such that

$$a^{-1} \in \text{Int } S. \quad (3.1)$$

Let  $\{\zeta_j\}$  denote the roots of  $\sigma(\zeta)$  (if  $\beta_k = 0$  then  $\infty$  is considered as one root) and set  $\sigma^* = \max_j |\zeta_j|$ . For  $a=0$  (3.1) is understood to mean that  $\sigma^* < 1$ . When  $a^{-1} \in \text{Int } S$ ,  $|\sigma(\zeta) - a\rho(\zeta)|$  is bounded away from zero on  $|\zeta| = 1$ .

Put

$$\inf_{|\zeta|>1} \text{Re} \frac{\rho(\zeta)}{\sigma(\zeta) - a\rho(\zeta)} = -b. \quad (3.2)$$

Since  $\rho(1) = 0$  we always have  $b \geq 0$ . In the notation of the previous chapter, (3.2) means for  $a^{-1} < b$  that  $D(a, b) \subset S$  and for  $a^{-1} > b$  that  $\mathbb{C} \setminus S \subset D(a, b)$ .

Assume that  $\{u_n\}$  satisfies (2.2) and  $\{v_n\}$  satisfies

$$h^{-1} \rho(E)v_n + \sigma(E)Av_n = g_{n+k}, \quad n \geq 0. \quad (3.3)$$

$$v_j = d_j \in D(A), \quad 0 \leq j < k.$$

Here  $\{v_n\}$  might e.g. be  $v_n = u(nh)$  where  $u(t)$  satisfies (2.1) or  $v_n = u_{n+1}$ .

Put  $f_{n+k} - g_{n+k} = q_{n+k}$ . Then we have

$$h^{-1} \rho(E)(u_n - v_n) + \sigma(E)(Au_n - Av_n) = q_{n+k}, \quad n \geq 0. \quad (3.4)$$

THEOREM 3.1. Let  $a, b$  be as in (3.1), (3.2) and assume that (2.5) and (2.6) hold. There exists a constant  $\delta > 0$  such that if

$$ah - b > -\delta \quad (3.5)$$

then we have for  $\{u_n - v_n\}$  satisfying (3.4)

$$|u_n - v_n| \leq C \left\{ e^n \max_{0 \leq j < k} \left[ |u_j - v_j| + h|Au_j - Av_j| \right] + h \sum_{j=k}^n e^{n-j} |q_j| \right\} \quad (3.6)$$

Here  $C, \theta$  are independent on  $n$  and.

$$\begin{aligned} &< 1 \text{ if } ah - b > 0 \\ &\theta = 1 \text{ if } ah - b = 0 \\ &> 1 \text{ if } ah - b < 0. \end{aligned}$$

PROOF. Define  $u_n - v_n = 0$  for  $n < 0$  and set  $q_m = h^{-1} \rho(E)(u_{m-k} - v_{m-k}) + \sigma(E)(Au_{m-k} - Av_{m-k})$  for  $m < k$ . Then (3.4) holds for all  $n \in \mathbb{Z}$ , with

$$q_m = 0 \text{ for } m < 0 \text{ and } h|q_m| \leq C \max_{0 \leq j < k} [ |u_j - v_j| + h|Au_j - Av_j| ] \text{ for } 0 \leq m \leq k-1.$$

Let us consider the case  $ah - b > 0$  first. Since  $a^{-1} \in \text{Int } S$   $(\sigma(\zeta) - a\rho(\zeta))^{-1}$  is analytic for  $|\zeta| > 1-\varepsilon$  with some  $\varepsilon > 0$ . Combining this with (3.2) we conclude that there exists  $\theta < 1$  such that

$$\operatorname{Re} \left\{ \left[ \sigma(\theta\zeta) - a\rho(\theta\zeta) \right]^{-1} \rho(\theta\zeta) \right\} \geq -b(\theta) \quad \text{for } |\zeta| > 1 \quad (3.7)$$

$$\text{and } ah - b(\theta) \geq \frac{1}{2}(ah - b) > 0.$$

Define  $[\sigma(E) - a\rho(E)]^{-1}$  using the power series

$$[\sigma(\zeta) - a\rho(\zeta)]^{-1} \sim \zeta^{-k} \{ \gamma_0 + \gamma_1 \zeta^{-1} + \gamma_2 \zeta^{-2} + \dots \}.$$

Since  $a^{-1} \in \text{Int } S$  the sequence  $\{\gamma_j\}$  is exponentially decaying and

$$[\sigma(E) - a\rho(E)]^{-1} \xi_{n+k} = \sum_{j=-\infty}^n \gamma_{n-j} \xi_j \text{ is well defined for all } \{\xi_j\} \in \ell^\infty.$$

For shortness put

$$Q_n = [\sigma(E) - a\rho(E)]^{-1} q_{n+k},$$

$$\Gamma x_n = [\sigma(E) - a\rho(E)]^{-1} \rho(E)x_n, \quad \text{and}$$

$$w_n = (u_n + ahAu_n) - (v_n + ahAv_n).$$

Since (2.5) means that  $|u_n - v_n| \leq L|w_n|$  it is sufficient to show (3.6) for  $|w_n|$ . The equation (3.4) can now be written as

$$\Gamma w_n + h(Au_n - Av_n) = h Q_n. \quad (3.8)$$

If  $\Gamma w_n = \sum_{j=0}^n c_{n-j} w_j$ , then multiplying (3.8) by  $\theta^{-2n} w_n$  and summing we get

$$\sum_{n=0}^N \left\langle \theta^{-n} w_n, \sum_{j=0}^n c_{n-j} \theta^{-(n-j)} \theta^{-j} w_j \right\rangle +$$

$$+ h \sum_{n=0}^N \left\langle \theta^{-2n} w_n, Au_n - Av_n \right\rangle = h \sum_{n=0}^N \left\langle \theta^{-2n} w_n, Q_n \right\rangle.$$

Using Parseval's identity to the sequence  $\tilde{w}_n = w_n$ ,  $n \leq N$ ,  $\tilde{w}_n = 0$  for  $n > N$ , (3.7) yields

$$\sum_{n=0}^N \left\langle \theta^{-n} w_n, \sum_{j=0}^N c_{n-j} \theta^{-(n-j)} \theta^{-j} w_j \right\rangle + b(\theta) \sum_{n=0}^N |\theta^{-n} w_n|^2 \geq 0. \quad (3.10)$$

The second term in (3.9) satisfies, by (2.6),

$$h \sum_{n=0}^N \left\langle \theta^{-2n} w_n, A u_n - A v_n \right\rangle \geq \alpha h \sum_{n=0}^N |\theta^{-n} w_n|^2. \quad (3.11)$$

Hence we have, combining (3.9), (3.10) and (3.11)

$$[\alpha h - b(\theta)] \left\{ \sum_{n=0}^N |\theta^{-n} w_n|^2 \right\} \leq h \left\{ \sum_{n=0}^N \theta^{-2n} |q_n|^2 \right\}^{1/2} \leq \gamma(\theta) \left\{ \sum_{n=0}^N \theta^{-2n} |q_n|^2 \right\}^{1/2}, \quad (3.12)$$

where  $\gamma(\theta) = \sup_{|\zeta|=0} |[\sigma(\zeta) - a\rho(\zeta)]^{-1}| < \infty$ . But (3.12) yields

$$|w_N| \leq [\alpha h - b(\theta)]^{-1} \gamma(\theta) h \sum_{n=0}^N \theta^{N-n} |q_n|.$$

To consider the case  $\alpha h - b \leq 0$  define a rational function  $r$  by

$$r(\zeta) = [b\sigma(\zeta^{-1}) + (1-ab)\rho(\zeta^{-1})] [\sigma(\zeta^{-1}) - a\rho(\zeta^{-1})]^{-1}. \quad (3.13)$$

By (3.2) we have  $\operatorname{Re} r(\zeta) \geq 0$  for  $|\zeta| < 1$ . Again  $r(\zeta)$  is analytic for  $|\zeta| < 1 + \varepsilon$  with some  $\varepsilon > 0$ . Also  $\operatorname{Re} r(0) \neq 0$ . Hence, using the harmonicity of  $\operatorname{Re} r$  we obtain

$$\begin{aligned} \operatorname{Re} r(\zeta) &\neq 0 \text{ on } |\zeta| = 1 \\ &> 0 \text{ for } |\zeta| < 1. \end{aligned} \quad (3.14)$$

Since  $r$  is rational  $\operatorname{Re} r(\zeta) = 0$  only for a finite number of points on  $|\zeta| = 1$ , and we conclude that there exists a nontrivial polynomial  $p$  such that

$$\operatorname{Re} r(\zeta) - |p(\zeta)|^2 \geq 0 \text{ for } |\zeta| \leq 1. \quad (3.15)$$

Using Parseval's identity we now obtain

$$\sum_{n=0}^N \left\langle w_n, \Gamma w_n \right\rangle + b \sum_{n=0}^N |w_n|^2 \geq \sum_{n=-\infty}^{\infty} |p(\zeta^{-1}) \tilde{w}_n|^2 \geq \delta |w_N|^2, \quad (3.16)$$

with some  $\delta > 0$ , where we used the fact that  $w_n = 0$  for  $n > N$ . Hence, (3.8) implies

$$|w_N|^2 + (\alpha h - b) \delta^{-1} \sum_{n=0}^N |w_n|^2 \leq h \delta^{-1} \sum_{n=0}^N |w_n| |q_n|. \quad (3.17)$$

For all  $n > 0$  this yields

$$|w_N|^2 \leq [-(\alpha h - b) \delta^{-1} + n/2] \sum_{n=0}^N |w_n|^2 + h^2 (2n\delta^2)^{-1} \sum_{n=0}^N |Q_n|^2. \quad (3.18)$$

Put  $d = -(\alpha h - b) \delta^{-1} + n/2$  and  $D = (2n\delta^2)^{-1} \gamma^2$ , where

$\gamma = \gamma(1) = \sup_{|\zeta|=1} |\sigma(\zeta) - \alpha\varphi(\zeta)|^{-1}$ . Then (3.18) implies

$$|w_n|^2 \leq d \sum_{m=0}^n |w_m|^2 + Dh^2 \sum_{m=0}^n |Q_m|^2. \quad (3.19)$$

Take  $n$  so small that  $d < 1$  and define  $\{\omega_n\}$  by  $\omega_0 = |w_0|^2$  and

$$\omega_n = d \sum_{m=0}^n \omega_m + Dh^2 \sum_{m=0}^n |Q_m|^2 \text{ for } n > 0.$$

Then (3.19) implies  $|w_n|^2 \leq \omega_n$  and we obtain the bound

$$\begin{aligned} |w_N|^2 &\leq (1-d)^{-N} |w_0|^2 + Dh^2 \sum_{n=0}^N (1-d)^{-(N-n)} |Q_n|^2 \leq \\ &\leq \left\{ C h \sum_{n=0}^N (1-d)^{-(N-n)/2} |Q_n|^2 \right\}^2. \end{aligned}$$

To see that in the case  $\alpha h - b = 0$  we actually may take  $\theta = 1$  we proceed as in the proof of Theorem 2 in [9]. Fix a positive integer  $M$ . If  $Q_n = 0$  for  $n \leq M$  there is nothing to show. Otherwise, estimate

$$|w_n| \leq (1/2)(\delta + \alpha h - b)^{-1} h \sum_{m=0}^M |Q_m| + (1/2)(\delta + \alpha h - b) \left[ h \sum_{m=0}^M |Q_m| \right]^{-1} |w_n|^2.$$

Put  $\delta(\delta + \alpha h - b) = 1/K$ ,  $(\alpha h - b)\delta^{-1} = H$  and  $R = h \sum_{m=0}^M |Q_m|$ , so that  $K \in (0, \infty)$ ,  $H \in (-1, 0]$ . From (3.17) we get

$$|w_N|^2 + H \sum_{n=0}^N |w_n|^2 \leq (1/2)KR^2 + (1/2)(1+H)R^{-1}h \sum_{n=0}^N |w_n|^2 |Q_n|, \quad (3.20)$$

for  $0 \leq N \leq M$ . Define  $\{\omega_N\}_0^M$  by replacing in (3.20)  $|w_n|^2$  by  $\omega_n$  and the inequality by equality. Since

$$1 + H - (1/2)(1+H)R^{-1}h |Q_N| \geq (1/2)(1+H) > 0,$$

$\omega_N$  is well defined. Assuming  $|w_n|^2 \leq \omega_n$  for  $n = 0, 1, \dots, N-1$  we have

$$|w_N|^2 + H \sum_{n=0}^N |w_n|^2 \leq \omega_{N-1} + H \sum_{n=0}^N \omega_n + (1/2)(1+H)R^{-1}h |Q_N| |w_N|^2.$$

Since  $H \leq 0$  this gives

$$|w_N|^2 + H|w_N|^2 \leq \omega_{N-1} + (1/2)(1+H)R^{-1}h|Q_N||w_N|^2$$

which shows that  $|w_N|^2 \leq \omega_N$ . But  $|w_0|^2 \leq \omega_0$  and so  $|w_N|^2 \leq \omega_N$  for  $0 \leq N \leq M$ . The conclusion follows, since

$$R \leq C \left\{ \max_{0 \leq j < k} \left[ |u_j - v_j| + h|Au_j - Av_j| \right] + h \sum_{n=k}^M |q_n| \right\}$$

and

$$\begin{aligned} \omega_M &= (1/2)KR^2(1+H)^{-M-1} \prod_{j=0}^M [1 - (1/2)R^{-1}h|Q_j|]^{-1} \leq \\ &\leq (e/2)KR^2(1+H)^{-M-1}. \end{aligned} \quad \square \quad (3.21)$$

REMARK 3.1. In most applications one chooses  $a^{-1} \in \text{Int } S$  so that  $b = 0$ . If one is interested in convergence when  $h \rightarrow 0$  the assumptions (2.5) and (2.6) must hold for  $h \in (0, h_0]$  with a fixed constant  $a$ . Since the constants in (3.6) generally depend on  $h$  we give the bound in another form, assuming  $b = 0$ . Put  $\theta = 1$  in (3.12), then for  $\alpha > 0$

$$\left\{ h \sum_{n=0}^{T/h} |u_n - v_n|^2 \right\}^{1/2} \leq L \alpha^{-1} \gamma \left\{ h \sum_{n=0}^{T/h} |q_n|^2 \right\}^{1/2}. \quad (3.22)$$

where  $\gamma = \sup_{|\zeta|=1} |\sigma(\zeta) - a\varphi(\zeta)|^{-1}$ .

For  $\alpha \leq 0$  we get from (3.21) that

$$|u_n - v_n| \leq (e/2)^{1/2} L \delta^{-1} \hat{\gamma} (1 - |\alpha| h \delta^{-1})^{-1 - T/2h} h \sum_{n=0}^{T/h} |q_n|, \quad 0 \leq n \leq T/h, \quad (3.23)$$

where  $\hat{\gamma} = \sum_{n=0}^{\infty} |\gamma_n|$ . Here  $h q_n$  denotes for  $n < k$  the initial errors in terms of  $|u_j - v_j|$ ,  $h|Au_j - Av_j|$ .

Observe that if  $\delta = 1/2$  the growth factor is of the "right" size,  $\exp[T|\alpha|(1 + O(h))]$ . The next example shows that this actually can happen. However, let us first point out that if we only know the Lipschitz-constant of  $A$ , say  $M$ , then for  $\alpha \leq 0$ , (2.5) is satisfied with  $L = (1 - |\alpha| h M)^{-1}$  and, hence, (2.6) holds with  $\alpha = -M(1 - |\alpha| h M)^{-1} = -M + O(h)$ . This means that (3.23) still gives the classical bound (1.1) with some  $C_1, C_2$ .

EXAMPLE 3.1. Consider one step methods  $(\rho, \sigma) = (\zeta - 1, (1-\beta)\zeta + \beta)$  with  $\beta \in [0, 1]$ . Choosing  $a = -\beta$  we have

$$\operatorname{Re}[\sigma(\zeta) - a\rho(\zeta)]^{-1}\rho(\zeta) = \operatorname{Re}(1 - \zeta^{-1}) \geq (1/2)|1 - \zeta^{-1}| \text{ for } |\zeta| \geq 1$$

Hence

$$\sum |P(E^{-1})w_n|^2 = (1/2) \sum |v w_n|^2 \geq (1/2)|w_N|^2,$$

if  $w_n = 0$  for  $n > N$ , and we have  $\delta = 1/2$ . Further,  $\gamma = \hat{\gamma} = 1$ . The choice  $a = -\beta = 0$  means that (2.5) holds with  $L = 1$ . For  $\beta > 0$  assume that  $A$  is Lipschitz-continuous with constant  $M$ , so that (2.5) holds with  $L = (1 - \beta hM)^{-1}$ . Clearly, the choice  $a = 0$  is possible for all methods with  $\beta < 1/2$ , but as  $\beta \rightarrow 1/2$  then  $\delta^{-1}\hat{\gamma} \rightarrow \infty$ .

REMARK 3.2. An  $\ell^2$ -version of Theorem 3.1 (of the form (3.22)) in the case  $ah - b > 0$  was given in [5]. A similar result for  $A_\infty$ -stable methods was independently found somewhat earlier by Odeh and Liniger [12]. They also used the exponential weighting technique to obtain the corresponding result for the  $\ell^\infty$ -norm. The case  $ah - b \leq 0$  was considered in [11]. Further, in [11] Theorem 3.1 was shown (with  $a = b = 0$ ) for G-stable methods, by slightly simplifying the proofs given in [9].

REMARK 3.3. Let  $A - \alpha I$  be maximally monotone with some  $\alpha > 0$  and assume that  $f(t) \rightarrow f_\infty$  as  $t \rightarrow \infty$ . Let  $u_\infty$  be the unique solution of  $Au_\infty = f_\infty$ , then the solution  $u(t)$  to (2.1) satisfies  $u(t) \rightarrow u_\infty$ , see Theorem 3.9 in [1]. From Theorem 3.1 we obtain for an  $A$ -stable method with  $\sigma^* < 1$  that also  $u_n \rightarrow u_\infty$  as  $n \rightarrow \infty$  if  $f_n \rightarrow f_\infty$  (apply (3.6) with  $v_n = u_\infty$ ).

4. THE CASE A = ∂φ

When A is the subdifferential of a lower semicontinuous convex functional then the solution of (2.1) has some additional smoothness properties, see Chapt. III3 in [1]. We shall study those multistep methods which carry these properties on the difference equations. Since we are considering only singlevalued operators we assume that Au is the Gateaux gradient of a proper convex functional  $\phi: H \rightarrow R \cup \{\infty\}$ . Hence we have

$$\phi(\xi) \geq \phi(x) + \langle \xi - x, Ax \rangle \quad \text{for all } \xi, x \in D(A). \quad (4.1)$$

Put  $\{\nabla x_n\} = \{x_n - x_{n-1}\}$ . Then multiplying  $Ax_n$  by  $\nabla x_n$  and summing we get an inequality using (4.1). For convenience we shall assume that  $\phi$  is nonnegative. We shall first have  $\sigma^* < 1$ . Hence  $\Gamma = \sigma(E)^{-1} \rho(E)$  and  $\sigma(E)^{-1}$  are well defined for all  $\{x_j\} \in l^\infty$ . Put  $u_n = u_0$  for  $n < 0$  and  $u_n = u_N$  for  $n > N$  with  $N > 0$  fixed. Define  $f_n$  for  $n < k$  so that (2.2) holds for all  $n \in \mathbb{Z}$ .

Hence

$$h^{-1} \Gamma u_n + A u_n = \sigma(E)^{-1} f_{n+k}. \quad (4.2)$$

Multiply (4.2) by  $\nabla u_n$ , sum up and use (4.1) to get

$$h^{-1} \sum_{n=1}^N \langle \nabla u_n, \Gamma u_n \rangle + \phi(u_N) \leq \phi(u_0) + \sum_{n=1}^N \langle \nabla u_n, \sigma(E)^{-1} f_{n+k} \rangle. \quad (4.3)$$

Assume that there exists constants  $0 < \lambda, \Lambda < \infty$  such that

$$\lambda \sum_{n=1}^N |\nabla u_n|^2 \leq \sum_{n=1}^N \langle \nabla u_n, \Gamma u_n \rangle \leq \Lambda \sum_{n=1}^N |\nabla u_n|^2 \quad (4.4)$$

holds for all  $N > 0$ . Then we can state

THEOREM 4.1. Let  $\sigma^* < 1$  and (4.4) hold. If (4.1) holds with  $\phi \geq 0$  then there exists a constant C, independent on N, such that

$$\begin{aligned} & \left[ h \sum_{n=1}^N h^{-2} |\nabla u_n|^2 \right]^{\frac{1}{2}} + [\phi(u_N)]^{\frac{1}{2}} \leq \\ & \leq C \left\{ h^{\frac{1}{2}} \max_{-k \leq \mu < 0} \left[ h^{-1} |\rho(E)u_\mu| + |\sigma(E)Au_\mu| \right] + [\phi(u_0)]^{\frac{1}{2}} + \left[ h \sum_{n=k}^N |f_n|^2 \right]^{\frac{1}{2}} \right\} \end{aligned} \quad (4.5)$$

PROOF. Using (4.4) the inequality (4.3) yields

$$h^{-1} \lambda \sum_{n=1}^N |\nabla u_n|^2 + \phi(u_N) \leq \phi(u_0) + \sum_{n=1}^N |\nabla u_n| |\sigma(E)^{-1} f_{n+k}|. \quad (4.6)$$

Hence we have

$$(1/2)h^{-1}\lambda \sum_{n=1}^N |\nabla u_n|^2 + \varphi(u_N) \leq (1/2\lambda) h \sum_{n=1}^N |\sigma(E)^{-1} h_n|^2 + \varphi(u_0) \quad (4.7)$$

and, since  $\sigma^* < 1$ ,  $\varphi(u) \geq 0$  and

$$|f_n| \leq \max_{-k \leq \mu < 0} [h^{-1} |\rho(E)u_\mu| + |\sigma(E)Au_\mu|] \quad \text{for } n < k,$$

the estimate (4.5) follows.  $\square$

Compare Theorem 4.1 e.g. to Theorem 3.6 in [1].

THEOREM 4.2. *The inequalities in (4.4) hold if and only if  $\sigma^* < 1$  and*

$$\operatorname{Re} \left\{ \left[ 1 - e^{-i\tau} \right] \left[ \frac{\rho}{\sigma}(e^{-i\tau}) - \lambda(1 - e^{i\tau}) \right] \right\} \geq 0 \quad \text{for } \tau \in [-\pi, \pi]. \quad (4.8)$$

Consider the trapezoidal rule for which we have  $\sigma^* = 1$ . When  $u_0 = 0$  we have  $Tu_n = u_n - 2u_{n-1} + 2u_{n-2} - \dots$  and hence

$$\sum_{n=1}^N \langle \nabla u_n, Tu_n \rangle = (1/2) \sum_{n=1}^N |\nabla u_n|^2 + \sum_{n=1}^N \langle (\omega \nabla u_n), \nabla u_n \rangle \geq (1/2) \sum_{n=1}^N |\nabla u_n|^2$$

where  $\omega = \{\omega_n\}_{n \geq 0} = \left\{ \frac{1}{2}, -1, 1, -1, \dots \right\}$ . We conclude that  $\sigma^* < 1$  is essential

for the existence of  $\Lambda$  but a positive  $\lambda$  may exist even when  $\sigma^* = 1$ . In fact, such a result can be proved but since  $\sigma^* < 1$  was used in Theorem 4.1 anyway we shall not go into it (multiply  $c_n$  by  $r^n < 1$  and let first  $s \nearrow 1$  and then  $r \nearrow 1$ ).

Proof of Theorem 4.2. By changing variables  $v_n = u_n - u_0$  we may assume that  $u_0 = 0$ .

Since  $\sigma^* < 1$  there exists  $\Lambda < \infty$  such that

$$\operatorname{Re} \left[ 1 - e^{-i\tau} \right] \frac{\rho}{\sigma}(e^{-i\tau}) \leq 2\Lambda(1 - \cos \tau).$$

Using Parseval's identity this gives the right hand side of (4.4) exactly in the same way as (4.8) does the left hand side.

Fix  $\sigma^* < \sigma < s < 1$  and  $M > N$  and define  $v \in \ell^2$  by

$$v_j = \begin{cases} u_j, & \text{for } j \leq M \\ s^j u_j, & \text{for } j > M. \end{cases}$$

Then we have

$$\begin{aligned} & \sum_{n=-\infty}^{\infty} \langle \nabla v_n, \Gamma v_n \rangle - \lambda \sum_{n=-\infty}^{\infty} |\nabla v_n|^2 = \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re} \left\{ \left[ 1 - e^{-i\tau} \right] \left[ \frac{\rho}{\sigma} (e^{-i\tau}) - \lambda (1 - e^{i\tau}) \right] \right\} |\hat{v}(\tau)|^2 d\tau, \end{aligned}$$

which means

$$\lambda \sum_{n=1}^N |\nabla u_n|^2 \leq \sum_{n=1}^N \langle \nabla u_n, \Gamma u_n \rangle + C(s) + D(s),$$

where  $C(s) = C_1 + C_2 + C_3$

$$\begin{aligned} &= \sum_{n=M+1}^{\infty} \left\langle (s^n - s^{n-1}) u_N, \sum_{j=0}^M c_{n-j} u_j \right\rangle \\ &+ \sum_{n=M+1}^{\infty} \left\langle (s^n - s^{n-1}) u_N, \sum_{j=M+1}^n c_{n-j} s^{-j} u_N \right\rangle \\ &- \lambda \sum_{n=M+1}^{\infty} (s^n - s^{n-1})^2 |u_N|^2 \end{aligned}$$

Since

$$\left| \sum_{j=0}^M c_{n-j} u_j \right| \leq K_1 \sigma^n$$

we have

$$|C_1| \leq K_2 (1-s)(1-s\sigma)^{-1}.$$

To estimate the second term write

$$\begin{aligned} \left| \sum_{j=M+1}^n c_{n-j} s^j \right| &\leq s^n \left| \sum_{j=0}^n c_j s^{-j} \right| + \left| \sum_{j=0}^M c_{n-j} s^j \right| \leq \\ &\leq s^n \frac{\rho}{\sigma} (s) + \sigma^n (1 - \sigma/s)^{-1} + K_3 \sigma^n. \end{aligned}$$

Hence

$$|C_2| \leq K_4 (1-s) \left[ (1-s^2)^{-1} \frac{\rho}{\sigma} (s) + (1-s\sigma)^{-1} (1-\sigma/s)^{-1} + (1-\sigma s)^{-1} \right].$$

Because we also have

$$|C_3| \leq K_5 (1-s)^2 (1-s^2)^{-1}$$

we obtain  $C_1 + C_2 + C_3 \rightarrow 0$  as  $s \neq 1$ .

Since  $D(s)$  only corrects  $C(s)$  for the index  $n = M+1$  we also have  $D(s) \rightarrow 0$  as  $s \neq 1$ , and the inequality follows.

On the contrary, if (4.8) does not hold then we have for some sequence  $\{w_j\}$  with compact support

$$\lambda \sum_{n=-\infty}^{\infty} |\nabla w_n|^2 > \sum_{n=-\infty}^{\infty} \langle \nabla w_n, \Gamma w_n \rangle.$$

Translating the sequence we may assume that  $\text{supp}\{w_n\} \subset \{1, 2, \dots, N-1\}$  with some  $N < \infty$ . Hence (4.4) does not hold either.  $\square$

COROLLARY 4.1. Assume that  $\rho$  has no other root on the circle than  $\zeta_1 = 1$  and that  $\sigma^* < 1$ . If  $(\rho, \sigma)$  is A-stable and

$$\text{Im} \frac{\rho}{\sigma}(e^{i\tau}) \geq 0 \quad \text{for } \tau \in [0, \pi] \quad (4.9)$$

then (4.4) holds.  $\square$

PROOF. Since A-stability gives

$$\text{Re} \frac{\rho}{\sigma}(e^{i\tau}) \geq 0 \quad \text{for } \tau \in [-\pi, \pi]$$

we get combining (4.9) and the strict root condition

$$\psi(\tau) = \text{Re}(1 - e^{-i\tau}) \frac{\rho}{\sigma}(e^{-i\tau}) > 0 \quad \text{for } \tau \neq 0.$$

But

$$(1/2)(1 - \cos \tau)^{-1} \psi(\tau) \rightarrow \frac{\rho'(1)}{\sigma(1)} = 1 \quad \text{as } \tau \rightarrow 0$$

and the existence of a positive  $\lambda$  follows from the continuity of  $\psi$ .  $\square$

Observe that (4.9) does not hold for all 2-step A-stable methods

$$\begin{cases} \rho(\zeta) = (1/2)(1+c)\zeta^2 - c\zeta + (1/2)(c-1) \\ \sigma(\zeta) = (1/4)(1+c+d)\zeta^2 + (1/2)(1-d)\zeta + (1/4)(1-c+d), \quad c \geq 0, d > 0 \end{cases}$$

but only when  $c^2 \geq d$ . But if e.g.  $c = 1$ ,  $d \geq 6$  then (4.8) does not hold either. Further, A-stability is not necessary for (4.8). This is seen e.g. by considering the backward differentiation formulas. These are

$$\text{given by } \sigma(\zeta) = \zeta^k, \quad \frac{\rho}{\sigma}(\zeta) = [1 - \zeta^{-1}] + \frac{1}{2}[1 - \zeta^{-1}]^2 + \dots + \frac{1}{k}[1 - \zeta^{-1}]^k.$$

Hence

$$\psi(\tau) = 2(1 - \cos \tau) \text{Re} \left\{ 1 + \dots + \frac{1}{k} [1 - e^{i\tau}]^{k-1} \right\} \geq 2(1 - \cos \tau) \lambda_k,$$

For  $k = 1, 2$  we have  $\lambda_k = 1$  and for  $k = 3, 4$  we can take  $\lambda_3 = 5/6$ ,  $\lambda_4 = 1/3$ , but for  $k = 5, 6$  a positive  $\lambda$  does not exist, take  $\tau = \pi/2$ . Finally, note that (4.5) implies the root condition on  $\rho$  and  $A_0$ -stability [2] on  $(\rho, \sigma)$ .

REMARK 4.1. Gradient nonlinearities are considered also in [12].

The stability property studied requires the existence of  $q = q(\lambda)$  for  $\lambda > 0$  such that the frequency condition

$$\operatorname{Re} \left\{ [1 + q(1 - e^{-i\tau})] \frac{\rho}{\sigma}(e^{-i\tau}) \right\} + \lambda > 0 \quad \text{for } \tau \in [-\pi, \pi]$$

holds.  $\square$

We shall give an application of Theorem 3.1 in the special case  $A = \partial\varphi$ . Assume  $\varphi \geq 0$  and  $K = \{v \in H \mid \varphi(v) = 0\} \neq \emptyset$ .

COROLLARY 4.2. Assume that  $(\rho, \sigma)$  is A-stable with  $\sigma^* < 1$ . If (4.1) holds, then for  $v_0 \in K$

$$h \sum_{n=0}^N \varphi(u_n) \leq C \left\{ \max_{0 \leq j < k} [|u_j - v_0| + h|Au_j|] + h \sum_{j=k}^N |f_j| \right\}^2. \quad (4.10)$$

In particular, for the homogeneous equation the method generates a minimizing sequence. When the method is not A-stable, we have to assume that  $\partial\varphi$  is Lipschitz-continuous:

$$|Au - Av| \leq M|u - v| \quad \text{for } u, v \in H.$$

COROLLARY 4.3. Assume that (3.1), (3.2) hold with  $a < 0$ ,  $b = 0$ . Take  $h$  such that

$$0 < h < (2|a|M)^{-1}, \quad (4.11)$$

then (4.10) holds for  $v_0 \in K$ .  $\square$

The gradient method generates a minimizing sequence for  $0 < h < 2/M$  (with fixed  $h$ , see Chapter 4.5 in [6]) although (4.11) requires  $0 < h < 1/M$  on the Euler method.

Proof of Corollary 4.3. (proof of Corollary 4.2 is similar).

Take  $v_0 \in K$ . Then (2.2) yields

$$h^{-1}\rho(E)(u_n - v_0) + \sigma(E)Au_n = f_{n+k}, \quad n \in \mathbb{Z}, \quad (4.12)$$

where we assume,  $f_n$  is defined for  $n < k$  by the choice  $u_n = v_0$  for  $n < 0$ . We shall apply Theorem 3.1 to (4.12), therefore we need to check (2.6) in the form

$$\langle u + ahAu - v, Au \rangle \geq 0 \quad \text{for } u \in H, v \in K. \quad (4.13)$$

Since for all  $u, x$  in  $H$  we have

$$0 \leq \varphi(u+x) = \varphi(u) + \int_0^1 \langle A(u+tx), x \rangle dt \leq \varphi(u) + \langle Au, x \rangle + (1/2)M|x|^2,$$

we get, choosing  $x = (-1/2M)Au$ ,

$$\varphi(u) \geq (1/2M)|Au|^2, \quad \text{for } u \in H. \quad (4.14)$$

Combining (4.1) and (4.14) we get

$$\langle u + ahAu - v, Au \rangle \geq (1 - 2|a|hM)\varphi(u), \quad (4.15)$$

which, together with (4.11), implies (4.13). From (4.12) we obtain, with  $\Gamma = [\sigma(E) - a\rho(E)]^{-1}\rho(E)$ ,

$$\begin{aligned} & h^{-1} \sum_{n=0}^N \langle u_n + ahAu_n - v_0, \Gamma(u_n + ahAu_n - v_0) \rangle + \\ & + \sum_{n=0}^N \langle u_n + ahAu_n - v_0, Au_n \rangle = \\ & = \sum_{n=0}^N \langle u_n + ahAu_n - v_0, [\sigma(E) - a\rho(E)]^{-1} f_{n+k} \rangle. \end{aligned} \quad (4.16)$$

Since  $b = 0$  the first sum in (4.16) is nonnegative. For the second term  $S_2$  we get using (4.15)

$$S_2 \geq (1 - 2|a|hM) \sum_{n=0}^N \varphi(u_n).$$

To estimate the third sum observe that (3.6) (with  $\theta = 1$ ) and the Lipschitz-continuity of  $A$  imply

$$|u_n + ahAu_n - v_0| \leq C \left\{ \max_{0 \leq j < k} [ |u_j - v_0| + |Au_j| ] + \sum_{j=k}^n |f_j| \right\}.$$

Finally,

$$\left\{ \sum_{n=0}^N |[\sigma(E) - a\rho(E)]^{-1} f_{n+k}|^2 \right\}^{1/2} \leq \sqrt{\left\{ \sum_{n=0}^N |f_n|^2 \right\}}.$$

and (4.10) follows.  $\square$

Monotonicity properties are preserved under Galerkin processes. The next example shows that this can also happen when discretizing differential operators.

EXAMPLE 4.1. Let  $a$  and  $b$  be nondecreasing locally Lipschitz-continuous functions  $\mathbb{R} \rightarrow \mathbb{R}$  with  $b(0) = 0$ . We shall consider the discretization of the space variable in the initial value problem

$$\begin{cases} \frac{\partial}{\partial t} u(t, x) - \frac{\partial}{\partial x} a\left(\frac{\partial}{\partial x} u(t, x)\right) + b(u(t, x)) = f(t), & t > 0 \\ u(0, x) = u_0(x), & x \in \mathbb{R}. \end{cases} \quad (4.17)$$

Put  $Au^v = -[a(u^{v+1} - u^v) - a(u^v - u^{v-1})]$  and  $Bu^v = b(u^v)$ , where  $\{u^v\} \in \ell^2$ . Since  $a$  is locally Lipschitz-continuous,  $A$  maps  $\ell^2$  in  $\ell^2$  and is hemi-continuous, i.e.  $A((1-t)x + ty) \rightarrow Ax$  as  $t \rightarrow 0$  for all  $x, y \in \ell^2$ . By Prop. 2.4 in [1]  $A$  is maximally monotone if it is monotone. We shall see that it is even cyclically monotone, that is, for all  $x_0, x_1, \dots, x_n$  with  $x_n = x_0$  one has

$$\sum_{i=1}^n \langle x_i - x_{i-1}, Ax_i \rangle \geq 0.$$

Take  $x_0, x_1, \dots, x_n = x_0$  such that  $\text{supp } x_i$  is compact for all  $i$ .

Put  $y_i^m = x_i^m - x_{i-1}^m$ , then

$$\begin{aligned} \sum_{i=1}^n \langle x_i - x_{i-1}, Ax_i \rangle &= \sum_{i=1}^n \sum_{m=-\infty}^{\infty} (y_i^m - y_{i-1}^m) a(y_i^m) = \\ &= \sum_{m=-\infty}^{\infty} \sum_{i=1}^n (y_i^m - y_{i-1}^m) [a(y_i^m) - a(y_n^m)] \geq 0 \end{aligned}$$

since the sum

$$\sum_{i=1}^n (y_i^m - y_{i-1}^m) [a(y_i^m) - a(y_n^m)]$$

is nonnegative for all  $m$  (assume that  $y_{i-1}^m \leq y_i^m$  for  $i < n$  and use the fact that  $a$  is nondecreasing). Hence  $A = \partial\phi$ , where  $\phi$  is a convex functional, namely

$$\phi(u) = \int_0^1 \langle \dot{A}tu, u \rangle dt.$$

Similarly one could consider higher order approximations

$$Au^v = \sum_{j=1}^J a_j (-1)^j \nabla^j a(E^j v^j u^v)$$

where  $a_j \geq 0$ . The treatment of  $B$  is obvious. Typically,  $a$  might be of the form  $a(r) = |r|^\alpha r$  for some  $\alpha > 0$ , see p.155 in [7].

REF E R E N C E S

- [1] Brezis, H.: *Operators maximaux monotones et semi-groupes de contraction dans les espaces de Hilbert*,  
North Holland, Amsterdam 1973.
- [2] Cryer, C.W.: *A new class of highly stable methods: A<sub>0</sub>-stable methods*,  
BIT 13, 153-159 (1973).
- [3] Dahlquist, G.: *Stability and error bounds in the numerical integration of ordinary differential equations*,  
Kungl.Tekn.Högsk.Handl. Stockholm, No. 130 (1959).
- [4] Dahlquist, G.: *Error analysis for a class of methods for stiff non-linear initial value problems*,  
Springer Lect. Not.Math. 506, 60-74, Berlin-Heidelberg-New York 1976.
- [5] Dahlquist, G., Nevanlinna, O.:  *$\ell_2$ -estimates of the error in the numerical integration of non-linear differential systems*,  
Dept. of Comp.Sci., Royal Institute of Technology, Report TRITA-NA-7607,  
(1976).
- [6] Daniel, J.W.: *The approximate minimization of functionals*,  
Prentice-Hall Inc., New-Jersey, 1971.
- [7] Lions, J.L.: *Quelques méthodes de résolution des problèmes aux limites non linéaires*,  
Dunod, Gauthier-Villars, Paris 1969.
- [8] Lošinskij, S.M.: *Error estimation in the numerical integration of ordinary differential equations, Part I*, (Russian),  
Izvestia Vyssich Zavedenij Matematika, 1958, No 5(6).
- [9] Nevanlinna, O.: *On error bounds for G-stable methods*,  
BIT 16, 79-84 (1976).
- [10] Nevanlinna, O.: *On the numerical solution of some Volterra equations on infinite intervals*,  
Institut Mittag-Leffler, Report No 2, 1976.
- [11] Nevanlinna, O.: *On multistep methods for nonlinear initial value problems with an application to minimization of convex functionals*,  
Inst. of Math. Helsinki Univ. of Tech., Report HTKK-MAT-A76 (1976).
- [12] Odeh, F., Liniger, W.: *Nonlinear fixed-h stability of linear multistep formulae*,  
IBM Res. Rep. RC 5717, November 11, 1975.
- [13] Uhlmann, W.: *Fehlerabschätzungen bei Anfangswertaufgaben gewöhnlicher Differentialgleichungssystems 1. Ordnung*,  
Zs.Angew.Math.Mech., 37, 88-99 (1957).