RAIRO
ANALYSE NUMÉRIQUE

PHILIP BRENNER
MICHEL CROUZEIX
VIDAR THOMÉE

Single step methods for inhomogeneous linear differential equations in Banach space

RAIRO – Analyse numérique, tome 16, n° 1 (1982), p. 5-26.
<http://www.numdam.org/item?id=M2AN_1982__16_1_5_0>

R A I R O Analyse numerique/Numerical Analysis
(vol 16, nº 1, 1982, p 5 a 26)

SINGLE STEP METHODS FOR INHOMOGENEOUS LINEAR DIFFERENTIAL EQUATIONS IN BANACH SPACE (*)

by Philip BRENNER (1), Michel CROUZEIX (2) and Vidar THOMÉE (1)

Abstract — *Considering the initial-value problem for the differential equation*

$$u(t) = Au(t) + f(t)$$

in a Banach space X, where A generates a bounded semigroup on X, we analyze single step discretizations of the form

$$u_{n+1} = r(kA)\, u_n + k \sum_{j=1}^{m} q_j(kA)\, f(nk + \tau_j k),$$

where k is the step size, r, q_1, , q_m are rational functions, bounded for Re $z \leqslant 0$, *and τ_j are quadrature points in* [0, 1]

Resume — *On considere le probleme aux conditions initiales pour l equation differentielle*

$$u'(t) = Au(t) + f(t)$$

dans un espace de Banach X, ou A engendre un semi-groupe borne sur X, et on analyse des discretisations a un pas du type

$$u_{n+1} = r(kA)\, u_n + k \sum_{j=1}^{m} q_j(kA)\, f(nk + \tau_j k),$$

ou k est le pas, r, q_1, , q_m sont des fonctions rationnelles, bornees pour Re $z \leqslant 0$, *et τ_j sont des points de quadrature sur* [0, 1]

1. INTRODUCTION

Let X be a Banach space and assume that A is a closed, densely defined linear operator on X which generates a bounded semigroup $E(t) = e^{tA}$ on X The solution of the initial value problem (with $u_t = du/dt$)

$$u_t = Au + f \quad \text{for} \quad t \geqslant 0\,, \qquad u(0) = v\,, \tag{1 1}$$

(*) Reçu le 19 mai 1981

(1) Department of Mathematics Chalmers University of Technology, S-412 96 Goteborg, Suede

(2) Laboratoire d'Analyse Numerique, Universite de Rennes, B P 25A, 35031 Rennes Cedex

R A I R O Analyse numerique/Numerical Analysis, 0399-0516/1982/85/$ 5 00

may then be expressed by means of Duhamel's principle as

$$u(t) = E(t)\, v + \int_0^t E(t - s)\, f(s)\, ds\,.$$

We shall be concerned with the discretization in time of the problem (1.1). For this purpose, let k be a small positive time increment and let $r, q_1, ..., q_m$ be rational functions which are bounded for $\mathrm{Re}\, z \leqslant 0$. Then, since A has its spectrum in $\mathrm{Re}\, z \leqslant 0$, $r(kA)$ and $q_j(kA)$ are well defined, and we may seek an approximate solution u_n at $t_n = nk$ of (1.1) by the recursion relation

$$u_{n+1} = E_k\, u_n + k(Q_k\, f)\,(t_n)\,, \quad n = 0, ..., \quad \text{with} \quad u_0 = v\,, \tag{1.2}$$

where

$$E_k = r(kA)\,, \quad (Q_k\, f)\,(t) = \sum_{j=1}^{m} q_j(kA)\, f(t + \tau_j\, k)\,,$$

with $\{\, \tau_j \,\}_1^m$ distinct quadrature points, for simplicity in [0, 1].

In order to express the degree of approximation of (1.2) we consider first the case when A is a bounded operator. We say that the scheme is accurate of order p if for any choice of f and v, with f sufficiently regular with respect to t, we have

$$\rho_n = u(t_{n+1}) - E_k\, u(t_n) - k(Q_k\, f)\,(t_n) = 0(k^{p+1}) \quad \text{as} \quad k \to 0\,, \tag{1.3}$$

that is, if the solution of (1.1) satisfies (1.2) with an error of order $0(k^{p+1})$. This will entail certain relations between r, the q_j and the τ_j which may be stated, for instance, in the form

$$r(z) - e^z = 0(z^{p+1}) \quad \text{as} \quad z \to 0\,, \tag{1.4}$$

and for $l = 0, ..., p - 1$,

$$\frac{l\,!}{z^{l+1}}\left(r(z) - \sum_{j=0}^{l} \frac{z^j}{j\,!}\right) - \sum_{j=1}^{m} \tau_j^l\, q_j(z) = 0(z^{p-l}) \quad \text{as} \quad z \to 0\,. \tag{1.5}$$

We observe that the global error $e_n = u(t_n) - u_n$ satisfies

$$e_{n+1} = E_k\, e_n + \rho_n\,, \quad n = 0, ..., \quad \text{with} \quad e_0 = 0\,.$$

Assuming that E_k is stable in X we shall hence be able to infer a $0(k^p)$ global error estimate from the local estimate (1.3).

We shall then turn to the case when A is an unbounded operator and discover

that in order for the analysis to yield an estimate of the form (1.3) we need to assume that u, in addition to being smooth in t, will have to satisfy relations like $u^{(l)}(t) \in D(A^{p+1-l})$ for $l = 0, \ldots, p+1$. In applications to partial differential equations, this often demands not only smothness of $u^{(l)}$ in the space variables, but also that these functions satisfy certain boundary conditions which are not natural to impose for $l = 0, \ldots, p-1$. In order to be able to avoid artificial assumptions on the solution we shall consider schemes which satisfy a stronger accuracy assumption than (1.3) : We say that the scheme is strictly accurate of order $p_0 \leqslant p$ if the truncation error vanishes for all f and v such that the solution is a polynomial in t of degree at most $p_0 - 1$. It will turn out that this condition is equivalent to demanding that the first p_0 relations in (1.5) hold with right hand sides replaced by zero. For schemes which are strictly accurate of order p, or under a not very restrictive additional condition, accurate of order p and strictly accurate of order $p-1$, we shall then be able to show the desired global error estimate.

The details of the above analysis are given in Section 2 below, where the estimates are expressed in terms of the solution u of (1.1), and in Section 3 where error bounds in terms of the data f and v are presented.

Our main motivation for this study is the application to numerical methods for partial differential equations, and to the situation when discretization also takes place in the space variables. It may be the case, for instance, that an initial value problem has been approximated in space by means of the finite element method, leaving us with a semidiscrete problem of the form (1.1), with A, f, and v replaced by A_h, f_h, v_h, depending on the small parameter h, and such that the error in the solution of this problem is bounded by ε_h. In order to produce a completely discrete method, the above scheme may now be applied to the semidiscrete problem and one hopes that the total error will be $O(\varepsilon_h + k^p)$. An example of a result of this nature is shown in Section 4 below.

The construction of schemes satisfying our above assumptions is the object of Section 5. It is seen, among other things, that if $\{\tau_j\}_1^m$ are chosen as the Gaussian points of order m in [0, 1], the scheme may be constructed to be accurate of order $2m$, but with no choice of $\{\tau_j\}_1^m$ can it be strictly accurate of order $m+2$.

A study which has many points in common with the present one was carried out in Crouzeix [2] in the context of Runge-Kutta methods in a Hilbert space.

Conditions for stability of operators of the form $E_k = r(kA)$ in general Banach spaces were discussed in Brenner and Thomée [1]. The present results together with those of [1] thus allow application to completely discrete schemes obtained from semidiscrete approximations with known error bounds in, say, the maximum norm.

2. ERROR ESTIMATES IN TERMS OF THE SOLUTION

We begin this section by deducing the conditions for the time discretization scheme (1.2) to be accurate of order p and strictly accurate of order $p_0 \leqslant p$, in the sense introduced above. Assuming thus A to be a bounded operator we obtain by Taylor series development of ρ_n with respect to k, for u and f sufficiently smooth with respect to t,

$$\rho_n = \sum_{l=0}^{p} \frac{k^l}{l!} u^{(l)}(t_n) - r(kA)\, u(t_n) - k \sum_{j=1}^{m} q_j(kA) \sum_{l=0}^{p-1} \frac{(\tau_j k)^l}{l!} f^{(l)}(t_n) + R_{n,p}$$

where

$$R_{n,p} = \int_{t_n}^{t_{n+1}} \frac{(t_{n+1} - s)^p}{p!} u^{(p+1)}(s)\, ds - k \sum_{j=1}^{m} q_j(kA) \int_{t_n}^{t_n + k\tau_j} \frac{(t_n + k\tau_j - s)^{p-1}}{(p-1)!} f^{(p)}(s)\, ds . \quad (2.1)$$

Using the fact that $f^{(l)} = u^{(l+1)} - Au^{(l)}$ this may be written

$$\rho_n = \sum_{l=0}^{p} \frac{k^l}{l!} h_l(kA)\, u^{(l)}(t_n) + R_{n,p}$$

where

$$h_0(z) = 1 - r(z) + z \sum_{j=1}^{m} q_j(z) ,$$

$$h_l(z) = 1 - l \sum_{j=1}^{m} \tau_j^{l-1} q_j(z) + z \sum_{j=1}^{m} \tau_j^l q_j(z) , \quad \text{for} \quad 1 \leqslant l \leqslant p-1 ,$$

and

$$h_p(z) = 1 - p \sum_{j=1}^{m} \tau_j^{p-1} q_j(z) .$$

Since $R_{n,p} = 0(k^{p+1})$ for small k, this representation of the truncation error immediately yields the following lemma.

LEMMA 1 : *The scheme* (1.2) *is accurate of order p if and only if*

$$h_l(z) = 0(z^{p+1-l}) \quad \textit{for} \quad l = 0, ..., p . \quad (2.2)$$

It strictly accurate of order $p_0 \leqslant p$ *if and only if*

$$h_l(z) = 0 \quad \textit{for} \quad l = 0, ..., p_0 - 1 . \quad (2.3)$$

We shall now turn to the error estimates, allowing A to be unbounded. For this purpose we shall first briefly discuss the representation of functions of A in terms of the semigroup $E(t)$.

Let $\tilde{M}$ denote the set of Laplace transforms of bounded measures on R_+,

$$g(z) = \tilde{\mu}(z) = \int_{R_+} e^{zt}\, d\mu(t)\,,$$

and recall (*cf.* [1]) that with A the generator of a bounded semigroup $E(t)$ on X, $g(A)$ may be represented as

$$g(A) = \int_{R_+} E(t)\, d\mu(t)\,.$$

Noting that μ is uniquely determined by g, we may set

$$m(g) = \int d\,|\,\mu\,|\,(t)\,,$$

and obtain that $g(kA)$ is a bounded operator on X, and for any $k > 0$,

$$\|\, g(kA)\,\| \leqslant \int_{R_+} \|\, E(kt)\,\|\, d\,|\,\mu\,|\,(t) \leqslant C_0\, m(g)\,, \quad \text{if} \quad \|\, E(t)\,\| \leqslant C_0\,.$$

Any rational function g which is bounded for $\operatorname{Re} z \leqslant 0$ belongs to $\tilde{M}$, as is seen by expansion into partial fractions. In this case $d\mu$ has the form $\Sigma_j\, p_j(t)\, e^{-\lambda_j t}\, dt$ where λ_j are the poles of g and p_j are polynomials; the above representation of $g(A)$ then reduces to the standard formulas for the resolvent of A and its powers. Note also that if $f, g \in \tilde{M}$ then $fg \in \tilde{M}$ and

$$(fg)\,(A) = f(A)\, g(A)\,.$$

Further, if $f, g \in \tilde{M}$ and $g(z) = f(z)\, z^l$ then $g(A)\, v = f(A)\, A^l\, v$ for $v \in D(A^l)$ (*cf.* Lemma 4 in [1]).

In particular, if (2.2) holds we have

$$\tilde{h}_l(z) = z^{-(p+1-l)}\, h_l(z) \in \tilde{M} \quad \text{for} \quad l = 0, \ldots, p\,,$$

and if (2.3) is satisfied as well, we may write the truncation error from (1.3) as

$$\rho_n = k^{p+1} \sum_{l=p_0}^{p} \frac{1}{l\,!}\, \tilde{h}_l(kA)\, A^{p+1-l}\, u^{(l)}(t_n) + R_{n,p}\,, \tag{2.4}$$

provided $u^{(l)}(t_n) \in D(A^{p+1-l})$. In order to estimate this expression we use the relation

$$k\varphi(t_n) = \int_{t_n}^{t_{n+1}} (\varphi(s) - (t_{n+1} - s)\,\varphi'(s))\,ds\,,$$

to obtain for $l = p_0, \ldots, p$,

$$k \,\|\, \tilde{h}_l(kA)\, A^{p+1-l}\, u^{(l)}(t_n) \,\| \leqslant$$
$$\leqslant \int_{t_n}^{t_{n+1}} (\| \tilde{h}_l(kA)\, A^{p+1-l}\, u^{(l)}(s) \| + \| kA\tilde{h}_l(kA)\, A^{p-l}\, u^{(l+1)}(s) \|)\,ds\,,$$

and hence since $\tilde{h}_l(z)$ and also $z\tilde{h}_l(z)$ belong to $\tilde{M}$,

$$k \,\|\, \tilde{h}_l(kA)\, A^{p+1-l}\, u^{(l)}(t_n) \,\| \leqslant$$
$$\leqslant C \int_{t_n}^{t_{n+1}} (\| A^{p+1-l}\, u^{(l)}(s) \| + \| A^{p-l}\, u^{(l+1)}(s) \|)\,ds\,.$$

For the remainder term $R_{n,p}$ we have at once from (2.1),

$$\| R_{n,p} \| \leqslant Ck^p \int_{t_n}^{t_{n+1}} (\| u^{(p+1)}(s) \| + \| f^{(p)}(s) \|)\,ds$$
$$\leqslant Ck^p \int_{t_n}^{t_{n+1}} (\| u^{(p+1)}(s) \| + \| Au^{(p)}(s) \|)\,ds\,,$$

and hence altogether

$$\| \rho_n \| \leqslant Ck^p \sum_{l=p_0}^{p+1} \int_{t_n}^{t_{n+1}} \| A^{p+1-l}\, u^{(l)}(s) \|\,ds\,. \tag{2.5}$$

We may now easily prove the following.

THEOREM 1 : *Assume that the scheme* (1.2) *is accurate of order* p *and strictly accurate of order* p_0 *and let* E_k *be stable in* X. *Then if* $u^{(l)} \in L^1(0, t_n; D(A^{p+1-l}))$ *for* $l = p_0, \ldots, p+1$ *we have*

$$\| u(t_n) - u_n \| \leqslant Ck^p \sum_{l=p_0}^{p+1} \int_0^{t_n} \| A^{p+1-l}\, u^{(l)}(s) \|\,ds\,.$$

Proof : Setting $e_n = u(t_n) - u_n$ we have since $e_0 = 0$,

$$e_n = \sum_{j=0}^{n-1} E_k^{n-1-j} \rho_j ,$$

and hence by the stability of E_k,

$$\| e_n \| \leqslant C \sum_{j=0}^{n-1} \| \rho_j \| .$$

The result is now an immediate consequence of (2.5).

Note that the error estimate of Theorem 1 requires $u^{(l)} \in D(A^{p+1-l})$ for $l = p_0, ..., p+1$ and t positive. In applications to initial-boundary value problems for partial differential equations this generally demands not only smoothness of the solution but also that its time derivatives satisfy certain boundary conditions. Although it is appropriate to assume $u^{(p)} \in D(A)$, the above conditions for $l < p$ are undesirable and the case $p_0 = p$ is therefore of special interest.

In our next result we shall show an optimal order error estimate without requiring artificial boundary conditions if the scheme is strictly accurate of order $p - 1$, only, but satisfies the additional condition

$$\sigma(z) = h_{p-1}(z)/(z(1 - r(z))) \in \tilde{M} . \tag{2.6}$$

Since $r(z) = 1 + z + 0(z^2)$ for small z, it follows in particular that (2.6) holds if $r(iy) \neq 1$ for $0 \neq y \in R \cup \{ \infty \}$, or more generally, if $r(iy) \neq 1$ for $0 \neq y \in R$ and $q_j(z) = 0(|z|^{-l})$ and $(r(z) - 1)^{-1} = 0(|z|^l)$ for large z and some $l \geqslant 0$. For example, this condition is satisfied for the first and second subdiagonal Padé approximations $r_{jk}(z)$, $0 \leqslant j - k \leqslant 2$, to e^z (*cf.* [1], p. 687), and also for the diagonal approximations $r_{11}(z)$ and $r_{22}(z)$, but is not valid for $r_{33}(z)$ as a simple computation shows. Also (2.6) will be fulfilled for schemes employing the restricted Padé approximants $R_k(z)$ of orders $k = 1, 2$, and 3 (*cf.* [1], p. 688).

THEOREM 2 : *Assume that the scheme* (1.2) *is accurate of order p, and strictly accurate of order* $p - 1$, *that* (2.6) *holds and that* E_k *is stable in X. Then under the appropriate regularity assumptions,*

$$\| u(t_n) - u_n \| \leqslant Ck^p \left\{ \| Au^{(p-1)}(0) \| + \int_0^{t_n} (\| Au^{(p)}(s) \| + \| u^{(p+1)}(s) \|) \, ds \right\} .$$

Proof : We have from (2.4),

$$\rho_n = \frac{k^{p+1}}{(p-1)!} \tilde{h}_{p-1}(kA) A^2 u^{(p-1)}(t_n) + \frac{k^{p+1}}{p!} \tilde{h}_p(kA) Au^{(p)}(t_n) + R_{n,p} .$$

The contribution to the global error of the last two terms is bounded as indicated in Theorem 1 with $p_0 = p$, and it remains to estimate

$$S_n = \sum_{j=0}^{n-1} E_k^{n-1-j} \frac{k^{p+1}}{(p-1)!} \tilde{h}_{p-1}(kA)\, A^2\, u^{(p-1)}(t_j)\,.$$

By the definition of σ we have $kA\tilde{h}_{p-1}(kA) = \sigma(kA)\,(I - E_k)$ and hence

$$S_n = \frac{k^p}{(p-1)!}\,\sigma(kA) \sum_{j=0}^{n-1} E_k^{n-1-j}(I - E_k)\, Au^{(p-1)}(t_j)$$

$$= \frac{k^p}{(p-1)!}\,\sigma(kA) \times$$

$$\times \left\{ Au^{(p-1)}(t_{n-1}) - \sum_{j=1}^{n-1} E_k^{n-j} \int_{t_{j-1}}^{t_j} Au^{(p)}(s)\, ds - E_k^n\, Au^{(p-1)}(0) \right\}.$$

We conclude

$$\| S_n \| \leqslant Ck^p \left\{ \| Au^{(p-1)}(t_n) \| + \| Au^{(p-1)}(0) \| + \int_0^{t_{n-1}} \| Au^{(p)}(s) \|\, ds \right\}$$

which is bounded as desired.

In the case with order of strict accuracy $p_0 \leqslant p - 2$ it is impossible in general to infer a $0(k^p)$ global error estimate without making assumptions of the type $u^{(l)} \in D(A^{p-l})$. Consider for example the problem

$$u_t = Au + p_0\, t^{p_0-1}\, w - t^{p_0}\, Aw \quad \text{for} \quad t \geqslant 0\,, \quad u(0) = 0\,,$$

with the solution $u(t) = t^{p_0} w$. Then

$$u(t_1) - u_1 = \rho_1 = k^{p+1}\, \tilde{h}_{p_0}(kA)\, A^{p+1-p_0}\, w\,.$$

In order to have $\rho_1 = 0(k^p)$ we need $kA\tilde{h}_{p_0}(kA)\, A^{p-p_0}\, w$ to be bounded. This is the case if $w \in D(A^{p-p_0})$ but not necessarily so if w is slightly less regular. To demonstrate this, let X be a Hilbert space and let $-A$ be self-adjoint, positive definite, and unbounded. We shall show that $\rho_1 = 0(k^p)$ implies $w \in D((-A)^{p-p_0-\varepsilon})$ for any $\varepsilon > 0$, and thus that if w fails to satisfy this requirement for some small positive ε, then optimal order convergence cannot take place. Let φ_j be the eigenfunctions and λ_j the corresponding eigenvalues of $-A$, and let $w = \Sigma_j\, c_j\, \varphi_j \in X$. Since $\tilde{h}_{p_0}$ is a rational function which does not vanish identically there exist positive γ and c such that $| x\tilde{h}_{p_0}(x) | \geqslant c$ for

$\gamma/2 \leqslant x \leqslant \gamma$ and hence

$$C \geqslant \| kA\tilde{h}_{p_0}(kA)\, A^{p-p_0} w \| = \Sigma_j \mid k\lambda_j\, \tilde{h}_{p_0}(k\lambda_j)\, \lambda_j^{p-p_0} c_j \mid^2$$
$$\geqslant c^2 \sum_{\gamma/2 \leqslant k\lambda_j \leqslant \gamma} \mid \lambda_j^{p-p_0} c_j \mid^2 .$$

Hence for $k = 2^{-(l+1)}$,

$$\sum_{2^l\gamma \leqslant \lambda_j \leqslant 2^{l+1}} \mid \lambda_j^{p-p_0-\varepsilon} c_j \mid^2 \leqslant C\, 2^{-2l\varepsilon} ,$$

so that

$$\Sigma_j \mid \lambda_j^{p-p_0-\varepsilon} c_j \mid^2 \leqslant \sum_{\lambda_j \leqslant \gamma} \mid \lambda_j^{p-p_0-\varepsilon} c_j \mid^2 + C \sum_{l=0}^{\infty} 2^{-2l\varepsilon} < \infty ,$$

which shows the desired conclusion $w \in D((-A)^{p-p_0-\varepsilon})$. In a similar way we can prove that $\rho_1 = 0(k^{p-\alpha})$ implies $w \in D((-A)^{p-p_0-\alpha-\varepsilon})$ for any $\varepsilon > 0$.

As a concrete example we may take $X = L_2(0, 1)$ and

$$u_t = u_{xx} + 2(tx(1-x) - t^2) \quad \text{in} \quad [0, 1] \times R_+ , \quad u(0) = 0 ,$$

with the exact solution $u(x, t) = t^2\, x(1-x)$. For $t > 0$ this function belongs to $D(A^q)$ only for $q < 5/4$, so with $p = 4$, $p_0 = 2$ we may not expect $0(k^4)$ convergence. In fact, with $w = x(1-x)$ we have $c_j \sim j^{-3}$ for j odd, and $c_j = 0$ for j even, and since $\lambda_j \sim j^2$ a simple calculation along the above lines shows $\| \rho_1 \| \geqslant ck^{13/4}$.

3. ERROR ESTIMATES IN TERMS OF DATA

Recall that the truncation error may be expressed in the form

$$\rho_n = k^{p+1} \sum_{l=0}^{p} \frac{1}{l!}\, \tilde{h}_l(kA)\, A^{p+1-l} u^{(l)}(t_n) + R_{n,p} , \tag{3.1}$$

where $R_{n,p}$ is defined by (2.1). For the purpose of estimating this in terms of the data of the problem we use the differential equation (1.1) to write

$$A^{p+1-l} u^{(l)} = u^{(p+1)} - \sum_{j=l}^{p} A^{p-j} f^{(j)} .$$

Inserting this into (3.1) we have

$$\rho_n = k^{p+1} \sum_{l=0}^{p-1} \frac{1}{l!}\, \tilde{\gamma}_l(kA)\, A^{p-l} f^{(l)}(t_n) + \tilde{R}_{n,p} , \tag{3.2}$$

where

$$\tilde{\gamma}_l(z) = -\, l\,! \sum_{j=0}^{l} \frac{1}{j\,!}\, \tilde{h}_j(z)\,, \quad l = 0, \ldots, p\,,$$

and

$$\tilde{R}_{n,p} = R_{n,p} - \frac{k^{p+1}}{p\,!}\, \tilde{\gamma}_p(kA)\,(f^{(p)}(t_n) - u^{(p+1)}(t_n))\,.$$

Before presenting precise bounds for the remainder term in terms of data we shall restate the accuracy conditions in terms of our newly introduced coefficients. Setting

$$\gamma_l(z) = z^{p-l}\, \tilde{\gamma}_l(z) \quad \text{for} \quad l = 0, \ldots, p\,,$$

we find easily from our definitions

$$\gamma_l(z) = \frac{l\,!}{z^{l+1}} \left(r(z) - \sum_{j=0}^{l} \frac{z^j}{j\,!} \right) - \sum_{j=1}^{m} \tau_j^l\, q_j(z) \quad \text{for} \quad l = 0, \ldots, p-1 \quad (3.3)$$

and

$$\gamma_p(z) = \frac{p\,!}{z^{p+1}} \left(r(z) - \sum_{j=0}^{p} \frac{z_j}{j\,!} \right).$$

We conclude at once from (3 3) and Lemma 1 the following result.

LEMMA 2 : *The scheme* (1.2) *is accurate of order p if and only if*

$$\gamma_l(z) = 0(z^{p-l}) \quad \textit{for} \quad l = 0, \ldots, p\,, \qquad (3.4)$$

and strictly accurate of order $p_0 \leqslant p$ *if and only if*

$$\gamma_l(z) = 0 \quad \textit{for} \quad l = 0, \ldots, p_0 - 1\,.$$

Note that for $l = p$ the condition (3.4) may also be written

$$r(z) = e^z + 0(z^{p+1}) \quad \text{as} \quad z \to 0\,. \qquad (3.5)$$

We have thus expressed the order of accuracy condition in the form stated in (1.4) and (1.5) of the introduction.

As a preparation for a global error estimate we shall now show that under our present assumptions,

$$\| \tilde{R}_{n,p} \| \leqslant Ck^{p+1} \left\{ \| \theta_p(kA)\, u^{(p+1)}(0) \| + \sup_{t_n \leqslant s \leqslant t_{n+1}} \| f^{(p)}(s) \| + \right.$$
$$\left. + Ck^p \int_{t_n}^{t_{n+1}} \int_0^s \| f^{(p+1)}(\sigma) \|\, d\sigma\, ds \right\} \quad (3.6)$$

where

$$\theta_p(z) = (e^z - r(z))/z^{p+1} .$$

Note that in view of (3.5), $\theta_p \in \tilde{M}$ so that in (3.6),

$$\| \theta_p(kA)\, u^{(p+1)}(0) \| \leqslant C \| u^{(p+1)}(0) \| . \tag{3.7}$$

and also that $u^{(p+1)}(0)$ may be expressed in terms of data; for $l \geqslant 0$ we have recursively $u^{(l+1)}(0) = Au^{(l)}(0) + f^{(l)}(0)$, with $u^{(0)}(0) = v$.

For the purpose of showing (3.6) we write $\tilde{R}_{n,p}$ as a sum of

$$\tilde{R}^1_{n,p} = -k \sum_{j=1}^{m} q_j(kA) \int_{t_n}^{t_n + k\tau_j} \frac{(t_n + k\tau_j - s)^{p-1}}{(p-1)!} f^{(p)}(s)\, ds - \frac{k^{p+1}}{p!} \tilde{\gamma}_p(kA) f^{(p)}(t_n)$$

and

$$\tilde{R}^2_{n,p} = \int_{t_n}^{t_{n+1}} \frac{(t_{n+1} - s)^p}{p!} u^{(p+1)}(s)\, ds - \frac{k^{p+1}}{p!} \tilde{\gamma}_p(kA)\, u^{(p+1)}(t_n) .$$

Since $\tilde{R}^1_{n,p}$ is obviously bounded as desired, it remains to consider $\tilde{R}^2_{n,p}$. We shall then use the fact that with $v_{p+1} = u^{(p+1)}(0)$,

$$u^{(p+1)}(t) = E(t)\, v_{p+1} + \int_0^t E(t-s) f^{(p+1)}(s)\, ds ,$$

to write

$$\tilde{R}^2_{n,p} = k^{p+1} E(t_n) \theta_p(kA)\, v_{p+1} + \int_{t_n}^{t_{n+1}} \frac{(t_n - s)^p}{p!} \int_0^s E(s-\sigma) f^{(p+1)}(\sigma)\, d\sigma\, ds +$$

$$+ \frac{k^{p+1}}{p!} \tilde{\gamma}_p(kA) \int_0^{t_n} E(t_n - s) f^{(p+1)}(s)\, ds ,$$

where the term in v_{p+1} has resulted from the simple identity

$$p!\, \theta_p(z) = \int_0^1 (1-\sigma)^p e^{\sigma z}\, d\sigma - \tilde{\gamma}_p(z) =$$

$$= k^{-p-1} \int_{t_n}^{t_{n+1}} (t_n - s)^p e^{(s-t_n)z/k}\, ds - \tilde{\gamma}_p(z) .$$

It now follows by obvious estimates that $\tilde{R}^2_{n,p}$ is bounded as stated.

We may now state the following result. Note again in the first estimate the artificial assumptions $f^{(l)} \in D(A^{p-l})$ for $p_0 < p$.

THEOREM 3 : *Assume that the scheme is accurate of order p, and strictly accurate of order* $p_0 \leqslant p$, *and let* E_k *be stable in* X. *Then under the appropriate regularity assumptions,*

$$\| u(t_n) - u_n \| \leqslant Ck^p \left\{ t_n \| \theta_p(kA)\, u^{(p+1)}(0) \| + t_n \sum_{l=p_0}^{p} \sup_{s \leqslant t_n} \| A^{p-l} f^{(l)}(s) \| + \right.$$
$$\left. + \int_0^{t_n} (t_n - s) \| f^{(p+1)}(s) \| \, ds \right\}. \quad (3.8)$$

If $p_0 = p - 1$ *and* (2.6) *holds we have*

$$\| u(t_n) - u_n \| \leqslant Ck^p \left\{ t_n \| \theta_p(kA)\, u^{(p+1)}(0) \| + \| f^{(p-1)}(0) \| + t_n \| f^{(p)}(0) \| + \right.$$
$$\left. + \int_0^{t_n} (t_n - s) \| f^{(p+1)}(s) \| \, ds \right\}. \quad (3.9)$$

Proof : The first estimate follows in a straightforward manner from the representation (3.2) for the truncation error, the estimate (3.6) for the remainder term, and the stability of E_k. In the latter case we have to estimate in addition,

$$S_n = k^{p+1} \sum_{j=0}^{n-1} E_k^{n-1-j}\, \tilde{\gamma}_{p-1}(kA)\, Af^{(p-1)}(t_j).$$

Since now $\tilde{\gamma}_{p-1}(z) = -\tilde{h}_{p-1}(z)$ we have as in the proof of Theorem 2,

$$\| S_n \| \leqslant Ck^p \left\{ \| f^{(p-1)}(0) \| + \int_0^{t_{n+1}} \| f^{(p)}(s) \| \, ds \right\},$$

which is bounded as desired.

Note again the inequality (3.7) bounding the first term on the right in (3.8) and (3.9). In fact, the proof of these inequalities without $\theta_p(kA)$ could be derived by a somewhat easier argument. In their present form they will be applied in Section 4.

4. TOTALLY DISCRETE SCHEMES

We shall briefly consider the application of our above results to the case when discretization also takes place with regard to the space X as would be the case when finite element approximations are used in the space variables. Thus let X_h be a family of finite dimensional spaces approximating X, with norms $\| \cdot \|_h$,

and assume that we are given corresponding linear operators $P_h : X \to X_h$ with

$$\| P_h v \|_h \leqslant C \| v \| \quad \forall v \in X ,$$

where $P_h v$ is thought of as an approximation of v. It could, for instance, be the case that X_h is a subspace of X, that $\| \cdot \|_h = \| \cdot \|$, and that P_h is a projection operator such that, with Y a dense subspace of X with norm $\| \cdot \|_Y$,

$$\| P_h v - v \| \leqslant \varepsilon_h \| v \|_Y \quad \forall v \in Y , \tag{4.1}$$

where ε_h is small with h. In applications to isoparametric finite elements one might have $X = L_2(\Omega)$ for some $\Omega \subset R^n$ and $X_h = L_2(\Omega_h)$ where Ω_h is an approximation to Ω, in which case $P_h v$ would approximate v in $\Omega \cap \Omega_h$.

Assume also that we are given approximations $A_h : X_h \to X_h$ of A which generate uniformly bounded semigroups e^{tA_h} on X_h. We may then consider the semidiscrete problem to find $u_h : [0, \infty) \to X_h$ such that

$$\frac{du_h}{dt} = A_h u_h + P_h f \quad \text{for} \quad t \geqslant 0 , \quad u_h(0) = P_h v , \tag{4.2}$$

and pose the corresponding completely discrete problem by application of our scheme (1.2), namely

$$\begin{aligned} u_{h,n+1} &= E_{kh} u_{h,n} + k(Q_{kh} P_h f)(t_n) \quad \text{for} \quad n = 0, 1, \ldots , \\ u_{h,0} &= P_h v , \end{aligned} \tag{4.3}$$

where

$$E_{kh} v_h = r(kA_h) v_h , \quad (Q_{kh} f_h)(t) = \sum_{j=1}^{m} q_j(kA_h) f_h(t_n + k\tau_j) .$$

Our purpose is now to show that under the appropriate regularity assumptions the combined error from both discretizations is $0(\varepsilon_h + k^p)$. In order to do so we shall need an assumption concerning the choice of A_h which is satisfied in typical applications. We introduce the « elliptic projection » $Q_h : Y \to X_h$ by

$$Q_h v = (I - A_h)^{-1} P_h (I - A) v , \tag{4.4}$$

which exists for $v \in D(A)$ since A_h generates a bounded semigroup. We also assume that, *cf.* (4.1),

$$\| Q_h v - P_h v \|_h \leqslant \varepsilon_h \| v \|_Y \quad \forall v \in D(A) \cap Y ,$$

and that the exact solution of (1.1) belongs to $C^1(0, T; Y)$ for any $T > 0$. Under these assumptions we shall prove an analogue of Theorem 2; a counterpart of Theorem 1 can be similarly derived.

THEOREM 4 : *Under the assumptions of Theorem 2 for the time discretization scheme (1.2) and under the present assumptions on the discretization in X, we have, if E_{kh} is uniformly stable in X, that under the appropriate regularity assumptions*

$$\| u_{h,n} - P_h u(t_n) \|_h \leqslant C\varepsilon_h \left\{ (1 + t_n) \sup_{s \leqslant t_n} \| u(s) \|_Y + t_n \sup_{s \leqslant t_n} \| u'(s) \|_Y \right\} +$$

$$+ Ck^p \left\{ \| (I - A) u^{(p-1)}(0) \| + \int_0^{t_n} (\| (I - A) u^{(p)}(s) \| + \| (I - A) u^{(p+1)}(s) \|) \, ds \right\}. \tag{4.5}$$

Proof : We find easily for the solution of the continuous problem

$$Q_h u'(t) - A_h Q_h u(t) = \tilde{f}_h(t) = P_h f(t) + (Q_h - P_h)(u'(t) - u(t))$$

where

$$\| \tilde{f}_h(t) - P_h f(t) \|_h \leqslant 2 \varepsilon_h \| u'(t) - u(t) \|_Y .$$

Considering the time discretized version of the equation satisfied by $Q_h u$, namely

$$\tilde{u}_{h,n+1} = E_{kh} \tilde{u}_{h,n} + k(Q_{kh} \tilde{f}_h)(t_n) \quad \text{for} \quad n = 0, 1, \ldots$$
$$\tilde{u}_{h,0} = Q_h v ,$$

we have by Theorem 2,

$$\| \tilde{u}_{h,n} - Q_h u(t_n) \|_h \leqslant Ck^p \left\{ \| A_h Q_h u^{(p-1)}(0) \|_h + \right.$$
$$\left. + \int_0^{t_n} (\| A_h Q_h u^{(p)}(s) \|_h + \| Q_h u^{(p+1)}(s) \|_h) \, ds \right\}.$$

Here

$$\| A_h Q_h u \|_h = \| A_h (I - A_h)^{-1} P_h (I - A) u \|_h \leqslant C \| (I - A) u \| ,$$

and similarly

$$\| Q_h u \|_h \leqslant C \| (I - A) u \| ,$$

so that

$$\| \tilde{u}_{h,n} - Q_h u(t_n) \|_h \leqslant Ck^p \left\{ \| (I - A) u^{(p-1)}(0) \| + \right.$$

$$\left. + \int_0^{t_n} (\| (I - A) u^{(p)}(s) \| + \| (I - A) u^{(p+1)}(s) \|) \, ds \right\}.$$

On the other hand, since $r(kA_h)$ is uniformly stable in X_h, we find

$$\| \tilde{u}_{h,n} - u_{h,n} \|_h \leqslant$$

$$\leqslant C \left\{ \| P_h v - Q_h v \|_h + k \sum_{l=0}^{n-1} \sum_{j=1}^{m} \| q_j(kA_h) (\tilde{f}_h(t_l + k\tau_j) - P_h f(t_l + k\tau_j)) \|_h \right\}$$

$$\leqslant C\varepsilon_h \left\{ (1 + t_n) \sup_{s \leqslant t_n} \| u(s) \|_Y + t_n \sup_{s \leqslant t_n} \| u'(s) \|_Y \right\}.$$

Together with the estimate

$$\| Q_h u(t_n) - P_h u(t_n) \|_h \leqslant \varepsilon_h \| u(t_n) \|_Y ,$$

this completes the proof of the theorem.

Note that in the case that $X_h \subset X$, $\| . \|_h = \| . \|$, and that (4.1) holds, Theorem 4 immediately bounds $\| u_{h,n} - u(t_n) \|$ by the right hand side of (4.5).

As an alternative to the above treatment we shall now indicate an analysis which uses the error estimates for the time discretization in terms of data, given in Theorem 3, and which assumes given an error estimate for the semidiscrete homogeneous equation rather than the one for the elliptic projection.

Thus let again $\{ X_h \}$ be a family of finite dimensional spaces approximating X, let $P_h : X \to X_h$ be uniformly bounded operators, and assume now that $E_h(t)$ is a given uniformly bounded family of semigroups on X_h which approximate $E(t)$ in the sense that, with Y a dense subspace of X,

$$\| E_h(t) P_h v - P_h E(t) v \|_h \leqslant \varepsilon_h(1 + \gamma t) \| v \|_Y \quad \forall v \in Y . \tag{4.6}$$

With A_h the generator of $E_h(t)$ we consider as before the semidiscrete problem (4.2) and its completely discrete analogue (4.3). Under the assumptions of Theorem 3 we shall now present an estimate for the error between the solutions of these two problems. Combined with an error estimate for the semidiscrete problem this would yield a complete error bound. We denote by Y_θ the interpolation space $Y_\theta = (X, Y)_{\theta,\infty}$ between our basic space X and its subspace Y.

THEOREM 5 : *In the present situation assume that the time discretization scheme is accurate of order p, strictly accurate of order* $p - 1$, *that* (2.6) *holds, and that* E_{kh} *is uniformly stable in* X_h. *Then under the appropriate regularity assumptions*

$$\| u_h(t_n) - u_{h,n} \|_h \leqslant C\varepsilon_h \left\{ (1 + t_n) \| v \|_Y + \| Av \|_Y + t_n \sum_{l=0}^{p-1} \| f^{(l)}(0) \|_{Y_{1-l/p}} \right\} +$$

$$+ Ck^p \left\{ t_n \| u^{(p+1)}(0) \| + t_n \sum_{l=0}^{p} \| f^{(l)}(0) \|_{Y_{1-l/p}} \right.$$

$$\left. + \| f^{(p-1)}(0) \| + \int_0^{t_n} (t_n - s) \| f^{(p+1)}(s) \| \, ds \right\}.$$

Proof : Direct application of Theorem 3 gives

$$\| u_h(t_n) - u_{h,n} \|_h \leqslant Ck^p \left\{ t_n \| \theta_p(kA_h) u_h^{(p+1)}(0) \|_h + \right.$$

$$\left. + \| P_h f^{(p-1)}(0) \|_h + t_n \| P_h f^{(p)}(0) \|_h + \int_0^{t_n} (t_n - s) \| P_h f^{(p+1)}(s) \|_h \, ds \right\}$$

where

$$u_h^{(p+1)}(0) = A_h^{p+1} P_h v + \sum_{l=0}^{p} A_h^{p-l} P_h f^{(l)}(0).$$

Since P_h is bounded, the terms containing f are bounded as stated. In order to estimate the first term on the right it suffices by (3.7) to bound

$$S = k^p \{ \theta_p(kA_h) u_h^{(p+1)}(0) - P_h \theta_p(kA) u^{(p+1)}(0) \},$$

or, with $\tilde{\theta}_j(z) = z^j \theta_p(z)$,

$$S = \tilde{\theta}_p(kA_h) A_h P_h v - P_h \tilde{\theta}_p(kA) Av + \sum_{l=0}^{p} k^l (\tilde{\theta}_{p-l}(kA_h) P_h - P_h \tilde{\theta}_{p-l}(kA)) f^{(l)}(0)$$

$$= S_{p+1} + \sum_{l=0}^{p} S_l.$$

In order to deal with the different terms in S we shall need the following lemma.

LEMMA 3 : *If* (4.6) *holds and* $g, g' \in \tilde{M}$ *we have*

$$\| g(kA_h) P_h v - P_h g(kA) v \|_h \leqslant \varepsilon_h (m(g) + \gamma k m(g')) \| v \|_Y.$$

Proof : We have

$$g(z) = \int_{R_+} e^{zt}\, d\mu(t) \quad \text{with} \quad \int_{R_+} d\,|\,\mu\,|\,(t) = m(g)\,,$$

and

$$g'(z) = \int_{R_+} e^{zt}\, t\, d\mu(t)\,,$$

so that

$$\int_{R_+} t\, d\,|\,\mu\,|\,(t) = m(g')\,.$$

Now

$$g(kA_h)\, P_h\, v - P_h\, g(kA)\, v = \int_{R_+} (E_h(kt)\, P_h - P_h\, E(kt))\, v\, d\mu(t)\,,$$

and hence

$$\begin{aligned} \|\, g(kA_h)\, P_h\, v - P_h\, g(kA)\, v\,\|_h &\leqslant \varepsilon_h \int_{R_+} (1 + \gamma kt)\, d\,|\,\mu\,|\,(t)\, \|\, v\,\|_Y \\ &\leqslant \varepsilon_h (m(g) + \gamma k m(g'))\, \|\, v\,\|_Y\,. \end{aligned}$$

Note that since $g(kA_h)$, P_h, and $g(kA)$ are bounded we have by interpolation, for k bounded,

$$\|\, g(kA_h)\, P_h\, v - P_h\, g(kA)\, v\,\|_h \leqslant C\varepsilon_h^{\theta}\, \|\, v\,\|\, Y_\theta \quad \text{for} \quad 0 \leqslant \theta \leqslant 1\,. \tag{4.7}$$

Note also that as a result of the lemma we have for the elliptic projection defined in (4.4),

$$\|\, Q_h\, v - P_h\, v\,\|_h = \|\, ((I - A_h)^{-1}\, P_h - P_h (I - A)^{-1})\, (I - A)\, v\,\|_h \leqslant C\varepsilon_h\, \|\, (I - A)\, v\,\|_Y\,.$$

We may now complete the proof of Theorem 5 by estimating the S_l, $l = 0, \ldots, p + 1$. Since $\tilde{\theta}_l$, $\tilde{\theta}_l' \in \tilde{M}$ we have by Lemma 3 and (4.7) for $l = 0, \ldots, p$,

$$\|\, S_l\,\|_h \leqslant Ck^l\, \varepsilon_h^{1 - l/p}\, \|\, f^{(l)}(0)\,\|_{Y_{1 - l/p}} \leqslant C(k^p + \varepsilon_h)\, \|\, f^{(l)}(0)\,\|_{Y_{1 - l/p}}\,.$$

In order to bound S_{p+1} we first note that we may replace $P_h\, v$ by discrete initial values $v_h = Q_h\, v$. For the difference in the solution between $v_h = Q_h\, v$

and $v_h = P_h v$ may be bounded as follows,

$$\| E_{kh}^n (Q_h - P_h) v \|_h \leqslant C\varepsilon_h (\| Av \|_Y + \| v \|_Y) .$$

With $v_h = Q_h v$ we have by a simple calculation

$$\begin{aligned} S_{p+1} &= \tilde{\theta}_p(kA_h) A_h Q_h v - P_n \tilde{\theta}_p(kA) Av \\ &= (\tilde{\theta}_p(kA_h) P_h - P_h \tilde{\theta}_p(kA)) Av + \tilde{\theta}_p(kA_h)(Q_h - P_h) v , \end{aligned}$$

and hence

$$\| S_{p+1} \|_h \leqslant C\varepsilon_h (\| Av \|_Y + \| v \|_Y) .$$

The proof of the theorem is now complete.

5. CONSTRUCTION OF ACCURATE SCHEMES

Recall from Section 1 that the scheme (1.2) is of order p if and only if

$$r(z) - e^z = 0(z^{p+1}) \quad \text{as} \quad z \to 0 \tag{i}$$

and

$$\gamma_l(z) = \frac{l!}{z^{l+1}} \left(r(z) - \sum_{j=0}^{l} \frac{z^j}{j!} \right) - \sum_{j=1}^{m} \tau_j^l q_j(z) = 0(z^{p-l}) \quad \text{as} \quad z \to 0 , \tag{ii}$$

$$\text{for} \quad l = 0, \ldots, p-1 ,$$

and is strictly accurate of order $p_0 \leqslant p$ if in addition

$$\gamma_l(z) = 0 \quad \text{for} \quad l = 0, \ldots, p_0 - 1 .$$

For the case that the number m of quadrature points is less than p we shall now give an alternative characterization of a scheme of order p which will be used to construct accurate schemes below.

LEMMA 4 : *Let $m < p$. Then the scheme* (1.2) *is accurate of order p if and only if* (i) *holds together with*

$$\gamma_l(z) = 0(z^{p-l}) \quad \textit{as} \quad z \to 0 \quad \textit{for} \quad l = 0, \ldots, m-1 , \tag{ii}'$$

and with $\omega(t) = \Pi_{j=1}^m (t - \tau_j)$,

$$\int_0^1 \omega(t)\, t^j\, dt = 0 \quad \textit{for} \quad j = 0, \ldots, p-m-1 . \tag{iii}$$

Proof : Let us first note that (iii) is equivalent to the existence of $b_1, ..., b_m$ such that, with Π_{p-1} all polynomials of degree at most $p-1$,

$$\int_0^1 \varphi(t)\, dt = \sum_{j=1}^{m} b_j\, \varphi(\tau_j) \qquad \forall \varphi \in \Pi_{p-1}\,. \tag{iii)'}$$

To show the necessity of the conditions it thus suffices to show (iii)' for $\varphi = t^l$, $l = 0, ..., p-1$. But by (i) and (ii),

$$\gamma_l(0) = \frac{1}{l+1} - \sum_{j=1}^{m} \tau_j^l\, q_j(0) = 0\,, \qquad l = 0, ..., p-1\,,$$

so that with $b_j = q_j(0)$,

$$\int_0^1 t^l\, dt = \frac{1}{l+1} = \sum_{j=0}^{m} b_j\, \tau_j^l\,.$$

We now turn to the sufficiency of the conditions and it suffices then to show that

$$\gamma_l(z) = 0(z^{p-l}) \quad \text{as} \quad z \to 0 \quad \text{for} \quad l = m, ..., p-1\,. \tag{5.1}$$

We have by integration by parts and by (i),

$$\frac{z^{l+1}}{l!} \int_0^1 e^{z(1-t)}\, t^l\, dt = e^z - \sum_{j=0}^{l} \frac{z^j}{j!} = r(z) - \sum_{j=0}^{l} \frac{z^j}{j!} + 0(z^{p+1}) \quad \text{as} \quad z \to 0$$

and hence

$$\gamma_l(z) = \int_0^1 e^{z(1-t)}\, t^l\, dt - \sum_{j=0}^{m} \tau_j^l\, q_j(z) + 0(z^{p-l}) \quad \text{as} \quad z \to 0\,.$$

For $\omega(t)$ as above we write $\omega(t) = \sum_{i=0}^{m} \alpha_i\, t^i$. Then since $\omega(\tau_j) = 0$ we obtain by expanding the integrand and using (ii)', for $l = 0, ..., p-m-1$,

$$\sum_{i=0}^{m} \alpha_i\, \gamma_{i+l}(z) = \int_0^1 e^{z(1-t)}\, t^l\, \omega(t)\, dt + 0(z^{p-m-l}) = 0(z^{p-m-l}) \quad \text{as} \quad z \to 0\,.$$

Since $\alpha_m = 1$ we conclude (5.1) by induction over l.

The above lemma provides a method for constructing a scheme which is accurate of order p, and strictly accurate of order m, if $m \geqslant p/2$: We first choose

$r(z)$ so that (i) holds, then select the distinct numbers $\{ \tau_j \}_1^m \subset [0, 1]$ so that (iii) is satisfied and finally determine the rational functions $\{ q_j(z) \}_1^m$ so that $\gamma_l(z) = 0$ for $l = 0, ..., m - 1$ or

$$\sum_{j=1}^{m} \tau_j^l \, q_j(z) = \frac{l\,!}{z^{l+1}} \left(r(z) - \sum_{j=0}^{l} \frac{z^j}{j\,!} \right) \quad \text{for} \quad l = 0, ..., m - 1 . \qquad \text{(ii)}''$$

Note that the matrix of this system is nonsingular since the τ_j are distinct, and that the $q_j(z)$ will have the same denominators as $r(z)$, which is advantageous for the implementation of the scheme. Note also that the condition $p \leqslant 2m$ is necessary for the existence of $\{ \tau_j \}_1^m$ so that (iii) holds; if $p = 2m$ the points are uniquely determined as the Gaussian points of order m on $[0, 1]$.

It is now natural to ask if the conditions (i), (ii)'' and (iii) (or (iii)') in fact imply strict accuracy of order higher than m. In this regard we have the following :

LEMMA 5 : *Assume that the scheme* (1.2) *is accurate of order* p *and strictly accurate of order* $m < p$ *where* m *is the number of quadrature points. Then it is strictly accurate of order* $m + 1$ *if and only if with* $\omega(t) = \Pi_{j=1}^{m}(t - \tau_j)$,

$$r(z) = \frac{\sum_{j=0}^{m} z^{m-j} \, \omega^{(j)}(1)}{\sum_{j=0}^{m} z^{m-j} \, \omega^{(j)}(0)} . \qquad (5.2)$$

The scheme cannot be strictly accurate of order $m + 2 \leqslant p$.

Proof : Recalling the definition of $\gamma_l(z)$ and $\omega(t) = \sum_{i=0}^{m} \alpha_i \, t^i$ we have since $\gamma_l(z) = 0$ for $l = 0, ..., m - 1$,

$$\gamma_m(z) = \sum_{i=0}^{m} \alpha_i \, \gamma_i(z) = z^{-(m+1)} \left\{ \sum_{i=0}^{m} \alpha_i \, i\,! \, z^{m-i} \, r(z) - \sum_{i=0}^{m} \sum_{j=0}^{i} \alpha_i \frac{i\,!}{j\,!} z^{m-i+j} \right\}$$

$$= z^{-(m+1)} \left\{ r(z) \sum_{i=0}^{m} z^{m-i} \, \omega^{(i)}(0) - \sum_{i=0}^{m} z^{m-i} \, \omega^{(i)}(1) \right\},$$

which shows that $\gamma_m(z) = 0$ if and only if (5.2) holds.

Similarly, if the scheme is strictly accurate of order $m + 1 < p$ we have with

$$\tilde{\omega}(t) = t\omega(t) = \sum_{i=0}^{m} \alpha_i \, t^{i+1} ,$$

$$\gamma_{m+1}(z) = z^{-(m+2)} \left\{ r(z) \sum_{i=0}^{m} z^{m-i} \, \tilde{\omega}^{(i+1)}(0) - \sum_{i=0}^{m} z^{m-i} \, \tilde{\omega}^{(i+1)}(1) \right\},$$

and strict accuracy of order $m + 2$ would imply that in addition to (5.2),

$$r(z) = \frac{\sum_{\iota=0}^{m} z^{m-\iota}\, \tilde{\omega}^{(\iota+1)}(1)}{\sum_{\iota=0}^{m} z^{m-\iota}\, \tilde{\omega}^{(\iota+1)}(0)}. \tag{5.3}$$

Since $\tilde{\omega}^{(m+1)}(0) = (m+1)\,! = (m+1)\,\omega^{(m)}(0)$ a comparison between (5.2) and (5.3) shows that we must have $\tilde{\omega}^{(\iota+1)}(0) = (m+1)\,\omega^{(\iota)}(0)$ for $i = 0, \ldots, m$. Since $\tilde{\omega}'(0) = \omega(0)$ this is impossible if $\omega(0) \neq 0$. But if $\omega(0) = 0$ we have since the τ_j are distinct that $\omega'(0) \neq 0$ and since $\tilde{\omega}''(0) = 2\,\omega'(0)$ we now conclude $m = 1$. In this case $\omega(t) = t$ and (5.2) and (5.3) both yield $r(z) = 1 + z$ which is not permissible.

For the case $p = 2m$ the function $r(z)$ defined by (5.2) is the diagonal Padé approximant $r_{m,m}(z)$ of e^z since this is then uniquely determined by (i). The particular case $m = 1$ corresponds to the Crank-Nicolson scheme

$$\left(I - \frac{1}{2}kA\right) u_{n+1} = \left(I + \frac{1}{2}kA\right) u_n + kf\left(t_n + \frac{1}{2}k\right).$$

For $m = 2$ we have

$$\left(I - \frac{1}{2}kA + \frac{1}{12}k^2A^2\right) u_{n+1} = \left(I + \frac{1}{2}kA + \frac{1}{12}k^2A^2\right) u_n +$$

$$+ \frac{1}{2}k\left[\left(I + \frac{\sqrt{3}}{6}kA\right) f\left(t_n + k\left(\frac{1}{2} - \frac{\sqrt{3}}{2}\right)\right) + \left(I - \frac{\sqrt{3}}{6}kA\right) f\left(t_n + k\left(\frac{1}{2} + \frac{\sqrt{3}}{2}\right)\right)\right],$$

which is accurate of order 4, and strictly accurate of order 3. It is easy to check that (2.6) holds so that the error estimate of Theorem 2 applies.

One way of generating schemes of type (1.2) is to use implicit Runge-Kutta methods (*cf.* [2]). If the method is of collocation type, condition (5.2) is satisfied and $\gamma_l(z) = 0$ for $l = 0, \ldots, m - 1$. An interesting class of such schemes is given by Nørsett [3]; these schemes satisfy the relation (5.2) and the rational functions r and q_j have exactly one pole, which is the same for all these functions. For related material, see also Nørsett and Wanner [4].

REFERENCES

1. P. BRENNER and V. THOMÉE, *On rational approximation of semi-groups,* SIAM J. Numer. Anal. *16,* 1979, 683-694.
2. M. CROUZEIX, *Sur l'approximation des équations différentielles opérationnelles linéaires par des méthodes de Runge-Kutta,* Thèse, Université Paris VI, 1975.
3. S. P. NØRSETT, *Runge-Kutta methods with a multiple real eigenvalue only,* BIT *16,* 1976, 388-399.
4. S. P. NØRSETT and G. WANNER, *Perturbed collocation and Runge-Kutta methods,* Report, Université de Genève, 1978.