N. R. Nassif

On a finite-element collocation method which reproduces the Padé table

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 15 (1974), No. 1-2, 105--109

Persistent URL: http://dml.cz/dmlcz/142336

1974 ACTA UNIVERSITATIS CAROLINAE - MATHEMATICA ET PHYSICA NO. 1-2 PAG. 105-109

On a Finite-Element Collocation Method Which Reproduces the Padé Table

N. R. NASSIF
Mathematics Department, American University, Beirut

In this paper, we derive Padé approximations of any order by using Hermite interpolatory bases and collocating appropriately to obtain one-step numerical solutions of the system of differential equations $dy/dt = -Sy$, $y(0) = y_0$.

I. Introduction

Consider the initial-value problem

$$y'(t) = -Sy\,, \quad 0 \leqq t \leqq T \tag{1}$$

$$y(0) = y_0 \tag{1'}$$

where $y\colon [0, T] \to R^n$ and $S = \{S_{ij}\}$ is an $n \times n$ matrix whose coefficients are independent of the variable t. The solution to (1) — (1′) is given by [4, p. 254], $y(t) = \exp(-tS)\,y_0$, which implies if $\Delta t \geqq 0$ that

$$y(t + \Delta t) = \exp(-\Delta t S)\,y(t)\,, \qquad 0 \leqq t, t + \Delta t \leqq T\,. \tag{2}$$

It is well-known [ibid, pp. 262—270], that one-step implicit and explicit schemes for solving numerically (1) — (1′) are obtained by considering Padé Matrix approximations to $\exp(-\Delta t S)$. Specifically consider for $0 \leqq l, r$ the (l, r) entry of the Padé Table composed of $P_{r,l}(z)/Q_{r,l}(z)$, where

$$P_{r,l}(z) = \sum_{k=0}^{l} \frac{(r + l - k)!\,(l)!}{(r + l)!\,(k)!\,(l - k)!}\,(-z)^k \tag{3}$$

and

$$Q_{r,l}(z) = \sum_{k=0}^{r} \frac{(r + l - k)!\,(r)!}{(r + l)!\,(k)!\,(r - k)!}\,z^k = P_{l,r}(-z)\,. \tag{3'}$$

The one-step scheme, based on this (r, l) entry is

$$Q_{r,l}(\Delta t S)\,z(t + \Delta t) = P_{r,l}(\Delta t S)\,z(t)\,. \tag{4}$$

If $r = 0$, (4) is *explicit*. Otherwise it is *implicit*, unless S reduces to a triangular matrix.

In this paper, our main result is that (4) can be obtained in the numerical solution of (1) — (1′) by a finite-element collocation technique in t. If we let

$$\pi_{\Delta t} = t_0 = 0 < t_1 = \Delta t < t_2 = 2\Delta t < \dots < t_m = m\Delta t = T$$

be a (not necessarily) uniform partition of the interval $[0, T]$. Let also l and r be two non-negative integers with $q = \min(r-1, l-1)$ and $p = \max(r-1, l-1)$.

We consider the following class of piecewise polynomials. For $l, r \geqq 1$ $H^{l,r} \equiv H^{l,r}([0, T]; R^n; \Pi_{\Delta t})$ denotes the subspace of $C^q([0, T]; R^n)$ which consists of those functions from $[0, T]$ to R^n which are polynomials of degree $(l + r - 1)$ on the intervals of $[0, T]$ determined by the partition $\Pi_{\Delta t}$.

For this purpose we construct on the interval [0,1], the polynomials of degree $l + r - 1, \Psi_k(t),\ 0 \leqq k \leqq l - 1$ and $\Phi_k(t),\ 0 \leqq k \leqq r - 1$, defined by the conditions

$$\Psi_k^{(j)}(0) = \delta_{jk}\,, \quad 0 \leqq j \leqq l - 1 \qquad \text{(5a)} \qquad \Psi_k^{(j)}(1) = 0\,, \quad 0 \leqq j \leqq r - 1 \qquad \text{(5b)}$$

$$\Phi_k^{(j)}(0) = 0\,, \quad 0 \leqq j \leqq l - 1 \qquad \text{(5c)} \qquad \Phi_k^{(j)}(1) = \delta_{kj}\,, \quad 0 \leqq j \leqq r - 1 \qquad \text{(5d)}$$

$$z \in H^{l,r} \Leftrightarrow z(t) = \sum_{k=0}^{l-1} (\Delta t)^k\, \Psi_k\left(\frac{t - t_i}{\Delta t}\right) z_i^k + \sum_{k=0}^{r-1} (\Delta t)^k\, \Phi_k\left(\frac{t - t_i}{\Delta t}\right) z_{i+1}^k\,,$$

$$t \in [t_i, t_{i+1}]\,. \qquad (6)$$

Note then that $z_i^k = z^{(k)}(t_i)$. (6), (5a), (5b), (5c), and (5d) insure that $z \in C^q([0, T]; R^n)$. If $l = 0$ and $r \geqq 1$ (or $r = 0$ and $l \geqq 1$) then $H^{0,r}$ (resp. $H^{l,0}$) consists of those functions which are polynomials of degree $l + r - 1$, on each subinterval $[t_i, t_{i+1}]$, $0 \leqq i \leqq m - 1$, and which are only piecewise continuous. That is $z \in H^{0,r}$ (resp. $H^{l,0}$) iff

$$z(t) = \sum_{k=0}^{r-1} (\Delta t)^k\, \Phi_k\left(\frac{t - t_i}{\Delta t}\right) z_{i+1}^k, \left(\text{resp. } z(t) = \sum_{k=0}^{l-1} (\Delta t)^k\, \Psi_k\left(\frac{t - t_i}{\Delta t}\right) z_i^k\right). \quad (6')$$

Note: $H^{l,r}$ represents exactly a special case of classical Hermite interpolation for two knots [1].

To construct the approximation $z(t)$ to the solution $y(t)$ of (1) — (1′) we start by rewriting (1), in a sequence of integral equations, and this by integrating (1) on $[t_i, t_{i+1}]$, obtaining,

$$y(t_{i+1}) - y(t_i) = -\int_{t_i}^{t_{i+1}} Sy(t)\,\mathrm{d}t\,, \qquad 0 \leqq i \leqq m - 1\,. \qquad (7)$$

Also by differentiating (1) j times one arrives at

$$y^{(j)}(t) = (-S)^j\, y(t) \qquad 0 \leqq t \leqq T, \qquad 1 \leqq j \qquad (8)$$

which implies,

$$y^{(j)}(t_i) = (-S)^j\, y(t_i)\,, \qquad 0 \leqq j \leqq p\,, \qquad 0 \leqq i \leqq m\,. \qquad (8')$$

Our requirements on $z(t)$ are (R1) $z \in H^{l,r}$, (R2) z satisfies the integral equations (8), (R3) z satisfies (1') and if in addition $l \geqq 2$, or $r \geqq 2$, or $l, r \geqq 2$, z satisfies (8'). See [3], for $l = r = 1,2$. Our results are as follows.

Theorem I. The difference scheme obtained from (R1), (R2) and (R3) is equivalent to (4).

Global error bounds can also be derived and this in any vector norm $\|x\|$, $x \in R^n$. Specifically, let $\|A\| = \sup\limits_{\|x\|=1} \|Ax\|$ be the usual associated matrix norm, $P_{r,l} \equiv P_{r,l}(\Delta t S)$, $Q_{r,l} \equiv Q_{r,l}(\Delta t S)$, and $G_{r,l} = [Q_{r,l}]^{-1} [P_{r,l}]$. Our next result is:

Theorem II. Assume S, Δt, are such that

$$\|(G_{r,l})^i (Q_{r,l})^{-1}\| \leqq K, \quad 1 \leqq i \leqq m \tag{9}$$

where K is independent of (Δt), then for $\Delta t \leqq \delta$, δ positive constant, we have

$$\|y(t) - z(t)\| \leqq M(\Delta t)^{l+r}, \quad 0 \leqq t \leqq T$$

where

$M = M(l, r, S, K, T, y)$. Thus one obtains global convergence of order $(\Delta t)^{l+r}$.

Remark: Assumption (9) is the usual stability assumption, in the sense of Richtmyer. It is for example satisfied when $r \geqq l$, and S is positive definite. In this case the stability is unconditional.

2. Proof of Theorem I: If $l = r = 0$, the proposition is obvious; one obtains then $z_{i+1} = z_i$. In the case $l = 0$ and $r \geqq 1$, or $r = 0$ and $l \geqq 1$, or $r, l \geqq 1$, then we obtain respectively, from (R1), (R2), and (R3),

$$z_{i+1} = \left\{1 + \sum_{k=0}^{l-1} \left[(\Delta t)^k \int_{t_i}^{t_{i+1}} \psi_k \left(\frac{t - t_i}{\Delta t}\right) \mathrm{d}t\right] (-S)^{k+1}\right\} z_i \tag{10}$$

$$\left\{1 - \sum_{k=0}^{r-1} (\Delta t)^k \left[\int_{t_i}^{t_{i+1}} \Phi_k \left(\frac{t - t_i}{\Delta t}\right) \mathrm{d}t\right] (-S)^{k+1}\right\} z_{i+1} = z_i \tag{10'}$$

$$\left\{1 - \sum_{k=0}^{r-1} \left[\int_{t_i}^{t_{i+1}} (\Delta t)^k \Phi_k \left(\frac{t - t_i}{\Delta t}\right) \mathrm{d}t\right] (-S)^{k+1}\right\} z_{i+1} =$$

$$= \left\{1 + \sum_{k=0}^{l-1} \left[\int_{t_i}^{t_{i+1}} (\Delta t)^k \psi_k \left(\frac{t - t_i}{\Delta t}\right) \mathrm{d}t\right] (-S)^{k+1}\right\} z_i . \tag{10''}$$

It can be proved without difficulty, in an otherwise lengthy argument that

$$\int_{t_i}^{t_{i+1}} \psi_k \left(\frac{t - t_i}{\Delta t}\right) \mathrm{d}t = \Delta t \, \frac{(l + r - k - 1)!\,(l)!}{(l + r)!\,(k + 1)!\,(l - 1 - k)!}$$

and

$$\int_{t_i}^{t_i+1} \Phi_k\left(\frac{t-t_i}{\Delta t}\right) \mathrm{d}t = \Delta t \frac{(l+r-k-1)!\ (r)!}{(l+r)!\,(k)!\,(r-1-k)!}\ .$$

We may then rewrite (10), (10′), (10″), thus verifying in all cases that

$$\left\{\sum_{k=0}^{r} \frac{(l+r-k)!\ (r)!}{(l+r)!\,(k)!\,(r-k)!} (\Delta t S)^k\right\} z_{i+1} = \left\{\sum_{k=0}^{l} \frac{(l+r-k)!\,(k)!}{(l+r)!\,(k)!\,(l-k)!} (-\Delta t S)^k\right\} z_i$$

precisely (4).

3. Error Estimates: We let $e(t) = y(t) - z(t)$, $y_i^k \equiv y^{(k)}(t_i)$, $e_i^k \equiv e^{(k)}(t_i)$, $0 \leqq k \leqq p$, with $e_i \equiv e_i^0$. To prove theorem 2, we obtain first some *local* error estimates. Precisely,

Lemma. For $0 \leqq k \leqq p$, $0 \leqq i \leqq m$,

$$\|e_i^k\| \leqq K(\Delta t)^{l+r+1} \frac{(l)!\ (r)!}{(l+r+1)!\,(l+r)!}\, i \sup_{0\leq t\leq T} \|y^{(l+r+k+1)}(t)\|\ . \tag{11}$$

Proof. From (2), it is clear that $y \in C^\infty([0, T]; R^n)$, and from [2],

$$y(t) = \sum_{k=0}^{l-1} (\Delta t)^k\, \Psi_k\left(\frac{t-t_i}{\Delta t}\right) y^k + \sum_{k=0}^{r-1} (\Delta t)^k\, \Phi_k\left(\frac{t-t_i}{\Delta t}\right) y_{i+1}^k +$$

$$+ (t-t_i)^l\,(t-t_{i+1})^r\,[y_i^{(l)}, y_{i+1}^{(r)}, y]$$

where $[y_i^{(l)}, y_{i+1}^{(r)}, y]$ is the usual notation for divided differences. Thus since y satisfies (7), (8) and (8′), define the truncation error,

$$Q_{r,l}\, e_{i+1} - P_{r,l}\, e_i = \sigma_i\ . \tag{12}$$

It follows then that

$$\sigma_i = -S \int_{t_i}^{t_i+1} (t-t_i)^l\,(t-t_{i+1})^r\,[y_i^{(l)}, y_{i+1}^{(r)}, y]\ \mathrm{d}t$$

and thus

$$e_i = \sum_{k=1}^{i} (G_{r,l})^{i-k} (Q_{r,l})^{-1} \sigma_{k-i}\,, \qquad 1 \leqq i \leqq m \tag{13}$$

(9) and (13) imply (11) for $k = 0$, since $[y_i^{(l)}, y_{i+1}^{(r)}, y] = y^{(l+r)}(\eta_i) \,/\, (l+r)!$

and

$$\int_{t_i}^{t_i+1} (t-t_i)^l\,(t_{i+1}-t)^r\ \mathrm{d}t = \frac{(r)!\,(l)!}{(l+r+1)!} (\Delta t)^{l+r+1}\ .$$

For $k \geqq 1$, consider the difference equation (4) obtained from (R1), (R2) and (R3).

By left-multiplication with $(-S)^k$, it implies that

$$Q_{r,l}(-S)^k z_{i+1} = P_{r,l}(-S)^k z_i\,, \quad \text{and from (R3)}$$

$$Q_{r,l}\, z_{i+1}^k = P_{r,l}\, z_i^k\,, \quad 0 \leqq i \leqq m-1,\ 1 \leqq k \leqq p \tag{14}$$

$$z_0^k = (-S)^k y_0\,. \tag{14'}$$

That is, the (r, l) Padé approximation to $y^{(k+1)}(t) = -Sy^{(k)}(t), y^{(k)}(0) = (-S)^k y_0 = y_0^k$. Therefore as for k $= 0$, (11) is true for $1 \leqq k \leqq p$. To complete the proof, observe that

$$e(t) = \sum_{k=0}^{l-1} (\Delta t)^k\, \Psi_k \left(\frac{t - t_i}{\Delta t}\right) e_i^k + \sum_{k=0}^{r-1} (\Delta t)^k\, \Phi_k \left(\frac{t - t_i}{\Delta t}\right) e_{i+1}^k +$$

$$+ \frac{(t - t_i)^l\,(t - t_{i+1})^r}{(l + r)!}\; y^{(l+r)}(\eta_i),\ \eta_i \in (t_i, t_{i+1}),\ 0 \leqq i \leqq m-1\,.$$

Thus, if

$$C = \max_{0 \leqq k \leqq l-1} \{ \sup_{0 \leqq \varkappa \leqq 1} |\Psi_k(\varkappa)|\}\,, \qquad D = \max_{0 \leqq k \leqq r-1} \{ \sup_{0 \leqq \varkappa \leqq 1} |\Phi_k(\varkappa)|\}\,,$$

and

$$\bar{C} = KT \frac{(l)!\,(r)!}{(l + r + 1)!\,(l + r)!} \max_{0 \leqq k \leqq p} \{ \sup_{0 \leqq t \leqq T} \|y^{(l+r+k+1)}(t)\|\}\,.$$

$$\text{If } \Delta t \leqq \delta < 1, \text{ then } \|e(t)\| \leqq M(\Delta t)^{l+r}\,,$$

$$\text{where } M = \max \left\{ \frac{(C + D)\,\bar{C}}{1 - \delta}\,; \quad \sup_{0 \leqq t \leqq T} \frac{\|y^{(l+r)}\,(t)\|}{(l + r)!} \right\}.$$

References

[1] Davis, P.: Interpolation and Approximation, Blaisdell.

[2] Isaacson, E., Keller, H. B.: Analysis of Numerical Methods, John Wiley & Sons.

[3] Kang, C. M.: On the Use of The Finite Element Method in Reactor Kinetics Problems. Ph.D. Thesis, Massachussets Institute of Technology (1971).

[4] Varga, R. S.: Matrix Iterative Analysis. Prentice-Hall (1962).