Journal of Computational and Applied Mathematics 131 (2001) 361–380

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.nl/locate/cam

An interpolatory approximation of the matrix exponential based on Faber polynomials

I. Moret *, P. Novati

Dipartimento di Scienze Matematiche, Università di Trieste, Piazzale Europa 1, 34127 Trieste, Italy

Received 14 August 1999; received in revised form 17 December 1999

Abstract

In this paper we introduce a method for the approximation of the matrix exponential obtained by interpolation in zeros of Faber polynomials. In particular, we relate this computation to the solution of linear IVPs. Numerical examples arising from practical problems are examined. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Given an $N \times N$ matrix A and an N-dimensional vector v, we consider the computation of $f(A)v$, where f is a given function. In particular, our attention is devoted to the case of the exponential, in relation to the solution of linear IVPs, such as

$$y'(t) - Ay(t) = f(t)v, \quad t \geqslant 0,$$

$$y(0) = 0.$$

We study here approximations belonging to the Krylov subspaces $K_m = \text{span}\{v, Av, \ldots, A^{m-1}v\}$ associated with A and v, namely of type $p_{m-1}(A)v$, where p_{m-1} is a polynomial of degree at most $m-1$. This approach turns out to be particularly convenient when A is large and sparse. The approximations proposed here have an interpolatory nature in the sense that they follow by interpolating f on suitable sets of points in the complex plane. Nevertheless, the procedure can be carried out without knowing explicitly the interpolation points. Some aspects of the approximation of matrix functions via polynomial interpolation were considered also in [18,19,26].

Our approach represents a generalization of methods recently proposed in the literature [1,4–6,9–11] where the Krylov subspaces are constructed by the Arnoldi or Lanczos algorithms and

* Corresponding author.

PII: S0377-0427(00)00261-2

f is interpolated in the so-called Arnoldi or Lanczos–Ritz values associated to A and v [22]. As well known the use of these basis generators may present in general various difficulties due to the growing computational costs and required storage and, in the Lanczos' case, to instability and possible breakdowns. Here, on the contrary, the interpolation points are the zeros of suitable polynomials which are defined beforehand using some information on the matrix A.

In particular, we are here interested in studying the approximation of $\exp(tA)v$, for $t > 0$, by interpolating on the zeros of Faber polynomials associated to a certain compact subset Ω of the complex plane which contains the spectrum of A, $\sigma(A)$. The convergence follows from the well-known fact that the zeros of Faber polynomials are uniformly distributed on Ω [26]. The construction of Ω makes use of a preliminary phase where the spectral properties of A need to be investigated. For this reason, in the context of the solution of algebraic linear systems, such kind of complex approximation techniques are often called hybrid methods. Another used terminology is Chebyshev-like methods. They have been studied by several authors (see for instance [7,8,14,15,25]) who gave various interesting motivations for this approach (cf. also [17,16]).

The present paper is organized as follows. In Section 2 the general interpolatory procedure is illustrated. Its application to the matrix exponential is discussed in Section 3, where we also relate the approximation of the matrix exponential to the solution of linear IVPs. This allows us to give a restarted version of the method. In Section 4 we introduce Faber polynomials and we consider the interpolation on their roots. In Section 5 we point out some computational details. Finally, Section 6 contains some numerical tests involving matrices arising from the semidiscretization of partial differential equations of parabolic type.

2. Interpolatory approximations

We start with some general considerations. Let the $N \times N$ real matrix A be given and let $\{v_1, v_2, \ldots, v_j, \ldots\}$ be an ordered system of vectors in R^N such that for any index $j \geqslant 1$:

$$Av_j = \sum_{i=1}^{j+1} h_{i,j} v_i. \tag{2.1}$$

Then, setting $h_{i,j} = 0$, for $i > j + 1$, for any given m we consider the $m \times m$ real upper Hessenberg matrix H_m having entries $h_{i,j}$ for $i, j = 1, 2, \ldots, m$. Accordingly, we have

$$AV_m = V_m H_m + h_{m+1,m} v_{m+1} e_m^{\mathrm{T}}, \tag{2.2}$$

where V_m is the $N \times m$ matrix $V_m = [v_1, v_2, \ldots, v_m]$. Here and below e_j is the jth vector of the canonical basis of R^m.

From now on let v be an N-dimensional real vector such that $v = \beta v_1$, for some scalar β. Having to compute $y = f(A)v$, where f is a given function, we consider the approximation

$$y_m = V_m f(H_m) \beta e_1. \tag{2.3}$$

Here below we point out the interpolatory nature of this approximation.

The following results easily follow by taking into account the Hessenberg structure of H_m.

Proposition 2.1. *For* $k = 0, 1, 2, \ldots, m-2$,

$$e_m^{\mathrm{T}} H_m^k e_1 = 0$$

and

$$e_m^{\mathrm{T}} H_m^{m-1} e_1 = \prod_{j=1}^{m-1} h_{j+1,j}.$$

Moreover, the following result concerning Hessenberg matrices is well known (see [20]).

Proposition 2.2. *Each eigenvalue of* H_m *has geometric multiplicity equal to* 1 *and* H_m *is* non−derogatory, *that is, the minimal polynomial of* H_m *is its characteristic polynomial.*

Using Lemma 3.2 in [22] and the above two propositions, one easily proves the following result which extends Lemma 3.1 and Theorem 3.3 in [22].

Proposition 2.3. *Let* $D \subset \mathbb{C}$ *be an open set and let* f *be an analytic in D. Assume that the spectra of* A *and of* H_m *are contained in D. Let* p_{m-1} *be the polynomial which interpolates f, in the Hermite sense, in the eigenvalues of* H_m, *repeated according to their multiplicity. Then*

$$f(H_m) = p_{m-1}(H_m)$$

and

$$V_m f(H_m)\beta e_1 = p_{m-1}(A)v.$$

In the sequel, $\| \ \|$ denotes the euclidean vector norm. The same notation is used for the corresponding induced matrix norm. Moreover, $\| \ \|_\Omega$ denotes the supremum norm on a suitable set Ω.

Assuming that $D \subseteq \Omega$ and that A is diagonalizable, i.e., XAX^{-1} is diagonal, by Proposition 2.3 for (2.3) we have the bound

$$\|y - y_m\| = \|(f(A) - p_{m-1}(A))v\| \leqslant \mathrm{cond}_2(X)\|f(z) - p_{m-1}(z)\|_\Omega \|v\|.$$

Estimates of $\|f(z) - p_{m-1}(z)\|_\Omega$ could be obtained by interpolation theory. Clearly, we demand that $\lim_{m\to\infty}\|f(z) - p_{m-1}(z)\|_\Omega = 0$.

Let $W(A)$ denote the *field of values* (*numerical range*) of A, i.e.

$$W(A) := \left\{ \frac{x^H A x}{x^t x}, x \in \mathbb{C}/\{0\}. \right\} \tag{2.4}$$

Let Γ be the boundary curve of a piecewise smooth bounded region G where f is analytic and assume that $W(A) \subset G$. In order to obtain error estimates for (2.3), one can also use the matrix version of the Cauchy Integral Theorem, i.e.,

$$f(A) - p_{m-1}(A) = \frac{1}{2\pi i}\int_\Gamma (f(z) - p_{m-1}(z))(zI - A)^{-1}\,\mathrm{d}z$$

and apply the following result from [24, Theorem 4.1]:

Proposition 2.4. Under the above assumptions

$$\|(zI - A)^{-1}\| \leqslant 1/\mathrm{dist}(z, W(A)).$$

Other estimates based on the so-called *ε-pseudospectrum* of A [28] can also be used (see [10]). In various cases, $f(A)v$ represents the solution of a particular equation and we can take into consideration, as a measure of the approximation, the corresponding residual.

Example 2.5. For instance, let us assume that for a complex z the matrix $(zI-A)$ is nonsingular and let us approximate $y=(zI-A)^{-1}v$ by $y_m=\beta V_m(zI-H_m)^{-1}e_1$, provided that $(zI-H_m)$ is nonsingular too. Clearly there is a monic polynomial π_{m-1} of degree $m-1$ such that

$$-(zI-H_m)^{-1} = \frac{\pi_{m-1}(zI-H_m)}{\mathrm{Det}(H_m - zI)}.$$

Hence, by Proposition 2.1, we easily obtain

$$-e_m^{\mathrm{T}}(zI-H_m)^{-1}e_1 = \frac{\prod_{j=1}^{m-1} h_{j+1,j}}{\mathrm{Det}(zI-H_m)}. \tag{2.5}$$

Then the residual vector is given by

$$v-(zI-A)y_m = -\beta h_{m+1,m}(e_m^{\mathrm{T}}(zI-H_m)^{-1}e_1)v_{m+1} = \left(\frac{\beta\prod_{j=1}^{m} h_{j+1,j}}{\mathrm{Det}(zI-H_m)}\right)v_{m+1}.$$

In this paper we consider the interpolation on the zeros of a family of polynomials generated through a recursion like (2.1). Namely, let $\{q_j(z)\}_{j=0}^{\infty}$, $q_0(z)\neq 0$, where q_j, for $j=0,1,\ldots$, has degree j, be a sequence of polynomials satisfying

$$zq_{j-1}(z) = \sum_{i=1}^{j+1} h_{i,j}q_{i-1}(z), \quad \text{for } j\geqslant 1, \tag{2.6}$$

where the $h_{i,j}$, for $i,j=1,2,\ldots$, are the given real parameters with $h_{j+1,j}\neq 0$. Accordingly, in (2.1) one can set $v_j=q_{j-1}(A)v$, defining a method of type (2.3). Clearly, for every m, we have

$$[zq_0(z), zq_1(z), \ldots, zq_{m-1}(z)] = [q_0(z), q_1(z), \ldots, q_{m-1}(z)]H_m + h_{m+1,m}q_m(z)e_m^{\mathrm{T}}. \tag{2.7}$$

Proposition 2.6. *Let $\pi(z)$ be the characteristic polynomial of H_m, then $q_m(z)=C\pi(z)$, for some constant C.*

Proof. From (2.7) we easily realize that $q_m(H_m)=0$. Since, by Proposition 2.2, $\pi(z)$ divides any other annihilating polynomial of H_m, the conclusion follows. □

By this proposition, in the corresponding method (2.3) the polynomial p_{m-1} interpolates f in the zeros of q_m.

In Section 4 we discuss the case where the polynomials q_j are chosen as the *ordinary Faber polynomials* associated to Ω.

3. The exponential case

In this section we deal with the approximation of the matrix exponential relating it to the solution of systems of differential equations.

Let us consider $y(t) = \exp(tA)v$, for some $t \geqslant 0$, which, referring to the previous notation, we approximate by

$$y_m(t) = \beta V_m \exp(tH_m)e_1. \tag{3.1}$$

Since $y(t)$ solves the initial value problem

$$\begin{aligned} &y'(t) - Ay(t) = 0, \quad t \geqslant 0, \\ &y(0) = v \end{aligned} \tag{3.2}$$

and $u(t) := \beta \exp(tH_m)e_1$ solves

$$\begin{aligned} &u'(t) - H_m u(t) = 0, \quad t \geqslant 0, \\ &u(0) = \beta e_1, \end{aligned} \tag{3.3}$$

we can consider the residual of system (3.2) at y_m, that is

$$r_m(t) = Ay_m(t) - y_m'(t) = AV_m u(t) - V_m u'(t). \tag{3.4}$$

Accordingly, using (2.2) and (3.3) we get

$$r_m(t) = \alpha_m(t)v_{m+1}, \tag{3.5}$$

where

$$\alpha_m(t) = h_{m+1,m}(e_m^{\mathrm{T}} \exp(tH_m)e_1). \tag{3.6}$$

Proposition 3.1. *We have*

$$\alpha_m(t) = (1 + \mathrm{O}(t))\frac{t^{m-1}(\prod_{j=1}^{m} h_{j+1,j})}{(m-1)!}$$

with the upper bound

$$|\alpha_m(t)| \leqslant \|H_m\|^{m-1} \max_{0 \leqslant s \leqslant t} \exp(\mu_2 s)\frac{t^{m-1}}{(m-1)!}, \tag{3.7}$$

where μ_2 denotes the 2-logarithmic-norm of H_m [3, p. 19].

Proof. In order to estimate $e_m^{\mathrm{T}}u(t)$, we recall (cf. [13]) that there are entire functions $\chi_{m-1}(t)$, $j = 0, \ldots, m-1$, such that

$$\exp(tH_m) = \sum_{k=0}^{m-1} \chi_{m-1}(t)H_m^k.$$

Then, using Proposition 2.1, we get

$$e_m^{\mathrm{T}} \exp(tH_m)e_1 = \left(\prod_{j=1}^{m-1} h_{j+1,j}\right) \chi_{m-1}(t).$$

Accordingly,

$$\alpha_m(t) = \left(\prod_{j=1}^{m} h_{j+1,j}\right) \chi_{m-1}(t).$$

Then, by Proposition 2.1, observing that, for $k = 0, \ldots, m-2$, the derivatives of χ_{m-1} are such that $\chi_{m-1}^{(k)}(0) = 0$ and $\chi_{m-1}^{(m-1)}(0) = 1$, we have

$$\chi_{m-1}(t) = (1 + \mathrm{O}(t))\frac{t^{m-1}}{(m-1)!}.$$

Inequality (3.7) follows easily by expansion of $e_m^{\mathrm{T}} \exp(tH_m)e_1$. □

When, for a suitable m, the approximation $y_m(t)$ has been computed, the procedure can be restarted, considering now the IVP

$$\begin{aligned} &(y - y_m)'(t) - A(y - y_m)(t) = r_m(t), \\ &(y - y_m)(0) = 0. \end{aligned} \tag{3.8}$$

Thus, here below, in the light of (3.8) and (3.5), we extend our attention to IVPs of the form

$$\begin{aligned} &y'(t) - Ay(t) = f(t)v, \quad t \geqslant 0, \\ &y(0) = 0, \end{aligned} \tag{3.9}$$

where $f(t)$ is a scalar function. The solution of this problem is

$$y(t) = \int_0^t f(s) \exp((t-s)A)v \,\mathrm{d}s$$

and we consider the approximation given by

$$y_m(t) = \beta V_m \int_0^t f(s) \exp((t-s)H_m)e_1 \,\mathrm{d}s. \tag{3.10}$$

Namely, $y_m(t) = V_m w(t)$ where now $w(t)$ solves

$$w'(t) - H_m w(t) = \beta f(t)e_1, \quad t > 0,$$

$$w(0) = 0.$$

This approach generalizes that of [1] where Arnoldi bases are used. It can also be viewed as a *reduced basis method* in the sense of [21].

Proceeding as before, we consider the residual of system (3.9) at $y_m(t)$, here denoted by $r_m^*(t)$, which is

$$r_m^*(t) = f(t)v - y_m'(t) + Ay_m(t) = h_{m+1,m}(e_m^{\mathrm{T}} w(t))v_{m+1}.$$

Since $e_m^{\mathrm{T}} w(t) = \int_0^t f(s)e_m^{\mathrm{T}} \exp((t-s)H_m)e_1 \,\mathrm{d}s$, referring to Proposition 3.1, we have

$$r_m^*(t) = \left(\int_0^t \alpha_m(t-s)f(s)\,\mathrm{d}s\right) v_{m+1}.$$

Accordingly, a restart of the procedure leads again to a problem of type (3.9). Here below we consider this restarted version of the method (for m fixed), when we choose $v_j = q_{j-1}(A)v$, being $\{q_{j-1}\}_{j=1}^{\infty}$ a sequence of polynomials satisfying (2.6).

For a fixed m, the restarted method produces a sequence of residuals $\{r_m^*(t)^{(k)}\}$, $k = 0, 1, 2, \ldots,$ where $r_m^*(t)^{(0)} := r_m(t)$ and

$$r_m^*(t)^{(1)} = \alpha_m^*(t)^{(1)} q_m(A)v$$

with

$$\alpha_m^*(t)^{(1)} = \int_0^t \alpha_m(t-s) f(s)\,\mathrm{d}s$$

and, for $k > 1$, $r_m^*(t)^{(k)}$ is the residual at the approximated solution of

$$z'(t) - Az(t) = r_m^*(t)^{(k-1)}, \quad t \geqslant 0,$$

$$z(0) = 0,$$

namely,

$$r_m^*(t)^{(k)} = \alpha_m^*(t)^{(k)} (q_m(A))^k v.$$

where

$$\alpha_m^*(t)^{(k)} = \int_0^t \alpha_m(t-s) \alpha_m^*(s)^{(k-1)}\,\mathrm{d}s. \tag{3.11}$$

The following result shows the convergence of the restarted procedure.

Proposition 3.2. *Let us set*

$$\alpha_t = \max_{0 \leqslant s \leqslant t} |\alpha(s)|.$$

Let m be any fixed positive integer, then

$$|r_m^*(t)^{(k)}| \leqslant \left(\frac{(t\alpha_t)^k}{k!} \right) \max_{0 \leqslant s \leqslant t} |f(s)| \|(q_m(A))^k v\|. \tag{3.12}$$

Proof. We proceed by induction. Clearly (3.12) holds for $k = 1$. Then, using (3.11) we easily obtain the result. □

4. Faber polynomials

Though Faber polynomials can be associated to more general sets [12,23], here we consider a compact set Ω in $\mathbb{C}$, bounded by a Jordan curve Γ. We denote by γ the *logarithmic capacity* of Ω.

Then (cf. [23]) we can consider the conformal surjection

$$\psi : \bar{\mathbb{C}} \setminus \{w\colon |w| \leqslant 1\} \to \bar{\mathbb{C}} \setminus \Omega, \quad \psi(\infty) = \infty, \ \psi'(\infty) = \gamma, \tag{4.1}$$

which has a Laurent expansion of the type

$$\psi(w) = \gamma w + c_0 + c_1 w^{-1} + c_2 w^{-2} + \cdots \ . \tag{4.2}$$

Since the boundary of Ω is assumed to be a Jordan curve, it is known that ψ has a continuous extension to $\{w \in \mathbb{C}\colon |w| \geqslant 1\}$. Let us set

$$\psi_0(w) := \psi(w) - \gamma w.$$

Then, from [12, Section 2] we have that

$$|\psi_0'(w)| \leqslant \gamma/|w|^2, \quad |w| > 1. \tag{4.3}$$

Now let $\phi : \bar{\mathbb{C}} \setminus \Omega \to \bar{\mathbb{C}} \setminus \{w: |w| \leqslant 1\}$ be the inverse mapping of ψ. The jth (*ordinary*) *Faber polynomial* associated to Ω is defined as the polynomial part of the Laurent expansion at ∞ of $[\phi(z)]^j$ (cf. [23, Section 2])

$$[\phi(z)]^j = z^j + \sum_{k=-\infty}^{j-1} \beta_{j,k} z^k, \quad j \geqslant 0,$$

that is,

$$F_j(z) := z^j + \sum_{k=0}^{j-1} \beta_{j,k} z^k, \quad j \geqslant 0.$$

For any $R \geqslant 1$, let Γ_R be the equipotential curve

$$\Gamma_R := \{z: |\phi(z)| = R\}$$

in $\bar{\mathbb{C}} \setminus \Omega$. We denote by Ω_R the compact set whose boundary is Γ_R. For our purposes we require that Ω (or some Ω_R) will contain the spectrum of A. Then, since we consider a real matrix A, from now on we assume that Ω is symmetric with respect to the real axis and convex. The same will be true for each compact Ω_R with $R \geqslant 1$ (cf. [25]).

Under our assumptions on Ω, the following further properties hold (cf. [25]):
(f1) all the coefficients c_j are real,
(f2) for $m \geqslant 0$, $|F_m(z)| \leqslant 2$, for $z \in \Omega$,
(f3) for $m \geqslant 0$, $(|w|^m - 1) < |F_m(\psi(w))| < 2|w|^m$, for $|w| > 1$.
Moreover, Faber polynomials can be defined recursively (cf. [2]) by

$$\begin{aligned} &F_0(z) = 1, \qquad \gamma F_1(z) = (z - c_0), \text{ and, for } m \geqslant 2, \\ &\gamma F_m(z) = (z - c_0) F_{m-1}(z) - (c_1 F_{m-2}(z) + \cdots + m c_{m-1} F_0(z)), \end{aligned} \tag{4.4}$$

where the coefficients $c_0, c_1, \ldots$ are those of expansion (4.2).

As well known, Faber polynomials can also be expressed by their generating function, that is we have

$$\frac{w\psi'(w)}{\psi(w) - z} = 1 + \sum_{j=1}^{\infty} F_j(z) w^{-j}, \quad z \in \Omega_r, \ r \geqslant 1, \ |w| > r. \tag{4.5}$$

According to (4.4), taking the Faber polynomials as the polynomials q_j in (2.6) and setting in (2.1)

$$v_j = F_{j-1}(A) v \quad \text{for } j \geqslant 1, \tag{4.6}$$

the entries of H_m, are given by

$$\begin{aligned} &h_{j,j} = c_0, \ h_{j+1,j} = \gamma, \ h_{1,j} = j c_{j-1} \quad \text{for every } j, \\ &\quad \text{and for } i \geqslant 2, \ h_{i,j} = c_{j-1}, \text{ for } 3 \leqslant j \leqslant i - 1. \end{aligned} \tag{4.7}$$

Moreover, by (4.4), it is $\beta = 1$.

As well known, in the particular case that Ω coincides with the closure of the internal part of an ellipse or with an interval in the complex plane, Faber polynomials are reduced to scaled and translated Chebyshev polynomials. We refer to [7,25] for a detailed description of these cases.

As a consequence of the well-known fact that the zeros of Faber polynomials are uniformly distributed on Ω [26], we have:

Proposition 4.1 (Cf. Tal-Ezer [26]). *Assume that R^* is the largest number such that $f(z)$ is analytic inside a boundary curve Γ_{R^*}. Let $p_{m-1}(z)$ be the interpolating polynomials in the zeros of $F_m(z)$, considering the respective multiplicities, then*

$$\limsup_{m\to\infty} \|f(z) - p_{m-1}(z)\|_{\Omega}^{1/m} = 1/R^*.$$

This is known as *maximal convergence property* for the sequence $\{p_{m-1}\}_{m=1}^{\infty}$.

Let us return to consider the exponential case and in particular residual (3.5), namely

$$r_m(t) = \alpha_m(t)F_m(A)v. \tag{4.8}$$

For the particular choice made here, it is often possible to get more precise estimates of $\alpha_m(t)$ and of $r_m(t)$.

Theorem 4.2. *Let Ω be symmetric with respect to the real axis and convex, for every $R > 1$,*

$$|\alpha_m(t)| \leqslant \frac{2(\exp(t\psi(R))R\gamma}{(R^m - 1)}. \tag{4.9}$$

Proof. We recall that, for every $R > 1$,

$$e_m^{\mathrm{T}} \exp(tH_m)e_1 = \frac{1}{2\pi i}\int_{\Gamma_R} \exp(tz)e_m^{\mathrm{T}}(zI - H_m)^{-1}e_1\,\mathrm{d}z,$$

hence, using (2.5), from (3.6) we get

$$\alpha_m(t) = \left(-\prod_{j=1}^{m} h_{j+1,j}\right)\frac{1}{2\pi i}\int_{\Gamma_R} (\exp(tz)/\det(zI - H_m))\,\mathrm{d}z.$$

Then, by (4.7), we get

$$\alpha_m(t) = -\gamma^m \frac{1}{2\pi i}\int_{\Gamma_R} (\exp(tz)/\det(zI - H_m))\,\mathrm{d}z$$

and, since $\det(zI - H_m) = \gamma^m F_m(z)$ (cf. (4.4) and Proposition 2.6),

$$\alpha_m(t) = \frac{-1}{2\pi i}\int_{\Gamma_R} (\exp(tz)/F_m(z))\,\mathrm{d}z,$$

that is

$$|\alpha_m(t)| \leqslant \frac{1}{2\pi}\int_{|w|=R} \frac{|(\exp(t\psi(w))\psi'(w)|}{|(F_m(\psi(w))|}\,\mathrm{d}w. \tag{4.10}$$

Then, observing that

$$|\exp(t\psi(w))| \leqslant \exp(t\psi(R)) \quad \text{for } |w| = R,$$

by (f3) and by our assumptions on Ω, from (4.10) we obtain

$$|\alpha_m(t)| \leqslant \frac{R\exp(t\psi(R))\max_{|w|=R}|\psi'(w)|}{(R^m-1)}. \tag{4.11}$$

Bound (4.9) follows from (4.11) recalling that, by (4.2) and (4.3),

$$|\psi'(w)| < 2\gamma \quad \text{for } |w| > 1. \quad \square$$

Here below we consider some cases often discussed in the literature (see e.g. [5,10,11]). For these cases, owing to the simple form of the mapping ψ, the previous general bound can be easily specialized. It is interesting to observe that the estimates are similar to those given for Krylov–Arnoldi approximations.

Proposition 4.3. *Let A be symmetric and negative semi-definite with eigenvalues in the interval $\Omega = [-4\gamma, 0]$, $\gamma > 0$. Then*

$$|\alpha_m(t)| \leqslant \frac{8t\gamma^2}{(m-1)}\exp\left(-\frac{1}{8}\frac{(m-1)^2}{\gamma t}\right)\|v\|_2, \quad 2 \leqslant m-1 \leqslant 2\gamma t, \tag{4.12}$$

$$|\alpha_m(t)| \leqslant 4\gamma\exp\left(\frac{(t\gamma)^2}{m-1} - 2t\gamma\right)\left(\frac{et\gamma}{m-1}\right)^{m-1}\|v\|_2, \quad m-1 \geqslant 2\gamma t. \tag{4.13}$$

Proof. From (4.11) we get immediately

$$|\alpha_m(t)| \leqslant \frac{2\gamma(\exp(t\psi(R))}{R^{m-1}(1-1/R)}. \tag{4.14}$$

In our case (cf. [7]), $\psi(w) = \gamma(w - 2 + w^{-1})$. Then, if $m-1 \geqslant 2\gamma t$, setting $R = (m-1)/\gamma t$ in (4.14), we easily get (4.13). Moreover, by (4.14), since $(1/R) \leqslant \exp(-(1-1/R))$, we also get

$$|\alpha_m(t)| \leqslant \frac{2\gamma\exp[t\psi(R) - (m-1)(1-1/R)]}{(1-1/R)}. \tag{4.15}$$

Hence, if $m-1 < 2\gamma t$, we set $R = 4\gamma t/(4\gamma t - m + 1)$. Since $\psi(R) = \gamma(R-1)^2/R$ and $1 < R \leqslant 2$, we have

$$\psi(R) \leqslant 2\gamma(R-1)^2/R^2. \tag{4.16}$$

In this way, using the relation $(1-1/R) = (m-1)/4\gamma t$ and inserting (4.16) in (4.15), we obtain (4.12) after simple computation. $\square$

Proposition 4.4. *Let A be a matrix with eigenvalues contained in the interval $\Omega = [\alpha\gamma - 2\gamma\mathrm{i}, \alpha\gamma + 2\gamma\mathrm{i}]$, $\gamma > 0$. Then*

$$|\alpha_m(t)| \leqslant 4\gamma\exp\left(\alpha\gamma t - \frac{(\gamma t)^2}{m-1}\right)\left(\frac{et\gamma}{m-1}\right)^{m-1}, \quad m-1 \geqslant 2\gamma t.$$

Proof. Now, the conformal mapping associated to Ω is $\psi(w) = \gamma(w + \alpha - w^{-1})$. The thesis follows straight from (4.14), setting therein $R = (m-1)/\gamma t$. $\square$

Proposition 4.5. *Assume that* $\Omega := \{z\colon |z+a| \leqslant a,\ a > 0\}$. *Then*

$$\alpha_m(t) = -\exp(-ta)\frac{t^{m-1}a^m}{(m-1)!}. \tag{4.17}$$

Proof. In this case the Faber polynomials are given by $F_j(z) = ((z/a)+1)^j$ (cf. [23, p. 133]) and one easily realizes that the interpolatory approximation coincides with the truncated Taylor expansion of $\exp(z)$ around $(-a)$. Any matrix H_m has entries

$$h_{j,j} = -a,\ h_{j+1,j} = a,\ h_{i,j} = 0 \quad \text{otherwise}.$$

Since

$$e_m^{\mathrm{T}} \exp(tH_m)e_1 = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \exp(tz) e_m^{\mathrm{T}} (zI - H_m)^{-1} e_1 \,\mathrm{d}z,$$

by Proposition 2.1, we obtain

$$\alpha_m(t) = \frac{-a^m}{2\pi \mathrm{i}} \int_\Gamma \frac{\exp(tz)}{(z+a)^m} \,\mathrm{d}z.$$

Hence, by the residue theory we get (4.17). □

If A is diagonalizable, i.e., XAX^{-1} is diagonal, using (f3) we get the bound

$$\|F_m(A)\| \leqslant \mathrm{cond}_2(X) \|F_m(\lambda)\|_{\Omega_r} \leqslant 2\,\mathrm{cond}_2(X) r^m, \quad \text{for } r \geqslant 1,$$

provided that $\sigma(A) \subseteq \Omega_r$. Other estimates are proposed below.

Proposition 4.6. *Let* Ω *be as in Proposition* 4.5 *and assume that* $W(A) \subseteq \{z : |z+a| \leqslant ra,\ r \geqslant 1\}$ (*see* (2.4)). *Then*

$$\|F_m(A)\| \leqslant \frac{r^m (m+1)^{m+1}}{m^m}. \tag{4.18}$$

Proof. For every $R > r$,

$$F_m(A) = \frac{1}{2\pi \mathrm{i}} \int_{|z+a|=Ra} F_m(z)(zI - A)^{-1} \,\mathrm{d}z.$$

Using Proposition 2.4 we obtain

$$\|F_m(A)\| \leqslant \frac{R^{m+1}}{(R-r)}$$

and, taking $R = r(m+1)/m$, we get the bound (4.18). □

An estimate of $\|F_m(A)\|$ for a general compact Ω, can be obtained as follows.

Proposition 4.7. *Assume that* $W(A) \subset \Omega_r$, *for some* $r \geqslant 1$. *Then*,

$$\|F_m(A)\| \leqslant 2r^m(2m+1)\left(\frac{m+1}{m}\right)^m. \tag{4.19}$$

Proof. Since

$$F_m(A) = \frac{1}{2\pi i}\int_{\Gamma_R} F_m(z)(zI - A)^{-1}\,dz, \quad R > r,$$

we get

$$\|F_m(A)\| \leqslant \frac{1}{2\pi}\int_{|w|=R} |F_m(\psi(w))||\psi'(w)|\|(\psi(w)I - A)^{-1}\|\,dw.$$

Hence, by Proposition 2.4, we obtain

$$\|F_m(A)\| \leqslant \frac{1}{2\pi}\int_{|w|=R} |F_m(\psi(w))|\left|\frac{\psi'(w)}{\psi(w) - u}\right|\,dw, \tag{4.20}$$

with $u \in \Omega_r$. Using (4.5), by (f2) and (f3), after simple computation one gets

$$R|(F_m(\psi(w))|\left|\frac{\psi'(w)}{\psi(w) - u}\right| \leqslant \frac{2R^m(R + r)}{(R - r)}, \quad u \in \Omega_r, \; |w| = R.$$

Then, setting $R = r(m + 1)/m$, from (4.20) we obtain (4.19). □

5. Some computational considerations

As mentioned before hybrid methods need a preliminary phase where estimates of the eigenvalues are achieved, in order to construct in a suitable way the set Ω containing $\sigma(A)$ (actually, in the case of the exponential, since it is analytic everywhere condition $\sigma(A) \subseteq \Omega$ is not essential for the convergence). To do this, in the general case, one of the several techniques proposed in the literature can be adopted. Among the others, we refer to the ones discussed in [25,17,16]. Clearly the obtained information can be re-used every time we want to apply an hybrid method to the same matrix. Nevertheless, there are also some important cases, when A represents the discretization of a differential operator, where information on the spectrum are a priori available. See for instance Example 1 below. Actually, this situation is not limited to simple cases (cf. also [14]), but an analytic study can give a priori eigenvalues estimates also for more general operators. Results upon this point will appear in a forthcoming paper.

After having defined the set Ω, we have to determine the Laurent expansion of ψ. We can proceed using the scheme proposed in [25], based on the resolution of the parameters problem relative to the Schwarz–Christoffel transformation associated to the mapping ψ, for which we refer to [27]. In order to solve this problem numerically, we employ the software SC Matlab Toolbox, written by T.A. Driscoll at MIT in 1995. Obviously, in addition to the capacity γ, only a finite number of coefficients of this expansion can be determined numerically, and so, fixing a priori this number, say p, instead of ψ we obtain the finite expansion of an approximated conformal mapping. So, formula (4.4) is a recurrence with a fixed finite number $p + 1$ of terms, and H_m is an Hessenberg matrix with upper bandwidth p. In the particular case that we compute only the first two coefficients of the Laurent expansion of ψ, that is c_0 and c_1, we work with scaled and translated Chebyshev polynomials (cf. [25]).

6. Numerical experiments

In order to illustrate the behavior of the method, we make a comparison with the Krylov method based on the Arnoldi algorithm (see e.g. [22]) on two examples arising from the semi-discretization, by the method of line (MOL), of partial differential equations of parabolic type. Obviously, when the restarted version is used, the comparison is made with the corresponding restarted version of the Krylov–Arnoldi method.

In all figures, the behavior of $\log_{10}\|r_m(t)\|_2$ with respect to the number of scalar products (taking into account of the sparsity pattern of A) is shown; $r_m(t)$ is clearly the mth residual of the corresponding IVP at time t. A continuous and a dotted line have been, respectively, used for Faber and Krylov method.

Thus, consider the following partial differential equation:

$$\frac{\partial u(x,t)}{\partial t} = Lu(x,t), \quad x \in E, \ t \geqslant 0,$$

$$u(x,0) = u_0, \quad x \in E,$$

$$u(x,t) = \sigma(x), \quad x \in \partial E, \ t > 0,$$

in which L is a second-order partial differential operator and E is an open bounded connected set. Semidiscretizing with respect to spatial variables using finite differences, a system of ordinary differential equations is achieved:

$$\begin{aligned} y'(t) &= Ay(t), \\ y(0) &= v, \end{aligned} \tag{6.1}$$

where w is a vector and A is a square matrix independent of t.

Example 6.1. In this first example let us consider the differential operator

$$L = \Delta - \tau_1 \frac{\partial}{\partial x} - \tau_2 \frac{\partial}{\partial y}, \quad \tau_1, \tau_2 \in \mathbb{R}.$$

Discretizing L on the cube $(0,1) \times (0,1) \times (0,1)$ with central differences on a uniform meshgrid of $(n+2) \times (n+2) \times (n+2)$ points with meshsize $h = 1/(n+1)$ along each direction, a nonsymmetric matrix A of order $N = n^3$ with particular block structure is obtained. It can be represented in the following way:

$$A := \frac{1}{h^2}\{I_n \otimes (I_n \otimes C_1) + [B \otimes I_n + I_n \otimes C_2] \otimes I_n\},$$

where I_n is the n-order matrix identity and

$$B := \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & \ddots & \ddots \\ & & \ddots & \ddots \end{bmatrix}, \qquad C_i := \begin{bmatrix} -2 & 1-\mu_i & & \\ 1+\mu_i & -2 & 1-\mu_i & \\ & 1+\mu_i & \ddots & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

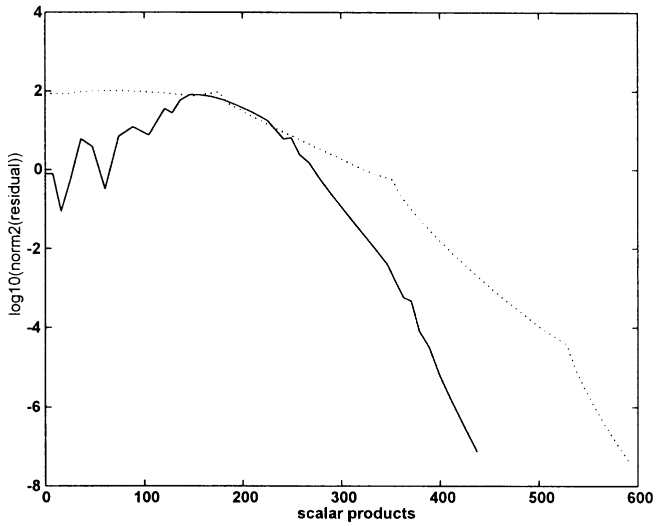

Fig. 1. $\mu_1 = 3$, $\mu_2 = 4$, $m = 10$.

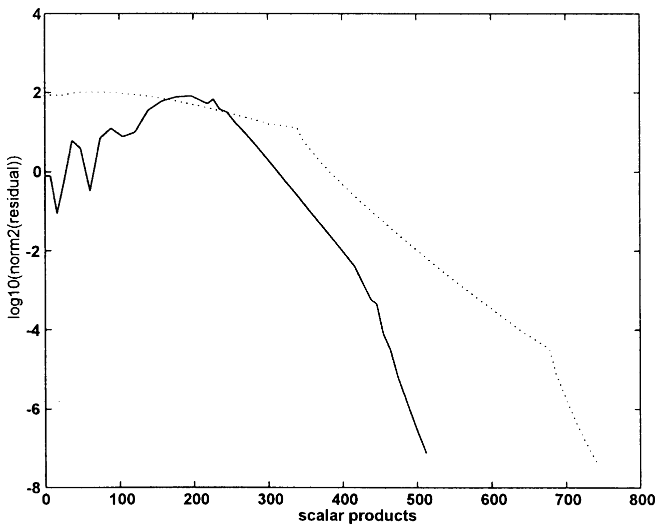

Fig. 2. $\mu_1 = 3$, $\mu_2 = 4$, $m = 15$.

for $i = 1, 2$, where $\mu_i := \tau_i(h/2)$. It is important to observe that in this case all the eigenvalues of A are explicitly known and $\sigma(A)$ is exactly contained in the rectangle

$$R(h, \mu_1, \mu_2) := \frac{1}{h^2}\left[-6 - 2\cos\left(\frac{\pi}{n+1}\right)\operatorname{Re}\delta, -6 + 2\cos\left(\frac{\pi}{n+1}\right)\operatorname{Re}\delta\right] \times \left[-2\mathrm{i}\cos\left(\frac{\pi}{n+1}\right)\operatorname{Im}\delta, 2\mathrm{i}\cos\left(\frac{\pi}{n+1}\right)\operatorname{Im}\delta\right], \qquad (6.2)$$

where $\delta := \sqrt{1 - \mu_1^2} + \sqrt{1 - \mu_2^2} + 1$.

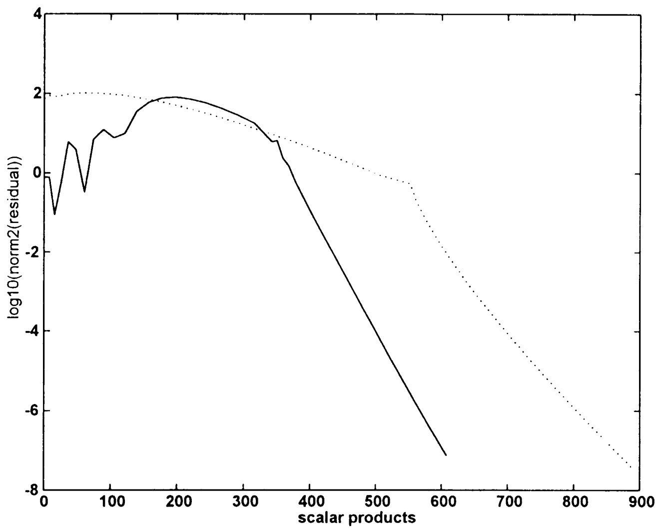

Fig. 3. $\mu_1 = 3$, $\mu_2 = 4$, $m = 20$.

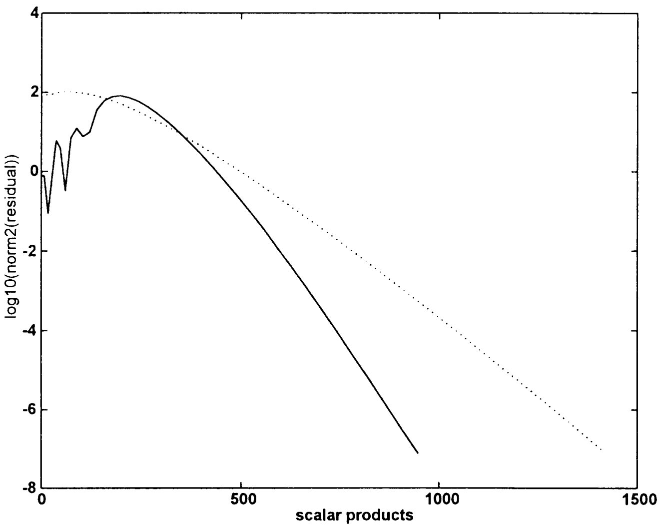

Fig. 4. $\mu_1 = 3$, $\mu_2 = 4$.

In particular, defining $n = 15$ ($N = 3375$), with $\mu_1 = 3$, $\mu_2 = 4$, $t = h^2$ and $v := (1, 1, \ldots, 1)^{\mathrm{T}}$, by (6.2) the convex hull of $\sigma(tA)$ is the rectangle

$$\Omega := \frac{1}{256} R(\tfrac{1}{16}, 3, 4) \approx [-7.9616, -4.0384] \times [-13.1453\mathrm{i}, 13.1453\mathrm{i}].$$

Computing $p = 6$ Laurent coefficients of ψ, in Figs. 1, 2, 3, 4 we observe the residual curves with restart $m = 10, 15, 20$ and without restart, respectively.

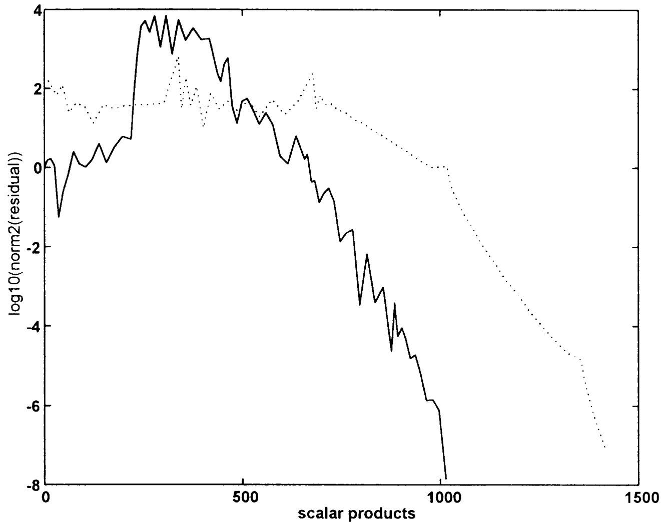

Fig. 5. $\mu_1 = 10$, $\mu_2 = 10$, $m = 15$.

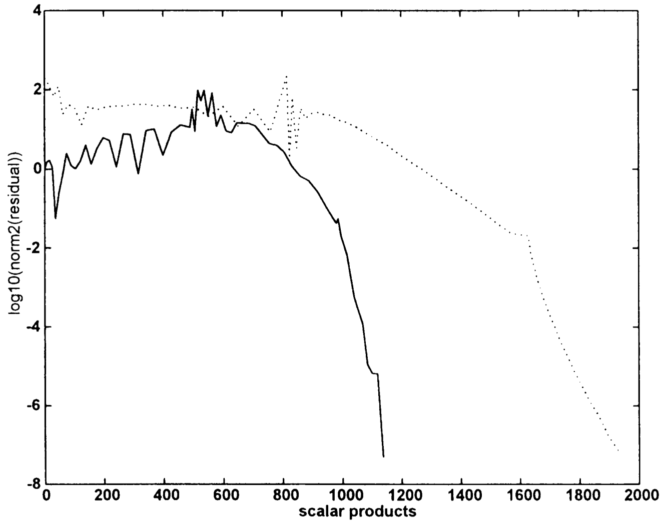

Fig. 6. $\mu_1 = 10$, $\mu_2 = 10$, $m = 25$.

Now, for the same problem with $\mu_1 = \mu_2 = 10$, the convex hull of $\sigma(tA)$ is the rectangle

$$\Omega := \tfrac{1}{256} R(\tfrac{1}{16}, 10, 10) \approx [-7.9616, -4.0384] \times [-39.0348i, 39.0348i]. \tag{6.3}$$

As before with $p = 6$ computed Laurent coefficients of ψ, in Figs. 5, 6, 7 we observe the residual curves with restart $m = 15, 25$ and without restart, respectively.

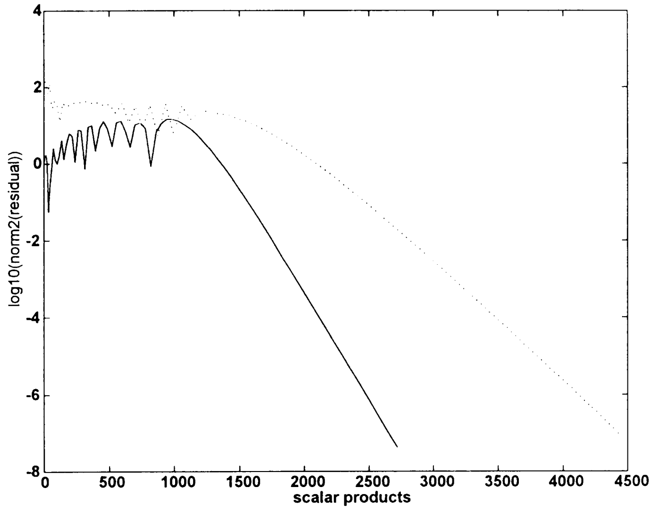

Fig. 7. $\mu_1 = 10$, $\mu_2 = 10$.

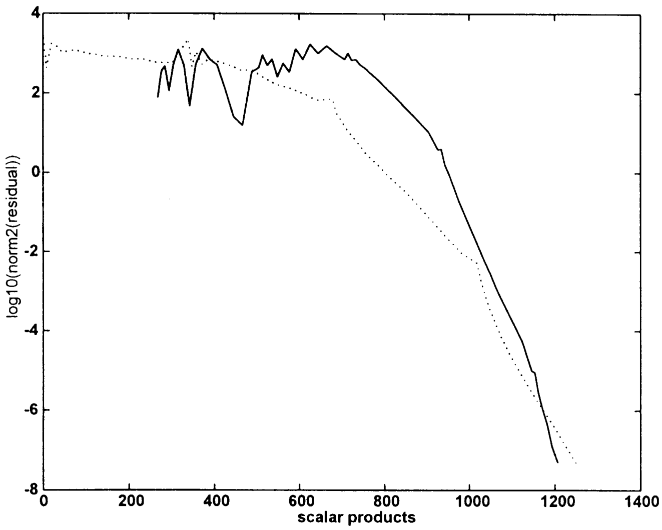

Fig. 8. $\mu = 8$, $\alpha = -2$, $m = 15$.

Example 6.2. In this second example we consider the differential operator

$$L = \Delta - \eta\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right) - \beta, \quad \gamma, \beta \in \mathbb{R}. \tag{6.4}$$

Discretizing as in Example 1 on the cube $(0,1)\times(0,1)\times(0,1)$ with uniform meshsize $h = 1/(n+1)$ along each direction, a nonsymmetric matrix A of order $N = n^3$ is obtained. Also in this case A is sparse with a particular block structure and can be represented by means of Kronecker products.

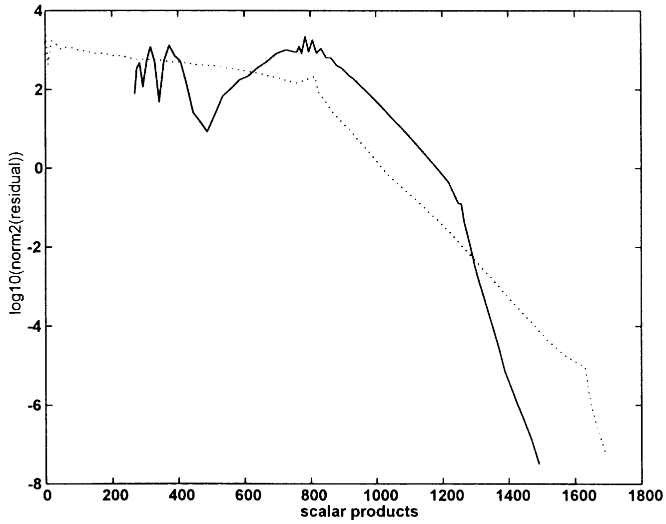

Fig. 9. $\mu = 8$, $\alpha = -2$, $m = 25$.

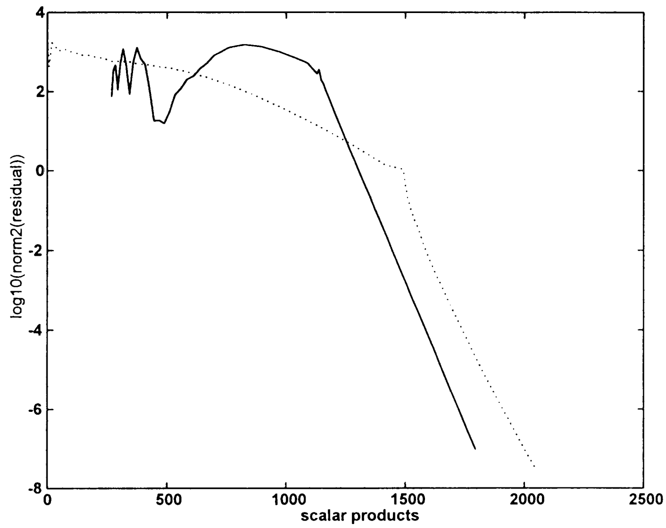

Fig. 10. $\mu = 8$, $\alpha = -2$, $m = 35$.

Let us set $n = 16$ ($N = 4096$), $v := (1, 1, \ldots, 1)^T$, and define the parameters $\mu := \eta(h/2)$ and $\alpha := \beta h^2$, $t := h^2$, setting in our experiments $\mu = 8$ and $\alpha = -2$. Following [25], by Arnoldi method we get a certain set $\{\lambda_i\}_{i=1,\ldots,s}$ of estimates of the spectrum and then we define the compact Ω as the polygon obtained joining the marginal points of this set. The cost of the computation of the p-truncated expansion of ψ is proportional to the number s of points that constitute the vertices of the polygon Ω, that is, the marginal points of $\{\lambda_i\}_{i=1,\ldots,s}$. In the experiments below we take $s = 14$, computing again $p=6$ Laurent coefficients of ψ. In Figs. 8, 9, 10 the residual curves with restart $m=15,25,35$, respectively are shown.

It is important to observe that in these last three figures the residual curve of Faber method is shifted to the right. This is due to the cost of the preliminary phase.

References

[1] E. Celledoni, I. Moret, A Krylov projection method for systems of ODEs, Appl. Numer. Math. 24 (1997) 365–378.

[2] J.H. Curtiss, Faber polynomials and Faber series, Amer. Math. Monthly 78 (1971) 577–596.

[3] K. Dekker, J.G. Verwer, Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations, North Holland, Amsterdam, 1984.

[4] V. Druskin, A. Greenbaum, L. Knizherman, Using nonorthogonal Lanczos vectors in the computation of matrix functions, SIAM J. Sci. Comput. 19 (1998) 38–54.

[5] V. Druskin, L. Knizherman, Krylov subspace approximations of eigenpairs and matrix functions in exact and computer arithmetic, Numer. Linear Algebra Appl. 2 (1995) 205–217.

[6] V. Druskin, L. Knizherman, Extended Krylov subspaces: approximation of the matrix square root and related functions, SIMAX, to appear.

[7] M. Eiermann, On semiiterative methods generated by Faber polynomials, Numer. Math. 56 (1989) 139–156.

[8] M. Eiermann, W. Niethammer, R.S. Varga, A study of semiiterative methods for nonsymmetric systems of linear equations, Numer. Math. 47 (1985) 505–533.

[9] E. Gallopoulos, Y. Saad, Efficient solution of parabolic equations by Krylov approximation methods, SIAM Sci. Statist. Comput. 13 (1992) 1236–1264.

[10] M. Hochbruck, C. Lubich, On Krylov subspace approximation to the matrix exponential operator, SIAM J. Numer. Anal. 34 (1995) 1911–1925.

[11] L. Knizherman, Calculation of functions of unsymmetric matrices using Arnoldi's method, USSR Comput. Math. Math. Phys. 31(1) (1991) 1–9 (English Edition by Pergamon Press).

[12] T. Kovari, C. Pomerenke, On Faber polynomials and Faber expansions, Math. Z. 99 (1967) 193–206.

[13] P. Lancaster, M. Tismenetsky, The Theory of Matrices, Academic Press, New York, 1985.

[14] T.A. Manteuffel, The Tchebychev iteration for nonsymmetric linear systems, Numer. Math. 28 (1977) 307–327.

[15] T.A. Manteuffel, Adaptive procedure for estimating parameters for the nonsymmetric Tchebychev iteration, Numer. Math. 31 (1978) 183–208.

[16] T.A. Manteuffel, G. Starke, On hybrid iterative methods for nonsymmetric systems of linear equations, Numer. Math. 73 (1996) 489–506.

[17] N.M. Nachtigal, L. Reichel, L.N. Trefethen, A hybrid GMRES algorithm for nonsymmetric linear systems, SIAM J. Matrix Anal. Appl. 13 (1992) 796–825.

[18] P. Novati, Metodi polinomiali per l' approssimazione di funzioni di matrici, Quad. Mat., Vol. 410, Dip. Sc. Mat., Univ. Trieste, 1997.

[19] P. Novati, A method based on Fejèr points for the computation of functions of nonsymmetric matrices. Quad. Mat., Vol. 440, Dip. Sc. Mat., Univ. Trieste, 1998.

[20] B.N. Parlett, Global convergence of the basic QR algorithm on Hessenberg matrices, Math. Comp. 22 (1968) 803–817.

[21] T.A. Porsching, M. Lin Lee, The reduced basis method for initial value problems, SIAM J. Numer. Anal. 24 (1987) 1277–1287.

[22] Y. Saad, Analysis of some Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal. 29 (1992) 209–228.

[23] V.I. Smirnov, N.A. Lebedev, Functions of a Complex Variable-Constructive Theory, Iliffe Books, London, 1968.

[24] M.N. Spijker, Numerical ranges and stability estimates, Appl. Numer. Math. 13 (1993) 241–249.

[25] G. Starke, R.S. Varga, A hybrid Arnoldi-Faber iterative method for nonsymmetric systems of linear equations, Numer. Math. 64 (1993) 213–240.

[26] H. Tal-Ezer, High degree polynomial interpolation in Newton form, SIAM J. Sci. Statist. Comput. 12 (1991) 648–667.
[27] L.N. Trefethen, Numerical computation of the Schwarz-Christoffel transformation, SIAM J. Sci. Statist. Comput. 1 (1980) 82–102.
[28] L.N. Trefethen, Pseudospectra of matrices, in: D.F. Griffiths, G.A. Watson (Eds.), Numerical Analysis, Longman, Harlow, OK, 1992.