



Numerical solution of nonclassical boundary value problems

Paola Boito¹ · Yuli Eidelman² · Luca Gemignani¹

Received: 9 February 2024 / Accepted: 12 September 2024
© The Author(s) 2024

Abstract

We provide a new approach to obtain solutions of linear differential problems set in a Banach space and equipped with nonlocal boundary conditions. From this approach we derive a family of numerical schemes for the approximation of the solutions. We show by numerical tests that these schemes are numerically robust and computationally efficient.

Keywords Differential problems · Rational approximation · Functions of matrices and operators

1 Introduction

In this work we focus on the solution of a class of linear differential problems of the form

$$\frac{dv}{dt} = Av, \quad 0 < t < T, \quad (1.1)$$

set in a Banach space, with the nonlocal integral condition

$$\frac{1}{T} \int_0^T v(t) dt = f,$$

where the function f and the linear, closed, possibly unbounded operator A are given.

Paola Boito and Luca Gemignani are members of GNCS-INDAM.

✉ Paola Boito
paola.boito@unipi.it
Yuli Eidelman
eideyu@tauex.tau.ac.il
Luca Gemignani
luca.gemignani@unipi.it

¹ Dipartimento di Matematica, Università di Pisa, Pisa, Italy

² School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv, Israel

Recall that, given Banach spaces X and Y equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, the norm of a bounded linear operator $A : D(A) \subset X \rightarrow Y$ is defined as

$$\|A\| = \sup_{x \in D(A), x \neq 0} \frac{\|Ax\|_Y}{\|x\|_X}.$$

Here $D(A)$ denotes the domain of A , which is a linear subspace of X . Saying that A is closed means that whenever $x_k \rightarrow x$ in $D(A)$ and $Ax_k \rightarrow y$ in Y , we have $Ax = y$.

The hypothesis that A is (possibly) unbounded with respect to the operator norm allows us to choose A as a differential operator. For instance, A could be taken as a second derivative w.r.t. a space variable, yielding the familiar form of the homogeneous heat equation, albeit with less common boundary conditions. The study of the heat equation with integral boundary conditions goes back to Cannon [5]; the existence and properties of the solution are investigated in [15, 17]. See also [11] and references therein. Another example of application, this time to system identification, is found in Bellman [1]. In the same vein, note that the evolution problem (1.1) fits into a more general framework of differential problems where the boundary conditions are expressed through the application of linear operators to the unknown function [13]; this includes integral conditions, but also periodic conditions, multipoint conditions and so on.

Let us recall that the theory of operator methods for differential problems provides a setting for the solution of problems such as (1.1): for a general presentation see for instance [7, 14] and references therein.

In a finite-dimensional setting, A is a linear operator acting on a finite-dimensional space, that is, a matrix. The solution to our differential problem can then be expressed as the action of a function of A on f , namely, $v(t) = \psi_t(A)f$, where $\psi_t(z) = \frac{Tze^{zt}}{e^zT-1}$. Note in passing that this function is closely related to the reciprocal of the φ_1 function, which is of interest in the development of exponential integrators: see e.g., [3, 4, 9, 10] and references therein.

A discussion of the finite-dimensional problem was proposed in [2] and [3]. More precisely, Boito et al. [2] introduced an expression for $v(t)$ based on a partial fraction decomposition of the function $\psi_t(z)$, thus motivating the development of a structured algorithm for the solution of shifted quasiseparable linear systems. It also presented an acceleration technique for this rational decomposition, which introduced a cubic polynomial term. In [3], generalization of this idea led to a family of mixed polynomial-rational approximations of $v(t)$, where the polynomial part is in fact a Bernoulli polynomial of arbitrary degree. In addition to its theoretical interest, this approximation formula allows for the design of effective numerical methods for the computation of $v(t)$, particularly when A is a structured matrix. Let us also mention that the approach based on Bernoulli polynomials for the solution of differential problems with nonlocal conditions was suggested for the first time in Eidelman et al. [6].

For the more general case where the differential problem is set in a Banach space, we prove the existence and uniqueness of the solution $v(t)$ and characterize it via a family of mixed polynomial-rational expansions w.r.t. the operator A . Each expansion contains a purely polynomial term of arbitrary degree, which is related to the Bernoulli

polynomials, followed by a series of rational terms. The evaluation of each rational term involves the inversion of A plus a multiple of the identity operator.

From this result we derive a general numerical procedure for computing an approximation of $v(t)$ up to a given tolerance (Algorithm 1). An interesting feature of this approach is the fact that successive rational terms can be computed independently: this allows us to fine-tune the accuracy of the approximation by adding further terms as needed, without the need to recompute the whole approximation. Moreover, in order to improve the efficiency of the implementation, one may also employ parallelization and/or structured methods for the inversion of families of shifted operators; a finite-dimensional example in presence of quasiseparable structure is found in Boito et al. [2].

Numerical tests highlight the effectiveness of this approach and suggest strategies for a good choice of the degree of the polynomial term and of the number of rational terms. After providing results for the matrix case, we focus on a model problem of parabolic equation. For this problem we investigate the behavior of the mixed approximation in combination with two different approaches: the classical method of lines, based on a finite-difference semi-discretization in space, and a “functional” approach where the expansion is applied directly to the infinite-dimensional operator. Each rational term is then computed as the solution of a boundary value problem. Among the proposed tests, this functional approach is the most innovative and promising application of our mixed approximation formula. While more computationally demanding, it circumvents the numerical obstacles encountered in the finite-dimensional discretization and proves to be more robust and flexible.

The paper is organized as follows. Section 2 recalls the problem under study and contains the main theoretical result of the paper, namely, the statement and proof for the mixed polynomial-rational expansion of the solution (Theorem 1). This is followed by Algorithm 1, which computes an approximation of $v(t)$. Sections 3 and 4 are devoted to numerical tests for the matrix case and for the parabolic model equation, respectively. Section 5 summarizes our contributions and sets out ideas for future work, particularly for further investigation and improvement of the functional approach.

2 The abstract nonlocal differential problem

Let X be a Banach space. Consider the linear homogeneous differential equation

$$\frac{dv}{dt} = Av, \quad 0 < t < T, \tag{2.1}$$

with the boundary conditions

$$\frac{1}{T} \int_0^T v(t) dt = f. \tag{2.2}$$

Here $A: D(A) \rightarrow X$ is a linear unbounded closed operator with domain $D(A) \subset X$, and $f \in D(A)$. By solution of the problem (2.1), (2.2) we mean a continuous

function $v : [0, T] \rightarrow X$ on $[0, T]$ with values in X and with $v(t) \in D(A)$, $0 < t < T$, such that $v \in C^1((0, T); X)$, that is, v is a continuously differentiable function, $Av \in C((0, T); X)$, and (2.1), (2.2) hold. Without loss of generality one can assume that $T = 2\pi$.

We say that $\lambda \in \mathbb{C}$ is a regular point of A if $A - \lambda I$, $I = I_X$, is bijective and $(A - \lambda I_X)^{-1}$ is a bounded operator, i.e.,

$$\| (A - \lambda I)^{-1} \| = \sup \left\{ \frac{\| (A - \lambda I)^{-1} y \|}{\| y \|} : y \neq 0 \right\} < +\infty.$$

The following result establishes the basis for the design of numerical schemes for computing a numerical approximation of the solution $v(t)$. A preliminary version of this result in the finite dimensional case has first appeared in [2, 3].

Theorem 1 *Assume that all the complex numbers*

$$\mu_k = ik, \quad k = \pm 1, \pm 2, \dots$$

are regular points of the operator A and there is a constant $C > 0$ such that

$$\| (A - \mu_k I)^{-1} \| \leq \frac{C}{|k|}, \quad k = \pm 1, \pm 2, \dots \tag{2.3}$$

Also, suppose that $f \in D(A^2)$. Then the problem (2.1), (2.2) has a unique solution which is given by the formula

$$v(t) = f + (t - \pi)Af + 2 \left(\sum_{k=1}^{\infty} \Sigma_k(Af) \cos kt + \sum_{k=1}^{\infty} \Upsilon_k(Af) \sin kt \right) \tag{2.4}$$

with

$$\begin{cases} \Sigma_k = A(A + ikI)^{-1}(A - ikI)^{-1}, \\ \Upsilon_k = k^{-1}A^2(A + ikI)^{-1}(A - ikI)^{-1}, \end{cases} \quad k = 1, 2, \dots \tag{2.5}$$

Moreover for any $n \geq 0$ under the additional assumption $f \in D(A^{2n+2})$ the solution $v(t)$ satisfies the formula

$$v(t) = p_n(t) + s_n(t) \tag{2.6}$$

with

$$p_n(t) = \sum_{k=0}^{2n+1} \frac{(2\pi)^k}{k!} B_k\left(\frac{t}{2\pi}\right) A^k f, \tag{2.7}$$

and

$$s_n(t) = (-1)^n 2 \sum_{k=1}^{\infty} \frac{1}{k^{2n}} (\Sigma_k(A^{2n+1} f) \cos kt + \Upsilon_k(A^{2n+1} f) \sin kt), \tag{2.8}$$

where $B_m(t)$ are the well-known Bernoulli polynomials:

$$B_m(t) = \sum_{j=0}^m \frac{1}{j+1} \sum_{k=0}^j (-1)^k \binom{j}{k} (t+k)^m.$$

Proof For any $f \in D(A^2)$ we consider the sequence of partial sums

$$v_n(t) = f + (t - \pi)Af + 2 \left(\sum_{k=1}^n \Sigma_k(Af) \cos kt + \sum_{k=1}^n \Upsilon_k(Af) \sin kt \right), \quad n = 0, 1, 2, \dots \tag{2.9}$$

At first we prove that the sequence $v_n(t)$ in (2.9) converges uniformly in t on $[0, 2\pi]$. Indeed using the first formula in (2.5) we have

$$\|\Sigma_k(Af) \cos(kt)\| \leq \|(A + ikI)^{-1}\| \|(A - ikI)^{-1}\| \|A^2 f\|, \quad k = 0, 1, 2, \dots \quad 0 \leq t \leq 2\pi$$

and from (2.3) follows the uniform convergence in $t \in [0, 2\pi]$ for the first series in (2.9). Using the second formula in (2.5) we get

$$\Upsilon_k(Af) \sin kt = k^{-1} A(A + ikI)^{-1} (A - ikI)^{-1} \sin kt (A^2 f), \quad k = 1, 2, \dots \quad 0 \leq t \leq 2\pi.$$

Using the equality $k^{-1}A = k^{-1}(A - ikI) + iI$ we get

$$\begin{aligned} \Upsilon_k(Af) \sin kt &= k^{-1}(A + ikI)^{-1} (A^2 f) \sin kt + i(A^2 + k^2 I)^{-1} (A^2 f) \sin kt, \quad 0 \leq t \leq 2\pi, \\ &k = 1, 2, \dots \end{aligned}$$

which implies the uniform convergence in $t \in [0, 2\pi]$ for the second series in (2.9). Thus there is a continuous function $v(t)$ on $[0, 2\pi]$ which is a uniform limit of the sequence $v_n(t)$ on $[0, 2\pi]$. Moreover using the equalities

$$\int_0^{2\pi} \cos kt \, dt = \int_0^{2\pi} \sin kt \, dt = 0, \quad k = 1, 2, \dots, \quad \int_0^{2\pi} (t - \pi) \, dt = 0$$

we get

$$\int_0^{2\pi} v_n(t) \, dt = 2\pi f, \quad n = 0, 1, 2, \dots$$

and passing to the limit as $n \rightarrow \infty$ we obtain the condition (2.2).

Now consider the sequences

$$Av_n(t) = Af + (t - \pi)A^2 f + 2 \left(\sum_{k=1}^n \Sigma_k(A^2 f) \cos kt + \sum_{k=1}^n \Upsilon_k(A^2 f) \sin kt \right), \quad n = 0, 1, 2, \dots \tag{2.10}$$

and

$$\frac{dv_n}{dt} = Af + 2 \left(\sum_{k=1}^n \Sigma_k(Af)(-k) \sin kt + \sum_{k=1}^n \Upsilon_k(Af)k \cos kt \right), \quad n = 0, 1, 2, \dots \tag{2.11}$$

Comparing (2.10) with (2.11) and using (2.5) we get

$$\frac{dv_n}{dt} + \phi(t)A^2 f = Av_n(t) + \phi_n(t)A^2 f$$

with

$$\phi(t) = t - \pi, \quad \phi_n(t) = - \sum_{k=1}^n \frac{2}{k} \sin kt$$

One can easily see that $\phi_n(t)$, $n = 1, 2, \dots$ is the sequence of the partial sums of the Fourier series of the function $\phi(t)$ on the segment $[0, 2\pi]$. Hence it follows that $\phi_n(t)$ converges to $\phi(t)$ as $n \rightarrow \infty$ uniformly in t on $[\delta, \gamma]$ for any δ, γ with $0 < \delta < \gamma < 2\pi$. One should prove that the sequences $Av_n(t)$ in (2.10) and $\frac{dv_n}{dt}$ in (2.11) converge uniformly in t on $[\delta, \gamma]$. To this end one can check the uniform convergence in t on $[\delta, \gamma]$ of the sums

$$\sum_{k=1}^n \Sigma_k \cos kt, \quad \sum_{k=1}^n \Upsilon_k \sin kt. \tag{2.12}$$

We apply the Abel transformation to these sums. We have

$$\sum_{k=1}^n \Sigma_k \cos kt = \sum_{k=1}^{n-1} (\Sigma_k - \Sigma_{k+1}) \sum_{j=1}^k \cos jt + \Sigma_n \sum_{j=1}^n \cos jt.$$

Using the formula

$$\sum_{j=1}^k \cos jt = -\frac{1}{2} + \frac{1}{2} \frac{\sin \frac{2k+1}{2}t}{\sin \frac{t}{2}}$$

we get

$$\begin{aligned} \sum_{k=1}^n \Sigma_k \cos kt &= -\frac{1}{2} \sum_{k=1}^{n-1} (\Sigma_k - \Sigma_{k+1}) + \frac{1}{2} \sum_{k=1}^{n-1} (\Sigma_k - \Sigma_{k+1}) \frac{\sin \frac{2k+1}{2}t}{\sin \frac{t}{2}} \\ &\quad - \frac{1}{2} \Sigma_n + \frac{1}{2} \Sigma_n \frac{\sin \frac{2n+1}{2}t}{\sin \frac{t}{2}} \end{aligned}$$

which implies

$$\sum_{k=1}^n \Sigma_k \cos kt = -\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_n \frac{\sin \frac{2n+1}{2}t}{\sin \frac{t}{2}} + \frac{1}{2} \sum_{k=1}^{n-1} (\Sigma_k - \Sigma_{k+1}) \frac{\sin \frac{2k+1}{2}t}{\sin \frac{t}{2}}.$$

Similarly for the second series we have

$$\sum_{k=1}^n \Upsilon_k \sin kt = \sum_{k=1}^{n-1} (\Upsilon_k - \Upsilon_{k+1}) \sum_{j=1}^k \sin jt + \Upsilon_n \sum_{j=1}^n \sin jt.$$

Using the formula

$$\sum_{j=1}^k \sin jt = \frac{1}{2} \cot \frac{t}{2} - \frac{1}{2} \frac{\cos \frac{2k+1}{2}t}{\sin \frac{t}{2}}$$

we get

$$\begin{aligned} \sum_{k=1}^n \Upsilon_k \sin kt &= \frac{1}{2} \cot \frac{t}{2} \sum_{k=1}^{n-1} (\Upsilon_k - \Upsilon_{k+1}) - \frac{1}{2} \sum_{k=1}^{n-1} (\Upsilon_k - \Upsilon_{k+1}) \frac{\cos \frac{2k+1}{2}t}{\sin \frac{t}{2}} + \\ &\quad \frac{1}{2} \cot \frac{t}{2} \Upsilon_n - \frac{1}{2} \Upsilon_n \frac{\cos \frac{2n+1}{2}t}{\sin \frac{t}{2}} \end{aligned}$$

which implies

$$\sum_{k=1}^n \Upsilon_k \sin kt = \frac{1}{2} \cot \frac{t}{2} \Upsilon_1 - \frac{1}{2} \Upsilon_n \frac{\cos \frac{2n+1}{2}t}{\sin \frac{t}{2}} - \frac{1}{2} \sum_{k=1}^{n-1} (\Upsilon_k - \Upsilon_{k+1}) \frac{\cos \frac{2k+1}{2}t}{\sin \frac{t}{2}}.$$

In order to prove convergence of (2.12), we need to study the behavior of Σ_k , Υ_k , $\Sigma_k - \Sigma_{k+1}$ and $\Upsilon_k - \Upsilon_{k+1}$ as $k \rightarrow \infty$. Using (2.5) we have

$$\begin{aligned} \Sigma_k - \Sigma_{k+1} &= A(A^2 + k^2I)^{-1} - A(A^2 + (k+1)^2I)^{-1} \\ &= A(A^2 + k^2I)^{-1}(A^2 + (k+1)^2I)^{-1}[(A^2 + (k+1)^2I) - (A^2 + k^2I)] \\ &= A(2k+1)(A^2 + k^2I)^{-1}(A^2 + (k+1)^2I)^{-1} \end{aligned}$$

and analogously

$$\Upsilon_k - \Upsilon_{k+1} = \frac{1}{k(k+1)}(A^2 + (3k^2 + 3k + 1)I)(A^2 + k^2I)^{-1}(A^2 + (k+1)^2I)^{-1}.$$

Now observe that from the identity

$$\frac{A}{k} = \frac{1}{k}(A - ikI) + iI$$

it follows

$$\frac{A}{k}(A - ikI)^{-1} = \frac{1}{k} + i(A - ikI)^{-1}. \tag{2.13}$$

Using the inequalities (2.3) together with (2.13) we get

$$\|\Sigma_k - \Sigma_{k+1}\|, \|\Upsilon_k - \Upsilon_{k+1}\| \leq \frac{K}{k^2}, \quad k = 1, 2, \dots, \tag{2.14}$$

where K is a constant. Moreover using (2.5) we have

$$\Sigma_k = (A + ikI)^{-1} + ik(A^2 + k^2I)^{-1}, \quad \Upsilon_k = \frac{1}{k} - k(A^2 + k^2I)^{-1}.$$

This by virtue of (2.3) implies

$$\lim_{n \rightarrow \infty} \Sigma_n = \lim_{n \rightarrow \infty} \Upsilon_n = 0. \tag{2.15}$$

Finally for any δ, γ with $0 < \delta < \gamma < 2\pi$ we have

$$\left| \sin \frac{t}{2} \right| \geq c > 0, \quad 0 < \delta \leq t \leq \gamma < 2\pi. \tag{2.16}$$

Thus combining the relations (2.14), (2.15), (2.16) together we conclude that the sums (2.12) converge uniformly in t on $[\delta, \gamma]$. Hence the same is true for the sequences (2.10), (2.11). Since A is a closed operator, we obtain that $Av_n(t) \rightarrow Av(t)$. Analogously, the uniform convergence of the derivatives implies that the derivative of the limit is the limit of the derivatives, and, hence, $\frac{dv_n}{dt} \rightarrow \frac{dv}{dt}$. Therefore, we conclude that $v(t)$ is a solution of the equation (2.1).

Proceeding in the same way as in [2, 3] with the assumption $f \in D(A^{2n+2})$ we obtain the formulas (2.6), (2.7), (2.8) for the solution $v(t)$.

Let us now prove the uniqueness of the solution. Let $v(t)$ be a solution of the linear homogeneous differential problem

$$\frac{dv}{dt} = Av, \quad 0 < t < T, \tag{2.17}$$

$$\frac{1}{2\pi} \int_0^{2\pi} v(t) dt = 0. \tag{2.18}$$

Note that by integrating (2.17) from 0 to 2π and applying (2.18) we obtain

$$v(2\pi) - v(0) = \int_0^{2\pi} \frac{dv}{dt} dt = A \int_0^{2\pi} v(t) dt = 0. \tag{2.19}$$

The function $v(t)$ is continuous on $[0, 2\pi]$ and has Fourier coefficients c_k , $k = 0, \pm 1, \pm 2, \dots$ given by

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} v(t) e^{-\mu_k t} dt, \quad k = \pm 1, \pm 2, \pm$$

and

$$c_0 = \frac{1}{2\pi} \int_0^{2\pi} v(t) dt = 0. \tag{2.20}$$

Multiplying the equation (2.17) by $\frac{1}{2\pi} e^{-\mu_k t}$ and integrating from 0 to 2π we get

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{dv}{dt} e^{-\mu_k t} dt = A \frac{1}{2\pi} \int_0^{2\pi} v(t) e^{-\mu_k t} dt, \quad k = \pm 1, \pm 2, \dots$$

Integrating by parts in the left hand side and applying (2.19) we get

$$\mu_k c_k = A c_k, \quad k = \pm 1, \pm 2, \dots$$

which implies

$$c_k = 0, \quad k = \pm 1, \pm 2, \dots \tag{2.21}$$

Thus from (2.21), (2.20) it follows that all the Fourier coefficients of the continuous function $v(t)$ are zeros. Hence $v(t) = 0$, $0 \leq t \leq 2\pi$. \square

Based on this theorem, we get a family of polynomial/rational approximations of the solution of (2.1), (2.2) of the form:

$$v(t) \simeq v_{n,\ell}(t) = p_n(t) + s_{n,\ell}(t) \tag{2.22}$$

where $s_{n,\ell}(t)$ is a partial sum of the series (2.8), i.e.

$$s_{n,\ell}(t) = (-1)^n 2 \sum_{k=1}^{\ell} \frac{1}{k^{2n}} (\Sigma_k(A^{2n+1} f) \cos kt + \Upsilon_k(A^{2n+1} f) \sin kt)$$

and

$$v(t) = v_{n,\ell}(t) + r_{n,\ell}(t)$$

with residual

$$r_{n,\ell}(t) = (-1)^n 2 \sum_{k=\ell+1}^{+\infty} \frac{1}{k^{2n}} (\Sigma_k(A^{2n+1} f) \cos kt + \Upsilon_k(A^{2n+1} f) \sin kt).$$

The bulk of the paper deals with theoretical and computational issues associated with the computation of $v_{n,\ell}(t)$. The results of a preliminary numerical experience reported in [2], [3] and [8] upon the finite dimensional case promote some considerations concerning the selection of the parameters n and ℓ .

1. In the finite dimensional case it is shown that the larger the degree of $p_n(t)$ is, the better the convergence of the series of $v(t)$ is. However, in finite precision arithmetic large values of n generally lead to stability issues in the computation of the approximant (2.22). Moreover, in view of Theorem 1 large values of n imply additional smoothness requirements upon the function $f(t)$ in (2.2). Therefore, in our numerical experience the better strategy is to set the value of n as small as possible, typically $n \in \{0, 1, 2, 3, 4\}$, and then determine the value of ℓ so as to obtain the desired accuracy.
2. An adaptive technique for the selection of the value of ℓ in the finite dimensional case is presented in Gemignani [8]. For a fixed $t \in [0, T]$ once the value of n and a threshold tolerance $\epsilon > 0$ are given, then we can determine the value of $\ell = \ell(t)$ by imposing the condition

$$\| v_{n,\ell-1}(t) - v_{n,\ell}(t) \| / \| v_{n,\ell}(t) \| \leq \epsilon.$$

In this way the approximation $v_{n,\ell}(t)$ is constructed incrementally by adding one term at a time until a fixed tolerance is reached.

According to these facts, we consider the following general procedure for computing an approximation $v_{n,\ell}(t)$ of $v(t)$ for $t \in [0, T]$.

Algorithm 1

- 1: **Select** the values of $n \in \{0, 1, 2, 3, 4\}$ and tol .
- 2: **Define** a coarse grid of points $\mathcal{I}_S = \{t_0, \dots, t_m\} \subset [0, T]$.
- 3: **Compute**, for any $t_i \in \mathcal{I}_S$, $v_{n,\ell_i}(t_i)$ as in (2.22) where ℓ_i is the minimum index such that

$$\| v_{n,\ell_i-1}(t_i) - v_{n,\ell_i}(t_i) \|_\infty / \| v_{n,\ell_i}(t_i) \|_\infty \leq tol \tag{2.23}$$

holds.

- 4: **Set** $\widehat{\ell} = \max_{1 \leq i \leq m} \ell_i$ be the length of our rational expansion.
 - 5: **return** $v_{n,\widehat{\ell}}(t)$ as the approximation of $v(t)$ over $[0, T]$.
-

The computational effort of **Algorithm 1** at step 3 basically amounts to determine $\Sigma_k(Af)$ and $\Upsilon_k(Af)$ for $k = 1, \dots, \widehat{\ell}$. The accuracy of the approximation $v_{n,\widehat{\ell}}(t)$ returned as output by **Algorithm 1** is measured by introducing a finer grid of points $\mathcal{I}_L \subset [0, T]$ and then setting

$$err = \max_{t_i \in \mathcal{I}_L} \| v(t_i) - v_{n,\widehat{\ell}}(t_i) \|_\infty / \| v(t_i) \|_\infty . \tag{2.24}$$

The selection of grid points can be problem dependent or related to specific properties of the solution function. For testing purposes we generally make use of equispaced grid of points. Theoretical estimates of the norm of the residual $r_{n,\ell}(t)$ in terms of the quantities (2.23) are currently unavailable. The feasibility and robustness of **Algorithm 1** is validated by numerical simulations shown in the next sections.

3 The matrix case: Numerical tests

Here it is assumed that $X = \mathbb{C}^N$, A is an $N \times N$ matrix with spectrum disjoint from the set $\{\pm ik \in \mathbb{C} : k = 1, 2, \dots\}$, f is an N -dimensional column and $v(t) = \text{col}(v_i(t))_{i=1}^N$ is a vector function. For testing purposes let us suppose that A is diagonalizable by a unitary congruence, that is, $A = Q^H D Q$ with $Q^H Q = I_N$ and $D = \text{diag}[\lambda_1, \dots, \lambda_N]$. Then it can be proved by direct calculations that the solution of (2.1),(2.2) can be expressed as

$$v(t) = Q^H \text{diag}[e^{\lambda_1 t} \psi_1(\lambda_1 2\pi), \dots, e^{\lambda_N t} \psi_1(\lambda_N 2\pi)] Q f, \tag{3.1}$$

where $\psi_1(z) = z/(e^z - 1)$ is a meromorphic function with poles $z_k = 2\pi \mu_k = 2\pi ik$, $k = \pm 1, \pm 2, \dots$. It is tacitly assumed that the computation of $v(t)$ by means of (3.1) gives the "exact" solution of (2.1),(2.2). Thus, such solution can be used in our numerical simulations to provide error estimates according to (2.24).

We have implemented the computation of the approximation $v_{n,\ell}(t)$ of $v(t)$ as described in (2.22) and **Algorithm 1**. The implementation is done in MATLAB and it always works in double precision, unless otherwise noted.

The coarse and fine grid of points are formed by equispaced nodes in the interval $[0, T]$. The input parameters are $m, n \in \mathbb{N}$ and $tol \in \mathbb{R}$. Set $\mathcal{I}_S = \{t_i = 2\pi(i-1)/(m-1) : 1 \leq i \leq m\}$ the set of m uniformly spaced points in the interval $[0, T] = [0, 2\pi]$. The finer grid of points is determined so that $\mathcal{I}_L = \{t_i = 2\pi(i-1)/(m-1)^2 : 1 \leq i \leq (m-1)^2 + 1\}$ includes \mathcal{I}_S .

Our test suite consists of the following set of rank-structured matrices:

1. Random unitary Hessenberg matrices. Test problems with known eigenvalues are generated as follows. A unitary diagonal matrix D is generated and its eigenvalues noted. A unitary matrix Q , random with respect to Haar measure, is generated, and the random unitary matrix $B = Q^H D Q$ formed. Then B is transformed to upper Hessenberg form by unitary congruence, that is, $P^H B P = H$, to yield an upper Hessenberg unitary matrix H with known eigenvalues, which is then factored into the form $H = (QP)^H D (QP)$. The matrix $V = (QP)^H$ is the unitary eigenvector matrix.

One could also choose other ways of generating an upper Hessenberg unitary matrix, such as applying a sequence of $N-1$ Givens transformations to the diagonal matrix D containing the eigenvalues.

2. Banded circulant matrices. We consider banded circulant matrices of the form

$$C = \begin{bmatrix} a_0 & & a_2 & a_1 \\ a_1 & \ddots & & a_2 \\ a_2 & \ddots & \ddots & \\ & \ddots & \ddots & \ddots \\ & & a_2 & a_1 & a_0 \end{bmatrix} \in \mathbb{C}^{N \times N}.$$

We recall that C can be diagonalized by a Discrete Fourier Transform. Specifically, if $F = (f_{j,\ell}), f_{j,\ell} = \frac{e^{i2\pi(j-1)(\ell-1)/N}}{\sqrt{N}}, 1 \leq j, \ell \leq N$, is the Fourier matrix of order N then $F^H C F = \text{diag}[\lambda_1, \dots, \lambda_N]$ with

$$\lambda_k = \sum_{j=0}^2 a_j e^{-i2\pi(k-1)j/N}, \quad 1 \leq k \leq N.$$

3. Shifted and scaled 1D Laplacian matrices. We consider $N \times N$ tridiagonal Toeplitz matrices generated via the MATLAB command `gallery` as

$$A = \sigma \text{gallery}('tridiag', N, 1, -2, 1) - \gamma \text{eye}(N) \tag{3.2}$$

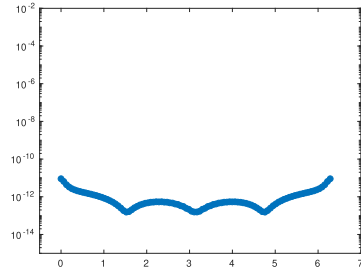
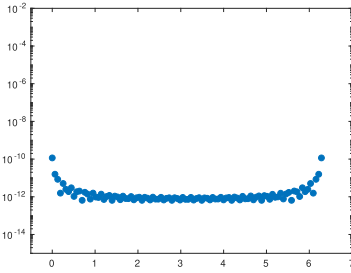
with $\sigma, \gamma \in \mathbb{R}^+$. These matrices can be diagonalized by means of a discrete sine transform $E = (\sqrt{2/(N+1)} \sin \frac{\pi i j}{N+1})_{i,j=1}^N$.

The figures and tables below show the results of our numerical experiments.

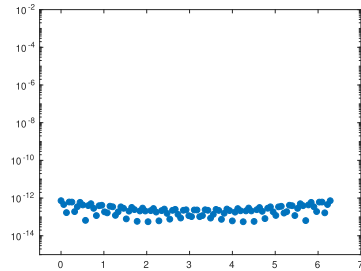
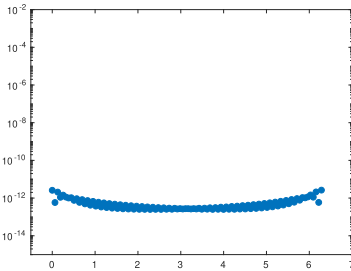
Random unitary Hessenberg matrices provide an easy test class. The eigenvalues are located on the unit circle in the complex plane, the norm of A is bounded by 1 and the solution of the shifted linear systems associated with the computation of $\Sigma_k(Af)$ and $\Upsilon_k(Af)$ in (2.4), (2.5) is generally well conditioned. Therefore, our proposed schemes perform quite well in this case. In Fig. 1 we illustrate the plots of the measured error over the grid \mathcal{I}_L for $N = 1024, \text{tol} = 1.0e-12$ and different values of $n \in \{1, 2, 3, 4\}$. Notice that the error is flattened out as n increases and, moreover, the stopping tolerance gives a quite precise measure of the norm of the residual. We guess that the error behaviour is determined by the decreasing of the algorithmic error in the computation of the mixed polynomial/rational approximation of the differential problem.

To experiment with more difficult tests we consider banded circulant matrices. For $a_0 = a_1 = a_2 = 1$ and even values of N , the matrix C has one eigenvalue equal to i . In view of Theorem 1 the solution of (2.1),(2.2) does not exist or is not unique. The implementation of our code reports "ill-conditioning" warnings in MATLAB's backslash command used to solve the shifted linear systems involved in the computation of the rational approximant $s_n(t)$.

Perturbed values of the coefficients make it possible to tune the conditioning of the shifted linear systems. In Fig. 2 we show the plots of the computed error for



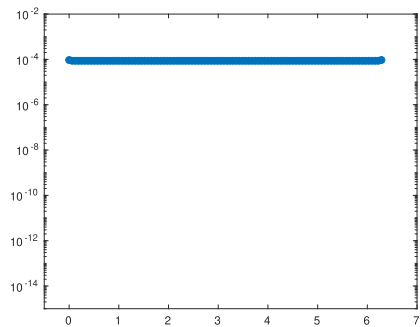
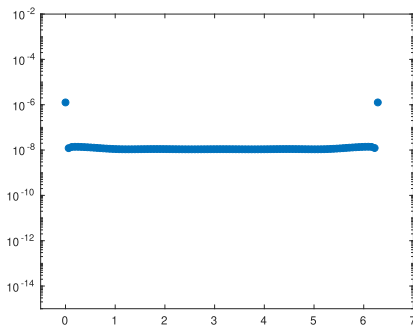
(a) Plot of computed error for $n = 1$. (b) Plot of computed error for $n = 2$.



(c) Plot of computed error for $n = 3$. (d) Plot of computed error for $n = 4$.

Fig. 1 Illustration of the measured error (2.24) for Test 1 with $tol = 1.0e - 12$ and $m = 10$ for different values of $n \in \{1, 2, 3, 4\}$

$N = 1024$, $tol = 1.0e - 12$, $n = 1$ and $a_0 = a_1 = 1, a_2 = 1 + 1.0e - 4$ and $a_0 = a_1 = 1, a_2 = 1 + 1.0e - 6$, respectively. This figure clearly indicates the impact of the conditioning of the shifted linear systems on the overall accuracy of the computed approximation.



(a)

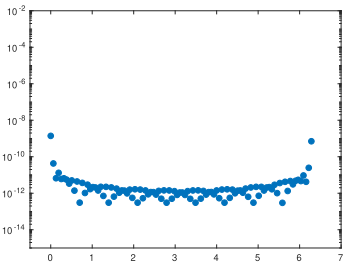
(b)

Fig. 2 Illustration of the measured error (2.24) for Test 2 with $tol = 1.0e - 12$, $m = 10$, $n = 1$, $a_0 = a_1 = 1, a_2 = 1 + 1.0e - 4$ in (a) and $a_0 = a_1 = 1, a_2 = 1 + 1.0e - 6$ in (b)

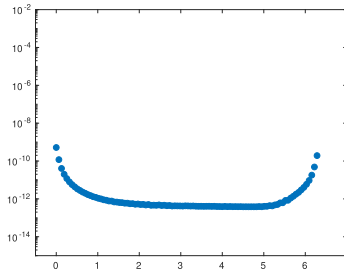
Differently, when N is odd, the proposed scheme performs quite satisfactorily. In Fig. 3 we show the measured error for Test 2 with $N = 1023$, $a_0 = a_1 = a_2 = 1$, $tol = 1.0e - 12$ and $m = 10$. Also, the acceleration of the convergence due to the increasing value of n is dramatic. For $n = 0$ the stopping criterion (2.23) is not fulfilled within 50000 iterations. Differently, for $n = 4$ we find $\hat{\ell} = 22$.

It is worth mentioning that (3.1) can be rewritten as $v(t) = \expm(At)v(0)$ where $\expm(A)$ is the matrix exponential of A . This relation provides an easy way based on the power method to give approximations of the solution vectors $v(t_i)$, $t_i \in \mathcal{I}_L$. Assume that $B = \expm(At_1)$ and $v(0)$ are available. Then we can determine the remaining vectors $v(t_i)$ by setting $v(t_{i+1}) = Bv(t_i)$, $i \geq 0$. In Fig. 4 we compare the accuracy of this power-based method with our algorithm applied for solving Test 2 with $a_0 = a_1 = a_2 = 1$, $tol = 1.0e - 12$, $n = 2$ and $m = 30$. The matrix B is computed by means of the MATLAB function `expm`.

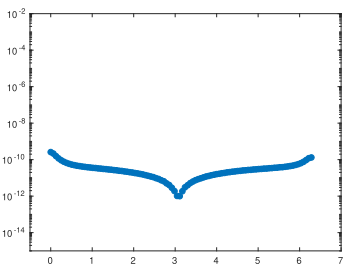
Matrices $A \in \mathbb{R}^{N \times N}$ of the form (3.2) are useful to check the dependence of the convergence rate and the accuracy w.r.t. the distribution of the eigenvalues and the norm of A . This set has already been considered for testing purposes in [2, 3]. For $\sigma = 1$ and $\gamma = 0$ all the proposed schemes with $n \in \{0, 1, 2, 3, 4\}$ perform well in terms of accuracy confirming the results in Boito et al. [2]. For $N = 1024$ and $tol = 1.0e - 12$ the scheme with $n = 0$ is unpractical since it requires more than 50000 terms to satisfy the stopping criterion. For $n = 1, 4$ we find $\hat{\ell} = 1714$ and



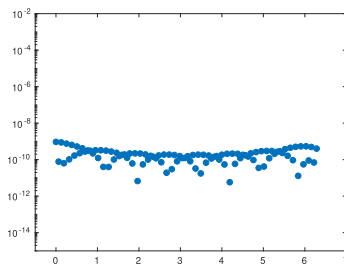
(a) Plot of computed error for $n = 1$.



(b) Plot of computed error for $n = 2$.



(c) Plot of computed error for $n = 3$.



(d) Plot of computed error for $n = 4$.

Fig. 3 Illustration of the measured error (2.24) for Test 2 with $a_0 = a_1 = a_2 = 1$, $tol = 1.0e - 12$ and $m = 10$ for different values of $n \in \{1, 2, 3, 4\}$

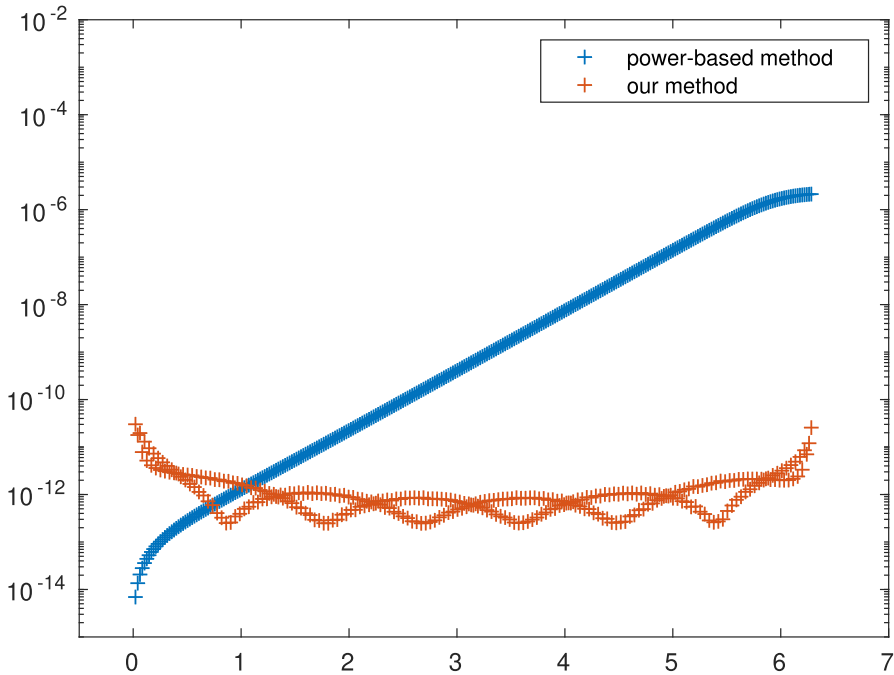


Fig. 4 Error plots generated by the power-based method and our algorithm applied or solving Test 2 with $a_0 = a_1 = a_2 = 1, tol = 1.0e - 12, n = 2$ and $m = 30$

$\hat{l} = 21$, respectively, showing again the efficiency of the convergence acceleration introduced by the Bernoulli polynomials.

Increasing the value of σ or γ introduces some numerical difficulties. For $\sigma = 1$ and $\gamma = 100$ the shifted linear systems are still well-conditioned but both the magnitude of eigenvalues and the norm of A increase. As a side effect we observe a deterioration of the convergence and the accuracy of our algorithm. For $n = 1, 2$ we find $\hat{l} = 8896$ and $\hat{l} = 457$, respectively. In Fig. 5 we show the plots of the computed error for shifted 1D Laplacian matrices with $N = 1024, tol = 1.0e - 12$, and $n = 1, 2$. The schemes with $n = 3, 4$ return inaccurate results. Similar behaviors for the error are found for $\sigma = 100$ and $\gamma = 0$.

In conclusion, numerical experiments with the finite dimensional case reveal that the computation of the approximant $v_{n,\ell}(t)$ given in (2.22) can be prone to numerical inaccuracies essentially due to:

1. a possibly large norm of the matrix A which makes unreliable the use of acceleration schemes based on Bernoulli polynomials;
2. the effects of a potential ill-conditioning of the shifted linear systems involved in the computation of the trigonometric expansion.

Moreover, the occurrence of large eigenvalues in magnitude determines two additional effects:

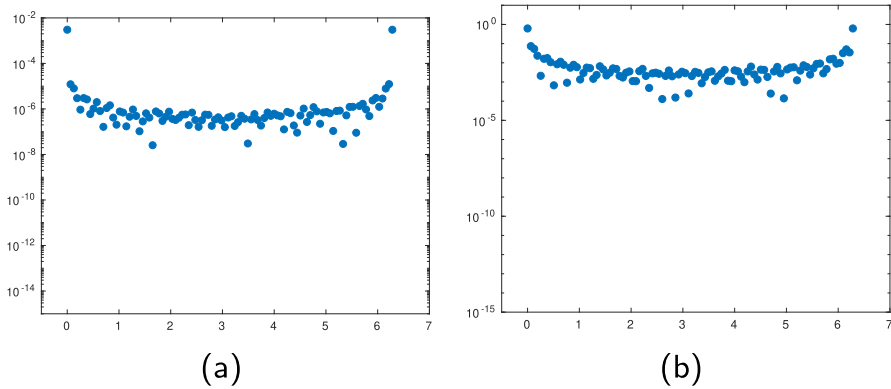


Fig. 5 Illustration of the measured error (2.24) for Test 3 with $\sigma = 1, \gamma = 100, tol = 1.0e - 12, m = 10, n = 1$ in (a) and $n = 2$ in (b)

1. the convergence generally deteriorates as more and more terms are needed into the expansion to evaluate the function at the eigenvalues;
2. since the convergence is slow, the stopping criterion can be satisfied even if the approximation is far from the "exact" value.

The analysis of these weak points motivates the extension of the results in [2, 3] to the functional setting given in Theorem 1. In particular, in the next section we will show that the extension makes it possible to apply our approach for solving certain differential problems with nonlocal boundary conditions.

4 The model one-dimensional problem for parabolic equation

Consider the differential problem

$$\begin{cases} u_t = \sigma u_{xx} + c(x)u, & 0 < x < 1, 0 < t < T, \\ u(0, t) = u(1, t) = 0, & 0 < t < T. \end{cases} \tag{4.1}$$

with the nonlocal condition

$$\frac{1}{T} \int_0^T u(x, t) dt = f(x), \quad f \in D(A). \tag{4.2}$$

In principle, several numerical schemes for computing the solution of (4.1), (4.2) can be devised based on Theorem 1. Specifically, we treat the problem (4.1) as a problem for an ordinary differential equation in the Banach space $X = L^2(0, 1)$ of the form

$$\frac{dv}{dt} = Av, \quad 0 < t < T, \tag{4.3}$$

Here $A = \sigma \frac{d^2}{dx^2} + c(x)$, where the domain $D(A)$ consists of the functions $w(x)$ with first derivative w' absolutely continuous on $[0, 1]$ such that $w'' \in L^2(0, 1)$, and satisfying the boundary conditions $w(0) = w(1) = 0$. It is worth noting that expansion (2.4) requires $f \in D(A^2)$, where $D(A^2)$ is the set of functions $w(x)$ on $[0, 1]$ such that $w'''(x)$ is absolutely continuous, $w^{(4)}(x) \in L^2(0, 1)$, and $w(0) = w(1) = w''(0) = w''(1) = 0$.

Notice that if we take $c(x) = 0$ for simplicity, then for any $f \in D(A^k)$ the values $A^k f$ are in fact the even-order derivatives of the given function f , namely $A^k f = f^{(2k)}$. Similarly, if $c(x)$ is constant, then $A^k f$ can be expressed as linear combinations of the even-order derivatives of the given function $f(x)$. So we can assume that the elements $g_j = A^j f$, $j = 0, \dots, k$ are given as well. Using Theorem 1, with the assumption $f \in D(A^{2n+3})$, we have that the solution $v(t)$ satisfies the formula

$$v(t) = v_{n,\ell}(t) + r_{n,\ell}(t) = p_n(t) + s_{n,l}(t) + r_{n,\ell}(t) \tag{4.4}$$

with

$$p_n(t) = \sum_{k=0}^{2n+1} \frac{(2\pi)^k}{k!} B_k \left(\frac{t}{2\pi} \right) g_k, \tag{4.5}$$

and

$$s_{n,l}(t) = (-1)^n 2 \sum_{k=1}^l \frac{1}{k^{2n}} \left(\Sigma_k A^2 g_{2n-1} \cos kt + \Upsilon_k A^2 g_{2n-1} \sin kt \right). \tag{4.6}$$

We transform the formula (4.6) to an equivalent form. We define the operators

$$V_k = (A^2 + k^2 I)^{-1}, \quad k = 1, 2, \dots \tag{4.7}$$

Using (2.5) we have that (4.6) takes the form

$$s_{n,l}(t) = (-1)^n 2 \sum_{k=1}^l \frac{1}{k^{2n}} \left(V_k g_{2n+2} \cos kt + \frac{1}{k} V_k g_{2n+3} \sin kt \right). \tag{4.8}$$

So the core of the algorithm is the evaluation of the quantities

$$V_k g = (A^2 + k^2 I)^{-1} g, \quad k = 1, 2, \dots, \quad g \in X.$$

The computation of these values may be done as follows.

1. Compute the solution $p^{(k)}$ of the equation

$$(A - (ki)I)p^{(k)} = g. \tag{4.9}$$

2. Compute the solution $w^{(k)}$ of the equation

$$(A + (ki)I)w^{(k)} = p^{(k)}. \tag{4.10}$$

For any $g \in X$ we have obviously $p^{(k)} \in D(A)$, $w^{(k)} \in D(A^2)$.

In our concrete case the evaluation of (4.9) and (4.10) is equivalent to the solution of two boundary value problems

$$\sigma \frac{d^2 p^{(k)}(x)}{dx^2} - ikp^{(k)}(x) + c(x)p^{(k)}(x) = g(x), \quad p^{(k)}(0) = p^{(k)}(1) = 0, \quad k = 1, 2, \dots, \tag{4.11}$$

$$\sigma \frac{d^2 w^{(k)}(x)}{dx^2} + ikw^{(k)}(x) + c(x)w^{(k)}(x) = p^{(k)}(x), \quad w^{(k)}(0) = w^{(k)}(1) = 0, \quad k = 1, 2, \dots \tag{4.12}$$

The procedure stated above for computing the approximation $v_{n,\ell}(t)$ of $v(t)$ allows us to explore different discretization methods. In particular, we have considered two schemes. The first scheme is purely numerical, whereas the second one uses a combination of symbolic and numerical computations.

- (I) The first, simplest approach is to discretize the operator A using finite differences methods. By using finite differences in space with equispaced points $x_i = ih = i/(N + 1)$, $0 \leq i \leq N + 1$, $h = 1/(N + 1)$, for the discretization of the second derivative we obtain the discrete analogue \hat{A} of the operator A ,

$$\hat{A} = \left(\text{diag}[c(x_1), \dots, c(x_N)] + \frac{\sigma}{h^2} \begin{bmatrix} -2 & 1 & & & \\ & 1 & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 1 & -2 \\ & & & & 1 & -2 \end{bmatrix} \right) \in \mathbb{R}^{N \times N}.$$

Replacing A with \hat{A} in the procedure for computing $v_{n,\ell}(t)$ is formally equivalent to applying our method for the solution of the first order system

$$\frac{du}{dt} = \hat{A}u(t), \tag{4.13}$$

with conditions

$$\frac{1}{T} \int_0^T u(t)dt = f, \tag{4.14}$$

where $f = [f(x_1), \dots, f(x_N)]^T$. The computational problem is of the form given in (2.1),(2.2) with $\hat{A} \in \mathbb{R}^{N \times N}$. This finite dimensional first order system (4.13), (4.14) can also be obtained by exploiting the classical method of lines [16] for solving (4.1), (4.2).

- (II) The second approach combines symbolical and numerical methods. The evaluation of the derivatives of the function f is performed symbolically. The computation of (4.5) is also carried out symbolically. The solution of (4.11) and (4.12) can be found by means of existing software for the solution of BVP's. In our implementation we used the MATLAB function `bvp4c` derived from

Kierzenka and Shampine [12] for solving these two-point boundary value problems. The code is a finite difference algorithm that implements the three-stage Lobatto IIIa formula. This is a collocation formula and the collocation polynomial provides a C^1 -continuous solution that is fourth order accurate uniformly in $[a, b]$. Mesh selection and error control are adaptively based on the residual of the continuous solution.

4.1 Numerical tests for the parabolic equation

Synthetic computational problems have been designed to test the two approaches outlined above. Our test suite consists of the following two problems:

1. The first model example is

$$\begin{cases} u_t = \sigma u_{xx}, & 0 < x < 1, 0 < t < 2\pi, \\ u(0, t) = u(1, t) = 0, & 0 < t < 2\pi. \end{cases} \tag{4.15}$$

$$\frac{1}{2\pi} \int_0^{2\pi} u(x, t) dt = f(x) \tag{4.16}$$

with the data

$$f(x) = 12 \frac{1 - e^{-2\pi\sigma(3\pi)^2}}{2\pi\sigma(3\pi)^2} \sin(3\pi x) - 7 \frac{1 - e^{-2\pi\sigma(2\pi)^2}}{2\pi\sigma(2\pi)^2} \sin(2\pi x).$$

Here the exact solution is

$$u(x, t) = 12e^{-\sigma(3\pi)^2 t} \sin(3\pi x) - 7e^{-\sigma(2\pi)^2 t} \sin(2\pi x).$$

2. The second test is concerned with the differential problem of the form

$$\begin{cases} u_t = u_{xx} + (4\pi^2 - 1)u(x), & 0 < x < 1, 0 < t < 2\pi, \\ u(0, t) = u(1, t) = 0, & 0 < t < 2\pi. \end{cases} \tag{4.17}$$

$$\frac{1}{2\pi} \int_0^{2\pi} u(x, t) dt = f(x) \tag{4.18}$$

with the data

$$f(x) = \frac{1 - e^{-2\pi}}{2\pi} \sin(2\pi x).$$

Here the exact solution is

$$u(x, t) = e^{-t} \sin(2\pi x).$$

In this section we present the results of numerical tests where approaches (I) and (II) are applied to model problems 1 and 2 above.

Table 1 Computed error and the minimum and the maximum values of the length ℓ_i of the rational approximation over $t_i \in \mathcal{I}_S$ as function of the degree n for the first model problem with $\sigma = 1.0e-6$, $N = 1000$ and $m = 100$

| n | 0 | 1 | 2 |
|--------------|----------|----------|----------|
| ℓ_{min} | 7798 | 177 | 34 |
| ℓ_{max} | 32716 | 223 | 46 |
| err | $1.8e-9$ | $1.8e-9$ | $1.8e-9$ |

Tests for Approach (I). Let us start by testing the accuracy of approach (I), which is based on semidiscretization in space. Here the error of the computed approximation $v_{n,\hat{\ell}}(t)$ is measured with respect to the exact values of the solution function $u(x, t)$ by

$$err = \max_{t_i \in \mathcal{I}_L} \|\hat{u}(t_i) - v_{n,\hat{\ell}}(t_i)\|_\infty / \|\hat{u}(t_i)\|_\infty,$$

where $\hat{u}(t) = [\hat{u}_1(t), \dots, \hat{u}_N(t)]^T$, $\hat{u}_i(t) = u(x_i, t)$, $1 \leq i \leq N$. The coarse and fine grid of points are defined by $\mathcal{I}_S = \{t_i = 2\pi(i - 1)/(m - 1) : 1 \leq i \leq m\}$ and $\mathcal{I}_L = \{t_i = 2\pi(i - 1)/(4(m - 1)) : 1 \leq i \leq 4(m - 1) + 1\}$, respectively.

In Table 1 we show the results obtained for the first model problem with $\sigma = 1.0e-6$, $N = 1000$ and $m = 100$. The results are generated by **Algorithm 1** with $tol = 1.0e-7$ for different values of n . We report the value of err and the minimum and the maximum values of the length ℓ_i of the rational approximation over $t_i \in \mathcal{I}_S$ as functions of the degree n .

The discretization error is of order $\sigma h^2 \max |u_{xxxx}|/12 \simeq 1.0e-8$. The shifted linear systems (2.5) are well conditioned. Increasing the value of tol does not improve the computed error, which is in accordance with a priori error bounds.

In Table 2 we repeat the same test with $\sigma = 0.01$. The discretization error is now of order $1.0e-4$. The error estimates change significantly since the scaling factor $\sigma(N + 1)^2$ highlights the pathologies observed in the previous section (compare with the results of Test 3). In particular, the use of fast expansions with increasing n becomes unreliable. For $n = 2$ **Algorithm 1** stops but the returned error is larger than 1.

Tests for Approach (II). Approach (II), on the other hand, is based on computing the expansion of $v(t)$ as given in Theorem 1 for the infinite dimensional operator $A = \sigma \frac{d^2}{dx^2} + c(x)$. Recall that in our implementation we used the MATLAB function

Table 2 Computed error and the minimum and the maximum values of the length ℓ_i of the rational approximation over $t_i \in \mathcal{I}_S$ as function of the degree n for the first model problem with $\sigma = 1.0e-2$, $N = 1000$ and $m = 100$. For $n = 2$ the computed error is larger than 1

| n | 0 | 1 | 2 |
|--------------|----------|----------|----|
| ℓ_{min} | 9093 | 112 | 21 |
| ℓ_{max} | 25977 | 284 | 33 |
| err | $2.2e-4$ | $6.4e-6$ | * |

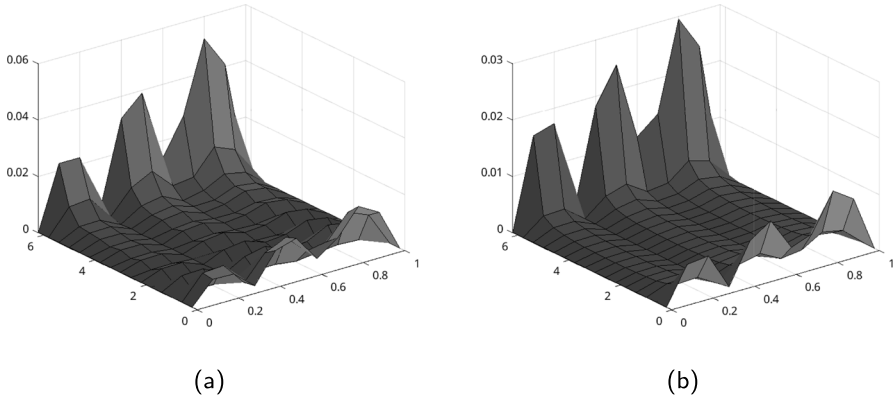


Fig. 6 Plot of $err(x, t)$ generated by solving the first model problem with $\sigma = 0.01$ for $n = 0$ with $\ell = 200$ in (a) and $\ell = 400$ in (b)

`bvp4c` for solving the two-point boundary value problems (4.11), (4.12). The number of boundary value problems to be solved depends on the value of ℓ in (2.22), which is fixed a priori. The output is an interpolating function $v_{n,\ell}(x, t)$. For the sake of graphical illustration the computed error is shown by plotting the function $err(x, t) = |v_{n,\ell}(x, t) - u(x, t)|$ over the domain $[0, 1] \times [0, 2\pi]$.

In Figs. 6, 7 and 8 we show the plots generated by solving the first model problem with $\sigma = 0.01$ for different values of n and ℓ .

In Fig. 9 we describe our results for the second model problem with $n = 1$ and $\ell = 20$ and $\ell = 40$.

In summary, our preliminary experience with the infinite dimensional (functional) approach indicates that, when applicable, expansions with increasing n allow the computation of quite accurate approximations of $u(x, t)$ already for small or moderate values of the truncation level ℓ .

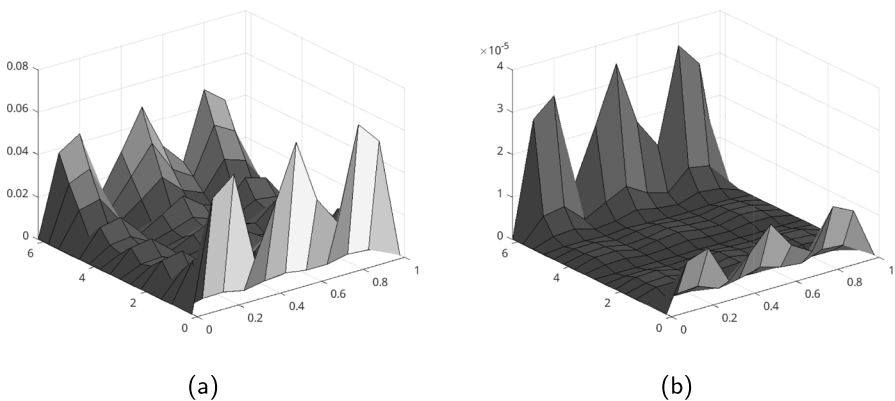


Fig. 7 Plot of $err(x, t)$ generated by solving the first model problem with $\sigma = 0.01$ for $n = 1$ with $\ell = 100$ in (a) and $\ell = 1000$ in (b)

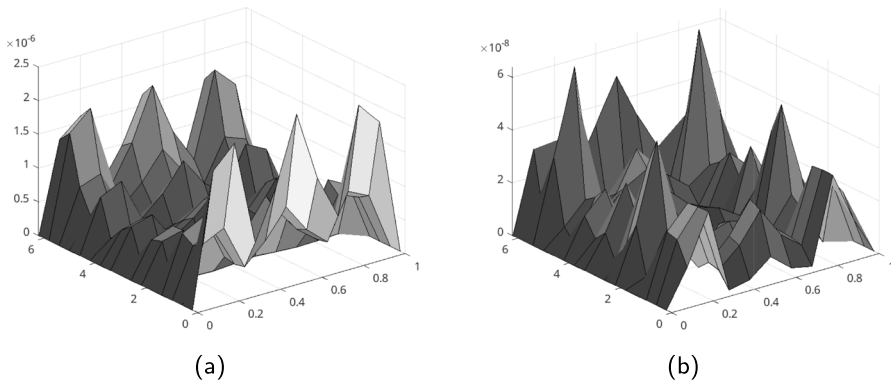


Fig. 8 Plot of $err(x, t)$ generated by solving the first model problem with $\sigma = 0.01$ for $n = 2$ with $\ell = 10$ in (a) and $\ell = 20$ in (b)

5 Conclusion and future work

In this paper we have devised a mixed polynomial/rational expansion for the solution of an abstract first-order differential problem with nonlocal conditions. We have proposed an algorithm that exploits this expansion to approximate numerically the solution and tested it on several examples, both in finite- and infinite-dimensional settings. Of particular interest are implementations where the expansion is directly applied to an infinite-dimensional operator and discretization only intervenes in the solution of the resulting family of shifted boundary value problems.

There are still many points for future developments. Specifically:

1. The derivation of precise residual estimates for this expansion also related with the design of efficient stopping criteria.
2. The study of alternative acceleration schemes not involving high-degree Bernoulli polynomials in order to relax the regularity assumptions.

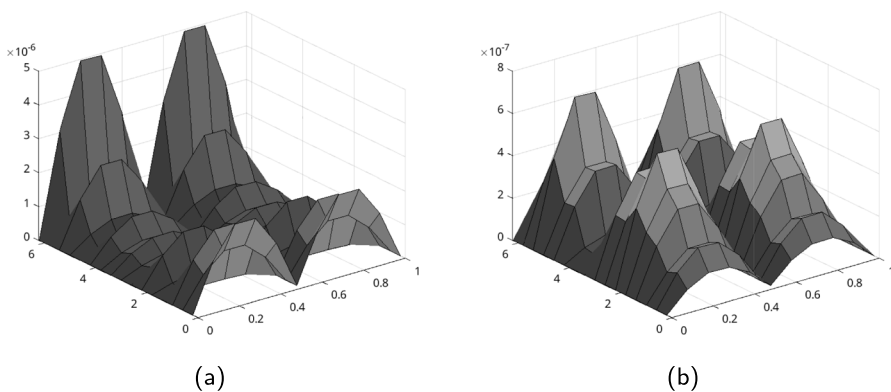


Fig. 9 Plot of $err(x, t)$ generated by solving the second model problem for $n = 1$ with $\ell = 20$ in (a) and $\ell = 40$ in (b)

3. The design of efficient numerical methods for solving the sequence of shifted boundary value problems (4.11), (4.12).
4. The application of our proposed method to more general differential problems such as the two-dimensional parabolic equation and the biharmonic equation.

We also point out that, for suitable classes of problems, one may apply the mixed polynomial/rational expansion presented here to devise hybrid approaches that combine symbolic/functional and purely numerical techniques. As an example, consider Problem 1 in Section 3, i.e., (4.15), (4.16). Starting from the finite-dimensional approach based on semi-discretization in space, we may incorporate more and more symbolic elements to circumvent potential sources of instability and improve the quality of the results. For instance, since the function $f(x)$ for (4.15), (4.16) is known analytically and the action of matrix A corresponds to a second derivative, one may pre-compute symbolically $f''(x)$ and $f^{(4)}(x)$, evaluate them on the points of the space grid, and use the resulting vectors in place of Af and A^2f in the polynomial/rational expansion. Moreover, we may also try evaluating the polynomial part of the expansion symbolically, if the polynomial degree is small (but high enough to be numerically troublesome). Preliminary experiments show that, with these modifications, the test presented in Table 2 for $n = 2$ yields an error that is no longer larger than 1, but comparable to case $n = 1$, with a smaller number of rational terms.

Acknowledgements Not applicable.

Author Contributions The authors contributed equally to the present work.

Funding Open access funding provided by Università di Pisa within the CRUI-CARE Agreement. P. Boito acknowledges the MUR Excellence Department Project awarded to the Department of Mathematics, University of Pisa, CUP I57G22000700001.

L. Gemignani and P. Boito are partially supported by European Union - NextGenerationEU under the National Recovery and Resilience Plan (PNRR) - Mission 4 Education and research - Component 2 From research to business - Investment 1.1 Notice Prin 2022 - DD N. 104 2/2/2022, titled Low-rank Structures and Numerical Methods in Matrix and Tensor Computations and their Application, proposal code 20227PCCKZ - CUP I53D23002280006. L. Gemignani is partially supported by the Spoke 1 "FutureHPC & BigData" of the Italian Research Center on High-Performance Computing, Big Data and Quantum Computing (ICSC) funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di "campioni nazionali di R&S (M4C2-19)" - Next Generation EU (NGEU).

Availability of supporting data Not applicable.

Declarations

Ethical approval Not applicable.

Competing interests The authors have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bellman, R.: A note on the identification of linear systems. *Proceedings of the American Mathematical Society* **17**(1), 68–71 (1966)
2. Boito, P., Eidelman, Y., Gemignani, L.: Efficient solution of parameter-dependent quasiseparable systems and computation of meromorphic matrix functions. *Numer. Linear Algebra Appl.* **25**(6), e2141 (2018)
3. Boito, P., Eidelman, Y., Gemignani, L.: Computing the reciprocal of a ϕ -function by rational approximation. *Adv. Comput. Math.* **48**(1), 1–28 (2022)
4. Botchev, M.A., Knizhnerman, L., Tyrtshnikov, E.E.: Residual and restarting in Krylov subspace evaluation of the φ function. *SIAM J. Sci. Comput.* **43**(6), A3733–A3759 (2021)
5. Cannon, J.R.: The solution of the heat equation subject to the specification of energy. *Q. Appl. Math.* **21**(2), 155–160 (1963)
6. Eidelman, Y.S., Sherstyukov, V.B., Tikhonov, I.V.: Application of Bernoulli polynomials in non-classical problems of mathematical physics. *Systems of Computer Mathematics and their Applications*, pages 223–226 (2017)
7. Engel, K.-J., Nagel, R., Brendle, S.: *One-parameter semigroups for linear evolution equations*, vol. 194. Springer, Berlin (2000)
8. Gemignani, L.: Efficient inversion of matrix ϕ -functions of low order. *Appl. Numer. Math.* **192**, 57–69 (2023)
9. Hochbruck, M., Ostermann, A.: Exponential integrators. *Acta Numer* **19**, 209–286 (2010)
10. Jimenez, J.C., de la Cruz, H., De Maio, P.A.: Efficient computation of phifunctions in exponential integrators. *J. Comput. Appl. Math.* **374**, 112758 (2020)
11. Kalenyuk, P.I., Kohut, I.V., Nytrebych, Z.M.: Problem with integral condition for a partial differential equation of the first order with respect to time. *J. Math. Sci.* **181**(3), 293–304 (2012)
12. Kierzenka, J., Shampine, L.F.: A BVP solver based on residual control and the MATLAB PSE. *ACM Trans. Math. Software* **27**(3), 299–316 (2001)
13. Opial, Z.: Linear problems for systems of nonlinear differential equations. *J. Differential Equations* **3**(4), 580–594 (1967)
14. Pazy, A.: *Semigroups of linear operators and applications to partial differential equations*, vol. 44. Springer Science & Business Media, Berlin (2012)
15. Popov, A.Y., Tikhonov, I.V.: Exponential solubility classes in a problem for the heat equation with a non-local condition for the time averages. *Sbornik: Mathematics* **196**(9), 1319–1348 (2005)
16. Schiesser, W.E.: *The numerical method of lines*. San Diego, Academic Press, Inc. CA. Integration of partial differential equations (1991)
17. Tikhonov, I.V.: Uniqueness theorems for linear non-local problems for abstract differential equations. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* **67**(2), 133–166 (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.