Rational Approximation for the Inverse of a ϕ-Function of Quasiseparable Matrices

P. Boito[1] Y. Eidelman[2] L. Gemignani[3]

[1]Dipartimento di Matematica
Università di Pisa

[2]School of Mathematical Sciences
Raymond and Beverly Sackler School of Exact Sciences
Tel-Aviv University

[3]Dipartimento di Informatica
Università di Pisa

NASCA 2018
Kalamata, July 2-6, 2018

Overview

- Matrix ϕ-functions

$$\phi_k(A), \quad k \geq 0,$$

  play an important role in the solution of certain differential equations (see e.g., [Higham 2008], [Hochbruck and Ostermann 2010]).
- Recursive definition:

$$\phi_{k+1}(z) := \frac{\phi_k(z) - 1/k!}{z}, \qquad \phi_0(z) = e^z.$$

- We focus on rational approximations of

$$\psi_1(z) := \phi_1(z)^{-1} = \frac{z}{e^z - 1}.$$

- Such approximations should allow for efficient computation of $\psi_1(A)\boldsymbol{b}$, where A is structured (e.g., quasiseparable).

A model problem

Consider the differential problem

$$\frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}(t) + \boldsymbol{p}, \quad 0 \leq t \leq \tau,$$

with $A \in \mathbb{R}^{d\times d}$ a given matrix, and $\boldsymbol{p} \in \mathbb{R}^d$ unknown.

We want to compute the solution $\boldsymbol{u}\colon [0,\tau] \to \mathbb{R}^d$ and the vector $\boldsymbol{p}$, with the conditions

$$\boldsymbol{u}(0) = \boldsymbol{u}_0 = \boldsymbol{g}, \quad \boldsymbol{u}(\tau) = \boldsymbol{h}.$$

A model problem

Assume that the points

$$2\pi ik/\tau, \quad k = \pm 1, \pm 2, \ldots$$

do not belong to the spectrum of A. Then we have

$$\boldsymbol{p} = \frac{1}{\tau}\psi_1(\tau A)\,(\boldsymbol{h} - \boldsymbol{g}) - A\boldsymbol{g}.$$

Note that $\boldsymbol{u}(t)$ can be computed as

$$\boldsymbol{u}(t) = w_t(A)(\boldsymbol{h} - \boldsymbol{g}) + \boldsymbol{g}, \quad 0 \le t \le \tau \tag{1}$$

with $w_t(z) = \frac{e^{zt}-1}{e^{\tau z}-1}, \quad 0 \le t \le \tau,\ z \in \mathbb{C}.$
See also

- [Eidelman, Tikhonov, Sherstyukov, *Application of Bernoulli polynomials in non-classical problems of mathematical physics*, in Systems of Computer Mathematics and their applications, 2017],
- [Tikhonov, Eidelman, *An inverse problem for a differential equation in a Banach space and the distribution of zeros of an entire function of Mittag-Leffler type*, Differ. Uravn. 2002]

Taylor expansion of $\psi(z)$

A classical approach to evaluation of $\psi_1(A)$ and $\psi_1(A)\boldsymbol{b}$ is based on truncated Taylor expansion:

$$\psi_1(z) = \sum_{k=0}^{+\infty} \frac{B_k}{k!} z^k, \qquad |z| < 2\pi.$$

Such approximations are quite accurate if $\|A\|$ is sufficiently small.

Here B_k denotes the kth Bernoulli number.
We will also denote as $B_k(t)$, $k \geq 0$, the Bernoulli polynomials.

Mixed polynomial-rational approximation

We propose a family of polynomial-rational approximations to $\psi_1(A)$ and related functions:

$$\psi_1(A) \simeq p_s(A) + \sum_{k=1}^{m} \gamma_{k,s} A^{\tau_s} (A^2 + k^2 I)^{-1},$$

for given integers $s > 1$ and $m \geq s$. Here $p_s(z)$ is a polynomial of degree $\ell = \ell(s)$ and $\tau_s = \tau(s) \in \mathbb{N}$.

Mixed polynomial-rational approximation

We propose a family of polynomial-rational approximations to $\psi_1(A)$ and related functions:

$$\psi_1(A) \simeq p_s(A) + \sum_{k=1}^{m} \gamma_{k,s} A^{\tau_s} (A^2 + k^2 I)^{-1},$$

for given integers $s > 1$ and $m \geq s$. Here $p_s(z)$ is a polynomial of degree $\ell = \ell(s)$ and $\tau_s = \tau(s) \in \mathbb{N}$.

- Better convergence rate and larger convergence domain w.r.t. polynomial approximations.
- Isn't matrix inversion expensive? In general yes, but not for banded or rank-structured matrices (Toeplitz-like, quasiseparable).
- In the quasiseparable case: efficient solution of shifted linear systems [B., Eidelman, Gemignani, NLAA 2018] for computing $\psi_1(A)\boldsymbol{b}$.

Approximation of $w_t(z)$

The solution $\boldsymbol{u}(t)$ of the model problem is given by

$$w_t(z) = \frac{e^{zt} - 1}{e^{2\pi z} - 1} = \frac{t}{2\pi} + y_t(z) - y_0(z),$$

where $y_t(z)$ is an auxiliary function. We prove

Lemma

Let $t \in [0, 2\pi]$ and $y_t(z)$ as above. Then we have

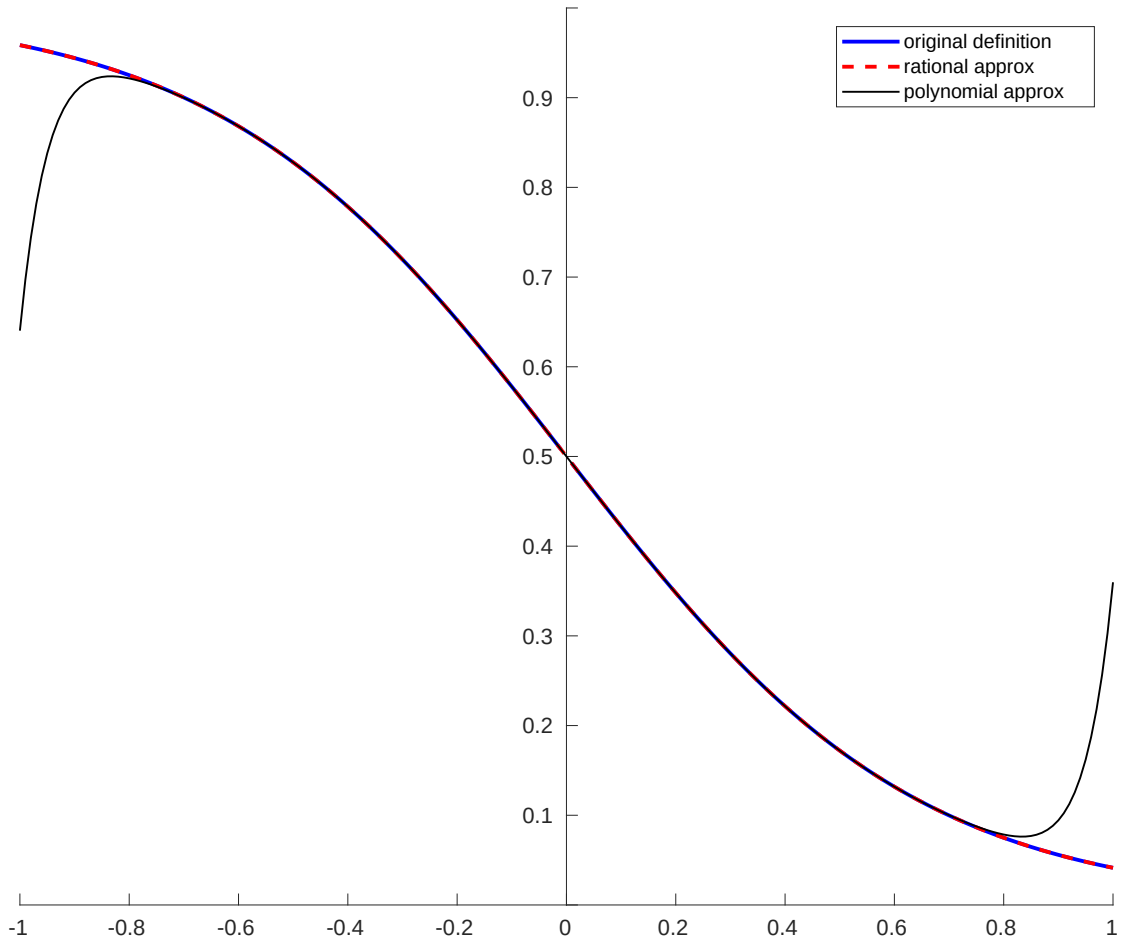
$$y_t(z) = p_{n,t}(z) + s_{n,t}(z), \quad n = 0, 1, 2, \ldots$$

$$\text{with} \qquad p_{n,t}(z) = \sum_{i=2}^{2n+1} \frac{(2\pi)^{i-1}}{i!} B_i\left(\frac{t}{2\pi}\right) z^{i-1}$$

$$\text{and} \qquad s_{n,t}(z) = \frac{(-1)^n}{\pi} \sum_{k=1}^{\infty} \frac{z^{2n}(z\cos(kt) + \frac{1}{k} z^2 \sin(kt)))}{k^{2n}(z^2 + k^2)}.$$

Approximation of $w_t(z)$

Here is a plot of $w_\pi(z)$ and its approximations (polynomial of degree 19, mixed rational of degree $(9, 10)$):



Approximation of $\psi_1(A)$

Let A be a matrix whose spectrum does not contain the poles of $\psi(z)$. Then:

$$\psi_1(A) = \phi_1(A)^{-1} = I - \frac{1}{2}A + A\,y_0\left(\frac{A}{2\pi}\right).$$

Theorem

For any fixed $n > 0$ it holds

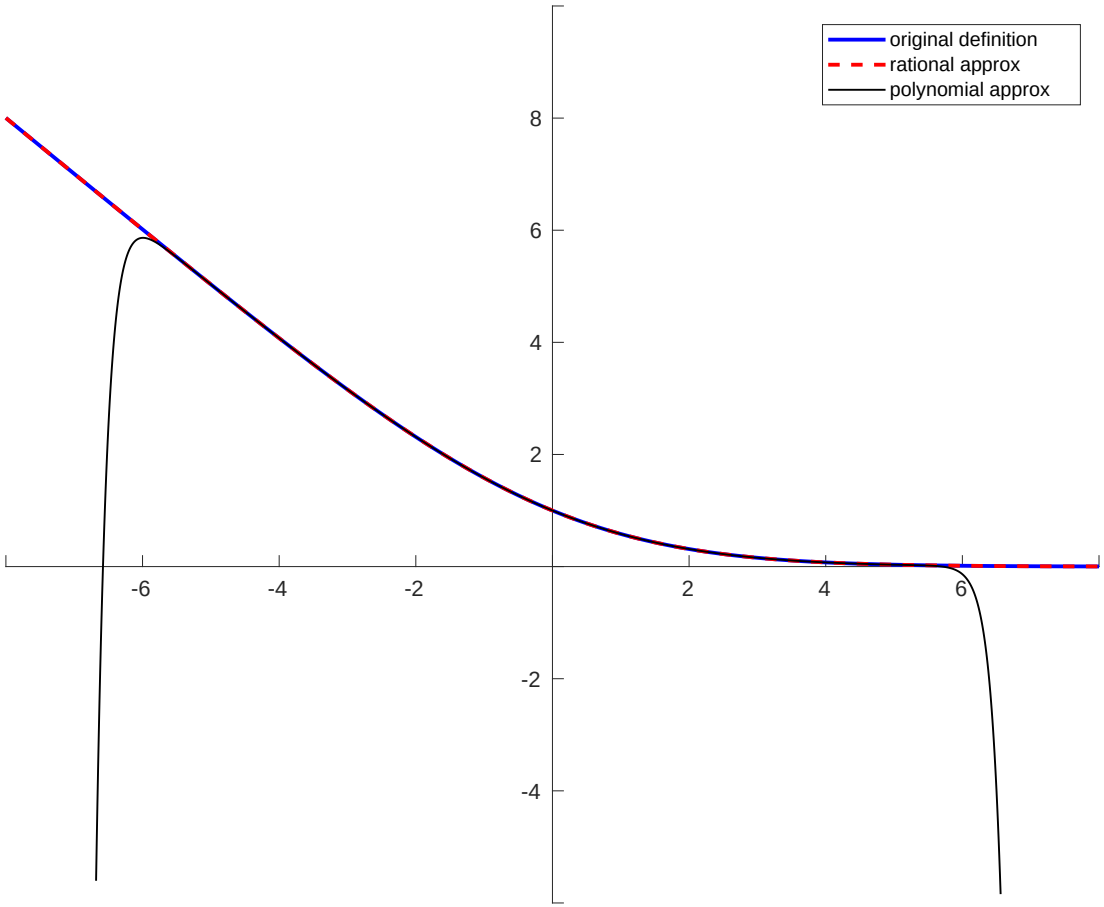
$$\psi_1(A) = p_n(A) + 2(-1)^n \sum_{k=1}^{\infty} \left(\frac{A}{2\pi}\right)^{2(n+1)} \frac{1}{k^{2n}} \left(\left(\frac{A}{2\pi}\right)^2 + k^2 I\right)^{-1},$$

where

$$p_n(A) = I - \frac{1}{2}A + \sum_{i=0}^{n-1} A^{2(i+1)} \frac{B_{2(i+1)}}{(2(i+1))!}.$$

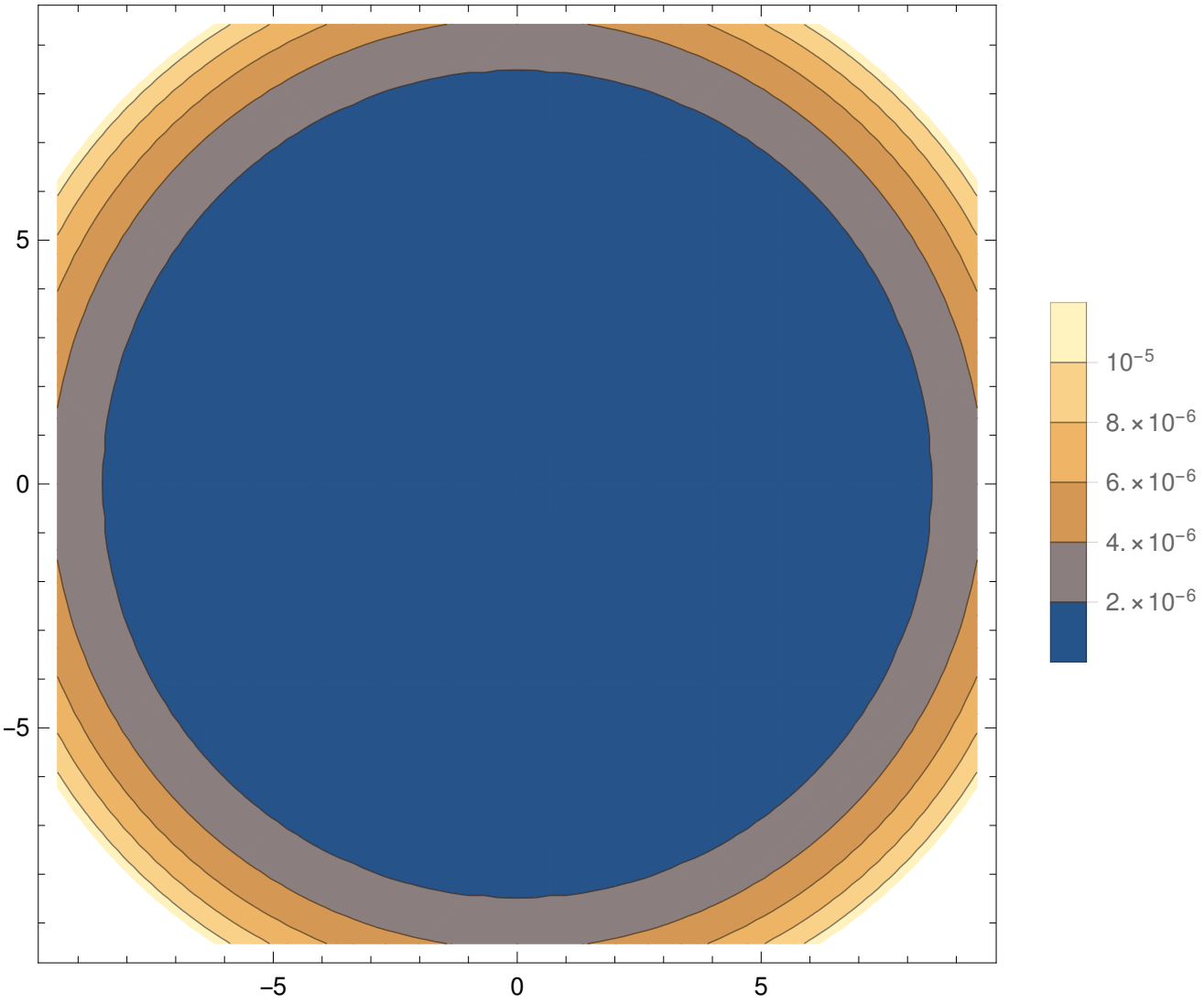
Approximation of $\psi_1(z)$ on $\mathbb{R}$

Here is a plot of $\psi_1(z)$ and its approximations (polynomial of degree 20, mixed rational of degree $(4, 16)$):



Approximation of $\psi_1(z)$ on $\mathbb{C}$

Absolute error for the mixed rational approximation of degree $(4, 16)$):



Approximation of $\psi_1(A)$ and $\psi_1(A)\boldsymbol{b}$

- We compare the accuracy of polynomial and rational approximations for computing both the matrix function $\psi_1(A)$ and the vector $\psi_1(A)\boldsymbol{b}$, where $A \in \mathbb{R}^{d\times d}$ is symmetric. and $\boldsymbol{b} \in \mathbb{R}^d$.
- We are interested in the case where A is structured, so that a linear system $A\boldsymbol{x} = \boldsymbol{f}$ can be solved in (soft) linear time with linear storage.
- For any given $n \geq 1$ the polynomial and rational approximations of $\psi_1(A)$ are $\psi_{n,0}(A)$ and $\psi_{m,n-m}(A)$, respectively. The corresponding normwise relative errors are

$$err\_p_n = \frac{\| \psi(A) - \psi_{n,0}(A) \|}{\| \psi_{n,0}(A) \|}, \quad err\_r_{n,m} = \frac{\| \psi(A) - \psi_{m,n-m}(A) \|}{\| \psi_{n,0}(A) \|}.$$

Approximation of $\psi_1(A)$ *and* $\psi_1(A)\boldsymbol{b}$

Example 1: A is the Toeplitz tridiagonal matrix generated as $A = \texttt{gallery}('\texttt{tridiag}', \texttt{d}, -1, 4, -1)$.

d	err_p_{50}	$err\_r_{50,3}$
256	3.33e−1	1.15e−12
512	3.33e−1	1.15e−12
1024	3.33e−1	1.15e−12
2048	3.33e−1	1.15e−12

Approximation of $\psi_1(A)$ *and* $\psi_1(A)\boldsymbol{b}$

Example 2: A is the quasiseparable matrix of order 1 generated as $A = 0.7 * \texttt{inv}(\texttt{gallery}('\texttt{tridiag}', \texttt{d}, \texttt{d}/2, [\texttt{d} : -1 : 1], \texttt{d}/2))$.

d	err_p_{50}	$err\_r_{50,3}$
256	2.25e−12	2.25e−12
512	6.42e−12	6.42e−12
1024	7.94e−3	2.78e−11
2048	2.80e77	2.00e−2

Approximation of $\psi_1(A)$ *and* $\psi_1(A)\boldsymbol{b}$

Example 3: A is the Kac-Murdock-Szegö Toeplitz matrix generated as $A = ((0.8)^{\mathtt{abs(i-j)}})$.

d	err_p_{50}	$err\_r_{50,3}$
256	6.68e−8	3.71e−14
512	7.40e−8	3.78e−14
1024	7.6e−8	3.8e−14
2048	7.65e−8	3.80e−14

Approximation of $\psi_1(A)$

Example 4: $A = \gamma Z$, where Z is the generator of the circulant matrix algebra of size $d \times d$.

γ	err_p_{50}	$err\_r_{50,3}$
2	1.24e−11	1.24e−11
4	1.33e−10	1.25e−11
8	6.90e4	1.26e−11
16	4.15e+18	7.00e−11
32	3.10e+32	4.20e−9
64	7.00e+46	9.00e−7

γ	$err\_r_{50,3}$	$err\_r_{100,3}$	$err\_r_{200,3}$	$err\_r_{400,3}$
64	9.00e−7	5.70e−9	5.30e−11	1.28e−11

Entry-wise bounds on $\psi_1(A)$

- Motivated by classical results on off-diagonal decay of functions of banded matrices (e.g., [Demko, Moss, Smith, Math. Comp. 84], [Benzi and Golub, BIT 1999]), we develop bounds on $|[\psi_1(A)]_{i,j}|$.
- Let $\varepsilon_{n,s}(z) = \psi_1(z) - p_n(z) - r_{n,s}(z)$ be the (n,s)-approximation error.
- In general we have

$$|[\psi_1(A)]_{i,j}| \leq |[p_n(A)]_{i,j}| + |[r_{n,s}(A)]_{i,j}| + |[\varepsilon_{n,s}(A)]_{i,j}|.$$

- Note that for A m-banded we have $|[p_n(A)]_{i,j}| = 0$ if $|i-j| > 2nm$.

Entry-wise bounds on $\psi_1(A)$

Theorem

Let $A \in \mathbb{R}^{d\times d}$ be a symmetric banded matrix with half-bandwidth m. For all $(n,s) \in \mathbb{N}\times\mathbb{N}$ it holds

$$|[\psi_1(A)]_{i,j}| \le |[r_{n,s}(A)]_{i,j}| + |[\varepsilon_{n,s}(A)]_{i,j}|, \quad |i-j| > 2mn,$$

where

$$|[r_{n,s}(A)]_{i,j}| \le \left\| \frac{A}{2\pi} \right\|_2^{2(n+1)} \sum_{\nu=j-2m(n+1)}^{j+2m(n+1)} \sum_{k=1}^{s} \frac{C_k}{k^{2n}} \lambda_k^{|i-\nu|}$$

and

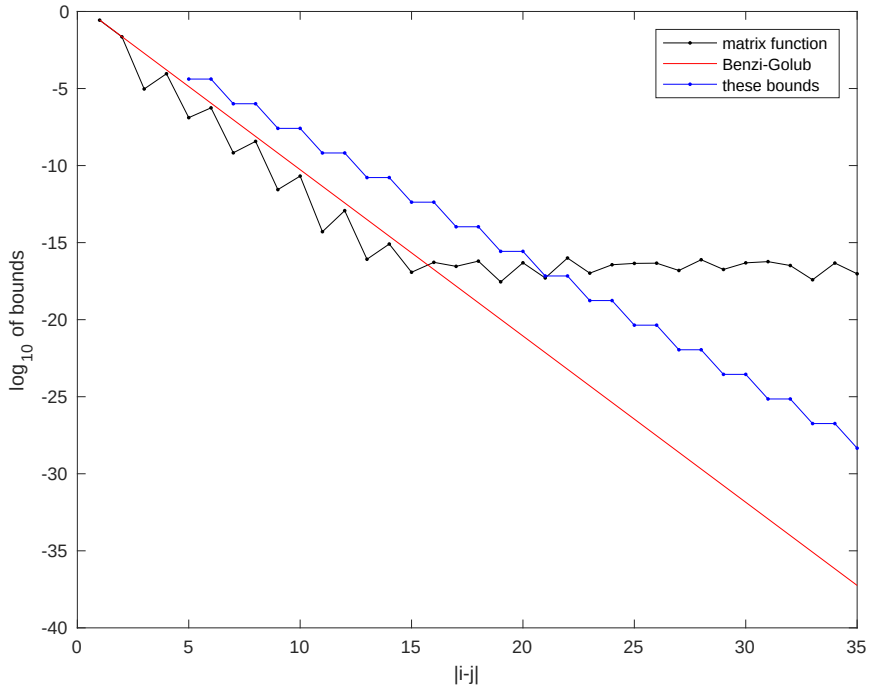
$$|[\varepsilon_{n,s}(A)]_{i,j}| \le \left\| \frac{A}{2\pi} \right\|_2^{2(n+1)} \left(\zeta(2n+2) - \sum_{k=1}^{s} \frac{1}{k^{2n+2}} \right),$$

with

$$a_k = k^2, \quad b_k = \left(\frac{\rho(A)}{2\pi}\right)^2 + k^2, \quad r_k = \frac{b_k}{a_k},$$

$$\lambda_k = \left(\frac{\sqrt{r_k}-1}{\sqrt{r_k}+1}\right)^{1/2m}, \quad C_k = \max\left\{ a_k^{-1}, \frac{(1+\sqrt{r_k})^2}{2a_k r_k} \right\}.$$

Entry-wise bounds on $\psi_1(A)$



Conclusions

- We have introduced a family of rational approximations of the inverse of the ϕ_1 function encountered in exponential integration methods.
- Such formulas are especially well-suited for computations involving structured matrices.
- Ideas for further developments:
  - rational Krylov methods for computation of ϕ-functions (see [Göckler and Grimm, SIMAX 2014]),
  - more numerical experiments and comparison with rational Carathéodory-Fejér approximation (see [Schmeltzer and Trefethen, ETNA 2007]).
- Reference: B., Eidelman, Gemignani, *Computing the Inverse of a $\phi-$Function by Rational Approximation*, arXiv: 1801.04573 [math.NA].