Functions of Matrices with Kronecker Sum Structure

Michele Benzi

Department of Mathematics and Computer Science
Emory University

Atlanta, Georgia, USA

2015 SIAM Conference on Applied Linear Algebra

Atlanta, GA, October 25-30, 2015

Motivation (cont.)

Let f be a function defined on the spectrum of $\mathcal{A}$, so that $f(\mathcal{A})$ is well defined.

Important examples include:

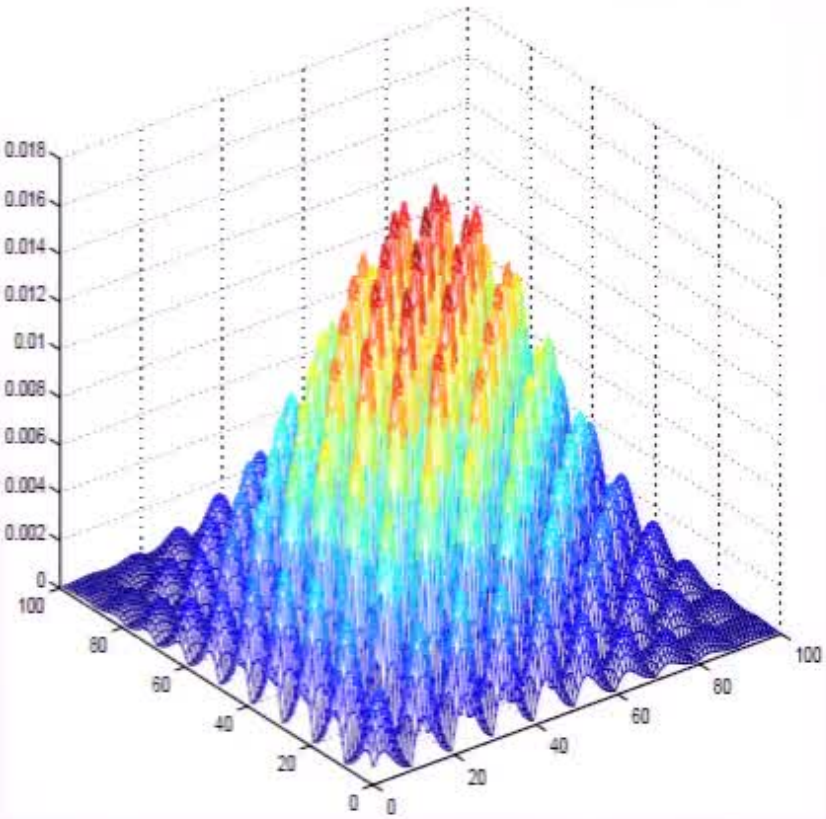
- The (negative) exponential: $f(z) = \mathrm{e}^{-tz}$, $t > 0$
- Negative fractional roots, e.g., $f(z) = z^{-\frac{1}{2}}$ (for $\mathcal{A}$ pos. def.)
- The ψ functions, e.g., $\psi_1(z) = \frac{\mathrm{e}^z - 1}{z}$
- Combinations like $f(z) = \frac{\sin(\sqrt{z}t)}{\sqrt{z}}$, $t > 0$

See N. J. Higham, *Functions of Matrices. Theory and Computation*, SIAM, 2008.

We are interested in

1. Studying the decay properties of $f(\mathcal{A})$
2. Efficiently computing the action $f(\mathcal{A})\, b$ for a given vector b

Motivation (cont.)



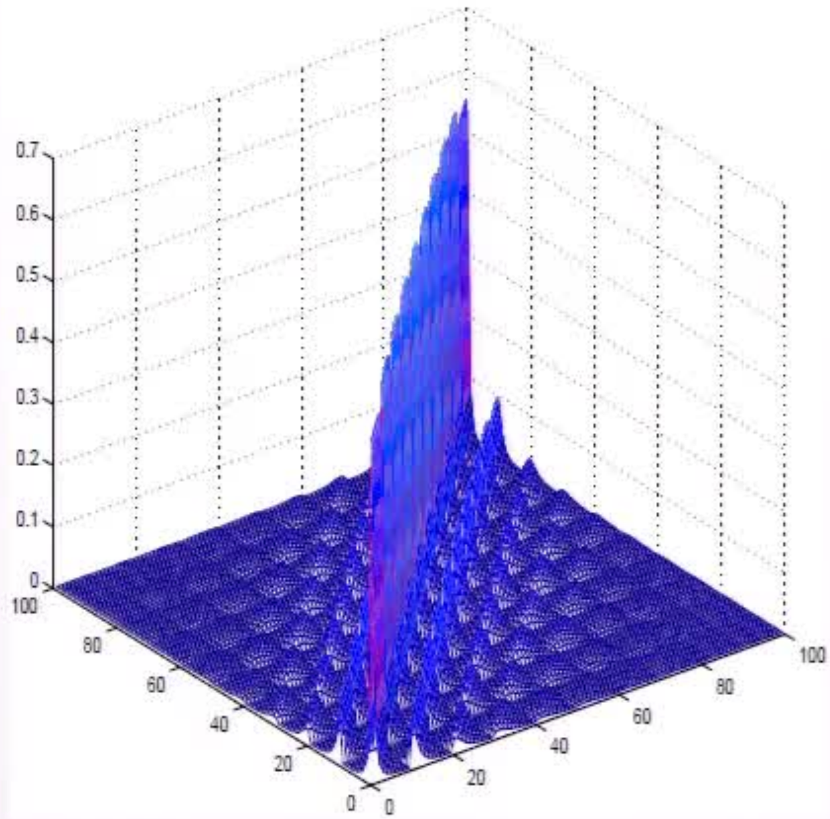


Figure: Three-dimensional decay plots for $[f(\mathcal{A})]_{ij}$ where $\mathcal{A}$ is the 5-point finite difference discretization of the negative Laplacian on the unit square on a 10×10 uniform grid with zero Dirichlet boundary conditions. Left: $f(\mathcal{A}) = \exp(-5\mathcal{A})$. Right: $f(\mathcal{A}) = \mathcal{A}^{-1/2}$.

Motivation (cont.)

The Kronecker sum of matrices has certain nice properties:

- $e^{M_1 \oplus M_2} = e^{M_1} \otimes e^{M_2}$
- If $\lambda \in \sigma(M_1)$ and $\mu \in \sigma(M_2)$, then $\lambda + \mu$ is an eigenvalue of $M_1 \oplus M_2$. Moreover, every eigenvalue of $M_1 \oplus M_2$ is of this form.
- If (λ, x) and (μ, y) are eigenpairs of M_1 and M_2, the eigenvector of $M_1 \oplus M_2$ corresponding to $\lambda + \mu$ is given by $\text{vec}(x \otimes y)$.
- $\sigma(M_1 \oplus M_2) = \sigma(M_2 \oplus M_1)$

See R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University press, 1991.

Motivation (cont.)

These properties suggest that one should be able to reduce analytical and computational problems for functions of Kronecker sums to analogous problems for the (much smaller) summands.

Some works in this direction:

C. Canuto, V. Simoncini, and M. Verani, *On the decay of the inverse of matrices that are sum of Kronecker products*, Linear Algebra Appl., 452 (2014), pp. 21–39.

I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromomskij, *$\mathcal{H}$-matrix approximation for the operator exponential with applications*, Numer. Math., 92 (2002), pp. 83–111.

I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromomskij, *Tensor-product approximation to the inverse and related operators in high-dimensional elliptic problems*, Computing, 74 (2005), pp. 131–157.

D. Kressner and C. Tobler, *Krylov subspace methods for linear systems with tensor product structure*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1688–1714.

Functions defined by integral transforms

Main idea: express the function $f(\mathcal{A})$, whenever possible, in terms of functions, like the exponential or the resolvent, for which we know how to exploit the Kronecker structure.

This leads naturally to two (related) classes of functions:

- Laplace–Stieltjes functions
- Cauchy–Stieltjes (or Markov-type) functions

The Laplace–Stieltjes functions are those functions that can be represented as Laplace transforms of nonnegative measures:

$$f(x) = \int_0^\infty e^{-\tau x} d\alpha(\tau)$$

where $\alpha(t)$ is nondecreasing and has at least one point of increase in $(0, \infty)$, and the integral converges for all $x > 0$.

Functions defined by integral transforms (cont.)

The Cauchy–Stieltjes functions are those functions that can be written as

$$f(z) = \int_\Gamma \frac{d\gamma(\omega)}{z-\omega}, \quad z \in \mathbb{C} \setminus \Gamma,$$

where γ is a (complex) measure supported on a closed set $\Gamma \subset \mathbb{C}$ and the integral is absolutely convergent.

Here we mostly consider Cauchy–Stieltjes functions of the form

$$f(x) = \int_{-\infty}^{0} \frac{d\gamma(\omega)}{x-\omega}, \quad x \in \mathbb{C} \setminus (-\infty, 0],$$

where γ is now a (possibly signed) real measure.

We note that functions of the form

$$f(x) = \int_0^\infty \frac{d\mu(\omega)}{x+\omega} = \int_{-\infty}^{0} \frac{d\gamma(\omega)}{x-\omega}, \quad \gamma(\omega) = -\mu(-\omega) \quad (x > 0),$$

are both of Laplace–Stieltjies and Cauchy–Stieltjes type, so the two classes overlap.

Functions defined by integral transforms (cont.)

Together, these two classes of functions cover many of the functions we are interested in: exponential, resolvent, negative fractional roots, ψ functions, and so forth. Also, these functional classes are closed under multiplication and linear combinations with positive coefficients.

For M Hermitian and positive definite and f in the Laplace–Stieltjes class, we can write

$$f(M) = \int_0^\infty \mathrm{e}^{-\tau M} \mathrm{d}\alpha(\tau).$$

In particular, the (i, j) entry of $f(M)$ is given by

$$[f(M)]_{ij} = \int_0^\infty [\mathrm{e}^{-\tau M}]_{ij} \, \mathrm{d}\alpha(\tau)$$

an expression which can be used to prove decay bounds on the entries of $f(M)$ starting from known decay bounds for the entries of $\mathrm{e}^{-\tau M}$.

Functions defined by integral transforms (cont.)

Similarly, for a given vector b we have

$$f(M)b = \int_0^\infty \mathrm{e}^{-\tau M} b \, \mathrm{d}\alpha(\tau).$$

and from this we can derive convergence estimates for Krylov subspace approximations to $f(M)b$ from known error bounds for the action of the matrix exponential on a vector.

When $M = \mathcal{A} = M_1 \oplus M_2$, we can further exploit the identity

$$\mathrm{e}^{M_1 \oplus M_2} = \mathrm{e}^{M_1} \otimes \mathrm{e}^{M_2}$$

to derive more precise bounds and estimates, and to design more efficient algorithms that exploit the Kronecker structure.

The same approach can be used for Cauchy–Stieltjes functions, exploiting known results for the resolvent $(\mathcal{A} - \omega I)^{-1}$.

Decay results for matrix functions

It is known that if $M = M^*$ is a banded matrix and $f(z)$ is an analytic function defined on an open set $\Omega \subseteq \mathbb{C}$ containing $\sigma(M)$, then there exist explicitly computable constants $K > 0$ and $\alpha > 0$ such that

$$|[f(M)]_{ij}| \leq K\, \mathrm{e}^{-\alpha|i-j|}, \quad 1 \leq i, j \leq n$$

i.e., the entries of $f(M)$ are bounded in an exponentially decaying manner away from the main diagonal.

Moreover, if $\{M_n\}$ is a sequence of $n \times n$ banded Hermitian matrices of increasing size such that $\sigma(M_n) \subset \Omega$ for all n and the distance between the spectra and the singularities of f remains bounded away from zero as $n \to \infty$, then K and α are independent of n.

Decay results for matrix functions (cont.)

This result also holds for general sparse matrices, with $|i - j|$ replaced by the graph distance. Results for non-Hermitian matrices also exist.

Such decay bounds and estimates are useful in many settings, from numerical analysis to quantum physics.

M. B. and G. H. Golub, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–418.

M. B. and N. Razouk, *Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices*, ETNA, 28 (2007), pp. 16–39.

M. B., P. Boito, and N. Razouk, *Decay properties of spectral projectors with applications to electronic structure*, SIAM Review, 55 (2013), pp. 3–64.

Decay results for matrix functions (cont.)

In the case of the matrix exponential, decay is actually superexponential, and the bounds above are not sharp.

Better estimates can be obtained by exploiting the following well known result by Hochbruck and Lubich (SINUM, 1997):

Theorem

Let M be a Hermitian positive semidefinite matrix with eigenvalues in the interval $[0, 4\rho]$. Then the error in the Arnoldi approximation of $\exp(-\tau M)v$ *with* $\|v\| = 1$*, namely*

$$\varepsilon_m := \|\exp(-\tau M)v - V_m \exp(-\tau H_m)e_1\|,$$

is bounded in the following ways:

i) $\varepsilon_m \leq 10 \exp(-m^2/(5\rho\tau))$, *for* $\rho\tau \geq 1$ *and* $\sqrt{4\rho\tau} \leq m \leq 2\rho\tau$

ii) $\varepsilon_m \leq 10(\rho\tau)^{-1} \exp(-\rho\tau) \left(\frac{e\rho\tau}{m}\right)^m$ *for* $m \geq 2\rho\tau$.

Decay results for matrix functions (cont.)

Applying this result with $v = e_j$ leads to the following estimates (B. and Simoncini, SIMAX 2015):

Theorem

Let M be as in the previous Theorem. Assume in addition that M is β-banded. For $i \neq j$, let $\xi = \lceil |i-j|/\beta \rceil$. Then

i) *For $\rho\tau \geq 1$ and $\sqrt{4\rho\tau} \leq \xi \leq 2\rho\tau$,*

$$|[\exp(-\tau M)]_{ij}| \leq 10 \exp\left(-\frac{1}{5\rho\tau}\xi^2\right);$$

ii) *For $\xi \geq 2\rho\tau$,*

$$|[\exp(-\tau M)]_{ij}| \leq 10 \frac{\exp(-\rho\tau)}{\rho\tau}\left(\frac{e\rho\tau}{\xi}\right)^{\xi}.$$

Decay results for matrix functions (cont.)

Theorem

Let $M = M^*$ be β-banded and positive definite, and let $\widehat{M} = M - \lambda_{\min} I$, with the spectrum of $\widehat{M}$ contained in $[0, 4\rho]$. Assume f is a Laplace–Stieltjes function. For $\xi = \lceil |i-j|/\beta \rceil \geq 2$, we have:

$$
\begin{aligned}
|[f(M)]_{ij}| &\leq \int_0^\infty \exp(-\lambda_{\min}\tau) |[\exp(-\tau \widehat{M})]_{ij}| \mathrm{d}\alpha(\tau) \\
&\leq 10 \int_0^{\frac{\xi}{2\rho}} \exp(-\lambda_{\min}\tau) \frac{\exp(-\rho\tau)}{\rho\tau} \left(\frac{\mathrm{e}\rho\tau}{\xi}\right)^{\xi} \mathrm{d}\alpha(\tau) \qquad (1) \\
&\quad +10 \int_{\frac{\xi}{2\rho}}^{\frac{\xi^2}{4\rho}} \exp(-\lambda_{\min}\tau) \exp\left(-\frac{\xi^2}{5\rho\tau}\right) \mathrm{d}\alpha(\tau) \qquad (2) \\
&\quad + \int_{\frac{\xi^2}{4\rho}}^{\infty} \exp(-\lambda_{\min}\tau) |[\exp(-\tau \widehat{M})]_{i,j}| \mathrm{d}\alpha(\tau) = I + II + III.
\end{aligned}
$$

Decay results for matrix functions (cont.)

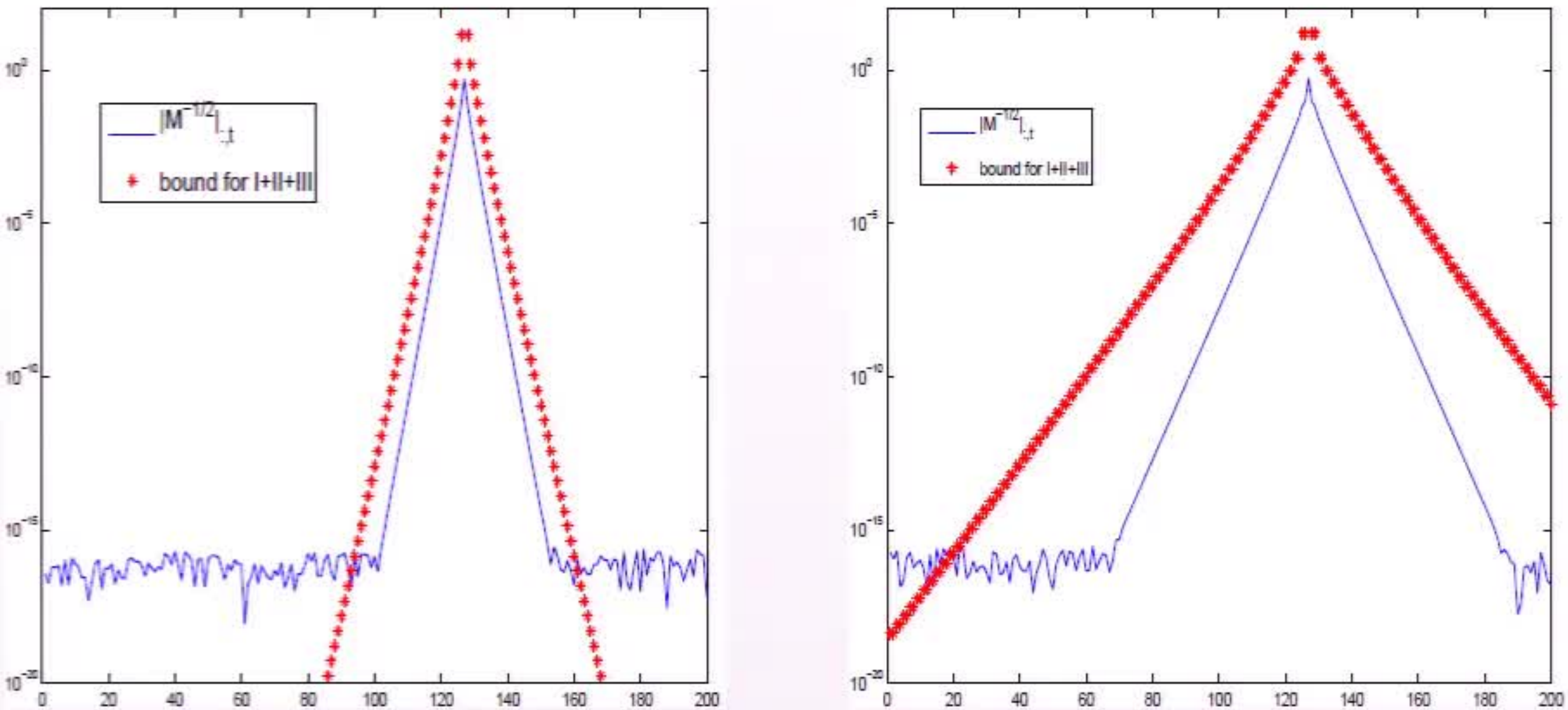


Figure: Estimates for $|[M^{-1/2}]_{:,j}|$, $j = 127$, using I+II and the upper bound for III. Size $n = 200$, Logarithmic scale. Left: $M = \text{tridiag}(-1, 4, -1)$. Right: $M = \text{pentadiag}(-0.5, -1, 4, -1, -0.5)$. Note the exponential rate of decay.

Decay results for matrix functions (cont.)

Similar results can be obtained for Cauchy–Stieltjes matrix functions, combined with (classical) decay bounds for the inverse of a banded SPD matrix due to Demko, Moss and Smith (Math. Comp., 1983).

These bounds can be extended to sparse (rather than banded) matrices using the notion of graph distance.

However, for Kronecker sums of matrices, the oscillatory decay behavior in $f(\mathcal{A})$ is better captured if we exploit the Kronecker structure.

For the special case of $\mathcal{A}^{-1}$ with $\mathcal{A} = M_2 \otimes I_{n_1} + I_{n_2} \otimes M_1$, see

C. Canuto, V. Simoncini, and M. Verani, *On the decay of the inverse of matrices that are sum of Kronecker products*, Linear Algebra Appl., 452 (2014), pp. 21–39.

Decay results for matrix functions (cont.)

Recall that

$$\exp(-\tau\mathcal{A}) = \exp(-\tau M_1) \otimes \exp(-\tau M_2), \quad \tau \in \mathbb{R}$$

when $\mathcal{A} = M_1 \otimes I + I \otimes M_2$.

We introduce a lexicographic ordering of the entries, so that each row or column index k of $\mathcal{A}$ corresponds to the ordered pair $k = (k_1, k_2)$ in a two-dimensional Cartesian grid. In other words, the generic entry of $\mathcal{A}$ has row index $(k_1 - 1)n + t_1$ and column index $(k_2 - 1)n + t_2$.

Furthermore, for $|i - j| > \sqrt{4\rho\tau}\beta$ define the function

$$\Phi(i,j) := \begin{cases} 10 \exp\left(-\frac{\xi^2}{5\rho\tau}\right), & \text{for} \quad \sqrt{4\rho\tau} \le \xi \le 2\rho\tau, \\ 10\frac{\exp(-\rho\tau)}{\rho\tau}\left(\frac{e\rho\tau}{\xi}\right)^{\xi}, & \text{for} \quad \xi \ge 2\rho\tau. \end{cases}$$

where again $\xi = \lceil |i - j|/\beta \rceil$.

Decay results for matrix functions (cont.)

Theorem

Let $\mathcal{A} = I \otimes M_1 + M_2 \otimes I$ with M_1 and M_2 Hermitian and positive semidefinite with bandwidth β_1, β_2 and spectrum contained in $[0, \rho_1]$ and $[0, \rho_2]$, respectively. Denote with Φ_ℓ the function described above, with $\rho = \rho_\ell$ and $\beta = \beta_\ell$ ($\ell = 1, 2$). Then for $t = (t_1, t_2)$ and $k = (k_1, k_2)$, with $|t_\ell - k_\ell| \geq \sqrt{4\rho_\ell \tau}\beta_\ell$, $\ell = 1, 2$, we have

$$|(\exp(-\tau \mathcal{A}))_{kt}| \leq \Phi_1(k_1, t_1)\, \Phi_2(k_2, t_2).$$

For f a Laplace–Stieltjies function, we can use this result to bound the entries of $f(\mathcal{A})$.

Note: A similar approach also works for Cauchy–Stieltjies functions (see our paper).

Decay results for matrix functions (cont.)

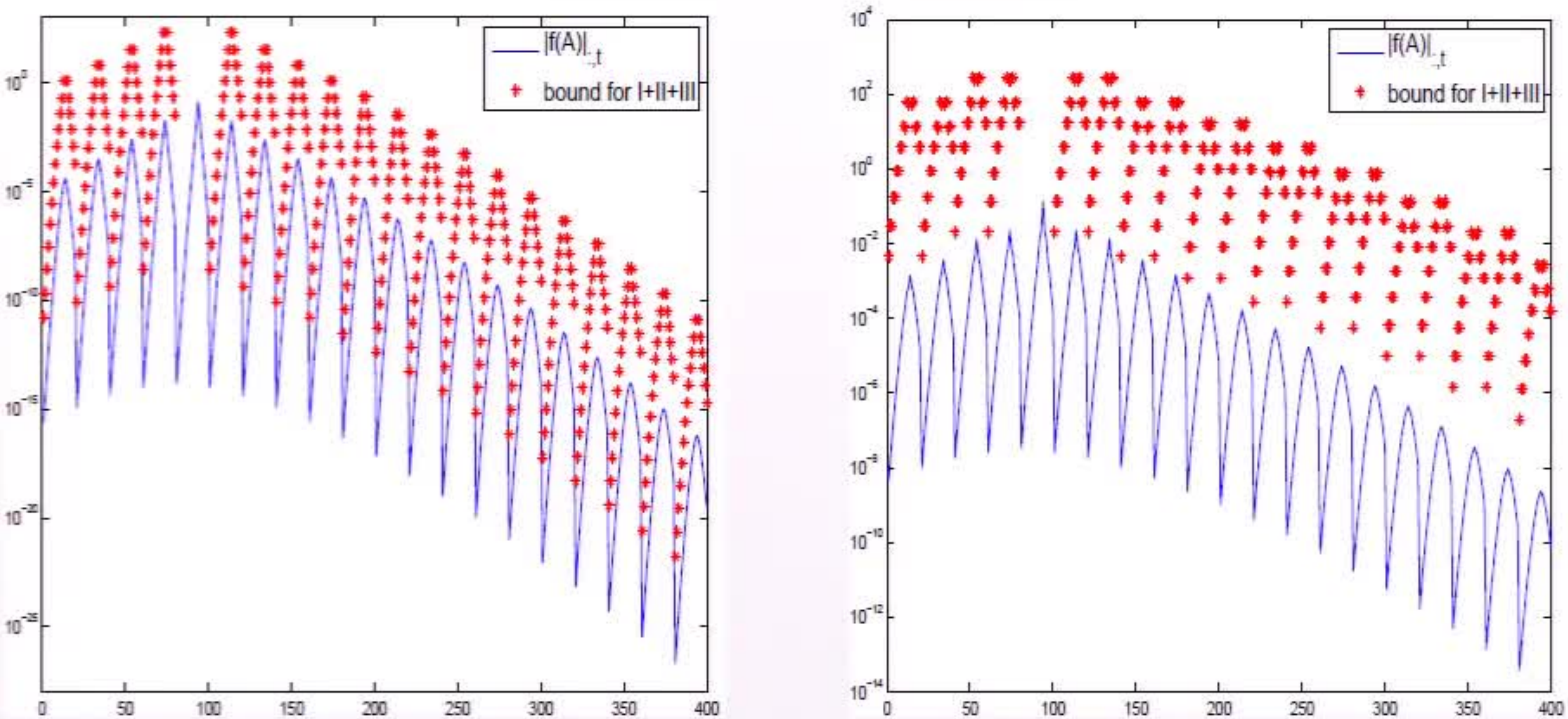


Figure: True decay and estimates for $|[\psi_1(\mathcal{A})]_{:,t}|$, $t = 94$, with $\psi_1(x) = (\mathrm{e}^x - 1)/x$ and $\mathcal{A} = M \otimes I + I \otimes M$ of size $n = 400$. Left: $M = \text{tridiag}(-1, 4, -1)$. Right: $M = \text{pentadiag}(-0.5, -1, 4, -1, -0.5)$.

Approximating $f(\mathcal{A})\,b$, with $\mathcal{A} = M_1 \oplus M_2$ (cont.)

In particular, $(P_m \otimes Q_m)b = \text{vec}((Q_m^T b_1)(b_2^T P_m))$.

Advantages of the structure-aware approximation:

- Memory requirements reduced from mn^2 to $2mn$
- Accurate approximation obtained with subspace of much smaller dimension
- $f(\mathcal{T}_m)$ can be computed without forming $\mathcal{T}_m$ using the eigendecomposition of T_1, T_2
- Applicable to non-Hermitian case as well

Limitation: right-hand side must have special structure.

There are, however, several applications where this structure is present.

An example ($n = 50$)

m	$\|f(\mathcal{A})b - x_m\|$	$\|f(\mathcal{A})b - x_m^{\otimes}\|$	$\frac{\|x_m - x_{m,old}\|}{\|x_m\|}$	$\frac{\|x_m^{\otimes} - x_{m,old}^{\otimes}\|}{\|x_m^{\otimes}\|}$
4	4.2422e-01	3.9723e-01	1.0000e+00	1.0000e+00
8	2.6959e-01	2.1025e-01	2.2710e-01	2.5313e-01
12	1.7072e-01	1.0365e-01	1.3066e-01	1.2971e-01
16	1.0324e-01	4.2407e-02	8.3444e-02	6.9960e-02
20	5.7342e-02	1.1176e-02	5.4224e-02	3.3969e-02
24	2.7550e-02	4.8230e-04	3.4054e-02	1.0935e-02
28	1.0351e-02	2.8883e-12	1.9296e-02	4.8230e-04
32	3.4273e-03	2.8496e-12	8.3585e-03	1.1366e-13
36	2.2906e-03	2.9006e-12	1.7514e-03	1.4799e-13
40	9.4368e-04	2.8119e-12	1.6283e-03	2.7323e-13
44	4.3935e-04	2.7593e-12	6.2797e-04	2.1786e-13
48	1.8744e-04	2.8235e-12	3.0332e-04	2.5965e-13

Table: Results for $f(x) = (e^{s\sqrt{z}} - 1)/z$ with $s = 10^{-3}$. Rank-1 right-hand side. Here $n = 50$.

An example ($n = 100$)

m	$\frac{\|x_m - x_{m,old}\|}{\|x_m\|}$	$\frac{\|x_m^{\otimes} - x_{m,old}^{\otimes}\|}{\|x_m^{\otimes}\|}$
4	1.0000e+00	1.0000e+00
8	2.3942e-01	2.7720e-01
12	1.5010e-01	1.6289e-01
16	1.0716e-01	1.0966e-01
20	8.1062e-02	7.8150e-02
24	6.3308e-02	5.7003e-02
28	5.0347e-02	4.1674e-02
32	4.0409e-02	2.9992e-02
36	3.2507e-02	2.0802e-02
40	2.6052e-02	1.3446e-02
44	2.0667e-02	7.5529e-03
48	1.6104e-02	2.9970e-03
52	1.2194e-02	3.1470e-04
56	8.8234e-03	1.1354e-12
60	5.9194e-03	3.4639e-13

Table: Same matrix function as before, case $n = 100$.

Convergence estimates for structured approximations

Let $\mathcal{A} = M \otimes I + I \otimes M$ (Hermitian), $\lambda_{\min}$, $\lambda_{\max}$ extreme eigenvalues of M, and $\widehat{\kappa} = \frac{\lambda_{\min}+\lambda_{\max}}{\lambda_{\min}+\lambda_{\min}}$.

Then the following holds:

- If f is a Laplace–Stieltjes function,

$$\|f(\mathcal{A}) - x_m^{\otimes}\| = \mathcal{O}\left(\exp\left(-\frac{2m}{\sqrt{\widehat{\kappa}}}\right)\right)$$

  for m and $\widehat{\kappa}$ sufficiently large.

- If f is a Cauchy–Stieltjes function,

$$\|f(\mathcal{A}) - x_m^{\otimes}\| \leq C\left(\frac{\sqrt{\widehat{\kappa}}-1}{\sqrt{\widehat{\kappa}}+1}\right)^m$$

  where C is computable and depends on f and M.

- For $\widehat{\kappa}$ large, the two estimates are equivalent.

Proofs based on similar techniques as those used for decay bounds.