ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

MATHICSE
Mathematics Institute of Computational Science and Engineering
School of Basic Sciences - Section of Mathematics

MATHICSE Technical Report
Nr. 26.2013
August 2013 (New 06.01.2014)

On the finite section method for computing exponentials of doubly-infinite skew-Hermitian matrices

Meiyue SHAO

http://mathicse.epfl.ch

Phone: +41 21 69 37648

Address:
EPFL - SB - MATHICSE (Bâtiment MA)
Station 8 - CH-1015 - Lausanne - Switzerland

Fax: +41 21 69 32545

On the finite section method for computing exponentials of doubly-infinite skew-Hermitian matrices

Meiyue Shao*

January 6, 2014

Abstract

Computing the exponential of large-scale skew-Hermitian matrices or parts thereof is frequently required in applications. In this work, we consider the task of extracting finite diagonal blocks from a doubly-infinite skew-Hermitian matrix. These matrices usually have unbounded entries which impede the application of many classical techniques from approximation theory. We analyze the decay property of matrix exponentials for several classes of banded skew-Hermitian matrices. Then finite section methods based on the decay property are presented. We use several examples to demonstrate the effectiveness of these methods.

Keywords. Matrix exponential, doubly-infinite matrices, finite section method, banded matrices, exponential decay

AMS subject classifications. 65F60

1 Introduction

In a number of scientific applications, especially in quantum mechanics, it is desirable to compute $\exp(\mathrm{i}A)$ where A is a self-adjoint operator. For example, $\exp(\mathrm{i}A)$ naturally appears in the solution of the time-dependent Schrödinger equation [7]. We refer to, e.g., [9] for applications from other domains. In practice, the operator A is often given in discretized form, i.e., a doubly-infinite Hermitian matrix under a certain basis, and a finite diagonal block of $\exp(\mathrm{i}A)$ is of interest. Suppose the $(-m : m, -m : m)$ block[1] of $\exp(\mathrm{i}A)$ is desired. A simple way to solve this problem is illustrated in Figure 1. We first compute the exponential of the $(-w : w, -w : w)$ block of A, where w is chosen somewhat larger than m, and then use its central $(2m+1) \times (2m+1)$ block to approximate the desired solution. In reference to similar methods for solving linear systems [15, 20], we call this approach *finite*

*ANCHP, MATHICSE, EPF Lausanne, CH-1015 Lausanne, Switzerland. Email: meiyue.shao@epfl.ch, telephone: +41-21-693 25 31.

[1]The MATLAB colon notation $i : j$ represents a set of consecutive integers $\{i, i+1, \ldots, j\}$.

section method. The diagonal blocks $(-m:m,-m:m)$ and $(-w:w,-w:w)$ are called the *desired window* and the *computational window*, respectively.

To our knowledge, much of the existing literature on infinite matrices is concerned with solving infinite dimensional linear systems, see e.g., [5, 6, 27] and the references therein. The matrix exponential problem for infinite matrices has also been studied [14, 16]. Despite the simplicity of the finite section method, it is crucial to ask how large the computational window needs to be, and whether this truncation produces sufficiently accurate approximation to the true solution. These questions are relatively easy to answer for bounded matrices, where standard polynomial approximation technique can be applied. But it turns out that the finite section method can also be applied to certain unbounded matrices, and still produces reliable solutions. For example, Figure 1 illustrates this for an unbounded Wilkinson-type matrix $W^-(1)$, see Section 3.2, for which the error decays quickly when the size of the computational window increases. In this paper we will explain this phenomenon and establish the finite section method with error estimates for several classes of doubly-infinite Hermitian matrices.

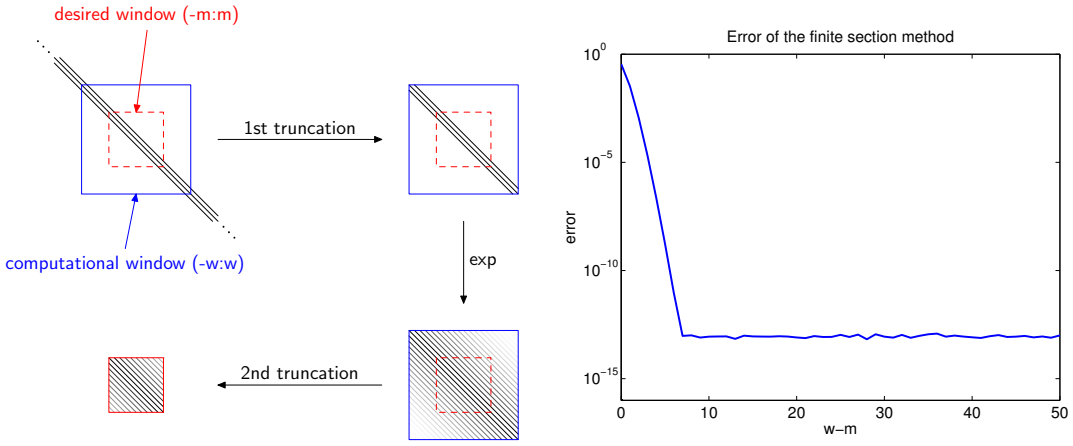


Figure 1: A pictorial illustration of the finite section method. In this case A is the Wilkinson-type matrix $W^-(1)$.

The rest of this paper is organized as follows. In Section 2, we discuss the decay property of $\exp(\mathrm{i}A)$ for a bounded matrix A and show how this can be used to analyze the finite section method. In Section 3, we first analyze decay of entries for Wilkinson-type matrices and derive the corresponding finite section method, and then discuss some extensions to more general unbounded matrices. Finally, numerical experiments are presented in Section 4 to demonstrate the reliability of finite section methods.

2 The Finite Section Method for Bounded Matrices

To analyze the accuracy of finite section methods, we start by discussing a relatively simple case—A is a bounded Hermitian matrix. To set the stage, let us first give a formal mathematical formulation of the problem, see also [1]. A doubly-infinite matrix is a two dimensional array $A = [a_{ij}]$ of complex numbers with $i,\ j \in \mathbb{Z}$. It is called Hermitian (or skew-Hermitian) if $a_{ij} = \overline{a_{ji}}$ (or $a_{ij} = -\overline{a_{ji}}$) for all $i,\ j \in \mathbb{Z}$. If there exists an even number b such that $a_{ij} = 0$ when $|i-j| > b/2$, then A is called b-*banded*, or *banded* in short. Matrix-matrix and matrix-vector multiplications are defined akin to those operations for finite matrices. That is,

$$\begin{aligned} (AB)_{ij} &= \sum_{k\in\mathbb{Z}} a_{ik}b_{kj}, \\ (Ax)_i &= \sum_{k\in\mathbb{Z}} a_{ik}x_k, \end{aligned} \tag{1}$$

where A, B are doubly-infinite matrices and $x = [x_i]$ is a doubly-infinite vector, provided that these summations converge absolutely. Evidently, multiplications involving banded matrices are always well-defined since all summations are finite. A doubly-infinite matrix A is called *bounded* if

$$\|A\|_2 := \sup_{\substack{x\in l^2(\mathbb{Z}) \\ \|x\|_2\le 1}} \|Ax\|_2 < +\infty;$$

in this case A can be interpreted as a continuous linear operator over $l^2(\mathbb{Z})$ [1]. By Gelfand's formula [12], we have

$$\rho(A) = \lim_{k\to\infty} \|A^k\|_2^{\frac{1}{k}} \le \|A\|_2,$$

indicating that the spectrum $\Lambda(A)$ is also bounded. For an analytic function $F(z)$ defined on a domain Ω which encloses $\Lambda(A)$ (i.e., the spectrum of A), the matrix function $F(A)$ is defined as

$$F(A) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial\Omega} F(z)(zI - A)^{-1}\,\mathrm{d}z.$$

In this paper we are interested in the exponential function $F(z) = \exp(\mathrm{i}z)$.

Throughout this section, we assume that the doubly-infinite matrix A is Hermitian and b-banded. In the following, we first recall the exponential decay property of $\exp(\mathrm{i}A)$ and then establish finite section methods based on this property.

2.1 The exponential decay property

It is well-known [3, 4, 5, 19] that when B is a finite banded matrix, the entries of $F(B)$ decay exponentially from the diagonal where F is an analytic function defined on a domain containing $\Lambda(B)$, see in particular [3, Section 2]. The decay properties easily carry over to functions of bounded doubly-infinite matrices. In the following we briefly recall these results.

Definition 1. *We say that a matrix $A = [a_{ij}]$ has the* exponential decay property *if there exist $K > 0$ and $\rho \in (0,1)$ such that*

$$|a_{ij}| \le K\rho^{-|i-j|}, \qquad \forall i,\ j. \tag{2}$$

The constant ρ is called the decay rate*. If for any $\rho \in (0,1)$ there exists a positive number K such that (2) holds for all i, j, we say that A decays* super-exponentially*.*

We remark that all finite matrices trivially have the exponential decay property by choosing sufficiently large K. Hence this property is meaningful only when K can be chosen moderately small and ρ is not too close to one. However, for a doubly-infinite matrix A, the exponential decay property is nontrivial since it implies the boundedness of A in $l^2(\mathbb{Z})$. In fact, both $\|A\|_1 := \sup_j \sum_{i\in\mathbb{Z}} |a_{ij}|$ and $\|A\|_\infty := \sup_i \sum_{j\in\mathbb{Z}} |a_{ij}|$ are bounded by $K(1+\rho)(1-\rho)^{-1}$ and hence the Schur test [24] implies $\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty} \le K(1+\rho)(1-\rho)^{-1}$.

Proofs of the exponential decay properties of certain matrix functions are usually built on polynomial approximation (see, e.g., [3, 4, 10]), i.e., approximate $F(B)$ by a matrix polynomial $p(B)$ where $p \in \mathcal{P}_k$ is a polynomial of degree at most k. The propositions below will be used in our analyses.

Lemma 1 (Bernstein)**.** [3, 21] *Let $\mathcal{E}_\chi$ ($\chi > 1$) be the* Bernstein ellipse *in the complex plane defined by*

$$\frac{\Re(z)^2}{\left(\chi+\chi^{-1}\right)^2} + \frac{\Im(z)^2}{\left(\chi-\chi^{-1}\right)^2} = \frac{1}{4}.$$

Then for any function F being analytic in the domain enclosed by $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$, we have

$$\inf_{p\in\mathcal{P}_k} \|F-p\|_\infty \le \frac{2M(\chi)}{\chi^k(\chi-1)}, \qquad (k \in \mathbb{N})$$

where

$$M(\chi) = \max_{z\in\mathcal{E}_\chi} |F(z)|.$$

Theorem 2 (Benzi-Golub)**.** [3] *Let B be a b-banded Hermitian matrix whose spectrum is contained in the interval $[-1,1]$. Then for any function F being analytic*[2] *inside the Bernstein ellipse $\mathcal{E}_\chi$ and continuous on $\mathcal{E}_\chi$, there exist constants $K > 0$ and $\rho \in (0,1)$ such that*

$$|[F(B)]_{ij}| \le K\rho^{|i-j|}.$$

More precisely, these constants can be chosen as

$$K = \max\left\{\frac{2\chi M(\chi)}{\chi-1},\ \|F(B)\|_2\right\} \quad \text{and} \quad \rho = \chi^{-\frac{2}{b}}.$$

[2] In [3], F is additionally required to be real analytic. But this extra assumption turns out to be unnecessary, see, e.g., [21, page 76].

Although Theorem 2 is derived only for finite matrices in [3], the same proof is valid for analytic functions of a doubly-infinite Hermitian b-banded matrix B as long as $\Lambda(B) \subset [-1, 1]$ holds. Now we consider a doubly-infinite b-banded Hermitian matrix A with spectrum $\Lambda(A) \subset [\lambda_0 - \Delta, \lambda_0 + \Delta]$.[3] In the following we show that $\exp(\mathrm{i}A)$ has the super-exponential decay property.

Corollary 3. *Suppose A is a doubly-infinite b-banded Hermitian matrix with $\Lambda(A) \subset [\lambda_0 - \Delta, \lambda_0 + \Delta]$. For any $\chi > 1$, let*

$$K = \frac{2\chi}{\chi - 1} \exp\Big[\frac{\Delta(\chi^2 - 1)}{2\chi}\Big] \quad \textit{and} \quad \rho = \chi^{-\frac{2}{b}}. \tag{3}$$

Then

$$\Big|\big[\exp(\mathrm{i}A)\big]_{ij}\Big| \le K\rho^{|i-j|}, \qquad \forall i,\, j \in \mathbb{Z}. \tag{4}$$

Moreover, $\exp(\mathrm{i}A)$ *has the super-exponential decay property.*

Proof. For $\chi > 1$, we set

$$M(\chi) = \max_{z \in \mathcal{E}_\chi} \big|\exp(\mathrm{i}\Delta z)\big| = \max_{\substack{x+\mathrm{i}y \in \mathcal{E}_\chi \\ x,y \in \mathbb{R}}} \big|\exp[-\Delta y]\big| = \exp\Big[\frac{\Delta(\chi^2 - 1)}{2\chi}\Big].$$

Applying Theorem 2 to $F(z) = \exp(\mathrm{i}\Delta z)$ with $B = (A - \lambda_0 I)/\Delta$, which has spectrum contained in $[-1, 1]$, yields

$$\Big|\big(\exp\big[\mathrm{i}(A - \lambda_0 I)\big]\big)_{ij}\Big| = |[F(B)]_{ij}| \le K\rho^{|i-j|}, \tag{5}$$

where $\rho = \chi^{-\frac{2}{b}}$ and

$$K = \max\left\{\frac{2\chi M(\chi)}{\chi - 1}, 1\right\} = \frac{2\chi M(\chi)}{\chi - 1} = \frac{2\chi}{\chi - 1} \exp\Big[\frac{\Delta(\chi^2 - 1)}{2\chi}\Big] > 1.$$

The bound (4) now follows from (5) using the fact that

$$|\exp(\mathrm{i}A)| = |\exp(\mathrm{i}\lambda_0)| \cdot \big|\exp\big[\mathrm{i}(A - \lambda_0 I)\big]\big| = \big|\exp\big[\mathrm{i}(A - \lambda_0 I)\big]\big|.$$

For any $\rho \in (0, 1)$ we choose $\chi = \rho^{-\frac{b}{2}}$ and $K = 2\chi \exp\big[\Delta(\chi^2 - 1)/(2\chi)\big]/(\chi - 1)$ according to (3) so that (4) holds for all $i,\, j \in \mathbb{Z}$. Therefore, by definition $\exp(\mathrm{i}A)$ has the super-exponential decay property. □

Since $F(z) = \exp(\mathrm{i}\Delta z)$ is an entire function, in Corollary 3 we can choose any χ from the interval $(1, +\infty)$ to bound $\big|[\exp(\mathrm{i}A)]_{ij}\big|$. Sometimes an upper bound of $\big|[\exp(\mathrm{i}A)]_{ij}\big|$ for a given entry (i, j) is of interest. Thus it is desirable to find a χ that minimizes $K\rho^{|i-j|}$. Such a choice is made by taking $\theta = 1$ and $d = |i - j|$ in the following theorem.

[3] Without loss of generality, we always assume that $\Delta > 0$.

Theorem 4. *Let b, d, Δ, and θ be positive numbers. Then the function*

$$g(\chi) = \Big(\frac{2\chi}{\chi-1}\Big)^{\theta} \exp\Big[\frac{\Delta(\chi^2-1)}{2\chi}\Big]\chi^{-\frac{2d}{b}}, \qquad (\chi > 1)$$

has a unique minimum at $\chi = \chi_*$, where χ_* is the unique root of the cubic equation

$$\chi^3 - \Big(1+\frac{4d}{b\Delta}\Big)\chi^2 + \Big(1+\frac{4d-2b\theta}{b\Delta}\Big)\chi - 1 = 0$$

in the interval $(1,+\infty)$. Moreover, we have

$$\lim_{d\to+\infty} \frac{\chi_*}{d} = \frac{4}{b\Delta}.$$

Proof. We first notice that $\lim_{\chi\to 1+} g(\chi) = \lim_{\chi\to+\infty} g(\chi) = +\infty$. Therefore $g(\chi)$ has at least one minimum in the interval $(1,+\infty)$ as $g(\chi)$ is continuously differentiable. Any minimizer χ_* satisfies the condition

$$0 = \frac{\mathrm{d}g(\chi)}{\mathrm{d}\chi}\bigg|_{\chi=\chi_*} = g(\chi_*)\Big[-\frac{\theta}{\chi_*(\chi_*-1)} + \frac{\Delta(\chi_*^2+1)}{2\chi_*^2} - \frac{2d}{b\chi_*}\Big].$$

Multiplying by $2\chi_*^2(\chi_*-1)/[g(\chi_*)\Delta]$ yields $h(\chi_*) = 0$, where

$$h(\chi) = \chi^3 - \Big(1+\frac{4d}{b\Delta}\Big)\chi^2 + \Big(1+\frac{4d-2b\theta}{b\Delta}\Big)\chi - 1.$$

Since $h(1) < 0$, $h(\chi)$ has either one root or three roots in the interval $(1,+\infty)$ according to the monotonicity of a real cubic function. If there are three roots in $(1,+\infty)$, the product of all these roots is greater than one. This contradicts Vieta's formulas which imply that the product of the roots of $h(\chi)$ is one. Therefore, $h(\chi)$ has only one root in $(1,+\infty)$. This root is also the unique minimizer of $g(\chi)$.

To obtain the asymptotic behavior of χ_*, we consider the function

$$\tilde{h}(\tilde{\chi}) = \frac{h(\chi)}{d^3} = \tilde{\chi}^3 - \Big(\frac{1}{d}+\frac{4}{b\Delta}\Big)\tilde{\chi}^2 + \Big(\frac{1}{d^2}+\frac{4}{bd\Delta}-\frac{2\theta}{d^2\Delta}\Big)\tilde{\chi} - \frac{1}{d^3},$$

where $\tilde{\chi} = \chi/d$. Since χ_* is the largest real root of $h(\chi)$, $\tilde{\chi}_* = \chi_*/d$ is also the largest real root of $\tilde{h}(\tilde{\chi})$. When $d = +\infty$, we have $\tilde{\chi}_* = 4/(b\Delta)$. Notice that $\tilde{\chi}_*$ is a continuous function of the coefficients of $\tilde{h}(\tilde{\chi})$ (see, e.g., [31]). Therefore, we obtain

$$\lim_{d\to+\infty} \frac{\chi_*}{d} = \lim_{d\to+\infty} \tilde{\chi}_* = \frac{4}{b\Delta}. \qquad \square$$

2.2 The finite section method

We have seen that $\exp(\mathrm{i}tA)$ has the (super-)exponential decay property when a doubly-infinite matrix A is banded Hermitian and bounded. We now make use of this property to establish the finite section method with guaranteed accuracy.

Suppose A is partitioned into the form

$$A = \begin{bmatrix} A_{11} & A_{12} & \\ A_{21} & A_{22} & A_{23} \\ & A_{32} & A_{33} \end{bmatrix},$$

where A_{22} corresponds to the computational window. The finite section method extracts the desired window from $\exp(\mathrm{i}A_{22})$. The matrix $\exp(\mathrm{i}A_{22})$ is the central diagonal block in $\exp(\mathrm{i}\,\mathrm{Diag}\,\{A_{11}, A_{22}, A_{33}\})$, and the latter can be viewed as the exact exponential of a perturbed matrix. Hence a perturbation analysis of the matrix exponential as in [28] is helpful for studying the truncation error in the finite section method.

Notice that the solution of the linear differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \mathrm{i}Ax(t), \qquad x(t) \in l^2(\mathbb{Z})$$

is given by [8, Theorem 2.1.10]

$$x(t) = \exp(\mathrm{i}tA)x(0).$$

Using this connection between the linear differential equation and the exponential function, it can be shown [8, Theorem 3.2.1] that

$$\exp(\mathrm{i}tA+\mathrm{i}tB)v - \exp(\mathrm{i}tA)v = \mathrm{i}\int_0^t \exp[\mathrm{i}(t-s)A]B\exp[\mathrm{i}s(A+B)]v\,\mathrm{d}s, \qquad \forall v \in l^2(\mathbb{Z}), \quad (6)$$

when both A and B are bounded. The following theorem is based on this perturbation bound.

Theorem 5. *Suppose that*

$$A = \begin{bmatrix} A_{11} & A_{12} & \\ A_{21} & A_{22} & A_{23} \\ & A_{32} & A_{33} \end{bmatrix}$$

is a doubly-infinite b-banded Hermitian matrix with $\Lambda(A) \subset [\lambda_0 - \Delta, \lambda_0 + \Delta]$, where A_{22} is the $(-w:w, -w:w)$ diagonal block of A. Let $\tilde{A} = \mathrm{Diag}\,\{A_{11}, A_{22}, A_{33}\}$ be a block diagonal approximation of A. Then for any $\chi > 1$, we have

$$\left|\left[\exp(\mathrm{i}A) - \exp(\mathrm{i}\tilde{A})\right]_{ij}\right| \le K\big(\rho^{|w-i|+|w-j|-\frac{b}{2}} + \rho^{|w+i|+|w+j|-\frac{b}{2}}\big), \qquad \forall i,\ j$$

where

$$K = \frac{b(b+2)}{4}\max\big\{\|A_{12}\|_2, \|A_{23}\|_2\big\}\Big(\frac{2\chi}{\chi-1}\Big)^2 \exp\Big[\frac{\Delta(\chi^2-1)}{2\chi}\Big] \quad \textit{and} \quad \rho = \chi^{-\frac{2}{b}}. \qquad (7)$$

Proof. Let

$$A_0 = \begin{bmatrix} A_{11} & A_{12} & \\ A_{21} & A_{22} & 0 \\ & 0 & A_{33} \end{bmatrix}, \qquad B = A_0 - A.$$

Then $\Lambda(A_0) \subset [\lambda_0 - \Delta, \lambda_0 + \Delta]$ because

$$\begin{aligned} \rho(A_0 - \lambda_0 I) &= \|A_0 - \lambda_0 I\|_2 \\ &= \max\left\{ \left\| \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix} \right] - \lambda_0 I \right\|_2, \|A_{33} - \lambda_0 I\|_2 \right\} \le \|A - \lambda_0 I\|_2 = \rho(A - \lambda_0 I). \end{aligned}$$

Similarly, it can be shown that $\Lambda(\tilde{A}) \subset [\lambda_0 - \Delta, \lambda_0 + \Delta]$. Using (6), we obtain

$$\exp(\mathrm{i}A_0)e_j - \exp(\mathrm{i}A)e_j = \mathrm{i}\int_0^1 \exp[\mathrm{i}(1-s)A]B\exp(\mathrm{i}sA_0)e_j \,\mathrm{d}s, \qquad \forall j \in \mathbb{Z}.$$

For each $s \in [0,1]$, we denote by $U(s) = \exp[\mathrm{i}(1-s)A]$ and $V(s) = \exp(\mathrm{i}sA_0)$. Then for i, $j \in \mathbb{Z}$, we conclude from the nonzero structure of B that

$$\begin{aligned} &\Big|\big[U(s)BV(s)\big]_{ij}\Big| \\ =&\left|\sum_{k=1}^{b/2}\sum_{\ell=-b/2+k}^{0} b_{w+k,w+\ell}U_{i,w+k}V_{w+\ell,j} + \sum_{k=-b/2+1}^{0}\sum_{\ell=1}^{b/2+k} b_{w+k,w+\ell}U_{i,w+k}V_{w+\ell,j}\right| \\ \le&\|A_{23}\|_2\Bigg(\sum_{k=1}^{b/2}\sum_{\ell=-b/2+k}^{0}|U_{i,w+k}V_{w+\ell,j}| + \sum_{k=-b/2+1}^{0}\sum_{\ell=1}^{b/2+k}|U_{i,w+k}V_{w+\ell,j}|\Bigg). \end{aligned}$$

Using (3) and (4), we have

$$\begin{aligned} |U_{i,w+k}| &\le \frac{2\chi}{\chi-1}\exp\Big[\frac{(1-s)\Delta(\chi^2-1)}{2\chi}\Big]\rho^{|w+k-i|}, \\ |V_{w+\ell,j}| &\le \frac{2\chi}{\chi-1}\exp\Big[\frac{s\Delta(\chi^2-1)}{2\chi}\Big]\rho^{|w+\ell-j|}, \end{aligned}$$

and thus

$$|U_{i,w+k}V_{w+\ell,j}| \le K_0\rho^{|w+k-i|+|w+\ell-j|} \le K_0\rho^{|w-i|+|w-j|-|k|-|\ell|} \le K_0\rho^{|w-i|+|w-j|-\frac{b}{2}},$$

where

$$K_0 = \Big(\frac{2\chi}{\chi-1}\Big)^2\exp\Big[\frac{\Delta(\chi^2-1)}{2\chi}\Big].$$

Hence, we obtain

$$\Big|\big[U(s)BV(s)\big]_{ij}\Big| = \frac{b(b+2)}{4}\|A_{23}\|_2K_0\rho^{|w-i|+|w-j|-\frac{b}{2}}.$$

Integrating over the interval $[0,1]$ eventually yields

$$\Big|\big[\exp(\mathrm{i}A_0)-\exp(\mathrm{i}A)\big]_{ij}\Big| \le \frac{b(b+2)}{4}\|A_{23}\|_2 K_0 \rho^{|w-i|+|w-j|-\frac{b}{2}} \le K\rho^{|w-i|+|w-j|-\frac{b}{2}}.$$

Following the same analysis above, we obtain another estimate

$$\Big|\big[\exp(\mathrm{i}\tilde{A})-\exp(\mathrm{i}A_0)\big]_{ij}\Big| \le \frac{b(b+2)}{4}\|A_{12}\|_2 K_0 \rho^{|w+i|+|w+j|-\frac{b}{2}} \le K\rho^{|w+i|+|w+j|-\frac{b}{2}}.$$

Then the theorem is proved from

$$\big|\exp(\mathrm{i}\tilde{A})-\exp(\mathrm{i}A)\big| \le \big|\exp(\mathrm{i}A_0)-\exp(\mathrm{i}A)\big| + \big|\exp(\mathrm{i}\tilde{A})-\exp(\mathrm{i}A_0)\big|. \qquad \square$$

Now we are ready to derive the finite section method for computing the diagonal block $\big[\exp(\mathrm{i}A)\big]_{(-m:m,-m:m)}$. Based on Theorem 5, we can choose $w > m$ such that the perturbations introduced at the $\pm w$th rows have only negligible impact on the desired window. For a given entry (i,j) in the desired window, we set $k = (i+j)/2$. Then

$$\begin{aligned}\Big|\big[\exp(\mathrm{i}A)-\exp(\mathrm{i}\tilde{A})\big]_{ij}\Big| &\le K\big(\rho^{|w-i|+|w-j|-\frac{b}{2}} + \rho^{|w+i|+|w+j|-\frac{b}{2}}\big)\\ &= K\big(\rho^{2(w-k)-\frac{b}{2}} + \rho^{2(w+k)-\frac{b}{2}}\big)\\ &\le K\big(\rho^{2(w-m)-\frac{b}{2}} + \rho^{2(w+m)-\frac{b}{2}}\big).\end{aligned}$$

The last inequality is based on the convexity of the function ρ^x. It is then desirable to minimize the right-hand-side in the above inequality. Substituting $d = 2(w-m) - b/2$ and $\theta = 2$ into Theorem 4, we obtain that the parameter χ_* that minimizes $K\rho^{2(w-m)-\frac{b}{2}}$ is the unique root in the interval $(1,+\infty)$ of the cubic equation

$$\chi_*^3 - \Big(1+\frac{4d}{b\Delta}\Big)\chi_*^2 + \Big(1+\frac{4(d-b)}{b\Delta}\Big)\chi_* - 1 = 0. \tag{8}$$

Such a choice is already sufficient for practical purpose since

$$K\rho^{2(w-m)-\frac{b}{2}} + K\rho^{2(w+m)-\frac{b}{2}} \le 2K\rho^{2(w-m)-\frac{b}{2}}$$

and usually $\rho^{2(w+m)-\frac{b}{2}} \ll \rho^{2(w-m)-\frac{b}{2}}$. Finally, we remark that the knowledge of the width of $\Lambda(A)$ is required in order to compute the constant K in (7). In practice a moderate overestimate of the width is sufficient since in Theorem 5 the closed interval $[\lambda_0-\Delta, \lambda_0+\Delta]$ is only required to contain $\Lambda(A)$. We demonstrate the finite section method in Algorithm 1.[4]

[4]In Step 4 of Algorithm 1, we label the indices of E with $-m:m$ instead of $1:(2m+1)$. We use this labeling convention for submatrices extracted from a doubly-infinite matrix, when there is no ambiguity.

Algorithm 1 (A priori) Finite section method for $\exp(\mathrm{i}A)$

Input: A is b-banded Hermitian and bounded, $m \in \mathbb{N}$, $\tau > 0$.

Estimate the extreme points of $\Lambda(A)$ and set

$$\Delta \leftarrow \frac{\sup \Lambda(A) - \inf \Lambda(A)}{2}.$$

Find the smallest integer w such that

$$K\big(\rho^{2(w-m)-1} + \rho^{2(w+m)-1}\big) \leq \tau \quad \text{and} \quad w \geq m,$$

where K and ρ are chosen optimally from (7) and (8) with $d = 2(w-m) - b/2$.

Compute $E \leftarrow \exp\big[\mathrm{i}A_{(-w:w,-w:w)}\big]$.

Output $E_{(-m:m,-m:m)}$.

Evidently, the effectiveness of Algorithm 1 depends on the quality of the a priori estimate. We remark that sometimes the estimate based on Theorem 5 can severely overestimate the truncation error, mainly because the bound in Theorem 2 is pessimistic. An extreme case occurs when A is diagonal and has a wide spectrum. Theorem 2 still provides a very large constant K which grows exponentially with Δ while the actual decay is arbitrarily fast. We will show another example in Section 3.2. Once the a priori estimate is too pessimistic, it might not be a good idea to identify the size of the computational window based on such a bound. A remedy for this issue will be proposed in the next section.

3 The Finite Section Method for Unbounded Matrices

In this section, we discuss how to derive the finite section method for unbounded self-adjoint matrices. Since unbounded matrices are conceptionally quite different from bounded matrices, in the following we first recall some preliminaries in functional analysis. Then the decay property of two classes of Wilkinson matrices are analyzed. This analysis is used to derive the finite section method for these matrices. Finally, we investigate some extensions to a certain class of diagonally dominant banded matrices.

3.1 The exponential of unbounded matrices

Unlike the bounded case for doubly-infinite matrices, unbounded Hermitian matrices do not necessarily represent self-adjoint operators on $l^2(\mathbb{Z})$. However, the self-adjointness is important even for defining the exponential function. Hence careful treatment is required for unbounded matrices. In the following, we recall some preliminaries in functional analysis. These results can be found in, e.g., [1, 8, 23].

In this section we only consider class of doubly-infinity banded Hermitian matrices A that can be expressed as the sum of three Hermitian matrices

$$A = D + N + R,$$

where D is diagonal and invertible, R is bounded, and $\|ND^{-1}\|_2 < 1$. We will show that A is self-adjoint, in the sense that it can be represented as a self-adjoint operator by choosing a suitable domain of definition in $l^2(\mathbb{Z})$. Let

$$\mathcal{D}_A = \left\{ x \in l^2(\mathbb{Z}) \colon \sum_{n=-\infty}^{+\infty} |d_{nn} x_n|^2 < +\infty \right\},$$

which is a dense subspace of $l^2(\mathbb{Z})$. Then A defines a symmetric operator $\phi[A] \colon \mathcal{D}_A \to l^2(\mathbb{Z})$, $x \mapsto Ax$, by the matrix-vector multiplication (1). Notice that

$$\|(A-D)x\|_2 \le \|(ND)^{-1}(Dx)\|_2 + \|Rx\|_2 \le \|(ND)^{-1}\|_2 \|Dx\|_2 + \|Rx\|_2 < \|Dx\|_2 + \|R\|_2 \|x\|_2$$

for all $x \in \mathcal{D}_A$ Then by the Kato-Rellich theorem [17, Theorem 4.4 in Chapter 5], $\phi[A]$ is essentially self-adjoint due to the fact that $\phi[D]$, which is defined on $\mathcal{D}_A$, is essentially self-adjoint [1]. By the spectral decomposition of the closure of $\phi[A]$,

$$\overline{\phi[A]} = \int_{-\infty}^{+\infty} \lambda \, \mathrm{d}P_\lambda,$$

we define $\exp\bigl(\mathrm{i}t\overline{\phi[A]}\bigr)$ as [23, Theorem VIII.6]

$$\exp\bigl(\mathrm{i}t\overline{\phi[A]}\bigr) = \int_{-\infty}^{+\infty} \exp(\mathrm{i}t\lambda) \, \mathrm{d}P_\lambda, \qquad t \in \mathbb{R},$$

where P is a projection-valued measure on $l^2(\mathbb{Z})$. Since $\exp\bigl(\mathrm{i}t\overline{\phi[A]}\bigr)$ is unitary in $l^2(\mathbb{Z})$ and hence bounded, it has a matrix representation [1] with respect to the standard basis $\{e_n\}_{n\in\mathbb{Z}}$. This matrix representation is denoted by $\exp(\mathrm{i}tA)$.

We have seen in Section 2.2 that the finite section method can be interpreted as introducing some perturbation in the matrix A. Since A is self-adjoint, by Stone's Theorem [23, Theorem VIII.8], $\{\exp(\mathrm{i}tA) \colon t \in \mathbb{R}\}$ is a strongly continuous unitary group on $l^2(\mathbb{Z})$ whose infinitesimal generator is $\mathrm{i}A$. As a consequence, (6) also holds [8, Theorem 3.2.1] for self-adjoint matrices A and B where A can be bounded or unbounded. This perturbation result plays an important role when analyzing the error of the finite section method.

3.2 Case study for Wilkinson-type W^- matrices

Our first example is the class of Wilkinson-type W^- matrices

$$W^-(\alpha) = \operatorname{Tridiag} \left\{ \begin{matrix} \cdots & \alpha & \cdots & \alpha & \alpha & \cdots & \alpha & \cdots \\ \cdots & n & \cdots & 1 & 0 & -1 & \cdots & -n & \cdots \\ \cdots & \bar{\alpha} & \cdots & \bar{\alpha} & \bar{\alpha} & \cdots & \bar{\alpha} & \cdots \end{matrix} \right\}.$$

Here we use the Tridiag notation to represent a tridiagonal matrix in terms of its three diagonals. Instead of handling the doubly-infinite matrix $W^-(\alpha)$, we start with its central $(2n+1)\times(2n+1)$ diagonal block

$$W_n^-(\alpha) = \mathrm{Tridiag}\begin{Bmatrix} \alpha & \cdots & \alpha & \alpha & \cdots & \alpha \\ n & \cdots & 1 & 0 & -1 & \cdots & -n \\ \bar{\alpha} & \cdots & \bar{\alpha} & \bar{\alpha} & \cdots & \bar{\alpha} \end{Bmatrix}.$$

Since all eigenvalues of $W_n^-(\alpha)$ are distinct [31], the corresponding eigenvectors are unique (up to scaling). We will show that these eigenvectors are highly localized once $|\alpha| \ll n$. The following lemma is a simplified version of [22, Lemma 4.1] tailored to $W_n^-(\alpha)$. The estimate provided here is slightly better than directly using the conclusion of [22, Lemma 4.1], since the special structure of $W_n^-(\alpha)$ is taken into account.

Lemma 6. *Let $\lambda_{-n} \geq \lambda_{-n+1} \geq \cdots \geq \lambda_{n-1} \geq \lambda_n$ be eigenvalues of $W_n^-(\alpha)$, with normalized eigenvectors x_{-n}, x_{-n+1}, …, x_{n-1}, x_n (i.e., $\|x_j\|_2 = 1$). Then the entries of these eigenvectors satisfy*

$$|x_j(i)| \leq \prod_{k=0}^{|i-j|-d_0} \frac{|\alpha|}{k+d_0-3|\alpha|} \tag{9}$$

for any integer $d_0 \geq 4|\alpha|$.

Proof. See Appendix A. □

Lemma 6 demonstrates that the entries $x_j(i)$ decay super-exponentially with respect to $|i-j|$. A more general conclusion for block tridiagonal matrices has been developed in [22], which aims at deriving eigenvalue perturbation bounds, see also [13] and [30]. By choosing $d_0 = \lceil 5|\alpha| \rceil$, we obtain a simpler (but much looser) exponential decay bound

$$|x_j(i)| \leq \min\{1, 2^{d_0-|i-j|}\}. \tag{10}$$

A notable observation is that neither (9) nor (10) depend on the size of $W_n^-(\alpha)$, apart from the fact that these bounds are not useful when $n < 2|\alpha|$. Now from the spectral decomposition

$$W_n^-(\alpha) = \sum_{k=-n}^{n} \lambda_k x_k x_k^*,$$

we immediately obtain that

$$\exp[\mathrm{i}\beta W_n^-(\alpha)] = \sum_{k=-n}^{n} \exp(\mathrm{i}\beta\lambda_k) x_k x_k^*$$

and hence $|\exp[\mathrm{i}\beta W_n^-(\alpha)]| \le |X_n||X_n|^*$ where $X_n = [x_{-n}, \dots, x_n]$ and $\beta \in \mathbb{R}$.[5] Because the product of two doubly-infinite matrices with exponentially decayed off-diagonals also has the exponential decay property (see, e.g., [19]), we conclude that $\exp[\mathrm{i}\beta W_n^-(\alpha)]$ has the exponential decay property. Lemmas 7 and 8 below give quantitative estimates of the decay.

Lemma 7. *Suppose two doubly-infinite matrices X and Y both have the exponential decay property, i.e.,*

$$|x_{ij}| \le K_X \rho_X^{|i-j|}, \quad |y_{ij}| \le K_Y \rho_Y^{|i-j|}, \qquad \forall i,\ j.$$

Then their product XY satisfies

$$|(XY)_{ij}| \le K_X K_Y \Big(\frac{2}{1-\rho_0^2} + |i-j| - 1 \Big) \rho_0^{|i-j|}, \tag{11}$$

where $\rho_0 = \max\{\rho_X, \rho_Y\}$.

Proof. The entries in the product XY can be bounded by

$$|(XY)_{ij}| \le \sum_{k\in\mathbb{Z}} |x_{ik} y_{kj}| \le K_X K_Y \sum_{k\in\mathbb{Z}} \rho_0^{|i-k|+|k-j|}.$$

Without loss of generality, we consider the case $i \le j$. Notice that

$$|i-k| + |k-j| = \begin{cases} |i-j|, & \text{if } i \le k \le j, \\ |i-j| + 2\min\{|i-k|, |j-k|\}, & \text{if } k < i \text{ or } k > j. \end{cases}$$

Therefore, we obtain

$$\begin{aligned} \sum_{k\in\mathbb{Z}} \rho_0^{|i-k|+|k-j|} &= \sum_{k=-\infty}^{i-1} \rho_0^{|i-j|+2|i-k|} + \sum_{k=i}^{j} \rho_0^{|i-j|} + \sum_{k=j+1}^{+\infty} \rho_0^{|i-j|+2|j-k|} \\ &= \Big(\frac{2}{1-\rho_0^2} + |i-j| - 1 \Big) \rho_0^{|i-j|}. \end{aligned}$$

□

Lemma 8. *Suppose two doubly-infinite matrices X and Y both have the exponential decay property of the form*

$$|x_{ij}| \le \min\left\{1, \rho^{|i-j|-d_0}\right\}, \quad |y_{ij}| \le \min\left\{1, \rho^{|i-j|-d_0}\right\}.$$

Then the product XY can be bounded by

$$|(XY)_{ij}| \le \begin{cases} \Big(|i-j| - 2d_0 - 1 + \dfrac{2}{1-\rho} \Big) \rho^{|i-j|-2d_0}, & \text{if } |i-j| \ge 2d_0, \\ 2d_0 - |i-j| - 1 + \dfrac{2}{1-\rho}, & \text{if } |i-j| < 2d_0. \end{cases} \tag{12}$$

[5] The inequality between matrices is understood entrywise.

Proof. Without loss of generality, we only need to prove the conclusion for $i \le j$. When $i \le j - 2d_0$, we have

$$\begin{aligned} |(XY)_{ij}| &\le \sum_{k=-\infty}^{i+d_0} |y_{kj}| + \sum_{k=i+d_0+1}^{j-d_0-1} |x_{ik}y_{kj}| + \sum_{k=j-d_0}^{+\infty} |x_{ik}| \\ &\le \Big(|i-j| - 2d_0 - 1 + \frac{2}{1-\rho}\Big)\rho^{|i-j|-2d_0}. \end{aligned}$$

When $j - 2d_0 < i \le j$, we use

$$\begin{aligned} |(XY)_{ij}| &\le \sum_{k=-\infty}^{j-d_0-1} |y_{kj}| + \sum_{k=j-d_0}^{i+d_0} |x_{ik}y_{kj}| + \sum_{k=i+d_0+1}^{+\infty} |x_{ik}| \\ &\le 2d_0 - |i-j| - 1 + \frac{2}{1-\rho} \end{aligned}$$

to obtain the conclusion. □

The estimates (11) and (12) can certainly be applied to finite matrices and yield decay bounds for $\exp\big[\mathrm{i}\beta W_n^-(\alpha)\big]$. To obtain an easily computable decay bound, we set

$$d_0 = \lceil 6|\alpha| \rceil, \qquad \rho = \frac{|\alpha|}{d_0 - 3|\alpha|},$$

and conclude from Lemma 8 that

$$\begin{aligned} \Big|\big(\exp[\mathrm{i}\beta W_n^-(\alpha)]\big)_{ij}\Big| &\le \Big(|i-j| - 2d_0 - 1 + \frac{2}{1-\rho}\Big)\rho^{|i-j|-2d_0} \\ &\le \big(|i-j| - 2d_0 + 2\big)\rho^{|i-j|-2d_0} \end{aligned}$$

when $|i-j| \ge 2d_0$, based on the fact that $\rho \le 1/3$. Then we consider the function

$$f(x) = (x+2)\Big(\frac{\mathrm{e}}{3}\Big)^x, \qquad (x \ge 0).$$

It can be easily verified that $\max_{x\ge 0} f(x) < 5$. Then applying the inequality

$$\big(|i-j| - 2d_0 + 2\big)\Big(\frac{1}{3}\Big)^{|i-j|-2d_0} \le 5\exp\big(-|i-j| + 2d_0\big), \tag{13}$$

the decay bound (12) on $\exp\big[\mathrm{i}\beta W_n^-(\alpha)\big]$ simplifies to

$$\Big|\big(\exp[\mathrm{i}\beta W_n^-(\alpha)]\big)_{ij}\Big| \le 5\exp\big(2\lceil 6|\alpha|\rceil - |i-j|\big) \le 5\exp\big(12\lceil |\alpha|\rceil - |i-j|\big)$$

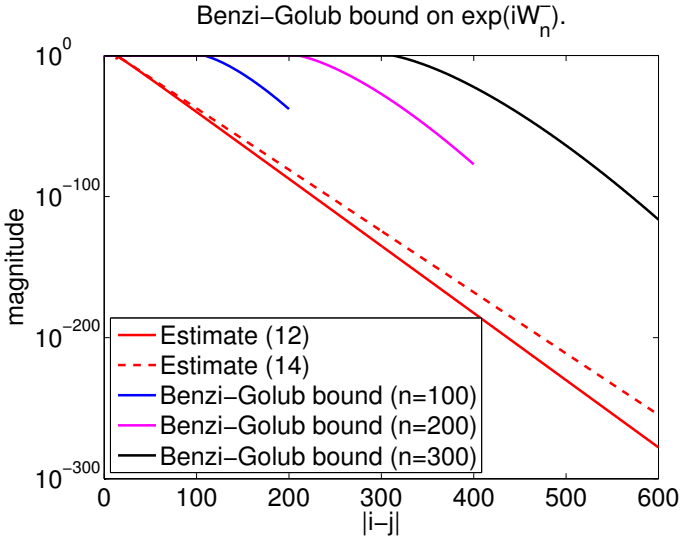


Figure 2: The Benzi-Golub bound (4) with optimally chosen χ deteriorates as n increases. Our estimates (12) and (14) are also provided for reference.

for $|i-j| \geq 2d_0$. Notice that $\exp\big(12\lceil|\alpha|\rceil - |i-j|\big) > 1$ when $|i-j| < 2d_0 < 12\lceil|\alpha|\rceil$. Taking into account that $\|\exp[\mathrm{i}\beta W_n^-(\alpha)]\|_2 = 1$, we combine the two cases into

$$\left|\big(\exp[\mathrm{i}\beta W_n^-(\alpha)]\big)_{ij}\right| \leq \min\left\{1, 5\exp\big(12\lceil|\alpha|\rceil - |i-j|\big)\right\}. \tag{14}$$

An important observation is that both (12) and (14) provide estimates independent of n. Finally, we remark that Theorem 2 can also be applied to $W_n^-(\alpha)$ for any given n. However, since $\Delta = \Theta(n)$, the estimate deteriorates as n increases, see Figure 2.

3.3 Case study for Wilkinson-type W^+ matrices

Let us consider another Wilkinson-type matrix

$$W^+(\alpha) = \mathrm{Tridiag}\left\{\begin{array}{ccccccccccc} \cdots & \alpha & \cdots & \alpha & \alpha & \cdots & \alpha & \cdots \\ \cdots & n & \cdots & 1 & 0 & 1 & \cdots & n & \cdots \\ \cdots & \bar{\alpha} & \cdots & \bar{\alpha} & \bar{\alpha} & \cdots & \bar{\alpha} & \cdots \end{array}\right\}$$

and its finite diagonal block

$$W_n^+(\alpha) = \mathrm{Tridiag}\left\{\begin{array}{ccccccc} \alpha & \cdots & \alpha & \alpha & \cdots & \alpha \\ n & \cdots & 1 & 0 & 1 & \cdots & n \\ \bar{\alpha} & \cdots & \bar{\alpha} & \bar{\alpha} & \cdots & \bar{\alpha} \end{array}\right\}.$$

Such a matrix has also been considered in [22, 30]. The spectral decomposition of $W_n^+(\alpha)$ can be constructed from the spectral decompositions of two smaller tridiagonal matrices, see [31] for details. Unlike the matrix $W_n^-(\alpha)$, the eigenvectors of $W_n^+(\alpha)$ do not have the decay property (2) when $\alpha \neq 0$, no matter how we order the eigenvectors [31]. However, it is still possible to establish a *bimodal* decay. To explain this, let the eigenvalues of $W_n^+(\alpha)$ be in the order $\lambda_{-n}, \ldots, \lambda_n$, where

$$\lambda_n \geq \lambda_{-n} \geq \lambda_{n-1} \geq \lambda_{-n+1} \geq \cdots \geq \lambda_1 \geq \lambda_{-1} \geq \lambda_0.$$

Then the corresponding normalized eigenvector matrix $X_n = [x_{-n}, \ldots, x_n]$ satisfies

$$|x_j(i)| \leq 2 \prod_{k=0}^{\big||i|-|j|\big|-d_0} \frac{|\alpha|}{k+d_0-3\alpha}.$$

This can be shown using the same techniques as in Lemma 6, see [26] for detailed derivation. The decay rate here is also independent of n. We can simplify this bound to an exponential decay bound of the form

$$|x_j(i)| \leq K \max\{\rho^{|i-j|}, \rho^{|i+j|}\} \leq K\big(\rho^{|i-j|} + \rho^{|i+j|}\big),$$

see Figure 3 as an illustration. The entries of X_n decay along its diagonal as well as along its anti-diagonal. We call this kind of decay property *bimodal exponential decay*. In contrast, we call the decay property (2) *unimodal exponential decay*.

To obtain the decay property of $\exp[\mathrm{i}\beta W_n^+(\alpha)]$, we provide two ways to make use of the existing results on the unimodal exponential decay. The first approach uses a *flipping trick*. Let $\Pi = [e_n, \ldots, e_{-n}]$. Then X_n can be split as $X_n = Y + \Pi Z$ where both Y and Z has the unimodal exponential decay property. Then we obtain

$$\big|\exp\big[\mathrm{i}\beta W_n^+(\alpha)\big]\big| \leq |X_n||X_n|^T \leq \big(|Y||Y|^T + \Pi|Z||Z|^T\Pi\big) + \big(\Pi|Z||Y|^T + |Y||Z|\Pi\big).$$

Applying Lemma 8, the corresponding matrix exponential admits a bimodal exponential decay bound

$$\big|[\exp[\mathrm{i}\beta W_n^+(\alpha)]]_{ij}\big| \leq \tilde{K}\big(\tilde{\rho}^{|i-j|} + \tilde{\rho}^{|i+j|}\big),$$

We will see in Section 3.4 that finite section methods based on bimodal decay can also be established, which naturally covers the unimodal diagonal decay property (2). Another approach uses a *shuffling trick* proposed in [25]. Let $\tilde{\Pi} = [e_0, e_{-1}, e_1, \ldots, e_{-n}, e_n]$. Notice that $\tilde{X} = \tilde{\Pi}^T X_n \tilde{\Pi}$ has the unimodal decay property. Then we conclude from

$$\big|\exp\big[\mathrm{i}\beta W_n^+(\alpha)\big]\big| \leq \tilde{\Pi}\big(|\tilde{X}||\tilde{X}|^T\big)\tilde{\Pi}^T$$

that $\exp[\mathrm{i}\beta W_n^+(\alpha)]$ has the bimodal exponential decay property. For $W_n^+(\alpha)$, the shuffling trick is not the first choice. It provides looser estimates compared to the flipping trick as the decay rate of $\tilde{X}$ is worse than that of Y or Z. But it would be useful when only the knowledge of the asymptotic decay is required.

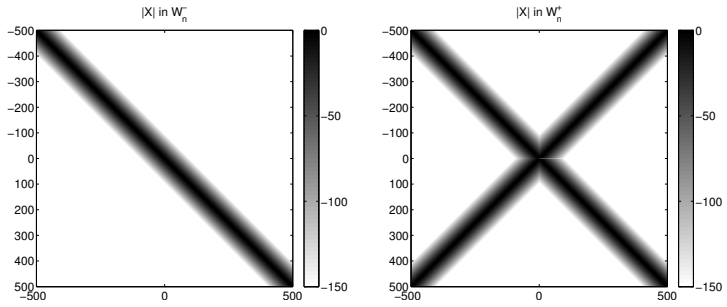


Figure 3: Localized eigenvectors of W_n^- and W_n^+ (for $n = 500$, $\alpha = 1$).

3.4 The finite section method

In Sections 3.2 and 3.3, we derived decay properties of two classes of finite Wilkinson-type matrices. Now we show that this kind of decay property is sufficient to guarantee the accuracy of the finite section method. Since the decay bounds are only available for finite matrices, we require a slightly different approach compared to the one in Section 2.2.

Consider two dynamical systems

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} = \mathrm{i}\begin{bmatrix} A_{11} & A_{12} & \\ A_{21} & A_{22} & A_{23} \\ & A_{32} & A_{33} \end{bmatrix}\begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \tilde{E}_{11} & 0 & 0 \\ 0 & \tilde{E}_{22} & 0 \\ 0 & 0 & \tilde{E}_{33} \end{bmatrix} = \mathrm{i}\begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix}\begin{bmatrix} \tilde{E}_{11} & 0 & 0 \\ 0 & \tilde{E}_{22} & 0 \\ 0 & 0 & \tilde{E}_{33} \end{bmatrix}$$

where E_{22} and $\tilde{E}_{22}$ are the central $(2w+1) \times (2w+1)$ diagonal blocks we apply the projection. Now we assume that the tridiagonal Hermitian matrix A has the bimodal exponential decay property in the truncated exponentials:

$$|[\tilde{E}_{22}(s)]_{ij}| \le K\big(\rho^{|i-j|} + \rho^{|i+j|}\big),$$

where the constants K and ρ are independent of w and $s \in [0,1]$. We have already seen from the previous two subsections that this assumption holds for $A = \beta W^{\pm}(\alpha)$. From (6), we obtain

$$\begin{aligned} &\left[E_{22} - \tilde{E}_{22}\right](1) \\ &= \mathrm{i}\int_0^1 \begin{bmatrix} 0 & I & 0 \end{bmatrix} \exp[\mathrm{i}(1-s)A] \begin{bmatrix} 0 & A_{12} & 0 \\ A_{21} & 0 & A_{23} \\ 0 & A_{32} & 0 \end{bmatrix} \begin{bmatrix} \tilde{E}_{11}(s) & 0 & 0 \\ 0 & \tilde{E}_{22}(s) & 0 \\ 0 & 0 & \tilde{E}_{33}(s) \end{bmatrix} \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} \mathrm{d}s \end{aligned}$$

$$= \mathrm{i}\int_0^1 \big[E_{21}(1-s)A_{12} + E_{23}(1-s)A_{32}\big]\tilde{E}_{22}(s)\,\mathrm{d}s. \tag{15}$$

Notice that $E_{21}(1-s)A_{12}$ has nonzero entries only in its first column and there exists an upper bound $\|E_{21}(1-s)A_{12}\|_2 \le \|A_{12}\|_2$. Using the bimodal exponential decay property of $\tilde{E}_{22}(s)$, we obtain that

$$\big\|[E_{21}(1-s)A_{12}\tilde{E}_{22}(s)]_{(:,j)}\big\|_2 \le \|A_{12}\|_2 \cdot K\big(\rho^{|j+w|} + \rho^{|j-w|}\big), \qquad j = -w, \ldots, w,$$

i.e., only the first and last several columns of this matrix can have nonnegligible entries. Similar property holds for $E_{23}(1-s)A_{32}\tilde{E}_{22}(s)$. Therefore the columns $(-m:m)$ in $E_{22}(1)$ and $\tilde{E}_{22}(1)$ agree with each other with accuracy at least $(\|A_{12}\|_2 + \|A_{32}\|_2)\cdot K\cdot\rho^{w-m}$. Certainly the $(2m+1)\times(2m+1)$ desired window is contained in this region. Therefore, by choosing a $(2w+1)\times(2w+1)$ computational window satisfying

$$\big(\|A_{12}\|_2 + \|A_{32}\|_2\big)\cdot K\big(\rho^{w-m} + \rho^{w+m}\big) \le \tau, \tag{16}$$

we ensure that $[\tilde{E}_{22}(1)]_{(-m:m,-m:m)}$ approximates the desired block in $E_{22}(1)$ with accuracy τ. Therefore, we can use (16) to find a suitable w a priori and obtain a finite section method similar to Algorithm 1.

Sometimes a priori estimates on the decay can be too pessimistic (e.g., if we apply the Benzi-Golub bound to $\exp[\mathrm{i}W_n^-(\alpha)]$). In some case we might even not have any concrete estimates despite the fact that we have the knowledge of asymptotic decay. Then algorithms such as Algorithm 1 become inappropriate. As a remedy of this issue, we propose a repeated-doubling approach based on the a posteriori error estimate using (15), as shown in Algorithm 2. In this algorithm no a priori knowledge about the detailed decay rate is required. The computational window at most doubles the smallest one that fulfills the accuracy requirement. Similar techniques on stopping criteria can be found in [11, 18].

Finally, we return to the problem left in the previous subsections—the decay property of doubly-infinite matrices $\exp[\mathrm{i}\beta W^{\pm}(\alpha)]$. Interestingly, this property can be derived as a consequence of the finite section method. For instance, let us consider $W^+(\alpha)$. To estimate the magnitude of the (i,j)-entry of $\exp[\mathrm{i}\beta W^+(\alpha)]$, we set $\tau = \epsilon K\big(\rho^{|i-j|} + \rho^{|i+j|}\big)$ and $m = \max\{|i|,|j|\}$, where ϵ can be any positive number. Using the finite section method above, we are able to find a suitable $(2w+1)\times(2w+1)$ window such that

$$\begin{aligned}\big|[\exp[\mathrm{i}\beta W^+(\alpha)]]_{ij}\big| &\le \tau + \big|[\exp[\mathrm{i}\beta W_w^+(\alpha)]]_{ij}\big| \\ &\le \tau + K\big(\rho^{|i-j|} + \rho^{|i+j|}\big) \\ &= (1+\epsilon)K\big(\rho^{|i-j|} + \rho^{|i+j|}\big).\end{aligned}$$

Since K and ρ can be chosen independent of w, letting $\epsilon \to 0+$, we conclude that the doubly-infinite matrix $\exp[\mathrm{i}\beta W^+(\alpha)]$ admits the same decay bound as the finite matrices $\exp[\mathrm{i}\beta W_w^+(\alpha)]$. Similar conclusions hold for $\exp[\mathrm{i}\beta W^-(\alpha)]$. Using the estimates (9) and (14), we obtain the following theorem (see [26] for a detailed proof).

Algorithm 2 (A posteriori) Finite section method for $\exp(\mathrm{i}A)$

Input: A is tridiagonal and Hermitian, $m \in \mathbb{N}$, $\tau > 0$.
Additionally, $\exp[\mathrm{i}A_{(-n:n,-n:n)}]$ is known to have the bimodal exponential decay property for all n.

1: Let $k \leftarrow 0$, $w^{(0)} \leftarrow 2m$.
2: Compute $E \leftarrow \exp\big[\mathrm{i}A_{(-w^{(k)}:w^{(k)},-w^{(k)}:w^{(k)})}\big]$.
3: Let $T \leftarrow \big|A_{(-w^{(k)}-1,-w^{(k)})}\big| \cdot \big\|E_{(-w^{(k)},-m:m)}\big\|_1 + \big|A_{(w^{(k)}+1,w^{(k)})}\big| \cdot \big\|E_{(w^{(k)},-m:m)}\big\|_1$.
4: **if** $T < \tau$ **then**
5: Output $E_{(-m:m,-m:m)}$.
6: **else**
7: Let $w^{(k+1)} \leftarrow 2w^{(k)}$, $k \leftarrow k+1$.
8: Go to step 2.
9: **end if**

Theorem 9. *For any real number β, the doubly-infinite matrices $\exp[\mathrm{i}\beta W^{\pm}(\alpha)]$ have the bimodal exponential decay property. Moreover, we have estimates*

$$\begin{aligned}
\big|[\exp[\mathrm{i}\beta W^{+}(\alpha)]]_{ij}\big| &\le \min\big\{1, 40\exp\big(12\lceil|\alpha|\rceil - \min\{|i-j|, |i+j|\}\big)\big\}, \\
\big|[\exp[\mathrm{i}\beta W^{-}(\alpha)]]_{ij}\big| &\le \min\big\{1, 5\exp\big(12\lceil|\alpha|\rceil - |i-j|\big)\big\},
\end{aligned}$$

for all i, $j \in \mathbb{Z}$.

3.5 More general unbounded matrices

We have seen that the eigenvector decay bounds, as established in Lemmas 6 and 7 play important roles in the derivation of the exponential decay property as well as finite section methods for Wilkinson-type matrices. In the following we extend our analyses to a more general class of unbounded matrices and establish finite section methods. We only consider the setting explained in Section 3.1. Additional requirements on the matrices will be discussed below.

To generalize the technique in Lemma 6, estimates on the eigenvalues in terms of diagonal entries are required. For any matrix A, finite or infinite, we define the *dominance factors* at its kth row as

$$\mu_k = \begin{cases} \frac{1}{|a_{kk}|}\sum_{j\neq k}|a_{kj}|, & \text{if } a_{kk} \neq 0, \\ +\infty, & \text{if } a_{kk} = 0. \end{cases}$$

The Gershgorin circle theorem [29] on a finite Hermitian matrix A states that

$$\Lambda(A) \subset \bigcup_k \big[a_{kk} - \mu_k|a_{kk}|, a_{kk} + \mu_k|a_{kk}|\big].$$

But even if A is diagonally dominant, in general we cannot further ensure that there exists an ordering of the eigenvalues λ_k of A satisfying

$$1 - \mu_k \le \frac{\lambda_k}{a_{kk}} \le 1 + \mu_k, \qquad \forall k, \tag{17}$$

when the Gershgorin disks are not separated. For instance,

$$A = \begin{bmatrix} 25 & 1 & 16 \\ 1 & 24 & 8 \\ 16 & 8 & 26 \end{bmatrix}$$

is such a counterexample to (17). To resolve this issue and establish a valid rowwise estimate similar to (17), we introduce the following concept.

Definition 2. *Let A be a strictly diagonally dominant matrix with dominance factors $\{\mu_k\}$. Then A is called* strong diagonally dominant *if there exists a set of numbers $\{\hat{\mu}_k\}$ satisfying $\mu_k \le \hat{\mu}_k < 1$ $(\forall k)$ and*

$$\begin{cases} (a_{ii} - a_{jj})\big[(a_{ii} - |\hat{\mu}_i a_{ii}|) - (a_{jj} - |\hat{\mu}_j a_{jj}|)\big] \ge 0, \\ (a_{ii} - a_{jj})\big[(a_{ii} + |\hat{\mu}_i a_{ii}|) - (a_{jj} + |\hat{\mu}_j a_{jj}|)\big] \ge 0, \end{cases} \qquad \forall i,\ j. \tag{18}$$

The numbers $\hat{\mu}_k$'s are called strong dominance factors *of A. The set*

$$\{z \in \mathbb{C} \colon |z - a_{kk}| \le |\hat{\mu}_k a_{kk}|\}$$

is called an extended Gershgorin disk *with respect to $\hat{\mu}_k$.*

The condition (18) has a geometrical interpretation—the leftmost/rightmost points of these extended Gershgorin disks follow the same order as their centers. This condition is used to limit the growth of off-diagonals compared to the diagonals A. With this new concept, we derive the following lemma, which provides a rowwise estimate similar to [2, Proposition 2].

Lemma 10. *Let A be an $N \times N$ diagonally dominant Hermitian matrix with $\hat{\mu}_k$ $(k = 1, \ldots, N)$ being its strong dominance factors. Then the ith smallest diagonal entry d_i and the ith smallest eigenvalue λ_i are related by*

$$1 - \hat{\mu}_i \le \frac{\lambda_i}{d_i} \le 1 + \hat{\mu}_i. \tag{19}$$

Proof. Without loss of generality, we assume that the diagonal entries of A are in increasing order, i.e., $d_i = a_{ii}$ for all i. By the Cauchy interlacing theorem, λ_i never exceeds the largest eigenvalue of $A_{(1:i,1:i)}$. Let μ_i be A's dominance factor at ith row. If $d_i < 0$, then by the Gershgorin circle theorem we have

$$\lambda_i \le \max_{1 \le j \le i} (1 - \mu_j) d_j \le \max_{1 \le j \le i} (1 - \hat{\mu}_j) d_j = (1 - \hat{\mu}_i) d_i.$$

If $d_i > 0$, we notice that Gershgorin disks centered in the left half plane never produce positive eigenvalues. Hence we have

$$\lambda_i \le \max_{\substack{j\le i\\ d_j>0}}(1+\mu_j)d_j \le \max_{\substack{j\le i\\ d_j>0}}(1+\hat{\mu}_j)d_j = (1+\hat{\mu}_i)d_i.$$

The two complementary estimates can be obtained by applying the same analysis to $-A$. □

Lemma 10 provides nice rowwise estimates for eigenvalues of strong diagonally dominant Hermitian matrices even when the Gershgorin disks overlap. Evidently, all finite diagonally dominant matrices are trivially strong diagonally dominant since we can choose $\hat{\mu}_k = (1+\max_k \mu_k)/2$, which is an upper bound independent of k. However, in many cases at least some $\hat{\mu}_k$ (e.g., which corresponds to an isolated Gershgorin disk) can be chosen not far larger than μ_k. In this case (19) becomes nearly as powerful as (17). With the help of this rowwise estimate, we now extend our analysis for Wilkinson-type matrices to more general cases. The following theorem, akin to Lemma 6, illustrates the decay in eigenvectors for nearly diagonally dominant matrices.

Theorem 11. *Suppose $A = \tilde{A} + R$ is an $N \times N$ Hermitian matrix where $\tilde{A}$ and R are both tridiagonal, and in addition, $\tilde{A}$ is diagonally dominant. μ_k and $\hat{\mu}_k$ ($k = 1, \ldots, N$) are $\tilde{A}$'s dominance factors and strong dominance factors, respectively. Let $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ be eigenvalues of A, with normalized eigenvectors $x_1, x_2, \ldots, x_N$. If the diagonal of $\tilde{A}$ is in increasing order, then the entries of x_j satisfy*

$$|x_j(i)| \le \begin{cases} \displaystyle\prod_{\substack{k=i\\ i}}^{k_0} \frac{\mu_k|\tilde{a}_{kk}| + \|R\|_2}{|\tilde{a}_{kk}-\tilde{a}_{jj}| - \mu_k|\tilde{a}_{kk}| - \hat{\mu}_j|\tilde{a}_{jj}| - 3\|R\|_2}, & (j \ge i), \\ \displaystyle\prod_{k=k_0}^{i} \frac{\mu_k|\tilde{a}_{kk}| + \|R\|_2}{|\tilde{a}_{kk}-\tilde{a}_{jj}| - \mu_k|\tilde{a}_{kk}| - \hat{\mu}_j|\tilde{a}_{jj}| - 3\|R\|_2}, & (j < i), \end{cases} \tag{20}$$

where k_0 is chosen between i and j such that it maximizes $|i-k_0|$ and ensures

$$|\tilde{a}_{kk}-\tilde{a}_{jj}| > 4\|R\|_2 + 2\mu_k|\tilde{a}_{kk}| + \hat{\mu}_j|\tilde{a}_{jj}|$$

for all k between i and k_0.

Proof. By Lemma 10, eigenvalues of $\tilde{A}$ satisfy $|\tilde{\lambda}_k - \tilde{a}_{kk}| \le \hat{\mu}_k|\tilde{a}_{kk}|$. Then Weyl's theorem implies that $|\lambda_k - \tilde{a}_{kk}| \le \|R\|_2 + \hat{\mu}_k|\tilde{a}_{kk}|$. The rest of proof mimics Lemma 6. We refer to [26] for details. □

Remark 1. *The bound in* (20) *cannot provide straightaway estimate without detailed knowledge of the matrix, mainly because we do not know how small the distance $|i-k_0|$*

can be. There exist matrices (e.g., Laplacian matrices) such that (20) *only provides trivial bounds* $|x_j(i)| \le 1$. *However, there are also matrices for which the decay property of eigenvectors can be well identified using* (20). *For example, for the Wilkinson-type matrix* $W_n^-(\alpha)$ *with* $n > 2|\alpha| > 0$, *we can introduce a perturbation*[6] $R = \operatorname{Diag}\{0, R_0, 0\}$ *with*

$$R_0 = W_n^-(\alpha) - W_n^-(0) + \epsilon \cdot e_0 e_0^*, \qquad (\epsilon > 0)$$

for a sufficiently small ϵ *so that* $W_n^-(\alpha) - R$ *is diagonally dominant. By setting* $\hat{\mu}_k = 2|\alpha|/|k + \epsilon|$ *for* $(-n \le k \le n)$, (20) *can then be simplified to*

$$|x_j(i)| \le \prod_{k=0}^{|i-j|-d_0} \frac{2|\alpha|}{k + d_0 - 8|\alpha|}.$$

for $d_0 > 10|\alpha|$. *This bound is worse than* (9) *since detailed information regarding the componentwise distribution in* R *is lost by crudely using* $\|R\|_2 \le 2|\alpha|$. *Nevertheless, this estimate is still asymptotically as good as* (9).

Remark 2. *If the diagonal of* $|\tilde{A}|$ *first decreases and then increases, just like the diagonal of* $W_n^+(\alpha)$, *the estimate* (20) *needs to be slightly adjusted accordingly. Roughly speaking,* $|x_j(i)|$ *is small if* $A(p,p)$ *and* $A(q,q)$ *are well-separated for all* p *close to* i *and* q *close to* j. *This can also be obtained by applying the shuffling trick. The theorem naturally extends to banded matrices, based on block versions of Lemmas 6 and 10 [29, Chapter 6].*

Despite that Theorem 11 is a more qualitative analysis rather than a sharp quantitative one, it is evident that certain types of (finite) diagonally dominant banded Hermitian matrices, possibly with small perturbations, have localized eigenvectors. More importantly, when A is extracted from an infinite matrix, the decay bound depends only on the location (i, j), but not on the size of A. Unfortunately, without detailed information of the decay, it would be difficult to derive a decay bound for $|\exp(\mathrm{i}A)| \le |X||X|^*$ as we have done in Lemmas 7 and 8. Here we only provide an intuitive explanation about the decay in $\exp(\mathrm{i}A)$. Let $A = X\Lambda X^*$ be the spectral decomposition of A, and Y be a b-banded approximation of X with accuracy $\|X - Y\|_2 \le \tau$. Then

$$\begin{aligned}\|\exp(\mathrm{i}A) - Y\exp(\mathrm{i}\Lambda)Y^*\|_2 &= \|X\exp(\mathrm{i}\Lambda)X^* - Y\exp(\mathrm{i}\Lambda)Y^*\|_2 \\ &\le \|X - Y\|_2\|\exp(\mathrm{i}\Lambda)\|_2\|X^*\|_2 + \|Y\|_2\|\exp(\mathrm{i}\Lambda)\|_2\|X^* - Y^*\|_2 \\ &\le 2\tau(1+\tau).\end{aligned}$$

Therefore $\exp(\mathrm{i}A)$ can be well approximated by a $(2b)$-banded matrix.

As seen in Section 3.4, to derive the finite section method for a doubly-infinite matrix, we only need the knowledge of decay properties in finite diagonal blocks. Hence when

[6] We set $R_0(0,0) = \epsilon \ne 0$ to guarantee the strict diagonal dominance of $W_n^-(\alpha) - R$. But this is not crucial since the analysis in Lemma 10 will not be completely ruined by a zero row.

Theorem 11 produces nontrivial bounds for all sufficiently large diagoanl blocks of a doubly-infinite matrix A, finite section methods can be applied to A. We classify such a kind of unbounded doubly-infinite matrices as follows.

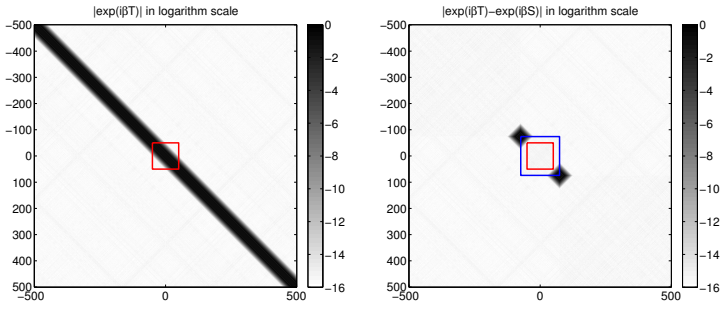
1. A is Hermitian and banded.
2. A is the sum of three Hermitian matrices $A = D + N + R$, where D is diagonal and invertible, R is bounded, and $\|ND^{-1}\|_2 < 1$.
3. $D + N$ is strong diagonally dominant; in addition, the diagonal of $D + N$ changes monotonicity at most once.
4. For each extended Gershgorin disk, its $(2\|R\|_2)$-neighborhood intersects only finitely many other extended Gershgorin disks.

Loosely speaking, the third condition indicates that the diagonal of A is nearly sorted so that we can apply a banded version of Theorem 11 to obtain the decay property for sufficiently large finite diagonal blocks of $\exp(\mathrm{i}A)$; the last condition ensures that finite sections of A have reasonably well-separated eigenvalues so that the eigenvector matrix has a certain decay property. For example, any Wilkinson-type matrix, or more generally, any banded Hermitian matrix A whose off-diagonal part (i.e., by setting all diagonal entries of A to zero) is bounded and $|a_{ii} - a_{jj}| = \Theta\big(\big||i| - |j|\big|^t\big)$ for some $t > 0$, belongs to this class. In principle, both Algorithm 1 and Algorithm 2 can be applied to unbounded self-adjoint matrices with slight modifications in the stopping criterion. We suggest that in general Algorithm 2 is preferred unless a reasonably accurate estimate of the decay is known in a priori.

Finally, we make a remark on the decay rate. If (15) can be bounded by a bimodal exponential decay (e.g., it is the case when A has bounded off-diagonals and the bimodal decay in $\tilde{E}_{22}$ is exponential and independent of w), then the distance $d = w - m$ stays constant when the user requires a larger m. However, if the decay of (15) is slower than exponential, to keep the same accuracy requirement $w - m$ will grow as m increases. This is the major reason why exponential decay is of great interest in finite section methods.

4 Numerical Experiments

In the following, we present numerical experiments for three examples to demonstrate the accuracy of finite section methods. We use reasonably large matrices to mimic infinite matrices. All experiments have been performed in MATLAB R2012a. The exponential function is computed via spectral decomposition (i.e., $\exp(\mathrm{i}A) = \exp(\mathrm{i}X\Lambda X^*) = X\exp(\mathrm{i}\Lambda)X^*$). It has been observed that sometimes even the componentwise accuracy of $\exp(\mathrm{i}A)$ is retained when the computed unitary matrix X has the exponential decay property.



(a) (b)

Figure 4: (a) Decay property of $\exp(10\mathrm{i}\,T_{500})$. The 101×101 desired window is marked. (b) Error of the finite section method ($w = 74$). Both the desired window and the computational window are marked.

Example 1. We first consider

$$T_n = \mathrm{Tridiag}\left\{ \begin{matrix} -1 & \cdots & -1 \\ 2 & \cdots \quad \cdots & 2 \\ -1 & \cdots & -1 \end{matrix} \right\} \in \mathbb{C}^{(2n+1)\times(2n+1)}$$

which are bounded with spectrum $\Lambda(T_n) \subset [0, 4]$ for all $n \in \mathbb{N}$. By Theorem 2, for any constant β, $\exp(\mathrm{i}\beta T_n)$ has the exponential decay property. Suppose $n = 500$, $m = 50$, and $\beta = 10$, i.e., the diagonal block $[\exp(\mathrm{i}\beta T_n)]_{(-50:50,-50:50)}$ is of interest. The desired (absolute) accuracy is $\tau = 10^{-8}$. The magnitude of $\exp(\mathrm{i}\beta T_n)$ is shown in Figure 4(a).

Algorithm 1 requires $w \geq 74$ to fulfill the condition $K\rho^{2(w-m)-1} \leq \tau$, with $\rho = \chi_*^{-1}$ chosen optimally from (8). As a comparison, the smallest possible computational window size to achieve accuracy 10^{-8} is $w_* = 69$. Algorithm 2 applied to this problem terminates after the first iterate, i.e., $w = 2m = 100$, which confirms the fact that $w < 2w_*$.

To visualize the error caused by truncation, let $S_{n,w}$ be a block diagonal approximation of T_n defined by

$$\begin{cases} S_{n,w}\big(\pm w, \pm(w+1)\big) = 0, \\ S_{n,w}\big(\pm(w+1), \pm w\big) = 0, \\ S_{n,w}(i,j) = T_n(i,j), \qquad \text{otherwise.} \end{cases}$$

It is comforting to see from Figure 4(b) that the error is localized around the corners of the computational window.

Example 2. Now we consider Wilkinson-type matrices $W^-(\alpha)$. In Figure 5, the exponential decay property of a 1001×1001 matrix (with $\alpha = 8$) is illustrated. It can be seen from Figure 5(g) that although the simplified bound (14) is asymptotically worse than the best bound on $|X||X|^*$ based on (9), the difference is insignificant for entries above 10^{-16}, since

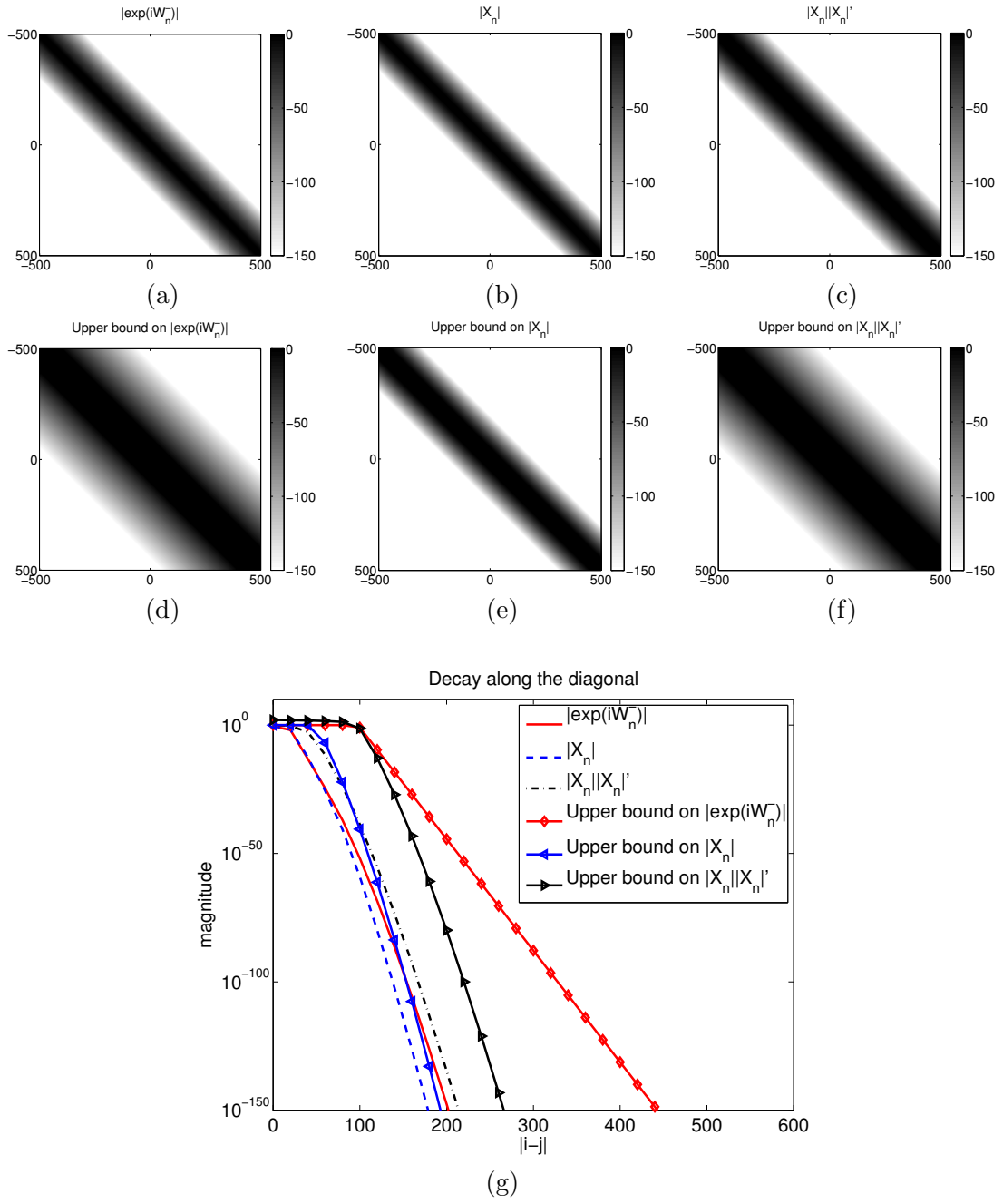


Figure 5: The exponential decay property of $\exp[\mathrm{i}W_n^-(\alpha)]$, X_n, and $|X_n||X_n|^*$ with $n = 500$ and $\alpha = 8$. The upper bounds of $\exp[\mathrm{i}W_n^-(\alpha)]$ and $|X_n|$ are given by the estimates (14) and (9), respectively, with $d_0 = 6\lceil\alpha\rceil = 48$. The upper bound on $|X_n||X_n|^*$ is obtained from the upper bound on $|X_n|$ by explicit multiplication.

Table 1: The distance between the computational window and the desired section (with accuracy $\tau = 10^{-8}$). $d = w - m$ is the a priori estimate while $d_* = w_* - m$ is the smallest distance obtained by enumeration.

	$\alpha = 1$		$\alpha = 2$		$\alpha = 4$		$\alpha = 8$	
β	d_*	d	d_*	d	d_*	d	d_*	d
1	7	33	9	45	12	70	18	119
2	9	33	12	46	18	71	27	119
4	11	34	16	47	25	71	41	120
8	12	35	17	47	26	72	43	121

the choice $d = 6\lceil|\alpha|\rceil$ produces a modest coefficient K. Another important fact is that the decay rate is independent of the matrix size, see Figure 6 for an illustration.

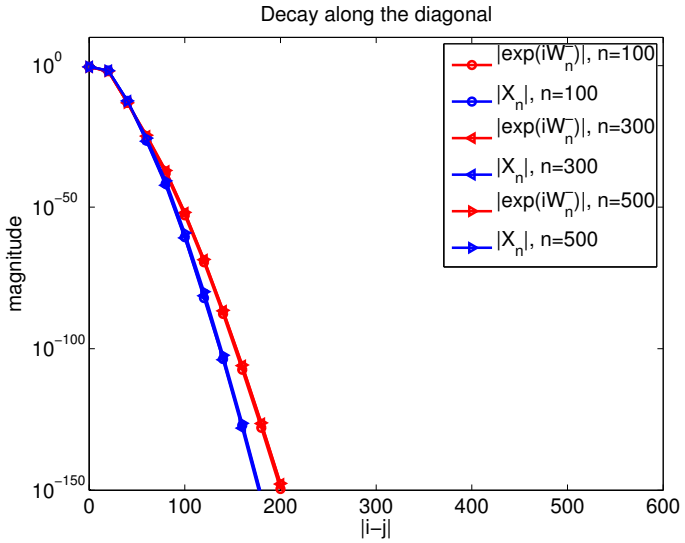


Figure 6: The decay rate is independent of the matrix size ($W_n^-(\alpha)$ for $\alpha = 8$).

Table 1 contains the distance between the computational window and the desired window. The experiments were performed with different matrix sizes ($n = 100, 200, \ldots, 500$) and different central block size ($m = 10, 20, 30$). But we only present those values for different α and β, because d is independent of m and n. Our estimates are quite conservative, but still produce computationally affordable w's. Another fact not shown in the table is that for fixed α and β, the desired $(2m + 1) \times (2m + 1)$ diagonal block extracted from matrices with different sizes agree quite well as expected.

Example 3. Our last example is another doubly-infinite matrix

$$A = \mathrm{Tridiag}\left\{ \begin{array}{ccccccccccc} \cdots & n^{\frac{3}{4}} & \cdots & 1 & 1 & \cdots & n^{\frac{3}{4}} & \cdots \\ \cdots & n & \cdots & 1 & 0 & 1 & \cdots & n & \cdots \\ \cdots & n^{\frac{3}{4}} & \cdots & 1 & 1 & \cdots & n^{\frac{3}{4}} & \cdots \end{array} \right\}.$$

A variant of Theorem 11 indicates that $\exp(\mathrm{i}A)$ has a bimodal decay property, while the decay is slower than the exponential decay. Suppose we would like to extract the central 101×101 diagonal block (i.e., $m = 50$) with absolute accuracy $\tau = 10^{-8}$. Algorithm 2 applied to this problem terminates at $w = 4m = 200$. Plots of X, $|X||X|^*$, $\exp(\mathrm{i}A)$ as well as the error are shown in Figure 7. Although the a priori estimate based on decay in eigenvectors is too pessimistic, Algorithm 2 still handles this difficult example quite well.

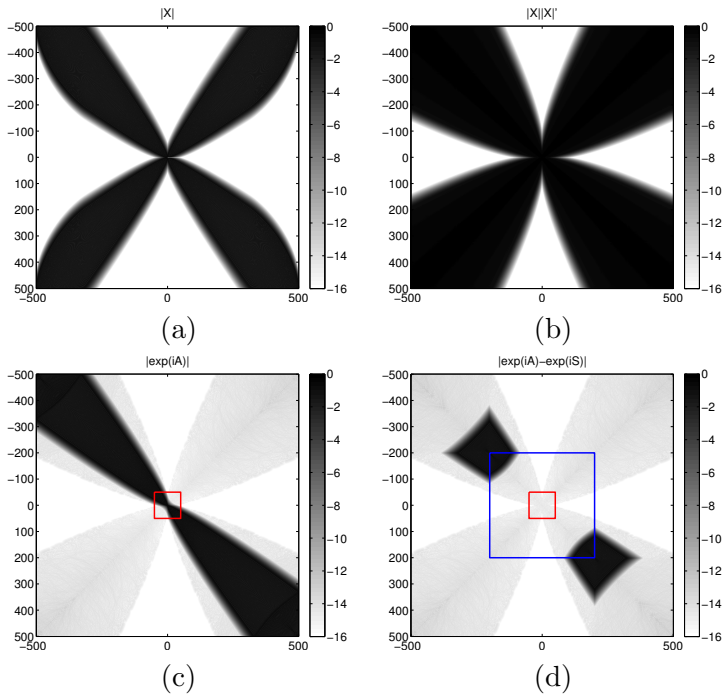


Figure 7: (a)–(c) The decay in X, $|X||X|^*$, and $\exp(\mathrm{i}A)$, respectively, in Example 3. The desired window in $\exp(\mathrm{i}A)$ is marked. (d) Error of the finite section method. Both the desired window and the computational window are marked. Here S is the block diagonal matrix by dropping the $\pm w$th sub-diagonal entries ($w = 200$) of A.

But we remark that Algorithm 2 does not work if there is no decay at all in a reasonably

computable range. For example, for

$$B = \text{Tridiag}\left\{\begin{matrix} \cdots & n^{1.9} & \cdots & 1 & 1 & \cdots & n^{1.9} & \cdots \\ \cdots & n^2 & \cdots & 1 & 0 & 1 & \cdots & n^2 & \cdots \\ \cdots & n^{1.9} & \cdots & 1 & 1 & \cdots & n^{1.9} & \cdots \end{matrix}\right\},$$

which is also self-adjoint. There is no obvious decay in a modest finite section of $\exp(\mathrm{i}B)$, see Figure 8. Hence Algorithm 2 cannot compute the finite section with $m = 50$ for this matrix unless a computational window with $w = 1600$ is affordable.

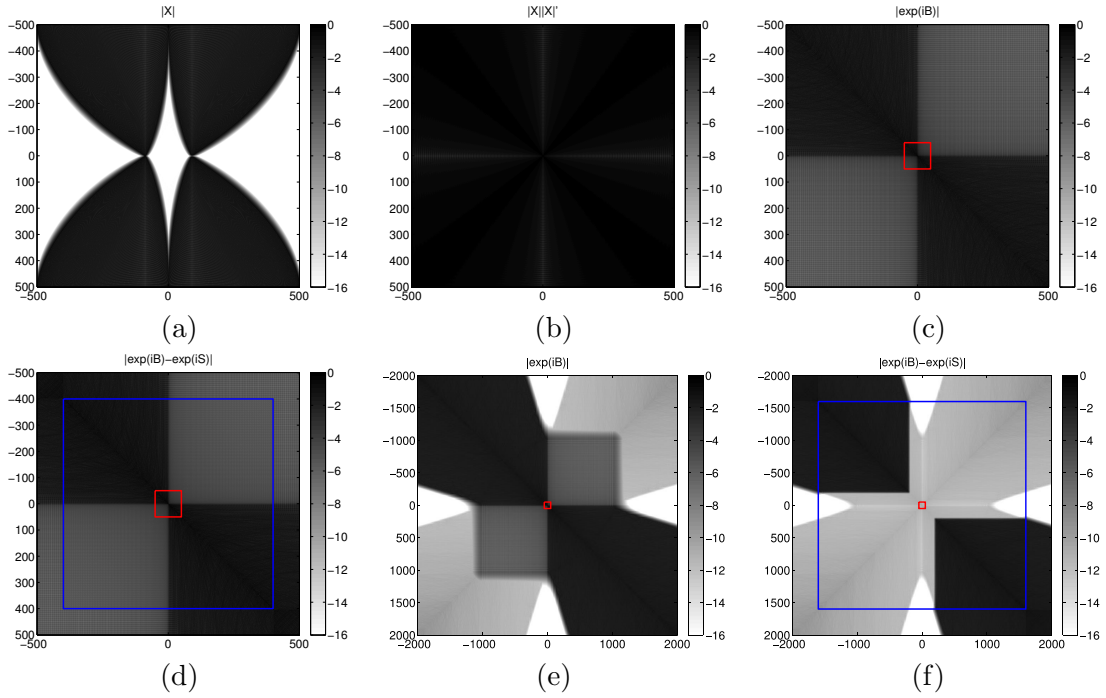


Figure 8: There is no decay in modest finite sections of X, $|X||X|^*$, and $\exp(\mathrm{i}B)$ in Example 3. The computational window is required to be very large in order to find a good approximation. The desired window and the computational window are marked.

5 Conclusions

We have shown that certain decay properties can be used to establish finite section methods for extracting a finite diagonal block of $\exp(\mathrm{i}A)$. For bounded and banded Hermitian matrices, the exponential decay property of $\exp(\mathrm{i}A)$ can be derived by polynomial approximation; for unbounded matrices, we identify localized eigenvectors in its finite diagonal blocks to obtain the decay property. We also show that the required distance to attain

a certain accuracy between the desired window and the computational window stays constant when the decay is exponential. Numerical experiments demonstrate the robustness of proposed finite section methods.

A Proof of Lemma 6

Proof. Only the nontrivial case $\alpha \neq 0$ is considered. We split $W_n^-(\alpha)$ into $W_n^-(\alpha) = W_n^-(0) + N_n(\alpha)$ where $N_n(\alpha)$ has zeros on its diagonal. Then by Weyl's theorem, we know that $\lambda_j \in (-j - 2|\alpha|, -j + 2|\alpha|)$ for all j. We rewrite $(W_n^-(\alpha) - \lambda_j I)x_j = 0$ as a set of equations:

$$\begin{aligned}
&(n - \lambda_j)x_j(-n) + \alpha x_j(-n+1) = 0 \\
&(n - 1 - \lambda_j)x_j(-n+1) + \bar{\alpha} x_j(-n) + \alpha x_j(-n+2) = 0 \\
&\cdots \\
&(-k - \lambda_j)x_j(k) + \bar{\alpha} x_j(k-1) + \alpha x_j(k+1) = 0 \\
&\cdots \\
&(-n + 1 - \lambda_j)x_j(n-1) + \bar{\alpha} x_j(n-2) + \alpha x_j(n) = 0 \\
&(-n - \lambda_j)x_j(n) + \bar{\alpha} x_j(n-1) = 0
\end{aligned} \tag{21}$$

Since the eigenvectors are normalized, the conclusion trivially holds when $|i - j| < d_0$. Hereafter we assume that $|i - j| \geq d_0$.

Suppose $i \leq j - d_0$. In the following we show by induction that for any integer k satisfying $-n \leq k \leq j - d_0$ we have

$$|x_j(k)| \leq \frac{|\alpha|}{-k + j - 3|\alpha|} |x_j(k+1)|. \tag{22}$$

The first equation in (21) yields

$$|x_j(-n)| = \frac{|\alpha|}{n - \lambda_j} |x_j(-n+1)| \leq \frac{|\alpha|}{n + j - 2|\alpha|} |x_j(-n+1)| \leq \frac{|\alpha|}{n + j - 3|\alpha|} |x_j(-n+1)|$$

because $n + j - 3|\alpha| \geq d_0 - 3|\alpha| \geq |\alpha| > 0$. Then for $-n < k \leq j - d_0$, we consider the equation

$$(-k - \lambda_j)x_j(k) = -\overline{\alpha} x_j(k-1) - \alpha x_j(k+1).$$

The induction hypothesis implies that $|x_j(k-1)| \leq |x_j(k)|$. Thus we obtain

$$|\alpha x_j(k+1)| \geq |(-k - \lambda_j)x_j(k)| - |\alpha x_j(k-1)| \geq (-k - \lambda_j - |\alpha|)|x_j(k)|.$$

The corresponding coefficient $-k - \lambda_j - |\alpha|$ is positive because

$$-k - \lambda_j - |\alpha| \geq -k + j - 3|\alpha| \geq d_0 - 3|\alpha| \geq |\alpha| > 0.$$

Hence we obtain

$$|x_j(k)| \leq \frac{|\alpha|}{-k+j-3|\alpha|}|x_j(k+1)|.$$

This finishes the proof of (22). Applying (22) repeatedly, we obtain

$$|x_j(i)| \leq |x_j(j-d_0+1)| \prod_{k=i}^{j-d_0} \frac{|\alpha|}{-k+j-3|\alpha|}.$$

Taking into account that $|x_j(j-d_0+1)| \leq 1$, we eventually arrive at

$$|x_j(i)| \leq \prod_{k=i}^{j-d_0} \frac{|\alpha|}{-k+j-3|\alpha|} = \prod_{k=0}^{j-i-d_0} \frac{|\alpha|}{k+d_0-3|\alpha|} = \prod_{k=0}^{|i-j|-d_0} \frac{|\alpha|}{k+d_0-3|\alpha|},$$

The case when $i \geq j+d_0$ can be analyzed in a similar manner by starting from the last equation in (21). □

Acknowledgement

This work was inspired by discussions with Prof. Gregory Chirikjian. The author is indebted to Prof. Daniel Kressner for many constructive comments, and to Prof. Zhaobo Huang for helpful discussions. The author is also grateful to the anonymous referee for valuable comments and suggestions.

References

[1] N. I. Akhiezer and I. M. Glazman. *Theory of Linear Operators in Hilbert Space*. Dover Publications, New York, NY, USA, 1993.

[2] J. L. Barlow and J. W. Demmel. Computing accurate eigensystems of scaled diagonally dominant matrices. *SIAM J. Numer. Anal.*, 27(3):762–791, 1990.

[3] M. Benzi and G. H. Golub. Bounds for the entries of matrix functions with applications to preconditioning. *BIT*, 39(3):417–438, 1999.

[4] M. Benzi and N. Razouk. Decay bounds and $O(n)$ algorithms for approximating functions of sparse matrices. *Electron. Trans. Numer. Anal.*, 28:16–39, 2007.

[5] P. J. Bickel and M. Lindner. Approximating the inverse of banded matrices by banded matrices with applications to probability and statistics. *Theory Probab. Appl.*, 56(1):1–20, 2012.

[6] A. Böttcher. C^*-algebra in numerical analysis. *Irish Math. Soc. Bulletin*, 45:57–133, 2000.

[7] C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Quantum Mechanics*. Wiley-VCH, Berlin, Germany, 1992.

[8] R. F. Curtain and H. J. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*. Springer-Verlag, New York, NY, USA, 1995.

[9] N. Del Buono, L. Lopez, and R. Peluso. Computation of the exponential of large sparse skew-symmetric matrices. *SIAM J. Sci. Comput.*, 27(1):278–293, 2005.

[10] S. Demko, W. F. Moss, and P. W. Smith. Decay rates for inverses of band matrices. *Math. Comp.*, 43:491–499, 1984.

[11] A. Frommer and V. Simoncini. Stopping criteria for rational matrix functions of Hermitian and symmetric matrices. *SIAM J. Sci. Comput.*, 30(3):1387–1412, 2008.

[12] I. M. Gelfand. Normierte Ringe. *Rec. Math. [Mat. Sbornik] N. S.*, 9(51):3–24, 1941.

[13] D. Gill and E. Tadmor. An $O(N^2)$ method for computing the eigensystem of $N \times N$ symmetric tridiagonal matrices by the divide and conquer approach. *SIAM J. Sci. Stat. Comput.*, 11(1):161–173, 1990.

[14] V. Grimm. Resolvent Krylov subspace approximation to operator functions. *BIT*, 52(3):639–659, 2012.

[15] R. Hagen, S. Roch, and B. Silbermann. *C^*-Algebras and Numerical Analysis*. Marcel Dekker, New York and Basel, 2001.

[16] M. Hochbruck and A. Ostermann. Exponential integrators. *Acta Numer.*, 19:209–286, 2010.

[17] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin, Germany, 1980.

[18] L. Knizhnerman and V. Simoncini. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numer. Linear Algebra Appl.*, 17(4):615–638, 2010.

[19] I. Krishtal, T. Strohmer, and T. Wertz. Localization of matrix factorizations. Technical report, 2013. Available at `http://arxiv.org/abs/1305.1618`.

[20] M. Lindner. *Infinite Matrices and Their Finite Sections: An Introduction to the Limit Operator Method*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2006. An introduction to the limit operator method.

[21] G. G. Lorentz. *Approximation of Functions*. AMS Chelsea Publishing, Providence, RI, USA, 2nd edition, 2005.

[22] Y. Nakatsukasa. Eigenvalue perturbation bounds for Hermitian block tridiagonal matrices. *Appl. Numer. Math.*, 62:67–78, 2012.

[23] M. Reed and B. Simon. *Functional Analysis*, volume I of *Methods of Modern Mathematical Physics*. Academic Press, London, UK, 1972.

[24] I. Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. Reine Angew. Math.*, 140:1–28, 1911.

[25] M. Seidel. *On some Banach Algebra Tools in Operator Theory*. PhD thesis, TU Chemnitz, 2011.

[26] M. Shao. *Dense and Structured Matrix Computations—the Parallel QR Algorithm and Matrix Exponentials*. PhD thesis, EPF Lausanne, 2013.

[27] P. N. Shivakumar and C. Ji. Upper and lower bounds for inverse elements of finite and infinite tridiagonal matrices. *Linear Algebra Appl.*, 247:297–316, 1996.

[28] C. F. Van Loan. The sensitivity of the matrix exponential. *SIAM J. Numer. Anal.*, 14(6):971–981, 1977.

[29] R. S. Varga. *Geršgorin and his circles*, volume 36 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2004.

[30] C. Vömel and B. N. Parlett. Detecting localization in an invariant subspace. *SIAM J. Sci. Comput.*, 33(6):3447–3467, 2011.

[31] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, New York, NY, USA, 1965.

Recent publications :

MATHEMATICS INSTITUTE OF COMPUTATIONAL SCIENCE AND ENGINEERING
Section of Mathematics
Ecole Polytechnique Fédérale
CH-1015 Lausanne

13.2013 P. CHEN, A. QUARTERONI:
Accurate and efficient evaluation of failure probability for partial different equations with random input data

14.2013 M. DISCACCIATI, P. GERVASIO, A. QUARTERONI:
Interface control domain decomposition (ICDD) methods for coupled diffusion and advection-diffusion problems

15.2013 D. KRESSNER, J. E. ROMAN:
Memory-efficient Arnoldi algorithms for linearizations of matrix polynomials in Chebyshev basis

16.2013 D. KRESSNER, M. MILOLOZA PANDUR, M. SHAO:
An indefinite variant of LOBPCG for definite matrix pencils

17.2013 A. ABDULLE, M. J. GROTE, C. STOHRER:
FE heterogeneous multiscale method for long time wave propagation

18.2013 A. ABDULLE, Y. BAI, G. VILMART:
An online-offline homogenization strategy to solve quasilinear two-scale problems at the cost of one-scale problems

19.2013 C.M. COLCIAGO, S. DEPARIS, A. QUARTERONI:
Comparison between reduced order models and full 3D models for fluid-structure interaction problems in haemodynamics

20.2013 D. KRESSNER, M. STEINLECHNER, B. VANDEREYCKEN:
Low-rank tensor completion by Riemannian optimization

21.2013 M. KAROW, D. KRESSNER, E. MENGI:
Nonlinear eigenvalue problems with specified eigenvalues

22.2013 T. LASSILA, A. MANZONI, A. QUARTERONI, G. ROZZA:
Model order reduction in fluid dynamics: challenges and perspectives

23.2013 M. DISCACCIATI, P. GERVASIO, A. QUARTERONI:
The interface control domain decomposition (ICDD) method for the Stokes problem

24.2013 V. LEVER, G. PORTA, L. TAMELLINI, M. RIVA:
Characterization of basin-scale systems under mechanical and geochemical compaction

25.2013 D. DEVAUD, A. MANZONI, G. ROZZA:
A combination between the reduced basis method and the ANOVA expansion: on the computation of sensitivity indices

26.2013 M. SHAO:
On the finite section method for computing exponentials of doubly-infinite skew-Hermitian matrices