MODERN MATHEMATICAL METHODS

VOLUME IV
ISSUE II

ISSN: 3023-5294

modernmathmeth.com

VOLUME IV ISSUE II

August 2026
https://modernmathmeth.com

MODERN MATHEMATICAL METHODS

MMM
MODERN MATHEMATICAL METHODS

Editor-in-Chief

Tuncer Acar
Department of Mathematics, Faculty of Science, Selçuk University, Konya, Türkiye
tunceracar@ymail.com

Managing Editors

Özlem Acar
Department of Mathematics, Faculty of Science, Selçuk University, Konya, Türkiye
acarozlem@ymail.com

Sadettin Kurşun
Department of Information Security Technology, İzmir Katip Çelebi University, İzmir, Türkiye
sadettinkursun@yahoo.com

Editorial Board

Costanza Conti
DIEF University of Florence, Italy

Danilo Costarelli
University of Perugia, Italy

Sang-Eon Han
Jeonbuk National University, Korea

Mokhtar Kirane
Khalifa University, UAE

Kunquan Lan
Toronto Metropolitan University, Canada

Hendrik Van Maldeghem
Ghent University, Belgium

Francisco Marcellán
University Carlos III de Madrid, Spain

Stefano De Marchi
University of Padova, Italy

Alain Miranville
Université de Poitiers, France

Manuel Lopez-Pellicer
Universitat Politécnica de Valencia, Spain

James Francis Peters
University of Manitoba, Canada

Michele Piconi
University of Perugia, Italy

Shahram Rezapour
Azarbaijan Shahid Madani University, Iran

Bessem Samet
King Saud University, Saudi Arabia

Wutiphol Sintunavarat
Thammasat University, Thailand

Masatomo Takahashi
Muroran Institute of Technology, Japan

Calogero Vetro
University of Palermo, Italy

Renhai Wang
Guizhou Normal University, China

Dariusz Wardowski
University of Lodz, Poland

Layout & Language Editors

Murat Bodur
Konya Technical University, Türkiye

Onur Köksal
Selçuk University, Türkiye

Editorial Staff

Eşref Erdoğan
Selçuk University, Türkiye

Metin Turgay
Selçuk University, Türkiye

Contents

MODERN MATHEMATICAL METHODS
4 (2026), No. 2, pp. 69-78
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

On the absence of an exceptional set in the main relation of Wiman-Valiron theory

ANDRIY BANDURA*, ANASTASIIA MOKHNAL, SVIATOSLAV DUBEI, AND OLEH SKASKIV

ABSTRACT. The object of investigation is a class of functions analytic in the right half-plane. The half-plane contains all points whose real part is greater than some fixed a and for all x from the interval $(a, +\infty)$ the corresponding supremum of modulus of the function F inside some vertical strip is finite. There is selected subclass consisting of those functions, for which the right-hand derivative of the supremum logarithm tends to infinity. For these functions for which there exists an auxiliary function with some local behavior. The following theorem is proved: If an analytic function belongs to the described class, then for each positive natural k the k-th order derivative equals k-th order power of the right-hand derivative of the supremum logarithm for every point.

Keywords: Analytic function, Wiman-Valiron theory, main relation, exceptional set.

2020 Mathematics Subject Classification: 30B50, 11F66.

1. INTRODUCTION

Let f be an entire function of the form

$$f(z) = f_0 + \sum_{k=1}^{+\infty} f_k z^k, \quad z \in \mathbb{C}. \tag{1.1}$$

For $r > 0$, we denote

$$M_f(r) = \max\{|f(z)|\colon |z| = r\},$$
$$\mu_f(r) = \max\{|f_k|r^k\colon k \geq 0\},$$
$$\nu_f(r) = \max\{k\colon |f_k|r^k = \mu_f(r)\}$$

the maximum modulus, maximal term and central index, respectively. The main theorem of the Wiman-Valiron theory [6, 7, 13, 23] states that for each non-constant entire function $f : \mathbb{C} \to \mathbb{C}$, for any point z, $|z| = r$, such that $|f(z)| = M_f(r)$, we have for given $n \in \mathbb{N}$

$$f^{(n)}(z) \sim \left(\frac{\nu_f(r)}{z}\right)^n f(z), \tag{1.2}$$

as $|z| = r \to +\infty$, $r \notin E$, where E is some set of finite logarithmic measure (f.l.m.), that is

$$\int_{E\cap[1,+\infty)} d\ln r < +\infty.$$

Received: 29.11.2025; Accepted: 26.02.2026; Published Online: 21.04.2026
*Corresponding author: Andriy Bandura; andriykopanytsia@gmail.com
DOI: 10.64700/mmm.94

Let us denote

$$K_f(r) = r(\ln M_f(r))'_+,$$

where $(\ln M_f(r))'_+$ is the right-hand derivative. From the theorem obtained in [14] for the entire Dirichlet series, it follows that for each entire function f

$$K_f(r) \sim \nu_f(r) \quad (r \to +\infty,\ r \notin E),$$

where $E = E(f)$ is some set of f.l.m.

Various theorems about analogues of relation (1.2) for an absolutely convergent Dirichlet series, by additional conditions on its positive monotonically increasing to infinity sequence of exponents, can be found in [4, 11, 13, 14, 18, 19]. In [11, 13, 14] for the entire Dirichlet series, for absolutely convergent in a half-plane Dirichlet series in [4, 17, 18, 19]. On the other hand, R. Gorenflo [5] found a class of entire functions for which an exceptional set is absent in relation (1.2). He proved the following statement.

Theorem 1.1 ([5]). *If $f(z) = \sum\limits_{n=0}^{+\infty} f_n z^n,\ a_n \geq 0\ (n \geq 0)$ is an entire function with perfectly regular growth, that is*

$$(\exists\, \varrho \in (0, +\infty))(\exists\, \sigma \in (0, +\infty))\colon \lim_{r \to +\infty} \frac{\ln M_f(r)}{r^{\varrho}} = \sigma,$$

then for any fixed $A, B > 0$, for each $n \in \mathbb{N},\ n < B$ one has

$$f^{(n)}(ze^{\tau}) = (1 + o(1))\Big(\frac{\nu_f(r)}{ze^{\tau}}\Big)^n e^{\nu_f(r)\tau} f(z) \text{ as } r \to +\infty \tag{1.3}$$

for arbitrary $\tau,\ |\tau| < A/\nu_f(r)$, and every point z, $|z| = r$, such that the inequality

$$|f(z)| \geq \alpha M_f(r),$$

holds with a given arbitrary $\alpha \in (0, 1]$.

Theorem 1.1 at $\tau = 0$ implies that for every entire function with non-negative Taylor coefficients of perfectly regular growth relation (1.2) holds as $r \to +\infty$ at every point $z,\ |z| = r$ such that the inequality

$$|f(z)| \geq \alpha M_f(r)$$

is true with a given $\alpha \in (0, 1]$.

R. R. London [8] obtained this result for a wider class of entire functions satisfying the following conditions: There exists a function $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+ := (0, +\infty)$ such that ϕ' is a positive and unbounded function, ϕ'' is a positive and continuous function,

$$(\exists\, \alpha, \beta \in \mathbb{R}_+)(\forall\, x \geq x_0)\colon \alpha\frac{\phi'(x)}{\phi(x)} < \frac{\phi''(x)}{\phi'(x)} < \beta\frac{\phi'(x)}{\phi(x)} \tag{1.4}$$

and

$$\ln M_f(r) \sim \phi(\ln r) \qquad (r \to +\infty). \tag{1.5}$$

In this paper, we will prove a slightly different result of this form. We essentially consider the class of entire functions which are bounded in the left half-planes

$$\Pi_b := \{z \in \mathbb{C}\colon\ \mathrm{Re}\, z < b\} \text{ for all } b \in \mathbb{R}.$$

Clearly, this class contains every function

$$F(z) = f(e^z),$$

where f is an entire function on the complex plane $\mathbb{C}$. Any absolutely convergent in the whole complex plane Dirichlet series with a sequence of positive exponents increasing to $+\infty$ represent the entire function, so it belongs to this class. The partial case of such series is the Riemann zeta function [9]. Note that a similar statement about the absence of an exceptional set was earlier proved in the case of validity of the asymptotic relation (Wiman's type theorem)

$$\begin{aligned} M(x,F) &= (1+o(1))B_F(x) \\ &= -(1+o(1))A_F(x) \end{aligned} \tag{1.6}$$

as $x \to +\infty$, where

$$B_F(x) = \sup\{\text{Re } F(z) \colon \text{Re } z = x\},$$

$$A_F(x) = \inf\{\text{Im } F(z) \colon \text{Re } z = x\}.$$

These results were presented in [3] and obtained in the class of entire functions which are bounded in the every left half-plane Π_b. For other results on the conditions under which the relation (1.6) holds as $x \to +\infty$ outside exceptional sets, see, for example, also [17, 20, 22]. In [20], this result is deduced for entire Dirichlet series with increasing to $+\infty$ of a sequence of positive exponents. A similar estimate was presented for functions represented by some positive integrals in [21]. In [17], the estimate was considered in the class of absolute convergent Dirichlet series in the left half-plane $\{z \in \mathbb{C} \colon x = \text{Re } z < 0\}$ as $x \to -0$ without some exceptional set.

2. Lemmas

Let $S(a)$, $-\infty \le a \le +\infty$, be a class of functions analytic in $\Pi(a) = \{z : a < \text{Re } z\}$ which are bounded in the vertical stripes, i.e.

$$(\forall x \in (a,+\infty)) \colon \ M(x,F) := \sup\{|F(t+iy)| : a < t \le x, y \in \mathbb{R}\} < +\infty.$$

By Maximum Modulus Principle the supremum in the vertical stripe is attained at its boundary, i.e. at the vertical line. Therefore, the boundedness in the vertical stripe is replaced by the boundedness on vertical line. Moreover, by Maximum Modulus Principle the function

$$M(x,F) = \sup\{|F(x+iy)| \colon y \in \mathbb{R}\}$$

is non-decreasing on $(a,+\infty)$, and by Hadamard Three Lines Theorem the function $\ln M(x,F)$ is convex on $(a,+\infty)$ (see [15, pp.14-16], [22, p.145, p.266]). Therefore, for all $x \in (a,+\infty)$ there exists the non-decreasing right-hand derivative

$$L(x) = L(x,F) \stackrel{def}{=} (\ln M(x,F))'_+$$

at the interval $(a,+\infty)$. Let us denote by $S_\infty(a)$ the class of the functions $F \in S(a)$ such that

$$L(x,F) \to +\infty \quad (x \to +\infty).$$

Similarly as in [3], by S_0, we denote the class of functions $F \in S_\infty(0)$ for which there exists a function $\delta(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ with the following properties

(1) $\delta(x) \nearrow +\infty \ \ (0 \le x \uparrow +\infty)$;
(2) $\frac{L(x,F)}{\delta(x)} \nearrow +\infty \ \ (0 \le x \uparrow +\infty)$;
(3) for all $x \ge x_0$ with sufficiently large x_0 one has

$$\left| L\left(x \pm \frac{\delta(x)}{L(x,F)}, F\right) - L(x,F) \right| \le \frac{L(x,F)}{\delta(x)}. \tag{2.7}$$

One should observe that the class S_0 is non-empty. Indeed, for every $p \in \mathbb{N}$ the function

$$F_p(z) = \exp\{\exp\{z^p\}\}$$

belongs to the class S_0. In paper [3], it was essentially proved Lemma 2.1 and also Lemma 2.2 which are formulated below.

Lemma 2.1. *If a function $\delta(\cdot)\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that*

(1) $\delta(x) \nearrow +\infty$;
(2) $\frac{L(x,F)}{\delta(x)} \nearrow +\infty$ $(0 \le x \uparrow +\infty)$

then condition (2.7) *and*

$$\left| L\left(x + \frac{\delta(x)}{L(x,F)}, F\right) - L(x,F) \right| \le \frac{L(x,F)}{\delta(x)} \quad (x \ge x_0) \tag{2.8}$$

are equivalent.

Lemma 2.2. *Let $F \in S_0$. Then, for all $x \ge x_0$ and every $\eta \in \mathbb{C}$, satisfying*

$$|\eta| \le \frac{\delta(x)}{L(x,F)},$$

the asymptotic relation

$$F(w+\eta) = (1+\omega(\eta))F(w)e^{\eta L(x,F)}$$

holds, where

$$|\omega(\eta)| < |\eta| c(x) \frac{L(x,F)}{\delta(x)},$$

$$c(x) := 1 + e(1+\varepsilon(x)),$$

and the point w is chosen such that $\mathrm{Re}\, w = x$ *and*

$$|F(w)| \ge \frac{M(x,F)}{1+\varepsilon(x)}$$

for a given function $\varepsilon(x)$ such that $\varepsilon(x) \to +0$ $(x \to +\infty)$.

We need also the modified Cauchy inequality. The complete proof of the following lemma can be found in the solution of Problem **236** (see, [10, Part III, Ch. 5, § 2, **236**, p.355]).

Lemma 2.3. *Let*

$$f(z) = \sum_{n=0}^{+\infty} f_n z^n$$

be an analytic function in the disk

$$\mathbb{D}_R = \{z\colon |z| < R\}, \quad R > 0.$$

If for all $z \in \mathbb{D}_R$ real part of the function f is upper bounded by some constant M, i.e.

$$\mathrm{Re}\ f(z) < M,$$

then for all $z \in \mathbb{D}_R$ one has

$$|f_n| R^n \le 2(M - \mathrm{Re}\ f_0).$$

3. Main Result

Let us prove the following main result of the present article.

Theorem 3.2. *Let $F \in S_0$. Then, for each $n \in \mathbb{N}$ the asymptotic relation*

$$F^{(n)}(w) = (1 + o(1))\, L^n(x, F)\, F(w) \qquad (x \to +\infty) \tag{3.9}$$

holds for every point w with $Re\, w = x$ such that the inequality

$$|F(w)| \geq \frac{M(x, F)}{1 + \varepsilon(x)}$$

is valid for a given positive function $\varepsilon(x)$ with $\varepsilon(x) \to +0$ $(x \to +\infty)$.

Proof. For proof we will use ideas from [4, 12, 16, 18, 19]. Denote

$$\psi(x) = \frac{\delta(x)}{L(x, F)}.$$

Let a given point w, and $x = Re\, w$ be such that condition

$$F(w) \geq \frac{M(x, F)}{1 + \varepsilon(x)} \ \ (x \geq x_0)$$

is satisfied. Then, for all $\eta \in \mathbb{C}$,

$$|\eta| < \frac{\psi(x)}{c(x)},$$

and for all $x \geq x_0$,

$$|\eta| c(x) \frac{L(x, F)}{\delta(x)} < 1,$$

from Lemma 2.2, we obtain

$$\begin{aligned} |F(w + \eta)| &\geq \big(1 - |\omega(\eta)|\big) |F(w) \mathrm{e}^{\eta L(x,F)}| \\ &\geq \left(1 - |\eta| c(x) \frac{L(x, F)}{\delta(x)}\right) |F(w) \mathrm{e}^{\eta L(x,F)}| > 0. \end{aligned}$$

Since the function

$$\frac{F^{'}(w + \tau)}{F(w + \tau)}$$

is an analytic function of the variable τ in the open disk

$$\left\{ \tau : |\tau| < \frac{\psi(x)}{c(x)} \right\}$$

for given w, we yield that the following function

$$f(\eta) := \int_0^\eta \frac{F^{'}(w + \tau)}{F(w + \tau)} d\tau - \eta L(x, F) \text{ with } f(0) = 0$$

is an analytic function in the disc

$$\left\{ \eta : |\eta| < \frac{\psi(x)}{c(x)} \right\}.$$

One should observe that

$$f^{'}(0) = \frac{F^{'}(w)}{F(w)} - L(x, F).$$

Further, let $q = q(x) \in \left(0, \frac{\psi(x)}{c(x)}\right)$ for fixed $x,\ x \geq x_0$. Then, for all η with $|\eta| \leq q$ we have

$$\begin{aligned}\operatorname{Re} f(\eta) &= \ln\left|\frac{F(w+\eta)}{F(w)}e^{-\eta L(x,F)}\right| = \ln|1+\omega(\eta)| \\ &\leq \ln(1+|\omega(\eta)|) \leq \ln\left(1+\frac{qc(x)}{\psi(x)}\right).\end{aligned}$$

Consider the function f in the disc $\mathbb{D}_q = \{\eta\colon |\eta| \leq q\}$ with $q < \frac{\psi(x)}{c(x)}$. We will apply now Lemma 2.3 and obtain

$$\begin{aligned}\left|\frac{F'(w)}{F(w)} - L(x,F)\right| &= |f'(0)| = |f_1| \\ &\leq 2\max\{\operatorname{Re} f(\eta)\colon |\eta| = q\} \\ &\leq 2\ln\left(1+\frac{qc(x)}{\psi(x)}\right) \leq \frac{2qc(x)}{\psi(x)}. \end{aligned} \tag{3.10}$$

Therefore, for all $w,\ Re\, w = x$ such that

$$F(w) \geq \frac{M(x,F)}{1+\varepsilon(x)} \quad (x \geq x_0)$$

and $L(x,F) \to +\infty$ as $x \to +\infty$, one has

$$\left|\frac{F'(w)}{F(w)}\frac{1}{L(x,F)} - 1\right| \leq \frac{2c(x)}{L(x,F)\psi(x)} = \frac{c(x)}{\delta(x)} = o(1). \tag{3.11}$$

Further, applying Lemma 2.3 to the function f at the disc $\mathbb{D}_q$ with arbitrary $q \in \left(0, \frac{\psi(x)}{c(x)}\right)$, for given arbitrary $k \in \mathbb{N}$ we obtain

$$\begin{aligned}\frac{1}{k!}\left|f^{(k)}(0)\right| &= |f_k| \leq \frac{2}{q^k}\max\{\operatorname{Re} f(\eta)\colon |\eta| = q\} \\ &\leq 2q^{-k}\ln(1+qc(x)/\psi(x)) \\ &\leq 2q^{-k+1}\frac{c(x)}{\psi(x)}.\end{aligned} \tag{3.12}$$

Continuing further, for all η with $|\eta| \leq q < \frac{\psi(x)}{c(x)}$ one has

$$\begin{aligned}F(w+\eta) &= F(w)\exp\{f(\eta)+\eta L(x,F)\} \\ &= F(w)\exp\left\{\frac{F'_A(w)}{F(w)}\eta + \sum_{k=2}^{+\infty}\frac{f^{(k)}(0)}{k!}\eta^k\right\} \\ &:= F(w)F_0(\eta),\end{aligned} \tag{3.13}$$

where $F_0(\eta)$ is an analytic function of the variable η and $F_0(0) = 1$. We put

$$F_0(\eta) = 1 + \sum_{k=1}^{+\infty} F_k\eta^k, \quad |\eta| \leq q. \tag{3.14}$$

Since for the function

$$f_1(\eta) := f(\eta) + \eta L(x,F)$$

we have

$$f_1^{(k)}(\eta) = f^{(k)}(\eta) \ \ (k \geq 2),$$

so

$$F_0(\eta) = 1 + \sum_{s=1}^{+\infty} \frac{1}{s!} \left(\frac{F'(z)}{F(z)}\eta + \sum_{k=2}^{+\infty} \frac{f^{(k)}(0)}{k!}\eta^k \right)^s$$

(3.15)
$$= 1 + \sum_{s=1}^{+\infty} \frac{1}{s!} \left(\sum_{k=1}^{+\infty} \frac{f_1^{(k)}(0)}{k!}\eta^k \right)^s .$$

Therefore, the Taylor coefficients in (3.14) are formed by sums of the following form

$$F_k = \sum_{\|\alpha\|_0 = k} B_\alpha \left(f_1'(0)\right)^{\alpha_1} \cdot \left(\frac{f_1''(0)}{2!} \right)^{\alpha_2} \cdot \ldots \cdot \left(\frac{f_1^{(k)}(0)}{k!} \right)^{\alpha_k}$$

with finite quantities of the summands, where $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{Z}_+^k$ is the multi-index, and $\|\alpha\|_0 := \sum_{j=1}^k j\alpha_j$ its weight, $B_\alpha \in \mathbb{R}_+$ is generated by finite sum of positive numbers by equating coefficients at η^k from (3.14) and (3.15). From the Taylor expansion

$$F(w+\eta) = \sum\nolimits_{k=0}^{+\infty} \frac{F^{(k)}(w)}{k!}\eta^k$$

and (3.13) it follows that

$$F^{(k)}(w) = F(w)k!F_k \ \ (k \geq 1).$$

Hence,

(3.16)
$$\frac{F^{(k)}(w)}{F(w)} \frac{1}{L^k(x,F)} - 1 = \left(\frac{F'(w)}{F(w)} \frac{1}{L(x,F)} \right)^k - 1 + \sum_{\substack{\|\alpha\|_0 = k \\ \alpha_1 < k}} B_\alpha \prod_{j=1}^k \left(\frac{f_1^{(j)}(0)}{j!} \right)^{\alpha_j} .$$

Since $c(x) = \mathcal{O}(1)$ and $\delta(x) \nearrow +\infty$ $(r \to +\infty)$, we have

$$\frac{c(x)}{\delta(x)} = o(1) \ \ (r \to +\infty).$$

So, from inequality (3.11), we have

$$\left| \left(\frac{F'(w)}{F(w)} \frac{1}{L(x,F)} \right)^k - 1 \right| = \left| \frac{F'(w)}{F(w)} \frac{1}{L(x,F)} - 1 \right| \sum_{j=0}^{k-1} \left| \frac{F'(w)}{F(w)} \frac{1}{L(x,F)} \right|^j$$

$$\leq \frac{c(x)}{\delta(x)} \sum_{j=0}^{k-1} \left(1 + \frac{c(x)}{\delta(x)} \right)^j$$

(3.17)
$$= \left(1 + \frac{c(x)}{\delta(x)} \right)^k - 1 = o(1)$$

as $x \to +\infty$. We put

$$q = \frac{1}{2} \frac{\psi(x)}{c(x)} = \frac{1}{2} \frac{\delta(x)}{c(x)L(x,F)}.$$

Using inequality (3.12), we get

$$\sum_{\substack{\|\alpha\|_0=k\\ \alpha_1<k}} B_\alpha \prod_{j=1}^{k}\left(\frac{f_1^{(j)}(0)}{j!}\right)^{\alpha_j} \le \sum_{\substack{\|\alpha\|_0=k\\ \alpha_1<k}} |B_\alpha|(qL(x,F))^{-k}\left(\frac{2qc(x)}{\psi(x)}\right)^{\|\alpha\|}$$

$$= \sum_{\substack{\|\alpha\|_0=k\\ \alpha_1<k}} |B_\alpha|\left(\frac{2c(x)}{\delta(x)}\right)^{k} = o(1) \tag{3.18}$$

as $x \to +\infty$. By applying relations (3.17) and (3.18), from equality (3.16) we obtain

$$\left|\frac{F^{(k)}(w)}{F(w)}\frac{1}{L^k(x,F)} - 1\right| = o(1)$$

as $x \to +\infty$. Thus, asymptotic relation (3.9) is proved. □

4. Some Corollaries and Open Problems

Let f be an entire transcendental function of the form (1.1). Since at $x = \ln r$

$$L(x,F) = K_f(r) \nearrow +\infty \quad (r \to +\infty)$$

for the function $F(z) = f(e^z)$, the function F belongs to the class $S_\infty(0)$. From Theorem 3.2, it follows the following corollary.

Corollary 4.1. *Let f be an entire transcendental function of the form (1.1), and $F(z) = f(e^z)$. If $F \in S_0$, then for given $n \in \mathbb{N}$ and for any point z, $|z| = r$, such that $|f(z)| = M_f(r)$, the asymptotic relation*

$$f^{(n)}(z) \sim \left(\frac{K_f(r)}{z}\right)^n f(z),$$

holds as $|z| = r \to +\infty$.

The proof of the Corollary 4.1 can be obtained from the proof of Theorem 3.2 using simple calculations and transformations.

Remark 4.1. *The condition $F \in S_0$ for the function $F(z) = f(e^z)$ can be written as the following properties: $K_f(r) \to +\infty$ $(r \to +\infty)$ and there exists a function $\delta(\cdot)\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that*

(1) $\delta(r) \nearrow +\infty \ (1 \le r \uparrow +\infty)$;
(2) $\frac{K_f(r)}{\delta(r)} \nearrow +\infty \ (1 \le r \uparrow +\infty)$;
(3) for all $r \ge r_0$ with sufficiently large r_0 one has

$$\left|K_f\left(r \pm \frac{\delta(r)}{K_f(r)}\right) - K_f(r)\right| \le \frac{K_f(r)}{\delta(r)}.$$

Question 4.1. *A natural question arises: How are the conditions (1.4), (1.5) and the condition (2.7) (also (2.8)) related to each other?*

Question 4.2. *It is well known that in the general class of entire functions it is impossible to obtain a basic relation (1.2) without exceptional sets. This fact leads to the following question: How much can the conditions of Gorenflo's Theorem or Theorem 3.2 be weakened so that the exceptional set is absent?*

Question 4.3. *What is analog of Theorem 3.2 for analytic functions in a vertical strip?*

Question 4.4. *What are analogs of relations*

$$M(x,F) = (1+o(1))B_F(x) = -(1+o(1))A_F(x) \text{ as } x \to +\infty$$

and

$$F^{(n)}(w) = (1+o(1))\, L^n(x,F)\, F(w) \qquad (\mathrm{Re}\, w = x \to +\infty)$$

for the class of entire vector-valued functions [1] *and for the class of entire slice regular functions of quaternionic variable* [2]*?*

Acknowledgments

The authors express their sincere gratitude and respect to the Armed Forces of Ukraine, who made it possible to complete the preparation of this article for publication.

References

[1] V. Baksa, A. Bandura and O. Skaskiv: *Maximum modulus of slice entire regular functions of quaternionic variable with bounded index*, ALTAY Conf. Proc. Math., **2** (1) (2025), 32–39.
[2] V. Baksa, A. Bandura and O. Skaskiv: *Growth estimates for analytic vector-valued functions in the unit ball having bounded* **L***-index in joint variables*, Constr. Math. Anal., **3** (1) (2020), 9–19.
[3] S. I. Dubei, O. B. Skaskiv and O. M. Trusevych: *On the absence of an exceptional set in the relation from Wiman's theorem*, Precarp. Bull. Shev. Sci. Soc. Number, **21** (79) (2025), 28–34.
[4] S. I. Dubei, O. B. Skaskiv: *On the main relation of the Wiman-Valiron theory and asymptotic h-density of an exceptional sets*, Precarp. Bull. Shev. Sci. Soc. Number, **19** (73) (2024), 18–23.
[5] R. Gorenflo: *Über ganze transzendente funktionen von regelmäßigem wachstum*, Math. Ann., **146** (3) (1962), 226–231.
[6] W. K. Hayman: *The local growth of power series: a survey of the Wiman–Valiron method*, Canad. Math. Bull., **17** (3) (1974), 317–358.
[7] W. K. Hayman: *Subharmonic functions: Volume 2*, Academic Press, London (1989).
[8] R. R. London: *The behaviour of certain entire functions near points of maximum modulus*, J. Lond. Math. Soc. (2), **12** (4) (1976), 485–504.
[9] M. Milgram, R. Hughes: *On a generalized moment integral containing Riemann's zeta function: Analysis and experiment*, Modern Math. Methods, **3** (1) (2025), 14–41.
[10] G. Polya, G. Szegö: *Problems and theorems in Analysis. I*, Springer-Verlag, Heidelberg (1964).
[11] T. M. Salo, O. B. Skaskiv and Ya. Z. Stasyuk: *On a central exponent of entire Dirichlet series*, Mat. Stud., **19** (1) (2003), 61–72.
[12] M. N. Sheremeta: *The Wiman-Valiron method for Dirichlet series*, Ukrainian Math. J., **30** (4) (1978), 376–383. Translated from Ukr. Mat. Zh., **30** (6) (1978), 488–497.
[13] M. N. Sheremeta: *Asymptotic properties of entire functions defined by Dirichlet series and of their derivatives*, Ukrainian Math. J., **31** (6) (1979), 558–564. Translated from Ukr. Mat. Zh., **31** (6) (1979), 723–730.
[14] M. N. Sheremeta: *On the derivative of an entire Dirichlet series*, Math. USSR Sb., **65** (1) (1990), 133–145.
[15] M. M. Sheremeta: *Entire Dirichlet series*, ISDO (1993). (In Ukrainian).
[16] O. Skaskiv, A. Bandura, T. Salo and S. Dubei: *Entire functions of several variables: analogs of Wiman's Theorem*, Axioms, **14** (3) (2025), Article ID: 216.
[17] O. B. Skaskiv, Ya. Z. Stasyuk: *On the Wiman theorem for absolutely convergent Dirichlet series*, Mat. Stud., **20** (2) (2003), 133–142.
[18] O. B. Skaskiv, Ya. Z. Stasyuk: *On derivatives of Dirichlet series*, Nauk. Visn. Chernivets'kogo Univ. Mat., **239** (2005), 107–111. (In Ukrainian).
[19] O. B. Skaskiv, Ya. Z. Stasyuk: *On equivalence of the logarithmic derivative and the central exponent of the Dirichlet series*, Mat. Stud., **33** (1) (2010), 49–55.
[20] O. B. Skaskiv: *A generalization of the little Picard theorem*, J. Math. Sci., **48** (5) (1990), 570–578. Translated from Teoriya Funktsii, Funktsional'nyi Analiz i Ikh Prilozheniya, **46** (1986), 90–100.
[21] O. Skaskiv, A. Bandura and A. Bodnarchuk: *About Borel type relation for some positive integrals*, ALTAY Conf. Proc. Math., **2** (1) (2025), 40–47.
[22] Sh. I. Strelitz: *Asymptotic properties of analytical solutions of differential equations*, Vilnius: Mintis (1972). (In Russian).
[23] H. Wittich: *Neuere untersuchungen über eindeutige analytische functionen*, Springer-Verlag, Berlin (1955).

Andriy Bandura
Ivano-Frankivsk National Technical University of Oil and Gas
Department of Physics and Mathematics
15 Karpatska Street, 76019, Lviv, Ukraine
Email address: andriykopanytsia@gmail.com

Sviatoslav Dubei
Ivan Franko National University of Lviv
Department of Function Theory and Functional Analysis
1 Universytetska Street, 79000, Lviv, Ukraine
Email address: sviatoslav.dubei@lnu.edu.ua

Anastasiia Mokhnal
Ivan Franko National University of Lviv
Faculty of Mechanics and Mathematics
1 Universytetska Street, 79000, Lviv, Ukraine
Email address: smohnal9@gmail.com

Oleh Skaskiv
Ivan Franko National University of Lviv
Department of Function Theory and Functional Analysis
1 Universytetska Street, 79000, Lviv, Ukraine
Email address: oleh.skaskiv@lnu.edu.ua

MODERN MATHEMATICAL METHODS
4 (2026), No. 2, pp. 79-84
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

More on remoteness functions of exact slow NIM

VLADIMIR GURVICH, VLADISLAV MAXIMCHUK, AND MARIYA NAUMOVA*

ABSTRACT. Given integer n and k such that $0 < k < n$ and n piles of stones, two players make moves alternately, reducing by every move exactly k piles from n, by one stone each. The player who must move but cannot is the loser. We call this game exact slow NIM and denote it $\text{NIM}^1_=(n, k)$. In recent articles [25, 26, 27, 28], a very simple explicit rule was suggested choosing a move in each position of this game. Furthermore, in case $n = k + 1$, it was proven that each move chosen by this rule reduces the Smith's remoteness function of the game exactly by one. This result allows us to compute the remoteness function efficiently, thus, solving the game. In the present article, we apply the same rule for two new cases: $k = 2, n = 4, 5$ and show that it still works efficiently: The remoteness function is reduced by one with almost every move. Exceptions are sparse and have a regular pattern. We suggest simple formulas explaining all known exceptions and, thus, solve the two considered new families of games too.

Keywords: Remoteness function, Sprague-Grundy function, slow NIM.

2020 Mathematics Subject Classification: 91A05, 91A46, 91A68.

1. INTRODUCTION

We assume that the reader is familiar with basic concepts of impartial game theory (see e.g., [1, 2] for an introduction) and also with the recent paper [26], where the Smith remoteness function of the game $\text{NIM}^1_=(n, k)$, so-called exact slow NIM, was computed for the case $n = k + 1$. Here, we extend this result and cover two more cases: $\text{NIM}^1_=(4, 2)$ and $\text{NIM}^1_=(5, 2)$.

1.1. **Exact Slow NIM.** Game exact slow NIM was introduced in [22] as follows: Given two integers n and k such that $0 < k \leq n$ and n piles containing $x_1, \dots, x_n$ stones. By one move it is allowed to reduce any k piles by exactly one stone each. Two players alternate turns. One who has to move but cannot is the loser in the normal version of the game and s/he is the winner of its misère version. In [22], this game was denoted $\text{NIM}^1_=(n, k)$. Here, we simplify this notation to $\text{NIM}(n, k)$, for short. Notice that this game is naturally partitioned into k independent subgames, since $x_1 + \ldots + x_n \mod k$ is an invariant: It cannot be changed by a move.

Game $\text{NIM}(n, k)$ is trivial if $k = 1$ or $k = n$. In the first case it ends after $x_1 + \ldots + x_n$ moves and in the second one–after $\min(x_1, \dots, x_n)$ moves. In each case everything depends only on the parity of the corresponding number and nothing on players' skills. All other cases are more complicated.

The game was solved for $k = 2$ and $n \leq 6$. In [24], an explicit formula for the Sprague-Grundy (SG) function was found for $n \leq 4$, for both the normal and misère versions. This formula allows us to compute the SG function in time linear in n, k and $\omega(x) = \sum_{i=1}^n \log(|x_i| + 1)$. Then, in [8] the P-positions of the normal version were found for $n \leq 6$. For the subgame

Received: 22.01.2026; Accepted: 19.04.2026; Published Online: 02.06.2026
*Corresponding author: Mariya Naumova; mnaumova@business.rutgers.edu
DOI: 10.64700/mmm.104

with even $x_1 + \ldots + x_n$ a very simple formula for the P-positions was found, which allows to verify in linear time if x is a P-position and, if not, to find a move from it to a P-position. The subgame with odd $x_1 + \ldots + x_n$ is more difficult. Still a (more sophisticated) formula for the P-positions was found, providing a recognition algorithm linear in time n, k and $\omega(x)$.

1.2. **Case $n = k + 1$, the Normal Version.** In [26], the normal version was solved in case $n = k + 1$ by the following simple rule:

- (o) if all piles are odd, keep a largest one and reduce all other,
- (e) if there exist even piles, keep a smallest one of them and reduce all other.

This rule is well-defined: It defines a unique move $x \to x'$ in each position x[1]. This rule and the corresponding moves is called the M-rule and M-moves; the sequence of successive M-moves is called the M-sequence. Of course, $n > 1$ is required. If $n = 1$, then x_1 will reach an even value in at most one M-move and stop. Since this case is trivial, we can assume that $n > 1$ without any loss of generality.

It is also easily seen that no M-move can result in a position whose all entries are odd. Hence, for an M-sequence, part (o) of the M-rule can be applied at most once, at the beginning; after this only part (e) works.

Given a position $x = (x_1, \ldots, x_n)$, assume that both players follow the M-rule and denote by $\mathcal{M}(x)$ the number of moves from x to a terminal position. In [26], it was proven that $\mathcal{M} = \mathcal{R}$, where $\mathcal{R}$ is the classical remoteness function introduced by Smith [39]. Thus, M-rule solves the game moreover, it allows to win as fast as possible in an N-position and to resist as long as possible in a P-position.

An algorithm computing $\mathcal{M} = \mathcal{R}$ (and in particular, the P-positions) in time polynomial in n, k and $\omega(x)$ was given in [26]. Let us also note that an explicit formula for the P-positions is known only for $n \leq 4$ and already for $n = 3$ it is pretty complicated; see [24] and also [26, Appendix].

1.3. **Related Works on NIM.** By definition, the present game NIM(n, k) is the exact slow version of the famous Moore's NIM$_k$ [34]. In the latter game a player, by one move, reduces arbitrarily (not necessarily by one stone) at most k piles from n. The case $k = 1$ corresponds to the classical NIM whose P-position was found by Bouton [7] for both the normal and misère versions.

Remark 1.1. *Actually, the Sprague-Grundy (SG) values of NIM were also computed in Bouton's paper, although were not defined explicitly in general. This was done later by Sprague* [41] *and Grundy* [18] *for arbitrary disjunctive compounds of impartial games; see also* [11, 39].

In fact, the concept of a P-position was also introduced by Bouton in [7]*, but only for the (acyclic) digraph of NIM, not for all impartial games. In its turn, this is a special case of the concept of a kernel, which was introduced for arbitrary digraphs by von Neumann and Morgenstern* [35]*. Also the misère version was introduced by Bouton in* [7]*, but only for NIM, not for all impartial games. The latter was done by Grundy and Smith* [17]*; see also* [11, 19, 20, 23, 39].

Moore [34] obtained an elegant explicit formula for the P-positions of NIM$_k$ generalizing the Bouton's case $k = 1$. Even more generally, the positions of the SG-values 0 and 1 were efficiently characterized by Jenkins and Mayberry [30]; see also [4, Section 4]. Also in [30], the SG function of NIM$_k$ was computed explicitly for the case $n = k + 1$ (in addition to the case $k = 1$). In general, no explicit formula, nor even a polynomial algorithm computing the SG-values (larger than 1) is known. The smallest open case: 2-values for $n = 4$ and $k = 2$.

[1] Obviously, permuting the piles with the same number of stones we keep the game unchanged.

The remoteness function of NIM_k was recently studied in [5] along with some other classical games: Euclid [9, 10, 16, 19, 20, 31, 32, 33, 37] and Wythoff [43] with some of its generalizations [6, 13, 14, 15, 16, 21, 33, 36, 38, 40, 42, 44]. Further generalizations of the exact slow NIM were considered in [25, 26, 27, 28].

Let us also mention the exact (but not slow) game $\text{NIM}_=(n,k)$ [3, 4] in which, by one move, exactly k from n piles are reduced, but by an arbitrary number of stones each, rather than by 1. The SG-function was efficiently computed in [3, 4] for $n \leq 2k$. Otherwise, even a polynomial algorithm looking for the P-positions is not known (unless $k = 1$, of course). The smallest open case is $n = 5$ and $k = 2$.

2. New Results

In this paper, we consider the remoteness function of the exact slow $\text{NIM}(n,k)$ for $(n,k) = (4,2)$ and $(n,k) = (5,2)$. We extend the M-rule, introduced in [26] for the case $n = k + 1$, to any k such that $1 < k < n$. The obtained GM-rule works efficiently for $\text{NIM}(4,2)$ and $\text{NIM}(5,2)$, that is, it frequently reduces the remoteness function $\mathcal{R}(x)$ by 1. The exceptions are sparse. Moreover, all exceptions found by computer form a regular pattern described by simple formulas. However, these formulas are not proven.

Let us recall that the remoteness function $\mathcal{R}(x)$ has the following properties: The player who makes a move from x can guarantee a victory in at most $\mathcal{R}(x)$ moves whenever $\mathcal{R}(x)$ is odd, that is, x is an N-position, and s/he can guarantee that the play will last at least $\mathcal{R}(x)$ moves if $\mathcal{R}(x)$ is even, in this case x is a P-position. In both cases the corresponding optimal move reduces $\mathcal{R}(x)$ by 1.

2.1. GM-Rule for $\text{NIM}(n,k)$ and Its Basic Properties.

Fix integers n,k such that $0 < k < n$, and consider a nonnegative integer vector $x = (x_1, \ldots, x_n)$. Assume that x is defined up to a permutation of its entries. Hence, without loss of generality, we will require monotonicity: $x_1 \leq \ldots \leq x_n$.

Denote by $m = m(x)$ the number of even entries of x. If $m \geq n-k$, choose the smallest $n-k$ of even entries. If $m < n-k$, choose all m even entries and add arbitrary $n-k-m$ others, for example the largest ones. Note that in both cases the choice may be not unique, since entries of x may be equal. In case of such ambiguity we use the following:

Tie-breaking rule: Always choose the rightmost entries, that is, x_i with the largest i.

The chosen $n-k$ entries of x will be called pivotal and their indices-pivots.

Given n,k and x, define a discrete dynamical system by the next GM-(n,k)-rule. By one step transform x to x' as follows: the $n-k$ pivotal entries of x are not changed, while the remaining k, non-pivotal, entries are reduced by 1. This transformation $x \to x'$ and the corresponding sequence $x = x^0 \to x^1 \to \ldots \to x^j \to \ldots$ will be called GM-(n,k)-move and GM-(n,k)-sequence, respectively. We will omit the arguments (n,k) whenever their values are unambiguous from the context and simply write GM-rule, GM-move, and GM-sequence, for short.

By definition, a GM-move $x \to x'$ is unique for any given x. We begin with several simple properties of a GM-move $x \to x'$.

Observation 2.1. *GM-moves respect monotonicity, that is,* $x'_1 \leq \ldots \leq x'_n$ *whenever* $x_1 \leq \ldots \leq x_n$.

Proof. It follows immediately from the tie-breaking rule. □

Recall that $m = m(x)$ denotes the number of even entries in x.

Observation 2.2. *GM-moves respect inequality* $n-k \leq m(x)$*, that is,* $n-k \leq m(x')$ *whenever* $n-k \leq m(x)$.

Proof. Let a move $x \to x'$ de-pivot $i \in [n] = \{1, \dots, n\}$, that is, i is a pivot of x but not of x'. Then, x_i is even and $x'_i = x_i$, since i is a pivot of x. In contrast, i is not a pivot of x'. Hence, by the tie-breaking rule, there exist $n - k$ pivots of x' each of which is either smaller than i, or larger than i, but its pivotal value equals $x_i = x'_i$. Thus, i may stop to be a pivot, after a GM-move $x \to x'$, only if a replacing pivot appears in x'. □

Observation 2.3. *For an arbitrary x^0, in the corresponding (unique) GM-sequence $x^0 \to x^1 \to \dots \to x^j \to \dots$ inequality $n - k \le m(x^j)$ holds for any $j \ge 1$.*

Proof. Suppose that $n - k > m = m(x^0)$. By definition, m even pivotal entries of x^j are not changed, while $n - k - m$ odd pivotal entries are reduced by 1, by the (unique) GM-move. Clearly, in one GM-move each of these $n - k - m$ entries becomes even. Then, by the GM-rule, it becomes a pivot. Moreover, it remains in this status in x^j until $n - k \ge m(x^j)$. Thus, the set of even pivots is expanding until their number reaches $n - k$, which happens just in one GM-move. After this even pivots may be replaced, but still, by Observation 2.2, their number remains $n - k$. In other words, the difference $m(x^j) - (n - k)$ in one GM-move becomes and then remains non-negative. □

Yet, it may decrease after the GM-move $x^j \to x^{j+1}$ for some j, because some pivots of x^j may disappear after this move, if $m(x^j) \ge n - k$. Also, we will assume that $m(x) \ge n - k$ holds, that is, all pivotal entries are even. By Observation 2.2 this property always holds after a GM-move. However, we do not know when this GM-move is optimal.

2.2. **Exact Slow NIM**$(4, 2)$**.** Our computations show that for this game a GM-move reduces the remoteness value $\mathcal{R}(x)$ by 1 (provided $m(x) \ge n - k = 2$) except the following cases:

$$x = (x_1, x_2, x_3, x_4) = (2a_1 + 1, 2a_2, 2a_3, 2(a_3 + a_2 - a_1)+),$$

where $0 \le a_1 < a_2 \le a_3$. Here, $y+$ means any integer number not smaller than y. Note that $m(x) \ge n - k = 2$. Furthermore, $\mathcal{R}(x) = 2(a_3 + a_2) - 1$ is always odd. In other words, only winning positions may be exceptional. In these exceptional positions the entries x_1 and x_4 are not changed by a (unique) move reducing $\mathcal{R}(x)$ by 1, while the GM-move $x \to x'$ has pivots 2 and 3.

For the GM-move we have $\mathcal{R}(x') = \mathcal{R}(x)+1$, whenever $x_4 > 2(a_3+a_2-a_1)$. This means that the GM-move is still optimal but it delays the fastest winning by 2 moves. If $x_4 = 2(a_3+a_2-a_1)$ then $\mathcal{R}(x') = \mathcal{R}(x)$, which means that the GM-move is losing in a winning position.

2.3. **Exact Slow NIM**$(5, 2)$**.** Our computations show that for this game a GM-move reduces the remoteness value $\mathcal{R}(x)$ by 1 (provided $m(x) \ge n - k = 3$) except for the following cases:

$$x = (x_1, x_2, x_3, x_4, x_5) = (2a_1, 2a_2 + 1, 2a_3, 2a_4, 2(a_4 + a_3 - a_2 + a_1)+),$$

where $0 \le a_1 \le a_2 < a_3 \le a_4$. Note that $m(x) \ge n - k = 3$. Furthermore, $\mathcal{R}(x) = 2(a_4 + a_3 + a_1) - 1$ is always odd. In other words, only winning positions may be exceptional. In these exceptional positions the entries x_1, x_2, and x_5 are not changed by a (unique) move reducing $\mathcal{R}(x)$ by 1, while the GM-move $x \to x'$ has pivots $1, 3$ and 4.

For the GM-move we have $\mathcal{R}(x') = \mathcal{R}(x) + 1$, whenever $x_5 > 2(a_4 + a_3 - a_2 + a_1)$. This means that the GM-move is still optimal but it delays the fastest winning by 2 moves. If $x_5 = 2(a_4 + a_3 - a_2 + a_1)$ then $\mathcal{R}(x') = \mathcal{R}(x)$, which means that the GM-move is losing in a winning position.

Acknowledgments

This research was prepared within the framework of the HSE University Basic Research Program.

References

[1] M. H. Albert, R. J. Nowakowski and D. Wolfe: *Lessons in play: An introduction to combinatorial game theory*, A. K. Peters, Massachusetts (2007).

[2] E. R. Berlekamp, J. H. Conway and R. K. Guy: *Winning ways for your mathematical plays*, **4**, A. K. Peters, Massachusetts (2004).

[3] E. Boros, V. Gurvich, N. B. Ho and K. Makino: *On the Sprague-Grundy function of extensions of proper NIM*, Internat. J. Game Theory, **50** (3) (2021), 635–654.

[4] E. Boros, V. Gurvich, N. B. Ho, K. Makino and P. Mursič: *Sprague-Grundy function of matroids and related hypergraphs*, Theor. Comput. Sci., **799** (2019), 40–58.

[5] E. Boros, V. Gurvich, K. Makino and M. Vyalyi: *Computing remoteness functions of Moore, Wythoff, and Euclid's games*, Int. J. Game Theory, **53** (4) (2024), 1315–1333.

[6] E. Boros, V. Gurvich and V. M. Oudalov: *A polynomial algorithm for a two-parameter extension of Wythoff NIM based on the Perron-Frobenius theory*, Int. J. Game Theory, **42** (4) (2013), 891–915.

[7] C. L. Bouton: *Nim, a game with a complete mathematical theory*, Ann. of Math., **3** (1–4) (1901-1902), 35–39.

[8] N. Chikin, V. Gurvich, K. Knop, M. Paterson and M. Vyalyi: *More about exact slow k-nim*, Integers, **21** (2021), 1–14.

[9] A. J. Cole, A. J. T. Davie: *A game based on the Euclidean algorithm and a winning strategy for it*, Math. Gaz., **53** (1969), 354–357.

[10] D. Collins: *Variations on a theme of Euclid*, Integers, **5** (2005), 1–12.

[11] J. H. Conway: *On numbers and games*, Academic Press, London (1976).

[12] H. S. M. Coxeter: *The golden section, Phyllotaxis and Wythoff's game*, Scripta Math., **19** (1953), 135–143.

[13] A. S. Fraenkel, I. Borosh: *A generalization of Wythoff's game*, J. Combin. Theory Ser. A, **15** (2) (1973), 175–191.

[14] A. S. Fraenkel: *How to beat your Wythoff games' opponent on three fronts*, Amer. Math. Monthly, **89** (1982), 353–361.

[15] A. S. Fraenkel: *Wythoff games, continued fractions, cedar trees and Fibonacci searches*, Theoret. Comput. Sci., **29** (1984), 49–73.

[16] A. S. Fraenkel: *Euclid and Wythoff games*, Discrete Math., **304** (2005), 65–68.

[17] P. M. Grundy, C. A. B. Smith: *Disjunctive games with the last player losing*, Math. Proc. Cambridge Philos. Soc., **52** (1956), 527–533.

[18] P. M. Grundy: *Mathematics of games*, Eureka, **2** (1939), 6–8.

[19] V. Gurvich: *On the misère version of game Euclid and miserable games*, Discrete Math., **307** (9-10) (2007), 1199–1204.

[20] V. Gurvich: *Miserable and strongly miserable impartial games*, RUTCOR Research Report 18-2011, Rutgers University (2011).

[21] V. Gurvich: *Further generalizations of the Wythoff game and minimum excludant*, Discrete Appl. Math., **160** (2012), 941–947.

[22] V. Gurvich, N. B. Ho: *Slow k-Nim*, RUTCOR Research Report 3-2015, Rutgers University (2015), `http://arxiv.org/abs/1508.05777`.

[23] V. Gurvich, N. B. Ho: *On tame, pet, domestic, and miserable impartial games*, Discrete Appl. Math., **243** (2018), 54–72.

[24] V. Gurvich, S. Heubach, N. B. Ho and N. Chikin: *Slow k-Nim*, Integers, **20** (2020), 1–19.

[25] V. Gurvich, M. Naumova: *GM-rule and its applications to impartial games*, arXiv 2311.03257, (2023).

[26] V. Gurvich, D. Martynov, V. Maximchuk and M. Vyalyi: *On remoteness functions of exact slow k-NIM with $k+1$ piles*, Integers, **24** (2024), 1–21.

[27] V. Gurvich, V. Maximchuk, G. Miheenkov and M. Naumova: *On remoteness functions of exact slow k-NIM with $k+1$ piles in normal and misère versions*, Games, **15** (6) (2024), Article ID:37.

[28] V. Gurvich, M. Naumova: *Screw discrete dynamical systems and their applications to exact slow NIM*, Discrete Appl. Math., **358** (2024), 382–394.

[29] V. Gurvich, M. Vyalyi: *On computational hardness of multidimensional subtraction games*, Algorithms, **14** (3) (2021), Article ID:71.

[30] T. A. Jenkyns, J. P. Mayberry: *The skeleton of an impartial game and the Nim-function of Moore's Nim_k*, Internat. J. Game Theory, **9** (1) (1980), 51–63.

[31] T. Lengyel: *A Nim-type games and continued fraction*, Fibonacci Quart., **41** (2003), 310–320.

[32] T. Lengyel: *On calculating the Sprague-Grundy function for the game Euclid*, Proc. 11th Int. Conf. on Fibonacci Numbers and Their Applications, Braunschweig (Germany) (2004), 169–175.

[33] R. Lenhardt: *Composite mathematical games*, Bachelor Thesis, Comenius University, Bratislava (2007).

[34] E. H. Moore: *A generalization of the game called Nim*, Ann. of Math., **11** (3) (1910), 93–94.
[35] J. von Neumann, O. Morgenstern: *Theory of games and economic behavior*, Princeton University Press, Princeton (1944).
[36] G. Nivasch: *More on the Sprague–Grundy function for Wythoff's game*, Games of No Chance 3, MSRI Publications, **56** (2009), 377–410.
[37] G. Nivasch: *The Sprague–Grundy function of the game Euclid*, Discrete Math., **306** (21) (2006), 2798–2800.
[38] G. Nivasch: *The Sprague–Grundy function for Wythoff's game: On the location of the q-values*, Master Thesis, Weizmann Institute of Sciences, Rehovot (2004).
[39] C. A. B. Smith: *Graphs and composite games*, J. Combin. Theory, **1** (1966), 51–81.
[40] E. L. Spitznagel Jr.: *Properties of a game based on Euclid's algorithm*, Math. Mag., **46** (1973), 87–92.
[41] R. Sprague: *Über mathematische Kampfspiele*, Tohoku Math. J., **41** (1936), 438–444.
[42] P. D. Straffin: *A Nim-type game, solution to problem # 1537*, Math. Mag., **71** (1998), 394–395.
[43] W. A. Wythoff: *A modification of the game of Nim*, Nieuw Arch. Wisk, **7** (1907), 199–202.
[44] A. M. Yaglom, I. M. Yaglom: *Challenging mathematical problems with elementary solutions vol. II*, Holden-Day, London (1967).

VLADIMIR GURVICH
NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS
MOSCOW, RUSSIA
RUTGERS UNIVERSITY
RUTCOR
PISCATAWAY, NJ, USA
Email address: vgurvich@hse.ru, vladimir.gurvich@gmail.com

VLADISLAV MAXIMCHUK
NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS (HSE)
MOSCOW, RUSSIA
Email address: vladislavmaximchuk3495@gmail.com

MARIYA NAUMOVA
RUTGERS UNIVERSITY
RUTGERS BUSINESS SCHOOL
PISCATAWAY, NJ, USA
Email address: mnaumova@business.rutgers.edu

MODERN MATHEMATICAL METHODS
4 (2026), No. 2, pp. 85-118
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

Linear algebra and power method combined with method of deflation applied to Toeplitz sinc matrices: A new approach for the computation of eigenvalues and eigenvectors

LUDWIG KOHAUPT* AND YAN WU

ABSTRACT. A new robust method for the computation of the eigenvalues and eigenvectors of real symmetric sinc $n \times n$-matrices $A(t) = A(t,n)$ with t in $0 < t < 1$ is developed. If the eigenvalues $\mu_j(t) = \mu_j(A(t)) = \mu_j(A(t,n))$, $j = 1,\ldots,n$ are arranged in descending order, one has $\mu_1(t) \to n \ (t \to 0)$ and $\mu_j(t) \to 0 \ (t \to 0)$, $j = 2,\ldots,n$. The new approach consists in computing only a limited number i_q of the eigenvalues and to set the remaining very small eigenvalues equal to zero, of which the corresponding eigenvectors are reconstructed via the Gram-Schmidt process. This opens the way for conceiving a new method for the computation of the eigenvalues and associated eigenvectors of severely ill-conditioned real symmetric Toeplitz sinc matrices that delivers good results.

Keywords: Condition number, eigenvalues and eigenvectors, method of deflation, power method, Toeplitz sinc matrix.

2020 Mathematics Subject Classification: 15B05, 65F15, 65F22, 65F25.

1. INTRODUCTION

The purpose of this introduction is to deal with the following questions:

Question 1. *Why is this paper written?*

Question 2. *What is its relation to other literature on the subject?*

Question 3. *What is new in this paper, and perhaps what makes it unique?*

Question 4. *What is its content?*

These questions are interrelated, and the answers are partly interwoven. Problems arise in the computation of eigenvalues and associated eigenvectors of Toeplitz sinc $n \times n$-matrices $A(t) = A(t,n)$ that are ubiquitous in digital signal processing. These are made up of the entries $s(0) := 1$ and $s(jt) := \sin(j\pi t)/(j\pi t)$, $j = 1,\ldots,n-1$ and are investigated for $0 < t < 1$. The eigenvalues $\mu_j(t) = \mu_j(A(t)) = \mu_j(A(t,n))$, $j = 1,\ldots,n$ are positive and are assumed to be arranged in descending order. Since $\mu_1(t) \to n \ (t \to 0)$ and $\mu_j(t) \to 0 \ (t \to 0)$, $j = 2,\ldots,n$, see [4], for the condition number $\kappa_2(t)$, one obtains $\kappa_2(t) = \mu_1(t)/\mu_n(t) \to \infty \ (t \to 0)$. This leads to numerical problems in the determination of eigenvalues and associated eigenvectors for small values of t since on a computer the set of real numbers $\mathbb{R}$ is replaced by machine numbers with only a restricted number of digital places.

Special Toeplitz matrices, namely those that commute with tridiagonal matrices were discussed by Grünbaum in [1]. However, apart from its credentials, it has some shortcomings.

Received: 05.01.2026; Accepted: 13.07.2026; Published Online: 24.07.2026
*Corresponding author: Ludwig Kohaupt; lkohaupt4@web.de
DOI: 10.64700/mmm.97

For example, in [8], the authors proved that the tridiagonal matrices used there have distinct eigenvalues whereas Grünbaum had to assume this. A weak point of this paper is also that it does not present numerical results. In [4], we discussed also the paper of Hertz [2] titled "Simple Bounds on the Extreme Eigenvalues of Toeplitz Matrices", and we concluded that the obtained results have merits, but have only limited applicability.

This paper is written in order to do without the restriction in [1] that the Toeplitz sinc matrix should commute with a tridiagonal matrix, and we choose a totally different approach to determine the eigenvalues and associated eigenvectors of a Toeplitz sinc $n \times n$-matrix $A(t) = A(t, n)$.

This new approach starts from the following observation. Since $\mu_j(t) \to 0 \ (t \to 0), \ j = 2, \ldots, n$, for small values of t in $0 < t < 1$ we determine an index i_q such that $\mu_j(t)$ is less than a prescribed value for $j = i_q + 1, \ldots, n$ and set $\mu_j(t) = 0, j = i_q + 1, \ldots, n$. In other words, we replace the eigenvalues $\mu_j(t)$, $j = 1, \ldots, n$ by $\mu_j(t)$, $j = 1, \ldots, i_q$ and $\mu_j(t) = 0$, $j = i_q + 1, \ldots, n$. Then, the eigenvalues $\mu_j(t)$, $j = 1, \ldots, i_q$ and associated eigenvectors can be determined without difficulty by the power method combined with the method of deflation, and for the eigenvalues $\mu_j(t) = 0$, $j = i_q + 1, \ldots, n$, many systems of orthonormal eigenvectors can easily be given where we choose a very simple one. Application of the Gram-Schmidt orthonormalization process to the eigenvectors corresponding to the eigenvalues $\mu_j(t)$, $j = 1, \ldots, i_q$ and $\mu_j(t) = 0$, $j = i_q + 1, \ldots, n$ delivers for the whole set of eigenvalues $\mu_j(t)$, $j = 1, \ldots, i_q, i_q + 1, \ldots, n$ and the associated eigenvectors $w_j(t)$, $j = 1, \ldots, i_q, i_q + 1, \ldots, n$ good approximations.

The important point is that we avoid-for small values of t in $0 < t < 1$-the difficulty caused by $\mu_j(t) \to 0 \ (t \to 0), \ j = 2, \ldots, n$ by selecting a limited number i_q of eigenvalues that are computed whereas simple eigenvectors for the other eigenvalues $\mu_j(t) = 0$, $j = i_q + 1, \ldots, n$ can easily be given. This idea seems to be unique, so far. This was also the reason why we commented only shortly on other methods for the computation of eigenvalues and associated eigenvectors of Toeplitz sinc matrices.

Next, we come to the content of this paper; it is organized as follows. In Section 2, Linear-Algebra Tools are collected consisting in material on Projections and Invariance of Vector Spaces under Linear Transformations in Subsection 2.1, Real Symmetric Matrices in Subsection 2.2, Gram-Schmidt Orthonormalization in Subsection 2.3, and Toeplitz Sinc Matrices in Subsection 2.4. These results are known, but the presentation is much more detailed than can be found in literature. In Section 3, the Power Method for a Real Symmetric Matrix with Non-Negative Eigenvalues is described which is used to compute its largest positive eigenvalue. In Section 4, the Method of Deflation is discussed. In combination with the power method, it allows one to determine all eigenvalues of a positive definite matrix. In Section 5, the obtained results are applied to Toeplitz sinc matrices $A(t, n)$ for $t = 0.1$ with $n = 5$ in Subsection 5.1, with $n = 10$ in Subsection 5.2, and with $n = 15$ in Subsection 5.3, and Subsection 5.4 contains a scrutiny of the computed eigenvectors. Finally, in Section 6, conclusions are drawn.

The style of the paper is expository in order to address a large readership. Only cited literature is given in the References.

2. Collection of Linear-Algebra Tools

This section collects those tools from Linear Algebra that play an important role in this paper. It is organized in subsections and contains, for instance, material on projections and invariance properties of linear transformations, on eigenvalues and eigenvectors of real symmetric matrices, on the Gram-Schmidt orthonormalization process as well as on Toeplitz sinc matrices.

The underlying vector spaces over the field $\mathbb{F} = \mathbb{R}$ resp. $\mathbb{F} = \mathbb{C}$ are supposed to be finite dimensional, and the operators (also termed mappings or transformations) are represented by matrices.

2.1. **Projections and Invariance of Vector Spaces under Linear Operators.** The first subsection is taken from [3, Chapter I, §3, Sections 3. and 4., pp.20-23] with many verbatim passages.

In a (finite-dimensional) vector space, the projection operator is introduced. It is idempotent and is associated with the notions complementary space or the direct sum of subspaces. Further, linear operators that map a subspace into itself are considered. One says that they let the corresponding subspace invariant or that the subspace is invariant under the linear operators.
(i) Projections
Let M, N be complementary subspaces in the vector space X; by this, we mean that

$$X = M \oplus N,$$

i.e., each $u \in X$ can be uniquely expressed in the form $u = u' + u''$ with $u' \in M$ and $u'' \in N$, see [3, Chapter I, §3, Section 4.]. We also say that X is the direct sum of M and N. The element u' is called the projection of u on M along N. If $v = v' + v''$ in the same sense, $\alpha\, u + \beta\, v$ has the projection $\alpha\, u' + \beta\, v'$ on M along N. If we set $u' = P\,u$, it follows that P is a linear operator on X to itself. P is called projection operator (or simply projection) on M along N. We have $P\,u = u$ if and only if $u \in M$ and $Pu = 0$ if and only if $u \in N$. The range of P is M and the null space of P is N. Since $P\,u \in M$, we have $P(Pu) = Pu$, that is, P is idempotent:

$$P^2 = P.$$

Conversely, any idempotent operator P is a projection. In fact, set $M = R(P)$, where $R(P)$ is the range of P and $N = R(Q) = R(I - Q)$. Then, $u' \in M$ implies that $u' = Pu$ for some $u \in X$ and therefore $Pu' = P^2u = Pu = u'$. Similarly, $u'' \in N$ implies $Pu'' = 0$. Hence, $u \in M \cap N = \{0\}$. Each $u \in X$ has the expression $u = u' + u''$ with $u' = Pu \in M$ and $u'' = (I - P)u \in N$. This shows that P is the projection on M along N.
(ii) Invariance of Vector Spaces under Linear Operators
This part is taken from [3, Chapter I, §3, Section 5.]. A subspace M of a vector space X is said to be invariant under a linear operator $T \in L(X)$ if $TM \subset M$. In this case, T induces a linear operator T_M on M to M defined by $T_M u = Tu$ for $u \in M$. If there are two invariant subspaces M, $N \subset X$ such that $X = M \oplus N$, T is said to be decomposed (or reduced) by the pair M, N.

2.2. **Real Symmetric Matrices.** In this subsection, we investigate the eigenvalues and associated eigenvectors of real symmetric $n \times n$-matrices B in the space of n-tuples $H = \mathbb{R}^{n\times n}$ endowed with the scalar product $(\cdot,\cdot) = (\cdot,\cdot)_2$. We shall see that there is a basis of real orthonormal eigenvectors $w_j, j = 1,\dots,n$ associated with the eigenvalues $\lambda_j = \lambda_j(B)$, $j = 1,\dots,n$ and standard eigenvector expansions for u and Bu with arbitrary $u \in H$. In the special case when $\lambda_j \neq 0$, $j = 1,\dots,m$ with $m < n$, we derive what can be called displayed eigenvector expansions for u and Bu with arbitrary $u \in H$. For $M_m := [w_1,\dots,w_m]$, one can show that $BM_m \subset M_m$ and that B is invertible on M_m.
(i) Eigenvalues and Eigenvectors
We have the following theorem.

Theorem 2.1 (Real eigenvalues and real eigenvectors of real symmetric mappings)**.** *Let the matrix $B = (b_{jk})_{j,k=1,\dots,n}$ be a real symmetric mapping in the space of n-tuples $H = \mathbb{R}^n$ endowed with the scalar product $(\cdot,\cdot) = (\cdot,\cdot)_2$. Then, the eigenvalues $\lambda_1,\dots,\lambda_n$ of B are real, and there is a corresponding orthonormal system of real eigenvectors $w_1,\dots,w_n$ with*

$$Bw_j = \lambda_j w_j, \quad (w_j, w_k) = \delta_{jk}, \quad j,k = 1,\dots,n.$$

Proof. First, we consider B in the space of n-tuples $\mathbb{F}^n = \mathbb{C}^n$. Then, the proof follows from [5, (15), p.91], where the eigenvectors can be elements of $\mathbb{C}^n$. Now, $w_j = Re\, w_j + iIm\, w_j = w_j^{(r)} + iw_j^{(i)}$, $j = 1, \dots, n$ with $w_j^{(r)} = Re\, w_j$ *and* $w_j^{(i)} = Im\, w_j$, $j = 1, \dots, n$. Since $w_j \neq 0$, one has $w_j^{(r)} \neq 0$ or $w_j^{(i)} \neq 0$ for $j = 1, \dots, n$. If one replaces $w_j \in \mathbb{C}^n$ by $w_j^{(r)}$ if $w_j^{(r)} \neq 0$ or $w_j \in \mathbb{C}^n$ by $w_j^{(i)}$ if $w_j^{(i)} \neq 0$ for $j = 1, \dots, n$, one obtains, after normalization, real normed eigenvectors w_j, $j = 1, \dots, n$. They are orthonormal or can be made orthonormal. To see this, consider two eigenvalues $\lambda_k \neq \lambda_l$. Then, $\lambda_k(w_k, w_l) = (\lambda_k w_k, w_l) = (Bw_k, w_l) = (w_k, Bw_l) = (w_k, \lambda_l w_l) = \lambda_l(w_k, w_l)$ leading to $(\lambda_k - \lambda_l)(w_k, w_l) = 0$ or $(w_k, w_l) = 0$. For $\lambda_k = \lambda_l$, the Gram-Schmidt procedure delivers orthonormal, real eigenvectors w_k, w_l. □

We mention that we have included the proof of Theorem 2.1 since the last case is usually not treated explicitly, at least not in engineering literature. From Theorem 2.1, we get the following theorem.

Theorem 2.2 (Standard eigenvector expansions). *Let the matrix $B = (b_{jk})_{j,k=1,\dots,n}$ be a real symmetric mapping in the space of n-tuples $H = \mathbb{R}^n$ endowed with the scalar product $(\cdot,\cdot) = (\cdot,\cdot)_2$. Let $\lambda_1, \dots, \lambda_n$ be the real eigenvalues of B, and let $w_1, \dots, w_n$ be a corresponding orthonormal system of real eigenvectors so that*

$$Bw_j = \lambda_j w_j, \quad (w_j, w_k) = \delta_{jk}, \quad j, k = 1, \dots, n.$$

Then, the following eigenvector expansions hold:

$$u = \sum_{j=1}^{n} (u, w_j) w_j, \quad u \in H,$$
$$Bu = \sum_{j=1}^{n} \lambda_j (u, w_j) w_j, \; u \in H.$$

Proof. The proof follows immediately from Theorem 2.1. □

(ii) Special Case When $\lambda_j = 0$ for $j = m+1, \dots, n$
Let $m \in \mathbb{N}$ with $1 \leq m < n$. Then, we consider the special case

$$\lambda_j \neq 0, \; j = 1, \dots, m, \quad \lambda_j = 0, \; j = m+1, \dots, n.$$

Set

$$M_m = [w_1, \dots, w_m], \quad M_0 = [w_{m+1}, \dots, w_n].$$

Then,

$$H = M_m \oplus M_0$$

with

$$M_0 = M_m^{\perp}.$$

This situation leads to the subsequent corollary.

Corollary 2.1 (Displayed eigenvector expansions in special case). *In the above special case, the following eigenvector expansions for u and Bu are valid:*

$$u = \sum_{j=1}^{m} (u, w_j) w_j + P_0 u, \quad u \in H$$

where $P_0 : H \mapsto N(B)$ is the projection of H on the null space $N(B)$ and where P_0 is defined by

$$P_0 u = \sum_{j=m+1}^{n} (u, w_j) w_j, \quad u \in H.$$

Further,

$$Bu = \sum_{j=1}^{m} \lambda_j (u, w_j) w_j, \quad u \in H.$$

Proof. We show that $P_0 : H \mapsto N(B)$ and that it is a projection. First,

$$\begin{aligned} B(P_0 u) =& B\left[\sum_{j=m+1}^{n} (u, w_j) w_j\right] = \sum_{j=m+1}^{n} (u, w_j) B w_j \\ =& \sum_{j=m+1}^{n} (u, w_j) 0 w_j = 0, \quad u \in H \end{aligned}$$

so that $B : H \mapsto N(B)$. Further, $u' := P_0 u$ is uniquely determined by the representation of $P_0 u$ and is therefore a projection according to Subsection 2.1. Next, we have

$$\begin{aligned} Bu =& B\left[\sum_{j=1}^{m} (u, w_j) w_j + \sum_{j=m+1}^{n} (u, w_j) w_j\right] \\ =& \sum_{j=1}^{m} (u, w_j) B w_j + \sum_{j=m+1}^{n} (u, w_j) B w_j \\ =& \sum_{j=1}^{m} (u, w_j) \lambda_j w_j + \sum_{j=m+1}^{n} (u, w_j) 0 w_j \\ =& \sum_{j=1}^{m} (u, w_j) \lambda_j w_j, \quad u \in H. \end{aligned}$$

□

By the way,

- we have $M_0 = M_m^{\perp}$ and thus $H = M_m \oplus M_0 = M_m \oplus M_m^{\perp}$,
- the standard eigenvector expansions in Theorem 2.2 remain valid, of course, but are less precise,
- the displayed eigenvector expansions for u and Bu in Corollary 2.1 were inspired by the first vector expansion in [6, Satz 7.3, p.34] as well as the first vector expansion in [6, Satz 7.2, p.32], as the case may be.

(iii) Invertibility of the Operator B on M_m

We have $BM_m \subset M_m$, that is, M_m is invariant under the linear operator B. This leads to the following theorem.

Theorem 2.3 (Eigenvector expansion of B^{-1} on M_m). *In the above special case, the $n \times n$-matrix B restricted to M_m is invertible and has the eigenvector expansion*

$$B^{-1} u = \sum_{j=1}^{m} \frac{1}{\lambda_j} (u, w_j) w_j, \; u \in M_m.$$

Proof. To prove this, let $u \in M_m$. Then,

$$\begin{aligned} u =& \sum_{j=1}^{m} (u, w_j) w_j = \sum_{j=1}^{m} \lambda_j (u, w_j) \frac{1}{\lambda_j} w_j \\ =& \sum_{j=1}^{m} \lambda_j (u, w_j) B^{-1} w_j = B^{-1} \left[\sum_{j=1}^{m} \lambda_j (u, w_j) w_j \right] \\ =& B^{-1}(Bu) = (B^{-1}B)u, \ u \in M_m \end{aligned}$$

so that B has a left inverse on M_m. Similarly,

$$\begin{aligned} u =& \sum_{j=1}^{m} (u, w_j) w_j = \sum_{j=1}^{m} \frac{1}{\lambda_j} (u, w_j) \lambda_j w_j \\ =& \sum_{j=1}^{m} \frac{1}{\lambda_j} (u, w_j) B w_j = B \left[\sum_{j=1}^{m} \frac{1}{\lambda_j} (u, w_j) w_j \right] \\ =& B(B^{-1}u) = (BB^{-1})u, \ u \in M_m \end{aligned}$$

so that B has a right inverse on M_m. On the whole, we have proven that B has an inverse on M_m that commutes with B. □

We mention that in the case $m = n$, one has $M_m = H$. Then, it follows that B is invertible on H.

2.3. **Gram-Schmidt Orthonormalization.** We have the following theorem.

Theorem 2.4 (Gram-Schmidt orthonormalization process)**.** *Let E be a vector space over the field $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ endowed with the scalar product $(\cdot, \cdot)$. Further, let $u_1, \ldots, u_m$ be a system of linearly independent vectors in E. Then, from the formulae*

$$\begin{aligned} v_1 &= u_1, & w_1 &= \tfrac{1}{\|v_1\|} v_1 \\ v_j &= u_j - \sum_{k=1}^{j-1} (u, w_k) w_k, & w_j &= \tfrac{1}{\|v_j\|} v_j, \ j = 2, \ldots, m \end{aligned}$$

an orthogonal system $v_1, \ldots, v_m$ and an orthonormal system $w_1, \ldots, w_m$ are obtained and are such that

$$M_j = [u_1, \ldots, u_j] = [v_1, \ldots, v_j] = [w_1, \ldots, w_j], \ j = 1, \ldots, m.$$

Proof. See [5, Section 7.1(3), p.111]. □

2.4. **Toeplitz Sinc Matrices.** In this subsection, we collect the important properties of Toeplitz sinc matrices obtained in [4] and [7]. They serve, in Section 4, to test whether the power method combined with the method of deflation are appropriate in computing the eigenvalues and associated eigenvectors with a given precision. Many verbatim passages in this subsection are taken from [4].

A Toeplitz sinc matrix $A(t) = A_n(t) = A(t,n)$ is defined as

$$A(t) = A(t,n) = \begin{bmatrix} s(0) & s(t) & s(2\,t) & \cdots & s((n-2)\,t) & s((n-1)\,t) \\ s(t) & s(0) & s(t) & \cdots & s((n-3)\,t) & s((n-2)\,t) \\ s(2\,t) & s(t) & s(0) & \cdots & & s((n-3)\,t) \\ & & & \vdots & & \\ s((n-1)\,t) & s((n-2)\,t) & & \cdots & s(t) & s(0) \end{bmatrix} \tag{2.1}$$

where $0 < t < 1$ and $s(0) = 1$ as well as $s(t) = sinc(t) = \sin(\pi t)/(\pi t)$. From [7, Theorem 2.2], it follows that this matrix is positive definite by setting there $t_1 = 0, t_i = (i-1)\,t,\ i = 2, \ldots, n$ and taking into account that $s(-t) = s(t)$ for $0 < t < 1$. As t gets smaller, the condition number of this matrix deteriorates quickly. As a rule, this leads to numerical problems, for instance, when one wants to compute the eigenvalues and associated eigenvectors. For the above Toeplitz sinc matrix, we state the following theorem.

Theorem 2.5. *Let the matrix $A(t) = A_n(t) = A(t,n)$ in (2.1) be given. Further, let the eigenvalues $\mu_j(t) = \mu_j(A(t)) = \mu_j(A_n(t)),\ j = 1, \ldots, n$ be arranged according to*

$$\mu_1(t) \geq \mu_2(t) \geq \cdots \geq \mu_n(t). \tag{2.2}$$

Then, the limits

$$A := \lim_{t\to 0} A(t) \tag{2.3}$$

as well as

$$\mu_j = \mu_j(A) := \lim_{t\to 0} \mu_j(t) = \lim_{t\to 0} \mu_j(A(t)), \quad j = 1, \ldots, n \tag{2.4}$$

exist and

$$A = A_n = \lim_{t\to 0} A(t) = \lim_{t\to 0} A_n(t) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ & & & \vdots & & \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}. \tag{2.5}$$

Further, the limits in (2.4) are eigenvalues of A, and with appropriate enumeration of the eigenvalues $\mu_j := \mu_j(A),\ j = 1, \ldots, n$, one has

$$\lim_{t\to 0} \mu_1(A(t)) = \lim_{t\to 0} \mu_1(A_n(t)) = \mu_1(A) = \mu_1 = n, \tag{2.6}$$

$$\lim_{t\to 0} \mu_j(A(t)) = \lim_{t\to 0} \mu_j(A_n(t)) = \mu_j(A) = \mu_j = 0,\ j = 2, \ldots, n. \tag{2.7}$$

Proof. See [4, Theorem 2.1]. □

We continue with comments on these results. For the special limits $\lim_{t\to 0} s(t)$ and $\lim_{t\to 1} s(t)$, one obtains

$$\lim_{t\to 0} s(t) = \lim_{t\to 0} \frac{\sin(\pi t)}{\pi t} = 1 \quad and \quad \lim_{t\to 1} s(t) = \lim_{t\to 1} \frac{\sin(\pi t)}{\pi t} = 0.$$

Further, the limit of the condition number $\kappa_2(t) = \mu_1(t)/\mu_n(t)$ is given by

$$\lim_{t\to 0} \kappa_2(t) = \infty$$

which indicates that problems might arise in numerical computations involving Toeplitz sinc matrices for small values of t in $0 < t < 1$. But, we have seen in Theorem 2.5 that the eigenvalues of the limit matrix $A = \lim_{t\to 0} A(t) = \lim_{t\to 0} A_n(t)$ are given by $\mu_1(A) = n$ and $\mu_j(A) = 0,\ j = 2,\dots,n$.

Moreover, a set of eigenvectors $u_1,\dots,u_n$ of A is given by

$$U = [u_1 \,|\, u_2,\dots,u_n] = \left[\begin{array}{c|cccccc} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -1 & 1 & \cdots & 0 & 0 \\ 1 & 0 & -1 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ & & & \vdots & & \\ 1 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 0 & \cdots & -1 & 1 \\ 1 & 0 & 0 & \cdots & 0 & -1 \end{array}\right]$$

with the property $rank(U) = n$ as well as

$$(u_1, u_1) = n, \quad (u_1, u_j) = 0,\ j = 2,\dots,n,$$

see [4, Section 8, for $n = 5$] where we used the corresponding normed vectors, instead. Here, u_1 is eigenvector associated with the largest eigenvalue $\mu_1 = \mu_1(A) = n$, and $u_j, j = 2,\dots,n$ are eigenvectors associated with the eigenvalues $\mu_j = \mu_j(A) = 0, j = 2,\dots,n$. Starting from these eigenvectors, applying the Gram-Schmidt process to $u_1,\dots,u_n$, orthonormal eigenvectors $w_1,\dots,w_n$ with $[w_1,\dots,w_n] = [u_1,\dots,u_n]$ can be found. This will be exploited in Section 5.

3. Power Method for Real Symmetric Matrices with Non-Negative Eigenvalues

This section discusses the power method for symmetric real $n \times n$-matrices B with non-negative eigenvalues $\lambda_j, j = 1,\dots,n$ arranged in descending order. Since this method determines the largest positive eigenvalue, it is clear that among the eigenvalues must be at least a positive one. Otherwise, all eigenvalues were equal to zero, a particular case that is not further investigated in this paper since it does not seem to have interesting applications.

In the following, we give a detailed account of the convergence of the power method exhibiting that - under mild conditions - it computes the largest positive eigenvalue and a corresponding normed eigenvector.

All other positive eigenvalues and associated orthonormal eigenvectors can be obtained by the power method when it is combined with the method of deflation, the last one being described in the next section. We take over most passages from [5, Subsection 10.1.1] on the power method for symmetric matrices $B \in \mathbb{R}^{n\times n}$ with non-negative eigenvalues $\lambda_j,\ j = 1,\dots,n$ arranged according to

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0. \tag{3.8}$$

However, the presentation is much more elaborate. The power method or von Mises method considers, first, for an arbitrarily chosen, but fixed initial vector $u^{(0)} \neq 0$ in $\mathbb{F}^n = \mathbb{R}^n$, the corresponding sequence of iterated vectors

$$u^{(\imath+1)} = Bu^{(\imath)},\ \imath = 0, 1, 2, \dots.$$

Using the orthonormal basis of eigenvectors of B, one can express, with $u^{(0)} = u$, these iterated vectors as

$$u^{(\imath)} = B^{\imath} u = \sum_{k=1}^{n} \lambda_k^{\imath} (u, w_k) w_k, \ \imath = 0, 1, 2, \ldots . \tag{3.9}$$

If the initial vector $u = u^{(0)}$ satisfies the hypothesis

$$u^{(1)} = Bu^{(0)} \neq 0 \quad \text{or equivalently} \quad \|u^{(1)}\|^2 = \|Bu\|^2 = \sum_{k=1}^{n} \lambda_k^2 |(u, w_k)|^2 > 0,$$

then there is a natural number m such that

$$\lambda_m (u, w_m) \neq 0, \ \lambda_k (u, w_k) = 0, \ k < m. \tag{3.10}$$

In the expression (3.9) for $u^{(\imath)}$, we assume that $\lambda = \lambda_m$ is the largest positive eigenvalue that occurs, and for sufficiently large values of $\imath$ the terms in $\lambda_k = \lambda$ dominate the other terms with $\lambda_k < \lambda$ as far as these occur in the summation. Accordingly, we split up the sum in (3.9), and writing

$$\lambda = \lambda_m > 0, \quad z = \sum_{\substack{k=m \\ \lambda_k = \lambda}}^{n} (u, w_k) w_k, \quad r^{(\imath)} = \sum_{\substack{k=m \\ \lambda_k < \lambda}}^{n} \left(\frac{\lambda_k}{\lambda} \right)^{\imath} (u, w_k) w_k, \tag{3.11}$$

we can express the iterated vectors $u^{(\imath)}$ as

$$u^{(\imath)} = \lambda^{\imath} (z + r^{(\imath)}), \ \imath = 1, 2, \ldots . \tag{3.12}$$

From the representation in (3.11), we see that the vector z is orthogonal to $r^{(\imath)}$, and so the relation

$$\|u^{(\imath)}\| = \lambda^{\imath} (\|z\|^2 + \|r^{(\imath)}\|^2)^{\frac{1}{2}}, \ \imath = 1, 2, \ldots \tag{3.13}$$

holds. The vector z is a linear combination of the orthonormal eigenvectors of B for the eigenvalue $\lambda = \lambda_m$, and by condition (3.10), $z \neq 0$, and so z is an eigenvector corresponding to the eigenvalue λ, i.e.,

$$Az = \lambda z, \ z \neq 0.$$

Let λ' be the largest positive eigenvalue of B with $\lambda' = \lambda_k < \lambda_m = \lambda$ for $k = m+1, \ldots, n$ if there is such a one, and otherwise let $\lambda' = 0$. Then, the vectors $r^{(\imath)}$ in (3.11) converge to zero according to the estimate

$$0 \leq q = \frac{\lambda'}{\lambda} < 1, \quad \|r^{(\imath)}\| \leq q^{\imath} \|r^{(0)}\| \to 0 \ (\imath \to \infty). \tag{3.14}$$

As a result of these considerations, we therefore have the convergence of the normalized iterated vectors

$$v^{(\imath)} = \frac{u^{(\imath)}}{\|u^{(\imath)}\|} = \frac{z + r^{(\imath)}}{\|z + r^{(\imath)}\|} \to \frac{z}{\|z\|} = w \ (\imath \to \infty),$$

i.e., convergence to the normalized eigenvector $w = z/\|z\|$ of B corresponding to the eigenvalue λ. We determine the corresponding approximations to the eigenvalue by means of the Rayleigh quotients of B in the form

$$\varrho(u^{(\imath)}) = \frac{(Bu^{(\imath)}, u^{(\imath)})}{\|u^{(\imath)}\|^2} = (Bv^{(\imath)}, v^{(\imath)}) = \varrho(v^{(\imath)}), \quad \imath = 0, 1, 2, \ldots,$$

and we continue with a two-sided estimate on $(Bu^{(\imath)}, u^{(\imath)})$ contained in the following lemma.

Lemma 3.1 (Two-sided estimate on $(Bu^{(\imath)}, u^{(\imath)})$). *From the representation*

$$u^{(\imath)} = B^{\imath} u = \sum_{k=1}^{n} \lambda_k^{\imath} (u, w_k) w_k, \; \imath = 0, 1, 2, \ldots,$$

we obtain the two-sided estimate

$$\lambda^{2\imath+1} \|z\|^2 \leq (Bu^{(\imath)}, u^{(\imath)}) \leq \sum_{k=1}^{n} \lambda_k^{2\imath+1} |(u, w_k)|^2 \leq \lambda \, \|u^{(\imath)}\|^2, \quad \imath = 0, 1, 2, \ldots.$$

Proof. One has

$$Bu^{(\imath)} = B[\,\lambda^{\imath}(z + r^{(\imath)})\,] = \lambda^{\imath}[Bz + Br^{(\imath)}\,] = \lambda^{\imath}[\lambda z + Br^{(\imath)}\,] = \lambda^{\imath+1} z + \lambda^{\imath} Br^{(\imath)},$$

where $z \perp Br^{(\imath)}$, $\imath = 0, 1, 2, \ldots$ due to the representation (3.11). Herewith, we can express the iterated vectors $u^{(\imath)}$ as

$$u^{(\imath)} = \lambda^{\imath}(z + r^{(\imath)}), \quad \imath = 0, 1, 2, \ldots$$

leading to

$$\begin{aligned} (Bu^{(\imath)}, u^{(\imath)}) =& (\lambda^{\imath+1} z + \lambda^{\imath} Br^{(\imath)}, \lambda^{\imath}(z + r^{(\imath)})) \\ =& \lambda^{2\imath+1}(z, z) + \lambda^{2\imath+1} \underbrace{(z, r^{(\imath)})}_{=0} + \lambda^{2\imath} \underbrace{(Br^{(\imath)}, z)}_{=0} + \lambda^{2\imath} \underbrace{(Br^{(\imath)}, r^{(\imath)})}_{\geq 0} \\ \geq& \lambda^{2\imath+1} \|z\|^2, \quad \imath = 0, 1, 2, \ldots \end{aligned}$$

so that we have the lower estimate

$$\lambda^{2\imath+1} \|z\|^2 \leq (Bu^{(\imath)}, u^{(\imath)}), \quad \imath = 0, 1, 2, \ldots.$$

Further,

$$Bu^{(\imath)} = BB^{\imath} u = \sum_{k=1}^{n} \lambda_k^{\imath} (u, w_k) Bw_k = \sum_{k=1}^{n} \lambda_k^{\imath+1} (u, w_k) w_k, \; \imath = 0, 1, 2, \ldots$$

so that with $\lambda_k \leq \lambda$,

$$(Bu^{(\imath)}, u^{(\imath)}) = \sum_{k=1}^{n} \lambda_k^{2\imath+1} |(u, w_k)|^2 \leq \lambda \sum_{k=1}^{n} \lambda_k^{2\imath} |(u, w_k)|^2 = \lambda \, \|u^{(\imath)}\|^2, \quad \imath = 0, 1, 2, \ldots.$$

Thus we get, indeed, the upper bound

$$(Bu^{(\imath)}, u^{(\imath)}) = \sum_{k=1}^{n} \lambda_k^{2\imath+1} |(u, w_k)|^2 \leq \lambda \, \|u^{(\imath)}\|^2, \quad \imath = 0, 1, 2, \ldots.$$

$\square$

From the upper bound in Lemma 3.1, we see that

$$0 \leq \lambda - \frac{(Bu^{(\imath)}, u^{(\imath)})}{\|u^{(\imath)}\|^2} = \lambda - \varrho(u^{(\imath)}) = \lambda - \varrho(v^{(\imath)}), \quad \imath = 0, 1, 2, \ldots,$$

i.e.,

$$0 \leq \lambda - \varrho(v^{(\imath)}), \quad \imath = 0, 1, 2, \ldots.$$

In the next lemma, we estimate the right-hand side further from above.

Lemma 3.2 (Estimate from above on $0 \le \lambda - \varrho(v^{(\imath)}),\ \imath = 0,1,2,\dots$). *For* $0 \le \lambda - \varrho(v^{(\imath)}),\ \imath = 0,1,2,\dots$, *we obtain the upper estimate*

$$0 \le \lambda - \varrho(v^{(\imath)}) \le \lambda\, q^{2\imath} \frac{\|r^{(0)}\|^2}{\|z\|^2}, \quad \imath = 0,1,2,\dots. \tag{3.15}$$

Proof. We depart from the expression for $(Bu^{(\imath)}, u^{(\imath)})/\|u^{(\imath)}\|^2$ yielding with $u^{(\imath)} = \lambda^\imath(z + r^{(\imath)})$,

$$\begin{aligned}\frac{(Bu^{(\imath)}, u^{(\imath)})}{\|u^{(\imath)}\|^2} &= \frac{(B[\,\lambda^\imath(z+r^{(\imath)})\,], \lambda^\imath[z+r^{(\imath)}])}{(\lambda^\imath(z+r^{(\imath)}), \lambda^\imath(z+r^{(\imath)}))} = \frac{(B(z+r^{(\imath)}), z+r^{(\imath)})}{(z+r^{(\imath)}, z+r^{(\imath)})}\\ &= \frac{(\lambda\, z + Br^{(\imath)}, z+r^{(\imath)})}{(z+r^{(\imath)}, z+r^{(\imath)})}, \quad \imath = 0,1,2,\dots.\end{aligned}$$

Now, $z \perp r^{(\imath)}$ leading to

$$(z+r^{(\imath)}, z+r^{(\imath)}) = \|z+r^{(\imath)}\|^2 = \|z\|^2 + \|r^{(\imath)}\|^2 \ge \|z\|^2$$

and

$$(\lambda z + Br^{(\imath)}, z + r^{(\imath)}) = \lambda\,\|z\|^2 + (Br^{(\imath)}, r^{(\imath)}).$$

Hereby,

$$\frac{(Bu^{(\imath)}, u^{(\imath)})}{\|u^{(\imath)}\|^2} = \frac{\lambda\,\|z\|^2 + (Br^{(\imath)}, r^{(\imath)})}{\|z\|^2 + \|r^{(\imath)}\|^2}.$$

This yields

$$\begin{aligned}0 \le& \lambda - \varrho(v^{(\imath)}) = \lambda - \varrho(u^{(\imath)}) = \lambda - \frac{(Bu^{(\imath)}, u^{(\imath)})}{\|u^{(\imath)}\|^2} = \lambda - \frac{\lambda\,\|z\|^2 + (Br^{(\imath)}, r^{(\imath)})}{\|z\|^2 + \|r^{(\imath)}\|^2}\\ =& \frac{\lambda\,[\,\|z\|^2 + \|r^{(\imath)}\|^2\,] - [\,\lambda\,\|z\|^2 + (Br^{(\imath)}, r^{(\imath)})\,]}{\|z\|^2 + \|r^{(\imath)}\|^2} = \frac{\lambda\,\|r^{(\imath)}\|^2 - (Br^{(\imath)}, r^{(\imath)})}{\|z\|^2 + \|r^{(\imath)}\|^2}\\ \le& \frac{\lambda\,\|r^{(\imath)}\|^2}{\|z\|^2 + \|r^{(\imath)}\|^2} \le \lambda\,\frac{\|r^{(\imath)}\|^2}{\|z\|^2} \le \lambda\, q^{2\imath}\frac{\|r^{(0)}\|^2}{\|z\|^2}, \quad \imath = 0,1,2,\dots.\end{aligned}$$

□

Accordingly, in the power method or the von Mises method, for initial vectors $u^{(0)}$ such that $u^{(1)} = Bu^{(0)} \ne 0$, we determine the sequence of corresponding normalized iterated vectors. These are calculated from the recurrence relation

$$v^{(0)} = \frac{u^{(0)}}{\|u^{(0)}\|}$$

and

$$v^{(\imath+1)} = \frac{1}{\|Bv^{(\imath)}\|}\, Bv^{(\imath)}, \quad \imath = 0,1,2,\dots \tag{3.16}$$

that follows immediately from the equation

$$v^{(\imath+1)} = \frac{u^{(\imath+1)}}{\|u^{(\imath+1)}\|} = \frac{Bu^{(\imath)}}{\|Bu^{(\imath)}\|} = \frac{Bv^{(\imath)}}{\|Bv^{(\imath)}\|}, \quad \imath = 0,1,2,\dots$$

The denominators $\|Bv^{(\imath)}\|$ are always non-zero due to (3.13) and $z \ne 0$. From the convergence of the normalized iterated vectors to the eigenvector $z/\|z\|$ of B corresponding to the eigenvalue $\lambda > 0$, we get the convergence

$$\|Bv^{(\imath)}\| \to \lambda \quad (\imath \to \infty)$$

since $\|Bw\| = |\lambda|\,\|w\| = \lambda$, and so the numbers $\|Bv^{(\imath)}\|$ also determine approximations to the eigenvalue λ. From these derivations, we have the following theorem.

Theorem 3.6 (Convergence theorem with a priori error estimates). *Let the matrix $B = (b_{jk}) \in \mathbb{R}^{n\times n}$ be symmetric with non-negative eigenvalues, $H = \mathbb{R}^n$ be endowed with the ordinary scalar product $(\cdot,\cdot) = (\cdot,\cdot)_2$, the eigenvalues λ_j, $j = 1,\ldots,n$ be arranged according to*

$$\lambda_1 \geq \lambda_2 \geq \ldots, \geq \lambda_n \geq 0,$$

and let the associated mutually orthonormal eigenvectors be

$$w_1, w_2, \ldots, w_n.$$

Then, for every initial vector $u = u^{(0)}$ such that $Bu^{(0)} \neq 0$, the corresponding sequence (3.16) of iterated vectors $v^{(\imath)}$ converges to the eigenvector $w = z/\|z\|$ of B corresponding to the largest eigenvalue $\lambda = \lambda_m > 0$ satisfying the condition (3.10). The following a priori error estimate holds:

$$\|v^{(\imath)} - w\| \leq \gamma q^{\imath}, \quad \imath = 0,1,2,\ldots \tag{3.17}$$

where $0 \leq q \leq \dfrac{\lambda'}{\lambda} < 1$ and $\gamma = \|r^{(0)}\|/\|z\|$. The sequence of corresponding Rayleigh quotients $\varrho(v^{(\imath)})$ converges to the eigenvalue λ with the a priori error estimate

$$0 \leq \frac{1}{\lambda}(\lambda - \varrho(v^{(\imath)}) \leq \gamma^2\, q^{2\imath}, \quad \imath = 0,1,2,\ldots. \tag{3.18}$$

Proof. The relation (3.17) is obtained as follows. We have

$$\|v^{(\imath)} - w\|^2 = 2 - \frac{1}{\|z\|}\,[(v^{(\imath)},z) + (z,v^{(\imath)})] = 2 - \frac{2}{\|z\|}\,(v^{(\imath)},z) \quad \imath = 0,1,2,\ldots.$$

Now, since $z \perp r^{(\imath)}$,

$$(v^{(\imath)},z) = \frac{(u^{(\imath)},z)}{\|u^{(\imath)}\|} = \frac{(z + r^{(\imath)},z)}{\|z + r^{(\imath)}\|} = \frac{\|z\|^2}{\|z + r^{(\imath)}\|}, \quad \imath = 0,1,2,\ldots.$$

This leads to

$$\begin{aligned}\|v^{(\imath)} - w\|^2 =& 2 - \frac{2}{\|z\|}\,\frac{\|z\|^2}{\|z + r^{(\imath)}\|} = 2 - \frac{2\,\|z\|}{\|z + r^{(\imath)}\|} = \frac{2\,\|z + r^{(\imath)}\| - 2\,\|z\|}{\|z + r^{(\imath)}\|}\\ \leq& \frac{2}{\|z\|}\,(\|z + r^{(\imath)}\| - \|z\|) = \frac{2}{\|z\|}\,\frac{\|z + r^{(\imath)}\|^2 - \|z\|^2}{\|z + r^{(\imath)}\| + \|z\|} \leq \frac{2}{\|z\|}\,\frac{\|r^{(\imath)}\|^2}{2\,\|z\|}\end{aligned}$$

for $\imath = 0,1,2,\ldots$. Hence, by means of (3.14), the a priori inequality follows

$$\|v^{(\imath)} - w\|^2 \leq \frac{\|r^{(\imath)}\|^2}{\|z\|^2} \leq q^{2\,\imath}\,\frac{\|r^{(0)}\|^2}{\|z\|^2}, \quad \imath = 0,1,2,\ldots.$$

This delivers (3.17). The estimate (3.18) follows directly from (3.15). $\square$

Since $\|v^{(\imath)}\| = 1$, the squares of the eigenvalue approximations $\|Bv^{(\imath)}\|$ are the Rayleigh quotients of the matrix B^2,

$$\|Bv^{(\imath)}\|^2 = (Bv^{(\imath)}, Bv^{(\imath)}) = (B^2 v^{(\imath)}, v^{(\imath)}) = \frac{(B^2 v^{(\imath)}, v^{(\imath)})}{\|v^{(\imath)}\|^2}, \quad \imath = 0,1,2,\ldots.$$

Here, it is of interest that these eigenvalue approximations form a monotone sequence as may be seen from this expression by using the Schwarz inequality

$$
\begin{aligned}
(3.19)\qquad \|Bv^{(\imath)}\| =&\frac{\|Bv^{(\imath)}\|^2}{\|Bv^{(\imath)}\|} = \frac{(Bv^{(\imath)}, Bv^{(\imath)})}{\|Bv^{(\imath)}\|} = \frac{(B^2v^{(\imath)}, v^{(\imath)})}{\|Bv^{(\imath)}\|} = \frac{(Bv^{(\imath+1)}, v^{(\imath)})}{\|v^{(\imath+1)}\|} \\
=&(Bv^{(\imath+1)}, v^{(\imath)}) \leq \|Bv^{(\imath+1)}\|, \quad \imath = 0, 1, 2, \ldots.
\end{aligned}
$$

In Section 5, we apply the power method to positive definite matrices B so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. Additionally, we choose $u^{(0)} = u$ such that $(u, w_1) \neq 0$ in (3.10) so that $\lambda_m = \lambda_1$. The further eigenvalues $\lambda_j > 0,\ j = 2, \ldots, n$ are determined by the power method combined with the method of deflation. The last one is discussed in the next section.

4. Method of Deflation

Let $B \in \mathbb{R}^{n\times n}$ be a positive definite matrix with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ arranged such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$, and let $w_1, w_2, \ldots, w_n$ be an associated orthonormal system of real eigenvectors.

Most important in the method of deflation is the use of projection operators $P_j \in \mathbb{R}^{n\times n},\ j = 1, \ldots, n$ in the vector space $H = \mathbb{R}^n$ endowed with the ordinary scalar product $(\cdot,\cdot) = (\cdot,\cdot)_2$. For this, we depart from the eigenvector expansions

$$
\begin{aligned}
u &= \sum_{j=1}^{n} (u, w_j) w_j,\ u \in H, \\
Bu &= \sum_{j=1}^{n} \lambda_j (u, w_j) w_j,\ u \in H.
\end{aligned}
$$

Next, we rewrite the expressions $(u, w_j)w_j,\ j = 1, \ldots, n$ in the form $(u, w_j)w_j = P_j u,\ j = 1, \ldots, n$ so that

$$
\begin{aligned}
u &= \sum_{j=1}^{n} P_j u,\ u \in H, \\
Bu &= \sum_{j=1}^{n} \lambda_j P_j u,\ u \in H,
\end{aligned}
$$

and use the obtained projection operators (i.e. $n\times n$- projection matrices) $P_j \in \mathbb{R}^{n\times n}$ to construct from $B = B_0$ and with the obtained projections P_j the deflated matrices $B_j = B_{j-1} - \lambda_j P_j,\ j = 1, 2, \ldots, n$ where it will be shown that B_j has the eigenvalues $\lambda_k > 0,\ k = j+1, \ldots, n$ and $\lambda_k = 0,\ k = 1, \ldots, j$. Now, the details follow.

Let $B = (B_{jk}) \in \mathbb{R}^{n\times n}$ be a positive definite matrix with eigenvalues $\lambda_j > 0,\ j = 1, \ldots, n$ in descending order and associated orthonormal eigenvectors that can be chosen real as we know from Theorem 2.1. Further, we have the above eigenvector expansions

$$
\begin{aligned}
u &= \sum_{j=1}^{n} (u, w_j) w_j,\ u \in H, \\
Bu &= \sum_{j=1}^{n} \lambda_j (u, w_j) w_j,\ u \in H.
\end{aligned}
$$

Next, let w_j^T be the transpose corresponding to the column vector w_j for $j = 1, \ldots, n$. Herewith, we rewrite $(u, w_j)w_j$ in the form

$$(u, w_j)w_j = (w_j^T u)w_j = w_j(w_j^T u) = (w_j w_j^T)u, \; u \in H$$

for $j = 1, \ldots, n$. The matrices

$$\boxed{P_j = w_j w_j^T \in \mathbb{R}^{n \times n}, \; j = 1, \ldots, n}$$

are projection operators with the property

$$P_j P_k = \delta_{jk} P_k, \; j, k = 1, \ldots, n.$$

Further, we have

$$u = \sum_{j=1}^{n} (u, w_j)w_j = \sum_{j=1}^{n} (P_j u) = (\sum_{j=1}^{n} P_j)u, \quad u \in H$$

or

$$\boxed{I = \sum_{j=1}^{n} P_j}$$

as well as

$$Bu = \sum_{j=1}^{n} \lambda_j (u, w_j)w_j = \sum_{j=1}^{n} \lambda_j (P_j u) = (\sum_{j=1}^{n} \lambda_j P_j)u, \quad u \in H$$

or

$$\boxed{B = \sum_{j=1}^{n} \lambda_j P_j.}$$

This expression is the starting point for the method of deflation which we take essentially from [5, Section 10.1.2, last paragraph on p. 185]. But, we restrict ourselves to the ordinary scalar product $(\cdot, \cdot) = (\cdot, \cdot)_2$ as opposed to the weighted scalar product $(u, v) = (Cu, v)_2, \; u, v \in \mathbb{R}^n$ with positive definite $C \in \mathbb{R}^{n \times n}$ applied there.

The derivation of the method of deflation is as follows: We rewrite the projection expansion of B in the form

$$B - \lambda_1 P_1 = \sum_{j=2}^{n} P_j.$$

With the deflated matrix $B_1 = B - \lambda_1 P_1$, we obtain

$$B_1 = \sum_{j=2}^{n} P_j.$$

Then, we get

$$B_1 w_1 = (B - \lambda_1 P_1)w_1 = Bw_1 - \lambda_1 w_1 = 0.$$

This means that the deflated matrix B_1 has the eigenvalues $\lambda_j > 0, j = 2, \ldots, n, \; \lambda_1 = 0$. Further, for the null space $N(B_1)$, we obtain $N(B_1) = [w_1]$. In the next step, we form the deflated matrix B_2 defined by

$$B_2 = \sum_{j=3}^{n} \lambda_j P_j = \sum_{j=2}^{n} \lambda_j P_j - \lambda_2 P_2 = B_1 - \lambda_2 P_2.$$

Then,

$$B_2 w_k = \sum_{j=3}^{n} \lambda_j P_j w_k = \lambda_j \delta_{jk} w_k,\ j,k = 1,\ldots,n$$

which means that the deflated matrix B_2 has the eigenvalues $\lambda_j > 0,\ j = 3,\ldots,n,\ \lambda_1 = 0,\ \lambda_2 = 0$. Further, for the null space $N(B_2)$, we obtain $N(B_2) = [w_1, w_2]$.

Generalizing this process, we determine the deflated matrix B_m defined by

$$B_m = \sum_{j=m+1}^{n} \lambda_j P_j = \sum_{j=m}^{n} \lambda_j P_j - \lambda_m P_m = B_{m-1} - \lambda_m P_m.$$

Then,

$$B_m w_k = \sum_{j=m+1}^{n} \lambda_j P_j w_k = \lambda_j \delta_{jk} w_k,\ j,k = 1,\ldots,n$$

which means that the deflated matrix B_m has the eigenvalues $\lambda_j > 0,\ j = m+1,\ldots,n,\ \lambda_1 = 0,\ldots,\lambda_m = 0$. Further, for the null space $N(B_m)$, we obtain $N(B_m) = [w_1,\ldots,w_m]$. For $m = n-1$, $B_m = B_{n-1}$ has the eigenvalues $\lambda_n > 0,\ \lambda_1 = 0,\ldots,\lambda_{n-1} = 0$. Further, $N(B_m) = B_{n-1} = [w_1,\ldots,w_{n-1}]$.

In the special case $m = n$, all eigenvalues of B_n are equal to zero. Further, $N(B_n) = [w_1,\ldots,w_n] = H = \mathbb{R}^n$.

5. Application to Toeplitz Sinc Matrices for $t = 0.1$

In Section 2, we considered symmetric matrices $B \in \mathbb{R}^{n\times n}$ with non-negative eigenvalues and an associated set of orthonormal eigenvectors. In Section 3, we studied symmetric matrices $B \in \mathbb{R}^{n\times n}$ with positive eigenvalues, i.e., real positive definite matrices.

In this section, we choose for $B \in \mathbb{R}^{n\times n}$ even more special matrices, namely Toeplitz sinc matrices $A_n(t) = A(t,n)$ that are real, positive definite and depend on the time variable t in $0 < t < 1$ as well as on $n \in \mathbb{N}$ with $n \geq 2$.

As we shall see later, the power method combined with the method of deflation will deliver the eigenvalues and associated eigenvectors, but not in descending order. However, in a subsequent step, the eigenvalues are sorted such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0,$$

i.e., in the desired descending order. At this point, we mention that the eigenvalues $\lambda_j > 0,\ j = 1,\ldots,n$ determined by the applied numerical procedure turn out to be all distinct, that is, one obtains

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$$

for $t = 0.1$ and $n = 5$, $n = 10$, $n = 15$ where we used as numerical procedure the power method combined with the method of deflation as well as for comparison reasons the Matlab program $eig.m$.

Further, we note that the eigenvalues and corresponding eigenvectors will be computed with a prescribed precision. Those eigenvalues in descending order that meet the prescribed precision criterion are denoted by $\lambda_j,\ j = 1,\ldots,i_q$, and the associated eigenvectors are also exact in the above sense. The inexact eigenvalues $\lambda_j,\ j = i_q+1,\ldots,n$ are replaced by $\lambda_j = 0,\ j = i_q+1,\ldots,n$, and associated eigenvectors can easily be provided.

By the prescribed precision, we mean that the computational results are exact in the 5 digits of the Matlab format shortE given by $x.xxxxe-006$. To obtain this, we form the quotient vector q of the Matlab floating-point relative accuracy $eps = 2.2204e-016$ and of the eigenvalue vector λ (in Matlab: $q = eps./\lambda$) whose components are sorted in descending order and are computed by the power method combined with the method of deflation. Then, we determine the index i_q as the number of the components $q(j)$ of q, such that $q(j-1) < 0.5e-0.006$ and $q(j) > 0.5e-0.006$ for $j = 2, \ldots, n$.

As a consequence, i_q is determined such that λ_j, $j = 1, \ldots, i_q$ in the Matlab format shortE $x.xxxxe{-}006$ with its 5 significant digits x (1 digit before the decimal point and 4 digits after the decimal point) are exact to the 5 displayed digits whereas the eigenvalues λ_j, $j = i_q + 1, \ldots, n$ are less than λ_{i_q} and can thus be set equal to zero.

In order to verify this result, we compute the eigenvalues and eigenvectors also by the Matlab program $eig.m$. Thereby, we obtain identical results in the 5 significant digits for the first i_q eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_{i_q}$ and the components of the associated eigenvectors which confirms that the index i_q in correctly determined. The computations are restricted to Toeplitz sinc matrices $A(t, n)$ with $t = 0.1$ and $n = 5$, $n = 10$, $n = 15$. But, it should pose no problem to carry out corresponding computations for other pairs (t, n). When in an applied Matlab program the input argument tol is required, we have chosen $tol = eps$.

Note that in Subsection 2.2, we have assumed that $\lambda_j = 0$, $j = m+1, \ldots, n$, whereas in this section we set $\lambda_j = 0$, $j = m+1, \ldots, n$ with $m = i_q$ due to the exactness demand. Further, the equality sign in this section - depending on the context - may mean only equality up to the exactness criterion in the above-mentioned sense. The largest eigenvalue of the limit matrix $A = \lim\limits_{\substack{0<t<1 \\ t\to 0}} A(t) = \lim\limits_{\substack{0<t<1 \\ t\to 0}} A(t,n)$ is denoted by $\mu_1 = \mu_1(A)$ as well as its normed eigenvector by w, and we write out the projection matrices P_j, $j = 1, \ldots, n$ only for $n = 5$. The computed matrices $A(t, n)$ and the corresponding computed eigenvalues $\mu_j(t)$, $j = 1, \ldots, n$ are denoted by B and λ_j, $j = 1, \ldots, n$, as the case may be.

5.1. **The Case $n = 5$.** For $t = 0.1$ and $n = 5$, we obtain

$$
\begin{aligned}
A &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{5\times 5}, \\
e &= [1 \quad 1 \quad 1 \quad 1 \quad 1]^T \in \mathbb{R}^5, \\
\mu_1 &= \mu_1(A) = 5, \\
w &= (1/\sqrt{n})\, e,
\end{aligned}
$$

leading to

$$
w = \begin{bmatrix} 4.4721e-001 \\ 4.4721e-001 \\ 4.4721e-001 \\ 4.4721e-001 \\ 4.4721e-001 \end{bmatrix} \in \mathbb{R}^5,
$$

$$
B = \begin{bmatrix}
1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\
9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\
9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\
8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\
7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000
\end{bmatrix} \in \mathbb{R}^{5\times 5}.
$$

The vector λ of eigenvalues in descending order computed by the Power Method combined with the Method of Deflation is given by

$$\lambda = \quad [4.6893e+000, \quad 3.0768e-001, \quad 3.0357e-003, \quad 8.8969e-006, \quad 8.0081e-009] \quad \in \mathbb{R}^5,$$

and the corresponding vector λ computed by the Matlab routine $eig.m$ reads also

$$\lambda = \quad [4.6893e+000, \quad 3.0768e-001, \quad 3.0357e-003, \quad 8.8969e-006, \quad 8.0081e-009] \quad \in \mathbb{R}^5.$$

The power method combined with the method of deflation delivers further

$$W = \begin{bmatrix} 4.3259e-001 & -6.2863e-001 & 5.4578e-001 & -3.2377e-001 & 1.2244e-001 \\ 4.5424e-001 & -3.2377e-001 & -2.5245e-001 & 6.2863e-001 & -4.7952e-001 \\ 4.6160e-001 & \underline{-4.9285e-012} & -5.2610e-001 & \underline{3.1519e-013} & 7.1424e-001 \\ 4.5424e-001 & 3.2377e-001 & -2.5245e-001 & -6.2863e-001 & -4.7952e-001 \\ 4.3259e-001 & 6.2863e-001 & 5.4578e-001 & 3.2377e-001 & 1.2244e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5}$$

with

$$W = [w_1,\, w_2,\, w_3,\, w_4,\, w_5]$$

where

$$w_j \in \mathbb{R}^5,\ j = 1, \ldots, n = 5$$

are the eigenvectors of B and where

$$P_j = w_j\, w_j^T,\ j = 1, \ldots, n = 5$$

are the associated matrix projections. The $eig.m$-routine delivers

$$W = \begin{bmatrix} 4.3259e-001 & -6.2863e-001 & 5.4578e-001 & 3.2377e-001 & -1.2244e-001 \\ 4.5424e-001 & -3.2377e-001 & -2.5245e-001 & -6.2863e-001 & 4.7952e-001 \\ 4.6160e-001 & \underline{1.2295e-015} & -5.2610e-001 & \underline{2.5060e-012} & -7.1424e-001 \\ 4.5424e-001 & 3.2377e-001 & -2.5245e-001 & 6.2863e-001 & 4.7952e-001 \\ 4.3259e-001 & 6.2863e-001 & 5.4578e-001 & -3.2377e-001 & -1.2244e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5}.$$

Comparing the results, setting the underlined (negligible) entries equal to zero, some normed eigenvectors $w_j,\ j = 2, \ldots, 5$ differ only in the factor -1. Therefore, both computed vectors $w_j,\ j = 1, \ldots, (n = 5)$ are equal as eigenvectors. Written out, we obtain the projection operators

$$P_1 = \begin{bmatrix} 1.8713e-001 & 1.9650e-001 & 1.9968e-001 & 1.9650e-001 & 1.8713e-001 \\ 1.9650e-001 & 2.0633e-001 & 2.0968e-001 & 2.0633e-001 & 1.9650e-001 \\ 1.9968e-001 & 2.0968e-001 & 2.1307e-001 & 2.0968e-001 & 1.9968e-001 \\ 1.9650e-001 & 2.0633e-001 & 2.0968e-001 & 2.0633e-001 & 1.9650e-001 \\ 1.8713e-001 & 1.9650e-001 & 1.9968e-001 & 1.9650e-001 & 1.8713e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5},$$

$$P_2 = \begin{bmatrix} 3.9518e-001 & 2.0353e-001 & 3.0982e-012 & -2.0353e-001 & -3.9518e-001 \\ 2.0353e-001 & 1.0482e-001 & 1.5957e-012 & -1.0482e-001 & -2.0353e-001 \\ 3.0982e-012 & 1.5957e-012 & 2.4290e-023 & -1.5957e-012 & -3.0982e-012 \\ -2.0353e-001 & -1.0482e-001 & -1.5957e-012 & 1.0482e-001 & 2.0353e-001 \\ -3.9518e-001 & -2.0353e-001 & -3.0982e-012 & 2.0353e-001 & 3.9518e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5},$$

$$P_3 = \begin{bmatrix} 2.9788e-001 & -1.3778e-001 & -2.8714e-001 & -1.3778e-001 & 2.9788e-001 \\ -1.3778e-001 & 6.3732e-002 & 1.3282e-001 & 6.3732e-002 & 1.3778e-001 \\ -2.8714e-001 & 1.3282e-001 & 2.7678e-001 & 1.3282e-001 & -2.8714e-001 \\ -1.3778e-001 & 6.3732e-002 & 1.3282e-001 & 6.3732e-002 & -1.3778e-001 \\ 2.9788e-001 & -1.3778e-001 & -2.8714e-001 & -1.3778e-001 & 2.9788e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5},$$

$$P_4 = \begin{bmatrix} 1.0482e-001 & -2.0353e-001 & -1.0205e-013 & 2.0353e-001 & -1.0482e-001 \\ -2.0353e-001 & 3.9518e-001 & 1.9814e-013 & -3.9518e-001 & 2.0353e-001 \\ -1.0205e-013 & 1.9814e-013 & 9.9344e-026 & -1.9814e-013 & 1.0205e-013 \\ 2.0353e-001 & -3.9518e-001 & -1.9814e-013 & 3.9518e-001 & -2.0353e-001 \\ -1.0482e-001 & 2.0353e-001 & 1.0205e-013 & -2.0353e-001 & 1.0482e-001 \end{bmatrix} \in \mathbb{R}^{5\times 5},$$

$$P_5 = \begin{bmatrix} 1.4993e-002 & -5.8714e-002 & 8.7455e-002 & -5.8714e-002 & 1.4993e-002 \\ -5.8714e-002 & 2.2994e-001 & -3.4249e-001 & 2.2994e-001 & -5.8714e-002 \\ 8.7455e-002 & -3.4249e-001 & 5.1014e-001 & -3.4249e-001 & 8.7455e-002 \\ -5.8714e-002 & 2.2994e-001 & -3.4249e-001 & 2.2994e-001 & -5.8714e-002 \\ 1.4993e-002 & -5.8714e-002 & 8.7455e-002 & -5.8714e-002 & 1.4993e-002 \end{bmatrix} \in \mathbb{R}^{5\times 5}.$$

This entails the standard eigenvector expansions for u and Bu, namely

$$\boxed{\begin{array}{l} u = \sum_{j=1}^{5} (u, w_j)\, w_j, \qquad u \in H = \mathbb{R}^5 \\ \text{and} \\ Bu = \sum_{j=1}^{5} \lambda_j\, (u, w_j)\, w_j, \quad u \in H = \mathbb{R}^5. \end{array}}$$

Correspondingly, using

$$\boxed{(u, w_j)\, w_j = (w_j^T\, u)\, w_j = w_j\, (w_j^T\, u) = (w_j\, w_j^T)\, u = P_j\, u,\ u \in H = \mathbb{R}^5, \quad j = 1, \ldots, n = 5,}$$

we obtain the standard projection-matrix expansions for the matrices I and B themselves, namely

$$\boxed{\begin{array}{l} I = \sum_{j=1}^{n=5} P_j \\ \text{and} \\ B = \sum_{j=1}^{n=5} \lambda_j\, P_j. \end{array}}$$

5.2. **The Case** $n = 10$**.** For $t = 0.1$ and $n = 10$, we obtain

$$\begin{aligned} A &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{10\times 10}, \\ e &= [1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1]^T \in \mathbb{R}^{10}, \\ \mu_1 &= \mu_1(A) = 10, \\ w &= (1/\sqrt{n})\, e, \end{aligned}$$

leading to

$$w = \begin{bmatrix} 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \\ 3.1623e-001 \end{bmatrix} \in \mathbb{R}^{10},$$

$$B = [B_l \,|\, B_r]$$

where

$$B_l = \begin{bmatrix} 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\ 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\ 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\ 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\ 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 \\ 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 \\ 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 \\ 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 \\ 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 \\ 1.0929e-001 & 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 \end{bmatrix} \in \mathbb{R}^{10\times 5},$$

$$B_r = \begin{bmatrix} 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 & 1.0929e-001 \\ 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 \\ 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 \\ 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 \\ 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 \\ 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\ 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\ 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\ 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\ 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 \end{bmatrix} \in \mathbb{R}^{10\times 5},$$

$$\begin{aligned} \lambda &= [\lambda_l \,|\, \lambda_r], \\ \lambda_l &= \ [\underline{7.8465e+000,\ 2.0423e+000,\ 1.0926e-001,\ 1.8886e-003,\ 1.5966e-005}] \ \in \mathbb{R}^5, \\ \lambda_r &= \ [\underline{7.5864e-008\, 2.1201e-010},\ 3.4535e-013,\ 1.1076e-015,\ 5.4224e-018] \ \in \mathbb{R}^5. \end{aligned}$$

$\boxed{\text{Determination of } i_q = 7 \text{ in Matlab:}}$

Form the quotient vector q of the Matlab quantity $eps = 2.2204e-016$ and of the sorted vector λ computed by the power method combined with the method of deflation. Then, determine i_q as the number of the components of q such that $q(j-1) < 0.5e-006$ and $q(j) > 0.5e-006$ for $j = 2\ldots,(n=10)$. As a consequence, i_q is determined such that $\lambda_j,\ j = 1,\ldots,i_q$ in the Matlab format shortE, given by $x.xxxxe-006$ with its 5 significant digits (1 digit before the decimal point and 4 digits after the decimal point) are exact whereas the eigenvalues $\lambda_j,\ j = i_q+1,\ldots,(n=10)$ are less than λ_{iq} as well as are considered being inexact; they therefore can be set equal to zero. The eigenvalues for $j = 1,\ldots,i_q = 7$ being exact in this sense are underlined above.

Correspondingly, the associated eigenvectors for $\lambda_j,\ j = i_q+1,\ldots,(n=10) = 8, 9, (n=10)$ are also considered being inexact. But, as we shall see below, sets of linearly independent eigenvectors associated with the eigenvalues $\lambda_j = 0,\ j = i_q+1,\dots,(n=10) = 8,9,(n=10)$ can be given without difficulty. Thus,

$$i_q = 7.$$

For comparison reasons, we apply also the Matlab routine $eig.m$ giving

$$\begin{aligned}\lambda_l &= \ [\underline{7.8465e+000,\ 2.0423e+000,\ 1.0926e-001,\ 1.8886e-003,\ 1.5966e-005}] \ \in \mathbb{R}^5,\\ \lambda_r &= \ [\underline{7.5864e-008,\ 2.1201e-010},\ 3.4521e-013,\ 2.7815e-016,\ 1.2737e-016] \ \in \mathbb{R}^5.\end{aligned}$$

Now, the columns of the following matrix U form a basis of eigenvectors of the limit matrix A (the first column is eigenvector associated with the largest eigenvalue $\mu_1 = \mu_1(A) = n = 10$, the other columns are eigenvectors associated with the eigenvalues $\mu_j = \mu_j(A) = 0,\ j = 2,\dots,(n=10)$:

$$U = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1\\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{10\times 10}.$$

Next, we replace the first $i_q = 7$ columns by the eigenvectors computed by the power method combined with the method of deflation. This leads to

$$U = [U_l \,|\, U_r]$$

where

$$U_l = \begin{bmatrix} 2.5812e-001 & 4.6675e-001 & 5.4006e-001 & -4.7703e-001 & 3.5511e-001\\ 2.9405e-001 & 3.9338e-001 & 2.2361e-001 & 1.1321e-001 & -3.9522e-001\\ 3.2291e-001 & 2.9807e-001 & -3.9039e-002 & 3.6088e-001 & -3.2820e-001\\ 3.4308e-001 & 1.8592e-001 & -2.2680e-001 & 3.3467e-001 & 4.1481e-002\\ 3.5346e-001 & 6.3176e-002 & -3.2461e-001 & 1.3187e-001 & 3.2904e-001\\ 3.5346e-001 & -6.3176e-002 & -3.2461e-001 & -1.3187e-001 & 3.2904e-001\\ 3.4308e-001 & -1.8592e-001 & -2.2680e-001 & -3.3467e-001 & 4.1481e-002\\ 3.2291e-001 & -2.9807e-001 & -3.9039e-002 & -3.6088e-001 & -3.2820e-001\\ 2.9405e-001 & -3.9338e-001 & 2.2361e-001 & -1.1321e-001 & -3.9522e-001\\ 2.5812e-001 & -4.6675e-001 & 5.4006e-001 & 4.7703e-001 & 3.5511e-001 \end{bmatrix} \in \mathbb{R}^{10\times 5},$$

$$U_r = \begin{bmatrix} -2.2686e-001 & 1.2336e-001 & 0 & 0 & 0\\ 5.0123e-001 & -4.3495e-001 & 0 & 0 & 0\\ -1.8605e-002 & 3.7956e-001 & 0 & 0 & 0\\ -3.8786e-001 & 2.3917e-001 & 0 & 0 & 0\\ -2.1569e-001 & -3.0713e-001 & 0 & 0 & 0\\ 2.1569e-001 & -3.0713e-001 & 0 & 0 & 0\\ 3.8786e-001 & 2.3917e-001 & 1.0000e+000 & 0 & 0\\ 1.8605e-002 & 3.7956e-001 & -1.0000e+000 & 1.0000e+000 & 0\\ -5.0123e-001 & -4.3495e-001 & 0 & -1.0000e+000 & 1.0000e+000\\ 2.2686e-001 & 1.2336e-001 & 0 & 0 & -1.0000e+000 \end{bmatrix} \in \mathbb{R}^{10\times 5}.$$

Gram-Schmidt orthonormalization with the last U as input and W as output delivers

$$W = [W_l \,|\, W_r]$$

where

$$W_l = \left[\begin{array}{rrrrr}
2.5812e-001 & 4.6675e-001 & 5.4006e-001 & -4.7703e-001 & 3.5511e-001 \\
2.9405e-001 & 3.9338e-001 & 2.2361e-001 & 1.1321e-001 & -3.9522e-001 \\
3.2291e-001 & 2.9807e-001 & -3.9039e-002 & 3.6088e-001 & -3.2820e-001 \\
3.4308e-001 & 1.8592e-001 & -2.2680e-001 & 3.3467e-001 & 4.1481e-002 \\
3.5346e-001 & 6.3176e-002 & -3.2461e-001 & 1.3187e-001 & 3.2904e-001 \\
3.5346e-001 & -6.3176e-002 & -3.2461e-001 & -1.3187e-001 & 3.2904e-001 \\
3.4308e-001 & -1.8592e-001 & -2.2680e-001 & -3.3467e-001 & 4.1481e-002 \\
3.2291e-001 & -2.9807e-001 & -3.9039e-002 & -3.6088e-001 & -3.2820e-001 \\
2.9405e-001 & -3.9338e-001 & 2.2361e-001 & -1.1321e-001 & -3.9522e-001 \\
2.5812e-001 & -4.6675e-001 & 5.4006e-001 & 4.7703e-001 & 3.5511e-001
\end{array}\right] \in \mathbb{R}^{10\times 5},$$

$$W_r = \left[\begin{array}{rr|rrr}
-2.2686e-001 & 1.2336e-001 & 2.0320e-002 & -4.8138e-002 & 2.7991e-002 \\
5.0123e-001 & -4.3495e-001 & -8.6248e-002 & 2.4622e-001 & -1.7692e-001 \\
-1.8605e-002 & 3.7956e-001 & 9.6871e-002 & -4.3938e-001 & 4.6964e-001 \\
-3.8786e-001 & 2.3917e-001 & 6.3933e-002 & 1.6843e-001 & -6.6518e-001 \\
-2.1569e-001 & -3.0713e-001 & -1.2715e-001 & 4.6341e-001 & 5.1796e-001 \\
2.1569e-001 & -3.0713e-001 & -2.3447e-001 & -6.3856e-001 & -1.9093e-001 \\
3.8786e-001 & 2.3917e-001 & 6.6404e-001 & 1.8419e-001 & 6.8279e-003 \\
1.8605e-002 & 3.7956e-001 & -6.2373e-001 & 1.8419e-001 & 6.8279e-003 \\
-5.0123e-001 & -4.3495e-001 & 2.7432e-001 & -1.5512e-001 & 6.8279e-003 \\
2.2686e-001 & 1.2336e-001 & -4.7907e-002 & 3.4776e-002 & -3.0424e-003
\end{array}\right] \in \mathbb{R}^{10\times 5}$$

with

$$W = [w_1,\, w_2,\, w_3,\, w_4,\, w_5,\, w_6,\, w_7|,\, w_8,\, w_9,\, w_{10}]$$

where

$$w_j \in \mathbb{R}^{10},\ j = 1, \ldots n = 10$$

and

$$P_j = w_j\, w_j^T,\ j = 1, \ldots, n = 10.$$

The $eig.m$-routine delivers

$$W = [W_l \,|\, W_r]$$

where

$$W_l = \left[\begin{array}{rrrrr}
2.5812e-001 & -4.6675e-001 & 5.4006e-001 & -4.7703e-001 & -3.5511e-001 \\
2.9405e-001 & -3.9338e-001 & 2.2361e-001 & 1.1321e-001 & 3.9522e-001 \\
3.2291e-001 & -2.9807e-001 & -3.9039e-002 & 3.6088e-001 & 3.2820e-001 \\
3.4308e-001 & -1.8592e-001 & -2.2680e-001 & 3.3467e-001 & -4.1481e-002 \\
3.5346e-001 & -6.3176e-002 & -3.2461e-001 & 1.3187e-001 & -3.2904e-001 \\
3.5346e-001 & 6.3176e-002 & -3.2461e-001 & -1.3187e-001 & -3.2904e-001 \\
3.4308e-001 & 1.8592e-001 & -2.2680e-001 & -3.3467e-001 & -4.1481e-002 \\
3.2291e-001 & 2.9807e-001 & -3.9039e-002 & -3.6088e-001 & 3.2820e-001 \\
2.9405e-001 & 3.9338e-001 & 2.2361e-001 & -1.1321e-001 & 3.9522e-001 \\
2.5812e-001 & 4.6675e-001 & 5.4006e-001 & 4.7703e-001 & -3.5511e-001
\end{array}\right] \in \mathbb{R}^{10\times 5},$$

$$W_r = \left[\begin{array}{rr|rrr}
-2.2686e-001 & -1.2336e-001 & -5.5684e-002 & -2.0045e-002 & -3.3536e-003 \\
5.0123e-001 & 4.3495e-001 & 2.8187e-001 & 1.3735e-001 & 3.2321e-002 \\
-1.8605e-002 & -3.7956e-001 & -5.0320e-001 & -3.8824e-001 & -1.3807e-001 \\
-3.8786e-001 & -2.3917e-001 & 2.3886e-001 & 5.4752e-001 & 3.4364e-001 \\
-2.1569e-001 & 3.0713e-001 & 3.2752e-001 & -2.9934e-001 & -5.4984e-001 \\
2.1569e-001 & 3.0713e-001 & -3.2760e-001 & -2.1738e-001 & 5.8704e-001 \\
3.8786e-001 & -2.3917e-001 & -2.3869e-001 & 4.9272e-001 & -4.1854e-001 \\
1.8605e-002 & -3.7956e-001 & 5.0308e-001 & -3.6461e-001 & 1.9227e-001 \\
-5.0123e-001 & 4.3495e-001 & -2.8183e-001 & 1.3140e-001 & -5.1671e-002 \\
2.2686e-001 & -1.2336e-001 & 5.5677e-002 & -1.9375e-002 & 6.1920e-003
\end{array}\right] \in \mathbb{R}^{10\times 5}.$$

Comparing the results, some normed eigenvectors $w_j,\ j = 2, \ldots, (i_q = 7)$ differ only in the factor -1. Therefore, both computed vectors $w_j,\ j = 1, \ldots, (i_q = 7)$ are equal as eigenvectors.

The eigenvectors $w_j,\ j = 1,\ldots,10$ computed by the power method combined with the method of deflation are exact in the 5 significant digital places whereas the eigenvectors $w_j,\ j = 8,9,10$ computed by the Matlab program $eig.m$ are inexact in these 5 digital places. Using the first method, this entails what could be called displayed eigenvector expansions for u and Bu, namely

$$\boxed{\begin{array}{l} u = \sum_{j=1}^{n=7} (u,w_j)\,w_j + P_{0,10}u, \quad u \in H = \mathbb{R}^{10} \\ \text{where} \\ P_{0,10}u = \sum_{j=8}^{n=10} (u,w_j)w_j, \qquad u \in H = \mathbb{R}^{10} \mapsto N(B). \\ \text{Further,} \\ Bu = \sum_{j=1}^{n=7} \lambda_j\,(u,w_j)w_j, \qquad u \in H = \mathbb{R}^{10}. \end{array}}$$

Correspondingly, using

$$\boxed{(u,w_j)\,w_j = (w_j^T\,u)\,w_j = w_j\,(w_j^T\,u) = (w_j\,w_j^T)\,u = P_j\,u,\ u \in H = \mathbb{R}^{10}, \quad j = 1,\ldots,n = 10,}$$

we obtain what could be called displayed projection-matrix expansions for the matrices I and B themselves, namely

$$\boxed{\begin{array}{l} I = \sum_{j=1}^{n=7} P_j + P_{0,10} \\ \text{where} \\ P_{0,10} = \sum_{j=8}^{n=10} P_j. \\ \text{Further,} \\ B = \sum_{j=1}^{n=7} \lambda_j\,P_j. \end{array}}$$

At this point, we want to mention that still the standard eigenvector expansions for u and Bu hold, namely

$$\boxed{\begin{array}{l} u = \sum_{j=1}^{n=10} (u,w_j)w_j, \qquad u \in H = \mathbb{R}^{10} \\ \text{and} \\ Bu = \sum_{j=1}^{n=10} \lambda_j\,(u,w_j)w_j, \quad u \in H = \mathbb{R}^{10} \end{array}}$$

as well as the standard matrix-projection expansions for the matrices I and B themselves, namely

$$\boxed{\begin{array}{l} I = \displaystyle\sum_{j=1}^{n=10} P_j \\ \text{and} \\ B = \displaystyle\sum_{j=1}^{n=10} \lambda_j \, P_j, \end{array}}$$

but they hide the interesting details.

5.3. **The Case $n = 15$.** For $t = 0.1$ and $n = 15$, we obtain

$$\begin{aligned} A =& \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{15\times 15}, \\ e =& [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]^T \in \mathbb{R}^{15}, \\ \mu_1 =& \mu_1(A) = 15, \\ w =& (1/\sqrt{n})\, e, \end{aligned}$$

leading to

$$w = \begin{bmatrix} 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \\ 2.5820e-001 \end{bmatrix} \in \mathbb{R}^{15},$$

$$B = [B_l \,|\, B_m \,|\, B_r\,]$$

where

$$B_l = \begin{bmatrix} 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\ 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\ 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\ 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\ 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 \\ 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 \\ 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 \\ 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 \\ 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 \\ 1.0929e-001 & 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 \\ 3.8982e-017 & 1.0929e-001 & 2.3387e-001 & 3.6788e-001 & 5.0455e-001 \\ -8.9421e-002 & 3.8982e-017 & 1.0929e-001 & 2.3387e-001 & 3.6788e-001 \\ -1.5591e-001 & -8.9421e-002 & 3.8982e-017 & 1.0929e-001 & 2.3387e-001 \\ -1.9809e-001 & -1.5591e-001 & -8.9421e-002 & 3.8982e-017 & 1.0929e-001 \\ -2.1624e-001 & -1.9809e-001 & -1.5591e-001 & -8.9421e-002 & 3.8982e-017 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$B_m = \begin{bmatrix} 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 & 1.0929e-001 \\ 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 \\ 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 \\ 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 \\ 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 \\ 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\ 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\ 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\ 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\ 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 \\ 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 \\ 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 & 9.3549e-001 \\ 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 & 8.5839e-001 \\ 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 & 7.5683e-001 \\ 1.0929e-001 & 2.3387e-001 & 3.6788e-001 & 5.0455e-001 & 6.3662e-001 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$B_r = \begin{bmatrix} 3.8982e-017 & -8.9421e-002 & -1.5591e-001 & -1.9809e-001 & -2.1624e-001 \\ 1.0929e-001 & 3.8982e-017 & -8.9421e-002 & -1.5591e-001 & -1.9809e-001 \\ 2.3387e-001 & 1.0929e-001 & 3.8982e-017 & -8.9421e-002 & -1.5591e-001 \\ 3.6788e-001 & 2.3387e-001 & 1.0929e-001 & 3.8982e-017 & -8.9421e-002 \\ 5.0455e-001 & 3.6788e-001 & 2.3387e-001 & 1.0929e-001 & 3.8982e-017 \\ 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 & 1.0929e-001 \\ 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 & 2.3387e-001 \\ 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 & 3.6788e-001 \\ 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 & 5.0455e-001 \\ 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 & 6.3662e-001 \\ 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 & 7.5683e-001 \\ 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 & 8.5839e-001 \\ 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 & 9.3549e-001 \\ 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 & 9.8363e-001 \\ 7.5683e-001 & 8.5839e-001 & 9.3549e-001 & 9.8363e-001 & 1.0000e+000 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$\begin{aligned} \lambda =& [\lambda_l \,|\, \lambda_m \,|\, \lambda_r], \\ \lambda_l =& \; [\underline{9.3119e+000,\ 4.9075e+000,\ 7.4593e-001,\ 3.3948e-002,\ 7.2882e-004}] \quad \in \mathbb{R}^5, \\ \lambda_m =& \; [\underline{9.2853e-006,\ 7.6801e-008,\ 4.3229e-010},\ 1.6988e-012,\ 4.1672e-015] \quad \in \mathbb{R}^5, \\ \lambda_r =& \; [3.0175e-017,\ 2.9844e-017,\ 2.9517e-017,\ 2.9194e-017,\ 2.8874e-017] \quad \in \mathbb{R}^5. \end{aligned}$$

Determination of $i_q = 8$ in Matlab:

Form the quotient vector q of the Matlab quantity $eps = 2.2204e - 016$ and of the sorted vector λ computed by the power method combined with the method of deflation. Then, determine i_q as the number of the components of q such that $q(j-1) < 0.5e - 006$ and $q(j) > 0.5e - 006$ for $j = 2\ldots,(n = 15)$. As a consequence, i_q is determined such that $\lambda_j,\ j = 1,\ldots,i_q$ in the Matlab format shortE $x.xxxxe - 006$ with its 5 significant digits (1 digit before the decimal point and 4 digits after the decimal point) are exact whereas the eigenvalues $\lambda_j,\ j = i_q + 1,\ldots,(n = 15)$ are less than λ_{iq} as well as are considered being inexact; they therefore can be set equal to zero. The eigenvalues for $j = 1,\ldots,i_q = 8$ being exact in this sense are underlined above.

Correspondingly, the associated eigenvectors for $\lambda_j,\ j = i_q + 1,\ldots,(n = 15) = 9,\ldots,(n = 15)$ are also considered being inexact. But, as we shall see below, sets of linearly independent eigenvectors associated with the eigenvalues $\lambda_j = 0,\ j = i_q + 1,\ldots,(n = 15) = 9,\ldots,(n = 15)$ can be given without difficulty. Thus,

$$i_q = 8.$$

For comparison reasons, we apply also the Matlab routine $eig.m$ giving

$$\begin{aligned}
\lambda_l &= \underline{[9.3119e+000,\ 4.9075e+000,\ 7.4593e-001,\ 3.3948e-002,\ 7.2882e-004}] \quad \in \mathbb{R}^5,\\
\lambda_m &= \underline{[9.2853e-006,\ 7.6801e-008,\ 4.3229e-010},\ 1.6991e-012,\ 4.4048e-015] \quad \in \mathbb{R}^5,\\
\lambda_r &= [5.6015e-016,\ 3.4319e-018,\ -7.6322e-017,\ -1.2156e-016,\ -2.0392e-016] \quad \in \mathbb{R}^5.
\end{aligned}$$

Now, the columns of the following matrix U form a basis of eigenvectors of the limit matrix A (the first column is eigenvector associated with the largest eigenvalue $\mu_1 = \mu_1(A) = n = 15$, the other columns are eigenvectors associated with the eigenvalues $\mu_j = \mu_j(A) = 0,\ j = 2,\ldots,(n = 15)$:

$$U = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1\\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix} \in \mathbb{R}^{15\times 15}.$$

Next, we replace the first $i_q = 8$ columns by the eigenvectors computed by the power method combined with the method of deflation. This leads to

$$U = [U_l \,|\, U_m \,|\, U_r]$$

where

$$U_l = \begin{bmatrix} 1.5416e-001 & -3.4966e-001 & 4.7804e-001 & 4.8946e-001 & 4.2678e-001 \\ 1.9085e-001 & -3.4298e-001 & 3.3329e-001 & 1.4066e-001 & -1.1587e-001 \\ 2.2573e-001 & -3.1907e-001 & 1.8468e-001 & -1.0801e-001 & -3.2307e-001 \\ 2.5709e-001 & -2.7861e-001 & 4.3562e-002 & -2.5063e-001 & -2.9188e-001 \\ 2.8334e-001 & -2.2336e-001 & -7.9298e-002 & -2.9208e-001 & -1.2684e-001 \\ 3.0314e-001 & -1.5605e-001 & -1.7448e-001 & -2.4741e-001 & 7.3476e-002 \\ 3.1545e-001 & -8.0229e-002 & -2.3468e-001 & -1.4015e-001 & 2.2874e-001 \\ 3.1963e-001 & -9.2175e-014 & -2.5528e-001 & 1.1263e-009 & 2.8656e-001 \\ 3.1545e-001 & 8.0229e-002 & -2.3468e-001 & 1.4015e-001 & 2.2874e-001 \\ 3.0314e-001 & 1.5605e-001 & -1.7448e-001 & 2.4741e-001 & 7.3476e-002 \\ 2.8334e-001 & 2.2336e-001 & -7.9298e-002 & 2.9208e-001 & -1.2684e-001 \\ 2.5709e-001 & 2.7861e-001 & 4.3562e-002 & 2.5063e-001 & -2.9188e-001 \\ 2.2573e-001 & 3.1907e-001 & 1.8468e-001 & 1.0801e-001 & -3.2307e-001 \\ 1.9085e-001 & 3.4298e-001 & 3.3329e-001 & -1.4066e-001 & -1.1587e-001 \\ 1.5416e-001 & 3.4966e-001 & 4.7804e-001 & -4.8946e-001 & 4.2678e-001 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$U_m = \begin{bmatrix} 3.3454e-001 & 2.3836e-001 & 1.5449e-001 & 0 & \\ -3.2607e-001 & -4.3382e-001 & -4.3364e-001 & 0 & 0 \\ -3.1912e-001 & -1.1578e-001 & 1.5983e-001 & 0 & 0 \\ -4.8217e-002 & 2.4844e-001 & 3.4139e-001 & 0 & 0 \\ 2.0726e-001 & 3.0528e-001 & 5.1350e-002 & 0 & 0 \\ 3.0126e-001 & 9.0343e-002 & -2.6264e-001 & 0 & 0 \\ 2.0949e-001 & -1.8209e-001 & -2.7273e-001 & 0 & 0 \\ 4.8494e-013 & -3.0115e-001 & 1.2309e-007 & 1.0000e+000 & 0 \\ -2.0949e-001 & -1.8209e-001 & 2.7273e-001 & -1.0000e+000 & 1.0000e+000 \\ -3.0126e-001 & 9.0343e-002 & 2.6264e-001 & 0 & -1.0000e+000 \\ -2.0726e-001 & 3.0528e-001 & -5.1350e-002 & 0 & 0 \\ 4.8217e-002 & 2.4844e-001 & -3.4139e-001 & 0 & 0 \\ 3.1912e-001 & -1.1578e-001 & -1.5983e-001 & 0 & 0 \\ 3.2607e-001 & -4.3382e-001 & 4.3364e-001 & 0 & 0 \\ -3.3454e-001 & 2.3836e-001 & -1.5449e-001 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$U_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1.0000e+000 & 0 & 0 & 0 & 0 \\ -1.0000e+000 & 1.0000e+000 & 0 & 0 & 0 \\ 0 & -1.0000e+000 & 1.0000e+000 & 0 & 0 \\ 0 & 0 & -1.0000e+000 & 1.0000e+000 & 0 \\ 0 & 0 & 0 & -1.0000e+000 & 1.0000e+000 \\ 0 & 0 & 0 & 0 & -1.0000e+000 \end{bmatrix} \in \mathbb{R}^{15\times 5}.$$

Gram-Schmidt orthonormalization with the last U as input and W as output delivers

$$W = [W_l \,|\, W_m \,|\, W_r]$$

where

$$W_l = \begin{bmatrix} 1.5416e-001 & -3.4966e-001 & 4.7804e-001 & 4.8946e-001 & 4.2678e-001 \\ 1.9085e-001 & -3.4298e-001 & 3.3329e-001 & 1.4066e-001 & -1.1587e-001 \\ 2.2573e-001 & -3.1907e-001 & 1.8468e-001 & -1.0801e-001 & -3.2307e-001 \\ 2.5709e-001 & -2.7861e-001 & 4.3562e-002 & -2.5063e-001 & 2.9188e-001 \\ 2.8334e-001 & -2.2336e-001 & -7.9298e-002 & -2.9208e-001 & -1.2684e-001 \\ 3.0314e-001 & -1.5605e-001 & -1.7448e-001 & -2.4741e-001 & 7.3476e-002 \\ 3.1545e-001 & -8.0229e-002 & -2.3468e-001 & -1.4015e-001 & 2.2874e-001 \\ 3.1963e-001 & -9.2166e-014 & -2.5528e-001 & 1.1263e-009 & 2.8656e-001 \\ 3.1545e-001 & 8.0229e-002 & -2.3468e-001 & 1.4015e-001 & 2.2874e-001 \\ 3.0314e-001 & 1.5605e-001 & -1.7448e-001 & 2.4741e-001 & 7.3476e-002 \\ 2.8334e-001 & 2.2336e-001 & -7.9298e-002 & 2.9208e-001 & -1.2684e-001 \\ 2.5709e-001 & 2.7861e-001 & 4.3562e-002 & 2.5063e-001 & -2.9188e-001 \\ 2.2573e-001 & 3.1907e-001 & 1.8468e-001 & 1.0801e-001 & -3.2307e-001 \\ 1.9085e-001 & 3.4298e-001 & 3.3329e-001 & -1.4066e-001 & -1.1587e-001 \\ 1.5416e-001 & 3.4966e-001 & 4.7804e-001 & -4.8946e-001 & 4.2678e-001 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$W_m = \left[\begin{array}{ccc|cc} 3.3454e-001 & 2.3836e-001 & 1.5449e-001 & 1.8810e-002 & 3.0811e-002 \\ -3.2607e-001 & -4.3382e-001 & -4.3364e-001 & -7.1290e-002 & -1.0422e-001 \\ -3.1912e-001 & -1.1578e-001 & 1.5983e-001 & 5.7138e-002 & 5.7890e-002 \\ -4.8217e-002 & 2.4844e-001 & 3.4139e-001 & 6.7868e-002 & 1.0746e-001 \\ 2.0726e-001 & 3.0528e-001 & 5.1350e-002 & -3.4973e-002 & -9.6212e-004 \\ 3.0126e-001 & 9.0343e-002 & -2.6264e-001 & -1.3299e-001 & -1.5319e-001 \\ 2.0949e-001 & -1.8209e-001 & -2.7273e-001 & -1.3676e-001 & -2.2452e-001 \\ 2.1841e-013 & -3.0115e-001 & -4.8001e-008 & 6.9414e-001 & 3.6622e-001 \\ -2.0949e-001 & -1.8209e-001 & 2.7273e-001 & -6.6148e-001 & 3.6622e-001 \\ -3.0126e-001 & 9.0343e-002 & 2.6264e-001 & 1.3542e-001 & -7.4602e-001 \\ -2.0726e-001 & 3.0528e-001 & -5.1350e-002 & 9.5257e-002 & 2.1427e-001 \\ 4.8217e-002 & 2.4844e-001 & -3.4139e-001 & 4.0423e-004 & 1.4368e-001 \\ 3.1912e-001 & -1.1578e-001 & -1.5983e-001 & -4.5697e-002 & -1.2195e-003 \\ 3.2607e-001 & -4.3382e-001 & 4.3364e-001 & 1.3928e-002 & -8.5197e-002 \\ -3.3454e-001 & 2.3836e-001 & -1.5449e-001 & 2.2594e-004 & 2.8778e-002 \end{array}\right] \in \mathbb{R}^{15\times 5},$$

$$W_r = \begin{bmatrix} 1.9648e-002 & -1.4242e-002 & -7.9476e-002 & 5.0415e-002 & -1.9917e-002 \\ 5.4351e-002 & 7.3854e-002 & 3.2376e-001 & -2.6729e-001 & 1.3593e-001 \\ 2.6320e-004 & -1.1387e-001 & -3.4982e-001 & 5.0804e-001 & -3.9528e-001 \\ 7.4893e-002 & -1.5865e-002 & -1.7421e-001 & -2.7900e-001 & 6.2895e-001 \\ 5.7608e-002 & 1.4555e-001 & 3.5567e-001 & -3.6645e-001 & -5.8037e-001 \\ 7.5675e-002 & 6.6402e-002 & 3.1861e-001 & 6.1479e-001 & 2.9670e-001 \\ 2.4707e-001 & -2.9156e-001 & -5.6428e-001 & -2.7855e-001 & -6.6574e-002 \\ 1.8374e-001 & 8.0959e-002 & 5.9911e-002 & 4.1446e-003 & 9.1637e-005 \\ 1.8374e-001 & 8.0959e-002 & 5.9911e-002 & 4.1446e-003 & 9.1637e-005 \\ 1.8374e-001 & 8.0959e-002 & 5.9911e-002 & 4.1446e-003 & 9.1637e-005 \\ 7.4713e-001 & 8.0959e-002 & 5.9911e-002 & 4.1446e-003 & 9.1637e-005 \\ 4.1956e-001 & -5.7053e-001 & 5.9911e-002 & 4.1446e-003 & 9.1637e-005 \\ 1.7197e-001 & 6.4891e-001 & -3.2177e-001 & 4.1446e-003 & 9.1637e-005 \\ 2.2900e-001 & -3.0769e-001 & 2.5379e-001 & -1.0729e-002 & 9.1637e-005 \\ 5.8597e-002 & 5.5201e-002 & -6.1827e-002 & 3.8970e-003 & -8.6567e-005 \end{bmatrix} \in \mathbb{R}^{15\times 5}$$

with

$$W = [w_1,\, w_2,\, w_3,\, w_4,\, w_5,\, w_6,\, w_7,\, w_8|,\, w_9,\, w_{10},\, w_{11},\, w_{12},\, w_{13},\, w_{14},\, w_{15}]$$

where

$$w_j \in \mathbb{R}^{15},\ j = 1, \ldots n = 15$$

and

$$P_j = w_j\, w_j^T,\ j = 1, \ldots, n = 15.$$

The $eig.m$-routine delivers

$$W_l = \begin{bmatrix} 1.5416e-001 & -3.4966e-001 & 4.7804e-001 & 4.8946e-001 & 4.2678e-001 \\ 1.9085e-001 & -3.4298e-001 & 3.3329e-001 & 1.4066e-001 & -1.1587e-001 \\ 2.2573e-001 & -3.1907e-001 & 1.8468e-001 & -1.0801e-001 & -3.2307e-001 \\ 2.5709e-001 & -2.7861e-001 & 4.3562e-002 & -2.5063e-001 & -2.9188e-001 \\ 2.8334e-001 & -2.2336e-001 & -7.9298e-002 & -2.9208e-001 & -1.2684e-001 \\ 3.0314e-001 & -1.5605e-001 & -1.7448e-001 & -2.4741e-001 & 7.3476e-002 \\ 3.1545e-001 & -8.0229e-002 & -2.3468e-001 & -1.4015e-001 & 2.2874e-001 \\ 3.1963e-001 & -7.2587e-017 & -2.5528e-001 & -1.5330e-015 & 2.8656e-001 \\ 3.1545e-001 & 8.0229e-002 & -2.3468e-001 & 1.4015e-001 & 2.2874e-001 \\ 3.0314e-001 & 1.5605e-001 & -1.7448e-001 & 2.4741e-001 & 7.3476e-002 \\ 2.8334e-001 & 2.2336e-001 & -7.9298e-002 & 2.9208e-001 & -1.2684e-001 \\ 2.5709e-001 & 2.7861e-001 & 4.3562e-002 & 2.5063e-001 & -2.9188e-001 \\ 2.2573e-001 & 3.1907e-001 & 1.8468e-001 & 1.0801e-001 & -3.2307e-001 \\ 1.9085e-001 & 3.4298e-001 & 3.3329e-001 & -1.4066e-001 & -1.1587e-001 \\ 1.5416e-001 & 3.4966e-001 & 4.7804e-001 & -4.8946e-001 & 4.2678e-001 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$W_m = \left[\begin{array}{ccc|cc} 3.3454e-001 & 2.3836e-001 & 1.5449e-001 & -9.0766e-002 & 4.7379e-002 \\ -3.2607e-001 & -4.3382e-001 & -4.3364e-001 & 3.5720e-001 & -2.4523e-001 \\ -3.1912e-001 & -1.1578e-001 & 1.5982e-001 & -3.6835e-001 & 4.2183e-001 \\ -4.8217e-002 & 2.4844e-001 & 3.4139e-001 & -1.6126e-001 & -1.2176e-001 \\ 2.0726e-001 & 3.0528e-001 & 5.1350e-002 & 2.7310e-001 & -3.5820e-001 \\ 3.0126e-001 & 9.0343e-002 & -2.6264e-001 & 2.5678e-001 & 1.4611e-001 \\ 2.0949e-001 & -1.8209e-001 & -2.7273e-001 & -1.0810e-001 & 2.9274e-001 \\ -6.6509e-012 & -3.0115e-001 & 1.8672e-007 & -3.1726e-001 & 2.5096e-002 \\ -2.0949e-001 & -1.8209e-001 & 2.7273e-001 & -1.0790e-001 & -3.3278e-001 \\ -3.0126e-001 & 9.0343e-002 & 2.6264e-001 & 2.5660e-001 & -1.2442e-001 \\ -2.0726e-001 & 3.0528e-001 & -5.1350e-002 & 2.7313e-001 & 3.4682e-001 \\ 4.8217e-002 & 2.4844e-001 & -3.4139e-001 & -1.6122e-001 & 1.2992e-001 \\ 3.1912e-001 & -1.1578e-001 & -1.5983e-001 & -3.6839e-001 & -4.2686e-001 \\ 3.2607e-001 & -4.3382e-001 & 4.3364e-001 & 3.5721e-001 & 2.4699e-001 \\ -3.3454e-001 & 2.3836e-001 & -1.5449e-001 & -9.0767e-002 & -4.7635e-002 \end{array}\right] \in \mathbb{R}^{15\times 5},$$

$$W_r = \begin{bmatrix} 1.8963e-002 & 2.2311e-003 & 9.7443e-003 & 1.4281e-002 & -2.1643e-004 \\ -1.2517e-001 & -1.6545e-002 & -5.8688e-002 & -1.0933e-001 & 4.7213e-003 \\ 3.2478e-001 & 5.6758e-002 & 1.1489e-001 & 3.4515e-001 & -4.0587e-002 \\ -3.8792e-001 & -1.2254e-001 & 5.2930e-003 & -5.4036e-001 & 1.6563e-001 \\ 1.7161e-001 & 1.8206e-001 & -3.4482e-001 & 3.1992e-001 & -3.6960e-001 \\ -5.9896e-002 & -1.5719e-001 & 4.5565e-001 & 2.5513e-001 & 4.8012e-001 \\ 3.0644e-001 & -4.1660e-002 & -7.5345e-003 & -5.2326e-001 & -3.8562e-001 \\ -3.6822e-001 & 3.7146e-001 & -4.2660e-001 & 1.8385e-001 & 2.6312e-001 \\ -4.0145e-002 & -6.0083e-001 & 1.7940e-001 & 2.1152e-001 & -2.3167e-001 \\ 2.3354e-001 & 5.5225e-001 & 3.7100e-001 & -1.8263e-001 & 8.3404e-002 \\ 1.3959e-001 & -3.2100e-001 & -4.9758e-001 & -4.6602e-002 & 2.3702e-001 \\ -4.7573e-001 & 1.1882e-001 & 2.3722e-001 & 1.2718e-001 & -4.1235e-001 \\ 3.7984e-001 & -2.7410e-002 & -2.7237e-002 & -7.0361e-002 & 2.9609e-001 \\ -1.3731e-001 & 3.9426e-003 & -1.4867e-002 & 1.7044e-002 & -1.0539e-001 \\ 1.9615e-002 & -3.4920e-004 & 4.1238e-003 & -1.5145e-003 & 1.5326e-002 \end{bmatrix} \in \mathbb{R}^{15\times 5}.$$

Comparing the results, both computed eigenvectors $w_j,\ j = 1, \ldots, (i_q = 8)$ are equal. The eigenvectors $w_j,\ j = 1, \ldots, 15$ computed by the power method combined with the method of deflation are exact in the 5 significant digital places whereas the eigenvectors $w_j,\ j = 9, \ldots, 15$ computed by the Matlab program $eig.m$ are inexact in these 5 digital places.

Using the first method, this entails what could be called displayed eigenvector expansions for u and Bu, namely

$$\boxed{\begin{array}{l} u = \displaystyle\sum_{j=1}^{n=8}(u, w_j)\, w_j + P_{0,15}u, \quad u \in H = \mathbb{R}^{15} \\ \text{where} \\ P_{0,15}u = \displaystyle\sum_{j=9}^{n=15}(u, w_j)w_j, \qquad u \in H = \mathbb{R}^{15} \mapsto N(B). \\ \text{Further,} \\ Bu = \displaystyle\sum_{j=1}^{n=8}\lambda_j(u, w_j)w_j, \qquad u \in H = \mathbb{R}^{15}. \end{array}}$$

Correspondingly, using

$$\boxed{(u, w_j)\, w_j = (w_j^T\, u)\, w_j = w_j\, (w_j^T\, u) = (w_j\, w_j^T)\, u = P_j\, u,\ u \in H = \mathbb{R}^{15}, \quad j = 1, \ldots, n = 15,}$$

we obtain what could be called displayed projection-matrix expansions for the matrices I and B themselves, namely

$$\boxed{\begin{array}{l} I = \displaystyle\sum_{j=1}^{n=8} P_j + P_{0,15} \\ \text{where} \\ P_{0,15} = \displaystyle\sum_{j=9}^{n=15} P_j. \\ \text{Further,} \\ B = \displaystyle\sum_{j=1}^{n=8}\lambda_j\, P_j. \end{array}}$$

At this point, we want to mention again that still the standard eigenvector expansions for u and Bu hold, namely

$$\boxed{\begin{array}{l} u = \displaystyle\sum_{j=1}^{n=15}(u, w_j)w_j, \qquad u \in H = \mathbb{R}^{15} \\ \text{and} \\ Bu = \displaystyle\sum_{j=1}^{n=15}\lambda_j(u, w_j)w_j, \quad u \in H = \mathbb{R}^{15} \end{array}}$$

as well as the standard matrix-projection expansions for the matrices I and B themselves, namely

$$\boxed{\begin{array}{l} I = \displaystyle\sum_{j=1}^{n=15} P_j \\ \text{and} \\ B = \displaystyle\sum_{j=1}^{n=15} \lambda_j\, P_j, \end{array}}$$

but, again, they hide the interesting details.

5.4. **Scrutiny of the Computed Eigenvectors and of Other Results.** Even when an operator $T(\kappa)$ in a finite-dimensional vector space is continuous in κ, the eigenvectors or eigenspaces are not necessarily continuous. That the eigenspaces can behave quite singularly even for a very smooth function $T(\kappa)$ is seen from an example due to Rellich. For these statements, see [3, Chapter II, §5.3, pp. 110-111]. Because of these remarks, we have to scrutinize as to whether the Power Method combined with the Method of Deflation for the Toeplitz sinc matrices $A(t) = A_n(t) = A(t,n)$ can be applied for the determination of the eigenvectors $w_j,\ j = 1,\ldots,n$. We remind the reader in this context that the eigenvalues $\lambda_j,\ j = i_q+1,\ldots,n$ are set equal to zero, that we have chosen $t = 0.1$, and that the computed orthonormal eigenvectors are denoted by $w_j,\ j = 1,\ldots,n$.

As abbreviation, we set $Dvec = A(t,n)W - W\Lambda$ which is equivalent to $A(t,n)w_j - \lambda_j w_j,\ j = 1,\ldots,n$ where the columns of W contain the computed eigenvectors and where Λ is a diagonal matrix containing in its diagonal the sorted computed eigenvalues $\lambda_1 > \lambda_1 > \cdots > \lambda_{i_q}$ and $\lambda_{i_q+1} = \ldots = \lambda_n = 0$. As further abbreviation, we set $mavec$ for the vector containing the maxima of the absolute values of the components of the columns of $Dvec$.
(i) Testing Results for $n = 5$
We obtain

$$Dvec = \begin{bmatrix} -1.1990e-014 & -1.5575e-012 & 1.7844e-012 & -1.8024e-017 & 1.1505e-016 \\ 5.7732e-015 & 7.2047e-013 & 9.1375e-013 & -4.0365e-017 & 2.4174e-017 \\ 1.1990e-014 & 1.5014e-012 & -1.0246e-014 & -2.8042e-018 & 6.2335e-017 \\ 5.3291e-015 & 7.2045e-013 & -9.3418e-013 & -1.1376e-017 & 6.5803e-017 \\ -1.1990e-014 & -1.5575e-012 & -1.8036e-012 & 7.2568e-017 & 1.2893e-016 \end{bmatrix} \in \mathbb{R}^{5\times 5},$$

and

$$mavec = [1.1990e-014,\ 1.5575e-012,\ 1.8036e-012,\ 7.2568e-017,\ 1.2893e-016] \in \mathbb{R}^5.$$

(ii) Testing Results for $n = 10$
We obtain

$$Dvec = [Dvec_l \,|\, Dvec_r]$$

where

$$Dvec_l = \begin{bmatrix} -8.8818e-016 & -5.4539e-012 & -4.7138e-012 & -1.6187e-014 & 2.1664e-014 \\ 0 & -2.2582e-012 & -3.9732e-012 & 1.7792e-014 & -4.9823e-015 \\ -4.4409e-016 & 3.9424e-013 & -3.0103e-012 & 1.4656e-014 & -1.6283e-014 \\ 0 & 2.2904e-012 & -1.8777e-012 & -1.9610e-015 & -1.5173e-014 \\ 4.4409e-016 & 3.2782e-012 & -6.3800e-013 & -1.5069e-014 & -5.9199e-015 \\ 0 & 3.2782e-012 & 6.3786e-013 & -1.5103e-014 & 6.0224e-015 \\ 0 & 2.2902e-012 & 1.8774e-012 & -1.9497e-015 & 1.5333e-014 \\ -4.4409e-016 & 3.9391e-013 & 3.0100e-012 & 1.4958e-014 & 1.6563e-014 \\ 0 & 2.2583e-012 & 3.9726e-012 & 1.7836e-014 & 5.3725e-015 \\ 0 & -5.4545e-012 & 4.7136e-012 & -1.6250e-014 & -2.1510e-014 \end{bmatrix} \in \mathbb{R}^{10\times 5},$$

$$Dvec_r = \begin{bmatrix} 1.7545e-016 & -2.2594e-016 & 1.1152e-014 & -1.5222e-014 & 1.0723e-014 \\ 1.4024e-016 & -8.8844e-017 & -5.6025e-014 & 7.6742e-014 & 1.3726e-014 \\ 1.6845e-016 & -1.0975e-016 & 1.0001e-013 & -1.3711e-013 & 4.5160e-014 \\ 2.0179e-016 & -1.8193e-016 & -4.7320e-014 & 6.5010e-014 & -9.7843e-015 \\ 2.6614e-016 & -2.0088e-016 & -6.5184e-014 & 8.9147e-014 & -1.6072e-014 \\ 2.1866e-016 & -2.7277e-016 & 6.5156e-014 & -8.9373e-014 & 3.2181e-014 \\ 2.5998e-016 & -9.2108e-017 & 4.7705e-014 & -6.5090e-014 & 2.5208e-014 \\ 1.1626e-016 & -9.5703e-017 & -1.0007e-013 & 1.3688e-013 & -3.0069e-014 \\ 1.6013e-016 & -3.7740e-017 & 5.6191e-014 & -7.6869e-014 & 2.7214e-014 \\ 1.0046e-016 & 7.4239e-017 & -1.0922e-014 & 1.5148e-014 & 1.5283e-015 \end{bmatrix} \in \mathbb{R}^{10\times 5},$$

and

$$mavec = [mavec_l \,|\, mavec_r]$$

where

$$\begin{aligned} mavec_l &= [8.8818e-016,\ 5.4545e-012,\ 4.7138e-012,\ 1.7836e-014,\ 2.1664e-014] \quad \in \mathbb{R}^5, \\ mavec_r &= [2.6614e-016,\ 2.7277e-016,\ 1.0007e-013,\ 1.3711e-013,\ 4.5160e-014] \quad \in \mathbb{R}^5. \end{aligned}$$

(iii) Testing Results for $n = 15$
We obtain

$$Dvec = [Dvec_l \,|\, Dvec_m \,|\, Dvec_r\,]$$

where

$$Dvec_l = \begin{bmatrix} -5.1070e-015 & -7.1854e-013 & 5.2358e-013 & -5.5722e-011 & -6.3905e-011 \\ -3.3307e-015 & -5.0160e-013 & 5.1312e-013 & 1.5129e-011 & -1.8365e-011 \\ -1.7764e-015 & -2.7800e-013 & 4.7687e-013 & 4.2181e-011 & 1.4103e-011 \\ 4.4409e-016 & -6.5947e-014 & 4.1565e-013 & 3.8108e-011 & 3.2723e-011 \\ 8.8818e-016 & 1.1924e-013 & 3.3241e-013 & 1.6561e-011 & 3.8135e-011 \\ 1.7764e-015 & 2.6168e-013 & 2.3104e-013 & -9.5931e-012 & 3.2303e-011 \\ 2.6645e-015 & 3.5238e-013 & 1.1718e-013 & -2.9865e-011 & 1.8299e-011 \\ 3.9968e-015 & 3.8313e-013 & -3.7470e-015 & -3.7414e-011 & -2.2524e-017 \\ 3.1086e-015 & 3.5250e-013 & -1.2390e-013 & -2.9865e-011 & -1.8299e-011 \\ 2.6645e-015 & 2.6223e-013 & -2.3784e-013 & -9.5933e-012 & -3.2303e-011 \\ 1.3323e-015 & 1.1924e-013 & -3.3883e-013 & 1.6561e-011 & -3.8136e-011 \\ 0 & -6.5059e-014 & -4.2163e-013 & 3.8108e-011 & -3.2723e-011 \\ -1.3323e-015 & -2.7711e-013 & -4.8167e-013 & 4.2181e-011 & -1.4103e-011 \\ -3.3307e-015 & -5.0049e-013 & -5.1731e-013 & 1.5129e-011 & 1.8365e-011 \\ -5.1070e-015 & -7.1765e-013 & -5.2680e-013 & -5.5722e-011 & 6.3905e-011 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$Fvec_m = \begin{bmatrix} 8.7083e-017 & -8.3483e-017 & 6.7980e-017 & 2.3852e-014 & 6.7613e-014 \\ -9.2601e-017 & -2.7581e-017 & -1.5402e-016 & -9.4089e-014 & -2.6609e-013 \\ -3.8362e-017 & -1.6132e-016 & 7.2962e-017 & 9.7183e-014 & 2.7430e-013 \\ -4.3969e-017 & 7.3537e-017 & 6.0734e-017 & 4.2218e-014 & 1.2030e-013 \\ -1.5728e-016 & -1.2473e-017 & -3.7211e-017 & -7.2109e-014 & -2.0344e-013 \\ -2.2853e-017 & 4.0571e-017 & -5.8071e-017 & -6.7177e-014 & -1.9122e-013 \\ -1.3398e-016 & 6.2083e-018 & -8.7916e-017 & 2.8681e-014 & 8.0496e-014 \\ -1.4081e-016 & 3.1602e-017 & -5.5577e-017 & 8.3283e-014 & 2.3645e-013 \\ -1.5282e-016 & -2.2218e-017 & -3.9310e-017 & 2.7934e-014 & 8.0505e-014 \\ -4.8842e-017 & -9.0342e-018 & 4.3674e-017 & -6.7428e-014 & -1.9117e-013 \\ -1.3168e-016 & 4.4989e-018 & -4.7067e-017 & -7.1451e-014 & -2.0350e-013 \\ -1.4477e-016 & -1.9973e-018 & -3.2573e-017 & 4.2429e-014 & 1.2021e-013 \\ -7.7979e-017 & -1.1565e-016 & 1.1062e-016 & 9.6018e-014 & 2.7447e-013 \\ -1.9811e-017 & -4.4871e-017 & 4.0796e-017 & -9.3506e-014 & -2.6616e-013 \\ -8.4387e-017 & -7.5298e-018 & 5.5178e-017 & 2.3768e-014 & 6.7703e-014 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

$$Dvec_r = \begin{bmatrix} 7.1290e-014 & 1.4221e-014 & -1.1300e-013 & 1.2664e-014 & -1.9473e-013 \\ -2.8071e-013 & -5.5565e-014 & 4.4363e-013 & -1.1547e-013 & -3.7196e-013 \\ 2.8887e-013 & 5.6894e-014 & -4.5845e-013 & 7.7955e-014 & -5.8111e-013 \\ 1.2738e-013 & 2.5923e-014 & -2.0083e-013 & 1.5469e-014 & -7.3921e-013 \\ -2.1427e-013 & -4.2103e-014 & 3.3878e-013 & -1.0785e-013 & -8.5831e-013 \\ -2.0239e-013 & -4.0816e-014 & 3.1769e-013 & -1.0637e-013 & -9.5018e-013 \\ 8.4299e-014 & 1.5973e-014 & -1.3563e-013 & -9.2369e-015 & -1.0035e-012 \\ 2.4998e-013 & 5.0085e-014 & -3.9545e-013 & 4.5670e-014 & -9.9696e-013 \\ 8.5917e-014 & 1.8079e-014 & -1.3499e-013 & -1.2350e-014 & -9.2147e-013 \\ -2.0172e-013 & -4.0183e-014 & 3.1761e-013 & -1.0986e-013 & -7.8879e-013 \\ -2.1570e-013 & -4.4263e-014 & 3.3751e-013 & -1.1107e-013 & -6.2147e-013 \\ 1.2677e-013 & 2.4952e-014 & -2.0178e-013 & 1.0368e-014 & -4.3230e-013 \\ 2.9114e-013 & 5.9550e-014 & -4.5806e-013 & 7.0337e-014 & -2.1132e-013 \\ -2.8162e-013 & -5.7121e-014 & 4.4202e-013 & -1.2048e-013 & 5.0226e-014 \\ 7.1755e-014 & 1.4516e-014 & -1.1374e-013 & 6.5880e-015 & 2.7011e-013 \end{bmatrix} \in \mathbb{R}^{15\times 5},$$

and

$$mavec = [mavec_l \,|\, mavec_m \,|\, mavec_r]$$

where

$$\begin{aligned} mavec_l &= [5.1070e-015,\ 7.1854e-013,\ 5.2680e-013,\ 5.5722e-011,\ 6.3905e-011] &\in \mathbb{R}^5, \\ mavec_m &= [1.5728e-016,\ 1.6132e-016,\ 1.5402e-016,\ 9.7183e-014,\ 2.7447e-013] &\in \mathbb{R}^5, \\ mavec_r &= [2.9114e-013,\ 5.9550e-014,\ 4.5845e-013,\ 1.2048e-013,\ 1.0035e-012] &\in \mathbb{R}^5. \end{aligned}$$

(iv) Evaluation of the Above Testing Results
The above testing results show that the computed eigenvectors are indeed good eigenvectors of the considered Toeplitz sinc matrices.
(v) Verification of the Eigenvector Expansions for u and Bu
The expansion formulas for u and Bu in Subsection 5.2 were tested with some random vectors u in Matlab. The results verified the respective computed expansions. The expansion formula for u means that $\mathbb{R}^{10}$ is the direct sum of the subspace spanned by $w_1, \ldots, w_7$ and that spanned by w_8, w_9, w_{10}.

Corresponding verifications for the expansion formulas in Subsections 5.1 and 5.3 are left to the reader.
(vi) Further Tests
The computed orthonormal vectors w_j, $j = i_q + 1, \ldots, n$ span the same orthogonal complement of the subspace spanned by the eigenvectors w_j, $j = 1, \ldots, i_q$ as the one spanned by the eigenvectors of the original matrix corresponding to those cutoff eigenvalues close to zero for $j = i_q+1, \ldots, n$. This is because the orthogonal complement of a subspace is unique. Therefore, we suspect that the computed orthonormal vectors might eventually span the same eigenspace as the original ones (or very closely). Some experiments were done, and the numerical results tend to agree. What was done is to construct the eigenvector matrix V with the first i_q columns from the i_q eigenvectors of the original Toeplitz sinc matrix and the last $n-i_q$ columns from the computed orthonormal vectors w_j, $j = i_q + 1, \ldots, n$. The diagonal matrix D has on the main diagonal the first i_q eigenvalues from the Toeplitz sinc matrix and then zeros for the rest. When $VDV^T - A$ was computed, the error is very small, the 2-norm of this residual is at $1e-14$. Therefore, the Gram-Schmidt process can help us reconstruct the eigenvectors corresponding to the eigenvectors close to zero.

6. Conclusions

A new approach for the computation of the eigenvalues and eigenvectors for a Toeplitz sinc matrix $A(t,n)$ is developed that circumvents the difficulties usually encountered for small values of t in $0 < t < 1$. This is important since Toeplitz sinc matrices are ubiquitous also outside the field of mathematics, namely in digital signal processing. The new approach essentially consists in computing only a limited number i_q of eigenvalues and eigenvectors by a conventional method as well as in setting the remaining very small eigenvalues equal to zero and in choosing from the many possible associated eigenvectors very simple ones. At this point we want to mention that this is different from truncating the expansions after i_q terms since then the expansion for $u \in H$ would be incorrect. The new method was scrutinized in several ways showing good agreement between theory and numerical results. The applied strategy leads to a new method for the computation of the eigenvalues and associated eigenvectors of real symmetric Toeplitz sinc matrices that delivers good results. For comparison reasons, we have computed the eigenvalues and eigenvectors also with the Matlab routine $eig.m$ showing only agreement in the first i_q results within the prescribed precision. We think that the power method combined with the method of deflation is also superior to the Matlab routine $eig.m$ in the following aspect: since the last one computes only the whole set of eigenvalues and eigenvectors, the first one is more flexible because it can compute any number of the n eigenvalues and eigenvectors, especially a small number of these quantities. It seems likely that the new method can be extended to other Toeplitz matrices or even to other classes of matrices. One can also imagine that the applied power method combined with the method of deflation may be replaced by different ones. It seems to be even possible that the new approach can be carried over to other sorts of problems.

Acknowledgments

One of the reviewers gave detailed hints for the improvement of the presentation. This is gratefully acknowledged by the authors.

References

[1] F. A. Grünbaum: *Toeplitz matrices commuting with tridiagonal matrices*, Linear Algebra Appl., **40** (1981), 25–35.
[2] D. Hertz: *Simple bounds on the extreme eigenvalues of Toeplitz matrices*, IEEE Trans. Inf. Theory, **38** (1) (1992), 175–176.
[3] T. Kato: *Perturbation theory for linear operators*, Springer-Verlag, New York (1966).
[4] L. Kohaupt, Y. Wu: *Lower estimates on the condition number of a Toeplitz sinc matrix and related questions*, Constr. Math. Anal., **5** (3) (2022), 168–182.
[5] F. Stummel, K. Hainer: *Introduction to numerical analysis*, Scottish Academic Press, Edinburgh (1980).
[6] F. Stummel, L. Kohaupt: *Eigenwertaufgaben in Hilbertschen Räumen. Mit Aufgaben und vollständigen Lösungen*, Logos Verlag, Berlin (2021).
[7] Y. Wu: *On the positiveness of a functional symmetric matrix used in digital filter design*, J. Circuits Syst. Comput., **13** (5) (2004), 1105–1110.
[8] Y. Wu, L. Kohaupt: *On the eigenvalue separation properties of real tridiagonal matrices*, Constr. Math. Anal., **6** (4) (2023), 210–221.

Ludwig Kohaupt
Berlin University of Applied Sciences and Technology (BHT)
Department of Mathematics
Luxemburger Str. 10, 13353 Berlin, Germany
Email address: lkohaupt4@web.de

YAN WU
GEORGIA SOUTHERN UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCES
65 GEORGIA AVE, STATESBORO, GA, 30460, USA
Email address: yan@georgiasouthern.edu

MODERN MATHEMATICAL METHODS
4 (2026), No. 2, pp. 119-133
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

A note concerning polyhyperbolic and related splines

JEFF LEDFORD*, LUKE PARIS, RYAN URBAN, AND ALEC VIDANES

ABSTRACT. This note concerns the interpolation problem with two parametrized families of splines related to polynomial spline interpolation. The authors address the questions of uniqueness and establish basic convergence rates for splines of the form $s_\alpha = p\cosh(\alpha\cdot) + q\sinh(\alpha\cdot)$ and $t_\alpha = p + q\tanh(\alpha\cdot)$ between the nodes where $p, q \in \Pi_{k-1}$.

Keywords: Spline interpolation, convergence rate, band limited data interpolation.

2020 Mathematics Subject Classification: 41A05, 41A15, 41A25.

1. INTRODUCTION

This note continues the study of splines, popularized by Schoenberg [15, 16, 17]. The goal of this paper is to develop basic properties of two parametrized families of splines. Specifically, we address questions of uniqueness and convergence rate similar to those found in [2, 3, 6, 13, 20]. The first family of splines satisfies $s \in C^{2k-2}[a,b]$ and

$$(D^2 - \alpha^2)^k s = 0 \text{ on } [a,b] \setminus X, \tag{1.1}$$

where $\alpha > 0$ and X is a partition of $[a,b]$. Following [9], we call these k-th order polyhyperbolic splines. These splines are examples of L-splines, [19, Ch. 10]. The hyperbolic designation was chosen because the homogeneous solution of (1.1) is given by

$$s_\alpha(x) = p(x)\cosh(\alpha x) + q(x)\sinh(\alpha x),$$

where p and q are polynomials of degree at most $k-1$. We note that, in the literature, splines corresponding to the differential equation $D^2(D^2 - \alpha^2)u = 0$ and are often called hyperbolic but they are also referred to by the terms tension or exponential. These splines have been studied extensively, see [11, 12, 13, 20] among many others. We also find a different notion for hyperbolic splines in [18], these correspond to the differential equation

$$(D^2 - (2k-1)^2)\cdots(D^2 - 3^2)(D^2 - 1)u = 0 \quad \text{or} \quad (D^2 - (2k)^2)\cdots(D^2 - 2^2)Du = 0.$$

Interpolation schemes which depend on parameters have appeared throughout the literature. Polyhyperbolic splines are similar to (but distinct from) splines in tension [13] and the GB splines of [7], although they do share the property that in the appropriate limiting sense, these schemes lead to polynomial spline interpolation. For the case studied here, we expect to recover cubic splines, since as $\alpha \to 0$, the differential operator reduces to D^4.

The second family of splines are built out of $\tanh$, specifically these are piecewise combinations of $t = p + q\tanh(\alpha\cdot)$, where p and q are polynomials. We can characterize the order these by the maximal degree of the polynomials; the k-th order splines have polynomials whose

Received: 04.09.2025; Accepted: 27.07.2026; Published Online: 31.07.2026
*Corresponding author: Jeff Ledford; ledfordjp@longwood.edu
DOI: 10.64700/mmm.78

degree does not exceed $k-1$. Owing to the popularity of $\tanh$ as an activation function in neural networks [1] and the use of splines for the same [4, 22], it's possible these splines have application in their own right. However in this note, we will concern ourselves with basic approximation properties of these splines. In principle, interpolating with either spline of order k corresponds to $2k-1$ degree polynomial spline interpolation. We will explore the problem for the values $k=1$ and $k=2$, which correspond to classical linear and cubic interpolation, respectively.

The rest of the paper is laid out as follows. Section 2 contains definitions and facts necessary to the sequel. Section 3 contains results for first order splines, while Section 4 contains those related to the second order splines. The main results are found in these sections as Proposition 3.2 and Theorem 4.2 and its corollary. B-spline bases for low degree splines are developed in Section 5. A related interpolation problem for higher degree splines is solved using a Fourier analysis approach in Section 6, whose main result is Theorem 6.4.

2. Definitions and Basic Facts

Throughout the sequel, we use standard notations for derivatives. For example, D^2u and u'' both correspond to the second derivative of the function u. The set of polynomials of degree at most k is denoted Π_k. By a partition of the interval $[a,b]$, we mean a sequence of increasing values $X=(x_j : 0\leq j\leq N)$ such that $a=x_0<\cdots<x_N=b$. Associated to a fixed X, we have $h_j := x_j - x_{j-1}$ and $\overline{h}(X) := \max\limits_{1\leq j\leq N} h_j$.

We denote the set of all k-th order polyhyperbolic splines by $S_\alpha^k(X)$, that is

$$S_\alpha^k(X) := \{s\in C^{2k-2}[a,b] : (D^2-\alpha^2)^k s = 0 \text{ on } [a,b]\setminus X\}.$$

We denote by $T_\alpha^k(X)$, the collection

$$T_\alpha^k(X) := \{\text{sech}(\alpha\cdot)u : u\in S_\alpha^k(X)\}.$$

Note that on each interval $[x_j,x_{j+1}]$, $t_\alpha\in T_\alpha^k(X)$ takes the form

$$t_\alpha(x) = p(x) + q(x)\tanh(\alpha x),$$

where $p,q\in\Pi_{k-1}$.

We use the notation $\|u\|_{L^\infty}$ to denote the usual L^∞ norm for a function u, while we use the notation $\|M\|_\infty$ to mean the maximum row sum of the matrix M. A diagonally dominant matrix is an $N\times N$ square matrix $[a_{ij}]$ which satisfies $|a_{ii}|-\sum_{j\neq i}|a_{ij}|>0$ for each $1\leq i\leq N$. The well-known Levy-Desplanques theorem proves this condition implies the invertibility of $[a_{ij}]$, see [21].

Generally, we will seek to interpolate the data $Y=(y_j : 0\leq j\leq N)$ on a fixed, but otherwise arbitrary, partition X using $s_\alpha\in S_\alpha^k(X)$ or $t_\alpha\in T_\alpha^k(X)$. When we want to emphasize the dependence on the data set Y, we write $s_\alpha[Y]$ or $t_\alpha[Y]$. Finally, we note the value of the constant C will be depend on its occurrence and is typically independent of the listed parameters unless stated otherwise.

3. Properties of First Order Splines

As one may expect, things are easier when $k=1$. We include these results as more or less a warm up for the more involved $k=2$ case. When $k=1$, we have $s_\alpha = \sum_{j=1}^N s_j$, where $s_j : [x_{j-1},x_j]\to\mathbb{R}$ is given by $s_j(x) = a_j\cosh(\alpha x) + b_j\sinh(\alpha x)$. Thus, we need only solve the

2×2 matrix equation

$$\begin{bmatrix} \cosh(\alpha x_{j-1}) & \sinh(\alpha x_{j-1}) \\ \cosh(\alpha x_j) & \sinh(\alpha x_j) \end{bmatrix} \begin{bmatrix} a_j \\ b_j \end{bmatrix} = \begin{bmatrix} y_{j-1} \\ y_j \end{bmatrix}.$$

The determinant of the matrix is

$$\cosh(\alpha x_{j-1}) \sinh(\alpha x_j) - \cosh(\alpha x_j) \sinh(\alpha x_{j-1}) = \sinh(\alpha h_j).$$

Since $h_j > 0$, this system has a unique solution which after simplification yields

$$s_j(x) = \frac{\sinh(\alpha(x_j - x))}{\sinh(\alpha h_j)} y_{j-1} + \frac{\sinh(\alpha(x - x_{j-1}))}{\sinh(\alpha h_j)} y_j. \tag{3.2}$$

A similar computation yields

$$t_j(x) = \frac{\tanh(\alpha x_j) - \tanh(\alpha x)}{\tanh(\alpha x_j) - \tanh(\alpha x_{j-1})} y_{j-1} + \frac{\tanh(\alpha x) - \tanh(\alpha x_{j-1})}{\tanh(\alpha x_j) - \tanh(\alpha x_{j-1})} y_j. \tag{3.3}$$

Expanding (3.2) and (3.3) in a Taylor series (in α) yields

$$s_\alpha\big|_{[x_{j-1}, x_j]}(x) = y_{j-1} + \frac{y_j - y_{j-1}}{h_j}(x - x_{j-1}) + \mathcal{O}(\alpha^2)$$

and

$$t_\alpha\big|_{[x_{j-1}, x_j]}(x) = y_{j-1} + \frac{y_j - y_{j-1}}{h_j}(x - x_{j-1}) + \mathcal{O}(\alpha^2).$$

We recognize the first two terms as the linear interpolant of (x_{j-1}, y_{j-1}) and (x_j, y_j). Thus we expect s_α to exhibit similar convergence properties to the linear spline $l[Y]$. The expansions above give us $s_0[Y] = t_0[Y] = l[Y]$. We may also estimate the rate of convergence by expanding a sufficiently smooth function f in a Taylor series about $x = x_{j-1}$, then on (x_{j-1}, x_j) we have

$$\begin{aligned} f(x) - t_\alpha(x) &= \left(f'(x_{j-1}) - \alpha h_j(\coth(\alpha h_j) + \tanh(\alpha x_j)) f[x_{j-1}, x_j]\right)(x - x_{j-1}) + \mathcal{O}(\overline{h}^2) \\ &= (f'(x_{j-1}) - f[x_{j-1}, x_j])(x - x_{j-1}) + \mathcal{O}(\overline{h}^2) \\ &= \mathcal{O}(\overline{h}^2). \end{aligned}$$

The last equation requires $f \in C^2([a,b])$, and follows from the Mean Value Theorem. A completely analogous result holds for s_α. We summarize these findings in the following propositions.

Proposition 3.1. *Let X be a partition of $[a,b]$, for any $\alpha > 0$ and data set Y, the interpolants $s_\alpha[Y]$ and $t_\alpha[Y]$ are unique. Furthermore they satisfy*

$$\lim_{\alpha \to 0} s_\alpha[Y](x) = l[Y](x) \quad \textit{and} \quad \lim_{\alpha \to 0} t_\alpha[Y](x) = l[Y](x)$$

for all $x \in [a,b]$, where $l[Y]$ is the linear spline interpolant.

Proposition 3.2. *Suppose that $f \in C^2([a,b])$ and X is a partition of $[a,b]$. If $t_\alpha \in T^1_\alpha(X)$ such that $t_\alpha|_X = f|_X$, then for α sufficiently small, we have*

$$\|f - t_\alpha\|_{L^\infty([a,b])} \le C\overline{h}^2,$$

where $C > 0$ depends on f and α. The same is true if t_α is replaced by $s_\alpha \in S^1_\alpha(X)$.

Remark 3.1. *Rather than expanding s_α in a Taylor series, we could use the fact that for any $f \in C^2([a,b])$, there exists $g \in C^2([a,b])$ with $f(x) = \cosh(\alpha x) g(x)$. Thus*

$$\|f - s_\alpha\|_{L^\infty([a,b])} = \|\cosh(\alpha \cdot)\,(g - t_\alpha)\,\|_{L^\infty([a,b])} \le C\overline{h}^2$$

for sufficiently small α.

4. Properties of Second Order Splines

The $k = 2$ problem is more involved. Just as in the case with cubic splines, we must impose endpoint conditions on our splines $s_\alpha \in S_\alpha^2(X)$ (or $t_\alpha \in T_\alpha^2(X)$). For these, we use those found in [13]:

$$\begin{aligned} \text{Type I: } & s_\alpha'(a) = y_0', s_\alpha'(b) = y_N', \\ \text{Type II: } & s_\alpha''(a) = s_\alpha''(b) = 0, \\ \text{Type III: } & s_\alpha''(a) = y_0'', s_\alpha''(b) = y_N''. \end{aligned}$$

Assuming that we are sampling a function $f : [a,b] \to \mathbb{R}$, so that $y_j = f(x_j)$, conditions I and III become

$$\begin{aligned} \text{Type I: } & s_\alpha'(a) = f'(a), s_\alpha'(b) = f'(b), \\ \text{Type III: } & s_\alpha''(a) = f''(a), s_\alpha''(b) = f''(b). \end{aligned}$$

In light of the remark that follows Proposition 3.2, we find it more convenient to work with $t_\alpha \in T_\alpha^2(X)$ in what follows, often suppressing the dependence on α. Since $t_\alpha\,|_{[x_{j-1},x_j]}\,(x) = p_j(x) + q_j(x)\tanh(\alpha x)$, where $p_j, q_j \in \Pi_1$, $t_\alpha''\,|_{[x_{j-1},x_j]} = D^2[q_j \tanh(\alpha \cdot)]$. Specifying $t_\alpha''(x_{j-1})$ and $t_\alpha''(x_j)$ allow us to solve for the coefficients of q_j then integrating twice and using the interpolation conditions allow us to solve for the coefficients of p_j. One may then generate a tridiagonal system to enforce smoothness of the first derivative, setting $t_j'' := t_\alpha''(x_j)$ we have

$$b_0 t_0'' + c_0 t_1'' = d_0, \tag{4.4a}$$

$$a_j t_{j-1}'' + b_j t_j'' + c_j t_{j+1}'' = d_j, \quad 1 \le j \le N-1, \tag{4.4b}$$

$$a_N t_{N-1}'' + b_N t_N'' = d_N. \tag{4.4c}$$

One needs to evaluate $t'\,|_{[x_{j-1},x_j]}$ at both endpoints to generate the coefficients in (4.4b). Setting

$$E_{j-1} = \frac{\alpha h_j \cosh(\alpha h_j) - \sinh(\alpha h_j)}{2\alpha^2 h_j(\tanh(\alpha x_j) - \tanh(\alpha x_{j-1}) - \alpha h_j \tanh(\alpha x_{j-1})\tanh(\alpha x_j))},$$

we have for $1 \le j \le N-1$

$$\begin{aligned} a_j &= \frac{\cosh(\alpha x_{j-1})}{\cosh(\alpha x_j)} E_{j-1} = \frac{h_j}{6} + \mathcal{O}\left((\alpha \overline{h})^2\right); \alpha \to 0, \\ c_j &= \frac{\cosh(\alpha x_{j+1})}{\cosh(\alpha x_j)} E_j = \frac{h_{j+1}}{6} + \mathcal{O}\left((\alpha \overline{h})^2\right); \alpha \to 0 \end{aligned}$$

and

$$d_j = \frac{y_{j+1} - y_j}{h_{j+1}} - \frac{y_j - y_{j-1}}{h_j}.$$

The term b_j is more complicated, setting

$$F_{j-1} = 4\alpha^2 h_j(\tanh(\alpha x_j) - \tanh(\alpha x_{j-1}) - \alpha h_j \tanh(\alpha x_{j-1})\tanh(\alpha x_j)),$$

which is twice the denominator of E_{j-1}, we can simplify the expression somewhat

$$\begin{aligned}b_j &= \left(\sinh(2\alpha x_j) - 2\alpha h_j - 2\tanh(\alpha x_{j-1})(\cosh^2(\alpha x_j) - \alpha^2 h_j^2)\right)/F_{j-1}\\ &\quad - \left(\sinh(2\alpha x_j) + 2\alpha h_{j+1} - 2\tanh(\alpha x_{j+1})(\cosh^2(\alpha x_j) - \alpha^2 h_{j+1}^2)\right)/F_j\\ &= \frac{h_j + h_{j+1}}{3} + \mathcal{O}\left((\alpha\bar{h})^2\right); \alpha \to 0.\end{aligned}$$

For type I endpoint conditions, the coefficients in (4.4a) and (4.4c) are given by

$$\begin{aligned}b_0 &= -\left(\sinh(2\alpha x_0) + 2\alpha h_1 - 2\tanh(\alpha x_1)(\cosh^2(\alpha x_0) - \alpha^2 h_1^2)\right)/F_0\\ &= \frac{h_1}{3} + \mathcal{O}\left((\alpha\bar{h})^2\right); \alpha \to 0,\\ c_0 &= \frac{\cosh(\alpha x_1)}{\cosh(\alpha x_0)}E_0 = \frac{h_1}{6} + \mathcal{O}\left((\alpha\bar{h})^2\right); \alpha \to 0,\\ d_0 &= \frac{y_1 - y_0}{h_1} - y_0',\\ a_N &= \frac{\cosh(\alpha x_{N-1})}{\cosh(\alpha x_N)}E_{N-1} = \frac{h_N}{6} + \mathcal{O}\left((\alpha\bar{h})^2\right); \alpha \to 0,\\ b_N &= \left(\sinh(2\alpha x_N) - 2\alpha h_N - 2\tanh(\alpha x_{N-1})(\cosh^2(\alpha x_N) - \alpha^2 h_N^2)\right)/F_{N-1}\\ &= \frac{h_N}{3} + \mathcal{O}\left((\alpha\bar{h})^2\right); \alpha \to 0,\\ d_N &= y_N' - \frac{y_N - y_{N-1}}{h_N}.\end{aligned}$$

The adjustments for the other types are

$$\begin{aligned}\text{II}: &\quad b_0 = b_N = 1, a_0 = c_0 = 0, d_0 = d_N = 0\\ \text{III}: &\quad b_0 = b_N = 1, a_0 = c_0 = 0, d_0 = y_0'', d_N = y_N''.\end{aligned}$$

Direct computation now provides the following analog of Proposition 3.1.

Theorem 4.1. *Let X be a partition of $[a,b]$. For sufficiently small α and data set Y with end condition type I, II or III, the interpolants $s_\alpha[Y] \in S_\alpha^2(X)$ and $t_\alpha[Y] \in T_\alpha^2(X)$ are unique. Furthermore,*

$$\lim_{\alpha\to 0} s_\alpha[Y](x) = \sigma[Y](x) \quad \text{and} \quad \lim_{\alpha\to 0} t_\alpha[Y](x) = \sigma[Y](x)$$

for all $x \in [a,b]$, where $\sigma[Y]$ is the cubic spline interpolant with the same type of endpoint condition.

Proof. The matrix (4.4) for $t_\alpha \in T_\alpha^2(X)$ is diagonally dominant for sufficiently small α. Now consider $\widetilde{Y} = (\operatorname{sech}(\alpha x_j)y_j : 0 \le j \le N)$, we have a unique $\tilde{t}_\alpha \in T_\alpha^2(X)$ which interpolates $\widetilde{Y}$ and $s_\alpha = \cosh(\alpha\cdot)t_\alpha \in S_\alpha^2(X)$ interpolates Y. □

We use an argument similar to the one given in [13] to prove a result similar to Proposition 3.2. We begin by setting $\delta = \sigma - t_\alpha$ and $\delta_j'' = \delta''(x_j)$. We use (4.4) and the tridiagonal system for σ to generate a system for δ'':

(4.5a)
$$\delta_0'' + \frac{c_0}{b_0}\delta_1'' = \frac{3b_0 - h_1}{3b_0}\sigma''(x_0) + \frac{6c_0 - h_1}{6b_0}\sigma''(x_1),$$

(4.5b)
$$\frac{a_j}{b_j}\delta_{j-1}'' + \delta_j'' + \frac{c_j}{b_j}\delta_{j+1}'' = \frac{6a_j - h_j}{6b_j}\sigma''(x_{j-1}) + \frac{3b_j - h_j - h_{j+1}}{3b_j}\sigma''(x_j) + \frac{6c_j - h_{j+1}}{6b_j}\sigma''(x_{j+1}),$$

(4.5c) $$\frac{a_N}{b_N}\delta''_{N-1} + \delta''_N = \frac{6a_N - h_N}{6b_N}\sigma''(x_{N-1}) + \frac{3b_N - h_N}{3b_N}\sigma''(x_N).$$

Now, we may write this as the matrix equation $(I+M)\delta'' = b$, where δ'' represents the vector with components δ''_j and b is the vector version of the right hand side (4.5). More computation with the coefficients in (4.4) shows that $\|M\|_\infty = \mathcal{O}(\frac{1}{2})$ as $\alpha \to 0$, hence for sufficiently small α

$$\|\delta''\|_\infty = \|(I+M)^{-1}b\|_\infty \le 3\|b\|_\infty.$$

To estimate $\|b\|_\infty$, we note that the Mean Value Theorem provides the estimate for the first and last rows $\mathcal{O}\left((\alpha\overline{h})^2\right)\|\sigma'''\|_{L^\infty([a,b]\setminus X)}$, while the triangle inequality provides the estimate for the middle rows $\mathcal{O}\left((\alpha\overline{h})^2\right)\|\sigma''\|_{L^\infty(X)}$ as $\alpha \to 0$. Thus, for sufficiently small α

(4.6) $$\|\delta''\|_\infty \le C(\alpha\overline{h})^2 \max\{\|\sigma'''\|_{L^\infty([a,b]\setminus X)}, \|\sigma''\|_{L^\infty(X)}\}.$$

Now the argument follows in a manner similar to Proposition 3.2. We may use Rolle's Theorem for the function δ, which allows us to find $\xi_j \in (x_{j-1}, x_j)$ such that $\delta'(\xi_j) = 0$, this together with the fact that $\delta(x_j) = 0$ allows us to write

$$\delta\big|_{[x_{j-1},x_j]}(x) = \int_{x_j}^x \delta'(u)\mathrm{d}u \quad \text{and} \quad \delta'\big|_{[x_{j-1},x_j]}(x) = \int_{\xi_j}^x \delta''(u)\mathrm{d}u.$$

We need only estimate $\|\delta''\|_{L^\infty([a,b])}$. To this end, note that we can write things in terms of the values at the partition

$$\begin{aligned}\delta''(x) &= \sigma''(x) - t''_\alpha(x)\\ &= w_1(x)\delta''_{j-1} + w_2(x)\delta''_j + z_1(x)\sigma''_{j-1} + z_2(x)\sigma''_j,\end{aligned}$$

where

$$\begin{aligned}
w_1(x) &= \frac{4\alpha^2 h_j^2 \cosh(\alpha x_{j-1})((-1+\alpha(x_j - x)\tanh(\alpha x_j))\tanh(\alpha x) + \tanh(\alpha x_j))}{\cosh^2(\alpha x)F_{j-1}}\\
z_1(x) &= \frac{x_j - x}{h_j} - w_1(x) = \mathcal{O}\left((\alpha\overline{h})^2\right)\|\sigma'''\|_{L^\infty([a,b]\setminus X)}; \alpha \to 0\\
w_2(x) &= \frac{4\alpha^2 h_j^2 \cosh(\alpha x_j)((1+\alpha(x - x_{j-1})\tanh(\alpha x_{j-1}))\tanh(\alpha x) - \tanh(\alpha x_{j-1}))}{\cosh^2(\alpha x)F_{j-1}}\\
z_2(x) &= \frac{x - x_{j-1}}{h_j} - w_2(x).
\end{aligned}$$

Thus

$$\|\delta''\|_{L^\infty([a,b])} = \mathcal{O}\left((\alpha\overline{h})^2\right)\max\left\{\|\sigma'''\|_{L^\infty([a,b]\setminus X)}, \|\sigma''\|_{L^\infty(X)}\right\}; \alpha \to 0.$$

Putting these estimates together yields the following theorem.

Theorem 4.2. *Let X be a partition for $[a,b]$. Given a data set Y, suppose that $t_\alpha[Y] \in T^2_\alpha(X)$ interpolates Y with type I (II, III) end conditions. For $i = 0, 1,$ or 2 and sufficiently small α,*

$$\|D^i(\sigma[Y] - t_\alpha[Y])\|_{L^\infty([a,b])} \le C(\alpha\overline{h})^{4-i}\max\left\{\|\sigma'''\|_{L^\infty([a,b]\setminus X)}, \|\sigma''\|_{L^\infty(X)}\right\},$$

where $\sigma[Y]$ is the cubic spline interpolant with type I (II, III) end conditions and $C > 0$ is a constant independent of $\overline{h}$.

Corollary 4.1. *Let X be a partition for $[a,b]$ and $f \in C^4[a,b]$. Suppose that $t_\alpha[f] := t_\alpha[f(X)] \in T^2_\alpha(X)$ interpolates $f(X)$ with type I (II, III) end conditions. For $i = 0, 1,$ or 2 and sufficiently small α,*

$$\|D^i(t_\alpha[f] - f)\|_{L^\infty([a,b])} \le C\overline{h}^{4-i},$$

where $C := C_{\alpha,f,X} > 0$. The same is true for $s_\alpha[f] \in S^2_\alpha(X)$.

Proof. The result for $t_\alpha[f]$ follows from Theorem 4.2, the triangle inequality, and known convergence rates for $\sigma[f] := \sigma[f(X)]$:

$$\|D^i(\sigma[f] - f)\|_{L^\infty([a,b])} \le C_{f,X}\overline{h}^{4-i}; \quad i = 0, 1, 2$$

and

$$\|D^3(\sigma[f] - f)\|_{L^\infty([a,b])} \le C_{f,X}\overline{h}$$

found in [3] and [6], respectively. The constants depend on the norms of various derivatives of f and the *mesh ratio*, $|X| = \overline{h}/\min h_j$. To see the result for $s_\alpha[f] \in S^2_\alpha(X)$, note that any function $f \in C^4[a,b]$ may be written $f(x) = \cosh(\alpha x)g(x)$, where $g \in C^4[a,b]$ as well. We have

$$\|D^i(s_\alpha[f] - f)\|_{L^\infty([a,b])} = \|D^i(\cosh(\alpha\cdot)(t_\alpha[g] - g))\|_{L^\infty([a,b])}$$

so the result follows from the Leibniz rule and the corresponding result for $t_\alpha \in T^2_\alpha(X)$. □

Remark 4.2. *One powerful aspect of modeling with cubic splines is the ability to preserve certain qualities of the underlying data set to be interpolated. Unfortunately, $s_\alpha[Y] \in S^2_\alpha(X)$ (or $t_\alpha[Y] \in T^2_\alpha(X)$) need not share properties of Y such as positivity, monotonicity, or convexity. As seen with interpolation with cubic splines, the trade off for preserving the shape of the data is giving up some smoothness of the interpolant. If we choose to specify the first derivatives rather than enforcing continuity of the second derivative at each internal $x_j \in X$, then we can solve (1.1) in terms of the values $Y = (y_j)$ and $Y' = (y'_j)$ on the interval $[x_{j-1}, x_j]$:*

$$\begin{aligned} s_\alpha(x) &= y_j + y'_j(x - x_j) \\ &\quad + \left(\frac{3(y_{j+1} - y_j) - h_{j+1}(2y'_j + y'_{j+1})}{h^2_{j+1}} + \mathcal{O}(\alpha^2)\right)(x - x_j)^2 \\ &\quad + \left(\frac{h_{j+1}(y'_j + y'_{j+1}) - 2(y_{j+1} - y_j)}{h^3_{j+1}} + \mathcal{O}(\alpha^2)\right)(x - x_j)^3 + \mathcal{O}(\alpha^2) \\ &= \sigma(x) + \mathcal{O}(\alpha^2 h^2_{\max}). \end{aligned} \tag{4.7}$$

Here σ is the cubic Hermite spline interpolant for the same data. In order for s_α to be approximately shape preserving, we may first generate a cubic interpolant with the property. Algorithms to preserve positivity, monotonicity, and convexity for cubic splines have been well established. See, for instance, [5] and the references therein. Essentially, these algorithms provide an interval from which to choose each parameter $y'_j \in Y'$ in such a way that certain inequalities hold for σ. In order to ensure that s_α has the same property, we first choose the parameters of σ to satisfy the strict inequalities. This gives us room to reduce α small enough so that the error term does not affect the relevant inequality for σ.

5. Local Bases

In this section, we develop a B-spline basis for both families $S^k_\alpha(X)$ and $T^k_\alpha(X)$, for $k = 1, 2$. These bases will have minimal support, in the $k = 1$ case, we need 2 intervals of the form $[x_j, x_{j+1}]$ while the $k = 2$ case requires 4 such intervals. We will also see that they are Riesz bases on their span. We begin with the following result.

Lemma 5.1. *Let $f \in S^k_\alpha(X)$ (or $T^k_\alpha(X)$), for $k = 1, 2$, with appropriate endpoint conditions specified. If supp$(f) \subset [x_{j_0}, x_{j_0+m}]$ where $m < 2k$, then $f = 0$.*

Proof. If $m < 2k$, we get a system of linear equations whose only solution is $f = 0$. When $k = 1$, we have the equations $f(x_0) = f(x_1) = 0$, solving the invertible system of equations for the coefficients yields $f = 0$. The $k = 2$ case works similarly. Suppose that $m = 3$. Then, we would need to solve for 12 coefficients using the 6 endpoint conditions: $f^{(i)}(a) = f^{(i)}(b) = 0$, for $i = 0, 1, 2$ and the 6 interior smoothness conditions. Hence the result follows from Theorem 4.1 and the fact that the system of equations for the coefficients is linear and homogeneous. $\square$

Let us now build the bases for $k = 1$. Interior bases ($1 \leq j \leq N-1$) are of the form

$$t_{\alpha,j}(x) = \begin{cases} \dfrac{\tanh(\alpha x) - \tanh(\alpha x_{j-1})}{\tanh(\alpha x_j) - \tanh(\alpha x_{j-1})}, & x \in [x_{j-1}, x_j] \\ \dfrac{\tanh(\alpha x_{j+1}) - \tanh(\alpha x)}{\tanh(\alpha x_{j+1}) - \tanh(\alpha x_j)}, & x \in [x_j, x_{j+1}] \end{cases}$$

and

$$s_{\alpha,j}(x) = \begin{cases} \dfrac{\sinh(\alpha(x - x_{j-1}))}{\sinh(\alpha h_j)}, & x \in [x_{j-1}, x_j] \\ \dfrac{\sinh(\alpha(x_{j+1} - x))}{\sinh(\alpha h_{j+1})}, & x \in [x_j, x_{j+1}] \end{cases}$$

and satisfy $t_{\alpha,j}(x_k) = s_{\alpha,j}(x_k) = \delta_{j,k}$, $0 \leq k \leq N$. The graphs are shown below.

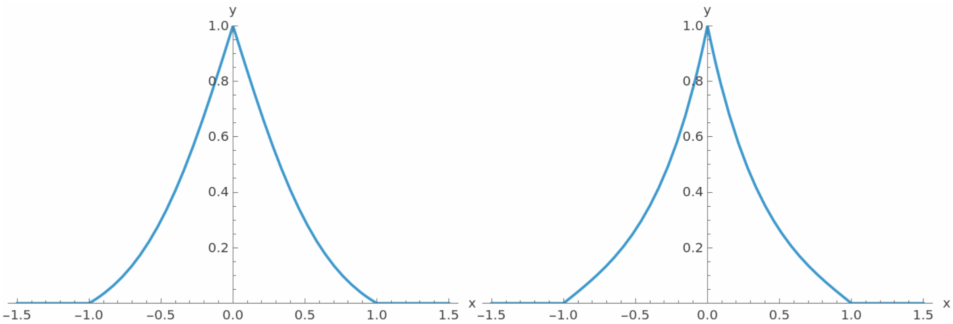


FIGURE 1. The graphs of $y = t_{1.5,0}(x)$ (left) and $y = s_{2.5,0}(x)$ (right) plotted with $h_j = 1$.

For $k = 2$, we present the uniform case $h_j \equiv h$ centered at $x = 0$. We have

$$s_\alpha(x) = \sum_{l=-2}^{1} \frac{s_l(x)}{e^{4\alpha h} - 4\alpha h e^{2\alpha h} - 1} \chi_{[lh,(l+1)h]}(x),$$

which is an even function, with

$$\begin{aligned} s_{-2}(x) &= e^{-\alpha x}(1 + \alpha(2h + x)) + e^{4\alpha h + \alpha x}(-1 + \alpha(2h + x)), \\ s_{-1}(x) &= -e^{-\alpha x}(1 + \alpha x) - 2e^{2\alpha h - \alpha x}(1 + \alpha(h + x)) \\ &\quad - 2e^{2\alpha h + \alpha x}(-1 + \alpha(h + x)) + e^{4\alpha h + \alpha x}(1 - \alpha x), \\ s_0(x) &= -e^{\alpha x}(1 - \alpha x) - 2e^{2\alpha h + \alpha x}(1 + \alpha(h - x)) \\ &\quad - 2e^{2\alpha h - \alpha x}(-1 + \alpha(h - x)) + e^{4\alpha h - \alpha x}(1 + \alpha x), \\ s_1(x) &= e^{\alpha x}(1 + \alpha(2h - x)) + e^{4\alpha h - \alpha x}(-1 + \alpha(2h - x)). \end{aligned}$$

Similarly for t_α we have

$$t_\alpha(x) = \sum_{l=-2}^{1} \frac{t_l(x)}{2\alpha h - \sinh(2\alpha h)} \chi_{[lh,(l+1)h]}(x)$$

where

$$\begin{aligned} t_{-2}(x) &= \text{sech}(\alpha x)\left[\sinh(\alpha(2h+x)) - \alpha(2h+x)\cosh(\alpha(2h+x))\right] \\ t_{-1}(x) &= 2\alpha h + 2\alpha x + \alpha x \text{sech}(\alpha x)\cosh(\alpha(2h+x)) \\ &\quad - \text{sech}(\alpha x)\sinh(\alpha(2h+x)) - 2\tanh(\alpha x), \\ t_0(x) &= -\text{sech}(\alpha x)\left[\alpha x\cosh(\alpha(2h-x)) - 2\alpha(h-x)\cosh(\alpha x)\right. \\ &\quad \left. + \sinh(\alpha(2h-x)) - 2\sinh(\alpha x)\right], \\ t_1(x) &= \text{sech}(\alpha x)\left[\sinh(\alpha(2h-x)) - \alpha(2h-x)\cosh(\alpha(2h-x))\right]. \end{aligned}$$

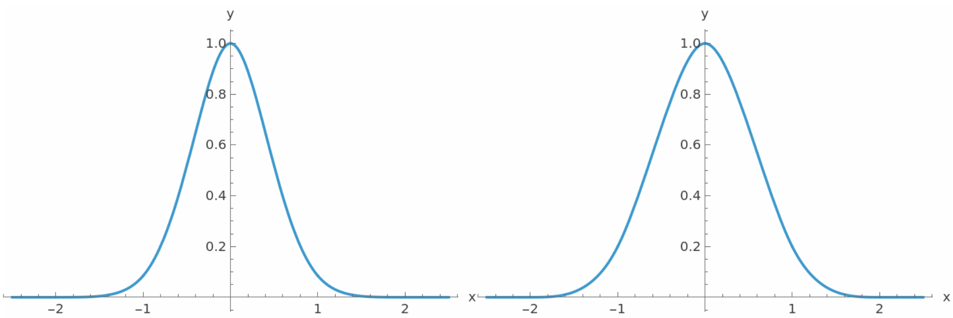


FIGURE 2. The graphs of $y = t_{1.5,0}(x)$ (left) and $y = s_{1.5,0}(x)$ (right) plotted with $h_j = 1$.

In general, for a k-th order spline, we could solve the $(2k^2) \times (2k^2)$ system

$$\begin{aligned} s^{(l)}(x_{j-k}) &= 0; \quad 0 \le l \le 2k-2 \\ s^{(l)}(x_m-) &= s^{(l)}(x_m+); \quad 0 \le l \le 2k-2, \quad j-k+1 \le m \le j-1 \\ s^{(l)}(x_j) &= \delta_{0,l}; \quad 0 \le l \le k-1 \end{aligned}$$

then extend to the right side of x_j.

Remark 5.3. *We note that $x \in [a,b]$ is in the support of at most $2k$ B-splines. For $x_j \in X$, we need only $2k-1$ B-splines.*

An important property of these splines is that they are stable. We justify this by providing estimates for the Riesz bounds. In what follows, we assume that our partition satisfies $0 < r \le h_j \le R$ for all j. We begin with $f \in S^1_\alpha(X)$, for which we have $f(x) = \sum_j c_j s_j(x)$ and

$$\begin{aligned} \int_a^b |f(x)|^2 \mathrm{d}x &= \sum_j \int_{x_j}^{x_{j+1}} \left\langle \begin{bmatrix} s_j^2(x) & s_j(x)s_{j+1}(x) \\ s_j(x)s_{j+1}(x) & s_{j+1}^2(x) \end{bmatrix} \begin{bmatrix} c_j \\ c_{j+1} \end{bmatrix}, \begin{bmatrix} c_j \\ c_{j+1} \end{bmatrix} \right\rangle \mathrm{d}x \\ &= \sum_j \left\langle \left(\int_{x_j}^{x_{j+1}} \begin{bmatrix} s_j^2(x) & s_j(x)s_{j+1}(x) \\ s_j(x)s_{j+1}(x) & s_{j+1}^2(x) \end{bmatrix} \mathrm{d}x \right) \begin{bmatrix} c_j \\ c_{j+1} \end{bmatrix}, \begin{bmatrix} c_j \\ c_{j+1} \end{bmatrix} \right\rangle. \end{aligned} \tag{5.8}$$

We may directly compute the eigenvalues of this 2×2 matrix, which yields

$$\lambda_{j,\text{min}} = h_{j+1}\frac{\alpha h_{j+1}+\sinh(\alpha h_{j+1})}{4\alpha h_{j+1}\cosh^2\left(\frac{\alpha h_{j+1}}{2}\right)} \geq r\frac{\alpha R+\sinh(\alpha R)}{4\alpha R\cosh^2\left(\frac{\alpha R}{2}\right)}$$

and

$$\lambda_{j,\text{max}} = h_{j+1}\frac{\alpha h_{j+1}+\sinh(\alpha h_{j+1})}{4\alpha h_{j+1}\sinh^2\left(\frac{\alpha h_{j+1}}{2}\right)} \leq R\frac{\alpha r+\sinh(\alpha r)}{4\alpha r\sinh^2\left(\frac{\alpha r}{2}\right)}.$$

Using these uniform bounds in (5.8) yields

$$r\frac{\alpha R+\sinh(\alpha R)}{4\alpha R\cosh^2\left(\frac{\alpha R}{2}\right)}\sum_j |c_j|^2 \leq \int_a^b |f(x)|^2\mathrm{d}x \leq 2R\frac{\alpha r+\sinh(\alpha r)}{4\alpha r\sinh^2\left(\frac{\alpha r}{2}\right)}\sum_j |c_j|^2.$$

A similar methodology can be used to find the bounds for $t_\alpha \in T^1_\alpha(X)$. In this case, the eigenvalues are in an interval governed by the two functions

$$\lambda_{\text{min}}(\alpha, r) = r + \frac{2r}{\tanh^2(\alpha r)} - \frac{2}{\alpha\tanh(\alpha r)} - 3\ln\left(\cosh(\alpha r)\right)$$

and

$$\lambda_{\text{max}}(\alpha, R) = R - \frac{\ln\left(\cosh(\alpha R)\right)}{\alpha\tanh(\alpha R)},$$

thus

$$\min\{\lambda_{\text{min}}, \lambda_{\text{max}}\}\sum_j |c_j|^2 \leq \int_a^b |f(x)|^2\mathrm{d}x \leq \max\{\lambda_{\text{min}}, \lambda_{\text{max}}\}\sum_j |c_j|^2.$$

One could extend the method above to the corresponding second order splines; however, the analysis is more tedious due to the integrals involved. For this reason, we use the Fourier techniques in the next section to analyze $s_\alpha \in S^2_\alpha(\mathbb{Z})$. The Riesz bounds are found in Corollary 6.2. The situation for $t_\alpha \in T^2_\alpha(\mathbb{Z})$ is somewhat worse, since the corresponding integrals above are not elementary and the Fourier transform does not have a closed form. Thus, we resort to a computer algebra system to approximate the lower bound $A(\alpha)$ and the upper bound $B(\alpha)$ for various α with $h_j = 1$.

α	$A(\alpha)$	$B(\alpha)$
0.001	0.121429	2.249999
0.01	0.121443	2.249910
0.1	0.122909	2.241043
1	0.244451	1.628952
10	0.144801	0.144802
100	0.014480	0.014480

The table above suggests that as $\alpha \to 0$,

$$A(\alpha) = \frac{17}{140} + \mathcal{O}(\alpha^2) \quad \text{and} \quad B(\alpha) = \frac{9}{4} - \mathcal{O}(\alpha^2).$$

As $\alpha \to \infty$, both bounds are $\mathcal{O}(\alpha^{-1})$.

6. Higher Order Splines

Having exhibited properties for splines of order $k = 1$ and $k = 2$, one is naturally led to seek results for higher order splines $S_\alpha^k(X)$ and $T_\alpha^k(X)$, where $k > 2$. The expectation is that properties from polynomial splines carry over for sufficiently small $\alpha > 0$. In the remainder of this note, we prefer to explore the interpolation problem when X is allowed to be infinite. This problem was studied in [9], and our goal is to establish the existence of spline interpolants for scattered data.

In particular, we assume that $X = (x_j : j \in \mathbb{Z})$ is a Complete Interpolating Sequence (CIS) which means that the sequence $(e^{ix_j \cdot} : j \in \mathbb{Z})$ is a Riesz basis for $L_2([-\pi, \pi])$. Among other things, we have the Riesz basis inequality

$$C_X^{-1} \sum_{j\in\mathbb{Z}} |a_j|^2 \leq \int_{-\pi}^{\pi} \left| \sum_{j\in\mathbb{Z}} a_j e^{-ix_j\xi} \right|^2 \mathrm{d}\xi \leq C_X \sum_{j\in\mathbb{Z}} |a_j|^2, \tag{6.9}$$

for some $C_X > 0$ and all $(a_j) \in \ell^2$. Associated to a Riesz basis X, we define for $n \in \mathbb{Z}$, the prolongation operator $A_n : L^2([-\pi, \pi]) \to L^2([-\pi, \pi])$ by

$$A_n[f](\xi) = A_n \left(\sum_{j\in\mathbb{Z}} a_j e^{-ix_j\xi} \right) := \sum_{j\in\mathbb{Z}} a_j e^{-2\pi i n x_j} e^{-ix_j\xi}; \quad |\xi| \leq \pi. \tag{6.10}$$

In light of 6.9, the operators A_n and their adjoints A_n^* are uniformly bounded.

We will make extensive use of the Fourier transform, which for $g \in L^1(\mathbb{R})$ is given by

$$\hat{g}(\xi) := \int_{\mathbb{R}} g(x) e^{-ix\xi} \mathrm{d}x,$$

and its extension to $g \in L^2(\mathbb{R})$ is denoted $\mathcal{F}[g]$. Using this convention, the inversion formula is given by

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{g}(\xi) e^{ix\xi} \mathrm{d}\xi$$

and Plancherel's formula is

$$\sqrt{2\pi} \|g\|_{L^2(\mathbb{R})} = \|\mathcal{F}[g]\|_{L^2(\mathbb{R})}.$$

We will restrict our attention to band-limited data. To this end, we assume that the data Y are samples taken from a function in the Paley-Wiener space

$$PW_\pi := \left\{ f \in L^2(\mathbb{R}) : \mathcal{F}[f] = 0 \text{ on } \mathbb{R} \setminus [-\pi, \pi] \right\}.$$

That is, $Y := Y[f(X)] = (f(x_j) : j \in \mathbb{Z})$, where $f \in PW_\pi$ and X is a CIS. Our argument closely follows the one found in [8]. However, our splines are not covered under the umbrella of regular interpolators found there. In the Fourier domain, we have that $s \in S_\alpha^k(X)$ enjoys the representation

$$\hat{s}(\xi) = (-1)^k (\xi^2 + \alpha^2)^{-k} \sum_{j\in\mathbb{Z}} a_j e^{-ix_j\xi},$$

in the distributional sense for some $(a_j) \in \ell^2$.

Our next result concerns two operators associated to the interpolation problem, $B_{\alpha,k}, M_{\alpha,k} : L^2([-\pi,\pi]) \to L^2([-\pi,\pi])$ which are defined as follows:

$$\begin{aligned} B_{\alpha,k}[g](\xi) &:= \sum_{n\neq 0} A_n^* \left(\left((\cdot + 2\pi n)^2 + \alpha^2 \right)^{-k} A_n[g] \right)(\xi) \\ M_{\alpha,k}[g](\xi) &:= (\alpha^2 + \xi^2)^{-k} g(\xi). \end{aligned}$$

Theorem 6.3. *Suppose that $\alpha > 0$ and $k \in \mathbb{N}$. For any $g \in L^2([-\pi,\pi])$ there exist constants $C_0, C_1 > 0$, independent of g, such that*

$$C_0 \|g\|^2_{L^2([-\pi,\pi])} \leq |\langle (M_{\alpha,k} + B_{\alpha,k})[g], g\rangle| \leq C_1 \|g\|^2_{L^2([-\pi,\pi])}.$$

Proof. We begin with the upper bound.

$$\begin{aligned} |\langle (M_{\alpha,k} + B_{\alpha,k})[g], g\rangle| &= \left| \int_{-\pi}^{\pi} (M_{\alpha,k} + B_{\alpha,k})[g](\xi)\overline{g(\xi)} \mathrm{d}\xi \right| \\ &\leq \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} |A_n^*(((\cdot + 2\pi n)^2 + \alpha^2)^{-k} A_n[g])(\xi)\overline{g(\xi)}| \mathrm{d}\xi \\ &\leq C_X \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} (\xi + 2\pi n)^2 + \alpha^2)^{-k} A_n[g](\xi)\overline{g(\xi)} \mathrm{d}\xi \\ &\leq C_X \sum_{n\in\mathbb{Z}} \sup_{|t|\leq\pi} (t + 2\pi n)^2 + \alpha^2)^{-k} \int_{-\pi}^{\pi} A_n[g](\xi)\overline{g(\xi)} \mathrm{d}\xi \\ &\leq C_X^2 \sum_{n\in\mathbb{Z}} \sup_{|t|\leq\pi} (t + 2\pi n)^2 + \alpha^2)^{-k} \int_{-\pi}^{\pi} |g(\xi)|^2 \mathrm{d}\xi \\ &\leq C_X^2 \left(\alpha^{-2k} + \sum_{n\neq 0} ((2|n| - 1)^2\pi^2 + \alpha^2)^{-k} \right) \|g\|^2_{L^2([-\pi,\pi])}. \end{aligned}$$

We have used the triangle inequality and then Tonelli's theorem to switch the sum and the integral in the first inequality. The second and fourth inequality follow from (6.9), while the third is monotonicity of the integral. Finally, the Extreme Value Theorem from Calculus provides the final inequality.

The lower bound is somewhat easier since the upper bound allows us to use the dominated convergence theorem to interchange the integral and the sum.

$$\begin{aligned} |\langle (M_{\alpha,k} + B_{\alpha,k})[g], g\rangle| &= \left| \int_{-\pi}^{\pi} (M_{\alpha,k} + B_{\alpha,k})[g](\xi)\overline{g(\xi)} \mathrm{d}\xi \right| \\ &= \left| \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} A_n^*(((\cdot + 2\pi n)^2 + \alpha^2)^{-k} A_n[g])(\xi)\overline{g(\xi)} \mathrm{d}\xi \right| \\ &= \left| \sum_{n\in\mathbb{Z}} \int_{-\pi}^{\pi} ((\xi + 2\pi n)^2 + \alpha^2)^{-k} A_n[g](\xi)\overline{A_n[g](\xi)} \mathrm{d}\xi \right| \\ &\geq \int_{-\pi}^{\pi} (\xi^2 + \alpha^2)^{-k} |g(\xi)|^2 \mathrm{d}\xi \\ &\geq (\pi^2 + \alpha^2)^{-k} \|g\|^2_{L^2([-\pi,\pi])}. \end{aligned}$$

The second equality may be justified by the dominated convergence theorem, and the third is just the adjoint relation. The first inequality follows from the positivity of the summands, while the final inequality follows from the monotonicity of the integral. □

As a Corollary, we can find the Riesz bounds for the local base s_α when $k=2$.

Corollary 6.2. *Let $f \in S^2_\alpha(\mathbb{Z})$, then $f(x) = \sum_j c_j s_j(x)$ and*

$$2\pi(\pi^2+\alpha^2)^{-2}\sum_j |c_j|^2 \le \int_a^b |f(x)|^2 \mathrm{d}x \le (2\pi)^3\left(\alpha^{-4}+\frac{1}{48}\right)\sum_j |c_j|^2.$$

Proof. We need only note that in this case Plancherel's theorem provides

$$\int_{\mathbb{R}} \left|\sum_j a_j e^{ij\xi}\right|^2 \mathrm{d}\xi = 2\pi\sum_j |a_j|^2,$$

so that $C_{\mathbb{Z}} = 2\pi$ and the upper bound is majorized by the quantity $\alpha^{-4} + \dfrac{2\zeta(4)}{\pi^4}$. □

Throughout the sequel, we consider an arbitrary but fixed $f \in PW_\pi$ sampled at a fixed but otherwise arbitrary CIS $X = (x_j : j \in \mathbb{Z})$. Since $\hat{f} \in L^2([-\pi,\pi])$, we may define the function φ_f via

$$\varphi_f(\xi) := (-1)^k (M_{\alpha,k}+B_{\alpha,k})^{-1}\hat{f}(\xi) = \sum_{j\in\mathbb{Z}} c[f]_j e^{-ix_j\xi}; \qquad |\xi| \le \pi. \tag{6.11}$$

Consider the function $I[f]$, defined by its Fourier transform:

$$\widehat{I[f]}(\xi) := (-1)^k(\alpha^2+\xi^2)^{-k}\varphi_f(\xi). \tag{6.12}$$

Since $\varphi_f \in L^2_{loc}(\mathbb{R})$, we have $\widehat{I[f]} \in (L^1 \cap L^2)(\mathbb{R})$, which means that the Fourier inversion theorem applies, and $I[f] \in (L^2 \cap C_0)(\mathbb{R})$. In particular, $I[f](x)$ is well-defined and for all $n \in \mathbb{Z}$, we have

$$\begin{aligned}
2\pi I[f](x_n) &= \int_{\mathbb{R}} \widehat{I[f]}(\xi) e^{ix_n\xi}\mathrm{d}\xi \\
&= (-1)^k \int_{-\pi}^{\pi} \sum_{j\in\mathbb{Z}} (\alpha^2+(\xi+2\pi j)^2)^{-k} A_j[\varphi_f](\xi) A_j[e^{ix_n\cdot}](\xi)\mathrm{d}\xi \\
&= (-1)^k \int_{-\pi}^{\pi} \sum_{j\in\mathbb{Z}} A_j^*\left((\alpha^2+(\cdot+2\pi j)^2)^{-k}A_j[\varphi_f](\cdot)\right)(\xi) e^{ix_n\xi}\mathrm{d}\xi \\
&= \int_{-\pi}^{\pi} (M_{\alpha,k}+B_{\alpha,k})[\varphi_f](\xi) e^{ix_n\xi}\mathrm{d}\xi \\
&= \int_{-\pi}^{\pi} \hat{f}(\xi) e^{ix_n\xi}\mathrm{d}\xi \\
&= 2\pi f(x_n).
\end{aligned}$$

Hence $I[f]$ interpolates f at X. Given the form of $\widehat{I[f]}$, it is clear that $I[f]$ satisfies $I[f] \in C^{2k-2}(\mathbb{R})$ and the ordinary differential equation

$$(D^2-\alpha^2)^k(I[f]) = \sum_{j\in\mathbb{Z}} c[f]_j \delta(\cdot - x_j), \tag{6.13}$$

where $c[f] \in \ell^2$ is chosen by (6.11). That is, $I[f]$ is a polyhyperbolic spline. As a consequence of the celebrated Paley-Wiener theorem, we can interpolate any sequence $(a_j) \in \ell^2$ by first finding $f \in PW_\pi$ which satisfies $f(x_j) = a_j$, then using Theorem 4.1 to get $c[f] \in \ell^2$.

We summarize these results in the following theorem.

Theorem 6.4. *Suppose that $b \in \ell^2$ and $X = (x_j)$ is a CIS. There exists a solution of (1.1), denoted $I[b]$, that satisfies*

(1) $I[b] \in (C_0^{2k-2} \cap L^2)(\mathbb{R})$, and
(2) $I[b](x_n) = b_n$ for all $n \in \mathbb{Z}$.

Remark 6.4. *If we require that our sampled function $f \in PW_\pi$ satisfies*

$$|f(x)| \leq C \cosh(\alpha x),$$

then we may solve the interpolation problem for $t_\alpha \in T_\alpha^k(X)$.

Having provided a solution to the interpolation problem, we may now follow the outline of [8, 10, 14] to prove the following recovery result concerning polyhyperbolic splines.

Theorem 6.5. *Suppose that $f \in PW_\pi$ and $X = (x_j)$ is a CIS. Let $I_k[f]$ denote the polyhyperbolic interpolant whose Fourier transform is defined in (6.12). Then, we have the following*

(1) $\lim_{k\to\infty} \|I_k[f] - f\|_{L^2(\mathbb{R})} = 0,$ and
(2) $\lim_{k\to\infty} |I_k[f](x) - f(x)| = 0,$ uniformly on $\mathbb{R}$.

Acknowledgments

The authors would like to thank the anonymous referees for their valuable suggestions, which improved both the mathematical quality and presentation of this article.

Financial disclosure. The funding for this work was provided by the Perspectives on Research In Science & Mathematics (PRISM) program at Longwood University.

References

[1] C. C. Aggarwal: *Neural Networks and Deep Learning: A textbook*, Springer, Cham (2018).
[2] J. H. Ahlberg, E. N. Nilson: *Convergence properties of the spline fit*, J. Soc. Indust. Appl. Math., **11** (1963), 95–104.
[3] G. Birkhoff, C. de Boor: *Error bounds for spline interpolation*, J. Math. Mech., **13** (5) (1964), 827–835.
[4] P. Bohra, J. Campos, H. Gupta, S. Aziznejad and M. Unser: *Learning activation functions in deep (spline) neural networks*, IEEE Open J. Signal Process., **1** (2020), 295–309.
[5] R. L. Dougherty, A. S. Edelman and J. M. Hyman: *Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation*, Math. Comp., **52** (186) (1989), 471–494.
[6] C. A. Hall: *On error bounds for spline interpolation*, J. Approx. Theory, **1** (1968), 209–218.
[7] B. I. Ksasov, P. Sattayatham: *GB-splines of arbitrary order*, J. Comput. Appl. Math., **104** (1) (1999), 63–88.
[8] J. Ledford: *Recovery of Paley-Wiener functions using scattered translates of regular interpolators*, J. Approx. Theory, **173** (2013), 1–13.
[9] J. Ledford: *Polyhyperbolic cardinal splines*, In Approximation Theory XV: San Antonio (2016), Springer Proc. Math. Stat., Springer, Cham (2017), 189–196.
[10] Y. Lyubarskii, W. R. Madych: *The recovery of irregularly sampled band limited functions via tempered splines*, J. Funct. Anal., **125** (1) (1994), 201–222.
[11] M. Maciejewski: *Hyperbolic splines with given derivatives at the knots*, Appl. Math. (Warsaw), **18** (2) (1984), 319–336.
[12] B. McCartin: *Theory of exponential splines*, J. Approx. Theory, **66** (1) (1991), 1–23.
[13] S. Pruess: *Properties of splines in tension*, J. Approx. Theory, **17** (1) (1976), 86–96.
[14] Th. Schlumprecht, N. Sivakumar: *On the sampling and recovery of bandlimited functions via scattered translates of the Gaussian*, J. Approx. Theory, **159** (1) (2009), 128–153.
[15] I. J. Schoenberg: *Contributions to the problem of approximation of equidistant data by analytic functions: Part A. On the problem of smoothing or graduation. A first class of analytic approximation formulae*, Quart. Appl. Math., **4** (1946), 45–99.

[16] I. J. Schoenberg, A. Whitney: *On Pólya frequence functions. III. The positivity of translation determinants with an application to the interpolation problem by spline curves*, Trans. Amer. Math. Soc., **74** (1953), 246–259.
[17] I. J. Schoenberg: *Cardinal spline interpolation*, SIAM, Philadelphia (1973).
[18] L. L. Schumaker: *On hyperbolic splines*, J. Approx. Theory, **38** (2) (1983), 144–166.
[19] L. L. Schumaker: *Spline functions: basic theory*, Cambridge University Press, Cambridge (2007).
[20] H. Späth: *Exponential spline interpolation*, Computing (Arch. Elektron. Rechnen), **4** (1969), 225–233.
[21] O. Taussky: *A recurring theorem on determinants*, Amer. Math. Monthly, **56** (10) (1949), 672–676.
[22] M. Unser: *A representer theorem for deep neural networks*, J. Mach. Learn. Res., **20** (2019), 1–30.

JEFF LEDFORD
LONGWOOD UNIVERSITY
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
FARMVILLE, VA, USA
Email address: ledfordjp@longwood.edu

LUKE PARIS
LONGWOOD UNIVERSITY
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
FARMVILLE, VA, USA
Email address: luke.paris@live.longwood.edu

RYAN URBAN
LONGWOOD UNIVERSITY
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
FARMVILLE, VA, USA
Email address: ryan.urban@live.longwood.edu

ALEC VIDANES
LONGWOOD UNIVERSITY
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
FARMVILLE, VA, USA
Email address: alec.vidanes@live.longwood.edu

MODERN MATHEMATICAL METHODS
4 (2026), No. 2, pp. 134-148
https://modernmathmeth.com/
ISSN 3023 - 5294

Research Article

Indefinite proximities inherent in dynamical systems: An axiomatic approach

JAMES F. PETERS, TANE VERGILI*, FATIH UÇAN, AND DIVAGAR VAKEESAN

ABSTRACT. This paper introduces indefinite proximities inherent in every collection of physical objects found in a dynamical system. The number of characteristics in the description of any dynamical system is unknown (Axiom 2.1 and Axiom 2.2). Hence, the description of a dynamical system is indefinite. Axiomatically, these indefinite proximities lead to a new form of Hausdorff topology, which is indefinite descriptively. The main results in this paper are

1. Every descriptive proximity space on a dynamical system is indefinite (Theorem 3.1).
2. Every dynamical system has an indefinite descriptive Hausdorff topology (Theorem 3.3).
3. The energy of a dynamical system varies with every clock tick (Theorem 5.4).

An application of these results is given in terms of the detection of those portions of a dynamical system that are stable and that have low energy dissipation.

Keywords: Characteristic, descriptive proximity space, dynamical system, Hilbert envelope, indefinite.

2020 Mathematics Subject Classification: 54E05, 54C50.

1. INTRODUCTION

This paper introduces an axiomatic approach in the study of measurable descriptive proximities that are inherent in the self-similarities in the parts, behaviours and waveforms of complex dynamical system. The consequences of this approach carry forward recent work on the descriptive approach in the study of dynamical systems [6, 7, 20, 23], especially chaotic dynamical systems [5]. This approach leads to the introduction of a number of new descriptive proximity relations that represent advances in proximity space theory [3, 14, 18] useful in the detection and measurement of those portions of dynamical systems that are stable and that have low energy dissipation.

All proximities between the parts and waveforms of dynamical systems are indefinite. This observation leads to the introduction of an indefinite descriptive proximity $\delta_{lim\Phi}$, which is a refinement of the relaxed descriptive proximity δ_{Φ_o} [6]. We observe that every descriptive proximity space on a dynamical system is indefinite (Theorem 3.1). Important results stemming from the introduction of an indefinite descriptive distance $d^{lim\Phi}$ (Definition 2.7) are given, namely, Indefinite Descriptive Hausdorff Topology (Lemma 3.2) and every dynamical system has an indefinite Hausdorff topology (Theorem 3.3). This paper also includes an application of $\delta_{lim\Phi}$ in detecting as well as measuring the stability, low energy dissipation portions of dynamical system waveforms.

Received: 13.11.2025; Accepted: 28.07.2026; Published Online: 08.08.2026
*Corresponding author: Tane Vergili; tane.vergili@ktu.edu.tr
DOI: 10.64700/mmm.87

Symbol	Meaning
d^{Φ}	Descriptive proximity distance: Def. 2.2
d_H^{Φ}	Descriptive Hausdorff distance: Def. 2.5
$(\mathcal{K}_{\Phi}, d_H^{\Phi}, \tau_H^{\Phi})$	Descriptive Hausdorff Topological Space: Def. 2.6
$d^{lim\Phi}$	Indefinite descriptive distance: Def. 2.7
(X, f, Φ)	Descriptive dynamical system: Def. 4.12
$Per(f, \Phi)$	Descriptive periodic points: Def. 4.14
δ_{Φ_o}	Relaxed descriptive proximity: Def. 5.17
$E_{m(t)}$	Waveform $m(t)$ energy: Def. 5.19
$E_{diss}(loc, t)$	Energy dissipation at location loc at time t: Def. 5.21

TABLE 1. Principal symbols used in this paper.

2. PRELIMINARIES

This section introduces characteristic-based proximities as well as characteristically proximal self-similarities in dynamical system behaviors. The earlier introduction of probe functions [14, p. 67] as a basis for the description of a nonempty set has been refined, based on earlier work on holomorphic functions [25].

Definition 2.1 (Characteristic)**.** *Let 2^X be a collection of subsets of a nonempty set X and $A \in 2^X$. Also, let $\mathbb{C}$ denote the complex plane. A characteristic of A is a mapping $\phi : 2^X \to \mathbb{C}$ defined by $\phi(A) = k \in \mathbb{C}$.*

A complete description of A with n characteristics is a set $\Phi(A)$ where

$$\Phi(A) = \{(\phi_1(a), \phi_2(a), \ldots, \phi_n(a)) \ : \ a \in A\} \subseteq \mathbb{C}^n.$$

Notice that $\Phi(A) \cap \Phi(B) \neq \emptyset$ implies there exist $a \in A$ and $b \in B$ such that $\Phi(\{a\}) = \Phi(\{b\})$. Throughout the rest of the paper, we will simply use $\Phi(a)$ instead of $\Phi(\{a\})$.

Definition 2.2 (Characteristically proximal sets)**.** *Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ (set of characteristic maps) and let X be a nonempty set with $A, B \in 2^X$. Consider the characteristic proximity mapping $d^{\Phi} : 2^X \times 2^X \to \mathbb{R}$ defined by*

$$d^{\Phi}(A, B) = \inf_{\substack{a \in A \\ b \in B}} \{|\Phi(a) - \Phi(b)|\} = r \in \mathbb{R}^{\geq 0}.$$

Then, we say that nonempty sets A, B are characteristically proximal, which is denoted by $A \ \delta_{\Phi} \ B$, provided $d^{\Phi}(A, B) = 0$.

Example 2.1. *Let $\boldsymbol{Fib} = \{\mathbf{1, 1, 2, 3, 5, 8, 13, 21, \ldots}\}$ (sequence of Fibonacci numbers). The pair of intersecting surfaces $\boldsymbol{S}_{\mathbb{R}} \cup \boldsymbol{S}_{\mathbb{C}}$ in Fig. 1 represent the real and the imaginary parts of a Hilbert cyclomatic exponential* [12]*, defined over sequences of reals x, y and a subsequence η of the Fibonacci numbers, e.g.,*

$$\boldsymbol{S}_{\mathbb{R}} = \boldsymbol{Re}\left[\boldsymbol{x}^{\mathbf{0.003}}\boldsymbol{e}^{\boldsymbol{i\pi y/2\eta+1}}\right].$$

In Fig. 1, let $\boldsymbol{imS}$ denote the surface derived from the imaginary part of the Hilbert cyclotomic exponential and let $\boldsymbol{reS}$ the surface derived from the real part of the Hilbert cyclotomic exponential. Further, let $\boldsymbol{reS^{+ve}}$ denote a characteristic of the upper surface peaks and let $\phi_{reS_{-ve}}$ denote a characteristic of the lower surface peaks. Also, let $X \subset reS^{+ve}$.

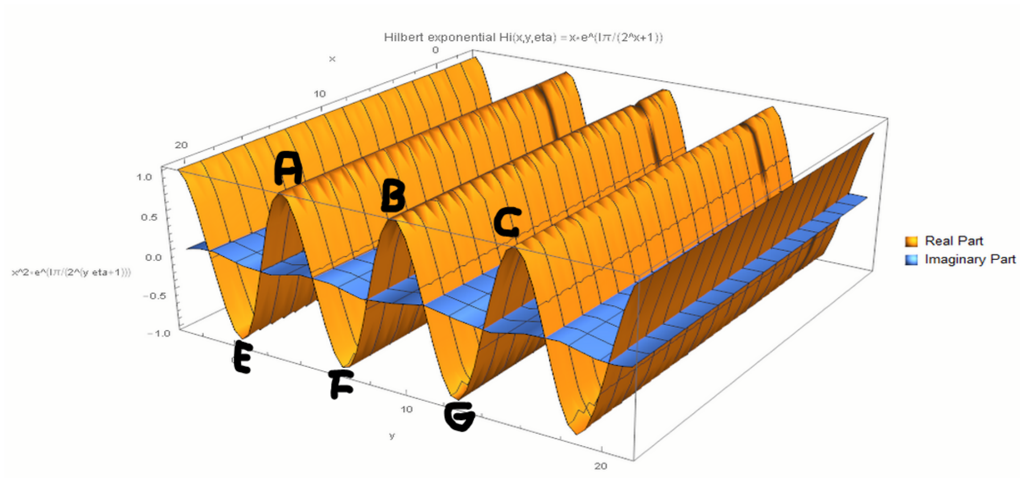


FIGURE 1. Hilbert sub-surfaces generated by a cyclomatic exponential [12].

Example 2.2 (Characteristics of Hilbert cyclomatic surface proximities)**.** *Consider, for the example, the following characteristics of the Hilbert cyclomatic surfaces in Fig. 1:*

$$\phi_{reS^{+ve}}(X) = 1, if\ X \subset reS^{+ve}, else\ \phi_{reS^{+ve}}(X) = 0 \text{ (upper surface characteristic).}$$

$$\phi_{reS^{-ve}}(X) = 1, if\ X \subset reS^{-ve}, else\ \phi_{reS^{-ve}}(X) = 0 \text{ (lower surface characteristic).}$$

Consider the following characteristic proximities

charProx-1: $\phi_{reS^{+ve}}(A) = \phi_{reS^{+ve}}(B) = \phi(C)_{reS^{+ve}} = 1$*, since* $X = \{A, B, C\} \subset reS^{+ve}$.

charProx-2: *From Def. 2.2 and charProx-1, we have* $A\ \delta_\Phi\ B$ *and* $B\ \delta_\Phi\ C$.

charProx-3: *Observe*

$$d^\Phi(A, B) = \inf_{\substack{a \in A \\ b \in B}} \{|\boldsymbol{\phi_{reS^{+ve}}(A) - \phi_{reS^{+ve}}(B)}|\} = 0.$$

Hence, our observation $A\ \delta_\Phi\ B$ *in charProx-2 is reinforced.*

charProx-4: $\phi_{reS^{+ve}}(E) = \phi_{reS^{+ve}}(F) = \phi_{reS^{+ve}}(G) = 0$*, since* $X = \{E, F, G\} \not\subset reS^{+ve}$.

charProx-5: $\phi_{reS^{-ve}}(E) = \phi_{reS^{-ve}}(F) = \phi(G)_{reS^{-ve}} = 1$*, since* $\{E, F, G\} \subset reS^{-ve}$. *Hence,* $E\ \delta_\Phi\ F$ *and* $F\ \delta_\Phi\ G$.

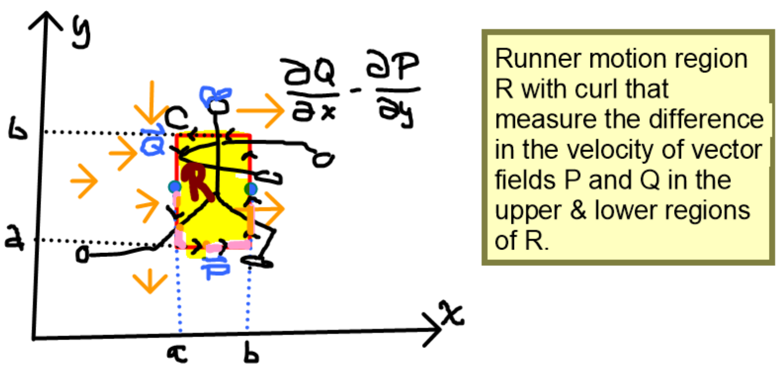


FIGURE 2. Runner vector field curl.

The following axioms enunciate the observations we have made about the characteristics of physical objects in dynamically changing systems.

Axiom 2.1 (Indefinite number of characteristics of a physical object)**.** *Let $\Phi(A)$ be a description of a nonempty set of physical objects A. The total number of characteristics of A is indefinite, i.e., not known.*

Axiom 2.2 (Unknown maximum size of the differences between sets of characteristics of physical objects)**.** *Let $|\Phi(A)|, |\Phi(B)|$ be the number of characteristics for nonempty sets A, B. The maximum size of the difference $|\Phi(A)| - |\Phi(B)|$ is not known.*

Remark 2.1 (N.B.)**.** *We use N.B. abbreviation for Latin nota bene (note well) to call attention to something important. This expression first appeared in 1711, see* [1].

N.B. 2.1 (Curl of a dynamically changing system (see [24, §11-4]))**.** *Observe, for example, in Fig. 2, that the total number of characteristics is unknown for a dynamically changing system represented by the curl of the runner vector field $\vec{V}f$, where*

$$\boldsymbol{curl\vec{V}f = \nabla \times \vec{V}f = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}},$$

which is the curl over a system surface region composed of sub-regions P, Q.

In describing the proximities between self-similarities in a physical chaotic system represented by the collection of subsets 2^X for a set X of system objects, the description of a subsets $P, Q \in 2^X$ would typically be incomplete. This is consistent with the view that the number of known characteristics of every physical object is indefinite (see Axiom 2.1) and the maximum number of characteristics of every pair of physical objects is also not known (see Axiom 2.2).

N.B. 2.2 (Origin of characteristic proximity)**.** *Characteristic functions were introduced by W. R. Hamilton* [9] *in 1837 as a means of deducing the properties of a system. Characteristically proximal sets were introduced informally* [16] *in 2012. This form of proximity is largely due to a collaboration with S. A. Naimpally that culminated in the topology of near sets* [14] *in 2013 and the introduction of object characteristics in the context of computational proximity* [17] *in 2016.*

Definition 2.3 (Characteristic proximity space, [14])**.** *For a nonempty set X, $\left(2^X, \delta_\Phi\right)$ is a characteristic proximity space.*

Definition 2.4 (Hausdorff distance, [10, 11])**.** *The Hausdorff distance, $d_H(Q, S)$, between a pair of compact subsets Q and S in $\mathbb{R}^m$ is defined by*

$$d_H(Q, S) = \max\left\{\sup_{q\in Q} D(q, S), \sup_{s\in S} D(Q, s)\right\} \geq 0,$$

where $D(p, -)$ or $D(-, p)$ denote the distance between a single point p and a given set.

One can also measure the descriptive Hausdorff distance of A and B in X by assuming that their complete descriptions $\Phi(A)$ and $\Phi(B)$ are compact subsets of $\mathbb{R}^n$.

Definition 2.5 (Descriptive Hausdorff distance)**.** *Let (X, δ_Φ) be a descriptive proximity space with a collection of n characteristics, and $A, B \in 2^X$ with compact complete descriptions $\Phi(A), \Phi(B)$ in $\mathbb{R}^m$. The descriptive Hausdorff distance $d_H^\Phi(A, B)$ between A and B is defined by*

$$d_H^\Phi(A, B) = d_H(\Phi(A), \Phi(B)).$$

Remark 2.2. *We assume that*

$$\boldsymbol{d^\Phi(A, B) = d_H(\Phi(A), \Phi(B)) = r} \in \mathbb{R}^{\geq 0}$$

for a set of known characteristics for $A, B \in 2^X$ for a set X. In a dynamical system X that is chaotic, X is inherently self-symmetric.

Given a descriptive proximity space (X, δ_Φ), let $\mathcal{K}_\Phi$ be a collection of all subsets of X with compact complete descriptions

$$\mathcal{K}_\Phi = \{A \in 2^X \ : \ \Phi(A) \text{ is compact}\}.$$

Then, d_H^Φ defines a metric on $\mathcal{K}_\Phi$.

Definition 2.6 (Descriptive Hausdorff topological space)**.** *Let the set I be an arbitrary index set. A descriptive Hausdorff topology, τ_H^Φ, induced by the descriptive Hausdorff metric d_H^Φ on $\mathcal{K}_\Phi$ has the following properties*

1º $\mathcal{K}_\Phi, \emptyset \in \tau_H^\Phi$.
2º $\{\mathcal{A}_i\}_{i\in I} \subseteq \tau_H^\Phi$ *implies* $\bigcup_{i\in I} \mathcal{A}_i \in \tau_H^\Phi$.
3º $\mathcal{A}, \mathcal{B} \in \tau_H^\Phi$ *implies* $\mathcal{A} \cap \mathcal{B} \in \tau_H^\Phi$.

The triple $(\mathcal{K}_\Phi, d_H^\Phi, \tau_H^\Phi)$ *is called a descriptive Hausdorff topological space.*

Corollary 2.1. *There is a descriptive Hausdorff topology τ_H^Φ on every collection of compact complete descriptions of a descriptive proximity space (X, δ_Φ).*

N.B. 2.3 (New form of descriptive proximity)**.** *From Axiom 2.1, the number of characteristics of a physical objects is indefinite. For this reason, we introduce a new form of descriptive proximity of a pair of sets of physical objects A, B in terms of the differences of the descriptions of A and B converging to 0 in the limit (see Definition 2.7).*

Definition 2.7 (Indefinite characteristic distance)**.** *Let X be a nonempty set. $A, B \in 2^X$ are indefinitely characteristically near sets of physical objects, provided*

$$d^{\lim \Phi}(A, B) = \lim_{|\phi(a)-\phi(b)|\to 0} d^\Phi(A, B) = 0.$$

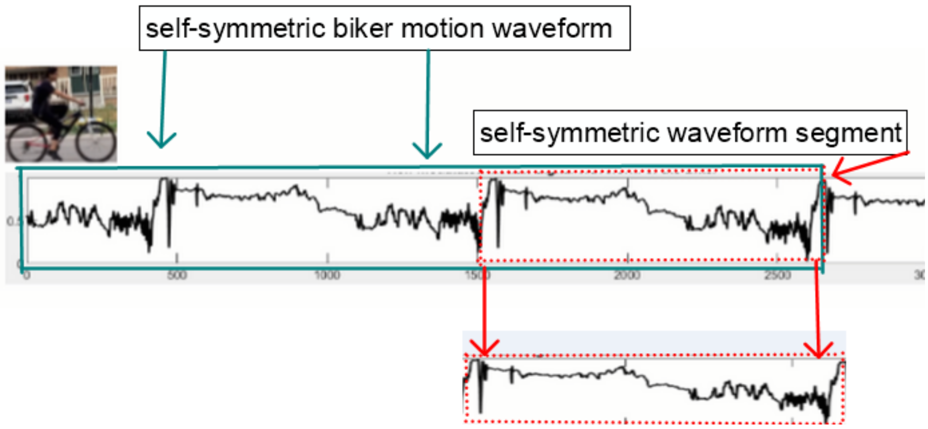


FIGURE 3. Self-similar biker motion waveform.

3. Chaotic Dynamical System

Dynamical chaos is different from randomness or disorder. Chaos is deterministic in descriptive set theory [23]. For example, vibrations of a mechanical system are quasi-periodic, changing from period doubling to chaos with period doubling [26]. Chaos in dynamics occurs if the cloud of representative points in the course of evolution in the phase space undergoes repetitive deformations of stretching, folding, and transversal compression [13]. Typically, nonlinear motion cascades to chaos [27].

Definition 3.8 (System)**.** *A system is a configuration of components functioning together as a whole and their relationships.*

Definition 3.9 (Dynamical system)**.** *A dynamical system is a physical system (collection of physical objects) that changes over time.*

Definition 3.10 (Chaotic dynamical system)**.** *A chaotic dynamical system (CDS) is a system that has the following properties*

1^o The waveform of a CDS is self-similar, i.e., the CDS contains parts with the same structure as the complete CDS.

2^o The self-similar fractals in the CDS satisfy the open set condition [5]*, i.e., there exists an integer n so that for every piece $A_i \in CDS$ with diameter ε, there are at most incomparable pieces $A_{j\leq n} \in CDS$ with diameter $\geq \varepsilon$ with distance ε from A_i* [22]*. Let $X \in 2^Y$ in a CDS with subsets 2^X in a descriptive Hausdorff proximity space $(\mathcal{K}^Y_\Phi, d^\Phi_H, \tau^\Phi_H)$. A description $\Phi(X)$ of a CDS X in a space Y contains the description of all the subsets of $A \in 2^X$ such that $d_H(A,B) < \varepsilon$ for all $B \in 2^Y$.*

3^o The number of known characteristics of the CDS is indefinite.

Example 3.3. *A vibrating system is an example of CDS. For example, in Fig. 3 a biker is a CDS with the following properties*

1^o The biker waveform in Fig. 3 is self-similar, i.e., the complete motion waveform pattern is repeated in its segments.

2^o The biker waveform satisfies the open set condition. To see this, let $X \in 2^Y$ in a biker motion system with subsets 2^X in a collection Y of moving systems (e.g., walkers and vehicles). This biker CDS resides in a descriptive Hausdorff proximity space $\mathcal{K}^X_\Phi$ that is a subspace of $\mathcal{K}^Y_\Phi$. Let $B \in 2^Y$. A description $\Phi(A)$ of a subset $A \in 2^X$ in space Y contains the description of all the subsets of $A \in 2^X$ such that $d_H(A,B) < \varepsilon$ for all $B \in 2^Y$.

3^o The number of known characteristics of the biker system is indefinite.

Lemma 3.1. *Every descriptive proximity space on a collection of physical objects is indefinite.*

Proof. Let (X, δ_Φ) be a descriptive proximity space on a collection of physical objects in 2^X. From Axiom 2.1, the number of known characteristics of subsets $A, B \in 2^X$ is indefinite. Then, from Definition 2.7, $d^\Phi(A,B) = d^{\lim \Phi}(A,B)$. Hence, $(2^X, d^\Phi)$ is indefinite. □

Theorem 3.1. *Every descriptive proximity space on a dynamical system is indefinite.*

Proof. Immediate from Axiom 2.1 and Lemma 3.1. □

Theorem 3.2. *Every descriptive proximity space on a chaotic dynamical system is indefinite descriptively.*

Proof. Immediate from Definition 3.10 and Theorem 3.1. □

Lemma 3.2 (Indefinite descriptive Hausdorff topology)**.** *Every descriptive Hausdorff topology on a sets of objects in a physical system is indefinite.*

Proof. Let 2^X be collections of subsets of a dynamical system X, $A, B \in 2^X$. Let $(X, \delta_{\lim \Phi})$ be an indefinite descriptive proximity space. From Axiom 2.2, all descriptive Hausdorff distances $d_H(\Phi(A), \Phi(B))$ approach zero, i.e., descriptive Hausdorff proximities between sets of physical objects in a dynamical system are indefinite. By replacing 2^X in Definition 2.6, and obtain an indefinite descriptive Hausdorff topology. □

Theorem 3.3. *Every dynamical system has an indefinite descriptive Hausdorff topology.*

Proof. Immediate from Definition 3.9 and Lemma 3.2. □

4. Descriptive Case

There are many definitions of chaos in the literature. Among these definitions, we often encounter definition of chaos in sense of R. L. Devaney made in 1986 [4], followed by studies of the relationship between individual chaos and collective chaos (see, e.g., [5, 8, 15, 22]). In this section, our aim is to analyze a chaotic dynamical systems in the descriptive sense by adding new features to it and observe them.

4.1. **Descriptive Dynamical System.** The fabric of a descriptive dynamical system is a family of interactions on a nonempty set of points (quanta) X. There is an orbit of each point $x \in X$ defined by a family of interactions on x with characteristics such as vector tail or vector tip. In this subsection, we consider X as a topological space together with a real-valued characteristic function $\Phi : X \to \mathbb{C}^m$ on it.

Definition 4.11 (Family of iterations)**.** *The dynamical system (X, f) on X defined by f is the family of iterations $\{f^n\}_{n\in\mathbb{Z}^+}$ with each f^n mapping from X to itself. For $x \in X$ and $n \in \mathbb{Z}^+$ the orbit of x is the set of points*

$$x, f(x), f^2(x), \ldots, f^n(x), \ldots$$

and is denoted by

$$Orb_f(x) = \{f^n(x) : n \in \mathbb{Z}^+\}.$$

Definition 4.12 (Descriptive dynamical system)**.** *A descriptive dynamical system is a pair $(\boldsymbol{S}, \boldsymbol{\Phi})$, which consists of a nonempty set S accompanied by a nonempty set characteristic functions such that*

$$\boldsymbol{\Phi} = \{\boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_i}, \ldots, \boldsymbol{\phi_n}\}\, |^{\boldsymbol{n \geq 1}} : \boldsymbol{\phi_i} : \boldsymbol{S} \to \mathbb{C}.$$

Definition 4.13 ([19])**.** *Given a descriptive dynamical system (X, f, Φ), a subset A of X is*

(1) a descriptive fixed subset of f, provided $\Phi(f(A)) = \Phi(A)$ and

$$\Phi(f^j(A)) \neq \Phi(A), \textit{for the set of descriptive periodic sets } j = 1, 2, \ldots, m-1.$$

(2) a descriptive period-m set of f if $\Phi(f^m(A)) = \Phi(A)$.

When $A = \{a\}$ is a singleton set, we simply replace the word set with the word point. In that case, we say that a point $a \in X$ is

(1) a descriptive fixed point of f, if $\Phi(f(a)) = \Phi(a)$,
(2) a descriptive period-m point of f if $\Phi(f^m(a)) = \Phi(a)$, and $\Phi(f^j(a)) \neq \Phi(a)$ for $j = 1, 2, \ldots, m-1$.

Definition 4.14 (Descriptive periodic points)**.** *Given a descriptive dynamical system (X, f, Φ),*

(1) the set of descriptive m-periodic points is given by

$$Per_m(f, \Phi) = \{a \in X : \Phi(f^m(a)) = \Phi(a)\}.$$

(2) the set of descriptive periodic points is given by

$$Per(f, \Phi) = \bigcup_{m\in\mathbb{Z}^+} Per_m(f, \Phi) = \left\{a \in X \ : \ \Phi(f^m(a)) = \Phi(a), m \in \mathbb{Z}^+\right\}.$$

Remark 4.3. *Here, one can understand from Definition 3.10 and Theorem 3.2 that every periodic m-set on a chaotic dynamical system is an indefinite descriptive periodic m-set. However, the converse may not be true. Even so, a periodic m-set is a descriptive periodic k-set for $k \leq m$.*

Definition 4.15. *The set-valued function $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is defined by $\bar{f}(A) = \{f(a) : a \in A\}$ so that the complete description of A is*

$$\begin{aligned}\Phi(\bar{f}(A)) &= \{\Phi(f(a)) \ : \ a \in A\} \\ &= \big\{\big(\phi_1(f(a)), \phi_2(f(a)), \ldots, \phi_n(f(a))\big) : a \in K\big\} \subseteq \mathbb{C}^n\end{aligned}$$

and its descriptive orbit is a collection of sets

$$\Phi(\bar{f}(A)), \Phi(\bar{f}^2(A)), \ldots, \Phi(\bar{f}^n(A)).$$

Remark 4.4. *For a subset A of X, the extension of A to $\mathcal{K}(X)$ is defined by*

$$e(A) = \{K \in \mathcal{K}(X) \ : \ K \subset A\}.$$

Also the following lemma is given in [21].

Lemma 4.3. *For two subsets A and B of X, the following holds*

- *$e(A) \neq \emptyset$ if and only if $A \neq \emptyset$,*
- *$e(A)$ is open subset of $\mathcal{K}(X)$, if $A \subseteq X$ is open,*
- *$e(A \cap B) = e(A) \cap e(B)$,*
- *$\bar{f}(e(A)) \subseteq e(\bar{f}(A))$,*
- *$\bar{f}^n = \bar{f}^n$, for all $m \in \mathbb{Z}^+$.*

Here, we present a descriptive approach to Devaney's definition of chaos by introducing new definitions: a dense collection of descriptive periodic sets, descriptive transitivity and descriptively sensitive.

Definition 4.16. *For $(\mathcal{K}(X), d_H)$, a mapping $\bar{f} : \mathcal{K}(X) \to \mathcal{K}(X)$ is said to be descriptively chaotic if*

- *the collection of $Per(\bar{f}, \Phi)$ descriptive periodic compact sets which is given by*
$$\bigcup_{m \in \mathbb{Z}^+} Per_m(\bar{f}, \Phi) = \big\{K \in \mathscr{K}(X) \ : \ \Phi(\bar{f}^m(K)) = \Phi(K), m \in \mathbb{Z}^+\big\}$$
is dense in $\mathscr{K}(X)$,
- *for every $e(U), e(V)$ which are respectively extensions of nonempty open pair $U, V \subseteq X$, there exists $n \in \mathbb{Z}^+$ such that $\bar{f}(e(U))$ and $e(V)$ are descriptively proximal, i.e., $\Phi(\bar{f}^n(e(U))) \cap \Phi(e(V)) \neq \emptyset$ ($\bar{f}$ is descriptively transitive), and*
- *there is a $\delta > 0$ (sensitive constant) such that for any elements A, B of $e(U)$ which is an extension of nonempty open subset U of X, there exists $n \in \mathbb{Z}^+$ so that $d^{\Phi}(\bar{f}^n(A), \bar{f}^n(B)) = d_H(\Phi(\bar{f}^n(A), \bar{f}^n(B))) > \delta$ ($\bar{f}$ is descriptively sensitive).*

Lemma 4.4. *If the collection of periodic sets in a Hausdorff metric space is dense, so is the collection of descriptively periodic sets.*

Proof. Immediately follows from the fact that any periodic set is also descriptively periodic set. □

Lemma 4.5. *If $\bar{f}$ is topologically transitive, so is descriptively transitive.*

Proof. Let $e(U), e(V)$ be extensions of two nonempty open subsets U, V of X respectively. Since $\bar{f}$ is topologically transitive, there exists a positive integer m such that $\bar{f}^m(e(U)) \cap e(V) \neq \emptyset$. In that case there exists a compact subset K in $\mathcal{K}(X)$ such that $K \in \bar{f}^m(e(U))$ and $K \in e(V)$. Since K is in the image of $\bar{f}^m(e(U))$, there is K' in $e(U)$ so that $\bar{f}^m(K') = K$. This implies $\Phi(\bar{f}^m(K')) = \Phi(, K)$ i.e., $\Phi(\bar{f}^m(e(U))) \cap \Phi(e(V)) \neq \emptyset$. □

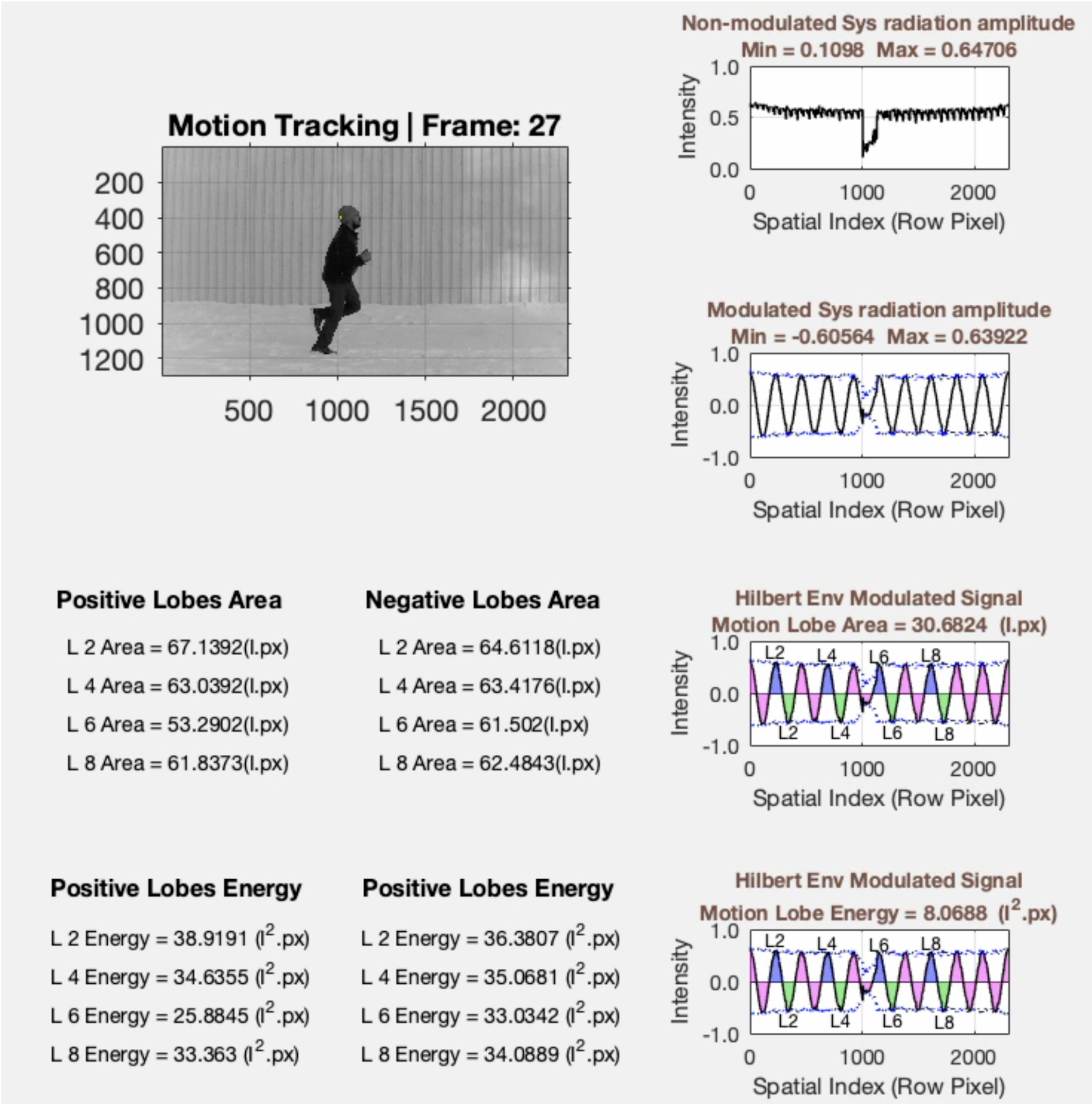


FIGURE 4. Relaxed proximities between Hilbert lobes.

5. APPLICATION: RELAXED DESCRIPTIVE PROXIMITIES

This application introduces a relaxed descriptive proximity (denoted by δ_{Φ_o}), which is a refinement of the indefinite proximity relation $\delta_{\lim\Phi}$ in the proof of Lemma. 3.1. δ_{Φ_o} serves as means of pinpointing non-equal descriptively close segments of motion waveforms emanating from vibrating dynamical systems. Let A, B be measurable bounded regions in a typical motion waveform of a dynamical system. A practical outcome of $\delta_{\Phi_o}(A, B)$ is the detection those portions of system motion that are stable and have low energy dissipation.

A pair of nonempty sets A, B have relaxed descriptive proximity (denoted by $A\delta_{\Phi_o}B$), provided the characteristics of A and B are close enough.

Definition 5.17 (Relaxed descriptive proximity, [6]). *Let $\varepsilon \in [0, 1]$ and 2^X denote the collection of subsets of a nonempty set X. The relaxed descriptive proximity relation δ_{Φ_o} is defined by*

$$A\delta_{\Phi_o}B \Leftrightarrow |\Phi(A) - \Phi(B)| < \epsilon, A, B \subset X.$$

The focus here is on proximal Hilbert envelope lobes inherent in time-constrained dynamical systems. Encapsulated in a Hilbert envelope, there are comparable regions called lobes.

Definition 5.18 (Hilbert envelope lobe). *A Hilbert envelope lobe (denoted by $H_{env_{lobe}}$) is a tiny bounded planar region attached to single waveform peak point on a waveform envelope, defined by* [2]

$$H_{env} = \sqrt{m(t)^2 + (-m(t))^2}.$$

Remark 5.5. *Hilbert envelope lobe regions provide a measure of motion waveform energy dissipation. The bigger a lobe region, the greater the motion energy. Our interest here is in identifying those motion waveforms having a high number of proximal Hilbert lobe energy regions that have a corresponding low energy dissipation (denoted by $E_{diss}(t)$ at time t).*

Expenditure of energy $E_{m(t)}$ by a dynamical system S_d is measured in terms of the area bounded by the motion $m(t)$ waveform emanating from S_d at time t, i.e.:

Definition 5.19 (Waveform energy). *A measure of dynamical system energy is the area of a finite planar region bounded by system waveform $m(t)$ curve at time t, defined by*

$$\boldsymbol{E_{m(t)} = \int_{t_0}^{t_1} |m(t)|^2 \, dt.}$$

For simplicity, we limit our consideration to elapsed time to clock tick, which is measured by an observed of a local mechanical clock.

Definition 5.20 (Elapsed time measured in clock ticks). *Let $p(t)$ be a characteristic function that models a cog in a rotating wheel in a mechanical clock with a velocity $\boldsymbol{\frac{dp}{dt}}$. A clock tick t_k is defined by*

$$\boldsymbol{t_k = \frac{dp(t)}{dt}|^{t=t_0}}.$$

N.B. 5.4 (Varying clock tick value). *The value of $\boldsymbol{t_k}$ will vary, since it depends on a mechanical system, which is a collection of wearable parts subject to varying friction.*

Axiom 5.3 (Instants clock). *Every system has its own instants clock, which is a cyclic mechanism that is a simple closed curve with an instant hand with one end of the instant hand at the centroid of the cycle and the other end tangent to a point indicating an elapsed time in the motion of a vibrating system. A clock tick occurs at every instant that a system changes its state.*

Axiom 5.4 (Waveform energy, [20]). *A measure of dynamical system energy is the area of a finite planar region bounded by system waveform $m(t)$ curve over a bounded temporal interval $[t_0, t_1]$., defined by*

$$\boldsymbol{E_m = \int_{t_0}^{t_1} |m(t)|^2 \, dt.}$$

Lemma 5.6 (Dynamical system energy, [20]). *Dynamical system energy is time-constrained and is always limited.*

Proof. Let E_m be the energy of a dynamical system, defined in Axiom 5.4. From Axiom 5.4, system energy always occurs in a bounded temporal interval $[t_0, t_1]$. Hence, E_m is time constrained. From Axiom 5.3, the length of a system waveform is finite, since, from Axiom 5.4, system duration is finite. From Axiom 5.4, system energy is derived from the area of a finite, bounded region. Consequently, system energy is always finite. □

From Def. 5.20 and Lemma 5.6, we obtain:

Theorem 5.4 (Time-constrained dynamical system energy [20])**.** *Let X be a dynamical system with waveform $m(t)$ at time t. The energy of X varies with every clock tick.*

Definition 5.21 (Energy dissipation [20])**.** *For a motion waveform $m(t)$ at time t, let loc identify a bounded region (called a Hilbert lobe) at location x in an enveloped waveform and let t be the time of occurrence of the lobe. A measure of motion dissipated energy is the mapping $E_{diss} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, which is defined by*

$$E_{diss}(loc, t) = |E(x, t) - E(x', t')| \geq 0$$

for quantized dissipation. There are two different forms of motion energy dissipation to consider, namely,

1^o ***Energy dissipation within one frame:*** *Let x, x' identify lobes in two different locations at time t in an enveloped waveform in a single video frame at elapsed time t. Then*

$$E_{diss}(loc, t) = |E(x, t) - E(x', t)|\,.$$

2^o **Energy dissipation between different frames:** *Let $x = x'$ identify a lobe in the same location at times t, t' in a pair video frames. Then*

$$E_{diss}(loc, t) = |E(x, t) - E(x, t')|\,.$$

The motivation for considering two different forms of $E_{diss}(loc, t)$ stems from Lemma 5.7, which associates the relaxed descriptive proximity of a pair of energy regions with corresponding energy dissipation that is minimal.

Lemma 5.7. *Let $\boldsymbol{E_{m(t)}}$,*

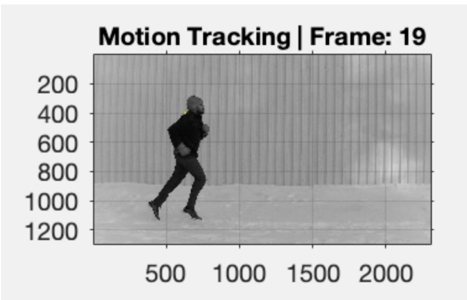
$$\boldsymbol{E_{m(t)}\ \delta_{\Phi_o}\ E_{m(t')} \Leftrightarrow E_{diss}(loc, t) < \varepsilon, \varepsilon \in [0, 1].}$$

That is, $E_{diss}(loc, t)$ is minimal.

Proof. It is known that energy dissipation $E_{diss}(t')$ in a motion curve is minimal, when the difference between $E_{m(t)}, E_{m(t')}, t < t'$ is small [20]. This occurs when $E_{m(t)}\ \delta_{\Phi_o}\ E_{m(t')}$, i.e., $E_{diss}(t) < \varepsilon, \varepsilon \in [0, 1]$. This gives the desired result. □

Theorem 5.5. *Every Hilbert enveloped motion curves $m(t), m(t')$ with $E_{m(t)}\ \delta_{\Phi_o}\ E_{m(t')}$ at times $t \neq t'$ has low energy dissipation.*

Proof. Immediate from Lemma 5.7. □



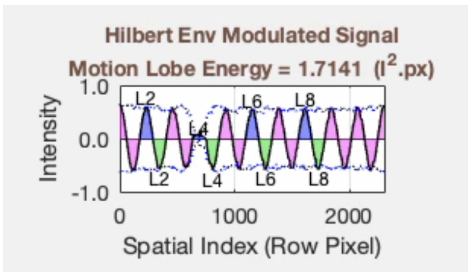


FIGURE 5. Source of Hilbert energy lobes in Table 2.

The sample lobe areas (a.k.a. motion energy regions) in Table 2 (see Fig. 5) lead to Example 5.4.

Lobe	+ve Lobe Areas (+ve lobe)	-ve Lobe Areas (-ve lobe)	Energy Dissipation $\boxed{E_{diss}}$
L2	38.952	36.2096	2.7424
L4	2.522	32.3247	29.8027
L6	33.1404	33.3126	0.1722
L8	33.6902	33.9613	0.2711

TABLE 2. Lobe energy for runner in IR video frame 19.

Example 5.4 (Relaxed proximities within a frame). *Let* $\varepsilon = 0.03$. *From Table 2, observe*

$$\begin{aligned} &1^o \;\; L2^{+ve} \;\; \not\delta_{\Phi_o} \; L2^{-ve}, \\ &2^o \;\; L4^{+ve} \;\; \not\delta_{\Phi_o} \; L4^{-ve}, \\ &3^o \;\; L6^{+ve} \; \delta_{\Phi_o} \; L6^{-ve}, \\ &4^o \;\; L8^{+ve} \; \delta_{\Phi_o} \; L8^{-ve}. \end{aligned}$$

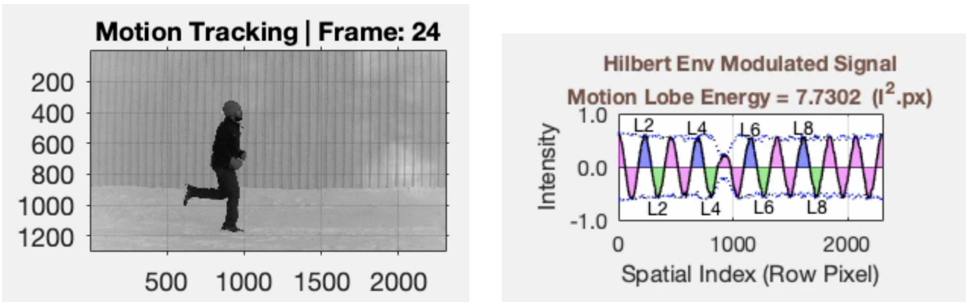


FIGURE 6. Source of Hilbert energy lobes in Table 3.

The sample lobe areas (a.k.a. motion energy regions) in Table 3 (see Fig. 6) and in Table 4 (see Fig. 7) are a source of important relaxed proximities between Hilbert lobes in terms of lobe energy. Lobe energy proximities between frames serve as an indication of the stability of system energy. Relaxed proximities between IR frames are considered in Example 5.5.

A check on relaxed proximities between Hilbert lobes in terms of lobe energy is considered next.

The sample lobe areas (a.k.a. motion energy regions) in Table 3 (see Fig. 6).

Lobe	+ve Lobe Areas (+ve lobe)	-ve Lobe Areas (-ve lobe)	Energy Dissipation $\boxed{E_{diss}}$
L2	38.7384	36.3319	2.4065
L4	34.5799	35.0868	0.5069
L6	32.8082	33.3491	0.5409
L8	33.0718	33.5572	0.4854

TABLE 3. Lobe energy for runner in IR video frame 24.

Example 5.5 (Relaxed proximities between frames). *Let* $\varepsilon = 0.2$ *for the dynamical system represented by motion* $m(t)$ *in Fig. 6 and motion* $m(t')$ *in Fig. 7. Also, let*

$$\text{(lobe) } L_k^{Tablek}, \text{(energy) } E_{\pm L_k}^{Tablek}, k \in \{19, 24, 26\}, n \in \{2, 4, 6, 8\}.$$

Observe

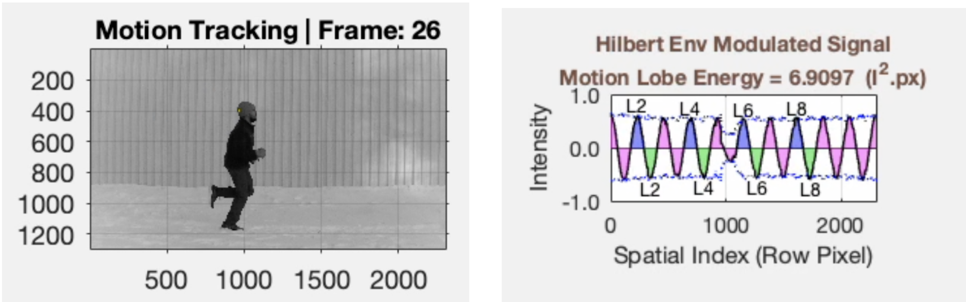


FIGURE 7. Source of Hilbert energy lobes in Table 4.

Lobe	+ve Lobe Areas (+ve lobe)	-ve Lobe Areas (-ve lobe)	Energy Dissipation $\boxed{E_{diss}}$
L2	39.0058	36.4091	2.5967
L4	34.3714	35.023	0.6516
L6	32.8635	33.424	0.5605
L8	33.1931	34.1811	0.988

TABLE 4. Lobe energy for runner in IR video frame 26.

1^o $+veL_4^{Table24}\ \delta_{\Phi_o}\ +veL_4^{Table26}$, *since*

$$\left|E_{+L_4}^{Table24} - E_{+L_4}^{Table26}\right| = |34.5799 - 34.3714| = 0.2085 < \epsilon.$$

2^o $-veL_4^{Table24}\ \delta_{\Phi_o}\ -veL_4^{Table26}$, *since*

$$\left|E_{+L_4}^{Table24} - E_{+L_4}^{Table26}\right| = |35.0868 - 35.023| = 0.0638 < \epsilon.$$

3^o $+veL_6^{Table24}\ \delta_{\Phi_o}\ +veL_6^{Table26}$, *since*

$$\left|E_{+L_6}^{Table24} - E_{+L_6}^{Table26}\right| = |32.8082 - 32.8635| = 0.0553 < \epsilon.$$

4^o $-veL_4^{Table24}\ \delta_{\Phi_o}\ -veL_4^{Table26}$, *since*

$$\left|E_{-L_6}^{Table24} - E_{-L_6}^{Table26}\right| = |33.3491 - 33.424| = 0.0749 < \epsilon.$$

Hence, $E_{m(t)}\delta_{\Phi_o}E_{m(t')}$. *From Theorem 5.5, system S at times t and t' has low energy dissipation.*

ACKNOWLEDGMENTS

The research has been supported by the Natural Sciences & Engineering Research Council of Canada (NSERC) discovery grant 185986 and Instituto Nazionale di Alta Matematica (INdAM) Francesco Severi, Gruppo Nazionale per le Strutture Algebriche, Geometriche e Loro Applicazioni grant 9 920160 000362, n.prot U 2016/000036 and Scientific and Technological Research Council of Turkey (TÜBİTAK) Scientific Human Resources Development (BIDEB) under grant no: 2221-1059B211301223.

REFERENCES

[1] J. Addison: *The works of Joseph Addison*, Oxford U. Press, Oxford (1891).
[2] A. Brandt: *Noise and vibration analysis. Signal analysis and experimental procedures*, Wiley, New York (2011).
[3] Á. Császár: *Proximities, screens, merotopies, uniformities. I.*, Acta Math. Hungar., **49** (3-4) (1987), 459–479.
[4] R. L. Devaney: *An introduction chaotic dynamical systems*, The Benjamin/Cummings Publishing, Menlo Park CA (1986).

[5] M. Edelman, E. E. N. Macau and M. A. F. Sanjuan: *Chaotic, fractional, and complex dynamics: New insights and perspectives*, Springer, Cham (2018).
[6] E. Erdag, J. F. Peters and Ö. Deveci: *The Jacobsthal-Padovan-Fibonacci p-sequence and its application in the concise representation of vibrating systems with dual proximal groups*, J. Supercomput., **81** (2025), Article ID:197.
[7] L. Gül, J. F. Peters and Ö. Deveci: *Dual proximal groups concisely representing complex Hosoya triangles*, Adv. Math. Phys., **2026** (2026), Article ID:1742124.
[8] M. S. Haider, J. F. Peters: *Temporal proximities: self-similar temporally close shapes*, Chaos solit. fractals, **151** (2021), Article ID:111237.
[9] W. R. Hamilton: *Essay on the theory of rays*, Trans. Royal Irish Academy, **15** (1) (1828), 69–174.
[10] F. Hausdorff: *Grundzüge der mengenlehre*, Veit and Company, Leipzig (1914).
[11] F. Hausdorff: *Set theory*, AMS Chelsea Publishing, Rhode Island (1957).
[12] D. Hilbert: *Ueber die theorie der algebraischen formen. (German) (On the Theory of Algebraic Forms)*, Math. Ann., **36** (4) (1890), 473–534.
[13] P. J. Hilton: *An introduction to homotopy theory*, Cambridge University Press, Cambridge (1953).
[14] S. A. Naimpally, J. F. Peters: *Topology with applications. Topological spaces via near and far*, World Scientific, Singapore (2013).
[15] E. Ozkan, B. Kuloglu and J. F. Peters: *k-Narayana sequence self-similarity. Flip graph views of k-Narayana self-similarity*, Chaos solit. fractals, **153** (2021), Article ID:111473.
[16] J. F. Peters, S. A. Naimpally: *Applications of near sets*, Notices Amer. Math. Soc., **59** (4) (2012), 536–542.
[17] J. F. Peters: *Computational proximity*, Intelligent Systems Reference Library, **102**, Springer, Cham (2016).
[18] J. F. Peters, T. Vergili: *Descriptive fixed set properties for ribbon complexes*, arXiv:2007.04394, (2020).
[19] J. F. Peters, T. Vergili: *Good coverings of proximal Alexandrov spaces. Path cycles in the extension of the Mitsuishi-Yamaguchi good covering and Jordan curve theorem*, arXiv:2104.05601, (2021).
[20] J. F. Peters, T. U. Liyanage: *Energy dissipation in Hilbert envelopes on motion waveforms detected in vibrating systems: An axiomatic approach*, Commun. Adv. Math. Sci., **7** (4) (2024), 178–186.
[21] H. Roman-Flores: *A note on transitivity in set-valued discrete systems*, Chaos solit. fractals, **17** (2003), 99–104.
[22] A. Schief: *Separation properties for self-similar sets*, Proc. Am. Math. Soc., **122** (1) (1994), 111–115.
[23] O. M. Sharkovsky: *Descriptive theory of deterministic chaos*, Ukrainian Math. J., **74** (12) (2022), 1950–1960.
[24] C. E. Springer: *Tensor and vector analysis. With applications to differential geometry*, The Ronald Press, New York (1962).
[25] E. M. Stein, R. Shakarchi: *Complex analysis. Princeton Lectures in Analysis*, Princeton University Press, Princeton (2003).
[26] Y. Wei, M. J. Wu: *A fractal measure of period-doubling leading to chaos-a computer-assisted quantitative analysis*, J. East China Norm. Univ. Natur. Sci. Ed., **1** (2) (1995), 27–45.
[27] R. Zhao, Y. Xu, Y. Jiao, Z. Li, Z. Chen and Z. Chen: *Nonlinear motion cascade to chaos in a rotor system based on energy transfer*, Nonlinear Dyn, **112** (2024), 10803–10821.

James F. Peters
University of Manitoba
Computational Intelligence Laboratory
Department of Electrical and Computer Engineering
WPG, MB, R3T 5V6, Canada
Email address: `james.peters3@umanitoba.ca`

Tane Vergili
Karadeniz Technical University
Department of Mathematics
Trabzon, Türkiye
Email address: `tane.vergili@ktu.edu.tr`

Fatih Uçan
Karadeniz Technical University
Department of Mathematics
Trabzon, Türkiye
Email address: `fatihucan@ktu.edu.tr`

DIVAGAR VAKEESAN
UNIVERSITY OF MANITOBA
COMPUTATIONAL INTELLIGENCE LABORATORY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
WPG, MB, R3T 5V6, CANADA
Email address: divagarv@myumanitoba.ca