School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

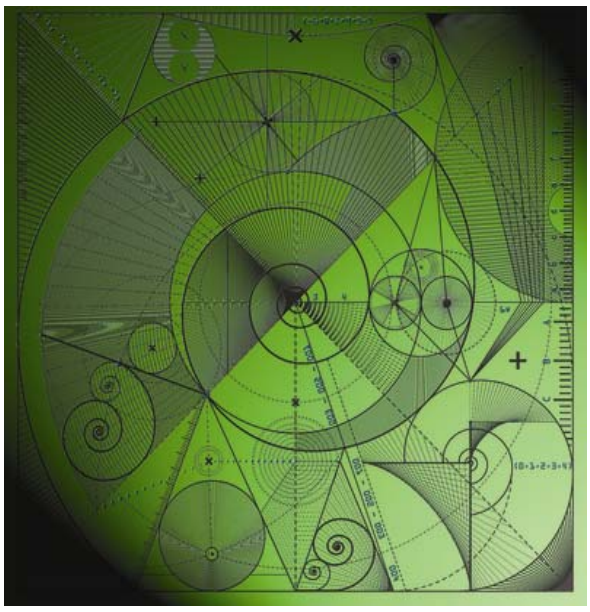
Preprint MPS-2011-08

30 June 2011

On the Spectra and Pseudospectra of a Class of Non-Self-Adjoint Random Matrices and Operators

by

Simon N. Chandler-Wilde, Ratchanikorn Chonchaiya and Marko Lindner



On the Spectra and Pseudospectra of a Class of Non-Self-Adjoint Random Matrices and Operators

Simon N. Chandler-Wilde[*], Ratchanikorn Chonchaiya[†] and Marko Lindner[‡]

June 30, 2011

Abstract. In this paper we develop and apply methods for the spectral analysis of non-self-adjoint tridiagonal infinite and finite random matrices, and for the spectral analysis of analogous deterministic matrices which are pseudo-ergodic in the sense of E. B. Davies (Commun. Math. Phys. 216 (2001), 687–704). As a major application to illustrate our methods we focus on the "hopping sign model" introduced by J. Feinberg and A. Zee (Phys. Rev. E 59 (1999), 6433–6443), in which the main objects of study are random tridiagonal matrices which have zeros on the main diagonal and random ± 1's as the other entries. We explore the relationship between spectral sets in the finite and infinite matrix cases, and between the semi-infinite and bi-infinite matrix cases, for example showing that the numerical range and p-norm ε-pseudospectra ($\varepsilon > 0$, $p \in [1, \infty]$) of the random finite matrices converge almost surely to their infinite matrix counterparts, and that the finite matrix spectra are contained in the infinite matrix spectrum Σ. We also propose a sequence of inclusion sets for Σ which we show is convergent to Σ, with the nth element of the sequence computable by calculating smallest singular values of (large numbers of) $n \times n$ matrices. We propose similar convergent approximations for the 2-norm ε-pseudospectra of the infinite random matrices, these approximations sandwiching the infinite matrix pseudospectra from above and below.

Mathematics subject classification (2000): Primary 47B80; Secondary 47A10, 47B36.
Keywords: random matrix, spectral theory, Jacobi matrix, operators on ℓ^p.

1 Introduction

In the last fifteen years there have been many studies of the spectra and pseudospectra of infinite random tridiagonal matrices in the non-self-adjoint case, and of the relationship of the spectral sets of these infinite matrices to those of corresponding large finite random $n \times n$ matrices (see e.g. [21, 16, 32, 17, 19, 39, 13, 22, 40, 31, 30] and the references therein). In this paper we contribute to this literature, introducing new methods of analysis and computation with emphasis throughout, as a major case study, on applying these techniques to understand the "hopping sign model" introduced by Feinberg and Zee [17], further studied in Holz, Orland and Zee [22], by ourselves previously in [5], and see also [11, 12] and [40, Section 37]. In this model the main object of study

[*]Email: S.N.Chandler-Wilde@reading.ac.uk
[†]Email: ratchanikorn@buu.ac.th
[‡]Email: marko.lindner@mathematik.tu-chemnitz.de

is the order n tridiagonal matrix given, for $n \geq 2$, by

$$A_n^b = \begin{pmatrix} 0 & 1 & & & \\ b_1 & 0 & 1 & & \\ & b_2 & 0 & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & b_{n-1} & 0 \end{pmatrix},$$

where $b = (b_1, \ldots, b_{n-1}) \in \mathbb{C}^{n-1}$ and each $b_j = \pm 1$. (For $n = 1$ we set $A_n^b = (0)$.)

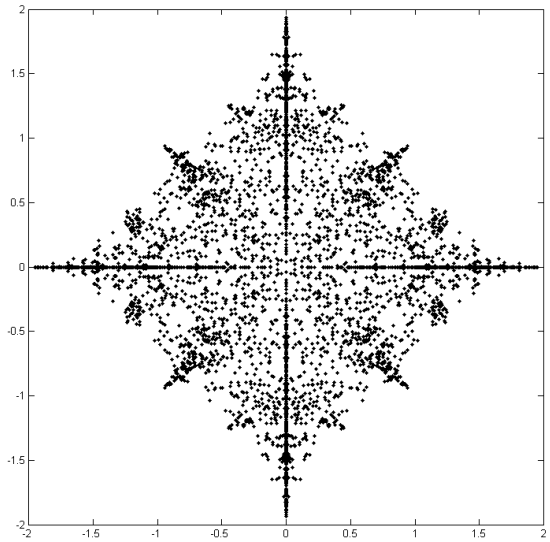


Figure 1: A plot of spec A_n^b, the set of eigenvalues of A_n^b, for a randomly chosen $b \in \{\pm 1\}^{n-1}$, with $n = 5000$ and the components b_j of b independently and identically distributed, with each b_j equal to 1 with probability 1/2. Note the symmetry about the real and imaginary axes by Lemma 3.4 below, and that the spectrum is contained in the square with corners at ± 2 and ± 2i by Lemma 3.1 below.

The objectives we set ourselves in this paper are to understand the behaviour of the spectrum and pseudospectrum of the matrix A_n^b, the spectrum and pseudospectrum of the corresponding semi-infinite and bi-infinite matrices, and the relationship between these spectral sets in the finite and infinite cases. Emphasis will be placed on asymptotic behaviour of the spectrum and pseudospectrum of the finite matrix A_n^b as $n \to \infty$, and we will be interested particularly in the case when the b_j are random variables, for example independent and identically distributed (iid), with $\Pr(b_j = 1) = 0.5$ for each j. (A visualisation of spec A_n^b for a realisation of this random matrix with $n = 5000$ is shown in Figure 1; cf. [17].) To be more precise, we will focus on the case when the vector $b \in \{\pm 1\}^{n-1}$ is the first $n-1$ terms of an infinite sequence $(b_1, b_2, \ldots)$, with each $b_j = \pm 1$, which is *pseudo-ergodic* in the sense introduced by Davies [14], which simply means that every finite sequence of ± 1's appears somewhere in $(b_1, b_2, \ldots)$ as a consecutive sequence. If the b_j are random variables then, for a large class of probability distributions for the b_j, in particular if each b_j is iid with $\Pr(b_j = 1) \in (0, 1)$ for each j, it is clear that the sequence $(b_1, b_2, \ldots)$ is pseudo-ergodic almost surely (with probability one). Thus, although pseudo-ergodicity is a purely deterministic property, our results assuming pseudo-ergodicity have immediate and significant corollaries for the case when A_n^b is a random matrix.

Our interest in studying this problem is in making a contribution to the understanding of the relationship between the spectral properties of finite random matrices and corresponding infinite random matrices in the difficult *non-normal* case. (We note that A_n^b is self-adjoint only in the special case that each $b_j = 1$, and it is an easy calculation that A_n^b is normal, i.e. A_n^b commutes with its transpose, only if $b_1 = b_2 = \ldots = b_{n-1}$.) For an interesting introduction to the behaviour of random matrices in the non-normal case see [40]. Our focus in this paper is on the particular matrix A_n^b and especially its infinite counterparts, but in the course of this investigation we develop and apply methods applicable to the study of spectral sets for the much larger classes of infinite tridiagonal or banded matrices.

Our study of the particular matrix A_n^b, with each $b_j = \pm 1$, is motivated by interest expressed in this class of random matrix in the physics literature [16, 22, 11, 12]. Despite this interest there are so far no rigorous mathematical results on the behaviour of the spectrum of A_n^b in the limit as $n \to \infty$. This paper makes steps in this direction. A further motivation for studying the particular matrix class A_n^b is that rigorous results are available on the asymptotics of the spectrum and resolvent norm for a related class of matrices, offering some hope that progress might be possible in this case also. This related class is the case when, rather than the first sub-diagonal consisting of random ± 1's, the diagonal has random ± 1's. Of course, the matrix is then upper-triangular, so that many computations become explicit; in particular the spectrum of the finite matrix is just $\{1, -1\}$ and the spectra of the corresponding infinite matrices can be explicitly calculated: see [39, 8, 29] for details. We shall see that the situation in the case studied in this paper is, in a number of respects, rather richer and the analysis more delicate. At the same time in a number of respects our results are more complete: for example, we are able to prove convergence of the pseudospectra of A_n^b to those of the corresponding infinite matrices, and to do this not just in a Hilbert space setting but in p-norm for $p \in [1, \infty]$.

The distinctive flavour of the results we develop in this paper, with their significant emphasis on pseudospectra and the relationship between finite random matrices and their infinite matrix counterparts, is in large part inspired by the paper by Trefethen, Contedini and Embree [39], by Part VI on random matrices in [40], and by results on convergence of the p-norm pseudospectra ($1 < p < \infty$) and numerical ranges of $n \times n$ Toeplitz matrices due to Böttcher, Grudsky, and Silbermann [1] and Roch [36], described recently in the monograph of Böttcher and Grudsky [2].

Let $\mathbb{N}$ denote the set of positive integers and $\mathbb{Z}$ the set of integers. Throughout, $\{\pm 1\}^{\mathbb{Z}}$, $\{\pm 1\}^{\mathbb{N}}$, and $\{\pm 1\}^m$, for $m \in \mathbb{N}$, denote the sets of vectors in $\ell^\infty(\mathbb{Z})$, $\ell^\infty(\mathbb{N})$ and $\mathbb{C}^m$, respectively, whose entries $b_j = \pm 1$. The related infinite-dimensional operators we study include the operators A_+^b, for $b = (b_1, b_2, \ldots) \in \ell^\infty(\mathbb{N})$, especially when each $b_j = \pm 1$. Here A_+^b acts on the sequence space $\ell^p(\mathbb{N})$, for $p \in [1, \infty]$, by the action

$$(A_+^b x)_i = \sum_{j \in \mathbb{N}} (A_+^b)_{ij} x_j, \quad i \in \mathbb{N}, \tag{1}$$

where $(A_+^b)_{ij} = b_{i-1}\delta_{i-1,j} + \delta_{i+1,j}$ and δ_{ij} is the usual Kronecker delta. In other words, A_+^b acts by multiplication by the infinite matrix

$$A_+^b = \begin{pmatrix} 0 & 1 & & \\ b_1 & 0 & 1 & \\ & b_2 & 0 & \ddots \\ & & \ddots & \ddots \end{pmatrix},$$

which has entry $(A_+^b)_{ij}$ in row i, column j, for $i, j \in \mathbb{N}$. (For simplicity, we make no distinction in our notation between A_+^b and its matrix representation.) A main aim of the paper will be to compute the spectrum, pseudospectrum, and numerical range of A_+^b in the case when $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic. We shall also study the same properties of the corresponding operator A^b which

acts on $\ell^p(\mathbb{Z})$, again focusing on the case when $b \in \{\pm 1\}^{\mathbb{Z}}$ is pseudo-ergodic. The action of A^b is given by the same formula (1) but now with $b \in \ell^\infty(\mathbb{Z})$ and with $\mathbb{N}$ replaced by $\mathbb{Z}$. In other words, A^b acts by mutiplication by the bi-infinite matrix

$$A^b := \begin{pmatrix} \ddots & \ddots & & & \\ \ddots & 0 & 1 & & \\ & b_{-1} & \boxed{0} & 1 & \\ & & b_0 & 0 & \ddots \\ & & & \ddots & \ddots \end{pmatrix}, \tag{2}$$

where the box marks the matrix entry at $(0,0)$. Our results will also apply, through the application of similarity transforms, to the more general matrices

$$A_n^{b,c} = \begin{pmatrix} 0 & c_1 & & & \\ b_1 & 0 & c_2 & & \\ & b_2 & 0 & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & b_{n-1} & 0 \end{pmatrix}, \tag{3}$$

in the case when $b_j = \pm 1$ and $c_j = \pm 1$, and to the corresponding infinite matrices

$$A_+^{b,c} = \begin{pmatrix} 0 & c_1 & & \\ b_1 & 0 & c_2 & \\ & b_2 & 0 & \ddots \\ & & \ddots & \ddots \end{pmatrix} \quad \text{and} \quad A^{b,c} := \begin{pmatrix} \ddots & \ddots & & & \\ \ddots & 0 & c_{-1} & & \\ & b_{-1} & \boxed{0} & c_0 & \\ & & b_0 & 0 & \ddots \\ & & & \ddots & \ddots \end{pmatrix}. \tag{4}$$

1.1 The Main Results

Let us summarise the main results that we obtain in this paper, first introducing a few key notations and definitions. Throughout, where B is a bounded linear operator on $\ell^p(S)$, for some $p \in [1,\infty]$, with $S = \mathbb{Z}$ or $\mathbb{N}$, or where B is a square matrix, we denote by $\operatorname{spec} B$ the *spectrum* of B, i.e. the set of $\lambda \in \mathbb{C}$ for which $B - \lambda I$ (I the identity matrix or operator) is not invertible. (We note that the spectra of A^b and A_+^b do not depend on $p \in [1,\infty]$, from general results on band operators (e.g. [26]); of course, when B is a matrix, the spectrum is just the set of eigenvalues of B.) Throughout, $\|x\|_p$, for $p \in [1,\infty]$, will be our notation for the standard p-norm of x, for $x \in \ell^p(S)$, with $S = \mathbb{Z}$ or $\mathbb{N}$, or $x \in \mathbb{C}^m$, for some $m \in \mathbb{N}$. Where B is an operator or matrix, $\|B\|_p$ will denote the norm of B induced by the vector norm $\|\cdot\|_p$, i.e. $\|B\|_p := \sup_{\|x\|_p=1} \|Bx\|_p$. With this notation, following e.g. [40], for $p \in [1,\infty]$ and $\varepsilon > 0$ we define the ℓ^p ε-pseudospectrum of B, $\operatorname{spec}_\varepsilon^p B$, by

$$\operatorname{spec}_\varepsilon^p B := \operatorname{spec} B \cup \{\lambda \in \mathbb{C} : \|(B-\lambda I)^{-1}\|_p > \varepsilon^{-1}\}.$$

When B is a bounded linear operator on $\ell^p(S)$, for some $p \in [1,\infty]$ and $S = \mathbb{Z}$ or $\mathbb{N}$, in general the spectrum of B is larger than the set of eigenvalues of B. We let $\operatorname{spec}_{\text{point}}^p B$ denote the set of eigenvalues of B considered as an operator on $\ell^p(S)$, i.e.

$$\operatorname{spec}_{\text{point}}^p B := \{\lambda \in \mathbb{C} : Bx = \lambda x, \text{ for some } x \in \ell^p(S) \text{ with } x \neq 0\}.$$

A key result we obtain on the spectra of our infinite matrices, in large part through limit operator arguments described in Section 2, is the following (cf. [14]): if $b, c, d \in \{\pm 1\}^{\mathbb{N}}$, $\tilde{b}, \tilde{c}, \tilde{d} \in \{\pm 1\}^{\mathbb{Z}}$, and b, $\tilde{b}$, cd, and $\tilde{c}\tilde{d}$ are all pseudo-ergodic, then

$$\operatorname{spec} A_+^b = \operatorname{spec} A_+^{c,d} = \operatorname{spec} A^{\tilde{b}} = \operatorname{spec} A^{\tilde{c},\tilde{d}} = \Sigma := \bigcup_{e \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec} A^e = \bigcup_{e \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec}^{\infty}_{\text{point}} A^e. \quad (5)$$

One surprising aspect of this formula is that the semi-infinite and bi-infinite matrices share the same spectrum, in contrast to many of the cases discussed in [39], this connected to the symmetries that we explore in Section 3.

We do not know a simple test for membership of the set Σ given by this characterisation (though see Figures 2 and 3 below for plots of known subsets of Σ, and see Section 4.3 for an algorithm for computing approximations to Σ). But this result implies that $\operatorname{spec} A^b \subset \Sigma$ for every $b \in \{\pm 1\}^{\mathbb{Z}}$ which gives the possibility of determining subsets of Σ by computing $\operatorname{spec} A^b$ for particular choices of b. In particular, as recalled in Section 2, when b is n-periodic for some $n \in \mathbb{N}$, i.e. $b_{j+n} = b_j$ for $j \in \mathbb{Z}$, $\operatorname{spec} A^b$ can be computed by calculating eigenvalues of an order n matrix (a periodised version of A_n^b). We compute $\pi_n \subset \Sigma$, for $n = 5, 10, ..., 30$ in Section 2, where π_n denotes the union of $\operatorname{spec} A_b$ over all n-periodic $b \in \{\pm 1\}^{\mathbb{Z}}$. We speculate at the end of the paper that

$$\pi_\infty := \bigcup_{n \in \mathbb{N}} \pi_n \quad (6)$$

is dense in Σ, and it has been shown recently in [9] that certainly π_∞ is dense in the unit disc $\mathbb{D} = \{z : |z| < 1\}$, which implies that $\overline{\mathbb{D}} \subset \Sigma$, as established slightly earlier directly from (5) in [5]. (Throughout, $\overline{S}$ denotes the closure of a set $S \subset \mathbb{C}$: for an element $z \in \mathbb{C}$, $\bar{z}$ denotes the complex conjugate.)

To obtain a first upper bound on Σ we compute the ℓ^2-numerical range, $W(A^b)$, of A^b when b is pseudo-ergodic. We show that, if $b, c, d \in \{\pm 1\}^{\mathbb{N}}$, $\tilde{b}, \tilde{c}, \tilde{d} \in \{\pm 1\}^{\mathbb{Z}}$, and b, $\tilde{b}$, cd, and $\tilde{c}\tilde{d}$ are all pseudo-ergodic, then

$$W(A_+^b) = W(A_+^{c,d}) = W(A^{\tilde{b}}) = W(A^{\tilde{c},\tilde{d}}) = \Delta := \{z = a + \mathrm{i}b : a, b \in \mathbb{R},\ |a| + |b| < 2\}.$$

Since the spectrum is necessarily contained in the closure of the numerical range, this implies that

$$\overline{\mathbb{D}} \subset \Sigma \subset \overline{\Delta}.$$

We point out that the numerical range of A_n^b converges to that of A^b, in particular that $W(A_n^b) \nearrow \Delta$ as $n \to \infty$, if b is pseudo-ergodic. (Here and throughout, for $T_n \subset \mathbb{C}$ and $T \subset \mathbb{C}$, the notation $T_n \nearrow T$ means that $T_n \subset T$ for each n and that $\operatorname{dist}(T, T_n) \to 0$ as $n \to \infty$, with $\operatorname{dist}(T, T_n)$ the Hausdorff distance defined in (16) below.)

The largest part of the paper (Section 4) is an investigation of the relationship between the finite and infinite matrix cases with respect to behaviour of spectra and pseudospectra. The spectral case is harder: our main result is to show that the spectra of the finite matrices are subsets of the infinite matrix spectra, precisely that, for every n and every $c \in \{\pm 1\}^{n-1}$,

$$\operatorname{spec} A_n^c \subset \pi_{2n+2} \subset \Sigma,$$

so that $\sigma_n := \bigcup_{c \in \{\pm 1\}^{n-1}} \operatorname{spec} A_n^c \subset \pi_{2n+2} \subset \Sigma$ and

$$\sigma_\infty := \bigcup_{n \in \mathbb{N}} \sigma_n \subset \pi_\infty \subset \Sigma. \quad (7)$$

We suspect that $\operatorname{spec} A_n^b \nearrow \Sigma = \operatorname{spec} A_+^b$ as $n \to \infty$, if $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic, and the numerical results in Figures 1 and 2, and other similar computations, are suggestive of a conjecture that $\operatorname{spec} A_n^b \nearrow \pi_\infty$, which set, as mentioned already, we speculate is dense in Σ.

We can prove neither of these last two conjectures about spectral asymptotics. On the other hand, our theoretical results for the pseudospectrum are fairly complete. We show first in Theorem 3.6 a pseudospectral version of (5), that, if $b, c, d \in \{\pm 1\}^{\mathbb{N}}$, $\tilde{b}, \tilde{c}, \tilde{d} \in \{\pm 1\}^{\mathbb{Z}}$, and b, $\tilde{b}$, cd, and $\tilde{c}\tilde{d}$ are all pseudo-ergodic, then, for $p \in [1, \infty]$ and $\varepsilon > 0$,

$$\operatorname{spec}_\varepsilon^p A_+^b = \operatorname{spec}_\varepsilon^p A_+^{c,d} = \operatorname{spec}_\varepsilon^p A^{\tilde{b}} = \operatorname{spec}_\varepsilon^p A^{\tilde{c},\tilde{d}} = \Sigma_\varepsilon^p := \bigcup_{e \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec}_\varepsilon^p A^e.$$

We then show that the pseudospectra of the large finite matrices are contained in and are well-approximated by the pseudospectra of the infinite matrices, and that this works for p-norm pseudospectra for the full range $p \in [1, \infty]$. Precisely, for $p \in [1, \infty]$ and $\varepsilon > 0$, we show that, if b is pseudo-ergodic, then

$$\operatorname{spec}_\varepsilon^p A_n^b \nearrow \Sigma_\varepsilon^p \tag{8}$$

as $n \to \infty$.

This last result, linking the pseudospectra of A_+^b with those of its finite sections A_n^b, is a somewhat unexpectedly satisfactory result. Even in the case in which the theory of the finite section method is arguably simplest and most well-understood, namely the case of the Toeplitz operator (a semi-infinite Toeplitz matrix), the limit as $n \to \infty$ of the ε-pseudospectra of the $n \times n$ finite section Toeplitz matrices has been calculated only relatively recently, and only for $p \in (1, \infty)$ [1, 2]. Moreover, except for the special case $p = 2$, this limit is not, in general, just the ℓ^p ε-pseudospectrum of the Toeplitz operator, but rather the union of the ℓ^p and ℓ^q ε-pseudospectra, with $p^{-1} + q^{-1} = 1$. (A component of the explanation of (8) is that we show in Lemma 3.2 that $\Sigma_\varepsilon^p = \Sigma_\varepsilon^q$ for $p^{-1} + q^{-1} = 1$.)

Equation (8) leads to characterisations of the spectrum Σ which, in principle, can be used for numerical approximation. Since $\bigcap_{\varepsilon > 0} \Sigma_\varepsilon^p = \Sigma$, it holds that

$$\Sigma = \lim_{\varepsilon \to 0} \Sigma_\varepsilon^p = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \operatorname{spec}_\varepsilon^p A_n^b, \tag{9}$$

for every $p \in [1, \infty]$ and pseudo-ergodic b. However, the formula (9) is not guaranteed to give useful results for any fixed ε and n as the convergence as $n \to \infty$ may be arbitrarily slow, as discussed in Section 4.3. In that section we develop alternative, much more useful, convergent sequences of computable, upper and lower bounds for Σ_ε^2 and a convergent sequence of computable upper bounds for Σ. We show firstly that

$$\sigma_{n,\varepsilon}^2 := \bigcup_{c \in \{\pm 1\}^{n-1}} \operatorname{spec}_\varepsilon^2 A_n^c \subset \Sigma_\varepsilon^2 \subset \sigma_{n,\varepsilon+\varepsilon_n}^2 = \bigcup_{c \in \{\pm 1\}^{n-1}} \operatorname{spec}_{\varepsilon+\varepsilon_n}^2 A_n^c,$$

giving explicit expressions for the ε_n which satisfy that $\varepsilon_n = O(n^{-1})$ as $n \to \infty$, and showing that $\sigma_{n,\varepsilon}^2 \nearrow \Sigma_\varepsilon^2$ and $\sigma_{n,\varepsilon+\varepsilon_n}^2 \searrow \Sigma_\varepsilon^2$ as $n \to \infty$. (The notation $T_n \searrow T$ means that $T \subset T_n$ for each n and that $\operatorname{dist}(T, T_n) \to 0$ as $n \to \infty$.) Then, taking the intersection over all ε, we deduce that

$$\sigma_n = \bigcup_{c \in \{\pm 1\}^{n-1}} \operatorname{spec} A_n^c \subset \Sigma \subset \overline{\sigma_{n,\varepsilon_n}^2},$$

and prove that

$$\overline{\sigma_{n,\varepsilon_n}^2} \searrow \Sigma \quad \text{as} \quad n \to \infty.$$

In a substantial series of numerical calculations, we compute these convergent upper bounds $\overline{\sigma_{n,\varepsilon_n}^2}$ for the spectrum Σ in Section 4.3, and through these calculations demonstrate that Σ is a strict subset of $\overline{\Delta}$.

All these results have implications for the behaviour of the spectral sets of A^b, A_+^b, A_n^b, $A^{b,c}$, $A_+^{b,c}$, and $A_n^{b,c}$, when the entries $b_j = \pm 1$ and $c_j = \pm 1$ are random, and we make explicit these

implications in a final Theorem 5.1, in the same section summarising succintly what we have established about the spectral sets Σ and Σ_ε^p (Theorem 5.2), and outlining a number of open problems.

In the course of this investigation, focused on a particular operator and matrix class, we develop results for the larger classes of tridiagonal or banded finite and infinite matrices. In particular, Theorem 4.4 shows that, for $p \in [1, \infty]$, $\varepsilon > 0$, the ℓ^p ε-pseudospectrum of a general, semi-infinite tridiagonal matrix is contained, for $\varepsilon' > \varepsilon$, in the ℓ^p ε'-pseudospectrum of its $n \times n$ finite section if n is sufficiently large. It also shows corresponding results relating the pseudospectra of a general bi-infinite matrix to that of its finite sections. In Section 2 we employ recent work [7, 8] on limit operator methods for the study of spectral sets for very general classes of infinite matrices. We make explicit in Theorems 2.1 and 2.9 the implications of this work for the essential spectrum, spectrum, and pseudospectra of bi-infinite and semi-infinite banded matrices with numerical (as opposed to operator-valued) entries. In Section 4.3 we make the first substantive application of a new method which generates sequences of inclusion sets for the spectra and pseudospectra of a tridiagonal operator, demonstrating, through this application, that these sequences of inclusion sets can in fact converge to the spectral sets that they enclose.

We mention that first versions of a number of the results in this paper are contained in the PhD thesis of the second author [10], and that a number of the results were announced (without proofs) in [5].

1.2 Pseudospectra and the Numerical Range

We shall need throughout the paper a number of properties of the ε-pseudospectra of a bounded linear operator B on a Banach space X, and of the pseudospectra of its adjoint operator B^* on the dual space X^* (dual in the sense e.g. of [24], so that X^* is the set of bounded anti-linear functionals, and the spectrum of B^* is the complex conjugate of the spectrum of B). We summarise these properties in this section, pointing out how the theory of pseudospectra in the Banach space setting has recently been significantly clarified by work of Shargorodsky [37]. The properties we shall need include the equivalent definitions encapsulated in the following theorem:

Theorem 1.1 *The ε-pseudospectrum of a bounded linear operator B on a Banach space X is defined, for $\varepsilon > 0$, by any one of the following equivalent definitions:*

(i) $\operatorname{spec}_\varepsilon B = \operatorname{spec} B \cup \{\lambda \in \mathbb{C} : \|(B - \lambda I)^{-1}\| > \varepsilon^{-1}\}$;

(ii) $\operatorname{spec}_\varepsilon B = \operatorname{spec} B \cup \{\lambda \in \mathbb{C} : \nu(B - \lambda I) < \varepsilon\}$, *where $\nu(C)$ is the lower norm of a bounded linear operator C, defined by $\nu(C) := \inf_{\|x\|=1} \|Cx\|$;*

(iii) $\operatorname{spec}_\varepsilon B$ *is the union of* $\operatorname{spec} B$ *and the set* $\operatorname{spec}_{\text{point},\varepsilon} B$ *of ε-pseudoeigenvalues of B, where λ is an ε-pseudoeigenvalue if there exists $x \in X$ with $\|x\| = 1$ and $\|(B - \lambda I)x\| < \varepsilon$;*

(iv) $\operatorname{spec}_\varepsilon B$ *is the union of* $\operatorname{spec}_{\text{point},\varepsilon} B$ *and the complex conjugate of* $\operatorname{spec}_{\text{point},\varepsilon} B^*$;

(v) $\operatorname{spec}_\varepsilon B = \bigcup_{\|E\|<\varepsilon} \operatorname{spec}(B + E)$, *the union taken over all bounded linear operators E with $\|E\| < \varepsilon$.*

For a proof of the equivalence of (i)-(v), and a useful short introduction to the pseudospectra of linear operators on Banach spaces, see [40, Section 4]. We will use the equivalence of (i)-(iv) throughout. The equivalence of the other definitions with (v), and the connection this makes with spectra of perturbed operators, is a significant motivation for the practical interest in pseudospectra. It is clear from the above definition that $\operatorname{spec}_\varepsilon B$ is an open set for $\varepsilon > 0$. An elementary but

important property of the lower norm is that

$$|\nu(A) - \nu(B)| \leq \|A - B\|, \tag{10}$$

for any bounded linear operators A and B on X.

In the case when, for some $N \in \mathbb{N}$, $X = \mathbb{C}^N$ and B is an $N \times N$ matrix, (i)-(v) are equivalent additionally to $\operatorname{spec}_\varepsilon B = \{\lambda \in \mathbb{C} : \nu(B - \lambda I) < \varepsilon\} = \operatorname{spec}_{\mathrm{point},\varepsilon} B$. If $\|\cdot\| = \|\cdot\|_2$, then, for every $N \times N$ matrix A, $\nu(A) = s_{\min}(A)$, the smallest singular value of A. Thus these definitions are additionally equivalent to [40]

$$\operatorname{spec}_\varepsilon B = \{\lambda \in \mathbb{C} : s_{\min}(B - \lambda I) < \varepsilon\}. \tag{11}$$

Note that (10) implies that

$$|s_{\min}(B - \lambda I) - s_{\min}(B - \mu I)| \leq |\lambda - \mu|, \quad \lambda, \mu \in \mathbb{C}. \tag{12}$$

It is equation (11) that we use for the numerical computations of pseudospectra in Section 4.3.

An alternative definition of the pseudospectrum is to replace the strict inequality $>$ in (i) by $\geq$, so that the ε-pseudospectrum is defined to be

$$\operatorname{Spec}_\varepsilon B = \operatorname{spec} B \cup \{\lambda \in \mathbb{C} : \|(B - \lambda I)^{-1}\| \geq \varepsilon^{-1}\}.$$

This has the attraction that $\operatorname{Spec}_\varepsilon B$, like $\operatorname{spec} B$, is a compact set for $\varepsilon > 0$. An interesting question is whether $\overline{\operatorname{spec}_\varepsilon B} = \operatorname{Spec}_\varepsilon B$, which hinges on the question of whether or not it is possible for the norm of the resolvent of B, $\|(B - \lambda I)^{-1}\|$, to take a finite constant value on a open set $G \subset \mathbb{C}$. Let us say that the Banach space X has the *strong maximum property* if, for every open set $G \subset \mathbb{C}$, every bounded linear operator B on X, and every $M > 0$, it holds that

$$(\|(B - \lambda I)^{-1}\| \leq M, \ \forall\, \lambda \in G) \Rightarrow (\|(B - \lambda I)^{-1}\| < M, \ \forall\, \lambda \in G).$$

If X has the strong maximum property, then no bounded linear operator on X can have a resolvent norm with a constant finite value on an open subset of $\mathbb{C}$, and it is easy to see that $\overline{\operatorname{spec}_\varepsilon B} = \operatorname{Spec}_\varepsilon B$. Recently, Shargorodsky [37] has shown, by constructing explicit counterexamples, that not every Banach space has the strong maximum property. But the following theorem from [37], which extends earlier work of [18], makes clear that the Banach spaces of relevance to this paper do have this property.

Theorem 1.2 *Suppose that X is a Banach space which is either finite-dimensional or is such that either X or X^* is complex uniformly convex (as defined e.g. in [37]). Then X has the strong maximum property. In particular, X has the strong maximum property if X is a Hilbert space, or if $X = \ell^p(S)$, for $S = \mathbb{N}$ or $\mathbb{Z}$ and $p \in [1, \infty]$.*

It is clear from (v) and standard operator perturbation arguments (see [40] for details) that, for $0 < \varepsilon < \varepsilon'$, $\operatorname{spec} B \subset \operatorname{spec}_\varepsilon B \subset \operatorname{spec}_{\varepsilon'} B$, and that

$$\varepsilon\mathbb{D} + \operatorname{spec} B \subset \operatorname{spec}_\varepsilon B. \tag{13}$$

In fact [40] $\varepsilon\mathbb{D} + \operatorname{spec} B = \operatorname{spec}_\varepsilon B$ if X is a Hilbert space and B is normal, i.e. $BB^* = B^*B$. Further [40]

$$\operatorname{spec} B = \bigcap_{\varepsilon > 0} \operatorname{spec}_\varepsilon B. \tag{14}$$

Generalising (13), it holds that [40]

$$\delta\mathbb{D} + \operatorname{spec}_\varepsilon B \subset \operatorname{spec}_{\delta+\varepsilon} B, \quad \text{for} \quad \varepsilon, \delta > 0. \tag{15}$$

For $S, T \subset \mathbb{C}$, let

$$\operatorname{dist}(S,T) := \max(\sup\{\operatorname{dist}(z,S) : z \in T\}, \sup\{\operatorname{dist}(z,T) : z \in S\}). \tag{16}$$

(This notion of distance, when applied to compact subsets of $\mathbb{C}$, is an instance of the Hausdorff distance between compact subsets of a metric space.) Given a sequence $T_n \subset \mathbb{C}$ and $T \subset \mathbb{C}$, let us write $T_n \to T$ if $\operatorname{dist}(T_n, T) \to 0$ as $n \to \infty$. Additionally, let us write $T_n \nearrow T$ if $T_n \to T$ and $T_n \subset T$ for each n, and write $T_n \searrow T$ if $T_n \to T$ and $T \subset T_n$ for each n. It is an easy calculation to show that

$$\operatorname{spec}_\varepsilon B \searrow \operatorname{spec} B \ \text{ as } \ \varepsilon \to 0^+. \tag{17}$$

Similarly, it holds for $\varepsilon > 0$ that $\operatorname{spec}_{\varepsilon'} B \searrow \operatorname{Spec}_\varepsilon B$, as $\varepsilon' \to \varepsilon^+$, and $\operatorname{spec}_{\varepsilon'} B \nearrow \operatorname{spec}_\varepsilon B$, as $\varepsilon' \to \varepsilon^-$. Thus, in the case where X has the strong maximum property so that $\overline{\operatorname{spec}_\varepsilon B} = \operatorname{Spec}_\varepsilon B$, it holds for $\varepsilon > 0$ that

$$\operatorname{spec}_{\varepsilon'} B \searrow \operatorname{spec}_\varepsilon B, \text{ as } \varepsilon' \to \varepsilon^+, \ \text{ and } \ \operatorname{spec}_{\varepsilon'} B \nearrow \operatorname{spec}_\varepsilon B, \text{ as } \varepsilon' \to \varepsilon^-, \tag{18}$$

so that $\operatorname{spec}_\varepsilon B$ depends continuously on ε.

The spectrum and ε-pseudospectra are connected to the numerical range. In the case that X is a Hilbert space with inner product $(\cdot,\cdot)$, and where B is a bounded linear operator on X, the *numerical range* or *field of values* of B, denoted $W(B)$, is the set

$$W(B) := \{(Bx, x) : x \in X, \|x\| = 1\}.$$

It is well known that this numerical range is a convex set and that $\operatorname{spec} B \subset \overline{W(B)}$, in fact $\operatorname{spec} B \subset W(B)$ if X is finite-dimensional. The relationship with the ε-pseudospectra is that, similarly, $\operatorname{spec}_\varepsilon B \subset W(B) + \varepsilon\mathbb{D}$, for $\varepsilon > 0$ [40, Section 17]. Let Y be a closed subspace of X, $P : X \to Y$ orthogonal projection onto Y, and let $B_Y := PB|_Y$. Then

$$W(B_Y) = \{(B_Y x, x) : x \in Y, \|x\| = 1\} = \{(Bx, x) : x \in Y, \|x\| = 1\} \subset W(B). \tag{19}$$

This observation is one component in the following result [20, Theorem 3.52]:

Theorem 1.3 *Suppose that X is a Hilbert space and that $(P_n)_{n\in\mathbb{N}}$ is a sequence of orthogonal projection operators on X that converges strongly to the identity operator ($P_n x \to x$ as $n \to \infty$, for every $x \in X$). Then, for every bounded linear operator B on X, where $B_n := P_n B|_{X_n}$ with $X_n = P_n(X)$, it holds that*

$$W(B_n) \nearrow W(B) \quad \text{as} \quad n \to \infty.$$

2 Results by Limit Operator Arguments

Let us start this section by establishing a few additional notations and definitions. Throughout the remainder of the paper, if B is a bounded linear operator on a Banach space X we will say that B is Fredholm if $B(X)$, the range of B, is closed and if, additionally, $\alpha(B) := \dim(\ker B)$, the dimension of the null-space of B, and $\beta(B) := \dim(X/B(X))$, the co-dimension of the range of B, are both finite, in which case we define the index of B by $\operatorname{ind} B := \alpha(B) - \beta(B)$. We will let $\operatorname{spec}_{\text{ess}} B$ denote the *essential spectrum* of B, i.e. the set of $\lambda \in \mathbb{C}$ for which $B - \lambda I$ is not Fredholm. Let M_b be the bounded linear operator which operates on the standard sequence space $\ell^p(\mathbb{Z})$, for $p \in [1, \infty]$, by multiplication by $b \in \ell^\infty(\mathbb{Z})$. Explicitly, for $y \in \ell^p(\mathbb{Z})$,

$$(M_b y)_j = b_j y_j, \quad j \in \mathbb{Z}.$$

Moreover, for $k \in \mathbb{Z}$ let V_k denote the shift operator defined by

$$(V_k y)_j = y_{j-k}, \quad j \in \mathbb{Z},$$

and note that $V_j M_b = M_{V_j b} V_j$, for $j \in \mathbb{Z}$, $b \in \ell^\infty(\mathbb{Z})$. In terms of these notations, the operators A^b and $A^{b,c}$, corresponding to the infinite matrices (2) and (4), can be written as

$$A^b = V_1 M_b + V_{-1} \quad \text{and} \quad A^{b,c} = V_1 M_b + M_c V_{-1}. \tag{20}$$

We will use these notations for $b, c \in \ell^\infty(\mathbb{Z})$, but especially for $b, c \in \{\pm 1\}^{\mathbb{Z}}$.

One major tool for computing the spectrum of the infinite matrices A^b and $A^{b,c}$, with $b, c \in \ell^\infty(\mathbb{Z})$, is the method of so-called limit operators [8, 26, 35]. In this method a bi-infinite matrix B is studied in terms of a family of infinite matrices that represents the behaviour of B at infinity. More precisely, let A be a banded matrix $A = (a_{ij})_{i,j\in\mathbb{Z}}$, with $\sup_{ij} |a_{ij}| < \infty$, so that the operator induced by A is a bounded operator on $\ell^p(\mathbb{Z})$, for all $p \in [1, \infty]$. We say that the operator induced by the matrix $B = (b_{ij})_{i,j\in\mathbb{Z}}$ is a *limit operator* of the operator induced by A if, for a sequence $h_1, h_2, \ldots$ of integers with $|h_k| \to \infty$, it holds that

$$a_{i+h_k, j+h_k} \;\to\; b_{ij} \qquad \text{as} \qquad k \to \infty,$$

for all $i, j \in \mathbb{Z}$. The set of all limit operators of A is denoted by $\sigma^{\mathsf{op}}(A)$. In some instances it is useful to think of $\sigma^{\mathsf{op}}(A)$ as the union of two subsets, as $\sigma^{\mathsf{op}}(A) = \sigma^{\mathsf{op}}_+(A) \cup \sigma^{\mathsf{op}}_-(A)$, where $\sigma^{\mathsf{op}}_\pm(A)$ denotes the subset of those limit operators associated with sequences h with $h_k \to \pm\infty$. It is an easy consequence of the Bolzano-Weierstrass theorem and a diagonal argument that each of $\sigma^{\mathsf{op}}_\pm(A)$ is non-empty, and it is clear that if $B = (b_{ij})$ is a limit operator of A then $\sup_{i-j=k} |b_{ij}| \leq \sup_{i-j=k} |a_{ij}|$, for every $k \in \mathbb{Z}$. In particular, if $A = A^{b,c}$ for some $b, c \in \{\pm 1\}^{\mathbb{Z}}$ and B is a limit operator of A, then $B = A^{\tilde{b},\tilde{c}}$ for some $\tilde{b}, \tilde{c} \in \{\pm 1\}^{\mathbb{Z}}$.

The following theorem, which applies in particular to A^b and to $A^{b,c}$, connects the essential spectrum with the set of limit operators. This result is a particular case of much more general results from [7], [8, Theorem 6.28, Corollary 6.49], which extend a main theorem on limit operators going back to [25, 34]. Note that the spectrum, as an operator on $\ell^p(\mathbb{Z})$, of an infinite banded matrix $A = (a_{ij})_{i,j\in\mathbb{Z}}$, with $\sup_{ij} |a_{ij}| < \infty$, does not depend on $p \in [1, \infty]$, and the same is true for the essential spectrum: moreover, if $\lambda \notin \operatorname{spec}_{\mathrm{ess}} A$, then $\operatorname{ind}(A - \lambda I)$ is also independent of p (see [28] or [8, Corollary 6.49]).

Theorem 2.1 *Let A be a banded matrix $A = (a_{ij})_{i,j\in\mathbb{Z}}$, with $\sup_{ij} |a_{ij}| < \infty$. Then*

$$\operatorname{spec}_{\mathrm{ess}} A = \bigcup_{B \in \sigma^{\mathsf{op}}(A)} \operatorname{spec} B = \bigcup_{B \in \sigma^{\mathsf{op}}(A)} \operatorname{spec}^\infty_{\mathrm{point}} B \tag{21}$$

and $\operatorname{spec}^p_\varepsilon B \subset \operatorname{spec}^p_\varepsilon A$, for all $\varepsilon > 0$, $p \in [1, \infty]$, and $B \in \sigma^{\mathsf{op}}(A)$. In particular, if $A \in \sigma^{\mathsf{op}}(A)$, in which case we say that A is self-similar*, then*

$$\operatorname{spec} A = \operatorname{spec}_{\mathrm{ess}} A = \bigcup_{B \in \sigma^{\mathsf{op}}(A)} \operatorname{spec} B = \bigcup_{B \in \sigma^{\mathsf{op}}(A)} \operatorname{spec}^\infty_{\mathrm{point}} B \;\; \textit{and} \;\; \operatorname{spec}^p_\varepsilon A = \bigcup_{B \in \sigma^{\mathsf{op}}(A)} \operatorname{spec}^p_\varepsilon B,$$

for $\varepsilon > 0$ and $p \in [1, \infty]$.

Recall that $\operatorname{spec}^\infty_{\mathrm{point}} B$ is the set of eigenvalues of B in $\ell^\infty(\mathbb{Z})$, so that $\lambda \in \operatorname{spec}^\infty_{\mathrm{point}} B$ iff $\lambda x = Bx$ has a non-trivial bounded solution.

One case where $A^{b,c}$ is self-similar is where (b, c) is periodic with some period $n \in \mathbb{N}$, i.e.

$$b_{j+n} = b_j, \quad c_{j+n} = c_j, \quad j \in \mathbb{Z}. \tag{22}$$

In this case the above theorem applied to $A^{b,c}$ reduces to $\operatorname{spec} A^{b,c} = \operatorname{spec}_{\mathrm{ess}} A^{b,c} = \operatorname{spec}^\infty_{\mathrm{point}} A^{b,c}$, and in fact it is well-known further, e.g. [15], that if $\lambda \in \operatorname{spec} A^{b,c}$ then $\lambda x = A^{b,c} x$ has a solution

which is not only bounded but also quasi-periodic, i.e. for some $\alpha \in \mathbb{C}$ with $|\alpha| = 1$, $x_{k+n} = \alpha x_k$, $k \in \mathbb{Z}$. It is easy to see that this implies that

$$\operatorname{spec} A^{b,c} = \bigcup_{|\alpha|=1} \operatorname{spec} \left(A_n^{b,c} + B_{n,\alpha}^{b,c} \right), \tag{23}$$

where $A_n^{b,c}$ is given by (3) (with $A_1^{b,c} := (0)$) and $B_{n,\alpha}^{b,c}$ is the $n \times n$ matrix whose entry in row i, column j is $\delta_{i,n}\delta_{j,1}\alpha c_n + \delta_{i,1}\delta_{j,n}\alpha^{-1}b_n$, where δ_{ij} is the Kronecker delta. We will abbreviate $B_{n,\alpha}^{b,c}$ as $B_{n,\alpha}^{b}$ in the case that $c = (1, ..., 1)$.

An important case where A^b is self-similar is where A^b is *pseudo-ergodic* in the sense of Davies [14]. The following is a specialisation of the definition from [14].

Definition 2.2 *Call $b \in \{\pm 1\}^{\mathbb{Z}}$ and the operator A^b* pseudo-ergodic *if, for every $N \in \mathbb{N}$ and every $w \in \{\pm 1\}^N$, there exists $J \in \mathbb{Z}$ such that $b_{n+J} = w_n$, for $n = 1, ..., N$.*

We see from this definition that A^b is pseudo-ergodic if and only if every finite sequence of ± 1's appears somewhere in the bi-infinite sequence b. The significance of this definition is that, for many cases where the entries b_n are random variables, the sequence b is pseudo-ergodic with probability one. In particular, the following lemma follows easily from the Second Borel Cantelli Lemma (e.g. [3, Theorem 8.16]), the argument sometimes called the 'Infinite Monkey Theorem'.

Lemma 2.3 *If the matrix entries b_n, for $n \in \mathbb{Z}$, are iid random variables taking the values ± 1 with $\Pr(b_n = 1) \in (0, 1)$, then A^b is pseudo-ergodic with probability one.*

The link to limit operators is provided by the following lemma (see [14, Lemma 6], [26, Corollary 3.70] or [8, Theorem 7.6]):

Lemma 2.4 *For $b \in \{\pm 1\}^{\mathbb{Z}}$, A^b is* pseudo-ergodic *if and only if $\sigma^{\mathsf{op}}(A^b) = \{A^c : c \in \{\pm 1\}^{\mathbb{Z}}\}$.*

Combining this lemma with Theorem 2.1 gives the following characterisation of the spectrum and pseudospectrum of A^b in the case when b is pseudo-ergodic:

Theorem 2.5 *If $b \in \{\pm 1\}^{\mathbb{Z}}$ and A^b is pseudo-ergodic, then*

$$\operatorname{spec} A^b = \operatorname{spec}_{\text{ess}} A^b = \bigcup_{c \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec} A^c = \Sigma := \bigcup_{c \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec}^{\infty}_{\text{point}} A^c \tag{24}$$

and

$$\operatorname{spec}^p_\varepsilon A^b \ = \Sigma^p_\varepsilon := \bigcup_{c \in \{\pm 1\}^{\mathbb{Z}}} \operatorname{spec}^p_\varepsilon A^c, \tag{25}$$

for $\varepsilon > 0$ and $p \in [1, \infty]$.

Limit operator ideas, the "Infinite Monkey" argument and the validity of the first two "=" signs in (24) are not new in the spectral theory of random matrices (see e.g. [4, 13, 14, 19, 33]). Equation (25) is previously shown, for a general class of pseudo-ergodic operators for the case $p = 2$ in [14]. What is more recent is the third "=" sign in the first of equations (24) and the extensions to $p \in [1, \infty]$, these shown in [7] and [8, Theorem 6.28, 7.6].

Note that the above theorem shows that the spectrum of A^b is the same set Σ for every pseudo-ergodic $b \in \{\pm 1\}^{\mathbb{Z}}$, and that $\operatorname{spec} A^c \subset \Sigma$ for every $c \in \{\pm 1\}^{\mathbb{Z}}$, and that similar statements hold

for the pseudospectrum $\mathrm{spec}^p_\varepsilon A^b$. In particular, $\mathrm{spec}\, A^c \subset \Sigma$ if $c \in \Pi_n$, for some $n \in \mathbb{N}$, where $\Pi_n := \{c \in \{\pm 1\}^{\mathbb{Z}} : c \text{ is } n\text{-periodic}\}$. Thus

$$\pi_n \;:=\; \bigcup_{c\in\Pi_n} \mathrm{spec}\, A^c \;=\; \bigcup_{c\in\Pi_n} \mathrm{spec}^\infty_{\mathrm{point}}\, A^c \subset \Sigma, \tag{26}$$

for every $n \in \mathbb{N}$: this is informative as π_n can be computed explicitly by (23) as the union of eigenvalues of $n \times n$ matrices. The following lemma carries out this computation for $n = 1, 2, 3$.

Lemma 2.6 *If $b \in \Pi_1$ with $b_0 = 1$, then $\mathrm{spec}\, A^b = [-2, 2]$ and $\mathrm{spec}\, A^{-b} = \mathrm{i}[-2, 2]$. If $b \in \Pi_2 \setminus \Pi_1$ then $\mathrm{spec}\, A^b = \tau_2 := \{x \pm \mathrm{i}x : -1 \le x \le 1\}$. If $b \in \Pi_3$, $b_0 = b_1 = 1$, and $b_2 = -1$, then*

$$\mathrm{spec}\, A^b = \tau_3 := \mathrm{i}[-1, 1] \cup \{x + \mathrm{i}y : -1/2 \le y \le 1/2,\ x^2 = 1 + 3y^2\}$$

while $\mathrm{spec}\, A^{-b} = \mathrm{i}\tau_3$. Thus

$$\pi_1 = [-2, 2] \cup \mathrm{i}[-2, 2], \quad \pi_2 = \pi_1 \cup \tau_2, \quad \pi_3 = \pi_1 \cup \tau_3 \cup \mathrm{i}\tau_3.$$

Note that $\max_{\lambda\in\pi_1} |\lambda| = 2$ while $\max_{\lambda\in\tau_j} |\lambda| = \sqrt{2}$, for $j = 2, 3$. For $j = 2$ this maximum is achieved at $\pm 1 \pm \mathrm{i}$, while for $j = 3$ this maximum is achieved at $\pm\sqrt{7}/2 \pm \mathrm{i}/2$.

Proof. If $b \in \Pi_1$ with $b_0 = \beta = \pm 1$ then, from (23), $\mathrm{spec}\, A^b = \cup_{|\alpha|=1} \mathrm{spec}\, B^b_{1,\alpha} = \{\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta}\beta : \theta \in \mathbb{R}\}$. So $\mathrm{spec}\, A^b = [-2, 2]$ if $\beta = 1$ and $\mathrm{spec}\, A^b = \mathrm{i}[-2, 2]$ if $\beta = -1$, and $\pi_1 = [-2, 2] \cup \mathrm{i}[-2, 2]$.

If $b \in \Pi_2 \setminus \Pi_1$ then, from (23), where $\beta = b_1 = \pm 1$,

$$\mathrm{spec}\, A^b = \bigcup_{\theta\in\mathbb{R}} \mathrm{spec} \begin{pmatrix} 0 & 1 - \mathrm{e}^{-\mathrm{i}\theta}\beta \\ \beta + \mathrm{e}^{\mathrm{i}\theta} & 0 \end{pmatrix} = \{\lambda \in \mathbb{C} : \lambda^2 = 2\mathrm{i}\sin\theta,\ \theta \in \mathbb{R}\}.$$

Thus $\mathrm{spec}\, A^b = \tau_2$ and $\pi_2 = \pi_1 \cup \tau_2$.

If $b \in \Pi_3$, $b_0 = b_1 = 1$, and $b_2 = -1$, then, from (23),

$$\mathrm{spec}\, A^b = \bigcup_{\theta\in\mathbb{R}} \mathrm{spec} \begin{pmatrix} 0 & 1 & -\mathrm{e}^{-\mathrm{i}\theta} \\ 1 & 0 & 1 \\ \mathrm{e}^{\mathrm{i}\theta} & 1 & 0 \end{pmatrix} = \{\lambda \in \mathbb{C} : \lambda^3 - \lambda = -2\mathrm{i}\sin\theta,\ \theta \in \mathbb{R}\}.$$

Writing $\lambda = x + \mathrm{i}y$, we see that $\lambda^3 - \lambda = -2\mathrm{i}\sin\theta$, for some $\theta \in \mathbb{R}$, iff

$$x(x^2 - 3y^2 - 1) = 0 \quad \text{and} \quad 3x^2y - y^3 - y \in [-2, 2].$$

But this implies that either $x = 0$ and $y^3 + y \in [-2, 2]$, or $x^2 = 3y^2 + 1$ and $8y^3 + 2y \in [-2, 2]$, and it follows that $\mathrm{spec}\, A^b = \tau_3$. That $\mathrm{spec}\, A^{-b} = \mathrm{i}\tau$ can be shown similarly, or follows from Lemma 3.4 below. Since $c \in \Pi_3$ iff $c = \pm V_j b$ for $j = 0, 1$ or 2, it follows that $\pi_3 = \pi_1 \cup \tau_3 \cup \mathrm{i}\tau_3$. ■

In Figure 2 we plot π_n for $n = 5, 10, ..., 30$, with π_n computed numerically in Matlab using the characterisation (23) (see [5] for small plots of π_n for $n = 1, 2, ..., 30$). For each n the set π_n, by the characterisation (23), consists of $k \le n2^n$ analytic arcs, and $\pi_n \subset \Sigma$. The visual impression that might be taken from this sequence of plots is that π_n "fills out" a large part of the square $\Delta := \{x + \mathrm{i}y : x, y \in \mathbb{R},\ |x| + |y| < 2\}$ as $n \to \infty$. But of course $\pi_\infty := \cup_{n\in\mathbb{N}} \pi_n$ is a countable union of analytic arcs, so that π_∞ has (two-dimensional) Lebesgue measure 0. Thus almost every point in Δ is not in π_∞ and so is not one of the points in the plots in Figure 2. Thus these figures provide no evidence that the Lebesgue measure of Σ is any larger than zero. And indeed it was conjectured in [22] that Σ has fractal dimension in the range $(1, 2)$ (and so Lebesgue measure zero). That this is not the case was shown in [5] by an application of Theorem 2.5, specifically by constructing a sequence $b \in \{\pm 1\}^{\mathbb{Z}}$ for which $\mathrm{spec}^\infty_{\mathrm{point}}\, A^b \supset \mathbb{D}$, the open unit disc. Of course, this implies by Theorem 2.5 the following result.

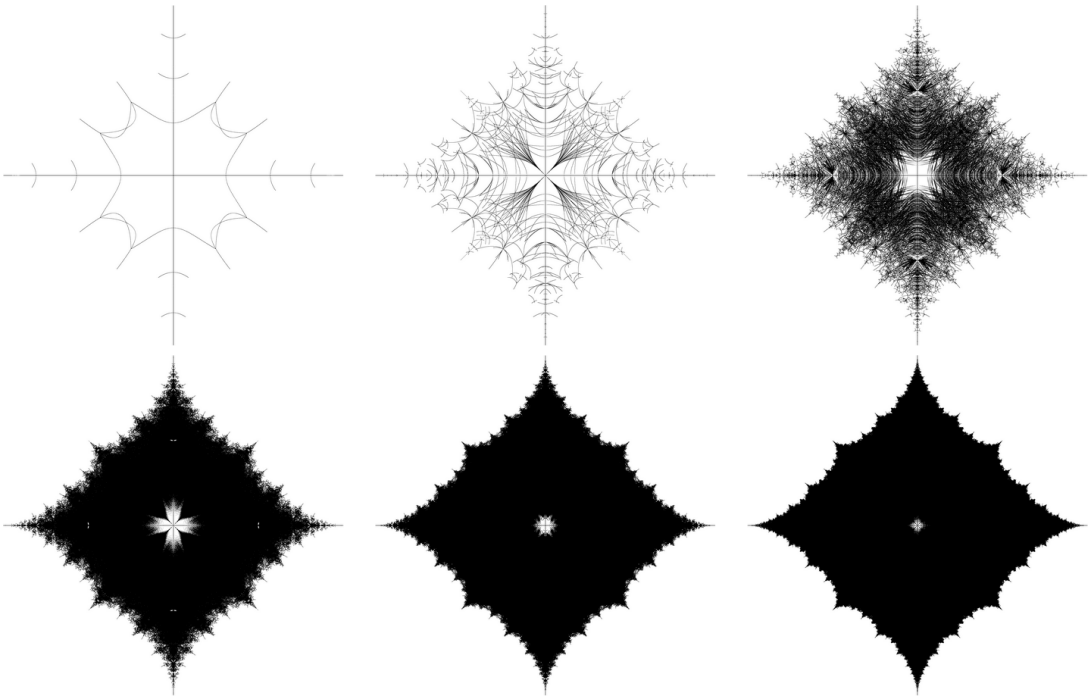


Figure 2: Our figure shows the sets π_n, as defined in (26), for $n = 5, 10, ..., 30$, computed using the characterisation (23), which is made explicit for $n = 1, 2$ and 3 in Lemma 2.6. In particular, $\pi_1 = [-2, 2] \cup \mathrm{i}[-2, 2]$ and, for each n, $\pi_1 \subset \pi_n$ and, by Lemma 3.1, $\pi_n \subset \overline{\Delta} = \{x + \mathrm{i}y : x, y \in \mathbb{R}, |x| + |y| \leq 2\}$.

Theorem 2.7 *[5, Proposition 2.1]* $\overline{\mathbb{D}} \subset \Sigma$.

Recently [9], an alternative proof of this theorem has been obtained, through a construction that shows that π_∞ is dense in $\mathbb{D}$. It is an open (and interesting) question as to whether π_∞ is dense in Σ. An interesting, related, case where the union of the spectra of all periodic operators is shown to be dense in the spectrum of the pseudo-ergodic case is studied in [29], but there are other pseudo-ergodic bi-infinite tridiagonal examples where this is not true.

The above results concern bi-infinite matrices, but similar results apply to the semi-infinite matrices A_+^b and $A_+^{b,c}$. We say that the operator induced by the bi-infinite matrix $B = (b_{ij})_{i,j\in\mathbb{N}}$ is a *limit operator* of the operator induced by the banded semi-infinite matrix $A_+ = (a_{ij})_{i,j\in\mathbb{N}}$ if, for a sequence $h_1, h_2, ...$ of integers with $h_k \to +\infty$, it holds that

$$a_{i+h_k,j+h_k} \;\to\; b_{ij} \qquad \text{as} \qquad k \to \infty,$$

for all $i, j \in \mathbb{Z}$. The set of all limit operators of A_+ is denoted by $\sigma^{\mathsf{op}}(A_+)$. An equivalent characterisation is that $\sigma^{\mathsf{op}}(A_+) = \sigma_+^{\mathsf{op}}(\tilde{A}_+)$, where, for any semi-infinite matrix A_+, $\tilde{A}_+$ is the bi-infinite matrix defined by $\tilde{A}_+ = (\tilde{a}_{ij})_{i,j\in\mathbb{Z}}$, where $\tilde{a}_{ij} := a_{ij}$, $i, j \in \mathbb{N}$, $\tilde{a}_{ij} := 0$, otherwise. The following version of Theorem 2.1 holds in the semi-infinite case. In its results on the pseudospectrum this theorem appears to be new and may be of independent interest. The arguments in this theorem and in later sections depend on the following lemma which, in its results for the pseudospectrum, generalises [40, Theorem 2.4(iii)] from the finite-dimensional Hilbert space case to an infinite-dimensional Banach space setting, and so may also be of independent interest.

Lemma 2.8 *Suppose that X is a Banach space which can be written as the direct sum of two closed subspaces as $X = X_1 \oplus X_2$, by which we mean that each $x \in X$ can be written in a unique way as $x = x_1 + x_2$ with $x_1 \in X_1$ and $x_2 \in X_2$, and that there exists a continuous projection operator $P_1 : X \to X_1$ (in which case $P_2 = I - P_1$ is a projection operator onto X_2). Suppose also that A is a bounded linear operator on X which has X_1 and X_2 as invariant subspaces, and let A_j denote A restricted to X_j, for $j = 1, 2$. Then $\operatorname{spec} A = \operatorname{spec} A_1 \cup \operatorname{spec} A_2$, $\operatorname{spec}_{\rm ess} A = \operatorname{spec}_{\rm ess} A_1 \cup \operatorname{spec}_{\rm ess} A_2$, and $\operatorname{spec}_\varepsilon A_j \subset \operatorname{spec}_\varepsilon A$, for $\varepsilon > 0$, and $j = 1, 2$. If, for some $p \in [1, \infty]$, it holds for every $x_1 \in X_1$ and $x_2 \in X_2$ that $\|x_1 + x_2\| = \|(\|x_1\|, \|x_2\|)\|_p$, then also $\operatorname{spec}_\varepsilon A = \operatorname{spec}_\varepsilon A_1 \cup \operatorname{spec}_\varepsilon A_2$, for $\varepsilon > 0$.*

Proof. The identities $\operatorname{spec} A = \operatorname{spec} A_1 \cup \operatorname{spec} A_2$ and $\operatorname{spec}_{\rm ess} A = \operatorname{spec}_{\rm ess} A_1 \cup \operatorname{spec}_{\rm ess} A_2$ are standard, see e.g. [23, 15]. By Theorem 1.1, $\operatorname{spec}_\varepsilon B = \operatorname{spec} B \cup \{\lambda \in \mathbb{C} : \nu(B - \lambda I) < \varepsilon\}$. Since $\nu(A_j - \lambda I) \geq \nu(A - \lambda I)$, for all $\lambda \in \mathbb{C}$ and $j = 1, 2$, it follows that $\operatorname{spec}_\varepsilon A_j \subset \operatorname{spec}_\varepsilon A$, for $\varepsilon > 0$, and $j = 1, 2$. If, for some $p \in [1, \infty]$, it holds for every $x_1 \in X_1$ and $x_2 \in X_2$ that $\|x_1 + x_2\| = \|(\|x_1\|, \|x_2\|)\|_p$, then, for every $\lambda \in \mathbb{C}$, where $B := A - \lambda I$ and $B_j := A_j - \lambda I$, for $j = 1, 2$, it holds for $x_1 \in X_1$ and $x_2 \in X_2$ that

$$\|B(x_1 + x_2)\| = \big\|\big(\|Bx_1\|, \|Bx_2\|\big)\big\|_p \geq \big\|\big(\nu(B_1)\|x_1\|, \nu(B_2)\|x_2\|\big)\big\|_p$$

so that

$$\nu(B) = \inf_{x_1 \in X_1, x_2 \in X_2} \frac{\|B(x_1 + x_2)\|}{\|x_1 + x_2\|} \geq \inf_{x_1 \in X_1, x_2 \in X_2} \frac{\big\|\big(\nu(B_1)\|x_1\|, \nu(B_2)\|x_2\|\big)\big\|_p}{\big\|\big(\|x_1\|, \|x_2\|\big)\big\|_p}.$$

But it is an easy calculation that this last infimum has the value $\min(\nu(B_1), \nu(B_2))$. Thus $\operatorname{spec}_\varepsilon A \subset \operatorname{spec}_\varepsilon A_1 \cup \operatorname{spec}_\varepsilon A_2$. ■

Theorem 2.9 *Let A_+ be a semi-infinite banded matrix $A_+ = (a_{ij})_{i,j \in \mathbb{N}}$, with $\sup_{ij} |a_{ij}| < \infty$. Then*

$$\operatorname{spec}_{\rm ess} A_+ = \bigcup_{B \in \sigma^{\rm op}(A_+)} \operatorname{spec} B = \bigcup_{B \in \sigma^{\rm op}(A_+)} \operatorname{spec}^\infty_{\rm point} B. \tag{27}$$

Further, $\operatorname{spec}^p_\varepsilon B \subset \operatorname{spec}^p_\varepsilon A_+$, for all $\varepsilon > 0$, $p \in [1, \infty]$, and $B \in \sigma^{\rm op}(A_+)$.

Proof. Given $\varepsilon > 0$ and $p \in [1, \infty]$, choose $\lambda > \|A_+\|_p + 2\varepsilon$ and apply Theorem 2.1 to the bi-infinite matrix $A = \tilde{A}_+ + B$, where $B = (b_{ij})_{i,j \in \mathbb{Z}}$ is defined by $b_{ij} := \lambda$, if $i = j \leq 0$, $b_{ij} := 0$, otherwise. Since $\sigma^{\rm op}(A) = \sigma^{\rm op}_+(A) \cup \sigma^{\rm op}_-(A) = \sigma^{\rm op}(A_+) \cup \{\lambda I\}$, we see, applying Lemma 2.8, that

$$\{\lambda\} \cup \operatorname{spec}_{\rm ess} A_+ = \operatorname{spec}_{\rm ess} A = \{\lambda\} \cup \bigcup_{B \in \sigma^{\rm op}(A_+)} \operatorname{spec} B = \{\lambda\} \cup \bigcup_{B \in \sigma^{\rm op}(A_+)} \operatorname{spec}^\infty_{\rm point} B. \tag{28}$$

Since $\lambda > \|A_+\|_p \geq \|B\|_p$ is not in $\operatorname{spec}_{\rm ess} A_+$ or in $\operatorname{spec} B$, for $B \in \sigma^{\rm op}(A_+)$, equation (27) follows. Similarly, applying Lemma 2.8, $\operatorname{spec}^p_\varepsilon A = \operatorname{spec}^p_\varepsilon(\lambda I) \cup \operatorname{spec}^p_\varepsilon A_+ = (\lambda + \varepsilon\mathbb{D}) \cup \operatorname{spec}^p_\varepsilon A_+$. It follows from Theorem 2.1 that, for $B \in \sigma^{\rm op}(A_+)$, $\operatorname{spec}^p_\varepsilon B \subset \operatorname{spec}^p_\varepsilon A = (\lambda + \varepsilon\mathbb{D}) \cup \operatorname{spec}^p_\varepsilon A_+$. Since $\operatorname{spec}^p_\varepsilon B \subset (\|B\|_p + \varepsilon)\mathbb{D}$ and $\lambda > \|A_+\|_p + 2\varepsilon \geq \|B\|_p + 2\varepsilon$, this implies that $\operatorname{spec}^p_\varepsilon B \subset \operatorname{spec}^p_\varepsilon A_+$. ■

One consequence of this result and Theorem 2.1 is the following lemma.

Lemma 2.10 *For $b, c \in \{\pm 1\}^{\mathbb{Z}}$, $0 \in \operatorname{spec}_{\rm ess} A^{b,c}$ and $0 \in \operatorname{spec}_{\rm ess} A^{b,c}_+$.*

Proof. It is easy to see that $0 \in \operatorname{spec}^\infty_{\rm point} A^{b,c}$ for every $b, c \in \{\pm 1\}^{\mathbb{Z}}$, and the result then follows from equations (21) and (27). ■

We extend the definition of pseudo-ergodic in Definition 2.2 to the semi-infinite case in the obvious way, replacing $\mathbb{Z}$ by $\mathbb{N}$ and A^b by A^b_+, so that $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic iff the sequence b contains every finite pattern of ± 1's. Then Lemma 2.3 holds with $\mathbb{Z}$ replaced by $\mathbb{N}$ and A^b replaced by A^b_+, and Lemma 2.4 holds with A^b replaced by A^b_+.

3 The Numerical Range and Symmetry Arguments

Let us first introduce some properties of and notation related to adjoint operators. Given a banded bi-infinite matrix $A = (a_{ij})_{i,j\in\mathbb{Z}}$, with $\sup_{ij} |a_{ij}| < \infty$, A^* will denote the matrix $A^* = (\bar{a}_{ji})_{i,j\in\mathbb{Z}}$. For $1 \le p < \infty$, where $q \in (1,\infty]$ satisfies $p^{-1} + q^{-1} = 1$, and identifying $\ell^q(\mathbb{Z})$ with $(\ell^p(\mathbb{Z}))^*$, the dual space of $\ell^p(\mathbb{Z})$ (in the sense e.g. of Kato [24], where the elements of the dual space are anti-linear functionals), it holds that $A^* : \ell^q(\mathbb{Z}) \to \ell^q(\mathbb{Z})$ is the adjoint of $A : \ell^p(\mathbb{Z}) \to \ell^p(Z)$. Further [24] A is invertible iff A^* is invertible and, if they are both invertible, then $\|A^{-1}\|_p = \|(A^*)^{-1}\|_q$. Similarly, A is Fredholm iff A^* is Fredholm and, if they are both Fredholm then $\operatorname{ind} A = -\operatorname{ind} A^*$.

In this section we first compute the numerical range of the operator A^b in the case when b is pseudo-ergodic, which gives an upper bound on the spectrum Σ of A^b. We then apply a variety of symmetry arguments to explore the relationship between spectral sets for matrices with one and two ±1 diagonals and between semi-infinite and bi-infinite matrices, and to explore the geometry of Σ and that of Σ_ε^p, the ε-pseudospectrum of A^b on $\ell^p(\mathbb{Z})$ when b is pseudo-ergodic. Our final result shows that, roughly speaking, in the pseudo-ergodic case, the spectral sets are the same whether the matrix is semi-infinite or bi-infinite, and whether the matrix has one or two ±1 diagonals.

These results are to some extent surprising: there is no expectation in general that the spectral sets associated with bi-infinite and corresponding semi-infinite matrices will be the same. A simple example is provided by the shift operator V_{-1}. This is a Laurent operator (a bi-infinite Toeplitz matrix) whose spectrum is the unit circle and whose ℓ^2 ε-pseudospectrum is the ε-neighbourhood of the unit circle. On the other hand the Toeplitz operator that is the shift operator restricted to $\ell^2(\mathbb{N})$ (a semi-infinite Toeplitz matrix) has spectrum that is the closed unit disc (e.g. [15]). An example closer to our case is studied in [39], where calculations are made of the spectra of random bi-diagonal bi-infinite and semi-infinite matrices, matrices which the authors term stochastic Laurent and Toeplitz operators, respectively, by which they mean a bi-infinite or semi-infinite matrix where each diagonal is either constant or has random entries, but with the random distribution constant along the diagonal. In the bi-diagonal case they study, which has the constant value 1 along the first superdiagonal and a random main diagonal, it is found [39] that the bi-infinite and semi-infinite matrices may or may not have the same spectra, this depending on the support of the probability density function for the random variables on the main diagonal.

Our first result is a computation of the numerical range. By $W(B)$ we denote the (2-norm) *numerical range* of the operator or matrix B, defined by (see Section 1.2) $W(B) := \{(Bx, x) : \|x\|_2 = 1\}$, where $(\cdot,\cdot)$ denotes the standard ℓ^2 inner-product on $\mathbb{C}^n$ or on $\ell^2(S)$, with $S = \mathbb{Z}$ or $\mathbb{N}$, as appropriate.

Lemma 3.1 *For $b \in \{\pm1\}^{\mathbb{Z}}$, $W(A_n^b) \subset W(A_+^b) \subset W(A^b) \subset \Delta := \{x + \mathrm{i}y : x, y \in \mathbb{R}, |x| + |y| < 2\}$, and $W(A^b) = \Delta$ if b is pseudo-ergodic. Similarly, $W(A_+^b) = \Delta$ if $b \in \{\pm1\}^{\mathbb{N}}$ is pseudo-ergodic, and $\Sigma \subset \overline{\Delta}$.*

Proof. For $b \in \{\pm1\}^{\mathbb{Z}}$ and $x \in \ell^2(\mathbb{Z})$ with $\|x\|_2 = 1$, defining $a = \bar{x}V_{-1}x$ we see that

$$(A^b x, x) = \sum_{k\in\mathbb{Z}} (b_{k-1}x_{k-1} + x_{k+1})\bar{x}_k = \sum_{k\in\mathbb{Z}} (b_k\bar{a}_k + a_k) = \sum_{k\in\mathbb{Z}} [\alpha_k(1 + b_k) + \mathrm{i}\beta_k(1 - b_k)],$$

where $\alpha_k = \Re(a_k)$ and $\beta_k = \Im(a_k)$. Thus

$$|\Re(A^b x, x)| + |\Im(A^b x, x)| \le \sum_{k\in\mathbb{Z}} \{|\alpha_k|(1 + b_k) + |\beta_k|(1 - b_k)\} \le 2\sum_{k\in\mathbb{Z}} |a_k| = 2\|a\|_1.$$

Now, since $x \in \ell^2(\mathbb{Z})$, $\bar{x}$ and $V_{-1}x$ must be linearly independent. Hence, by the Cauchy-Schwarz inequality, $\|a\|_1 = \|\bar{x}V_{-1}x\|_1 < \|x\|_2\|V_{-1}x\|_2 = 1$. We have shown that $W(A^b) \subset \Delta$; it follows

that $W(A_n^b) \subset W(A_+^b) \subset W(A^b)$ from (19). From this it follows, from standard properties of the numerical range (see the end of Section 1.2) that $\Sigma \subset \overline{W(A^b)} \subset \overline{\Delta}$. But, since $\pi_1 = [-2,2] \cup \mathrm{i}[-2,2] \subset \Sigma$, this implies that ± 2 and $\pm 2\mathrm{i}$ are in $\overline{W(A^b)}$, if b is pseudo-ergodic. Hence, if b is pseudo-ergodic, then, for every $\eta > 0$ there exist points $r, s, t, u \in W(A^b)$ with $|2 - r| < \eta$, $|-2 - s| < \eta$, $|2\mathrm{i} - t| < \eta$ and $|-2\mathrm{i} - u| < \eta$. Since $W(A^b)$ is convex, this implies that $\Delta \subset W(A^b)$, and so $W(A^b) = \Delta$. A similar argument, using that $\Sigma = \operatorname{spec} A_+^b \subset \overline{W(A_+^b)}$ if $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic (that $\Sigma = \operatorname{spec} A_+^b$ is established in Theorem 3.6 below), shows that $W(A_+^b) = \Delta$ if $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic. ■

Our next result elucidates the relationship between the spectral properties of matrices with one and two ± 1 diagonals. One obvious symmetry result we use already in this lemma is that, since the coefficients $b, c \in \{\pm 1\}^{\mathbb{Z}}$ are real-valued, the spectrum and pseudospectrum of $A^{b,c}$ are symmetric about the real axis.

Lemma 3.2 *For $a, b, c \in \{\pm 1\}^{\mathbb{Z}}$,*

$$M_a A^{b,c} M_a^{-1} = A^{bd,cd},$$

where $d = aV_{-1}a$, so that

$$\operatorname{spec} A^{b,c} = \operatorname{spec} A^{bd,cd} = \operatorname{spec} A^{bc}, \quad \operatorname{spec}_{\mathrm{ess}} A^{b,c} = \operatorname{spec}_{\mathrm{ess}} A^{bd,cd} = \operatorname{spec}_{\mathrm{ess}} A^{bc}.$$

Further, for $\lambda \notin \operatorname{spec}_{\mathrm{ess}} A^{b,c}$, $\operatorname{ind}(A^{b,c} - \lambda I) = 0$, and, for $\lambda \notin \operatorname{spec} A^{b,c}$ and $p \in [1, \infty]$, where $q \in [1, \infty]$ is given by $p^{-1} + q^{-1} = 1$,

$$\|(A^{b,c} - \lambda I)^{-1}\|_p = \|(A^{bd,cd} - \lambda I)^{-1}\|_p = \|(A^{bc} - \lambda I)^{-1}\|_p = \|(A^{b,c} - \lambda I)^{-1}\|_q,$$

so that, for $\varepsilon > 0$,

$$\operatorname{spec}_\varepsilon^p A^{b,c} = \operatorname{spec}_\varepsilon^p A^{bd,cd} = \operatorname{spec}_\varepsilon^p A^{bc} = \operatorname{spec}_\varepsilon^q A^{b,c}.$$

Moreover, for $1 \le p \le r \le 2$ and $\varepsilon > 0$,

$$\operatorname{spec}_\varepsilon^r A^{b,c} \subset \operatorname{spec}_\varepsilon^p A^{b,c}.$$

Proof. For $a, b, c \in \{\pm 1\}^{\mathbb{Z}}$, recalling (20) and noting that $M_a^{-1} = M_a$,

$$M_a A^{b,c} M_a^{-1} = M_a V_1 M_{ab} + M_{ac} V_{-1} M_a = V_1 M_{V_{-1}a} M_{ab} + M_{ac} M_{V_{-1}a} V_{-1} = V_1 M_{bd} + M_{cd} V_{-1} = A^{bd,cd}.$$

In particular, choosing a so that $d = c$, this identity reduces to $M_a A^{b,c} M_a^{-1} = A^{bc}$, while, choosing a so that $d = bc$, this identity reduces to $M_a A^{b,c} M_a^{-1} = A^{c,b} = (A^{b,c})^*$. The remaining results, except the last equation, follow since M_a is an isometric isomorphism, and using the properties of the adjoint listed immediately at the beginning of the section, and standard properties of Fredholm operators, e.g. [24, 23]. The last inclusion follows from the interpolation theorem of Riesz-Thorin, often called the Riesz convexity theorem [38, Chapter V, Theorem 1.3], which implies that, for $\lambda \notin \operatorname{spec} A^{b,c}$,

$$\|(A^{b,c} - \lambda I)^{-1}\|_r \le \max(\|(A^{b,c} - \lambda I)^{-1}\|_p, \|(A^{b,c} - \lambda I)^{-1}\|_q) = \|(A^{b,c} - \lambda I)^{-1}\|_p.$$

■

Note that this lemma implies that, for $1 \le p \le 2 \le q \le \infty$, where $p^{-1} + q^{-1} = 1$,

$$\operatorname{spec}_\varepsilon^2 A^{b,c} \subset \operatorname{spec}_\varepsilon^p A^{b,c} = \operatorname{spec}_\varepsilon^q A^{b,c} \subset \operatorname{spec}_\varepsilon^1 A^{b,c} = \operatorname{spec}_\varepsilon^\infty A^{b,c}.$$

In general, for a non-self-adjoint operator or matrix A, it need not hold that $\operatorname{spec}_\varepsilon^r A \subset \operatorname{spec}_\varepsilon^p A$ for any distinct $p, r \in [1, \infty]$.

Exactly the same results hold in the semi-infinite case. Precisely, where M_a^+ denotes the operator on $\ell^p(\mathbb{N})$ of multiplication by $a \in \{\pm 1\}^{\mathbb{N}}$, Lemma 3.2 holds also with $\mathbb{Z}$ replaced by $\mathbb{N}$, M_a replaced by M_a^+, and all other operators replaced by their semi-infinite counterparts. Similarly, where D_n^a is the diagonal matrix with the vector $a = (a_1, ..., a_n)$ on the diagonal, the following finite dimensional version of the above lemma holds.

Lemma 3.3 *For $n \in \mathbb{N}$, $a \in \{\pm 1\}^n$, and $b, c \in \{\pm 1\}^{n-1}$,*

$$D_n^a A_n^{b,c} D_n^a = A_n^{bd,cd},$$

where $d = (a_1 a_2, ..., a_{n-1} a_n)$, so that

$$\operatorname{spec} A_n^{b,c} = \operatorname{spec} A_n^{bd,cd} = \operatorname{spec} A_n^{bc}.$$

Further, for $p \in [1, \infty]$ and $\varepsilon > 0$, where $q \in [1, \infty]$ is given by $p^{-1} + q^{-1} = 1$,

$$\operatorname{spec}_\varepsilon^p A_n^{b,c} = \operatorname{spec}_\varepsilon^p A_n^{bd,cd} = \operatorname{spec}_\varepsilon^p A_n^{bc} = \operatorname{spec}_\varepsilon^q A_n^{b,c}.$$

Moreover, for $1 \le p \le r \le 2$ and $\varepsilon > 0$, $\operatorname{spec}_\varepsilon^r A_n^{b,c} \subset \operatorname{spec}_\varepsilon^p A_n^{b,c}$.

A first application of the above lemmas is the following symmetry result (cf. [22]).

Lemma 3.4 *For $b \in \{\pm 1\}^{\mathbb{Z}}$, $\varepsilon > 0$, and $p \in [1, \infty]$, $\operatorname{spec} A^b$, $\operatorname{spec}_{\text{ess}} A^b$, $\operatorname{spec}_\varepsilon^p A^b$, $\operatorname{spec} A_n^b$, and $\operatorname{spec}_\varepsilon^p A_n^b$ are invariant under reflection in the real and imaginary axes. Further, where $S(b)$ denotes any one of these sets, it holds that $S(-b) = \mathrm{i} S(b)$. The set Σ, which is the set $\operatorname{spec} A^b = \operatorname{spec}_{\text{ess}} A^b$ in the case that b is pseudo-ergodic, and, for $\varepsilon > 0$ and $p \in [1, \infty]$, the set Σ_ε^p, which is the set $\operatorname{spec}_\varepsilon^p A^b$ for b pseudo-ergodic, are invariant under reflection in either axis and under rotation by 90^0.*

Proof. We prove the results for A^b using Lemma 3.2; the proof for A_n^b using Lemma 3.3 is similar. That the entries of the matrix A^b are real implies the symmetry about the real axis. Defining $a \in \{\pm 1\}^{\mathbb{Z}}$ by $a_k = (-1)^k$, $k \in \mathbb{Z}$, so that $d = aV_{-1}a$ is the constant sequence $d = (..., -1, -1, ...)$, it follows from Lemma 3.2 that $M_a A^b M_a^{-1} = -A^b$, which implies that the sets $\operatorname{spec} A^b$, $\operatorname{spec}_{\text{ess}} A^b$, and $\operatorname{spec}_\varepsilon^p A^b$ are also invariant under reflection in the origin, so that they are also invariant under reflection in the imaginary axis. Defining, instead, $a \in \ell^\infty(\mathbb{Z})$ by $a_k = \mathrm{i}^k$, we obtain, similarly, that $M_a A^b M_a^{-1} = A^{db,\bar{d}}$, where $d = \bar{a} V_{-1} a$ so that $d_k = \mathrm{i}$. Thus $M_a A^b M_a^{-1} = -\mathrm{i} A^{-b}$, and we see that $S(-b) = \mathrm{i} S(b)$, where $S(b)$ denotes one of $\operatorname{spec} A^b$, $\operatorname{spec}_{\text{ess}} A^b$, or $\operatorname{spec}_\varepsilon^p A^b$. Where $S(b)$ again denotes one of these sets, since b is pseudo-ergodic iff $-b$ is pseudo-ergodic, that $S(b) = S(-b) = \mathrm{i} S(b)$ follows from Theorem 2.5. ■

The following lemma further elucidates the relationship between the spectral properties of semi-infinite and bi-infinite matrices. In this lemma for $p \in [1, \infty]$ we let $\ell_o^p(\mathbb{Z})$ denote the closed subspace of odd elements of $\ell^p(\mathbb{Z})$, i.e. $x \in \ell_o^p(\mathbb{Z})$ iff $x_{-k} = -x_k$, $k \in \mathbb{Z}$, and let $\ell_e^p(\mathbb{Z})$ denote the closed subspace of even elements of $\ell^p(\mathbb{Z})$, i.e. $x \in \ell_e^p(\mathbb{Z})$ iff $x_{-k} = x_k$, $k \in \mathbb{Z}$, so that $\ell^p(\mathbb{Z}) = \ell_o^p(\mathbb{Z}) \oplus \ell_e^p(\mathbb{Z})$. It is convenient to equip $\ell_o^p(\mathbb{Z})$ with the norm $\|x\| := 2^{-1/p} \|x\|_p$, so that the extension operator $E : \ell^p(\mathbb{N}) \to \ell_o^p(\mathbb{Z})$ given by $(Ex)_k = x_k$, $k \in \mathbb{N}$, $(Ex)_0 = 0$, and $(Ex)_{-k} = -x_k$, $k \in \mathbb{N}$, is an isometric isomorphism, as is the restriction operator $P : \ell_o^p(\mathbb{Z}) \to \ell^p(\mathbb{N})$ given by $(Px)_k = x_k$, $k \in \mathbb{N}$. (This change of norm does not effect the value of the induced norm of a bounded linear operator A on $\ell_o^2(\mathbb{Z})$, and so does not affect the definition of $\operatorname{spec}_\varepsilon^p A$.) Further, let $R : \ell^p(\mathbb{Z}) \to \ell^p(\mathbb{Z})$ be the reflection operator given by $(Rx)_k = x_{-1-k}$, $k \in \mathbb{Z}$.

Lemma 3.5 *Suppose $b \in \{\pm 1\}^{\mathbb{Z}}$ with $b_k = 1$, $k \le 0$, and let $c = Rb$. Then, for $p \in [1, \infty]$, $A^{b,c}$ maps $\ell_o^p(\mathbb{Z})$ to $\ell_o^p(\mathbb{Z})$ and maps $\ell_e^p(\mathbb{Z})$ to $\ell_e^p(\mathbb{Z})$. Further, where $A_o^{b,c}$ denotes the restriction of $A^{b,c}$ to $\ell_o^p(\mathbb{Z})$,*

$$A_o^{b,c} = E A_+^b P. \tag{29}$$

Thus $\operatorname{spec} A_+^b = \operatorname{spec} A_o^{b,c} \subset \operatorname{spec} A^{b,c} = \operatorname{spec} A^{bc}$ and $\operatorname{spec}_\varepsilon^p A_+ = \operatorname{spec}_\varepsilon^p A_o^{b,c} \subset \operatorname{spec}_\varepsilon^p A^{b,c} = \operatorname{spec}_\varepsilon^p A^{bc}$, for $\varepsilon > 0$ and $p \in [1, \infty]$.

Proof. For $x \in \ell_o^2(\mathbb{Z})$, $(A^{b,c}x)_0 = b_{-1}x_{-1} + c_0 x_1 = x_{-1} + x_1 = 0$ and, for $k \in \mathbb{N}$, $(A^{b,c}x)_{-k} = b_{-k-1}x_{-k-1} + c_{-k}x_{-k+1} = -c_k x_{k+1} - b_{k-1}x_{k-1} = -(A^{b,c}x)_k$, so that $A^{b,c} : \ell_o^p(\mathbb{Z}) \to \ell_o^p(\mathbb{Z})$.

Similarly, for $x \in \ell^2_e(\mathbb{Z})$ and $k \in \mathbb{N}$, $(A^{b,c}x)_{-k} = b_{-k-1}x_{-k-1} + c_{-k}x_{-k+1} = c_k x_{k+1} + b_{k-1}x_{k-1} = (A^{b,c}x)_k$, so that $A^{b,c} : \ell^p_e(\mathbb{Z}) \to \ell^p_e(\mathbb{Z})$. Further, for $k \in \mathbb{N}$, $(EA^b_+ x)_k = b_{k-1}x_{k-1} + x_{k+1} = b_{k-1}x_{k-1} + c_k x_{k+1} = (A^{b,c}_o x)_k$, so that (29) holds. Since E and P are isometric isomorphisms and $E = P^{-1}$, it follows that $\operatorname{spec} A^b_+ = \operatorname{spec} A^{b,c}_o$ and that $\operatorname{spec}^p_\varepsilon A_+ = \operatorname{spec}^p_\varepsilon A^{b,c}_o$, for $\varepsilon > 0$ and $p \in [1,\infty]$. The remaining results follow from Lemma 2.8 and Lemma 3.2. ■

Putting the results from the previous section and this section together gives the following characterisations of the spectrum, essential spectrum, and pseudospectrum in the pseudo-ergodic case.

Theorem 3.6 *If $b, c, d \in \{\pm 1\}^{\mathbb{N}}$, $e, f, g \in \{\pm 1\}^{\mathbb{Z}}$, and b, cd, e, and fg are pseudo-ergodic, then $\operatorname{spec} A^b_+ = \operatorname{spec} A^{c,d}_+ = \operatorname{spec} A^e = \operatorname{spec} A^{f,g} = \operatorname{spec}_{\text{ess}} A^b_+ = \operatorname{spec}_{\text{ess}} A^{c,d}_+ = \operatorname{spec}_{\text{ess}} A^e = \operatorname{spec}_{\text{ess}} A^{f,g} = \Sigma$ and, for $\varepsilon > 0$ and $p \in [1,\infty]$, where $q \in [1,\infty]$ is given by $p^{-1} + q^{-1} = 1$, $\operatorname{spec}^p_\varepsilon A^b_+ = \operatorname{spec}^p_\varepsilon A^{c,d}_+ = \operatorname{spec}^p_\varepsilon A^e = \operatorname{spec}^p_\varepsilon A^{f,g} = \Sigma^p_\varepsilon = \Sigma^q_\varepsilon$. Further, for $1 \le p \le r \le 2$ and $\varepsilon > 0$, $\Sigma^r_\varepsilon \subset \Sigma^p_\varepsilon$.*

Proof. From Lemma 3.2 and the remarks following that lemma we have that $\operatorname{spec} A^{c,d}_+ = \operatorname{spec} A^{cd}_+$, $\operatorname{spec}_{\text{ess}} A^{c,d}_+ = \operatorname{spec}_{\text{ess}} A^{cd}_+$, $\operatorname{spec} A^{f,g} = \operatorname{spec} A^{fg}$, and $\operatorname{spec}_{\text{ess}} A^{f,g} = \operatorname{spec}_{\text{ess}} A^{fg}$. From Theorems 2.5, 2.9, and the remarks at the end of Section 2, we have moreover that if b and e are pseudo-ergodic then $\operatorname{spec} A^e = \operatorname{spec}_{\text{ess}} A^e = \operatorname{spec}_{\text{ess}} A^b_+ = \Sigma$. This implies that $\Sigma \subset \operatorname{spec} A^b_+$, and that $\operatorname{spec} A^b_+ \subset \Sigma$ follows from Lemma 3.5 which, together with Lemma 3.2, gives that $\operatorname{spec} A^b_+ \subset \operatorname{spec} A^{b,Rb} = \operatorname{spec} A^{bRb} \subset \Sigma$. The results for the pseudospectrum are shown similarly, again using Theorems 2.5, 2.9, the remarks at the end of Section 2, and Lemma 3.2. ■

4 The relationship between the spectra and pseudospectra of finite and infinite matrices

An obvious method to try to calculate the spectrum of an infinite matrix is to study the spectra of large finite submatrices of the infinite matrix and hope that these provide good approximations. In particular, one can apply this idea to the infinite matrix A^b_+, and hope that the spectrum of the $n \times n$ matrix A^b_n, the intersection of the first n rows and columns of A^b_+, will approximate the spectrum of A^b_+ well for n large.

In general the spectrum and pseudospectrum of an infinite banded matrix may or may not be well-approximated by the spectra and pseudospectra of its finite submatrices (see [30] and the references therein for some discussion, with emphasis on the case of tridiagonal pseudo-ergodic matrices). In particular, there need be no relationship at all between the spectrum of a bi-infinite matrix and the spectra of its finite sections. A simple example is provided by the Laurent operator that is the shift operator V_{-1} with matrix representation $(a_{ij})_{ij\in\mathbb{Z}}$, with $a_{ij} = \delta_{i,j+1}$, whose spectrum is the unit circle. The Toeplitz matrices that are its $n \times n$ finite sections, $(a_{ij})_{1\le i,j\le n}$, clearly have zero as the only eigenvalue.

The purpose of this section is to show that, for the particular class of pseudo-ergodic operators we are studying, there is a perhaps surprisingly close (given that our pseudo-ergodic operators are not self-adjoint or normal) connection between the spectral sets in the finite and infinite case. This connection is particularly close for the pseudospectra.

4.1 That the finite matrix spectral sets are contained in the infinite matrix counterparts

For $n \in \mathbb{N}$, introduce the $n \times n$ matrices

$$I_n \;=\; \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \qquad \text{and} \qquad J_n \;=\; \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix},$$

so that I_n is the order n identity matrix. The proof of the following result uses a similar construction to that of the bi-infinite matrix $A^{b,c}$ in the proof of Lemma 3.5.

Theorem 4.1 *If b is pseudo-ergodic then, for $n \in \mathbb{N}$,*

$$\operatorname{spec} A_n^b \subset \sigma_n := \bigcup_{f \in \{\pm 1\}^{n-1}} \operatorname{spec} A_n^f \;\subset \pi_{2n+2} \subset \operatorname{spec} A^b = \Sigma.$$

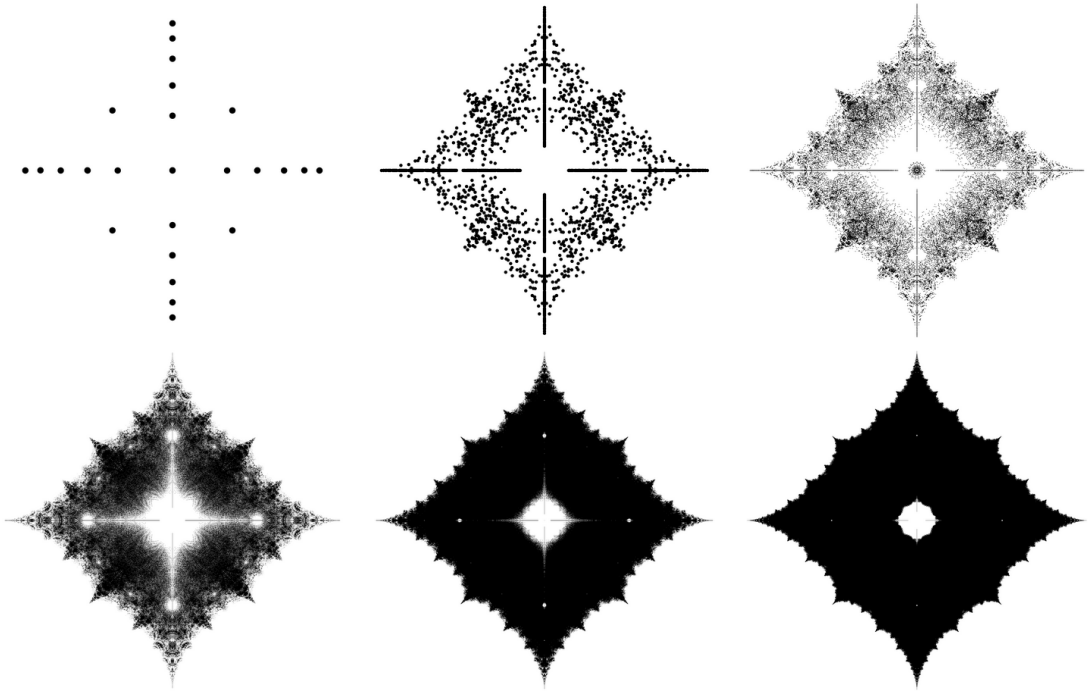


Figure 3: Our figure shows the sets σ_n of all $n \times n$ matrix eigenvalues, as defined in Theorem 4.1, for $n = 5, 10, ..., 30$. Note that in the first pictures (with only a few eigenvalues), we have used heavier pixels for the sake of visibility.

Proof. If $\lambda \in \operatorname{spec} A_n^f$, for some $f \in \{\pm 1\}^{n-1}$, then $A_n^f x = \lambda x$ for some non-zero $x \in \mathbb{C}^n$. Put $\widehat{x} := J_n x$ and $\widehat{A_n^f} := J_n A_n^f J_n$. Then

$$\widehat{A_n^f}\widehat{x} \;=\; J_n A_n^f J_n J_n x \;=\; J_n A_n^f x \;=\; J_n \lambda x = \lambda \widehat{x}$$

and hence, using repeated reflections, i.e. by putting

$$A^{c,d} := \begin{pmatrix} \ddots & 1 & & & & & & \\ 1 & \boxed{\widehat{A_n^f}} & & & & & & \\ & -1 & 1 & & & & & \\ & & 0 & 1 & & & & \\ & & 1 & \boxed{A_n^f} & & & & \\ & & & -1 & 1 & & & \\ & & & & 0 & 1 & & \\ & & & & 1 & \boxed{\widehat{A_n^f}} & & \\ & & & & & -1 & 1 & \\ & & & & & & \ddots & \end{pmatrix} \quad \text{and} \quad \widetilde{x} := \begin{pmatrix} \vdots \\ \boxed{\widehat{x}} \\ 0 \\ \boxed{x} \\ 0 \\ \boxed{\widehat{x}} \\ \vdots \end{pmatrix},$$

we get $A^{c,d}\widetilde{x} = \lambda\widetilde{x}$ with $\widetilde{x} \in \ell^\infty(\mathbb{Z})$, so that λ is an eigenvalue of $A^{c,d}$ as an operator on $\ell^\infty(\mathbb{Z})$. Thus, applying Lemma 3.2 and Theorem 2.5, and noting that both c and d are periodic, with period $2n+2$, we see that $\lambda \in \operatorname{spec} A^{c,d} = \operatorname{spec} A^{cd} \subset \pi_{2n+2} \subset \Sigma = \operatorname{spec} A^b$. ■

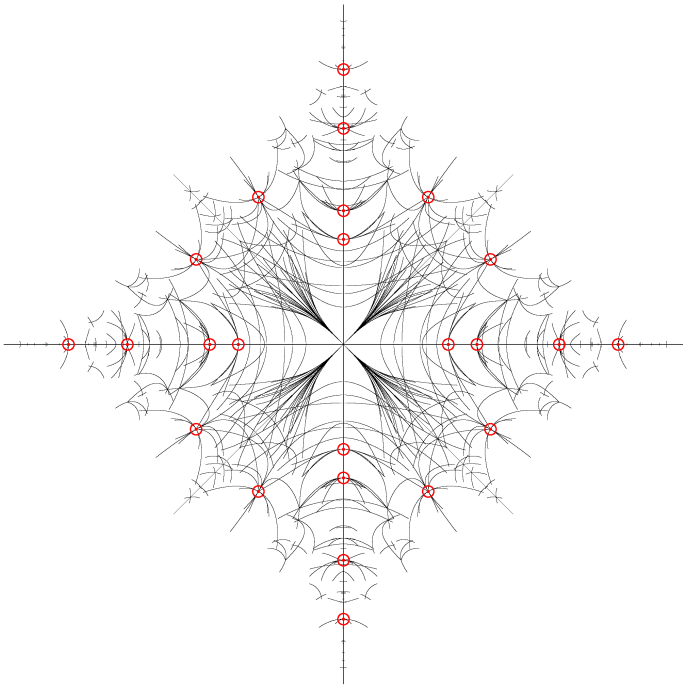


Figure 4: An illustration of the inclusion $\sigma_4 \subset \pi_{10}$, which holds by Theorem 4.1. (The points in σ_4 are indicated by circled dots.) For similar figures for other values of n see [27].

In Figure 3 we plot the sets σ_n, for $n = 5, 10, ..., 30$ (note that each set σ_n is invariant under reflection in either axis or under rotation by 90^0, by Lemma 3.4, and see [5] for smaller plots of these sets for $n = 1, ..., 30$). By the above theorem, $\sigma_n \subset \pi_{2n+2}$ for each n, so that

$$\sigma_\infty := \bigcup_{n\in\mathbb{N}} \sigma_n \subset \pi_\infty.$$

The inclusion $\sigma_n \subset \pi_{2n+2}$ is illustrated for $n = 4$ in Figure 4.

An interesting question, alluded to already in Section 2, is whether π_∞, which is contained in Σ, or σ_∞, which is a countable subset of π_∞, are dense in Σ, the spectrum of A^b for b pseudo-ergodic. Of course, we do not know what Σ is, so that this question is difficult to resolve! We do know however (Theorem 2.7) that the unit disc $\mathbb{D} \subset \Sigma$, and we can consider the question as to whether π_∞ or σ_∞ are dense in $\mathbb{D}$. Recall that the sets π_n, for $n = 5, 10, ..., 30$, are plotted already in Figure 2. Studying Figures 2 and 3, it appears that there is a "hole" in both σ_n and π_n around the origin, though these holes appear to be reducing in size as n increases. And in fact, as mentioned already in Section 2, it has been shown recently that π_∞ is dense in $\mathbb{D}$. Further, it appears to us plausible, comparing the two figures, to conjecture that σ_∞ is dense in π_∞ and so dense in $\mathbb{D}$.

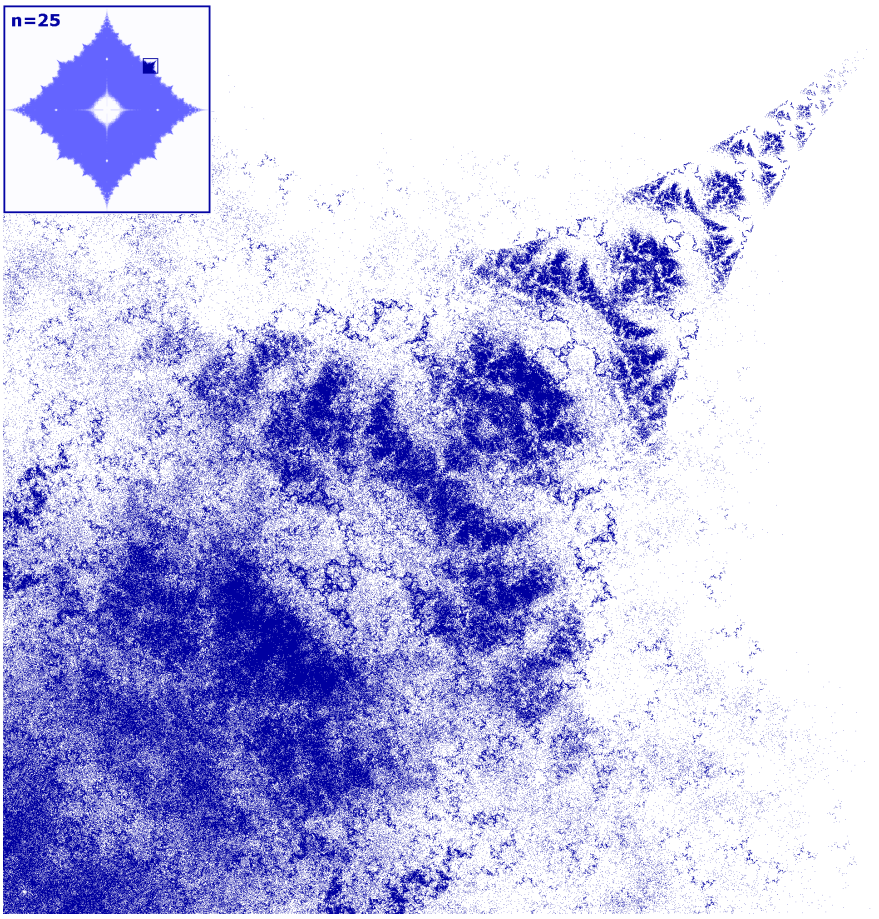


Figure 5: This is a zoom into σ_{25} – the 5th picture of Figure 3. The location of this zoom is near the point $1 + i$, which is the midpoint of the northeast edge of the square $W(A^b) = \Delta$. The picture clearly suggests self-similar features of the set σ_{25}.

Figure 5, taken from [5], zooms into the part of the set σ_{25} around $1 + \mathrm{i}$. Intriguingly this set, the collection of all eigenvalues of a set of 2^{24} matrices of size 25×25 ($25 \times 2^{24} = 419,430,400$ eigenvalues in all!), appears to have a self-similar structure. We have no explanation for these beautiful geometrical patterns, and it is not clear to us how to gain insight into the geometry of this set.

In the next theorem and corollary we show the analogue of Theorem 4.1 for pseudospectra.

Theorem 4.2 *If b is pseudo-ergodic and $n \in \mathbb{N}$ then, for all $\lambda \in \mathbb{C} \setminus \Sigma$, $f \in \{\pm 1\}^{n-1}$, and $p \in [1, \infty]$,*

$$\|(A_n^f - \lambda I_n)^{-1}\|_p \;\leq\; \|(A^b - \lambda I)^{-1}\|_p.$$

Proof. Let $\lambda \in \mathbb{C} \setminus \Sigma$ and $f \in \{\pm 1\}^{n-1}$, so that $A_n^f - \lambda I_n$ is invertible by Theorem 4.1, and let $p \in [1, \infty]$. Put $M := \|(A_n^f - \lambda I_n)^{-1}\|_p$. For every $\delta > 0$, there exists an $x = (x_1, ..., x_n)^\top \in \mathbb{C}^n$ such that $\|x\|_p = 1$ and $y := (A_n^f - \lambda I_n)x$ has $\|y\|_p < \frac{1}{M-\delta}$. Now let $c, d \in \{\pm 1\}^{\mathbb{Z}}$ be the sequences in the matrix $A^{c,d}$ introduced in the proof of Theorem 4.1. At this point our current proof has to bifurcate depending on the value of p.

Case 1: $p = \infty$
Define $\widetilde{x} \in \ell^\infty(\mathbb{Z})$ exactly as in the proof of Theorem 4.1. Then $\widetilde{y} := (A^{c,d} - \lambda I)\widetilde{x}$ is of the form $\widetilde{y} = (\cdots, y^\top, 0, (J_n y)^\top, 0, y^\top, 0, (J_n y)^\top, \cdots)^\top \in \ell^\infty(\mathbb{Z})$ and $\|\widetilde{y}\|_\infty = \|y\|_\infty$, as well as $\|\widetilde{x}\|_\infty = \|x\|_\infty$, so that

$$\left\|(A^{c,d} - \lambda I)^{-1}\right\|_\infty \geq \frac{\|\widetilde{x}\|_\infty}{\|\widetilde{y}\|_\infty} = \frac{\|x\|_\infty}{\|y\|_\infty} > M - \delta.$$

Case 2: $p < \infty$
For any $m \in \mathbb{N}$, let $\widetilde{x}^{(m)}$ be the sequence $\widetilde{x}$ from case 1, but with all entries of index outside $\{-m(n+1), \ldots, m(n+1)\}$ put to zero (where we suppose that the sequence $\widetilde{x}$ is numbered so that at index zero there is one of the 0 entries between x and $\widehat{x}$ of $\widetilde{x}$, so that $\widetilde{x}_0^{(m)} = 0$). Then $\widetilde{y}^{(m)} := (A^{c,d} - \lambda I)\widetilde{x}^{(m)}$ is the same as $\widetilde{y}$ from Case 1 for entries with index between $-m(n+1)+1$ and $m(n+1)-1$, is zero outside $\{-m(n+1)-1, \ldots, m(n+1)+1\}$ and we have $\widetilde{y}^{(m)}_{-m(n+1)} = x_1$ if m is even and $\widetilde{y}^{(m)}_{-m(n+1)} = x_n$ if m is odd, while $\widetilde{y}^{(m)}_{m(n+1)} = -x_1$ if m is even and $\widetilde{y}^{(m)}_{m(n+1)} = -x_n$ if m is odd. As a result, we find that

$$\|\widetilde{x}^{(m)}\|_p^p = 2m\,\|x\|_p^p \qquad \text{and} \qquad \|\widetilde{y}^{(m)}\|_p^p = 2m\,\|y\|_p^p + \begin{cases} 2\,|x_1|^p & \text{if } m \text{ is even,} \\ 2\,|x_n|^p & \text{if } m \text{ is odd.} \end{cases}$$

From $\|x\|_p = 1$, $\|y\|_p < \frac{1}{M-\delta}$ and $|x_1|, |x_n| \leq \|x\|_p = 1$ we hence get that $\|\widetilde{x}^{(m)}\|_p = 2m$ and $\|\widetilde{y}^{(m)}\|_p^p < 2m\frac{1}{(M-\delta)^p} + 2$, so that

$$\|(A^{c,d} - \lambda I)^{-1}\|_p^p \;\geq\; \frac{\|\widetilde{x}^{(m)}\|_p^p}{\|\widetilde{y}^{(m)}\|_p^p} \;>\; \frac{2m}{\frac{2m}{(M-\delta)^p} + 2} \;=\; \frac{1}{\frac{1}{(M-\delta)^p} + \frac{1}{m}}.$$

In either case, Case 1 or 2, these inequalities hold for all $\delta > 0$ and all $m \in \mathbb{N}$. Hence, and applying Lemma 3.2,

$$\|(A_n^f - \lambda I_n)^{-1}\|_p = M \leq \|(A^{c,d} - \lambda I)^{-1}\|_p = \|(A^{cd} - \lambda I)^{-1}\|_p \leq \|(A^b - \lambda I)^{-1}\|_p,$$

where the last inequality follows by [8, Theorem 5.12(ix)], since A^{cd} is a limit operator of A^b by Lemma 2.4. ■

The following corollary is immediate.

Corollary 4.3 *If b is pseudo-ergodic and $n \in \mathbb{N}$ then, for all $\varepsilon > 0$ and $p \in [1, \infty]$,*

$$\sigma_{n,\varepsilon}^p := \bigcup_{c \in \{\pm 1\}^{n-1}} \operatorname{spec}_\varepsilon^p A_n^c \;\subseteq\; \operatorname{spec}_\varepsilon^p A^b = \Sigma_\varepsilon^p,$$

and in particular $\operatorname{spec}_\varepsilon^p A_n^b \;\subset\; \operatorname{spec}_\varepsilon^p A^b$.

4.2 Convergence of the finite matrix spectral sets to their infinite matrix counterparts

As we have remarked at the beginning of this section, it is not clear that the spectrum of a general banded matrix should have anything to do with the spectra of its finite submatrices. In particular, it need not be the case either that the spectrum of a large finite submatrix is contained in a neighbourhood of the spectrum of the corresponding infinite matrix, or that the converse statement is true. But the situation is somewhat more positive for the pseudospectrum, namely that, as we show for a general tridiagonal matrix as our first result of this section (and our method of argument applies to banded matrices more generally), the ε-pseudospectrum of the infinite matrix is contained in the ε'-pseudospectrum of an appropriately chosen $n \times n$ submatrix, for a given $\varepsilon' > \varepsilon$, provided n is sufficiently large. The argument is based on a standard and rather obvious idea: the point is that every eigenvector, or approximate eigenvector, of the infinite matrix is, when truncated in a careful way, also an approximate eigenvector of the finite matrix.

The opposite statement is, in general, false; an approximate eigenvector of a large finite matrix is an approximate eigenvector also of an infinite matrix, but the infinite matrix that it is an approximate eigenvector of need not be the infinite matrix whose spectrum one wishes to approximate! (One recent result which expresses this idea very precisely in the ℓ^2 case for a version of the finite section method for the class of general pseudo-ergodic tridiagonal matrices is [30, Theorem 2.14].) But, for the pseudo-ergodic operators A^b and A^b_+ that we are studying, we have also shown, in Corollary 4.3 and Theorem 3.6, that $\operatorname{spec}^p_\varepsilon A^b_n \subset \operatorname{spec}^p_\varepsilon A^b = \operatorname{spec}^p_\varepsilon A^p_+$. Putting this result together with Theorem 4.4 proves that $\operatorname{spec}^p_\varepsilon A^b_n \nearrow \operatorname{spec}^p_\varepsilon A^b$ (using the notation of the introduction and Section 1.2).

Theorem 4.4 *Suppose that $A = (a_{ij})_{i,j\in\mathbb{Z}}$ is a bi-infinite tridiagonal matrix with $M := \sup_{ij} |a_{ij}| < \infty$. Define the semi-infinite matrix A_+ by $A_+ = (a_{ij})_{i,j\in\mathbb{N}}$ and, for $\ell, m \in \mathbb{N}$ with $\ell \le m$, define the finite matrix of order $m + 1 - \ell$ by $A_{\ell,m} = (a_{ij})_{i,j\in\{\ell,\ldots,m\}}$. Then, for every $\varepsilon' > \varepsilon > 0$ and $p \in [1, \infty]$, there exists $N \in \mathbb{N}$ such that*

$$\operatorname{spec}^p_\varepsilon A \subset \operatorname{spec}^p_{\varepsilon'} A_{\ell,m}, \qquad \text{for } \ell \le -N \text{ and } m \ge N, \tag{30}$$

and

$$\operatorname{spec}^p_\varepsilon A_+ \subset \operatorname{spec}^p_{\varepsilon'} A_{1,m}, \qquad \text{for } m \ge N. \tag{31}$$

Proof. We will prove (30). The proof of (31) is similar.

As a first step we will show that, given some $\varepsilon > 0$ and $p \in [1, \infty]$, for every $\lambda \in \operatorname{spec}^p_\varepsilon A$ there exists $N \in \mathbb{N}$ (depending on λ) such that $\lambda \in \operatorname{spec}^p_\varepsilon A_{\ell,m}$ if $\ell \le -N$ and $m \ge N$. We then combine this result with a compactness argument to obtain the proof of the theorem.

So suppose that $\varepsilon > 0$, $p \in [1, \infty]$, and that $\lambda \in \operatorname{spec}^p_\varepsilon A$. Then, by Theorem 1.1(iv), either $\lambda \in \operatorname{spec}^p_{\mathrm{point},\varepsilon} A$ or $\bar\lambda \in \operatorname{spec}^q_{\mathrm{point},\varepsilon} A^*$, where $p^{-1} + q^{-1} = 1$.

Suppose first that $\lambda \in \operatorname{spec}^p_{\mathrm{point},\varepsilon} A$, i.e. that there exists $x \in \ell^p(\mathbb{Z})$ with $\|x\|_p = 1$ and $\tilde\varepsilon := \|y\|_p < \varepsilon$, where $y := (A - \lambda I)x$. In the case $p < \infty$, let $\widetilde{x} := (x_\ell, \ldots, x_m)^T$ and $\widetilde{y} := (A_{\ell,m} - \lambda I_{m+1-\ell})\widetilde{x}$, so that $\widetilde{y}_k = y_k$, $k = \ell + 1, \ldots, m - 1$. Since $|x_k| \to 0$ as $|k| \to \infty$, it is easy to see that we can, given $\delta > 0$, choose N such that $\|\widetilde{y}\|_p < \tilde\varepsilon + \delta$ and $\|\widetilde{x}\|_p > 1 - \delta$ whenever $\ell \le -N$ and $m \ge N$. But this implies that $\lambda \in \operatorname{spec}^p_\varepsilon A_{\ell,m}$ if N is large enough and $\ell \le -N$ and $m \ge N$. In the case $p = \infty$ we have to modify this argument slightly. Given $\ell \le -N$ and $m \ge N$ put $\widetilde{x} = (\widetilde{x}_\ell, \ldots \widetilde{x}_m)^T$ with $\widetilde{x}_k := \omega_N(k) x_k$, $k = \ell, \ldots, m$, and $\omega_N(k) := \max(0, 1 - |k|/(N-1))$, $k \in \mathbb{Z}$, and let $\widetilde{y} := (A_{\ell,m} - \lambda I_{m+1-\ell})\widetilde{x}$. Then, for $i = \ell, \ldots, m$,

$$\begin{aligned} |\widetilde{y}_i| &= \left|a_{i,j-1}\widetilde{x}_{j-1} + a_{ij}\widetilde{x}_j + a_{i,j+1}\widetilde{x}_{j+1}\right| \\ &= \left|\omega_N(j) y_j + a_{i,j-1}(\omega_N(j-1) - \omega_N(j)) x_{j-1} + a_{i,j+1}(\omega_N(j+1) - \omega_N(j)) x_{j+1}\right| \\ &\le |y_i| + 2M\|x\|_\infty/(N-1) < \varepsilon + 2M/(N-1), \end{aligned}$$

since $|\omega_N(j) - \omega_N(j+1)| \leq (N-1)^{-1}$, for $k \in \mathbb{Z}$. Since also, for each $k \in \ell, ..., m$, $\widetilde{x}_k \to x_k$ as $N \to \infty$, it is clear that, for every $\delta > 0$, if N is chosen large enough, then $\|\widetilde{x}\|_\infty \geq 1 - \delta$, and also $\|\widetilde{y}\|_\infty < \tilde{\varepsilon} + \delta$. But this implies that $\lambda \in \mathrm{spec}^p_\varepsilon A_{\ell,m}$ if N is large enough and $\ell \leq -N$ and $m \geq N$.

If $\bar{\lambda} \in \mathrm{spec}^q_{\mathrm{point},\varepsilon} A^*$ then essentially the identical argument shows that $\bar{\lambda} \in \mathrm{spec}^q_\varepsilon A^*_{\ell,m}$. But this implies that $\lambda \in \mathrm{spec}^p_\varepsilon A_{\ell,m}$ [40, Section 4]. This completes the proof of the first step.

To finish the proof of the theorem we argue as follows. Given $\varepsilon' > \varepsilon > 0$ and $p \in [1, \infty]$, let $\eta := (\varepsilon' - \varepsilon)/2$, and $\varepsilon^* = \varepsilon + \eta$. Let $S := \overline{\mathrm{spec}^p_\varepsilon A}$, and let $O := \{\lambda + \eta\mathbb{D} : \lambda \in S\}$. Then O is an open cover of the compact set S, and so has a finite subcover, i.e. there exists a finite set $\Lambda \subset \overline{\mathrm{spec}^p_\varepsilon A}$ with $S \subset \bigcup_{\lambda \in \Lambda}(\lambda + \eta\mathbb{D}) = \eta\mathbb{D} + \Lambda$. Now $\Lambda \subset \overline{\mathrm{spec}^p_\varepsilon A} \subset \mathrm{spec}^p_{\varepsilon^*} A$. Applying the result shown in the first step, we see that we can choose N so that, for $\ell \leq -N$ and $m \geq N$, $\Lambda \subset \mathrm{spec}^p_{\varepsilon^*} A_{\ell,m}$. Thus $\mathrm{spec}^p_\varepsilon A \subset S \subset \eta\mathbb{D} + \Lambda \subset \eta\mathbb{D} + \mathrm{spec}^p_{\varepsilon^*} A_{\ell,m} \subset \mathrm{spec}^p_{\varepsilon'} A_{\ell,m}$, by (15). ■

To apply this result, for $\ell, m \in \mathbb{Z}$ with $\ell \leq m$, let $A^b_{\ell,m}$ denote $A_{\ell,m}$, the matrix of order $m+1-\ell$ as defined in the above theorem, in the case that $A = A^b$. So, in particular, $A^b_{1,n} = A^b_n$ for $n \in \mathbb{N}$.

Corollary 4.5 *If $b \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic then, for every $\varepsilon > 0$ and $p \in [1, \infty]$,*

$$\mathrm{spec}^p_\varepsilon A^b_n = \mathrm{spec}^p_\varepsilon A^b_{1,n} \nearrow \mathrm{spec}^p_\varepsilon A^b_+ = \Sigma^p_\varepsilon, \qquad \text{as } n \to \infty.$$

If $b \in \{\pm 1\}^{\mathbb{Z}}$ is pseudo-ergodic then, for every $\varepsilon > 0$ and $p \in [1, \infty]$,

$$\mathrm{spec}^p_\varepsilon A^b_{\ell,m} \nearrow \mathrm{spec}^p_\varepsilon A^b = \Sigma^p_\varepsilon, \qquad \text{as } \ell \to -\infty \text{ and } m \to \infty.$$

Proof. We will prove the second of these statements. The proof of the first is similar. From Corollary 4.3 and Theorem 4.4, given any $\varepsilon' \in (0, \varepsilon)$ there exists $N \in \mathbb{N}$ such that

$$\mathrm{spec}^p_{\varepsilon'} A \subset \mathrm{spec}^p_\varepsilon A_{\ell,m} \subset \mathrm{spec}^p_\varepsilon A, \qquad \text{for } \ell \leq -N \text{ and } m \geq N. \tag{32}$$

Since, from (18), $\mathrm{spec}^p_{\varepsilon'} A \nearrow \mathrm{spec}^p_\varepsilon A$ as $\varepsilon' \to \varepsilon^-$, it follows that $\mathrm{spec}^p_\varepsilon A^b_{\ell,m} \nearrow \mathrm{spec}^p_\varepsilon A^b$ (which is equal to Σ^p_ε by Theorem 3.6), as $\ell \to -\infty$ and $m \to \infty$. ■

A similar result holds for the convergence of the numerical range, as an instance of the general result Theorem 1.3. Note that while convergence of the pseudospectra needs that b is pseudo-ergodic, to ensure that the matrix pseudospectra are contained in the operator pseudospectra, the corresponding inclusion (19) for numerical ranges holds for *any* bounded linear operator, so that we need no constraint on b. The following is thus an immediate corollary of Theorem 1.3 and Lemma 3.1.

Corollary 4.6 *If $b \in \{\pm 1\}^{\mathbb{N}}$ then*

$$W(A^b_n) = W(A^b_{1,n}) \nearrow W(A^b_+), \qquad \text{as } n \to \infty,$$

with $W(A^b_+) = \Delta$ if b is pseudo-ergodic. If $b \in \{\pm 1\}^{\mathbb{Z}}$ then

$$W(A^b_{\ell,m}) \nearrow W(A^b), \qquad \text{as } \ell \to -\infty \text{ and } m \to \infty.$$

with $W(A^b) = \Delta$ if b is pseudo-ergodic.

4.3 Quantitative convergent approximations to the spectrum and pseudospectrum

In this section we present numerical algorithms for approximating Σ and Σ^2_ε which are, respectively, from Theorem 3.6, the spectrum and the ℓ^2 ε-pseusdospectrum of both A^b and A^b_+ in the case when b is pseudo-ergodic.

The previous subsection already provides potential methods for computing these sets. We have that, if $b \in \{\pm 1\}^n$ is pseudo-ergodic, then

$$\Sigma_\varepsilon^2 = \lim_{n\to\infty} \operatorname{spec}_\varepsilon^2 A_n^b. \tag{33}$$

This then implies, by (18), that

$$\Sigma = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \operatorname{spec}_\varepsilon^2 A_n^b. \tag{34}$$

In principle, these equations can be used as the basis of algorithms for computing Σ_ε^p and Σ. In particular, to approximate Σ_ε^2 one uses the sequence of sets $\operatorname{spec}_\varepsilon^2 A_n^b$, $n = 1, 2, ...$, which can be computed as described in Section 1.2. The difficulty with this scheme is that one has no idea of the rate of convergence of $\operatorname{spec}_\varepsilon^2 A_n^b$ to Σ. Indeed it is clear that it can be arbitrarily slow: to see this consider that if $c \in \{\pm 1\}^{\mathbb{N}}$ is pseudo-ergodic, then so is $b \in \{\pm 1\}^{\mathbb{N}}$ if $b_m = c_m$ for all sufficiently large m. But this means that it can hold that b is pseudo-ergodic and that $b_m = 1$, for $1 \le m \le N$, with N arbitrarily large. If this is the case then A_n^b is self-adjoint and thus, and by Lemma 3.1, $\operatorname{spec} A_n^b \subset (-2, 2)$ and $\operatorname{spec}_\varepsilon^2 A_n^b = \operatorname{spec} A_n^b + \varepsilon\mathbb{D} \subset (-2, 2) + \varepsilon\mathbb{D}$, for $n \le N$. So if, e.g., $N = 10^9$ then, while ultimately $\operatorname{spec}_\varepsilon^2 A_n^b \to \Sigma_\varepsilon^2$, there is no early sign of this.

The situation with (34) is rather worse. This equation implies that there exists some sequence of positive reals $\varepsilon_n \to 0$ for which it holds that

$$\operatorname{spec}_{\varepsilon_n}^2 A_n^b \to \Sigma,$$

but provides neither a recipe for choosing the ε_n nor any guarantee of the rate of convergence.

The source of the difficulty regarding the rate of convergence can be traced back to Theorem 4.4 and its proof, this theorem a key ingredient in the proof of Corollary 4.5 and so of (33). This theorem guarantees that, for every $\varepsilon' > 0$, $\Sigma_\varepsilon^2 = \operatorname{spec}_\varepsilon^2 A_+^b \subset \operatorname{spec}_{\varepsilon'}^2 A_n^b$ for all n sufficiently large, but gives no idea of how large n should be. And indeed we have argued above that there is no upper bound on how large n may need to be for this equation to hold for a given pseudo-ergodic b.

This difficulty has been resolved in recent work by the authors [6], who quantify, for general tridiagonal matrices, by a sharpened version of the arguments of Theorem 4.4, adapted particularly to the case $p = 2$, exactly how ε' should depend on n in (31), but at the expense of replacing in this equation the pseudospectrum of a single $n \times n$ submatrix by the union of the pseudospectra of *all* possible $n \times n$ principal submatrices. The results in [6] are much more general, but we will restrict the exposition here to how these results apply to the bi-infinite matrix A^b with $b \in \{\pm 1\}^Z$. Using the notation of Corollary 4.5, the result shown in [6] (or see [10, Corollary 3.7]) is the following when applied to A^b:

Theorem 4.7 *For $b \in \{\pm 1\}^{\mathbb{Z}}$, $\varepsilon > 0$, and $n \in \mathbb{N}$,*

$$\operatorname{spec}_\varepsilon^2 A^b \subset \bigcup_{\ell\in\mathbb{Z}} \operatorname{spec}_{\varepsilon+\varepsilon_n}^2 A_{\ell,\ell+n-1}^b,$$

where $\varepsilon_n = 4\sin\theta_n \le 2\pi/(n+2)$, with θ_n the unique solution in the interval $\left(\dfrac{\pi}{2(n+3)}, \dfrac{\pi}{2(n+2)}\right]$ of the equation $2\cos((n+1)\theta) = \cos((n-1)\theta)$. Further,

$$\operatorname{spec} A^b \subset \overline{\bigcup_{\ell\in\mathbb{Z}} \operatorname{spec}_{\varepsilon_n}^2 A_{\ell,\ell+n-1}^b} = \bigcup_{\ell\in\mathbb{Z}} \operatorname{Spec}_{\varepsilon_n}^2 A_{\ell,\ell+n-1}^b.$$

An important point is that the unions of pseudospectra over $\ell \in \mathbb{Z}$ in the above equations reduce to finite unions, because there are only 2^{n-1} distinct $n \times n$ matrices A_n^c with $c \in \{\pm 1\}^{n-1}$. In

the notation introduced in Corollary 4.3, it must hold that $\bigcup_{\ell\in\mathbb{Z}} \operatorname{spec}^2_\eta A^b_{\ell,\ell+n-1} \subset \sigma^2_{n,\eta}$, for every $\eta > 0$. For small values of n, ε_n in the above theorem can be calculated explicitly, in particular

$$\varepsilon_1 = 2 \text{ and } \varepsilon_2 = \sqrt{2}. \tag{35}$$

Example 4.8 As a first example of application of the above theorem, consider the case when $b_m = 1$ for each m. Then $A^b_{\ell,\ell+n-1} = A^b_{1,n} = A^b_n$ for each ℓ. Further, this matrix is self-adjoint, so that $\operatorname{spec}^2_\eta A^b_n = \operatorname{spec} A^b_n + \eta\mathbb{D}$, for every $\eta > 0$. Thus the statements of the theorem reduce to

$$\operatorname{spec} A^b \subset \operatorname{spec} A^b_n + \varepsilon_n\overline{\mathbb{D}} \;\text{ and }\; \operatorname{spec}^2_\varepsilon A^b \subset \operatorname{spec} A^b_n + (\varepsilon+\varepsilon_n)\mathbb{D},\ \varepsilon > 0. \tag{36}$$

In this simple case we can compute the above sets explicitly, to check that the above inclusions hold, finding that $\operatorname{spec} A^b = [-2,2]$, $\operatorname{spec}^2_\varepsilon A^b = [-2,2] + \varepsilon\mathbb{D}$, and $\operatorname{spec} A^b_n = \left\{2\cos\frac{j\pi}{n+1} : j = 1,...,n\right\}$. Elementary calculations show that the inclusions (36) do hold in this case, in fact one can calculate (see [10, Section 3.2.2] for details), if ε_n were replaced with $\varepsilon^*_n \le \varepsilon_n$ in the above inclusions, the smallest value of ε^*_n for which the inclusions would still hold. This is $\varepsilon^*_1 = 2$, $\varepsilon^*_n = 2\sin(\pi/(2(n+1)))$ if n is even (in particular $\varepsilon^*_2 = 1$), and $\varepsilon^*_n = \sin(\pi/(n+1))$ if $n \ge 3$ is odd. Thus $\varepsilon_n/\varepsilon^*_n = 1$ for $n = 1$ (the bound (36) is sharp for $n = 1$) and $\varepsilon_n/\varepsilon^*_n \to 2$ as $n \to \infty$. □

The main example of interest to us here is the case where $b \in \{\pm1\}^{\mathbb{Z}}$ is pseudo-ergodic. Recall from Theorem 3.6 that $\operatorname{spec} A^b = \Sigma$ and $\operatorname{spec}^2_\varepsilon A^b = \Sigma^2_\varepsilon$ in that case. Combining Theorem 4.7 with Theorem 4.1, Corollary 4.3, (13), (18) and (17), we obtain the following result.

Theorem 4.9 *For $\varepsilon > 0$ and $n \in \mathbb{N}$,*

$$\sigma_n \subset \Sigma \subset \overline{\sigma^2_{n,\varepsilon_n}} \subset \overline{\Sigma^2_{\varepsilon_n}} \text{ and } \Sigma^2_\varepsilon \subset \sigma^2_{n,\varepsilon+\varepsilon_n} \subset \Sigma^2_{\varepsilon+\varepsilon_n}, \tag{37}$$

where ε_n is defined as in Theorem 4.7. Further,

$$\sigma_n + \varepsilon\mathbb{D} \subset \Sigma^2_\varepsilon, \text{ for } \varepsilon > 0, \quad \text{and} \quad \Sigma^2_{\varepsilon-\varepsilon_n} \subset \sigma^2_{n,\varepsilon} \subset \Sigma^2_\varepsilon, \text{ for } \varepsilon > \varepsilon_n. \tag{38}$$

Moreover, as $n \to \infty$, $\overline{\sigma^2_{n,\varepsilon_n}} \searrow \Sigma$ and, for $\varepsilon > 0$, $\sigma^2_{n,\varepsilon+\varepsilon_n} \searrow \Sigma^2_\varepsilon$ and $\sigma^2_{n,\varepsilon} \nearrow \Sigma^2_\varepsilon$.

In most respects this result is superior to Corollary 4.5. It provides both upper and lower bounds for Σ^2_ε, moreover these converge to Σ^2_ε as $n \to \infty$ at guaranteed convergence rates (at least as fast as $\Sigma_{\varepsilon-\varepsilon_n}$ and $\Sigma_{\varepsilon+\varepsilon_n}$, respectively). Further, the theorem provides an upper bound which is convergent to Σ, at least as fast as $\Sigma^2_{\varepsilon_n}$. Of course, that the convergence rates are guaranteed is at a cost: evaluating $\sigma^2_{n,\eta}$ for some $n \in \mathbb{N}$ and $\eta > 0$ requires exponentially large computation for n large. Precisely, using the characterisation (11), we see that

$$\sigma^2_{n,\eta} = \{\lambda \in \mathbb{C} : S_n(\lambda) < \eta\}, \tag{39}$$

where

$$S_n(\lambda) := \min_{c\in\{\pm1\}^{n-1}} s_{\min}(A^c_n - \lambda I_n), \quad \lambda \in \mathbb{C}. \tag{40}$$

Clearly, computing $S_n(\lambda)$ for a particular λ, to check membership of $\sigma^2_{n,\eta}$, requires calculation of the smallest singular value of 2^{n-1} matrices of order n. Note that it follows from (12) that

$$|S_n(\lambda) - S_n(\mu)| \le |\lambda - \mu|, \quad \lambda, \mu \in \mathbb{C}. \tag{41}$$

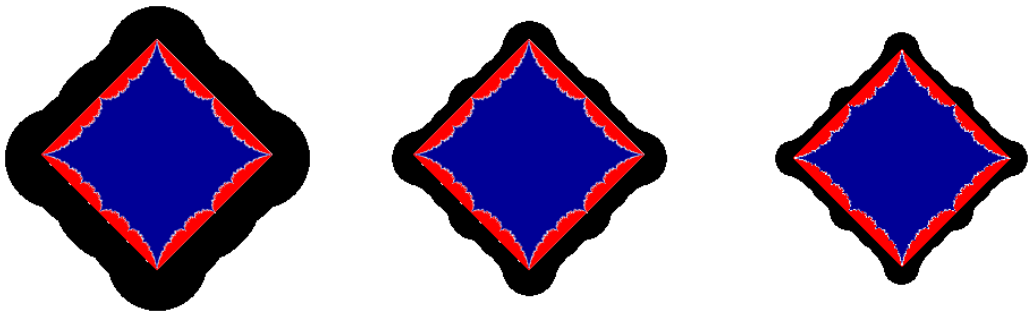


Figure 6: Plots, for $n = 6, 12$ and 18, of the sets $\overline{\sigma^2_{n,\varepsilon_n}}$, which are inclusion sets for $\Sigma = \operatorname{spec} A^b$, when $b \in \{\pm 1\}^{\mathbb{Z}}$ is pseudo-ergodic. Also shown, overlaid in red, is the square Δ, with corners at ± 2 and $\pm 2\mathrm{i}$, which is $W(A^b)$, the numerical range of A^b. Overlaid on top of that in blue is the set $\pi_{30} \cup \mathbb{D}$ which, by definition and Theorem 2.7, is a subset of Σ.

In Figure 6 we plot $\overline{\sigma^2_{n,\varepsilon_n}}$, for $n = 6, 12$, and 18. Each of these sets contains Σ, by Theorem 4.9, and note that each set is invariant under reflection in either axis or under rotation by 90^0, by Lemma 3.4. On the same figure we plot the square $\overline{\Delta}$ which, by Lemma 3.1, also contains Σ. It appears that, for $n \leq 18$, $\Delta \subset \sigma^2_{n,\varepsilon_n}$. If this were to hold for all $n \in \mathbb{N}$ then it would follow, from Theorem 4.9, which tells us that $\overline{\sigma^2_{n,\varepsilon_n}} \searrow \Sigma$, and Lemma 3.1, which tells us that $\Sigma \subset \overline{\Delta}$, that $\Sigma = \overline{\Delta}$. It seems impossible from these plots to take an educated guess as to whether or not $\Delta \subset \sigma^2_{n,\varepsilon_n}$ holds for all n, not least because the convergence rate of $\overline{\sigma^2_{n,\varepsilon_n}}$ to Σ may be slow: Theorem 4.9 tells us that $\operatorname{dist}(\overline{\sigma^2_{n,\varepsilon_n}}, \Sigma) \leq \operatorname{dist}(\Sigma^2_{\varepsilon_n}, \Sigma)$ but it follows from (13) that $\operatorname{dist}(\Sigma^2_{\varepsilon_n}, \Sigma) \geq \varepsilon_n \approx 2\pi/(n+2)$.

We have not been able to produce similar plots to those in Figure 6 for much larger values of n because of the large computational cost. But it is feasible to compute $S_n(\lambda)$ for a single λ for larger n. We have carried out this computation for $\lambda = 1.5 + 0.5\mathrm{i}$, a quarter of the way along one of the sides of Δ. Computing in standard double-precision floating point arithmetic we find that

$$S_{34}(1.5 + 0.5\mathrm{i}) = 0.17201954132506... > \varepsilon_{34} = 0.169830415547956... \,. \tag{42}$$

This implies that $1.5 + 0.5\mathrm{i} \notin \sigma^2_{34,\varepsilon_{34}}$ and so $1.5 + 0.5\mathrm{i} \notin \Sigma$, which of course implies that Σ is a strict subset of $\overline{\Delta}$. In fact, in view of (41) and the symmetries of Σ noted in Lemma 3.2, the inequality (42) implies more, namely that

$$(\pm(1.5 \pm 0.5\mathrm{i}) + \eta\mathbb{D}) \cap \Sigma = \emptyset,$$

for $\eta = \varepsilon_{34} - S_{34}(1.5 + 0.5\mathrm{i}) = 0.0021891257771....$

We note that the computation required to evaluate $S_{34}(1.5 + 0.5\mathrm{i})$ and so establish that $1.5 + 0.5\mathrm{i} \notin \Sigma$ is considerable: we need to evaluate the smallest singular value of $2^{33} \approx 8.6 \times 10^9$ matrices of order 34 (of course these computations are ideally suited for parallel implementation). We note that it seems to be necessary to use n as large as 34, in that other computations show that $S_{33}(1.5 + 0.5\mathrm{i}) < \varepsilon_{33}$, so that $1.5 + 0.5\mathrm{i} \in \sigma^2_{33,\varepsilon_{33}}$.

5 The Random Case and Concluding Remarks

We finish this paper by spelling out the implications of the above results for the finite matrices A^b_n and $A^{b,c}_n$, the bi-infinite matrices A^b and $A^{b,c}$, and the semi-infinite matrices A^b_+ and $A^{b,c}_+$, in the case when the entries of b and c are random variables taking the values ± 1.

Theorem 5.1 *Suppose that the entries of* $b \in \{\pm 1\}^{\mathbb{Z}}$ *are iid random variables, with* $\Pr(b_m = 1) \in (0,1)$*. Then:*

(i) $\operatorname{spec} A^b \subset \Sigma$, $\operatorname{spec} A^b_+ \subset \Sigma$, *with* $\operatorname{spec}_{\text{ess}} A^b = \operatorname{spec} A^b = \operatorname{spec}_{\text{ess}} A^b_+ = \operatorname{spec} A^b_+ = \Sigma$ *almost surely.*

(ii) $W(A^b_+) \subset W(A^b) \subset \Delta$, *with* $W(A^b) = W(A^b_+) = \Delta$ *almost surely.*

(iii) *For* $n \in \mathbb{N}$, $\operatorname{spec} A^b_n \subset \Sigma$ *and* $W(A^b_n) \subset \Delta$, *and, as* $n \to \infty$, $W(A^b_n) \nearrow \Delta$, *almost surely.*

(iv) *For* $\varepsilon > 0$ *and* $p \in [1,\infty]$, $\operatorname{spec}^p_\varepsilon A^b \subset \Sigma^p_\varepsilon$, $\operatorname{spec}^p_\varepsilon A^b_+ \subset \Sigma^p_\varepsilon$, *with* $\operatorname{spec}^p_\varepsilon A^b = \operatorname{spec}^p_\varepsilon A^b_+ = \Sigma^p_\varepsilon$ *almost surely.*

(v) *For* $\varepsilon > 0$, $p \in [1,\infty]$, *and* $n \in \mathbb{N}$, $\operatorname{spec}^p_\varepsilon A^b_n \subset \Sigma^p_\varepsilon$ *and, as* $n \to \infty$, $\operatorname{spec}^p_\varepsilon A^b_n \nearrow \Sigma^p_\varepsilon$, *almost surely.*

Similarly, if $b, c \in \{\pm 1\}^{\mathbb{Z}}$, *and the entries of* bc *are iid random variables, with* $\Pr(b_m c_m = 1) \in (0,1)$, *then (i)-(v) hold with* A^b, A^b_+, A^b_n, *replaced by* $A^{b,c}$, $A^{b,c}_+$, *and* $A^{b,c}_n$, *respectively.*

Proof. To see that (i)-(v) hold, note that, by Lemma 2.3 and the remarks at the end of Section 2, the condition of the theorem imply that b and also $b_+ := (b_1, b_2, ...)$ are pseudo-ergodic with probability one. Then (i) follows from the definition of Σ in Theorem 2.5, and from Lemma 3.5 and Theorem 3.6. That (ii) and (iii) hold follows from Lemma 3.1, Theorem 4.1 and Corollary 4.6. That (iv) holds follows from the definition of Σ^p_ε in Theorem 2.5, and from Lemma 3.5 and Theorem 3.6. Finally, (v) follows from corollaries 4.3 and 4.5. That (i)-(v) hold for the case where A^b, A^b_+, A^b_n are replaced by $A^{b,c}$, $A^{b,c}_+$, and $A^{b,c}_n$, respectively, and the entries of bc are iid random variables, with $\Pr(b_m c_m = 1) \in (0,1)$, follow using the same results, with the help of lemmas 3.2 and 3.3 and the observations between these lemmas on the semi-infinite case, and noting first that these assumptions imply that bc and $(b_1c_1, b_2c_2, ...)$ are pseudo-ergodic, with probability one. ■

Of course, in the above theorem $\Delta = \{z = a + \mathrm{i}b : a, b \in \mathbb{R}, |a| + |b| < 2\}$ and Σ and Σ^p_ε are as defined in Theorem 2.5. The following theorem summarises, for the convenience of the reader, what we have established in the sections above about the compact set Σ and the bounded open sets Σ^p_ε. Recall that σ_∞ and π_∞ are defined by (7) and (6), respectively, $\sigma^p_{n,\varepsilon}$ is defined in Corollary 4.3 (and see (39) for $p = 2$), ε_n is defined in Theorem 4.7, and $S_n(\lambda)$ in (40).

Theorem 5.2 *For* $\varepsilon > 0$ *and* $p \in [1,\infty]$, *where* $q \in [1,\infty]$ *is given by* $p^{-1} + q^{-1} = 1$*:*

(i) $\overline{\mathbb{D}} \subset \Sigma \subset \overline{\Delta}$, *and* Σ *is a strict subset of* $\overline{\Delta}$ *provided* $\varepsilon_{34} - S_{34}(1.5 + 0.5\mathrm{i}) > 0$, *for which see (vi).*

(ii) $\sigma_\infty \subset \pi_\infty \subset \Sigma \subset \Sigma^p_\varepsilon \subset \Sigma^p_{\varepsilon'}$, *for* $\varepsilon' > \varepsilon$. *(See Figures 2, 3, 4 and 5 for visualisations of* σ_n *and* π_n, *for* $n \in \mathbb{N}$, *and their interrelation.)*

(iii) Σ *and* Σ^p_ε *are invariant under reflection in the real and imaginary axes and under rotation by* 90^0.

(iv) $\Sigma^p_\varepsilon = \Sigma^q_\varepsilon$ *and* $\Sigma^p_\varepsilon \subset \Sigma^r_\varepsilon$ *if* $1 \le r \le p \le 2$, *so that* $\Sigma^2_\varepsilon = \bigcap_{r \in [1,\infty]} \Sigma^r_\varepsilon$.

(v) *As* $n \to \infty$, $\overline{\sigma^2_{n,\varepsilon_n}} \searrow \Sigma$, $\sigma^2_{n,\varepsilon+\varepsilon_n} \searrow \Sigma^2_\varepsilon$, *and* $\sigma^2_{n,\varepsilon} \nearrow \Sigma^2_\varepsilon$. *(See Figure 6 for visualisations of* $\sigma^2_{n,\varepsilon_n}$, *for* $n = 6, 12, 18$.*)*

(vi) *For* $\lambda = 1.5 + 0.5\mathrm{i}$, *provided* $\eta = \varepsilon_{34} - S_{34}(\lambda) > 0$ *(and floating point calculations give* $\eta \approx 0.00219$*), it holds that* $\pm(1.5 \pm 0.5\mathrm{i}) + \eta\mathbb{D} \cap \Sigma = \emptyset$.

Proof. Part (i) follows from Theorem 2.7 (taken from [5]) and Lemma 3.1, and that Σ is a strict subset of Δ holds, as discussed at the end of 4.3, provided $\eta = \varepsilon_{34} - S_{34}(1.5 + 0.5\mathrm{i}) > 0$. Part (ii) is Theorem 4.1, with $\Sigma^p_\varepsilon \subset \Sigma^p_{\varepsilon'}$ because $\Sigma^p_\varepsilon = \mathrm{spec}^p_\varepsilon A^b$ if $b \in \{\pm 1\}^{\mathbb{Z}}$ is pseudo-ergodic (Theorem 3.6). Part (iii) is Lemma 3.4, (iv) is from 3.2, (v) is part of Theorem 4.9, and (vi) is from the end of Section 4.3. ■

It is clear from the above results that we understand well, in Theorem 5.1, the interrelation between the numerical ranges and pseudospectra of the semi-infinite, bi-infinite, and finite random matrix cases, and have shown that the almost sure spectrum is the same set Σ for the semi-infinite and bi-infinite cases, and contains the spectrum in the finite matrix case. Interesting open questions are whether or not, similarly to the analogous results for the pseudospectra, $\mathrm{spec}\, A^b_n \nearrow \Sigma$ almost surely as $n \to \infty$, which would imply that σ_∞ is dense in Σ, so that π_∞ is dense in Σ. (That σ_∞ is dense in Σ was conjectured in [5].) Note that, if it does hold that $\mathrm{spec}\, A^b_n \nearrow \Sigma$ almost surely, then both Figs 2 and 3 are visualisations of sequences of sets converging to Σ.

Regarding the geometry of Σ (and of the pseudospectra Σ^p_ε), we have some information in Theorem 5.2, including in the last part of this theorem establishing a computable sequence of sets converging from above to Σ (a sequence of three of these plotted in Figure 6). However there is much that is not known. Is Σ connected (which would imply, by general results on pseudospectra [40, Theorem 4.3], that also Σ^p_ε is connected)? In fact, is Σ simply-connected? What is the geometry of the boundary of Σ, and the geometry of the sets σ_n, the finite-dimensional analogues of Σ (cf. Figure 5)? We have conjectured in [5] that Σ is a simply-connected set which is the closure of its interior and which has a fractal boundary, which is plausible from, or at least consistent with, Figure 6, if it holds that $\overline{\sigma_\infty} = \Sigma$. Our methods and results provide no information about what is a usual concern of research on random matrices, to obtain asymptotically in the limit as $n \to \infty$ the pdf of the density of eigenvalues, except, of course, that we have shown in Theorem 5.1(iii) that the support of this pdf is a subset of Σ.

There are many possibilities for applying the methods introduced in this paper to much larger classes of random (or pseudo-ergodic) operators. For some steps in this direction we refer the reader to [30, 6, 9].

Acknowledgements. We are grateful to Estelle Basor from the American Institute of Mathematics for drawing our attention to this beautiful operator class, and are grateful for feedback on our work in progress from Brian Davies and Eugene Shargorodsky (KCL) and from Titus Hilberdink and Michael Levitin (Reading), including the feedback that prompted the computation of $S_{34}(1.5 + 0.5\mathrm{i})$ in Section 4.3. In regard to this computation, we are grateful for the assistance with programming and parallel implementation of the calculations from Roman Unger of TU Chemnitz. We also acknowledge the financial support of a Higher Education Strategic Scholarship for Frontier Research from the Thai Ministry of Higher Education to the second author, of Marie-Curie Grant MEIF-CT-2005-009758 of the EU to the third and first authors, and Marie-Curie Grant PERG02-GA-2007-224761 of the EU to the third author.

References

[1] A. Böttcher, S. M. Grudsky and B. Silbermann: Norms of inverses, spectra, and pseudospectra of large truncated Wiener-Hopf operators and Toeplitz matrices, *New York J. Math.* **3** (1997), 1-31.

[2] A. Böttcher and S. M. Grudsky: *Spectral Properties of Banded Toeplitz Matrices*, SIAM, Philadelphia 2005.

[3] M. Capinski and E. P. Kopp: *Measure, Integral and Probability*, Springer Verlag 2004.

[4] R. Carmona and J. Lacroix: *Spectral Theory of Random Schrodinger Operators*, Birkhäuser, Boston 1990.

[5] S. N. Chandler-Wilde, R. Chonchaiya and M. Lindner: Eigenvalue problem meets Sierpinski triangle: computing the spectrum of a non-self-adjoint random operator, to appear in *Operators and Matrices*. Preprint at arXiv:1003.3946v3

[6] S. N. Chandler-Wilde, R. Chonchaiya and M. Lindner: Upper bounds on the spectra and pseudospectra of Jacobi and related operators, in preparation.

[7] S. N. Chandler-Wilde and M. Lindner: Sufficiency of Favard's condition for a class of band-dominated operators on the axis, *J. Funct. Anal.* **254** (2008), 1146–1159.

[8] S. N. Chandler-Wilde and M. Lindner: Limit Operators, Collective Compactness, and the Spectral Theory of Infinite Matrices, *Memoirs AMS* **210** (2011), No. 989.

[9] S. N. Chandler-Wilde and E. B. Davies: Spectrum of a Feinberg-Zee random hopping matrix, in preparation.

[10] R. Chonchaiya: Computing the Spectra and Pseudospectra of Non-Self-Adjoint Random Operators Arising in Mathematical Physics, *PhD Thesis, University of Reading, UK,* 2010.

[11] G. M. Cicuta, M. Contedini and L. Molinari: Non-Hermitian tridiagonal random matrices and returns to the origin of a random walk, *J. Stat. Phys.* **98** (2000), 685–699.

[12] G. M. Cicuta, M. Contedini and L. Molinari: Enumeration of simple random walks and tridiagonal matrices. *J. Phys. A Math. Gen.* **35** (2002), 1125–1146.

[13] E. B. Davies: Spectral properties of non-self-adjoint matrices and operators, *Proc. Royal Soc. A.* **457** (2001), 191–206.

[14] E. B. Davies: Spectral theory of pseudo-ergodic operators, *Commun. Math. Phys.* **216** (2001), 687–704.

[15] E. B. Davies: *Linear Operators and their Spectra*, Cambridge University Press, 2007.

[16] J. Feinberg and A. Zee: Non-Hermitean Localization and De-Localization, *Phys. Rev. E* **59** (1999), 6433–6443.

[17] J. Feinberg and A. Zee: Spectral Curves of Non-Hermitean Hamiltonians, *Nucl. Phys. B* **552** (1999), 599–623.

[18] J. Globevnik: Norm-constant analytic functions and equivalent norms, *Illinois J. Math.* **20** (1976), 503-506.

[19] I. Goldsheid and B. Khoruzhenko: Eigenvalue curves of asymmetric tridiagonal random matrices, *Electronic Journal of Probability* **5** (2000), 1–28.

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

[21] N. Hatano and D. R. Nelson: Vortex Pinning and Non-Hermitian Quantum Mechanics, *Phys. Rev. B* **56** (1997), 8651-8673.

[22] D.E. Holz, H. Orland and A. Zee: On the remarkable spectrum of a non-Hermitian random matrix model, *J. Phys. A Math. Gen.* **36** (2003), 3385–3400.

[23] K. Jörgens: *Linear Integral Operators*, Pitman, Boston, 1982.

[24] T. Kato: *Perturbation Theory for Linear Operators*, 2nd edition, Springer, New York, 1980.

[25] B. V. Lange and V. S. Rabinovich: On the Noether property of multidimensional discrete convolutions, *Mat. Zametki* **37** (1985), 407–421 (Russian, English transl. *Math. Notes* **37** (1985), 228–237).

[26] M. Lindner: *Infinite Matrices and their Finite Sections: An Introduction to the Limit Operator Method*, Frontiers in Mathematics, Birkhäuser 2006.

[27] M. Lindner: Fredholm Theory and Stable Approximation of Band Operators and Generalisations, *Habilitation thesis, TU Chemnitz, Germany*, 2009.

[28] M. Lindner: Fredholmness and index of operators in the Wiener algebra are independent of the underlying space, *Operators and Matrices* **2** (2008), 297–306.

[29] M. Lindner: A note on the spectrum of bi-infinite bi-diagonal random matrices, *Journal of Analysis and Applications* **7** (2009), 269-278.

[30] M. Lindner and S. Roch: Finite sections of random Jacobi operators, to appear in *SIAM J. Numer. Anal.*. Preprint at arXiv:1011.0907v1

[31] C. Martínez Adame: On the spectral behaviour of a non-self-adjoint operator with complex potential, *Math. Phys. Anal. Geom.* **10** (2007), 81–95.

[32] D. R. Nelson and N. M. Shnerb: Non-Hermitian localization and population biology, *Phys. Rev. E* **58** (1998), 1383-1403.

[33] L. A. Pastur and A. L. Figotin: *Spectra of Random and Almost-Periodic Operators*, Springer, Berlin 1992.

[34] V. S. Rabinovich, S. Roch and B. Silbermann: Fredholm theory and finite section method for band-dominated operators, *Integral Equations Operator Theory* **30** (1998), 452–495.

[35] V. S. Rabinovich, S. Roch and B. Silbermann: *Limit Operators and Their Applications in Operator Theory*, Birkhäuser 2004.

[36] S. Roch: Numerical ranges of large Toeplitz matrices, *Lin. Alg. Appl.* **282** (1998), 185–198.

[37] E. Shargorodsky: On the level sets of the resolvent norm of a linear operator, *Bull. London Math. Soc.* **40** (2008), 493-504.

[38] E. M. Stein and G. Weiss: *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, 1971.

[39] L. N. Trefethen, M. Contedini and M. Embree: Spectra, pseudospectra, and localization for random bidiagonal matrices, *Comm. Pure Appl. Math.* **54** (2001), 595–623.

[40] L. N. Trefethen and M. Embree: *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*, Princeton University Press, Princeton, NJ, 2005.

Authors' addresses:

Simon N. Chandler-Wilde `S.N.Chandler-Wilde@reading.ac.uk`
Department of Mathematics and Statistics
University of Reading
Reading, RG6 6AX
UK

Ratchanikorn Chonchaiya ratchanikorn@buu.ac.th
Department of Mathematics
Faculty of Science
Burapha University
Longhard Bangsaen Road
Muang, Chonburi
20131, THAILAND

Marko Lindner marko.lindner@mathematik.tu-chemnitz.de
Fakultät Mathematik
TU Chemnitz
D-09107 Chemnitz
GERMANY