

# Transfer matrix analysis of non-Hermitian Hamiltonians: asymptotic spectra and topological eigenvalues

Lars Koekenbier and Hermann Schulz-Baldes

**Abstract.** Transfer matrix techniques are used to provide a new proof of Widom’s results on the asymptotic spectral theory of finite block Toeplitz matrices. Furthermore, a rigorous treatment of the skin effect, spectral outliers, the generalized Brillouin zone and the bulk-boundary correspondence in such systems is given. This covers chiral Hamiltonians with topological eigenvalues close to zero, but no line-gap.

## 1. Overview

This paper is about the spectral theory of translation invariant block-tridiagonal operators of the form

$$H = \begin{pmatrix} \ddots & \ddots & & & & \\ \ddots & V & T & & & \\ & R & \ddots & \ddots & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \end{pmatrix}, \quad (1.1)$$

where  $R, V, T \in \mathbb{C}^{L \times L}$  are all  $L \times L$  complex matrices. This non-Hermitian Hamiltonian  $H$  is viewed as a bounded operator on the Hilbert space  $\ell^2(\mathbb{Z}, \mathbb{C}^L)$ . It is a widely studied model in non-Hermitian quantum mechanics [1, 18, 46] in which, more specifically, the skin effect is discussed [46–48] as well as the existence of robust topological eigenvalues via a so-called *bulk-boundary correspondence* [4, 5, 20, 28, 30, 31, 41, 44, 45, 47]. These latter two physical effects are studied on finite, but large volume compressions  $H_N$  of  $H$ , namely the restrictions of  $H$  to  $\ell^2(\{1, \dots, N\}, \mathbb{C}^L) \cong \mathbb{C}^{NL}$ . It is well known that the spectra of  $H$  and  $H_N$  have little in common, and also differ from the spectrum of the (right) half-space restriction  $H^R$  of  $H$  to  $\ell^2(\mathbb{N}, \mathbb{C}^L)$ . The operator  $H$  is often also called a *Laurent operator*,  $H^R$  a Toeplitz operator and  $H_N$  a finite Toeplitz matrix, also a compression or section of  $H$ .

---

*Mathematics Subject Classification 2020:* 19K56 (primary); 15B05, 81Q99 (secondary).

*Keywords:* banded Toeplitz matrices, transfer matrices, topological bound states.

There is an extensive mathematical literature, including the books [8, 9, 15, 17, 26, 27, 38], about the spectra of Toeplitz operators and matrices. These sources mainly deal with general (fully occupied) Toeplitz operators with focus on the scalar case  $L = 1$ , but also cover some results on the block case  $L \geq 2$ . Reference [8] provides a detailed study of banded scalar Toeplitz matrices, a case that can be cast in the tridiagonal block form (1.1) and is hence recovered by the present analysis (see Example 3 below). Topological non-Hermitian systems, however, can only be described by the general matrix-valued case  $L \geq 2$ . The asymptotic spectral theory of  $H_N$  with  $L \geq 2$  was studied fifty years ago by Widom in the breakthrough work [42], but there seem to be only few further developments since (let us mention Delvaux's work [13] though; also remarkable is Widom's work on Szego asymptotics of the Toeplitz determinants [43] with numerous follow-ups nicely described in [2], but this does not seem to be directly relevant for the asymptotic spectral theory).

This work provides an elementary approach based on transfer matrix methods which are well known for selfadjoint block Jacobi matrices. For the scalar case  $L = 1$ , there are countless contributions by both the physics and mathematics community, starting with work of Dyson, Schmidt, and others in the 1950's. The block Jacobi case  $L \geq 2$  is heavily used in the solid state physics community where it is at the basis for numerical studies of Anderson localization, see the review [21]. This is tightly connected to the field of products of (symplectic) random matrices; see, e.g., [10]. Important elements of the mathematical theory (restricted to the Hermitian case) are matrix-valued Sturm–Liouville-type oscillation theory [34] and Weyl theory [35] (these works contain references to various earlier contributions, in particular, from the Russian school). These techniques also transpose to block CMV matrices [12] and the slightly more general scattering zippers [25]. As to transfer matrix techniques for the non-Hermitian case (1.1), there is the work [16] for the scalar case  $L = 1$ , but for  $L \geq 2$ , we are only aware of the non-rigorous works of the physics community [11, 20, 22–24, 28, 46].

In this work transfer matrix methods are implemented for non-Hermitian block Toeplitz matrices. They lead to new self-contained proofs of Widom's main results on asymptotic spectra [42], albeit under somewhat different hypotheses. The arguments also invoke Widom's determinant formula (which is re-derived below), but it is then used to explicitly construct eigenfunctions of the finite-volume operators  $H_N$  rather than appealing to potential theory. Along the way, it is also shown that non-Hermitian Bloch theory is essentially a rediscovery of Widom's results by several theoretical physicists [22, 45, 46]. Furthermore, a new criterion for topological zero-energy eigenvalues (bound states) is derived for Hamiltonians having a chiral symmetry. This is rooted in an interplay of Widom's results with index-theoretic statements (essentially only the well-known Noether–Gohberg–Krein index theorem). In particular, this allows to show that such chiral Hamiltonians can have topological

approximate zero modes (namely eigenvalues close to 0) without having a line-gap, a situation that to the best of our knowledge is not covered by the physics literature. Summing up, we hope that this work opens up an alternative perspective for mathematicians interested in the spectral theory of Toeplitz matrices, and on the other hand brings some of the existing deep mathematical results to the attention of the theoretical physics community.

Let us now start with a short overview of the main results. Crucial is the following hypothesis, that is assumed throughout the whole text.

**Hypothesis A.** Both  $T$  and  $R$  are invertible matrices.

It allows for the construction of the  $2L \times 2L$  transfer matrix at complex energy  $E \in \mathbb{C}$  by

$$\mathcal{T}^E = \begin{pmatrix} (E\mathbf{1} - V)T^{-1} & -R \\ T^{-1} & 0 \end{pmatrix}, \quad (1.2)$$

and implies that  $\mathcal{T}^E$  is invertible. In Section 2, it is recalled how the transfer matrix allows for the construction of solutions of the eigenvalue equations for  $H$ ,  $H^R$ , and  $H_N$ . Of crucial importance are the eigenvalues  $z_1(E), \dots, z_{2L}(E)$  of  $\mathcal{T}^E$ , ordered such that  $|z_l(E)| \leq |z_{l+1}(E)|$  for  $l = 1, \dots, 2L - 1$ . Degeneracies are collected in the set

$$\mathcal{F} = \{E \in \mathbb{C} : \mathcal{T}^E \text{ has a degenerate eigenvalue}\}. \quad (1.3)$$

**Hypothesis B.** The set  $\mathcal{F}$  is assumed to be finite.

The spectra of  $\mathcal{T}^E$  furthermore allow for the introduction of the following two subsets of  $\mathbb{C}$ :

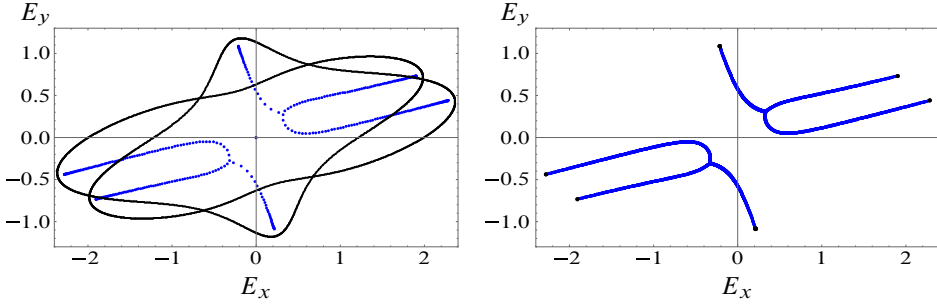
$$\Sigma = \{E \in \mathbb{C} : \mathcal{T}^E \text{ has at least one eigenvalue of modulus 1}\}, \quad (1.4)$$

$$\Lambda = \{E \in \mathbb{C} : \text{two middle eigenvalues } z_L(E), z_{L+1}(E) \text{ of } \mathcal{T}^E \\ \text{have the same modulus}\}. \quad (1.5)$$

It is relatively elementary to see that  $\Sigma = \text{spec}(H)$  is precisely the spectrum of the Laurent operator  $H$ . Theorem 1 below will show that  $\Lambda$  is the attractor of the bulk states of  $H_N$ , under suitable further hypotheses. Other than for Hermitian Hamiltonians, the two sets  $\Sigma$  and  $\Lambda$  are typically very different. This is illustrated in the numerical example in Figure 1.

**Hypothesis C.**  $|z_{L-1}(E)| < |z_L(E)| = |z_{L+1}(E)| < |z_{L+2}(E)|$  for all but finitely many  $E \in \partial\Lambda$ .

Here  $\partial\Lambda$  denotes the boundary of  $\Lambda$ . It is shown in Sections 4.1 and 4.6 that Hypothesis C implies that  $\Lambda = \partial\Lambda$  and that it is given by finite union of analytic arcs. To formulate the last hypothesis, consider the sets  $I_0 = \{1, \dots, L\}$  and



Plot of the spectrum of the symbol  
and the finite-volume Hamiltonian

Plot of the sets  $\Lambda$  and  $\mathcal{F}$

**Figure 1.** Plots for the same model and parameters described in Figures 4 and 5 below, except  $s = 1.4$ . On the left, the black curve shows  $\Sigma$  and in blue are the eigenvalues  $E = E_x + iE_y$ ,  $E_x, E_y \in \mathbb{R}$ , of  $H_N$  for  $N = 200$ . The right plot shows the set  $\Lambda$ . There is a twice degenerate approximate zero mode visible in the left plot, as predicted by Theorem 2.

$I_1 = \{1, \dots, L-1, L+1\}$  and let  $\mathcal{R}_{I_j}^E$  denote the Riesz projection of  $\mathcal{T}^E$  onto the  $L$  eigenvalues  $z_l(E)$  with  $l \in I_j$  where  $j = 0, 1$ . (It is irrelevant in the sequel that the definition is ambiguous for  $E \in \Lambda$ ; the notation  $\mathcal{R}_{I_j}^E$  instead of  $\mathcal{R}_j^E$  may seem awkward and clumsy at this point, but fits nicely in the general framework developed below.) Then introduce

$$q_{I_j}(E) = \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_{I_j}^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right), \quad E \in \mathbb{C} \setminus \mathcal{F}.$$

**Hypothesis D.** The functions  $q_{I_0}$  and  $q_{I_1}$  are not constantly zero locally near points in  $\partial\Lambda \setminus \mathcal{F}$ .

Intuitively, a zero of  $q_{I_j}$  at  $E$  indicates that eigenstates constructed as in Section 2.3 and lying in the range of  $\mathcal{R}_{I_j}^E$  satisfy the Dirichlet boundary conditions on the left and right end. Finally, let us introduce the discrete set

$$\Gamma = \{E \in \mathbb{C} \setminus \Lambda : q_{I_0}(E) = 0\}. \quad (1.6)$$

It is referred to as the set of spectral outliers. The next theorem is the main result about spectral asymptotics of the  $H_N$ , essentially contained in [42] because it is shown in Section 4 that Hypothesis C implies [42, Conditions A and B].

**Theorem 1.** *Suppose that Hypotheses A–D hold. Then, for all  $E \in \partial\Lambda \cup \Gamma$  but a finite number of  $E$ , there exist  $E_N \in \text{spec}(H_N)$  such that  $\lim_{N \rightarrow \infty} E_N = E$ .*

Section 4.6 contains a more precise formulation of this result. It also provides an example of a Hamiltonian for which Hypotheses C and D are violated and the set  $\Lambda$  has non-empty interior. Such non-generic models typically have some symmetries leading to degeneracies in the spectrum of  $\mathcal{T}^E$ . Figure 1 illustrates Theorem 1. It contains numerical plots for a Hamiltonian  $H$  with  $L = 2$  that is described in detail in Section 5.4. Here the focus is on the qualitative aspects. The continuous curve in the left plot is the spectrum  $\Sigma = \text{spec}(H)$  of the periodic operator, and the dots show the eigenvalues of  $H_N$  for  $N = 200$  computed with the Mathematica eigenvalue solver (there are thus 400 points, counting multiplicities). The right plot shows that set  $\Lambda$  computed from the  $4 \times 4$  transfer matrix by implementing (1.5), together with bigger dots showing the points in  $\mathcal{F}$  (at the ends of the arcs). One readily observes the good agreement of  $\Lambda$  with  $\text{spec}(H_N)$  as predicted by Theorem 1. Further examples are dealt with in Section 5 which also contains a discussion of the skin effect and the generalized Brillouin zone.

A more careful inspection of Figure 1 also shows that there is an eigenvalue of  $H_N$  close to 0, actually there are two such eigenvalues. These are the approximate zero modes advertised above. In fact, the Hamiltonian in question has a chiral symmetry. For even  $L$ , such a symmetry is implemented by the matrix  $K = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}$  with blocks of size  $\frac{L}{2}$ . Then  $H$  is said to have a *chiral symmetry* if  $KRK = -R$ ,  $KVK = -V$ , and  $KTK = -T$ . The symmetry immediately implies that  $\Sigma = -\Sigma$ ,  $\Lambda = -\Lambda$ , and  $\Gamma = -\Gamma$ , which is also visible in Figure 1. Furthermore, one can readily check that  $R$ ,  $V$ , and  $T$  are block off-diagonal in the grading of  $K$  and we will denote the upper right entries by  $R_+$ ,  $V_+$ , and  $T_+$  and the lower left ones by  $R_-$ ,  $V_-$ , and  $T_-$ . The following result gives a criterion for  $0 \in \Gamma$  involving two winding numbers associated to merely these latter six matrices (in physics jargon, bulk invariants). It is proved in Section 4.4. By Theorem 1, this then implies that  $H_N$  has an eigenvalue near 0, hence establishing a bulk-boundary correspondence.

**Theorem 2.** *Let  $H$  be chiral and satisfy Hypothesis A. Further let  $W_{\pm}(s) \in \mathbb{Z}$  denote the winding numbers of  $z \in \mathbb{T} \mapsto \det((sz)^{-1}R_{\pm} + V_{\pm} + (sz)T_{\pm})$ . If  $W_+(s) = -W_-(s) \neq 0$  for some  $s > 0$ , then  $0 \in \Gamma$ .*

To apply the criterion, one first has to search for an  $s > 0$  such that  $W_+(s) = -W_-(s)$ , and then check that this number is non-vanishing. This is carried out in Section 5.4. The criterion in Theorem 2 differs from that in other works mentioned above; in particular, also the one in the supplementary material of [28]. Let us also stress that  $\Sigma = \text{spec}(H)$  may encircle 0 so that, again in physics jargon, the system has a point gap, but no line gap. Therefore, standard index-theoretic arguments assuring the existence of zero modes [20, 31] do not apply in this situation. Once available, these zero modes can potentially be used for technological innovations, e.g., trapped light modes [29] or topological sensors [11].

The paper is organized as follows. In Section 2 transfer matrices and their spectra are introduced. Section 3 is essentially index-theoretic, but also already establishes links to the Riesz projections. Then in Section 4 the sets  $\Lambda$  and  $\Gamma$  are analyzed in detail and Theorems 1 and 2 are proven. Finally, Section 5 offers a physics perspective on these results and provides various numerical examples.

## 2. Set-up and basic properties

In this section, many of the objects are introduced and some basic properties are described.

### 2.1. Hamiltonians and their symbols

Throughout this text, the non-Hermitian Hamiltonian  $H$  on the Hilbert space  $\ell^2(\mathbb{Z}, \mathbb{C}^L)$  is the Laurent operator given by (1.1). It is hence specified by the three  $L \times L$  matrices  $R$ ,  $V$ , and  $T$ . The operator  $H$  is selfadjoint if and only if  $V^* = V$  and  $R = T^*$ . The finite-volume approximation of  $H$  on  $\ell^2(\{1, \dots, N\}, \mathbb{C}^L)$  will be denoted by  $H_N \in \mathbb{C}^{NL \times NL}$  (in the mathematics literature,  $H_N$  is called a *section* or a *compression* of  $H$ ). Moreover,  $H^R$  and  $H^L$  are the right and left half-space restrictions of  $H$  to the half-line Hilbert spaces  $\ell^2(\mathbb{N}, \mathbb{C}^L)$  and  $\ell^2(\mathbb{N}_{\leq}, \mathbb{C}^L)$  respectively. The operators  $H^R$  and  $H^L$  are called *Toeplitz operators* in the mathematics literature. Note that  $\ell^2(\mathbb{Z}, \mathbb{C}^L) = \ell^2(\mathbb{N}_{\leq}, \mathbb{C}^L) \oplus \ell^2(\mathbb{N}, \mathbb{C}^L)$  and that one has

$$H - H^L \oplus H^R = |1\rangle R \langle 0| + |0\rangle T \langle 1|, \quad (2.1)$$

where the Dirac bra-ket notation was used for partial isometries  $|n\rangle: \mathbb{C}^L \rightarrow \ell^2(\mathbb{Z}, \mathbb{C}^L)$  onto the matrix degrees of freedom over the lattice sites  $n \in \mathbb{Z}$ . In particular, (2.1) implies that  $H^L \oplus H^R$  is a compact perturbation of  $H$ . The next example shows under which conditions banded scalar Laurent and Toeplitz operators fit into the present framework.

**Example 3.** A Laurent operator with scalar entries and finite band width can be cast into the form (1.1) and satisfy Hypothesis A, provided it is balanced with as many upper as lower diagonals. More precisely, let  $t_{-L}, t_{1-L}, \dots, t_{L-1}, t_L \in \mathbb{C}$  with  $t_{\pm L} \neq 0$  and consider the Hamiltonian  $H$  on  $\ell^2(\mathbb{Z})$  with entries  $H_{l,k} = \langle l|H|k\rangle = \chi(|l-k| \leq L) t_{l-k}$  where  $\chi$  denotes the indicator function. Then set

$$R = \begin{pmatrix} t_L & t_{L-1} & \cdots & t_1 \\ 0 & t_L & & \vdots \\ \vdots & & \ddots & t_{L-1} \\ 0 & \cdots & 0 & t_L \end{pmatrix},$$

$$V = \begin{pmatrix} t_0 & t_{-1} & \cdots & t_{-L+1} \\ t_1 & t_0 & & \vdots \\ \vdots & & \ddots & t_{-1} \\ t_{L-1} & \cdots & t_1 & t_0 \end{pmatrix},$$

$$T = \begin{pmatrix} t_{-L} & 0 & \cdots & 0 \\ t_{1-L} & t_{-L} & & \vdots \\ \vdots & & \ddots & 0 \\ t_{-1} & \cdots & t_{1-L} & t_{-L} \end{pmatrix}.$$

Hence, clearly  $R$  and  $T$  are invertible due to  $t_{\pm L} \neq 0$ .

**Example 4.** A special class of periodic scalar banded operators  $H$  can also be dealt with using Hamiltonians of the form (1.1). Suppose that the matrix entries  $H_{l,k} = \langle l|H|k \rangle$  satisfy  $H_{l+p,k+p} = H_{l,k}$  for all  $k, l \in \mathbb{Z}$  and for an integer  $p$  called the *period*. If, moreover,  $H_{l,k} = 0$  for  $|k - l| > p$ , that is, the band width is  $2p + 1$ , then one can set

$$R = \begin{pmatrix} H_{1,-p+1} & H_{1,-p+2} & \cdots & H_{1,0} \\ 0 & H_{2,-p+2} & & \vdots \\ \vdots & & \ddots & H_{p-1,0} \\ 0 & \cdots & 0 & H_{p,0} \end{pmatrix},$$

$$T = \begin{pmatrix} H_{1,p+1} & 0 & \cdots & 0 \\ H_{2,p+1} & H_{2,p+2} & & \vdots \\ \vdots & & \ddots & 0 \\ H_{p,p+1} & \cdots & H_{p,2p} & H_{p,2p+1} \end{pmatrix}.$$

If furthermore  $V$  is simply the compression of  $H$  onto the sites  $\{1, \dots, p\}$ , then  $H$  coincides with (1.1). If all the diagonal entries of  $R$  and  $T$  are non-vanishing, then Hypothesis A is satisfied and the transfer matrix methods can be used. Note that the entries  $H_{l,k}$  may also be square matrices instead of scalars, without altering the reasoning above. This example may seem somewhat particular because the band width and periodicity have to coincide. It is, however, possible to deal with more general periodic banded matrices by applying the reduced transfer matrices similarly to [14, 22, 32, 37]. This will be developed elsewhere.

The symbol of  $H$  is a matrix-valued meromorphic function given by

$$H(z) = Rz^{-1} + V + Tz, \quad (2.2)$$

and furthermore the associated symbol is defined by

$$\tilde{H}(z) = H(z^{-1}). \quad (2.3)$$

Let us stress that these definitions differ from a large part of the literature where  $H$  and  $\tilde{H}$  are exchanged [8, 9, 13, 17, 42, 43]. The reason for the modification is that it leads to nice consistent notations in connection with the transfer matrix, most notably in (2.12) and (2.13) below. The associated symbol also specifies an operator  $\tilde{H}$  which is defined as in (1.1), but with  $R$  and  $T$  exchanged. One then also has corresponding half-space operators  $\tilde{H}^R$  and  $\tilde{H}^L$ , as well as finite-volume restrictions  $\tilde{H}_N$ . Using a spatial flip implemented by a selfadjoint unitary, one sees that  $\tilde{H}$  and  $\tilde{H}_N$  are unitarily equivalent to  $H$  and  $H_N$  respectively. This does not apply to the half-sided operator though. Indeed, consider the unitary  $W: \ell^2(\mathbb{N}_{\leq}, \mathbb{C}^L) \rightarrow \ell^2(\mathbb{N}, \mathbb{C}^L)$  given by  $W|_{-n} = |n+1\rangle$  for  $n \geq 0$ , for which one then has

$$WH^LW^* = \tilde{H}^R, \quad WH^RW^* = \tilde{H}^L. \quad (2.4)$$

Also the adjoint  $H^*$  will be of relevance and its associated symbol will be denoted by

$$\hat{H}(z) = T^*z^{-1} + V^* + R^*z.$$

For selfadjoint  $H$ , one has  $\hat{H}(z) = H(z)$ . For the half-sided operators, one then checks

$$\hat{H}^R = (H^R)^* = (H^*)^R, \quad \hat{H}^L = (H^L)^* = (H^*)^L. \quad (2.5)$$

Of crucial importance will be the resolvent of the symbol

$$E \in \mathbb{C} \mapsto (H(z) - E\mathbf{1})^{-1}$$

and the associated symbol

$$E \in \mathbb{C} \mapsto (\tilde{H}(z) - E\mathbf{1})^{-1}$$

which are meromorphic functions in  $z$  and  $E$  with values in  $\mathbb{C}^{L \times L}$ . The determinants

$$D(z, E) = \det_L(H(z) - E\mathbf{1}), \quad \tilde{D}(z, E) = \det_L(\tilde{H}(z) - E\mathbf{1}) \quad (2.6)$$

can be factorized as

$$D(z, E) = z^{-L}d(z, E), \quad \tilde{D}(z, E) = z^{-L}\tilde{d}(z, E),$$

where

$$d(z, E) = \det_L(zH(z) - zE\mathbf{1}), \quad \tilde{d}(z, E) = \det_L(z\tilde{H}(z) - zE\mathbf{1}),$$

are polynomials in  $z$  of degree  $2L$ .



## 2.2. Spectra of periodic (Laurent) operators

The spectrum of the periodic operator  $H$  can readily be computed by the discrete Fourier transform  $\mathcal{F}: \ell^2(\mathbb{Z}) \rightarrow L^2(\mathbb{T})$  where  $\mathbb{T} \subset \mathbb{C}$  is the unit circle equipped with the Lebesgue measure, trivially extended to the vector-valued case. It shows that the spectrum  $\Sigma = \text{spec}(H)$  of  $H$  satisfies

$$\Sigma = \bigcup_{z \in \mathbb{T}} \text{spec}(H(z)). \quad (2.7)$$

As  $z \in \mathbb{T} \mapsto H(z) \in \mathbb{C}^{L \times L}$  is a real analytic function (with an analytic extension to a neighborhood of  $\mathbb{T}$  in  $\mathbb{C}$ ), analytic perturbation theory implies that all eigenvalues of  $H(z)$  are real analytic in  $z$  except at level crossings, where the branches can be chosen analytic in the  $L$ th root (Puiseux expansion) [19]. In conclusion,  $\text{spec}(H)$  is given by a set of closed curves in  $\mathbb{C}$ .

## 2.3. Transfer matrices

Another standard element of the theory of Hermitian block Jacobi operators is the transfer matrix. It allows for the construction of the formal solutions of the Schrödinger equation  $H\psi = E\psi$  for a complex energy  $E$  and a state  $\psi = (\psi_n)_{n \in \mathbb{Z}}$  with  $\psi_n \in \mathbb{C}^L$  which is not necessarily in the Hilbert space. For the periodic operator  $H$  given in (1.1), the Schrödinger equation is

$$T\psi_{n+1} + V\psi_n + R\psi_{n-1} = E\psi_n, \quad n \in \mathbb{Z}.$$

As  $T$  is invertible (see Hypothesis A), this can be rewritten as

$$\begin{pmatrix} T\psi_{n+1} \\ \psi_n \end{pmatrix} = \begin{pmatrix} (E\mathbf{1} - V)T^{-1} & -R \\ T^{-1} & 0 \end{pmatrix} \begin{pmatrix} T\psi_n \\ \psi_{n-1} \end{pmatrix}, \quad (2.8)$$

where the second line of this equation is tautological. Therefore, one introduces the transfer matrix at complex energy  $E$  by (1.2) which can also be factorized as

$$\mathcal{J}^E = \begin{pmatrix} E\mathbf{1} - V & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \begin{pmatrix} T^{-1} & 0 \\ 0 & R \end{pmatrix}.$$

Then one can iterate (2.8) to obtain

$$\begin{pmatrix} T\psi_{n+1} \\ \psi_n \end{pmatrix} = (\mathcal{J}^E)^n \begin{pmatrix} T\psi_1 \\ \psi_0 \end{pmatrix}. \quad (2.9)$$

If  $T$  is not invertible, one can construct a reduced transfer matrix of size  $2r$  where  $r$  is the rank of  $T$ , see [22]. For sake of simplicity, this is not carried out here,

but let us stress that this allows to analyze many more Laurent and Toeplitz operators of finite band width by transfer matrix techniques. Note that  $\det_{2L}(\mathcal{T}^E) = \det_L(T)^{-1} \det_L(R)$ , so that the transfer matrix is invertible for all  $E \in \mathbb{C}$  as also  $R$  is invertible (see Hypothesis A). The inverse of  $\mathcal{T}^E$  is explicitly given by

$$(\mathcal{T}^E)^{-1} = \begin{pmatrix} T & 0 \\ 0 & R^{-1} \end{pmatrix} \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & E\mathbf{1} - V \end{pmatrix} = \begin{pmatrix} 0 & T \\ -R^{-1} & R^{-1}(E\mathbf{1} - V) \end{pmatrix}.$$

Associated to  $\tilde{H}(z)$  and  $\hat{H}(z)$ , there are two further naturally associated transfer matrices:

$$\tilde{\mathcal{T}}^E = \begin{pmatrix} (E\mathbf{1} - V)R^{-1} & -T \\ R^{-1} & 0 \end{pmatrix}, \quad \hat{\mathcal{T}}^E = \begin{pmatrix} (E\mathbf{1} - V^*)(R^*)^{-1} & -T^* \\ (R^*)^{-1} & 0 \end{pmatrix}.$$

Straightforward algebra shows

$$(\hat{\mathcal{T}}^{\bar{E}})^* \mathcal{I} \mathcal{T}^E = \mathcal{I}, \quad \mathcal{I} = \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}. \quad (2.10)$$

Note that this implies  $(\mathcal{T}^E)^{-1} = \mathcal{I}^* (\hat{\mathcal{T}}^{\bar{E}})^* \mathcal{I}$ . If the Hamiltonian is Hermitian and the energy  $E \in \mathbb{R}$  is real, the transfer matrix thus satisfies the  $\mathcal{I}$ -unitary relation  $\mathcal{T}^* \mathcal{I} \mathcal{T} = \mathcal{I}$ . The above identity is hence the non-Hermitian analogue of the  $\mathcal{I}$ -unitarity relation. There is yet another identity that will be used below, namely one readily checks

$$\begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix}^{-1} \tilde{\mathcal{T}}^E \begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix} = (\mathcal{T}^E)^{-1}. \quad (2.11)$$

Of great importance will be the resolvent of the transfer matrix. It can be computed by the Schur complement formula:

$$\begin{aligned} (\mathcal{T}^E - z\mathbf{1})^{-1} &= \begin{pmatrix} (E\mathbf{1} - V)T^{-1} - z\mathbf{1} & -R \\ T^{-1} & -z\mathbf{1} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} S^{-1} & -z^{-1}S^{-1}R \\ z^{-1}T^{-1}S^{-1} & -z^{-1}\mathbf{1} - z^{-2}T^{-1}S^{-1}R \end{pmatrix} \end{aligned}$$

where

$$S = (E\mathbf{1} - V)T^{-1} - z\mathbf{1} - z^{-1}RT^{-1} = (E\mathbf{1} - H(z))T^{-1}.$$

Replacing the latter condition shows

$$\begin{aligned} &(\mathcal{T}^E - z\mathbf{1})^{-1} \\ &= \begin{pmatrix} T & 0 \\ 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} (E\mathbf{1} - H(z))^{-1} & -z^{-1}(E\mathbf{1} - H(z))^{-1} \\ z^{-1}(E\mathbf{1} - H(z))^{-1} & -z^{-1}R^{-1} - z^{-2}(E\mathbf{1} - H(z))^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & R \end{pmatrix}. \end{aligned} \quad (2.12)$$

Next let us compute the characteristic polynomial of the transfer matrix. Recalling the identity

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(D) \det(A - BD^{-1}C),$$

one finds

$$\begin{aligned} \det_{2L}(\mathcal{T}^E - z\mathbf{1}) &= (-z)^L \det_L((E - V)T^{-1} - z\mathbf{1} - z^{-1}RT^{-1}) \\ &= z^L \det_L(T)^{-1} \det_L(V - E\mathbf{1} + zT + z^{-1}R) \\ &= z^L \det_L(T)^{-1} \det_L(H(z) - E\mathbf{1}) \\ &= z^L \det_L(T)^{-1} D(z, E) \\ &= \det_L(T)^{-1} d(z, E). \end{aligned} \quad (2.13)$$

This identity also follows by taking the determinant of (2.12). This shows that  $\Sigma = \text{spec}(H)$  is indeed given by both (1.4) and (2.7). Together with (2.11), one also deduces

$$\det_{2L}(\mathcal{T}^E - z\mathbf{1}) = \det_{2L}((\tilde{\mathcal{T}}^E)^{-1} - z\mathbf{1}). \quad (2.14)$$

#### 2.4. Spectral properties of transfer matrices

Let  $z_1(E), \dots, z_{2L}(E)$  be the  $2L$  eigenvalues of  $\mathcal{T}^E$ , listed with their multiplicity and ordered such that  $|z_l(E)| \leq |z_{l+1}(E)|$ . As they are the zeros of characteristic polynomial and  $\mathcal{T}^E$  is invertible, it follows from (2.13) that they are also the zeros of  $z \mapsto D(z, E)$  (also of  $z \mapsto d(z, E)$ ). Similarly, the eigenvalues  $\tilde{z}_1(E), \dots, \tilde{z}_{2L}(E)$  of  $\tilde{\mathcal{T}}^E$ , ordered such that  $|\tilde{z}_l(E)| \leq |\tilde{z}_{l+1}(E)|$ , are the zeros of  $z \mapsto \tilde{D}(z, E)$ . From (2.14), one then deduces the equalities

$$\tilde{z}_k(E) = \frac{1}{z_{2L+1-k}(E)}. \quad (2.15)$$

In particular,

$$\text{spec}(\tilde{\mathcal{T}}^E) = \text{spec}(\mathcal{T}^E)^{-1}, \quad (2.16)$$

which also follows directly from (2.11). Moreover, it follows from identity (2.10) that

$$\text{spec}(\mathcal{T}^E)^{-1} = \overline{\text{spec}(\widehat{\mathcal{T}}^{\bar{E}})}.$$

If the Hamiltonian is Hermitian and the energy  $E \in \mathbb{R}$  is real, then the spectrum satisfies  $\text{spec}(\mathcal{T}^E)^{-1} = \overline{\text{spec}(\mathcal{T}^E)}$ . At several instances it will be useful to diagonalize  $\mathcal{T}^E$ . Of course, this is not possible for all  $E$  as, in general, there may be Jordan blocks

at level crossings. If all Jordan blocks are trivial, let the basis change  $\mathcal{M}^E$  be chosen such that

$$(\mathcal{M}^E)^{-1} \mathcal{T}^E \mathcal{M}^E = \text{diag}(z_1(E), \dots, z_{2L}(E)) = \text{diag}(\tilde{z}_{2L}(E)^{-1}, \dots, \tilde{z}_1(E)^{-1}). \quad (2.17)$$

Recall that  $\mathcal{F} \subset \mathbb{C}$  defined in (1.3) consists of the set of  $E \in \mathbb{C}$  for which  $\mathcal{T}^E$  has a degenerate eigenvalue. Away from the set  $\mathcal{F}$ , diagonalization is always possible with a matrix  $\mathcal{M}^E$  which is unique up to normalization (right multiplication by an invertible diagonal matrix).

**Lemma 5.**  $\mathcal{F}$  is either finite or  $\mathcal{F} = \mathbb{C}$ .

*Proof.* Recall that the eigenvalues of  $\mathcal{T}^E$  are precisely the zeros of the polynomial  $p_E(z) = d(z, E) = \det_L(z(H(z) - E\mathbf{1}))$ . Further recall that the polynomial  $p_E$  has zeros of order greater than one if and only if the resultant  $r_E$  of  $p_E$  and its derivative  $p'_E$  is zero. The resultant is the determinant of the Sylvester matrix and hence a polynomial expression in the coefficients of the polynomials  $p_E$  and its derivative  $p'_E$  which in turn are polynomials in  $E$ . Therefore, the resultant  $r_E$  is also polynomial in  $E$ . It is thus either zero for all or just for a finite set of  $E$ . ■

Hypothesis B states that we only consider the first of the two cases in Lemma 5.

**Lemma 6.** Let  $z_i(E)$  and  $z_j(E)$  be two continuous choices of eigenvalues of  $\mathcal{T}^E$ . If  $|z_i(E)| = |z_j(E)|$  for all  $E$  in some open subset of  $\mathbb{C}$ , then there exists a scalar  $c \in \mathbb{T}$  such that  $z_i(E) = cz_j(E)$  for all  $E \in \mathbb{C}$ .

*Proof.* Note that the  $z_i(E)$  are the zeros of the polynomial  $p_E(z) = \det_L(z(H(z) - E\mathbf{1}))$ . Since zeros of polynomials are continuous in the coefficients of the polynomial and the coefficients of  $p_E(z)$  are polynomial in  $E$ , the  $z_i(E)$  can indeed be chosen continuous in  $E$ . Moreover, for such a choice the  $z_i(E)$  are well known to be holomorphic in  $E$  on  $\mathbb{C} \setminus V$ , where  $V$  is some branch cut [19].

Let now  $U \subset \mathbb{C}$  be the open subset on which  $|z_i(E)| = |z_j(E)|$ . Then  $z_i(E)/z_j(E)$  is holomorphic on  $\mathbb{C} \setminus V$  and has constant modulus equal to 1 on the open non-empty set  $U \setminus V$ . By the maximum modulus principle, it follows that  $z_i(E)/z_j(E)$  is constant on  $\mathbb{C} \setminus V$ ; notably there is a scalar  $c \in \mathbb{T}$  such that  $z_i(E) = cz_j(E)$  for all  $E \in \mathbb{C} \setminus V$ . By continuity, this also extends to  $V$ . ■

### 3. Spectra of half-sided (Toeplitz) operators

This section collects results on the spectrum of the half-sided (Toeplitz) operators  $H^R$  and  $H^L$ . This involves two technical elements: the Noether–Gohberg–Krein index theorem and the study of eigenvalues using the intersection theory of the boundary conditions with the contracting directions of the transfer matrices.

### 3.1. Winding number and index theorem

According to (2.1), the direct sum  $H^R \oplus H^L$  differs from  $H$  only by a compact operator. Hence,  $\Sigma = \text{spec}(H) \subset \text{spec}(H^R) \cup \text{spec}(H^L)$  by a Weyl sequence argument (which is feasible because  $\Sigma = \text{spec}(H)$  has empty interior). Furthermore, a direct consequence of analytic Fredholm theory (see, e.g., [36, Appendix D]) shows that the components of  $\mathbb{C} \setminus \Sigma$  either only contain discrete spectrum of  $H^R \oplus H^L$  or each point in  $\mathbb{C} \setminus \Sigma$  is an eigenvalue of  $H^R \oplus H^L$  (each of these eigenvalues are eigenvalues of either  $H^R$  or  $H^L$ , but not necessarily both). Which one of the two cases applies can be read off the winding number of the determinant of the symbol shifted by a complex energy  $E \in \mathbb{C} \setminus \Sigma$ , more precisely defined by

$$\text{Wind}^E(H) = \int_{\mathbb{T}} \frac{dz}{2\pi i} \partial_z \log(\det_L(H(z) - E\mathbf{1})).$$

Due to (2.13) and (2.12), this is also given by

$$\begin{aligned} \text{Wind}^E(H) &= \int_{\mathbb{T}} \frac{dz}{2\pi i} \partial_z \log(z^{-L} \det_L(R) \det_{2L}(\mathcal{T}^E - z\mathbf{1})) \\ &= -L + \sum_{|z|<1} m_z(\mathcal{T}^E), \end{aligned} \quad (3.1)$$

where  $m_z(\mathcal{T})$  denotes the algebraic multiplicity of  $z$  as eigenvalue of  $\mathcal{T}$  so that the sums are actually finite. Indeed, a classical theorem (due to Noether, Gohberg, and Krein and holding even for continuous symbols, see, e.g., [9, Theorem 6.5] where there is a different sign due to the altered definition (2.2) of the symbol) is the following.

**Theorem 7.** *For all  $E \in \mathbb{C}$ ,*

$$H^R - E\mathbf{1} \text{ Fredholm} \iff \det_L(H(z) - E\mathbf{1}) \neq 0 \text{ for all } z \in \mathbb{T}, \quad (3.2)$$

*and, if this holds, its index is given by*

$$\text{Ind}(H^R - E\mathbf{1}) = \text{Wind}^E(H). \quad (3.3)$$

This shows that  $\Sigma = \text{spec}(H)$  is actually equal to the essential spectrum  $\text{spec}_{\text{ess}}(H^R)$  of  $H^R$  (which by definition consists of all  $E$  such that  $H^R - E\mathbf{1}$  is a Fredholm operator, so we do *not* use Weyl's notion of essential spectrum here). Moreover,  $\Sigma = \text{spec}(H)$  consists of curves in  $\mathbb{C}$  that split  $\mathbb{C}$  in several connected components; picking a point  $E$  in one of these components, one has a winding number  $\text{Wind}^E(H)$ ; if it is non-zero, then the whole component of  $\mathbb{C} \setminus \Sigma$  belongs to the spectrum of  $H^R$ ; if the winding vanishes, one has to analyze the spectrum in that

component by other means; the unbounded component of  $\mathbb{C} \setminus \Sigma$  will never be in the essential spectrum of  $H^R$  though. Thus, one gets the following.

**Corollary 8.** *One has*

$$\text{spec}(H^R) \supset \Sigma \cup \{E \in \mathbb{C} \setminus \Sigma : \text{Wind}^E(H) \neq 0\}.$$

The same arguments apply to  $\tilde{H}^R = WH^LW^*$ . Moreover, (3.2) and (2.15) also imply

$$H^R - E\mathbf{1} \text{ Fredholm} \iff \tilde{H}^R - E\mathbf{1} \text{ Fredholm} \iff H^L - E\mathbf{1} \text{ Fredholm}. \quad (3.4)$$

As

$$0 = \text{Ind}(H - E\mathbf{1}) = \text{Ind}(H^R - E\mathbf{1}) + \text{Ind}(H^L - E\mathbf{1}) \quad (3.5)$$

by (2.1), one has

$$\text{spec}(H^R) \cap \text{spec}(H^L) \supset \Sigma \cup \{E \in \mathbb{C} \setminus \Sigma : \text{Wind}^E(H) \neq 0\}.$$

The discrete spectra of  $H^R$  and  $H^L$  lie outside of this set and are, in general, not the same. It is one of the objects of this work to study these so-called *outliers* of the spectrum. Under suitable conditions, zero energy outliers will be shown to be stable and of topological nature.

### 3.2. Uniform point spectrum

The spectrum of half-sided (Toeplitz) operators  $H^R$  and  $H^L$  can have open subsets consisting entirely of eigenvalues, and we will refer to such regions as *uniform point spectrum*. This is well known to occur for the example of the half-sided shift operator on  $\ell^2(\mathbb{N})$ , and the arguments transpose to other cases. Here several mathematical perspectives on the phenomenon are offered. First of all, the index theorem can also be used to determine eigenvalues for the half-sided operators  $H^R$  and  $H^L$ . Indeed, for  $E \notin \Sigma = \text{spec}(H)$ , Theorem 7 implies

$$\text{Wind}^E(H) = \dim(\text{Ker}(H^R - E\mathbf{1})) - \dim(\text{Ker}((H^R)^* - \bar{E}\mathbf{1})) \quad (3.6)$$

$$= \dim(\text{Ker}((H^L)^* - \bar{E}\mathbf{1})) - \dim(\text{Ker}(H^L - E\mathbf{1})), \quad (3.7)$$

where the first equality follows directly from (3.3) and the definition of the index and the second one, moreover, uses (2.4) as well as the important fact

$$\text{Wind}^E(H) = -\text{Wind}^E(\tilde{H}). \quad (3.8)$$

For non-vanishing winding number, (3.6) and (3.7) imply a dichotomy (with defect).

**Proposition 9.** *If  $\text{Wind}^E(H) > 0$ , then  $\text{Ker}(H^R - E\mathbf{1}) \neq \{0\}$  and  $E$  is an eigenvalue of  $H^R$ . If  $\text{Wind}^E(H) < 0$ , then  $\text{Ker}(H^L - E\mathbf{1}) \neq \{0\}$  and  $E$  is an eigenvalue of  $H^L$ .*

Note that the winding number is locally constant on  $\mathbb{C} \setminus \Sigma$  and only changes on  $\Sigma$ . As, moreover,  $\text{spec}((H^R)^*) = \overline{\text{spec}(H^R)}$ , a connected component of  $\mathbb{C} \setminus \Sigma$  belongs entirely to the spectrum of  $H^R$  if the winding number is non-vanishing in this component. The same statements hold for  $H^L$ . Note also that these facts align exactly with analytic Fredholm theory as described in [36].

Another perspective on (3.6) and (3.7) (actually only a disguise of the above) invokes the bulk-boundary correspondence [31] for an associated Hermitian Hamiltonian given by

$$h^E = \begin{pmatrix} 0 & H - E\mathbf{1} \\ H^* - \bar{E}\mathbf{1} & 0 \end{pmatrix}. \quad (3.9)$$

By construction,  $h^E$  has chiral symmetry; namely,  $Jh^E J = -h^E$ , where  $J = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}$  is the third Pauli matrix in the grading of (3.9). It is possible to write  $h^E$  again in the form (1.1) with blocks of doubled size  $2L$ . The off-diagonal entry  $H - E\mathbf{1}$  has a winding number  $\text{Wind}^E(H)$  which in this context is also called the *first Chern number*  $\text{Ch}_1(h^E)$  of the chiral Hamiltonian  $h^E$ , see [31]. The bulk-boundary correspondence for one-dimensional chiral systems (see, e.g., [31], here applied to a periodic and not merely covariant system) states that the so-called *bulk invariant*  $\text{Wind}^E(H)$  is connected to boundary states via

$$\text{Wind}^E(H) = -\text{Sig}(J|_{\text{Ker}(h^{E,R})}). \quad (3.10)$$

Here  $\text{Sig}$  denotes the signature of a quadratic form, given by the restriction of  $J$  to the subspace  $\text{Ker}(h^{E,R})$  where  $h^{E,R}$  is the half-space restriction of  $h^E$  on  $\ell^2(\mathbb{N}, \mathbb{C}^{2L})$  (with Dirichlet boundary conditions). Because  $(H^*)^R = (H^R)^*$  by (2.5), one thus deduces

$$h^{E,R} = \begin{pmatrix} 0 & H^R - E\mathbf{1} \\ (H^R)^* - \bar{E}\mathbf{1} & 0 \end{pmatrix},$$

and using this in (3.10) then directly leads to (3.6). The second identity (3.7) then follows from  $\text{Wind}^E(\tilde{H}) = \text{Sig}(J|_{\text{Ker}(\tilde{h}^{E,R})})$  combined with  $\tilde{H}^R = W^* H^L W$ , see (2.4).

Yet another proof of Proposition 9 is rooted in the spectral analysis of the transfer matrix, and this is hence more in line with the remainder of the paper. Let us first focus on  $H^R$ . It has Dirichlet boundary conditions at 0. Then the geometric multiplicity of  $E$  as eigenvalue of  $H^R$  is given by

$$\dim(\text{Ker}(H^R - E\mathbf{1})) = \dim(\mathcal{E}_r^E \cap \mathcal{D}_0), \quad (3.11)$$

where  $\mathcal{E}_r^E \subset \mathbb{C}^{2L}$  is the span of the eigenvectors of  $\mathcal{T}^E$  with modulus strictly less than  $r$  and  $\mathcal{D}_0 \subset \mathbb{C}^{2L}$  is the range of  $\begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}$ , namely fixing the Dirichlet boundary conditions

at 0. Indeed, the Schrödinger equation  $H^R \psi = E \psi$  for  $\psi = (\psi_n)_{n \geq 1}$  becomes on the first site  $V \psi_1 + T \psi_2 = E \psi_1$  which can be rewritten as  $\mathcal{T}^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} T = \mathcal{T}^E \begin{pmatrix} T \\ 0 \end{pmatrix} = \begin{pmatrix} E \mathbf{1} \\ \mathbf{1}^{-V} \end{pmatrix}$  just as in (2.8). Hence, if  $\begin{pmatrix} v \\ 0 \end{pmatrix} \in \mathcal{E}_1^E \cap \mathcal{D}_0$ , then  $(\mathcal{T}^E)^n \begin{pmatrix} v \\ 0 \end{pmatrix}$  is exponentially decreasing and  $\psi_n = \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* (\mathcal{T}^E)^n \begin{pmatrix} v \\ 0 \end{pmatrix}$  is thus a square integrable eigenstate of  $H^R$  with energy  $E$ . Similar statements hold for the eigenvalues of  $H^L$ . Now, as  $\dim(\mathcal{D}_0) = L$ , the intersection  $\mathcal{E}_1^E \cap \mathcal{D}_0$  is always non-trivial if  $\dim(\mathcal{E}_1^E) > L$ . However, according to (3.1), one has

$$\dim(\mathcal{E}_1^E) = L + \text{Wind}^E(H). \quad (3.12)$$

Hence, if  $\text{Wind}^E(H) > 0$ , then  $\dim(\mathcal{E}_1^E \cap \mathcal{D}_0) \geq 1$  and  $E$  is an eigenvalue of  $H^R$ . In a similar manner, one can show that  $\text{Wind}^E(H) < 0$  implies that  $E$  is an eigenvalue of  $H^L$ .

### 3.3. Outliers of half-sided operators

This section is about so-called *spectral outliers* of  $H^R$  and  $H^L$ , namely eigenvalues  $E$  in regions where the winding number vanishes, but the intersection in (3.11) is nevertheless non-trivial. It is hence of interest to introduce the associated (skew) Riesz projections

$$\mathcal{R}_r^E = \oint_{\mathbb{T}_r} \frac{dz}{2\pi i} (z \mathbf{1} - \mathcal{T}^E)^{-1}, \quad (3.13)$$

provided that  $\mathbb{T}_r \cap \text{spec}(\mathcal{T}^E) = \emptyset$ . Then  $\mathcal{E}_r^E = \text{Ran}(\mathcal{R}_r^E)$  still for  $\mathbb{T}_r \cap \text{spec}(\mathcal{T}^E) = \emptyset$ , and

$$\text{rk}(\mathcal{R}_r^E) = \dim(\text{Ran}(\mathcal{R}_r^E)) = \sum_{|z| < r} m_z(\mathcal{T}^E).$$

Moreover, using the basis change  $\mathcal{M}^E$  as in (2.17),

$$\mathcal{R}_r^E = \mathcal{M}^E \begin{pmatrix} \mathbf{1} & 0 \\ 0 & 0 \end{pmatrix} (\mathcal{M}^E)^{-1},$$

where  $\mathbf{1}$  is of the size  $\text{rk}(\mathcal{R}_r^E)$ . Changing variables, one also has

$$\mathcal{R}_r^E = \oint_{\mathbb{T}} \frac{dz}{2\pi i} (z \mathbf{1} - r^{-1} \mathcal{T}^E)^{-1}, \quad (3.14)$$

To further compute  $\mathcal{R}_r^E$ , let us replace (2.12) for the resolvent of  $\mathcal{T}^E$ . One finds

$$\mathcal{R}_r^E = \begin{pmatrix} T & 0 \\ 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} Q_r^{(0)}(E) & -Q_r^{(1)}(E) \\ Q_r^{(1)}(E) & R^{-1} - Q_r^{(2)}(E) \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & R \end{pmatrix}, \quad (3.15)$$



where

$$Q_r^{(j)}(E) = \oint_{\mathbb{T}_r} \frac{dz}{2\pi i z^j} (H(z) - E\mathbf{1})^{-1}, \quad j \in \mathbb{Z}. \quad (3.16)$$

Then

$$Q_r^{(1)}(E) = \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_r^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = -T^{-1} \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}^* \mathcal{R}_r^E \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix} R^{-1},$$

and similarly for the other entries. Changing variables,  $Q_r^{(j)}(E)$  can also be rewritten as

$$Q_r^{(j)}(E) = r^{1-j} \oint_{\mathbb{T}} \frac{dz}{2\pi i z^j} (H(rz) - E\mathbf{1})^{-1}, \quad j \in \mathbb{Z}.$$

There are similar objects  $\tilde{\mathcal{R}}_r^E$ ,  $\tilde{Q}_r^{(j)}(E)$ , etc., constructed from  $\tilde{H}$  instead of  $H$ , e.g.,

$$\tilde{Q}_r^{(j)}(E) = \oint_{\mathbb{T}_r} \frac{dz}{2\pi i z^j} (\tilde{H}(z) - E\mathbf{1})^{-1}, \quad j \in \mathbb{Z}, \mathbb{T}_r \cap \text{spec}(\tilde{\mathcal{T}}^E) = \emptyset.$$

As above,

$$\tilde{\mathcal{R}}_r^E = \begin{pmatrix} R & 0 \\ 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} \tilde{Q}_r^{(0)}(E) & -\tilde{Q}_r^{(1)}(E) \\ \tilde{Q}_r^{(1)}(E) & T^{-1} - \tilde{Q}_r^{(2)}(E) \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 \\ 0 & T \end{pmatrix}.$$

A change of variable  $z \mapsto z^{-1}$  also shows that

$$\tilde{Q}_r^{(j)}(E) = Q_{r^{-1}}^{(2-j)}(E). \quad (3.17)$$

In particular, for  $j = 1$  and  $r = 1$ , one has

$$\tilde{Q}_1^{(1)}(E) = Q_1^{(1)}(E).$$

Let us now consider  $H^L$ . As  $WH^LW^* = \tilde{H}^R$  with the unitary reflection  $W$ , one has

$$\dim(\text{Ker}(H^L - E\mathbf{1})) = \dim(\text{Ker}(\tilde{H}^R - E\mathbf{1})) = \dim(\tilde{\mathcal{E}}_1^E \cap \mathcal{D}_0), \quad (3.18)$$

where now  $\tilde{\mathcal{E}}_r^E = \text{Ran}(\tilde{\mathcal{R}}_r^E)$ . Using (2.11), one also finds

$$\begin{aligned} \mathbf{1} - \tilde{\mathcal{R}}_r^E &= \oint_{\mathbb{T}_r} \frac{dz}{2\pi i} (z^{-1}\mathbf{1} - (z\mathbf{1} - \tilde{\mathcal{T}}^E)^{-1}) \\ &= \oint_{\mathbb{T}_r} \frac{dz}{2\pi i} z^{-2} (z^{-1}\mathbf{1} - (\tilde{\mathcal{T}}^E)^{-1})^{-1} \end{aligned}$$

$$\begin{aligned}
&= \oint_{\mathbb{T}_{r^{-1}}} \frac{dz}{2\pi i} (z\mathbf{1} - (\tilde{\mathcal{T}}^E)^{-1})^{-1} \\
&= \oint_{\mathbb{T}_{r^{-1}}} \frac{dz}{2\pi i} \begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix} (z\mathbf{1} - \mathcal{T}^E)^{-1} \begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix} \mathcal{R}_{r^{-1}}^E \begin{pmatrix} 0 & R \\ T^{-1} & 0 \end{pmatrix}^{-1}.
\end{aligned}$$

Actually, this identity also follows from (3.17) replaced into the representation (3.15), or even directly from (2.11). Equivalently,

$$\mathbf{1} - \mathcal{R}_r^E = \begin{pmatrix} 0 & T \\ R^{-1} & 0 \end{pmatrix} \tilde{\mathcal{R}}_{r^{-1}}^E \begin{pmatrix} 0 & T \\ R^{-1} & 0 \end{pmatrix}^{-1}.$$

From this, one deduces

$$\text{Ran}(\mathbf{1} - \mathcal{R}_r^E) = \begin{pmatrix} 0 & T \\ R^{-1} & 0 \end{pmatrix} \text{Ran}(\tilde{\mathcal{R}}_{r^{-1}}^E). \quad (3.19)$$

By essentially the same arguments, one finds from (2.10) that

$$\mathcal{I}^*(\mathcal{R}_r^E)^*\mathcal{I} = \mathbf{1} - \hat{\mathcal{R}}_r^{\bar{E}},$$

with  $\hat{\mathcal{R}}_r^{\bar{E}}$  defined as in (3.13) with  $\mathcal{T}^E$  replaced by  $\tilde{\mathcal{T}}^E$ .

**Lemma 10.** *Recall that  $\mathcal{D}_0$  is the range of  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ . For all  $r > 0$  with  $\mathbb{T}_r \cap \text{spec}(\mathcal{T}^E) = \emptyset$ ,*

$$\text{Ran}(\tilde{\mathcal{R}}_{r^{-1}}^E) \cap \mathcal{D}_0 = \text{Ran}(\mathbf{1} - \mathcal{R}_r^E) \cap \mathcal{D}_0.$$

*Proof.* Let  $v \in \mathbb{C}^N$  be such that

$$\begin{pmatrix} Rv \\ 0 \end{pmatrix} \in \text{Ran}(\tilde{\mathcal{R}}_{r^{-1}}^E) \cap \mathcal{D}_0.$$

Due to (3.19), one then has

$$\begin{pmatrix} 0 \\ v \end{pmatrix} = \begin{pmatrix} 0 & T \\ R^{-1} & 0 \end{pmatrix} \begin{pmatrix} Rv \\ 0 \end{pmatrix} \in \text{Ran}(\mathbf{1} - \mathcal{R}_r^E).$$

As the latter subspace is a direct sum of generalized eigenspaces of  $\mathcal{T}^E$ , it is invariant under  $\mathcal{T}^E$ . Therefore,

$$\mathcal{T}^E \begin{pmatrix} 0 \\ v \end{pmatrix} = \begin{pmatrix} -Rv \\ 0 \end{pmatrix} \in \text{Ran}(\mathbf{1} - \mathcal{R}_r^E).$$

As the latter vector is clearly also in  $\mathcal{D}_0$ , one concludes that  $\text{Ran}(\tilde{\mathcal{R}}_{r^{-1}}^E) \cap \mathcal{D}_0 \subset \text{Ran}(\mathbf{1} - \mathcal{R}_r^E) \cap \mathcal{D}_0$ . The inverse inclusion is shown in the same manner. ■

**Proposition 11.** *Suppose that  $\text{spec}(\mathcal{T}^E) \cap \mathbb{T} = \emptyset$ , or equivalently  $E \notin \Sigma$ . Then  $E$  is an eigenvalue of  $H^R$  or  $H^L$  if and only if  $\text{Ker}(Q_1^{(1)}(E)) \neq \{0\}$ . More precisely,*

$$\dim(\text{Ker}(Q_1^{(1)}(E))) = \dim(\text{Ker}(H^R - E\mathbf{1})) + \dim(\text{Ker}(H^L - E\mathbf{1})).$$

*Proof.* Suppose  $E$  is an eigenvalue of  $H^R$  with (non-zero and square integrable) eigenvector  $\psi = (\psi_n)_{n \geq 1}$ . Then  $\psi$  has to decay, i.e., the vector  $\begin{pmatrix} T\psi_{n+1} \\ \psi_n \end{pmatrix}$  lies in the decaying subspace  $\mathcal{E}_1^E$  of  $\mathcal{T}^E$ . In particular,

$$\begin{pmatrix} T\psi_1 \\ 0 \end{pmatrix} \in \text{Ran}(\mathcal{R}_1^E) \cap \mathcal{D}_0.$$

Moreover, since  $\mathcal{R}_1^E$  is an idempotent, one has

$$\begin{pmatrix} T\psi_1 \\ 0 \end{pmatrix} = \mathcal{R}_1^E \begin{pmatrix} T\psi_1 \\ 0 \end{pmatrix} = \begin{pmatrix} TQ_1^{(0)}(E)T\psi_1 \\ Q_1^{(1)}(E)T\psi_1 \end{pmatrix}.$$

Thus, indeed  $T\psi_1 \in \text{Ker}(Q_1^{(1)}(E))$  and  $T\psi_1$  is non-zero, since otherwise successive application of the transfer matrix would imply  $\psi = 0$ . Similarly, supposing  $E$  is an eigenvalue of  $\tilde{H}^R$  with eigenvector  $\tilde{\psi}$ , one has

$$\begin{pmatrix} R\tilde{\psi}_1 \\ 0 \end{pmatrix} \in \text{Ran}(\tilde{\mathcal{R}}_1^E) \cap \mathcal{D}_0$$

and

$$\begin{pmatrix} R\tilde{\psi}_1 \\ 0 \end{pmatrix} = \tilde{\mathcal{R}}_1^E \begin{pmatrix} R\tilde{\psi}_1 \\ 0 \end{pmatrix} = \begin{pmatrix} R\tilde{Q}_1^{(0)}(E)R\tilde{\psi}_1 \\ \tilde{Q}_1^{(1)}(E)R\tilde{\psi}_1 \end{pmatrix} = \begin{pmatrix} RQ_1^{(2)}(E)R\tilde{\psi}_1 \\ Q_1^{(1)}(E)R\tilde{\psi}_1 \end{pmatrix}.$$

Hence,  $0 \neq R\tilde{\psi}_1 \in \text{Ker}(Q_1^{(1)}(E))$ . Moreover,  $\text{Ran}(\tilde{\mathcal{R}}_1^E) \cap \mathcal{D}_0 = \text{Ran}(\mathbf{1} - \mathcal{R}_1^E) \cap \mathcal{D}_0$  by Lemma 10. As  $\text{Ran}(\mathcal{R}_1^E)$  and  $\text{Ran}(\mathbf{1} - \mathcal{R}_1^E)$  are linearly independent, it follows that also  $\begin{pmatrix} T\psi_1 \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} R\tilde{\psi}_1 \\ 0 \end{pmatrix}$  are linearly independent. As both first components lie in  $\text{Ker}(Q_1^{(1)}(E))$ , one concludes together with (3.11) and (3.18) that

$$\dim(\text{Ker}(Q_1^{(1)}(E))) \geq \dim(\text{Ker}(H^R - E\mathbf{1})) + \dim(\text{Ker}(H^L - E\mathbf{1})).$$

Now, let  $0 \neq v \in \text{Ker}(Q_1^{(1)}(E))$ . Then

$$\mathcal{R}_1^E \begin{pmatrix} v \\ 0 \end{pmatrix} = \begin{pmatrix} TQ_1^{(0)}(E)v \\ Q_1^{(1)}(E)v \end{pmatrix} = \begin{pmatrix} TQ_1^{(0)}(E)v \\ 0 \end{pmatrix} \in \text{Ran}(\mathcal{R}_1^E) \cap \mathcal{D}_0$$

and similarly

$$(\mathbf{1} - \mathcal{R}_1^E) \begin{pmatrix} v \\ 0 \end{pmatrix} = \begin{pmatrix} (\mathbf{1} - TQ_1^{(0)}(E))v \\ 0 \end{pmatrix} \in \text{Ran}(\mathbf{1} - \mathcal{R}_1^E) \cap \mathcal{D}_0 = \text{Ran}(\tilde{\mathcal{R}}_1^E) \cap \mathcal{D}_0,$$

where the last claim follows from Lemma 10. As at least one of the last two vectors is non-vanishing, one concludes that  $E$  is an eigenvalue of  $H^R$  or of  $\tilde{H}^R$  due to (3.11) and (3.18). Moreover, if  $\mathcal{S}^E = (\text{Ker}(Q_1^{(1)}(E))) \subset \mathcal{D}_0$ , then

$$\begin{aligned} \dim(\text{Ker}(Q_1^{(1)}(E))) &= \dim(\text{Ran}(\mathcal{R}_1^E) \cap \mathcal{S}^E) + \dim(\text{Ran}(\mathbf{1} - \mathcal{R}_1^E) \cap \mathcal{S}^E) \\ &\leq \dim(\text{Ran}(\mathcal{R}_1^E) \cap \mathcal{D}_0) + \dim(\text{Ran}(\mathbf{1} - \mathcal{R}_1^E) \cap \mathcal{D}_0) \\ &= \dim(\text{Ker}(H^R - E\mathbf{1})) + \dim(\text{Ker}(\tilde{H}^R - E\mathbf{1})), \end{aligned}$$

concluding the proof due to (3.18).  $\blacksquare$

**Corollary 12.** *The sum of the geometric multiplicities of  $E$  as eigenvalue of  $H^R$  or  $H^L$  is at most  $L$ :*

$$\dim(\text{Ker}(H^R - E\mathbf{1})) + \dim(\text{Ker}(H^L - E\mathbf{1})) \leq L.$$

**Example 13.** Let us consider  $L = 1$ , namely a scalar periodic operators with  $T, V, R \in \mathbb{C}$  with  $T, R \neq 0$  (this is the Hatano–Nelson model, see Section 5.1). Then

$$\text{spec}(H) = \{Re^{-ik} + V + Te^{ik} : k \in [-\pi, \pi)\}$$

is the boundary of an open ellipse  $\mathbb{E} \subset \mathbb{C}$ . Using the residue theorem, one has

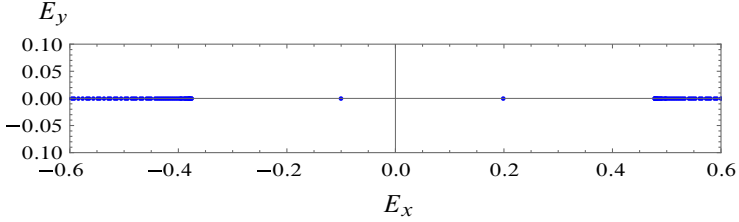
$$Q_1^{(1)}(E) = \frac{1}{2\pi i T} \oint_{\mathbb{T}} dz \frac{1}{z^2 + (V - E)T^{-1}z + RT^{-1}} = \begin{cases} 0, & E \in \mathbb{E}, \\ c_E, & E \in \mathbb{C} \setminus \bar{\mathbb{E}}, \end{cases}$$

with

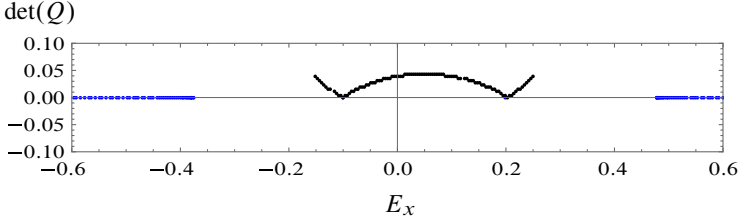
$$c_E = ((E - V)^2 - 4RT)^{-\frac{1}{2}} \neq 0$$

with the root taken using the first branch. Hence, Proposition 11 tells us that for  $E \in \mathbb{E}$  either  $H^R$  or  $H^L$  has an eigenvalue. Of course, as  $|\text{Wind}^E(H)| = 1$  for  $E \in \mathbb{E}$ , this merely reproduces the implication of Theorem 7 (which is actually stronger because it also determines whether  $H^R$  or  $H^L$  must have an eigenvalue).

**Example 14.** Let us consider a selfadjoint Hamiltonian with  $L = 2$ . It is specified by two matrices  $T = R^* = \begin{pmatrix} t_{0,0} & t_{0,1} \\ t_{1,0} & t_{1,1} \end{pmatrix}$  and  $V = \begin{pmatrix} v_{0,0} & v_{0,1} \\ v_{1,0} & v_{1,1} \end{pmatrix} = V^*$ . For suitable parameters inspired by the so-called *SSH model* [31], the Hamiltonian has two bands and two eigenvalues (bound states) in the gap. One can then compute the  $2 \times 2$  matrix  $Q_1^{(1)}(E)$  and its determinant numerically and its zeros lie exactly at these eigenvalues. This is illustrated in Figure 2, and can also be understood using intersection theory of Lagrangian subspaces, see [34].



Plot of the spectrum of the finite-volume Hamiltonian


 Plot of the spectrum and  $\det(Q)$ 

**Figure 2.** Plot of the central part of the spectrum of the selfadjoint Hamiltonian described in Example 14 with parameters are  $t_{0,1} = 1.0$ ,  $t_{1,0} = 0.1$ ,  $t_{0,0} = t_{1,1} = 0$  and  $v_{0,0} = 0.2$ ,  $v_{0,1} = 0.5t = \overline{v_{1,0}}$ ,  $v_{1,1} = -0.1$ . The bottom plot also shows the value of  $|\det(Q_1^{(1)}(E))|$  as a function of the  $E \in \mathbb{R}$ . One clearly sees the two zeros, in accordance with Proposition 11.

### 3.4. Topological zero modes of half-sided chiral operators

This section first defines what a chiral symmetry is and provides some structural information that it induces on the objects introduced so far. It then shows how index-theoretic arguments imply the existence of zero eigenvalues for suitable half-sided (Toeplitz) operators with chiral symmetry. Throughout, the dimension  $L$  will be supposed to be even. Then on each fiber there is selfadjoint unitary of vanishing signature given by  $K = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ . Its fiberwise extension  $\mathbf{1} \otimes K$  will also be denoted by  $K$ . The chiral symmetry of  $H$  then reads

$$KHK = -H. \quad (3.20)$$

Due to the chiral symmetry (3.20), it follows that  $\Sigma = \text{spec}(H) = -\Sigma$ . Moreover, one deduces  $KVK = -V$ ,  $KTK = -T$  and  $KRK = -R$ . Hence, each of these matrices is off-diagonal in the grading of  $K$ . Let us introduce the notations

$$V = \begin{pmatrix} 0 & V_+ \\ V_- & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & T_+ \\ T_- & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & R_+ \\ R_- & 0 \end{pmatrix}.$$

It is also natural and useful to introduce a notation for the off-diagonal entries of the symbol:

$$H(z) = \begin{pmatrix} 0 & H_+(z) \\ H_-(z) & 0 \end{pmatrix}, \quad H_{\pm}(z) = R_{\pm}z^{-1} + V_{\pm} + T_{\pm}z, \quad (3.21)$$

and then consider  $H_{\pm}$  also as block operators of the form (1.1), but on the Hilbert space  $\ell^2(\mathbb{Z}, \mathbb{C}^{\frac{k}{2}})$ . As above, one also has  $\tilde{H}_{\pm}(z)$  and  $\hat{H}_{\pm}(z)$ . Next, let us consider the winding numbers. As  $\det(K(H(z) - E\mathbf{1})K) = (-1)^L \det(H(z) + E\mathbf{1})$  and  $L$  is even, one has for  $E \in \mathbb{C} \setminus \Sigma$

$$\text{Wind}^E(H) = \oint_{\mathbb{T}} \frac{dz}{2\pi i} \partial_z \log \det(K(H(z) - E\mathbf{1})K) = \text{Wind}^{-E}(H).$$

Of particular importance is the winding number at zero energy. Due to the off-diagonal form of  $H(z)$ , one has

$$\text{Wind}^0(H) = \text{Wind}^0(H_+) + \text{Wind}^0(H_-). \quad (3.22)$$

These latter two winding numbers imply the following extension of the bulk-boundary correspondence as described in Section 3.2.

**Proposition 15.** *Let  $H$  have the chiral symmetry (3.20). Then  $\text{Wind}^0(H_{\pm}) > 0$  implies that  $\dim(\text{Ker}(H_{\pm}^R)) \geq 1$  and hence also  $\dim(\text{Ker}(H^R)) \geq 1$ . Similarly,  $\text{Wind}^0(H_{\pm}) < 0$  implies  $\dim(\text{Ker}(H_{\pm}^L)) \geq 1$  and  $\dim(\text{Ker}(H^L)) \geq 1$ .*

*Proof.* As in (3.9), one has

$$h^0 = \begin{pmatrix} 0 & 0 & 0 & H_+ \\ 0 & 0 & H_- & 0 \\ 0 & (H_-)^* & 0 & 0 \\ (H_+)^* & 0 & 0 & 0 \end{pmatrix}. \quad (3.23)$$

Now, (3.10) holds and implies that  $\text{Wind}^0(H) = -\text{Sig}(J|_{\text{Ker}(h^0, R)})$ . Next let us recall the identities  $((H_{\pm})^*)^R = \hat{H}_{\pm}^R$  and  $((H_{\pm})^*)^L = \hat{H}_{\pm}^L$ , see (2.5). Together with (3.22), one therefore has

$$\begin{aligned} & \text{Wind}^0(H_+) + \text{Wind}^0(H_-) \\ &= \dim(\text{Ker}(H_-^R)) + \dim(\text{Ker}(H_+^R)) - \dim(\text{Ker}(\hat{H}_+^R)) - \dim(\text{Ker}(\hat{H}_-^R)). \end{aligned}$$

Moreover, using  $J' = \text{diag}(\mathbf{1}, -\mathbf{1}, \mathbf{1}, -\mathbf{1})$  in the grading of (3.23), the selfadjoint Hamiltonian  $h^0$  also has the further chiral symmetry  $J'h^0J' = -h^0$ . For this symmetry, (3.10) combined with  $\text{Wind}^0((H_-)^*) = -\text{Wind}^0(H_-)$  leads to

$$\begin{aligned} & \text{Wind}^0(H_+) - \text{Wind}^0(H_-) \\ &= \dim(\text{Ker}(\hat{H}_-^R)) + \dim(\text{Ker}(H_+^R)) - \dim(\text{Ker}(\hat{H}_+^R)) - \dim(\text{Ker}(H_-^R)). \end{aligned}$$

Adding and subtracting these identities then leads to the two identities

$$\text{Wind}(H_{\pm}, 0) = \dim(\text{Ker}(H_{\pm}^R)) - \dim(\text{Ker}(\hat{H}_{\pm}^R)),$$

which imply the first claim. For the second one, the same reasoning is applied to  $\tilde{H}$  for which the winding number is given by (3.8). Because of the relations  $WH_{\pm}^L W^* = \tilde{H}_{\pm}^R$  and  $WH_{\pm}^R W^* = \tilde{H}_{\pm}^L$  as in (2.4) hold, one obtains

$$\begin{aligned} & \text{Wind}^0(H_+) + \text{Wind}^0(H_-) \\ &= \dim(\text{Ker}(\hat{H}_+^L)) + \dim(\text{Ker}(\hat{H}_-^L)) - \dim(\text{Ker}(H_+^L)) - \dim(\text{Ker}(H_-^L)) \end{aligned}$$

and

$$\begin{aligned} & \text{Wind}^0(H_+) - \text{Wind}^0(H_-) \\ &= \dim(\text{Ker}(\hat{H}_+^L)) + \dim(\text{Ker}(H_+^L)) - \dim(\text{Ker}(\hat{H}_-^L)) - \dim(\text{Ker}(H_-^L)). \end{aligned}$$

Hence,

$$\text{Wind}^0(H_{\pm}) = \dim(\text{Ker}(\hat{H}_{\pm}^L)) - \dim(\text{Ker}(H_{\mp}^L)),$$

implying the second claim. ■

On first sight, the statements of Proposition 15 and Proposition 9 resemble each other. There is, however, a crucial difference: Proposition 9 is stable under perturbations in  $E$  (because the winding number  $\text{Wind}^E(H)$  is so), while Proposition 15 merely concerns the eigenvalue 0, the so-called *zero modes* or *zero energy bound states*. Let us discuss two distinct scenarios for the application of Proposition 15 (which in physical terminology explained in Section 5 are compatible with line-gapped and point-gapped systems respectively). If  $\text{Wind}^0(H) = 0$ , then the component of  $\mathbb{C} \setminus \Sigma$  containing 0 only contains discrete spectrum (by analytic Fredholm theory) and the zero eigenvalues resulting from Proposition 15 (in the case  $\text{Wind}^0(H_+) \neq 0$ ) are robust under perturbations conserving the chiral symmetry (and are merely moved away from the origin if the perturbations break the chiral symmetry). On the other hand, if  $\text{Wind}^0(H) \neq 0$ , then Proposition 9 implies that the component of  $\mathbb{C} \setminus \Sigma$  containing 0 consists entirely of spectrum of  $H^R$  (or  $H^L$ ) and then the supplementary eigenvalues resulting from Proposition 15 are covered by this spectrum and seem to be of little interest. It will, however, be shown in Section 5.4 that finite-volume approximations do have approximate zero modes resulting from  $\text{Wind}^0(H_+) \neq 0$ . This is based on Widom's theory of asymptotic spectra described in Section 4.

The remainder of this section will provide an alternative proof of Proposition 15 based on transfer matrix methods (similar to the arguments towards the end of Section 3.2, leading to Proposition 9). This is crucially based on the following property

of the transfer matrix of a chiral Hamiltonian:

$$\begin{aligned} (K \otimes \sigma_3) \mathcal{T}^E (K \otimes \sigma_3) &= \begin{pmatrix} K(E - V)T^{-1}K & KRK \\ -KT^{-1}K & 0 \end{pmatrix} \\ &= \begin{pmatrix} -(E + V)T^{-1} & -R \\ T^{-1} & 0 \end{pmatrix} = \mathcal{T}^{-E}. \end{aligned}$$

This implies  $\det_{2L}(\mathcal{T}^E - z\mathbf{1}) = \det_{2L}(\mathcal{T}^{-E} - z\mathbf{1})$  so that  $\text{spec}(\mathcal{T}^E) = \text{spec}(\mathcal{T}^{-E})$ . Furthermore,  $(K \otimes \sigma_3) \mathcal{T}^0 (K \otimes \sigma_3) = \mathcal{T}^0$ , notably the zero energy transfer matrix has invariant subspaces. Indeed, one readily checks

$$\mathcal{T}^0 = \begin{pmatrix} -VT^{-1} & -R \\ T^{-1} & 0 \end{pmatrix} = \begin{pmatrix} -V_+T_+^{-1} & 0 & 0 & -R_+ \\ 0 & -V_-T_-^{-1} & -R_- & 0 \\ 0 & T_-^{-1} & 0 & 0 \\ T_+^{-1} & 0 & 0 & 0 \end{pmatrix} = \mathcal{T}_+^0 \hat{\oplus} \mathcal{T}_-^0,$$

where the direct sum is understood in the obvious way (regrouping the first and fourth columns and lines, as well as the second and third ones) and these  $2 \times 2$  blocks are given by

$$\mathcal{T}_+^0 = \begin{pmatrix} -V_+T_+^{-1} & -R_+ \\ T_+^{-1} & 0 \end{pmatrix}, \quad \mathcal{T}_-^0 = \begin{pmatrix} -V_-T_-^{-1} & -R_- \\ T_-^{-1} & 0 \end{pmatrix}.$$

Note that  $\mathcal{T}_+^0$  and  $\mathcal{T}_-^0$  are the transfer matrices at  $E = 0$  for  $H_+$  and  $H_-$  respectively. In complete analogy to (3.11) for  $E = 0$ , one now has for these two subsystems

$$\dim(\text{Ker}(H_{\pm}^R)) = \dim(\mathcal{E}_{1,\pm}^0 \cap \mathcal{D}_0), \quad (3.24)$$

where  $\mathcal{E}_{1,\pm}^0 \subset \mathbb{C}^L$  is the generalized eigenspace of all eigenvalues of  $\mathcal{T}_{\pm}^0$  with modulus less than 1. Moreover, similarly to (3.12), one has

$$\dim(\mathcal{E}_{1,\pm}^0) = \frac{L}{2} + \text{Wind}^0(H_{\pm}).$$

Hence, if  $\text{Wind}^0(H_{\pm}) > 0$  is positive, then one has  $\dim(\mathcal{E}_{1,\pm}^0) > \frac{L}{2}$  which enforces  $\dim(\mathcal{E}_{1,\pm}^0 \cap \mathcal{D}_0) \geq 1$  so that, due to (3.24), 0 is an eigenvalue of  $H_{\pm}^R$ . This provides the promised alternative proof of Proposition 15.

**Remark 16.** Let us briefly show how Proposition 15 relates to Proposition 11. For  $0 \notin \Sigma = \text{spec}(H)$  the latter result states that

$$\dim(\text{Ker}(Q_1^{(1)}(0))) = \dim(\text{Ker}(H^R)) + \dim(\text{Ker}(H^L)).$$



The matrix-valued function  $E \mapsto Q_r^{(1)}(E)$  satisfies

$$\begin{aligned} K Q_r^{(1)}(E) K &= \oint_{\mathbb{T}_r} \frac{dz}{2\pi i z} K (H(z) - E \mathbf{1})^{-1} K \\ &= \oint_{\mathbb{T}_r} \frac{dz}{2\pi i z} (-H(z) - E \mathbf{1})^{-1} = -Q_r^{(1)}(-E), \end{aligned}$$

implying, in particular, that  $Q_r^{(1)}(0)$  is chiral. It is hence off-diagonal in the grading of  $K$  with two off-diagonal entries  $Q_{r,\pm}^{(1)}(0)$  given by (3.16) from  $H_{\pm}$ :

$$Q_r^{(1)}(0) = \oint_{\mathbb{T}_r} \frac{dz}{2\pi i z} \begin{pmatrix} 0 & H_+(z) \\ H_-(z) & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & Q_{r,-}^{(1)}(0) \\ Q_{r,+}^{(1)}(0) & 0 \end{pmatrix}.$$

One hence has

$$\dim(\text{Ker}(Q_r^{(1)}(0))) = \dim(\text{Ker}(Q_{r,+}^{(1)}(0))) + \dim(\text{Ker}(Q_{r,-}^{(1)}(0))).$$

Furthermore,  $0 \notin \Sigma$  is equivalent to  $\det(H(e^{ik})) = \det(H_+(e^{ik})) \det(H_-(e^{ik})) \neq 0$  for all  $e^{ik} \in \mathbb{T}$ , which is hence also equivalent to  $0 \notin \text{spec}(H_+)$  and  $0 \notin \text{spec}(H_-)$ . Therefore, one can apply Proposition 11 also separately to  $H_{\pm}$ , giving

$$\dim(\text{Ker}(Q_{1,\pm}^{(1)}(0))) = \dim(\text{Ker}(H_{\pm}^R)) + \dim(\text{Ker}(H_{\pm}^L)).$$

Together these equations merely reflect that  $\text{Ker}(H^R) = \text{Ker}(H_+^R) \oplus \text{Ker}(H_-^R)$ .

## 4. Widom's theory of asymptotic spectra revisited

This section is the mathematical core of the paper, providing the proofs of Theorems 1 and 2.

### 4.1. The set $\Lambda$ of energies with middle eigenvalues of equal modulus

Let us first note that the set  $\Sigma = \text{spec}(H)$  can be written as

$$\Sigma = \{E \in \mathbb{C} : |z_l(E)| = 1 \text{ for some } l = 1, \dots, 2L\},$$

see also (1.4). In the Hermitian case in which the transfer matrix is  $\mathcal{I}$ -unitary, the existence of one eigenvalue on the unit circle implies that there is also a second eigenvalue on the unit circle and in the corresponding two-dimensional subspace one can construct eigenfunctions of finite-volume approximations  $H_N$  at nearby energies

(by well-known techniques described in the proof of Proposition 38, or by the procedure in Section 4.8). Hence, the points in  $\Sigma$  are indeed near the spectrum of  $H_N$ . However, for a non-Hermitian  $H$  one has generically only one eigenvalue on the unit circle and the fundamental solutions constructed by the transfer matrices do not allow to construct eigenfunctions for  $H_N$  satisfying both boundary conditions. From this perspective, it is natural to introduce the set  $\Lambda$  of complex energies  $E$  where the middle two eigenvalues  $|z_L(E)| \leq |z_{L+1}(E)|$  of the transfer matrix  $\mathcal{T}^E$  (or equivalently zeros of  $D(z, E) = \det_L(H(z) - E\mathbf{1})$ ) coincide:

$$\Lambda = \{E \in \mathbb{C} : |z_L(E)| = |z_{L+1}(E)|\}, \quad (4.1)$$

see (1.5). (In Widom's work [42, p. 312], this set is denoted by  $C_2$ .) Due to (2.16), one can also write this set as

$$\Lambda = \{E \in \mathbb{C} : \text{middle two eigenvalues of } \tilde{\mathcal{T}}^E \text{ are of same modulus}\}.$$

This section is about proving topological and analytical properties of this set  $\Lambda$ . A first property is the following: if  $E$  is in  $\Lambda$ , then either  $E \in \Sigma$  or  $E \notin \Sigma$ . In the latter case,  $\det(H(e^{ik}) - E\mathbf{1}) \neq 0$  for all  $e^{ik} \in \mathbb{T}$  so that  $\text{Wind}^E(H)$  is well defined and, due to (3.1), non-vanishing. Hence,

$$\Lambda \subset \Sigma \cup \{E \in \mathbb{C} : \text{Wind}^E(H) \neq 0\}. \quad (4.2)$$

**Proposition 17.** *The set  $\Lambda$  satisfies  $\Lambda \subset \text{spec}(H^R)$  and  $\Lambda \subset \text{spec}(H^L)$ .*

*Proof.* Rewriting (4.2) gives together with Theorem 7 that

$$\begin{aligned} \Lambda &\subset \{E \in \mathbb{C} : \det_L(H(z) - E\mathbf{1}) = 0 \text{ for some } z \in \mathbb{S}^1 \text{ or } \text{Wind}^E(H) \neq 0\} \\ &= \{E \in \mathbb{C} : H^R - E\mathbf{1} \text{ is not Fredholm or is Fredholm with } \text{Ind}(H^R - E\mathbf{1}) \neq 0\} \\ &\subset \text{spec}(H^R). \end{aligned}$$

The same holds for  $H^L$ . ■

**Remark 18.** Let us consider the special case of Hermitian  $H$  (namely  $T = R^*$  and  $V = V^*$ ). Then the spectrum of the  $\mathcal{I}$ -unitary matrix  $\mathcal{T}^E$  has a reflection symmetry on  $\mathbb{T}$  and therefore the middle two eigenvalues are of the same modulus if and only if they lie on the unit circle, notably

$$\Lambda = \{E \in \mathbb{C} : \text{spec}(\mathcal{T}^E) \cap \mathbb{T} \neq \emptyset\}.$$

This is a well-known characterization of the spectrum of periodic Hermitian operators, namely  $\Sigma = \Lambda$ . There are close connections to the standard Brillouin zone, see the discussion in Section 5.3 below. Furthermore, the inclusion  $\Lambda \subset \lim_{N \rightarrow \infty} \text{spec}(H_N)$

follows from a standard Weyl sequence argument. For systems without any bound states (also called *outliers*), one has  $\Lambda = \lim_{N \rightarrow \infty} \text{spec}(H_N)$ , namely the set  $\Gamma$  introduced in Section 4.3 is empty. On the other hand, for Hermitian chiral SSH-type topological insulators (of the type described in Section 5.4), it is known that there are spectral outliers at zero energy, similar to the eigenvalues already encountered in Example 14. This paper provides non-Hermitian chiral Hamiltonians with zero energy outliers.

In the remainder of this section, various other properties of the set  $\Lambda$  are proved.

**Lemma 19.** *The set  $\Lambda$  is compact.*

*Proof.* As the  $z_j(E)$  are continuous in  $E$ , the set  $\Lambda$  is closed. Moreover, by Proposition 17,  $\Lambda \subset \text{spec}(H^R)$ . As  $H^R$  is bounded, this also shows that  $\Lambda$  is a bounded set. ■

**Lemma 20.** *The set  $\Lambda$  does not contain isolated points.*

*Proof.* Suppose  $E$  is an isolated point of  $\Lambda \setminus \mathcal{F}$  so that  $|z_L(E)| = |z_{L+1}(E)|$  and, in particular,  $z_L(E) \neq z_{L+1}(E)$ . These two eigenvalues are isolated and hence holomorphic at  $E$ . Let us denote these two holomorphic eigenvalues by  $\zeta_L(E)$  and  $\zeta_{L+1}(E)$ . This eliminates discontinuities of the functions  $z_L(E)$  and  $z_{L+1}(E)$  resulting from labeling according to the modulus. Note that, if there are only two eigenvalues with modulus  $|z_L(E)|$  at  $E$ , then  $\{z_L(E + \varepsilon), z_{L+1}(E + \varepsilon)\} = \{\zeta_L(E + \varepsilon), \zeta_{L+1}(E + \varepsilon)\}$  locally, namely for  $\varepsilon$  small enough. Note now that the labeling of the eigenvalues is determined only up to permutation of the labels among eigenvalues of the same modulus. Since  $E \in \Lambda$  is isolated (because  $\mathcal{F}$  is finite), it follows that one could always have chosen a labeling such that  $|\zeta_L(E + \varepsilon)| \neq |\zeta_{L+1}(E + \varepsilon)|$  for all  $\varepsilon \neq 0$  sufficiently small. Now,  $\varepsilon = 0$  is a local maximum or minimum of the function  $\varepsilon \mapsto |\zeta_L(E + \varepsilon)|/|\zeta_{L+1}(E + \varepsilon)|$ , because  $E$  is isolated in  $\Lambda$  so that the quotient does not take the value 1 in a some pointed neighborhood  $B_\delta(0) \setminus \{0\}$ . But then the maximum modulus principle would imply that  $|\zeta_L(E + \varepsilon)| = |\zeta_{L+1}(E + \varepsilon)|$  for all  $\varepsilon$  in an open neighborhood of 0. Then also  $|z_L(E + \varepsilon)| = |z_{L+1}(E + \varepsilon)|$  for  $\varepsilon \in B_\delta(0)$ , a contradiction to the isolation of  $E$ .

Suppose now  $E \in \Lambda \cap \mathcal{F}$  is an isolated point of  $\Lambda$ . Then not necessarily all of the  $z_l$  are holomorphic in a neighborhood of  $E$ . However, since they are the zeros of a polynomial whose coefficients are polynomials in  $E$ , they can be given by a Puiseux series in  $E$  with non-zero convergence radius. Hence, there exist  $p_{l,E} \in \mathbb{N}$  (compare [19]) such that the functions

$$\varepsilon \mapsto \tau_l^E(\varepsilon) = \zeta_l(E + \varepsilon^{p_{l,E}}) \quad (4.3)$$

are holomorphic in a neighborhood of  $\varepsilon = 0$ , where the  $\zeta_l$  are again a choice of eigenvalue making the  $\tau_l^E$  holomorphic. Repeating the previous argument for the pair of functions  $\varepsilon \mapsto \tau_L^E(\varepsilon^{p_{L+1,E}})$  and  $\varepsilon \mapsto \tau_{L+1}^E(\varepsilon^{p_{L,E}})$  shows that also a point  $E \in \Lambda \cap \mathcal{F}$  cannot be isolated in  $\Lambda$ .  $\blacksquare$

**Lemma 21.** *The set  $\partial\Lambda$  locally consists of a finite union of analytic arcs.*

*Proof.* Let  $E \in \partial\Lambda$ . Then

$$|z_{L-a-1}(E)| < |z_{L-a}(E)| = \cdots = |z_{L+b}(E)| < |z_{L+b+1}(E)|$$

for some integers  $a \geq 0$  and  $b \geq 1$ . Using the function  $\tau_l^E$  defined in (4.3), let us introduce the holomorphic functions

$$\chi_{l,k}^E(\varepsilon) = \frac{\tau_l^E(\varepsilon^{p_{k,E}})}{\tau_k^E(\varepsilon^{p_{l,E}})} = \frac{\zeta_l(E + \varepsilon^{p_{l,E} p_{k,E}})}{\zeta_k(E + \varepsilon^{p_{k,E} p_{l,E}}}.$$

(Instead of the power  $p_{l,E} p_{k,E}$  it is also possible to work with  $q_{l,k} = \text{lcm}(p_{l,E}, p_{k,E})$ , namely the lowest common multiple of  $p_{l,E}$  and  $p_{k,E}$ .) There is an open neighborhood  $V \subset \mathbb{C}$  of 0 such that for all  $\varepsilon \in V$  one has

$$|\tau_1^E(\varepsilon)|, \dots, |\tau_{L-a-1}^E(\varepsilon)| < |\tau_{L-a}^E(\varepsilon)|, \dots, |\tau_{L+b}^E(\varepsilon)| < |\tau_{L+b+1}^E(\varepsilon)|, \dots, |\tau_{2L}^E(\varepsilon)|.$$

Let now  $\mathcal{P} \subset \{L-a, \dots, L+b\}^2$  be the set of pairs  $(l, k)$  of indices  $l < k$  such that  $\chi_{l,k}$  is not locally constant around 0 and  $\zeta_l, \zeta_k$  is the pair of middle eigenvalues somewhere on  $E + V \setminus \{0\}$ . Note that  $\mathcal{P}$  is non-empty, since  $E \in \partial\Lambda$  (because then in some direction of  $\varepsilon$  the middle two eigenvalues need to have different modulus so that the corresponding  $\chi$  cannot be constant). For all  $(l, k) \in \mathcal{P}$ , let us set

$$\gamma_{l,k} = \{\varepsilon \in V : |\chi_{l,k}^E(\varepsilon)| = 1\}.$$

Since all  $\chi_{l,k}^E$  are holomorphic and non-constant around 0, there is an  $n \in \mathbb{N}$  such that

$$\chi_{l,k}^E(\varepsilon) = \chi_{l,k}^E(0)(1 + c_n(E)\varepsilon^n + \mathcal{O}(\varepsilon^{n+1})), \quad (4.4)$$

with  $c_n(E) \neq 0$ . Note that  $E \mapsto c_1(E)$  is analytic on  $\mathbb{C} \setminus \mathcal{F}$  and cannot vanish identically around points  $E \in \partial\Lambda \setminus \mathcal{F}$ , for the following reason. For all  $\varepsilon$  and  $\delta$  small enough,

$$\chi_{l,k}^E(\varepsilon + \delta) = \frac{z_l(E + \varepsilon + \delta)}{z_k(E + \varepsilon + \delta)} = \chi_{l,k}^{E+\delta}(\varepsilon).$$

Since both  $\chi_{l,k}^E$  and  $\chi_{l,k}^{E+\delta}$  are analytic around 0, they have a convergent power series expression there. Comparing the terms of order  $\varepsilon$ , one finds

$$c_1(E + \delta) = \sum_{m>0} m c_m(E) \delta^{m-1}.$$

This is the power series expansion of  $c_1$  around  $E$ . Now, if  $c_1$  would vanish identically around  $E$ , it follows from the expansion that  $c_m(E) = 0$  for all  $m$ . However, this means that  $\chi_{l,k}^E$  has to be constant around 0, which it is not. Therefore,  $c_1$  does not vanish identically around any  $E \in \partial\Lambda \setminus \mathcal{F}$ . Hence, the map  $E \mapsto c_1(E)$  has a discrete set of zeros. (Recall that  $\mathcal{F}$  is discrete by Hypothesis B.) One simply has  $n = 1$  except on a discrete set of points  $E$ , but these points with  $n \geq 2$  are singular points of interest. Namely, (4.4) implies that  $\gamma_{l,k}$  is the union of  $n$  analytic curves in  $V$  going through  $0 \in V$ , possibly after choosing  $V$  sufficiently smaller. Consider now  $W = \bigcap_{(l,k) \in \mathcal{P}} V^{p_{l,E} p_{k,E}}$  where  $V^q = \{\varepsilon^q : \varepsilon \in V\}$ . Then  $E + W$  is a neighborhood of  $E$ . Recalling that  $\tau_l^E(\varepsilon) = \zeta_l(E + \varepsilon^{p_{l,E}})$  with  $p_{l,E} > 0$ , one finds that on the open neighborhood  $E + W$ , the points in  $\partial\Lambda$  are locally near  $E$  precisely given by  $E + W \cap \bigcup_{(l,k) \in \mathcal{P}} (\gamma_{l,k})^{p_{l,E} p_{k,E}}$ . This latter set consists of a finite union of analytic arcs. ■

**Remark 22.** Let us consider a point  $E \in \Lambda \cap \mathcal{F} = \{E \in \Lambda : z_L(E) = z_{L+1}(E)\}$  for which the Puiseux expansion of some of the middle eigenvalues is non-trivial, namely that  $p > 1$  in (4.3). Then necessarily  $\mathcal{T}^E$  has a Jordan block for one of the middle eigenvalues (which is hence degenerate, reflecting again that  $E \in \mathcal{F}$ ). If there are only two middle eigenvalues and the coefficient in (4.4) is  $n = 1$ , then  $\mathcal{P}$  consists of just this one pair and there is just one set  $\gamma$  given by an analytic line. Taking the power the gives a set  $\gamma^{p^2}$ , that is a cusp or half-line. This is observed in numerical examples in Figures 3 and 4 at the extremities of  $\Lambda$ .

**Corollary 23.** *If the interior of  $\Lambda$  is empty, the set  $\Lambda$  locally consists of a finite union of analytic arcs.*

*Proof.* As  $\Lambda$  is closed and the closure is always given by the disjoint union  $\bar{\Lambda} = \partial\Lambda \dot{\cup} \overset{\circ}{\Lambda}$  of the boundary and the interior, it follows from the hypothesis that  $\Lambda = \partial\Lambda$  so that the claim holds by Lemma 21. ■

Let us provide a simple criterion assuring that  $\Lambda$  has empty interior so that the above corollary applies.

**Lemma 24.** *If  $T$  and  $R$  have no eigenvalues of same modulus, the set  $\Lambda \subset \mathbb{C}$  has empty interior. If  $T$  and  $R$  have no double eigenvalues, then Hypothesis B holds.*

*Proof.* This will follow from an asymptotic analysis of the spectrum of the transfer matrices  $\mathcal{T}^E$  and  $\tilde{\mathcal{T}}^E$  at  $|E| \rightarrow \infty$ . For this purpose, let us set  $\xi = \frac{1}{E}$  and  $\mathcal{S}^\xi = \frac{1}{E} \mathcal{T}^E$ . The function  $\xi \in \mathbb{C} \setminus \{0\} \mapsto \mathcal{S}^\xi$  is analytic and has a removable singularity at 0, namely

$$\mathcal{S}^\xi = \begin{pmatrix} T^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \mathcal{O}(\xi).$$

By analytic perturbation theory [19],  $\mathcal{S}^\xi$  has  $L$  non-vanishing eigenvalues  $\mu_1(\xi), \dots, \mu_L(\xi)$  of modulus of order 1, and the others are of order  $\xi$ . More precisely, if  $\Gamma_1, \dots, \Gamma_L$  are the eigenvalues of  $T$ , then for an appropriate choice of the labeling one has  $|\mu_l(\xi) - \frac{1}{\Gamma_l}| = \mathcal{O}(\xi)$  for  $l = 1, \dots, L$ . As  $\text{spec}(\mathcal{T}^E) = E \text{spec}(\mathcal{S}^E)$ , the spectrum of  $\mathcal{T}^E$  has  $L$  eigenvalues  $E\mu_1(\frac{1}{E}), \dots, E\mu_L(\frac{1}{E})$  of order  $E$  and  $L$  eigenvalues of order 1. As  $E\mu_l(\frac{1}{E}) = \frac{E}{\Gamma_l} + \mathcal{O}(1)$  for  $l = 1, \dots, L$ , the hypothesis that all  $\Gamma_l$  have different moduli implies that these  $L$  eigenvalues all have different modulus for  $E$  sufficiently large. Moreover, one can run the same argument for  $\tilde{\mathcal{T}}^E$  based on the hypothesis on the spectrum of  $R$ . Due to (2.16), this shows that none of the eigenvalues of  $\mathcal{T}^E$  have equal modulus for  $E$  sufficiently large. Due to Lemma 6, this implies the first claim. The second follows in a similar manner. ■

## 4.2. Scaling operation

Let us introduce an invertible scaling matrix by

$$S_N = \text{diag}(s, s^2, \dots, s^N), \quad s \neq 0.$$

Then

$$S_N^{-1} H_N S_N = \begin{pmatrix} V & sT & & \\ s^{-1}R & \ddots & \ddots & \\ & \ddots & \ddots & sT \\ & & s^{-1}R & V \end{pmatrix}. \quad (4.5)$$

In particular, for all  $s > 0$ ,

$$\text{spec}(H_N) = \text{spec}(S_N^{-1} H_N S_N).$$

This suggests to introduce the half-sided block-Toeplitz matrix  $H^{s,R}$  with entries as in (4.5). Let us stress that for  $s \neq 1$ , the operators  $H^{s,R}$  and  $H^R$  are not conjugate to each other, and are therefore typically *not* isospectral. The symbol of  $H^{s,R}$  will be denoted by  $H^s(z)$ . Comparing with (2.2), it is hence obtained by scaling

$$H^s(z) = s^{-1}Rz^{-1} + V + sTz = H(sz).$$

Its associated symbol is according to (2.3) given by

$$\tilde{H}^s(z) = H^s\left(\frac{1}{z}\right) = H\left(\frac{s}{z}\right) = \tilde{H}\left(\frac{z}{s}\right) = \tilde{H}^{\frac{1}{s}}(z).$$

Then the corresponding determinants defined by  $D^s(z, E) = \det_L(H^s(z) - E\mathbf{1})$  as in (2.6) satisfy

$$D^s(z, E) = D(sz, E), \quad \tilde{D}^s(z, E) = \tilde{D}(sz, E).$$

Moreover, let  $\mathcal{T}^{s,E}$  and  $\tilde{\mathcal{T}}^{s,E}$  denote the transfer matrices associated to the rescaled Hamiltonian symbols  $H^s$  and  $\tilde{H}^s$ . One finds

$$\mathcal{T}^{s,E} = s^{-1}\mathcal{T}^E, \quad \tilde{\mathcal{T}}^{s,E} = s\tilde{\mathcal{T}}^E. \quad (4.6)$$

Then there are also associated Riesz projections

$$\mathcal{R}_r^{s,E} = \oint_{\mathbb{T}_r} \frac{dz}{2\pi i} (z\mathbf{1} - \mathcal{T}^{s,E})^{-1}.$$

They do not contain supplementary information though because the range and kernel of  $\mathcal{T}^{s,E}$  is independent of  $s$  due to (4.6). Actually, also the change of variable formula (3.14) shows

$$\mathcal{R}_r^{s,E} = \mathcal{R}_{sr}^E.$$

Therefore, Proposition 11 immediately implies the following.

**Corollary 25.** *Suppose that  $\text{spec}(\mathcal{T}^E) \cap \mathbb{T}_s = \emptyset$ . Then*

$$\dim(\text{Ker}(Q_s^{(1)}(E))) = \dim(\text{Ker}(H^{s,R} - E\mathbf{1})) + \dim(\text{Ker}(H^{s,L} - E\mathbf{1})).$$

Let us now start with the application of the scaling to the spectral analysis. The argument of Proposition 17 applies to each  $H^{s,R}$  and therefore

$$\Lambda \subset \bigcap_{s>0} \text{spec}(H^{s,R}).$$

The inverse implication holds for  $L = 1$ , see Corollary 28 below. In the following, it will be shown that a modification is necessary in the matrix-valued case  $L \geq 2$ . Due to (3.1) applied to  $H^s$ , the scaling (4.6) implies that, as long as  $\text{spec}(s\mathcal{T}^E) \cap \mathbb{T}^1 = \emptyset$  or equivalently  $H^{s,R} - E\mathbf{1}$  is a Fredholm operator (see Theorem 7),

$$\text{Ind}(H^{s,R} - E\mathbf{1}) = \text{Wind}^E(H^s) = -L + \sum_{|z|<1} m_z(\mathcal{T}^{s,E}) = -L + \sum_{|z|<s} m_z(\mathcal{T}^E).$$

This implies that one has  $\text{spec}(s\mathcal{T}^E) \cap \mathbb{T}^1 \neq \emptyset$  or  $\text{Wind}^E(H^s) \neq 0$  for all  $s > 0$  if and only if the two middle eigenvalues of  $\mathcal{T}^E$  are of same modulus. One therefore has the characterization

$$\begin{aligned} \Lambda &= \{E \in \mathbb{C} : \text{for all } s > 0, \text{ either } H^{s,R} - E\mathbf{1} \text{ is not Fredholm} \\ &\quad \text{or } \text{Wind}^E(H^s) \neq 0\} \\ &= \{E \in \mathbb{C} : \text{for all } s > 0, \text{ either } H^{s,R} - E\mathbf{1} \text{ is not Fredholm} \\ &\quad \text{or } \text{Ind}(H^{s,R} - E\mathbf{1}) \neq 0\}. \end{aligned} \quad (4.7)$$

One can carry out the same argument with  $H^L$  (or simply use (3.4) and (3.5)) to deduce

$$\Lambda = \{E \in \mathbb{C} : \text{for all } s > 0, \text{ either } H^{s,L} - E\mathbf{1} \text{ is not Fredholm} \\ \text{or } \text{Ind}(H^{s,L} - E\mathbf{1}) \neq 0\}. \quad (4.8)$$

Hence, the set  $\Lambda$  is equivalently characterized by either  $H^R$  or  $H^L$  and their scaling.

Now, let  $E \in \bigcap_{s>0} \text{spec}(H^{s,R})$ . Then, for all  $s > 0$ , one of the following three mutually exclusive cases holds true:

- (i)  $H^{s,R} - E\mathbf{1}$  is not Fredholm;
- (ii)  $H^{s,R} - E\mathbf{1}$  is Fredholm with non-vanishing index;
- (iii)  $H^{s,R} - E\mathbf{1}$  is Fredholm with vanishing index, but  $\text{Ker}(H^{s,R} - E\mathbf{1}) \neq \{0\}$  is non-trivial.

In the last point (iii), one can also replace  $\text{Ker}(H^{s,R} - E\mathbf{1}) \neq \{0\}$  by the statement that  $H^{s,R} - E\mathbf{1}$  is not invertible (or, as matter of fact, that  $\text{Ker}((H^{s,R} - E\mathbf{1})^*) \neq \{0\}$ ). Comparing with (4.7), one sees that  $\Lambda$  contains all  $E \in \bigcap_{s>0} \text{spec}(H^{s,R})$  satisfying either (i) or (ii). It is hence natural to introduce the set of  $E$  for which (iii) holds, which is done in the next section.

### 4.3. The set $\Gamma$ of spectral outliers

As motivated in the end of the last section, let us introduce the set

$$\Gamma^R = \{E \in \mathbb{C} \setminus \Lambda : \text{for all } s > 0 \text{ with } \text{Ind}(H^{s,R} - E\mathbf{1}) = 0, \text{ one has} \\ \text{Ker}(H^{s,R} - E\mathbf{1}) \neq \{0\}\},$$

where the condition

$$\text{Ind}(H^{s,R} - E\mathbf{1}) = 0$$

includes the Fredholm property of  $H^{s,R} - E\mathbf{1}$ . Alternatively,

$$\Gamma^R = \{E \in \mathbb{C} \setminus \Lambda : \text{for all } s > 0 \text{ with } \text{Wind}^E(H^s) = 0, \text{ one has} \\ \text{Ker}(H^{s,R} - E\mathbf{1}) \neq \{0\}\},$$

where again the condition  $\text{Wind}^E(H^s) = 0$  contains that the winding number is well defined. By construction, the discussion at the end of Section 4.2 thus implies

$$\bigcap_{s>0} \text{spec}(H^{s,R}) = \Lambda \cup \Gamma^R. \quad (4.9)$$

Moreover, if the defining property for  $E$  being in  $\Gamma^R$  holds for some  $s$ , it actually holds for all  $s$  in some interval, as shows the proof of the following lemma.



**Lemma 26.**  $\Gamma^R = \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Ind}(H^{s,R} - E\mathbf{1}) = 0 \text{ and } H^{s,R} - E\mathbf{1} \text{ not invertible}\}$ .

*Proof.* By definition,

$$\Gamma^R \subset \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Ind}(H^{s,R} - E\mathbf{1}) = 0 \text{ and } H^{s,R} - E\mathbf{1} \text{ not invertible}\}.$$

Let now  $E$  be an element of the right-hand side and let  $s > 0$  be such that  $H^{s,R} - E\mathbf{1}$  is Fredholm with index zero, but also not invertible. This means that  $H^{s,R} - E\mathbf{1}$  has a non-trivial kernel. This can be reformulated in terms of the  $L$ -dimensional subspace  $\mathcal{E}_1^{s,E} \subset \mathbb{C}^{2L}$  containing all generalized eigenvectors of  $\mathcal{T}^{s,E}$  with eigenvalues of modulus strictly less than 1. In fact, just as in (3.11),

$$\dim(\text{Ker}(H^{s,R} - E\mathbf{1})) = \dim(\mathcal{E}_1^{s,E} \cap \mathcal{D}_0).$$

Now,  $\mathcal{T}^{s,E} = s^{-1}\mathcal{T}^E$  and thus  $\mathcal{E}_1^{s,E}$  is independent of  $s$  as long as no eigenvalue of  $\mathcal{T}^{s,E}$  touches the unit circle (in which case the Fredholm condition is violated). Hence,  $\text{Ker}(H^{s,R} - E\mathbf{1})$  is also independent of  $s$  as long as one has  $s^{-1}|z_L(E)| < 1 < s^{-1}|z_{L+1}(E)|$ , i.e., for  $s \in (|z_L(E)|, |z_{L+1}(E)|)$ . ■

In connection with Lemma 26, let us note that  $E \in \mathbb{C} \setminus \Lambda$ , so  $E \notin \Lambda$ , always implies that there exists  $s$  with  $\text{Ind}(H^{s,R} - E\mathbf{1}) = \text{Wind}^E(H^s) = 0$ . Hence,  $\Gamma^R$  is characterized by the singularity of  $H^{s,R} - E\mathbf{1}$  for such  $s$ . Furthermore, as  $\mathcal{E}_1^{s,E} = \mathcal{E}_s^E$  for  $s \in (|z_L(E)|, |z_{L+1}(E)|)$ , the proof of Lemma 26 also allows to rewrite the definition of the set  $\Gamma^R$  merely in terms of spectral properties of the transfer matrix  $\mathcal{T}^E$ :

$$\Gamma^R = \{E \in \mathbb{C} : |z_L(E)| < |z_{L+1}(E)| \text{ and } \mathcal{E}_s^E \cap \mathcal{D}_0 \neq \emptyset \text{ for } s \in (|z_L(E)|, |z_{L+1}(E)|)\}.$$

From this representation, one immediately deduces the following.

**Proposition 27.** *For  $L = 1$ , one has  $\Gamma^R = \emptyset$ .*

*Proof.* For  $E \in \Gamma^R$  the  $2 \times 2$  transfer matrix  $\mathcal{T}^E$  given in (1.2) has two distinct eigenvalues. For  $s$  satisfying  $|z_1(E)| < s < |z_2(E)|$ , the subspace  $\mathcal{E}_s^E \subset \mathbb{C}^2$  is 1-dimensional and spanned by an eigenvector of  $\mathcal{T}^E$ . For  $E \in \Gamma^R$ , this eigenvector has to be constant multiple of  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ . However,  $\mathcal{T}^E$  has no eigenvector of this form as can be readily seen by looking at the lower component of the eigenvalue equation for  $\mathcal{T}^E$ . ■

Proposition 27 together with (4.9) directly implies the next result which goes back to the work of Schmidt and Spitzer [33] (more precisely, it is the result  $A = C$  therein).

**Corollary 28.** *For  $L = 1$ , one has*

$$\Lambda = \bigcap_{s>0} \text{spec}(H^{s,R}). \quad (4.10)$$

Another well-known result in the case  $L = 1$  states that  $\Lambda$  is connected [40].

**Proposition 29.** *The set  $\Gamma^R$  is discrete in  $\mathbb{C}$ .*

*Proof.* Let us rewrite the definition of  $\Gamma^R$  as

$$\Gamma^R = \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Wind}^E(H^s) = 0 \text{ and} \\ \dim(\text{Ker}(H^{s,R} - E\mathbf{1})) \geq 1\}.$$

The condition  $\text{Wind}^E(H^s) = 0$  is open in  $E$  (and actually also in  $s$ ). On the other hand,  $\dim(\text{Ker}(H^{s,R} - E\mathbf{1})) \geq 1$  implies that  $\dim(\text{Ker}(Q_s^{(1)}(E))) \geq 1$  by Corollary 25, so that

$$\Gamma^R \subset \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Wind}^E(H^s) = 0 \text{ and} \\ \det(Q_s^{(1)}(E)) = 0\}.$$

But  $E \mapsto \det(Q_s^{(1)}(E))$  is analytic (except at points where the Fredholm condition is not satisfied) and not identically 0. Hence, its zeros form a discrete set and this implies that also  $\Gamma^R$  is discrete. ■

Up to now, only the right-sided Toeplitz operators  $H^{s,R}$  were considered. As  $\Lambda$  can also be expressed in terms of the  $H^{s,L}$  by (4.8), let us also introduce

$$\Gamma^L = \{E \in \mathbb{C} \setminus \Lambda : \text{for all } s > 0 \text{ with } \text{Ind}(H^{s,L} - E\mathbf{1}) = 0 \text{ one has} \\ \text{Ker}(H^{s,L} - E\mathbf{1}) \neq \{0\}\}. \quad (4.11)$$

Then all of the above properties directly transpose. In particular,

$$\bigcap_{s>0} \text{spec}(H^{s,L}) = \Lambda \cup \Gamma^L,$$

and the equivalents of Lemma 26 and Propositions 27 and 29 hold. Note that it is possible (and not in contradiction to Proposition 11) that  $\Gamma^R \cap \Gamma^L \neq \emptyset$ , but this is non-generic (it is generic within the set of chiral Hamiltonians though, see Section 4.4). Finally, let us set

$$\Gamma = \Gamma^R \cup \Gamma^L.$$

Then the proof of Proposition 29 combined with Corollary 25 shows that

$$\Gamma = \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Wind}^E(H^s) = 0 \text{ and} \\ \det(Q_s^{(1)}(E)) = 0\}$$

$$= \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } |z_L(E)| < s < |z_{L+1}(E)| \text{ and } \det(Q_s^{(1)}(E)) = 0\}. \quad (4.12)$$

It will be shown in (4.21) below that this coincides with the definition (1.6) given in the introduction. Let us point out that this set  $\Gamma$  also appears in [13, equation (1.14)] where it is denoted  $G_0$ . In Widom's work [42] there is no notation for it, but it appears in [42, Theorem 6.1] as  $\det(Q_s^{(1)})$  is the function denoted by  $E$  in [42]. Widom also shows that  $\det(Q_s^{(1)})$  is equal to a certain Fredholm determinant, but this will not be of relevance in this work.

#### 4.4. Zero as spectral outlier of chiral Hamiltonians

This short section is about the sets  $\Lambda$  and  $\Gamma$  for block Hamiltonians (1.1) satisfying the chiral symmetry (3.20). In particular,  $L$  is even and all the structural properties of the Hamiltonian and the transfer matrix discussed in Section 3.4 hold. Hence,  $\text{spec}(\mathcal{T}^E) = \text{spec}(\mathcal{T}^{-E})$  and this implies that

$$\Lambda = -\Lambda, \quad \Gamma = -\Gamma, \quad (4.13)$$

namely both  $\Lambda$  and  $\Gamma$  are reflection symmetric subsets of  $\mathbb{C}$ . Hence, the discrete set  $\Gamma$  comes in pairs of points  $(E, -E)$ , except for the origin 0 which is the only point that is invariant under reflection. It is hence of crucial interest whether  $0 \in \Gamma$  or  $0 \notin \Gamma$ . The following is Theorem 2.

**Proposition 30.** *Let  $H$  have the chiral symmetry (3.20) with off-diagonal entries  $H_{\pm}$  as in (3.21) and suppose that  $0 \notin \Lambda$ . If  $\text{Wind}^0(H_+^s) = -\text{Wind}^0(H_-^s) \neq 0$  for some  $s > 0$ , then  $0 \in \Gamma$ .*

*Proof.* By Proposition 15, the hypothesis implies that there exists an  $s$  such that both  $\text{Ker}((H^s)^R) \neq \{0\}$  and  $\text{Ker}((H^s)^L) \neq \{0\}$  and, moreover,  $\text{Ind}(H^s) = \text{Wind}^0(H^s) = 0$ . These facts together imply that  $0 \in \Gamma = \Gamma^R \cup \Gamma^L$  by Lemma 26 and (4.11). ■

#### 4.5. Widom's determinant formula

**Theorem 31.** *Let  $E \notin \mathcal{F}$ . Then the characteristic polynomial is given by Widom's formula*

$$\det_{NL}(H_N - E\mathbf{1}) = \sum_I G_I(E)^{N+1} q_I(E) \quad (4.14)$$

where the sum runs over all index sets  $I \subset \{1, \dots, 2L\}$  of cardinality  $L$  and the two functions are defined by

$$G_I(E) = (-1)^L \det(R) \prod_{i \in I} \frac{1}{z_i(E)} \quad (4.15)$$

and

$$q_I(E) = \det_L \left( \oint_{\gamma_I^0} \frac{dz}{2\pi iz} (H(z) - E\mathbf{1})^{-1} \right) = \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right), \quad (4.16)$$

where  $\gamma_I^0$  is a path in  $\mathbb{C}$  which has a winding number 1 around 0 as well as each  $z_i(E)$  for  $i \in I$  and a winding number 0 around all other eigenvalues of  $\mathcal{T}^E$ .

In contradistinction to Widom's work, in the present context it is possible to replace  $\gamma_I^0$  in the definition of  $q_I(E)$  by the path  $\gamma_I$  which has a winding number 1 around each  $z_i(E)$  for  $i \in I$  and a winding number 0 around all other eigenvalues of  $\mathcal{T}^E$  as well as 0. This holds because  $z(H(z) - E\mathbf{1})$  is a matrix-valued polynomial in  $z$  and its eigenvalues are the same as the eigenvalues of  $\mathcal{T}^E$  by (2.13), so that 0 is not an eigenvalue due to the invertibility of  $\mathcal{T}^E$ . Let us also stress that, if  $E$  is such that there are eigenvalues of same modulus, the ordering of the eigenvalues is ambiguous and hence the definition of  $G_I(E)$  and  $q_I(E)$  as well. However, Widom's formula does not inherit this ambiguity because it contains a sum over all index sets  $I$  of cardinality  $L$ . Another way to circumvent the ambiguity is to choose one labeling at a point  $E$  and then to analytically extend it to a neighborhood of  $E$  (which is possible because  $E \notin \mathcal{F}$ ).

Let us stress that Theorem 31 is just a special case of Widom's [42, Theorem 6.2] which also covers the case of non-invertible off-diagonal coefficients. The proof of Theorem 31 given in [42] is based on another formula from [3] and is rather involved. Based on transfer matrix methods, it is here possible to provide a relatively short proof of Widom's formula. It will be shown below that this strategy also leads to a few new identities within the present restricted context.

Let us start out by noting that

$$\det_{NL}(H_N - E\mathbf{1}) = (-1)^{NL} \det_L(T)^N \det_L \left( \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}^* (\mathcal{T}^E)^N \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right). \quad (4.17)$$

Indeed, one can start by constructing all solutions at  $E$  satisfying the Dirichlet boundary conditions at 1 from (2.9) and then check whether any of these solution satisfying the Dirichlet boundary condition at  $N$ . Hence, the multiplicity of the zero of the determinant on the right-hand side is equal to the multiplicity of  $E$  as an eigenvalue of  $H_N$ . Thus, both sides of (4.17) vanish exactly at the eigenvalues of  $H_N$ . Moreover, as both sides are polynomials of degree  $NL$  in  $E$ , comparing the leading coefficient  $E^{NL}$  immediately implies the identity. For the computation the last determinant, one can use the following lemma with  $M = 2L$ . It is [42, Lemma 6.3], but again an alternative proof is provided.

**Lemma 32.** *Let  $A: \mathbb{C}^M \rightarrow \mathbb{C}^L$ ,  $B: \mathbb{C}^M \rightarrow \mathbb{C}^M$ , and  $C: \mathbb{C}^L \rightarrow \mathbb{C}^M$  be linear maps with  $B$  diagonalizable. Then there is a spectral decomposition  $B = \sum_{i=1}^M b_i P_i$  where the  $P_i$  are rank one idempotents onto the eigenvector associated to eigenvalue  $b_i$  and*

$$\det_L(ABC) = \sum_{|I|=L} \left( \prod_{i \in I} b_i \right) \det_L \left( \sum_{i \in I} AP_i C \right)$$

where the sum runs over subsets  $I$  of  $\{1, 2, \dots, M\}$  of cardinality  $|I| = L$ .

*Proof.* The existence of the spectral decomposition follows immediately from the diagonalizability of  $B$ . Moreover, the exterior power  $\bigwedge^L B: \bigwedge^L \mathbb{C}^M \rightarrow \bigwedge^L \mathbb{C}^M$  is also diagonalizable, since  $B$  is diagonalizable, so for the same reason we have a decomposition  $\bigwedge^L B = \sum_I b_I P_I$ , where the  $P_I$  are rank one projections onto the eigenvector associated to eigenvalue  $b_I$  of  $\bigwedge^L B$ . Note that, given a basis of eigenvectors  $(v_i)_{i=1}^M$  of  $B$  for  $\mathbb{C}^M$ , the eigenvalues  $b_I$  can be indexed by choices  $I$  of subsets of  $\{1, 2, \dots, M\}$  of cardinality  $L$  and can then exactly be given by  $b_I = \prod_{i \in I} b_i$ . Moreover, an eigenvector  $v_I$  associated to  $b_I$  can be given by  $v_{i_1} \wedge v_{i_2} \wedge \dots \wedge v_{i_L}$  with  $i_j \in I$  and  $i_j < i_k$  if  $j < k$ . Hence,

$$P_I = \bigwedge^L \sum_{i \in I} P_i$$

which can be checked explicitly by evaluation on a basis of eigenvectors of  $\bigwedge^L B$ . Let  $(e_i)_{i=1}^L$  be a basis of  $\mathbb{C}^L$ . One finds

$$\begin{aligned} & \det_L(ABC) e_1 \wedge e_2 \wedge \dots \wedge e_L \\ &= \left( \bigwedge^L A \bigwedge^L B \bigwedge^L C \right) e_1 \wedge e_2 \wedge \dots \wedge e_L \\ &= \left( \bigwedge^L A \left( \sum_I b_I P_I \right) \bigwedge^L C \right) e_1 \wedge e_2 \wedge \dots \wedge e_L \\ &= \sum_I b_I \left( \bigwedge^L A P_I \bigwedge^L C \right) e_1 \wedge e_2 \wedge \dots \wedge e_L \\ &= \sum_I \left( \prod_{i \in I} b_i \right) \left( \bigwedge^L A \bigwedge^L \sum_{i \in I} P_i \bigwedge^L C \right) e_1 \wedge e_2 \wedge \dots \wedge e_L \\ &= \sum_{|I|=L} \left( \prod_{i \in I} b_i \right) \det_L \left( \sum_{i \in I} AP_i C \right) e_1 \wedge e_2 \wedge \dots \wedge e_L, \end{aligned}$$

which implies the claim. ■

*Proof of Theorem 31.* First of all, the particular form (1.2) of the transfer matrix allows to rewrite (4.17) as

$$\det_{NL}(H_N - E\mathbf{1}) = (-1)^{NL} \det_L(T)^{N+1} \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* (\mathcal{T}^E)^{N+1} \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right).$$

Next, recall that the eigenvalues of the transfer matrix  $\mathcal{T}^E$  are  $z_i(E)$  which by the assumption  $E \notin \mathcal{F}$  are pairwise distinct. Then,  $\mathcal{T}^E$  is diagonalizable and an application of Lemma 32 gives

$$\begin{aligned} \det_{NL}(H_N - E\mathbf{1}) \\ = (-1)^{NL} \det_L(T)^{N+1} \sum_I \left( \prod_{i \in I} z_i(E)^{N+1} \right) \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right), \end{aligned}$$

where the sum is as in the statement of the theorem and  $\mathcal{R}_I^E$  is the Riesz projection associated to the eigenvalues  $z_i(E)$  with  $i \in I$ , namely

$$\mathcal{R}_I^E = \oint_{\gamma_I^0} \frac{dz}{2\pi i} (z\mathbf{1} - \mathcal{T}^E)^{-1},$$

with  $\gamma_I^0$  also as above (which may here be replaced by  $\gamma_I$  because  $\mathcal{T}^E$  is invertible). As to the product of eigenvalues, let us use

$$\left( \prod_{i \in I} z_i(E) \right) \left( \prod_{i \in I^c} z_i(E) \right) = \det_{2L}(\mathcal{T}^E) = \frac{\det_L(R)}{\det_L(T)},$$

where  $I^c = \{1, \dots, 2L\} \setminus I$  is the complementary set of  $I$ . Hence, replacing the definition of  $G_I(E)$ , one finds

$$G_I(E)G_{I^c}(E) = \det_L(R) \det_L(T), \quad (4.18)$$

so that

$$\det_{NL}(H_N - E\mathbf{1}) = (-1)^L \sum_I G_{I^c}(E)^{N+1} \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right).$$

Due to

$$\begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_{I^c}^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathbf{1}_{2L} \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = 0,$$

this implies

$$\det_{NL}(H_N - E\mathbf{1}) = \sum_I G_I(E)^{N+1} \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right),$$

because the indices in  $\sum_I$  and  $\sum_{I^c}$  run through the same set. Finally, replacing (2.12) gives

$$\begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = \oint_{\gamma_I^0} \frac{dz}{2\pi i} (H(z) - E\mathbf{1})^{-1}.$$

Taking the determinant shows the second equality in (4.16), completing the proof. ■

The above strategy of proof can be slightly modified to obtain several similar identities in the present context. Let us spell out two such modifications. Starting directly from (4.17), one finds using first (4.18) and then (2.12) that

$$\begin{aligned}
 \det_{NL}(H_N - E\mathbf{1}) &= (-1)^{NL} \det_L(T)^N \sum_I \left( \prod_{i \in I} z_i(E)^N \right) \det_L \left( \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right) \\
 &= \sum_I G_{I^c}(E)^N \det_L \left( \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right) \\
 &= \sum_I G_I(E)^N \det_L \left( \mathbf{1} - \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}^* \mathcal{R}_I^E \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} \right) \\
 &= \sum_I G_I(E)^N \det_L(\mathbf{1} - T Q_I^{(0)}(E)), \tag{4.19}
 \end{aligned}$$

where

$$Q_I^{(j)}(E) = \oint_{\gamma_I^0} \frac{dz}{2\pi i z^j} (H(z) - E\mathbf{1})^{-1}.$$

By a similar calculation, one also finds

$$\begin{aligned}
 \det_{NL}(H_N - E\mathbf{1}) &= (-1)^{NL} \frac{\det_L(T)^{N+1}}{\det_L(-R)} \det_L \left( \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix}^* (\mathcal{T}^E)^{N+2} \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix} \right) \\
 &= \frac{1}{\det_L(-RT)} \sum_I G_I(E)^{N+2} \det_L(\mathbf{1} + Q_I^{(2)}(E)R). \tag{4.20}
 \end{aligned}$$

One of the crucial facts about Widom's formula (4.14) explored in the next section is that for  $E \notin \Lambda$  and in the limit  $N \rightarrow \infty$  one of the summands is much larger than the others. In fact, it is clear from the definition (4.15) that for  $I_0 = \{1, \dots, L\}$  and any  $I \neq I_0$  one has

$$\left| \frac{G_I(E)}{G_{I_0}(E)} \right| = \left| \frac{\prod_{i \in I_0} z_i(E)}{\prod_{i \in I} z_i(E)} \right| < 1,$$

because the middle two eigenvalues have distinct moduli for  $E \notin \Lambda$  and one always has  $G_I(E) \neq 0$ . Moreover, the corresponding factors  $Q_{I_0}^{(j)}(E)$  are connected to  $Q_s^{(j)}(E)$  as defined in (3.16), namely for  $E \notin \Lambda$  and  $|z_L(E)| < s < |z_{L+1}(E)|$

$$Q_{I_0}^{(0)}(E) = Q_s^{(0)}(E), \quad Q_{I_0}^{(1)}(E) = Q_s^{(1)}(E), \quad Q_{I_0}^{(2)}(E) = Q_s^{(2)}(E).$$

Now, equating (4.14) with (4.19) and (4.20), then dividing by  $G_{I_0}(E)^{N+1}$  and taking the limit  $N \rightarrow \infty$ , one finds for  $s$  still satisfying  $|z_L(E)| < s < |z_{L+1}(E)|$  that

$$\det_L(Q_s^{(1)}(E)) = \frac{1}{G_{I_0}(E)} \det_L(\mathbf{1} - T Q_s^{(0)}(E)) = \frac{G_{I_0}(E)}{\det_L(-RT)} \det_L(\mathbf{1} - T Q_s^{(2)}(E)).$$

These identities do not seem to be contained in Widom's works, but are likely restricted to the present set-up in which the Hypothesis **A** holds. This also allows to rewrite (4.12) as

$$\begin{aligned}
\Gamma &= \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Wind}^E(H^s) = 0 \text{ and} \\
&\quad \det_L(\mathbf{1} - TQ_s^{(0)}(E)) = 0\} \\
&= \{E \in \mathbb{C} \setminus \Lambda : \text{there exists } s > 0 \text{ with } \text{Wind}^E(H^s) = 0 \text{ and} \\
&\quad \det_L(\mathbf{1} - TQ_s^{(2)}(E)) = 0\} \\
&= \{E \in \mathbb{C} \setminus \Lambda : q_{I_0}(E) = 0\}, \tag{4.21}
\end{aligned}$$

where the third equality follows because  $E \in \mathbb{C} \setminus \Lambda$  implies that  $|z_L(E)| < |z_{L+1}(E)|$  so that one can choose a scaling parameter  $s > 0$  such that  $\text{Wind}^E(H^s) = 0$  and for such  $s$  we have  $q_{I_0}(E) = \det(Q_s^{(1)}(E))$ .

#### 4.6. Limit of the spectra of the finite-volume operators

This section provides a more detailed description and discussion of Theorem 1. The following definition will be used to describe the asymptotic spectra.

**Definition 33** (e.g., [9]). Let  $(\Sigma_N)_{N \geq 1}$  be a sequence of subset  $\Sigma_N \subset \mathbb{C}$ . Then the uniform and partial limit sets are respectively defined by

$$\begin{aligned}
\text{u-lim}_{N \rightarrow \infty} \Sigma_N &= \{E \in \mathbb{C} : \text{there exists } E_N \in \Sigma_N \text{ such that } E_N \rightarrow E\}, \\
\text{p-lim}_{N \rightarrow \infty} \Sigma_N &= \{E \in \mathbb{C} : \text{there exists } (N_n)_{n \geq 1} \text{ and } E_{N_n} \in \Sigma_{N_n} \text{ such that} \\
&\quad E_{N_n} \rightarrow E \text{ and } N_n \rightarrow \infty\}.
\end{aligned}$$

**Theorem 34.** *Suppose that Hypotheses **A–D** hold. Then*

$$\partial\Lambda \cup \Gamma = \text{closure}(\text{u-lim}_{N \rightarrow \infty} \text{spec}(H_N)) = \text{p-lim}_{N \rightarrow \infty} \text{spec}(H_N).$$

Clearly, Theorem 34 implies Theorem 1. The proof of Theorem 34 will be given at the end of Section 4.8. In fact, Sections 4.7 and 4.8 will provide criteria the construction of eigenstates near  $E \in \Gamma$  and  $E \in \Lambda$  respectively. The criterion on  $\Lambda$  will invoke Hypothesis **C** in a crucial manner (more precisely, to show that the quotient in (4.24) is not constant). Let us here further elucidate Hypotheses **C** and **D**. It was already stated in Section 1 that Hypothesis **C** holds generically. Example 37 further down provides a model that violates Hypothesis **C**. It is a nilpotent perturbation of a model with symmetries which enforce more than two of the eigenvalues  $z_l(E)$  to be of equal modulus on an open set in  $\mathbb{C}$ . First of all, let us prove that Hypothesis **C** implies that  $\mathring{\Lambda} = \emptyset$  which is actually [42, Condition B].



**Lemma 35.** *Hypothesis C implies that  $\Lambda$  has empty interior, namely  $\Lambda = \partial\Lambda$ .*

*Proof.* Suppose that  $\overset{\circ}{\Lambda} \neq \emptyset$ . Then for all  $E \in \partial\Lambda$  with vanishing distance  $d(E, \overset{\circ}{\Lambda}) = 0$  there would be an open neighborhood  $U$  such that on the open set  $U \cap \overset{\circ}{\Lambda} \neq \emptyset$  one has  $|z_L(E')| = |z_{L+1}(E')|$ ; Lemma 6 then implies that  $z_L(E'') = cz_{L+1}(E'')$  for all  $E'' \in U$  and some constant  $c \in \mathbb{T}$  because all other eigenvalues are separated from the two middle ones by Hypothesis C and the continuity of the  $z_l$ ; but then also  $U \setminus \overset{\circ}{\Lambda}$  lies in  $\Lambda$ , in contradiction to  $E \in \partial\Lambda$ . ■

Another result assuring empty interior of  $\Lambda$  is Lemma 24. Let us also recall from Lemma 21 that  $\partial\Lambda$  locally consists of a finite union of analytic arcs. The finite number of points that are excluded in Hypothesis C are precisely the intersection points of these analytic arcs. As to Hypothesis D for  $q_{I_0}$ , it actually only addresses the bounded components of  $\mathbb{C} \setminus \Lambda$ .

**Lemma 36.**  *$q_{I_0}$  is not constantly zero on the unbounded component of  $\mathbb{C} \setminus \Lambda$ .*

*Proof.* Due to the particular form of  $\mathcal{T}^E$  given in (1.2), one concludes that there is some constant  $C > 0$  such that, for all  $E$  large enough,  $|z_l(E)| \geq C|E|$  for all  $l = L+1, \dots, 2L$  and, using the same bound for  $\tilde{\mathcal{T}}^E$  and invoking (2.15),  $|z_l(E)| \leq C^{-1}|E|^{-1}$  for all  $l = 1, \dots, L$ . Therefore,  $\gamma_{I_0}^0$  in the first formula in (4.16) can be chosen to be the unit circle  $\mathbb{T}$ . Therefore,

$$q_{I_0}(E) = \frac{(-1)^L}{E^L} \det_L \left( \oint_{\mathbb{T}} \frac{dz}{2\pi iz} \left( \mathbf{1} - \frac{H(z)}{E} \right)^{-1} \right) = \frac{(-1)^L}{E^L} + \mathcal{O}(E^{-L-1}),$$

because  $\mathbf{1} - \frac{H(z)}{E}$  is invertible for  $|E|$  large enough so that the integral is merely equal to 1, up to corrections of higher order in  $\frac{1}{E}$ . ■

Next let us provide the promised example of a Hamiltonian for which Hypotheses C and D do not hold.

**Example 37.** For  $L = 3$ , choose

$$R = \begin{pmatrix} r & 0 & 0 \\ 0 & -r & 0 \\ 0 & 0 & 2r \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} t & 0 & 0 \\ 0 & -t & 0 \\ 0 & 0 & \frac{t}{2} \end{pmatrix},$$

where  $r, t$  are non-vanishing complex numbers. Then  $H$  is a direct sum of three Hatano–Nelson models (see Section 5.1), and two of them merely differ by a sign. By Theorem 34, the finite-volume spectra converge to the union of the  $\Lambda$ -sets of the three models with  $L = 1$  which are given by line segments (again Section 5.1). On the other hand, let us next show that  $\Lambda$  for  $H$  has open interior. The eigenvalues of the transfer

matrix  $\mathcal{T}^E$  can be computed as in Section 5.1, but here the only important point is that two merely differ by a sign, namely the analytic branches of these eigenvalues are of the form  $z'_1(E)$ ,  $z'_2(E)$ ,  $-z'_1(E)$ ,  $-z'_2(E)$ ,  $2z'_1(E)$ ,  $2z'_2(E)$ . Hence, one readily checks that  $\Lambda$  has open interior. This is not a contradiction to Theorem 34 though, because by Lemma 35 Hypothesis C does not hold. Furthermore, also Hypothesis D is violated because for any  $E$  in the interior of  $\Lambda$  the lowest three eigenvalues necessarily include all eigenvalues of one of the three direct summands, so that corresponding to this summand, the block-diagonal matrix  $\mathcal{R}_I^E$  contains a  $2 \times 2$  identity matrix which by (4.16) implies that  $q_{I_0}(E) = 0$ . All the above facts readily transpose to a wider class of models. For example, one can fill the (strict) upper triangles of  $R$ ,  $V$ , and  $T$  with arbitrary entries without modifying the sets  $\Lambda$  as well as the spectra.

As already stressed above, the proofs in Sections 4.7 and 4.8 explicitly construct eigenfunctions using Widom's formula. The aim of the following is to compare this to the standard procedure of constructing the solutions using the transfer matrices (which determines the fundamental solutions). The following result spells out what can be achieved near a point  $E \in \Lambda$ . An analogous statement can be obtained for  $E \in \Gamma$ .

**Proposition 38.** *Let  $E \in \Lambda$ . Then there is a normalized quasimode  $\varphi^E \in \mathbb{C}^{NL}$  of  $H_N$  such that*

$$\|(H_N - E\mathbf{1})\varphi^E\| \leq \frac{C}{\sqrt{N}},$$

for some constant  $C$  independent of  $N$ . In particular,

$$\|(H_N - E\mathbf{1})^{-1}\| \geq \frac{\sqrt{N}}{C}.$$

*Proof.* Let us choose the scaling  $s$  in (4.5) such that the two middle eigenvalues of  $\mathcal{T}^E$  both have modulus 1. The bound will be shown for  $S_N H_N S_N^{-1}$  which will simply be denoted by  $H_N$  again. Then  $s$  enters into the constant  $C$ . Let us start constructing the state  $\varphi^E$  from its values  $\varphi_m^E, \varphi_{m+1}^E \in \mathbb{C}^L$  at sites  $m$  and  $m+1$  (say for  $m = \frac{N}{2}$  if  $N$  even) by choosing the vector  $\begin{pmatrix} T\varphi_{m+1}^E \\ \varphi_m^E \end{pmatrix}$  that lies both in the subspace of non-expanding directions of  $\mathcal{T}^E$  and in the subspace of non-expanding directions of  $(\mathcal{T}^E)^{-1}$ . By hypothesis, these two subspaces are both of dimension  $L+1$ , so that the intersection is at least of dimension 2. Then construct  $\varphi^E$  by application of the transfer matrices, both to the right and to the left of  $m$ . One has

$$\sum_{n=1, \dots, N} \|\varphi_n^E\|^2 \leq cN.$$

After normalization of  $\varphi_{m+1}^E, \varphi_m^E$  by  $\sqrt{N}$ , one obtains the desired state for which  $\|\varphi_N^E\| \leq cN^{-\frac{1}{2}}$  and  $\|\varphi_1^E\| \leq cN^{-\frac{1}{2}}$ . This state satisfies the Schrödinger equation

everywhere exactly, except at these two boundary sites 1 and  $N$ . This leads to the desired bound. Then

$$1 = \|\varphi^E\| = \|(H_N - E\mathbf{1})^{-1}(H_N - E\mathbf{1})\varphi^E\| \leq \|(H_N - E\mathbf{1})^{-1}\| \frac{C}{\sqrt{N}},$$

implies the second bound.  $\blacksquare$

If now  $H_N$  is normal, the spectral theorem implies

$$\|(H_N - E\mathbf{1})^{-1}\| = \text{dist}(E, \text{spec}(H_N))^{-1}$$

so that one can conclude from Proposition 38 that there is an eigenvalue in a neighborhood of  $E$  of size of order  $\mathcal{O}(N^{-\frac{1}{2}})$ . However, for non-normal operators it is well known [39] that  $\|(H_N - E\mathbf{1})^{-1}\|$  is much larger than  $\text{dist}(E, \text{spec}(H_N))^{-1}$ , and then it is not possible to use the approximate eigenfunctions constructed in Proposition 38 to conclude that  $H_N$  has spectrum near  $E$ . This difficulty cannot be circumvented by general methods, and is addressed by Widom's formula in the next two sections.

#### 4.7. Construction of eigenfunctions close to $\Gamma$

In this section it will be proved that the multiplicity of a zero of  $q_{I_0}$  determines the number of eigenvalues of  $H_N$  that are attracted to it. Due to (4.21) this proves the part of Theorem 34 involving the set  $\Gamma$ . As to notation, let us denote by  $\text{ord}_f(E)$  the order of the point  $E \in \mathbb{C}$  as zero of a holomorphic function  $f$ ; if  $E$  is not a zero of  $f$ , then  $\text{ord}_f(E) = 0$ .

**Proposition 39.** *Assume that  $q_{I_0}$  is not the zero function on  $\mathbb{C} \setminus \Lambda$ . Then, for all  $E \in \mathbb{C} \setminus \Lambda$  and all  $\rho > 0$  small enough, there exists an  $N_0 \in \mathbb{N}$  such that for all  $N > N_0$*

$$\#\text{spec}(H_N) \cap B_\rho(E) = \text{ord}_{q_{I_0}}(E).$$

For the proof, let us start with a preparatory result that is also used in the next section. The eigenvalues  $z_i(E)$  of the transfer matrix  $\mathcal{T}^E$  are locally holomorphic on  $\mathbb{C} \setminus \mathcal{F}$  up to relabeling as explained in the proof of Lemma 4.1. Here locally holomorphic means holomorphic up to branch cuts. They are, however, not necessarily holomorphic at  $\mathcal{F}$ , but merely have a Puiseux expansion there. By a procedure as for the  $z_i$  in the proof of Lemma 4.1 we can assume that the  $G_I$  are locally holomorphic on  $\mathbb{C} \setminus \mathcal{F}$ , but not necessarily at points in  $\mathcal{F}$ . The next lemma shows at which points in  $\mathcal{F}$  the functions  $G_I$  are nevertheless holomorphic. Note that this is in line with standard results on the analyticity of Riesz projections as derived in [19].

**Lemma 40.** *Let  $I \subset \{1, \dots, 2L\}$  consist of  $L$  numbers and  $E \in \mathbb{C}$  be such that for all  $i \in I$  and  $j \notin I$  one has  $z_i(E) \neq z_j(E)$ . Then  $G_I$  and  $q_I$  are holomorphic at  $E$ . In particular,  $G_I$  and  $q_I$  are locally holomorphic on  $\mathbb{C} \setminus \mathcal{F}$ .*

*Proof.* Note that the choice of  $E$  assures that the path  $\gamma_I^0$  as used in Theorem 31 exists. Next

$$\begin{aligned}
G_I(E) &= \exp \operatorname{Log} \left( (-1)^L \det_L(R) \prod_{i \in I} \frac{1}{z_i(E)} \right) \\
&= \exp \left( \oint_{\gamma_I^0} \operatorname{Log} \left( z^L \det_L(H(z) - E\mathbf{1}) \prod_{i \in I} (z - z_i(E))^{-1} \right) \frac{dz}{2\pi i z} \right) \\
&= \exp \left( \oint_{\gamma_I^0} \operatorname{Log}(\det_L(H(z) - E\mathbf{1})) \frac{dz}{2\pi i z} \right. \\
&\quad \left. + \oint_{\gamma_I^0} \operatorname{Log} \left( z^L \prod_{i \in I} (z - z_i(E))^{-1} \right) \frac{dz}{2\pi i z} \right) \\
&= \exp \left( \oint_{\gamma_I^0} \operatorname{Log}(\det_L(H(z) - E\mathbf{1})) \frac{dz}{2\pi i z} \right).
\end{aligned}$$

From this expression, it follows immediately that  $G_I$  is holomorphic in  $E$ . The claim on  $q_I$  follows similarly from (4.16). ■

**Corollary 41.**  $G_{I_0}$  and  $q_{I_0}$  are locally holomorphic on  $\mathbb{C} \setminus \Lambda$ .

*Proof.* For  $E \in \mathbb{C} \setminus \Lambda$ , one has  $|z_L(E)| < |z_{L+1}(E)|$ ; so, for  $E \in \mathcal{F} \setminus \Lambda$ , one has  $z_i(E) = z_j(E)$  either for  $i, j \geq L + 1$  or  $i, j \leq L$ . The claim thus follows from Lemma 40. ■

*Proof of Proposition 39.* For all  $E \in \mathbb{C} \setminus \Lambda$  and  $I \neq I_0$ , one has

$$\left| \frac{G_I(E)}{G_{I_0}(E)} \right| = \left| \frac{\prod_{i \in I_0} z_i(E)}{\prod_{i \in I} z_i(E)} \right| < 1.$$

Using Widom's formula (4.14) and the fact that the functions  $G_I$  are continuous, one deduces that

$$\frac{\det(H_N - E\mathbf{1})}{G_{I_0}(E)^{N+1}} - q_{I_0}(E) = \sum_{I \neq I_0} \left( \frac{G_I(E)}{G_{I_0}(E)} \right)^{N+1} q_I(E) \rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad (4.22)$$

uniformly on compact subset of  $\mathbb{C} \setminus (\Lambda \cup \mathcal{F})$ . Let us next analyze the convergence also on compact subsets of  $\mathbb{C} \setminus \Lambda$  which contain points in  $\mathcal{F}$ . For that purpose, let  $f_N$  denote the function on the left-hand side of (4.22). By Corollary 41, it is holomorphic in all  $E \in \mathbb{C} \setminus \Lambda$ . Let next  $E \in \mathcal{F} \setminus \Lambda$  and choose  $\rho > 0$  such that the closed ball  $B_{2\rho}(E)$  of size  $2\rho$  around  $E$  has trivial intersection with  $\Lambda$ . Then for all  $E' \in B_\rho(E)$ ,

Cauchy's integral formula gives

$$\begin{aligned} |f_N(E')| &= \left| \oint_{\partial B_{2\rho}(E)} \frac{dz}{2\pi i} \frac{f_N(z)}{z - E'} \right| \leq \frac{1}{2\pi\rho} \oint_{\partial B_{2\rho}(E)} dz |f_N(z)| \\ &\leq 2 \max_{z \in \partial B_{2\rho}(E)} |f_N(z)|, \end{aligned}$$

which converges to 0 by the above, uniformly in  $E'$ . In conclusion,

$$\lim_{N \rightarrow \infty} \frac{\det(H_N - E\mathbf{1})}{G_{I_0}(E)^{N+1}} = q_{I_0}(E),$$

uniformly on compact subsets of  $\mathbb{C} \setminus \Lambda$ . Note that the functions  $G_{I_0}$  and  $q_{I_0}$  are locally holomorphic on  $\mathbb{C} \setminus \Lambda$  by Corollary 41. Moreover,  $\det_{NL}(H_N - E\mathbf{1})$  is polynomial in  $E$  and thus an entire function. Since  $G_{I_0}$  has no zeros because the transfer matrix is invertible for all  $E$ , it follows that also the expression  $\det_{NL}(H_N - E\mathbf{1})G_{I_0}(E)^{-N-1}$  is locally holomorphic in  $E \in \mathbb{C} \setminus \Lambda$ . The statement of the proposition now follows from Hurwitz's theorem and the fact that the zeros of  $\det_{NL}(H_N - E\mathbf{1})G_{I_0}(E)^{-N-1}$  are exactly the eigenvalues of  $H_N$ . ■

**Remark 42.** It is possible to generalize Proposition 39 so that the limit of the spectrum of  $H_N$  can be studied in cases where Hypotheses C and D do not hold. For that purpose, one can consider a generalized  $\Lambda$ -set  $\hat{\Lambda} = \bigcup_{I,J} \partial\{E \in \mathbb{C} : |G_I(E)| = |G_J(E)|\}$ . On the connected components of the complement of  $\hat{\Lambda}$ , one has either  $|G_I| < |G_J|$  or  $|G_I| = |G_J|$ . Now, there is an  $I$  such that locally  $q_I \neq 0$  and if one replaces  $I_0$  with  $I$  in (4.22), then the limit again converges uniformly on compact subsets, but now possibly only for subsequences. Carrying out the argument as in the above proof, one obtains a similar result as in Proposition 39, but now on the set  $\mathbb{C} \setminus \hat{\Lambda}$  and possibly only for subsequences. Details will be provided elsewhere.

#### 4.8. Construction of eigenfunctions close to $\partial\Lambda$

The first lemma provides a criterion for an energy  $E \in \partial\Lambda$  to attract an eigenvalue of  $H_N$ . It is slightly more general than required for the proof of Theorem 34.

**Lemma 43.** *Let  $E \in \mathbb{C} \setminus \mathcal{F}$  be such that*

$$|z_{L-1}(E)| < |z_L(E)| = |z_{L+1}(E)| < |z_{L+2}(E)|,$$

*and suppose that  $q_{I_0}(E)$  and  $q_{I_1}(E)$  are non-zero and  $G_{I_0}/G_{I_1}$  is locally non-constant around  $E$ , where  $I_1 = \{1, \dots, L-1, L+1\}$ , as in Section 1. Then there are constants  $C > 0$  and  $n, N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$  there is an  $\varepsilon \in \mathbb{C}$  with  $|\varepsilon| \leq CN^{-\frac{1}{n}}$  and  $E + \varepsilon \in \text{spec}(H_N)$ .*

*Proof.* Note that the supposed equality of the middle two eigenvalues implies that  $E \in \Lambda$ . Let us start out by using Widom's formula (4.14) which implies that, if  $E + \varepsilon \notin \mathcal{F}$ , then  $E + \varepsilon \in \text{spec}(H_N)$  if and only if

$$0 = \sum_I G_I(E + \varepsilon)^{N+1} q_I(E + \varepsilon). \quad (4.23)$$

If  $\varepsilon = 0$  solves this equation, then it is the desired solution. In the following, it will hence be supposed that  $\varepsilon = 0$  does not solve (4.23). Next using that

$$\varepsilon \mapsto G_{I_0}(E + \varepsilon)/G_{I_1}(E + \varepsilon)$$

is non-constant in a neighborhood of  $\varepsilon = 0$ , (4.23) can be rewritten to

$$\begin{aligned} & \left( \frac{G_{I_1}(E)}{G_{I_0}(E)} \cdot \frac{G_{I_0}(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right)^{N+1} \\ &= - \frac{1}{q_{I_0}(E + \varepsilon)} \left( \frac{G_{I_1}(E)}{G_{I_0}(E)} \right)^{N+1} \sum_{I \neq I_0} \left( \frac{G_I(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right)^{N+1} q_I(E + \varepsilon). \end{aligned}$$

Taking a (suitable branch of the) logarithm and dividing by  $N + 1$ , one sees that (4.23) for  $\varepsilon \neq 0$  (namely  $E + \varepsilon \in \text{spec}(H_N)$ ) is equivalent to  $g_N(\varepsilon) = \frac{1}{N+1}$ , where

$$g_N(\varepsilon) = \frac{\text{Log}\left(\frac{G_{I_1}(E)}{G_{I_0}(E)} \cdot \frac{G_{I_0}(E+\varepsilon)}{G_{I_1}(E+\varepsilon)}\right)}{\text{Log}\left(-q_{I_0}(E + \varepsilon)^{-1} \left(\frac{G_{I_1}(E)}{G_{I_0}(E)}\right)^{N+1} \sum_{I \neq I_0} \left(\frac{G_I(E+\varepsilon)}{G_{I_1}(E+\varepsilon)}\right)^{N+1} q_I(E + \varepsilon)\right)}.$$

Let us stress that the function  $g_N$  is well defined and holomorphic on some connected, simply connected, open neighborhood of  $\varepsilon = 0$  because all  $G_I$  and  $q_I$  are,  $q_{I_0}(E)$  is non-zero and the denominator is non-zero at  $\varepsilon = 0$  by the assumption that  $\varepsilon = 0$  does not solve (4.23). Furthermore, let us note that  $g_N(0) = 0$  and that 0 is an isolated zero of  $g_N$  since the denominator of  $g_N$  around  $\varepsilon = 0$  is bounded and  $G_{I_0}/G_{I_1}$  is holomorphic and locally non-constant around  $E$ .

Next, because  $|z_{L-1}(E)| < |z_L(E)| = |z_{L+1}(E)| < |z_{L+2}(E)|$  and the  $G_I$  are continuous,

$$\left| \frac{G_I(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right| = \left| \frac{\prod_{i \in I_1} z_i(E + \varepsilon)}{\prod_{i \in I} z_i(E + \varepsilon)} \right| < 1,$$

for all  $I \notin \{I_0, I_1\}$  and all  $\varepsilon$  in a neighborhood of 0. It follows that

$$\lim_{N \rightarrow \infty} \sum_{I \neq I_0} \left( \frac{G_I(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right)^{N+1} q_I(E + \varepsilon) = q_{I_1}(E + \varepsilon),$$

uniformly on compact subsets of a neighborhood of  $\varepsilon = 0$ . Since both  $q_{I_0}$  and  $q_{I_1}$  are continuous and non-zero at  $E$  and since  $G_{I_1}(E)/G_{I_0}(E) = z_L(E)/z_{L+1}(E) \in \mathbb{T}^1$ ,

there exists an  $M \in \mathbb{R}_{>0}$  and an  $N_1 \in \mathbb{N}$  such that for all  $N > N_1$  and all  $\varepsilon$  in an open neighborhood of 0

$$\left| \text{Log}(-q_{I_0}(E + \varepsilon)^{-1} \left( \frac{G_{I_1}(E)}{G_{I_0}(E)} \right)^{N+1} \sum_{I \neq I_0} \left( \frac{G_I(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right)^{N+1} q_I(E + \varepsilon)) \right| < M.$$

Moreover, the  $G_I$  are holomorphic around  $E$  since  $E \notin \mathcal{F}$ , and  $G_{I_0}/G_{I_1}$  is not locally constant around  $E$  by assumption. It follows that

$$\frac{G_{I_1}(E)}{G_{I_0}(E)} \cdot \frac{G_{I_0}(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} = 1 + c_n \varepsilon^n + \mathcal{O}(\varepsilon^{n+1}) \quad (4.24)$$

for some  $n \in \mathbb{N}$  and non-zero  $c_n \in \mathbb{C}$ . Combining these results, one finds that

$$|g_N(\varepsilon)| > \frac{1}{M} \left| \text{Log} \left( \frac{G_{I_1}(E)}{G_{I_0}(E)} \cdot \frac{G_{I_0}(E + \varepsilon)}{G_{I_1}(E + \varepsilon)} \right) \right| = \frac{1}{M} |c_n \varepsilon^n (1 + \mathcal{O}(\varepsilon))| > \frac{c'}{M} |\varepsilon|^n, \quad (4.25)$$

for some  $c' > 0$ , all  $N > N_1$  and all  $\varepsilon$  in a neighborhood of 0.

Next, recall that  $g_N$  is constructed in such a way that, whenever  $g_N(\varepsilon)$  exists and  $E + \varepsilon \notin \mathcal{F}$ , then  $g_N(\varepsilon) = \frac{1}{N+1}$  if and only if  $E + \varepsilon \in \text{spec}(H_N)$ . Since  $\mathcal{F}$  is finite and  $E \notin \mathcal{F}$ , one can now choose an  $\varepsilon$ -neighborhood  $V$  of 0 on which the lower bound (4.25) holds and which, moreover, satisfies  $(E + V) \cap \mathcal{F} = \emptyset$ . As all  $g_N$  are holomorphic around  $\varepsilon = 0$  and have  $\varepsilon = 0$  as an isolated zero, it follows from the open mapping theorem and the fact that  $\frac{1}{N+1} \rightarrow 0$  as  $N \rightarrow \infty$  that there exists an  $N_2 \in \mathbb{N}$  such that for all  $N > N_2$  the equations  $g_N(\varepsilon) = \frac{1}{N+1}$  have solutions  $\varepsilon_N$ . Moreover, the lower bound (4.25) tells us that  $\varepsilon_N \rightarrow 0$  as  $N \rightarrow \infty$ . Because  $V$  is a neighborhood of 0, there thus is an  $N_3 \geq N_2$  such that for all  $N > N_3$  one has  $\varepsilon_N \in V$ . Using the lower bound (4.25) again, one concludes that for all  $N > \max\{N_1, N_3\}$

$$\frac{1}{N} > \frac{1}{N+1} = g_N(\varepsilon_N) > \frac{c'}{M} |\varepsilon_N|^n,$$

and  $E + \varepsilon_N \in \text{spec}(H_N)$  since  $\frac{1}{N+1} = g_N(\varepsilon_N)$  and  $\varepsilon_N \in V$  such that  $E + \varepsilon_N \notin \mathcal{F}$ . ■

The previous lemma can under suitable conditions be used to prove that  $\partial\Lambda$  attracts spectrum of the Hamiltonians  $H_N$ .

**Proposition 44.** *Suppose Hypotheses A–D hold. Then, for all but finitely many  $E \in \partial\Lambda$ , there are constants  $C > 0$  and  $N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$  there is an  $\varepsilon \in \mathbb{C}$  with  $|\varepsilon| \leq \frac{C}{N}$  and  $E + \varepsilon \in \text{spec}(H_N)$ .*

*Proof.* Note that the functions  $q_I$  are all locally holomorphic on  $\mathbb{C} \setminus \mathcal{F}$ . Since  $q_{I_0}$  and  $q_{I_1}$  are not locally constantly zero on  $\partial\Lambda \setminus \mathcal{F}$ , their zeros on this set are isolated. Because  $\Lambda$  is compact, it thus follows that  $q_{I_0}$  and  $q_{I_1}$  have only finitely many zeros

on  $\partial\Lambda \setminus \mathcal{F}$ . Moreover,  $\frac{G_{I_0}(E)}{G_{I_1}(E)} = \frac{z_{L+1}(E)}{z_L(E)}$  is locally non-constant around all  $E \in \partial\Lambda \setminus \mathcal{F}$  because  $E \in \partial\Lambda$  and outside of  $\Lambda$  its modulus is less than 1. Since  $\mathcal{F}$  is finite, an application of the previous lemma now implies that for all but finitely many  $E \in \partial\Lambda$  there are constants  $C > 0$  and  $n, N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$  there is an  $\varepsilon \in \mathbb{C}$  with  $|\varepsilon| \leq CN^{-\frac{1}{n}}$  and  $E + \varepsilon \in \text{spec}(H_N)$ . As the quotient on the left-hand side of (4.24) is equal to  $\frac{z_{L+1}(E)z_L(E+\varepsilon)}{z_L(E)z_{L+1}(E+\varepsilon)}$ , the proof of Lemma 21 shows that for all but finitely many  $E$  one can take  $n = 1$  in the proof of Lemma 43. ■

*Proof of Theorem 34.* First recall  $\Lambda = \partial\Lambda$  by Lemma 35. Therefore, Proposition 39 implies that

$$(\text{u-lim}_{N \rightarrow \infty} \text{spec}(H_N)) \cap (\mathbb{C} \setminus \partial\Lambda) = (\text{p-lim}_{N \rightarrow \infty} \text{spec}(H_N)) \cap (\mathbb{C} \setminus \partial\Lambda) = \Gamma.$$

Furthermore, Proposition 44 shows that one has a sequence of eigenvalues converging to all but a finite number of points in  $\partial\Lambda$ , hence  $\partial\Lambda \setminus \text{u-lim}_{N \rightarrow \infty} \text{spec}(H_N)$  is finite. Since  $\Gamma$  is discrete and  $\Lambda = \partial\Lambda$  is closed and has no isolated points, it follows that

$$\text{closure}(\text{u-lim}_{N \rightarrow \infty} \text{spec}(H_N)) = \partial\Lambda \cup \Gamma.$$

Note that for any sequence  $(\Sigma_n)_{n \geq 1}$  of subsets its partial limit  $\text{p-lim}_{n \rightarrow \infty} \Sigma_n$  is closed. Hence,

$$\text{closure}(\text{u-lim}_{N \rightarrow \infty} \text{spec}(H_N)) \subset \text{p-lim}_{N \rightarrow \infty} \text{spec}(H_N) \subset \partial\Lambda \cup \Gamma,$$

concluding the proof. ■

## 5. Applications to non-Hermitian quantum systems

This section applies the above results to several concrete models that have been studied in the physics literature. These models all describe quasi-one-dimensional non-Hermitian quantum systems with periodic coefficients. In applications, the non-Hermitian character typically results from either dissipative terms or external driving forces. The spectra of the infinite-volume operators and finite-volume approximations are computed numerically and are, as expected, often very different. This is illustrated in Section 5.1 in the example of the Hatano–Nelson model [18] and one of its variants having next-nearest neighbor hopping. Furthermore, the eigenfunctions of the finite-volume operators are typically localized at one or the other end of the system, an effect called the *skin effect* (see Section 5.2). Next, Section 5.3 briefly defines and discusses the generalized Brillouin zone. In the following Section 5.4, particular focus is on finite systems with topological invariants protected by a (chiral) symmetry. These invariants imply the existence of (approximate zero energy) bound states, and this is again illustrated in a concrete model.



### 5.1. Hatano–Nelson-like models

The Hatano–Nelson model [18] is the most simple strictly one-dimensional and translation invariant non-Hermitian Hamiltonian, it is namely tridiagonal with constant scalar entries on the diagonal. This corresponds to the case  $L = 1$  with matrix entries  $R, T \in \mathbb{C} \setminus \{0\}$  and  $V \in \mathbb{C}$ . As the symbol is scalar, one can readily write out the spectrum  $\Sigma^s = \text{spec}(H^s)$  of the periodic scaled Hamiltonian  $H^s$ :

$$\Sigma^s = \{s^{-1}Re^{-ik} + V + sTe^{ik} : k \in [-\pi, \pi)\}. \quad (5.1)$$

This set is an ellipse in the complex plane which, for the critical scaling parameter  $s_c$  given by  $s_c^2 = |RT^{-1}|$ , degenerates to a line segment. As the Schmidt–Spitzer result (4.10) applies in the scalar case considered here, one deduces

$$\Lambda = \Sigma^{s_c},$$

namely  $\Lambda$  is a line segment. Of course, this can also be obtained directly from the definition (4.1) of the set  $\Lambda$  as the set of energies where the middle two eigenvalues of the transfer matrix have the same modulus. Indeed, it follows from (5.1) for  $s = s_c$  that each point  $E \in \Lambda$  is reached for two different values of  $k$ . By (2.13), these two numbers on the unit circle are also the two eigenvalues  $z_1(E), z_2(E) \in \mathbb{T}$  of the transfer matrices  $\mathcal{T}^{s_c, E}$ , so that  $E \in \Lambda$ . Yet another way to compute  $\Lambda$  starts from the formulas for the two eigenvalues of  $\mathcal{T}^E$ :

$$z_{1/2}(E) = \frac{(E - V)T^{-1}}{2} \pm \sqrt{\frac{(E - V)^2 T^{-2}}{4} - \frac{R}{T}}.$$

Then  $\Lambda = \{E \in \mathbb{C} : |z_1(E)| = |z_2(E)|\}$  can be computed explicitly using the fact that for any complex number  $\xi \in \mathbb{C}$  one has the equivalence

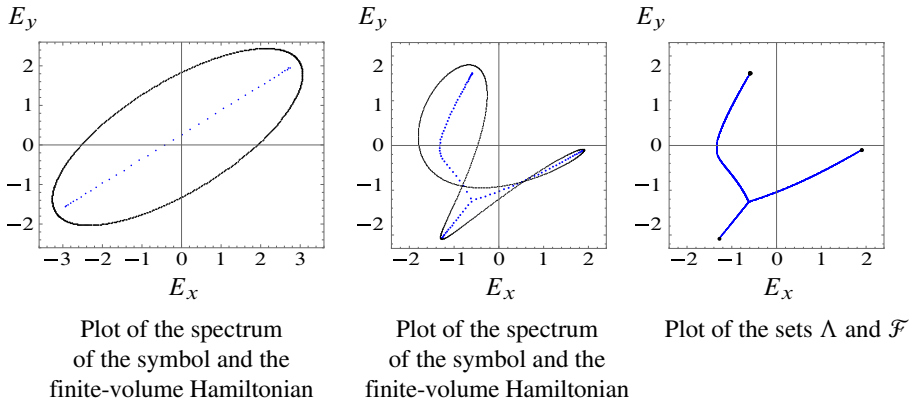
$$\Re e(\xi) = 0 \iff \xi^2 \leq 0$$

where the right-hand side means that  $\xi^2$  is real and non-positive. Elementary algebra then leads to

$$\Lambda = \left\{ E \in \mathbb{C} : \frac{(E - V)^2}{RT} \leq 4 \right\}.$$

This is yet another representation of  $\Lambda$  as a line segment. Note that this formula is intrinsically independent of scaling  $(R, T) \mapsto (s^{-1}R, sT)$  and that for the well-known case of the discrete Laplacian where  $R = T = 1$  and  $V = 0$ , one recovers  $\Lambda = \{E \in \mathbb{R} : |E| \leq 2\}$ . Finally, let us note that the set of outliers  $\Gamma = \emptyset$  is empty by Proposition 27. All of this is illustrated in the numerical example in Figure 3.

Next let us consider a non-Hermitian scalar Hamiltonian with five diagonals, namely add next-nearest neighbor hopping terms. As described in Example 3 this

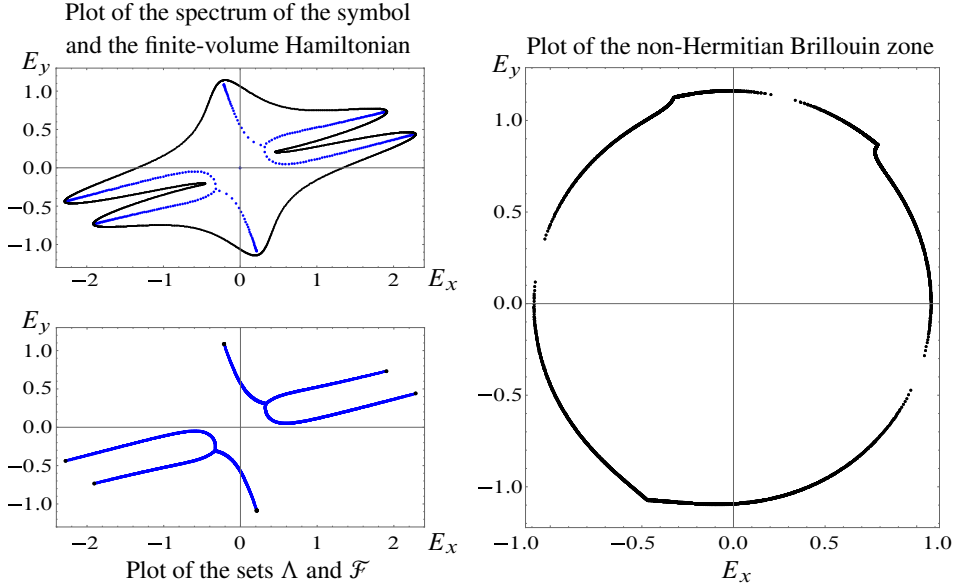


**Figure 3.** The first plot shows the numerical computation of the curve  $\Sigma$  and the spectrum of the Hatano–Nelson model with  $T = 0.5 + i$ ,  $V = -0.1 + 0.2i$ , and  $R = 2.5$ , and  $N = 40$ . The second plot provides the spectrum of the symbol and finite-volume Hamiltonian  $H_N$  for the next-nearest neighbor version with  $t_{-2} = 1.0$ ,  $t_{-1} = -0.5i$ ,  $t_0 = -0.3 - 0.3i$ ,  $t_1 = 1.0$ , and  $t_2 = 0.5i$  with  $N = 50$ , and for the same parameters the last plot shows the set  $\Lambda$  as well as  $\{E \in \Lambda : z_L(E) = z_{L+1}(E)\} \subset \mathcal{F} \cap \Lambda$  by larger dots (this latter set consists of merely the three end points of the segments).

model depends on five complex parameters  $t_{-2}, \dots, t_2 \in \mathbb{C}$  with  $t_{\pm 2} \neq 0$  and can be cast in the form (1.1) with blocks of size  $L = 2$ . The Schmidt–Spitzer result (4.10) applies still, but it does not allow to compute the set  $\Lambda$  as explicitly. One rather has to recourse to the definition (4.1) for a numerical computation: for  $E$  on a suitable grid of a region of the complex plane, one computes the eigenvalues of  $\mathcal{T}^E$  and simply checks whether the middle two eigenvalues are of approximately the same modulus. A first example is provided in Figure 3, others in Figures 4 and 5. Let us mention that another numerical algorithm for the computation of  $\Lambda$  for scalar banded Toeplitz matrices (as described in Example 3) is described and implemented in the recent paper [6].

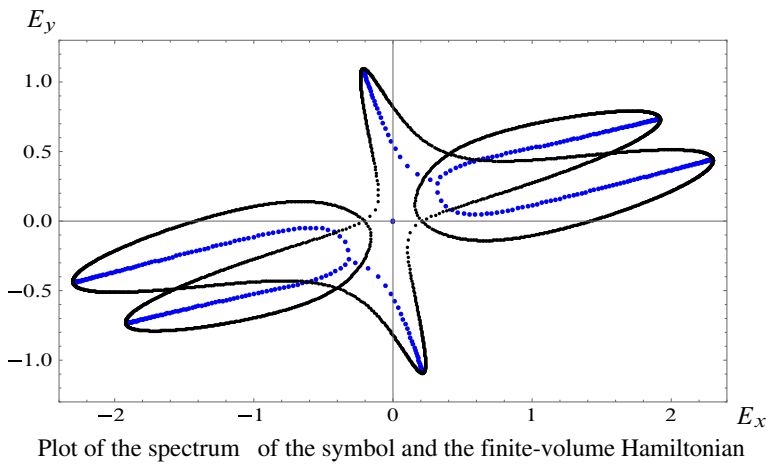
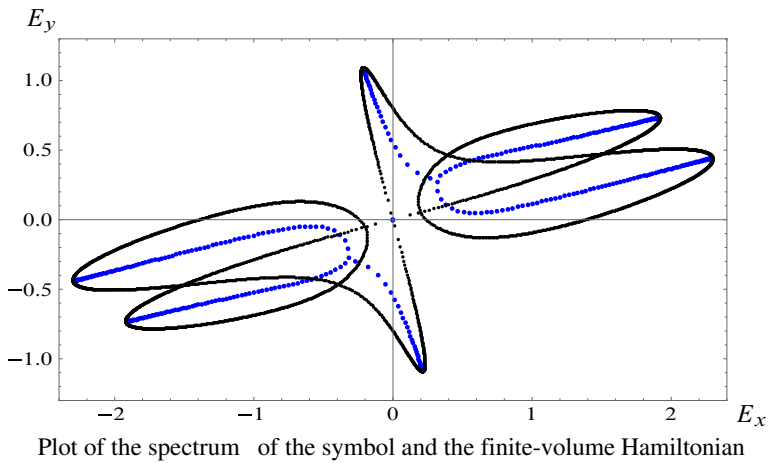
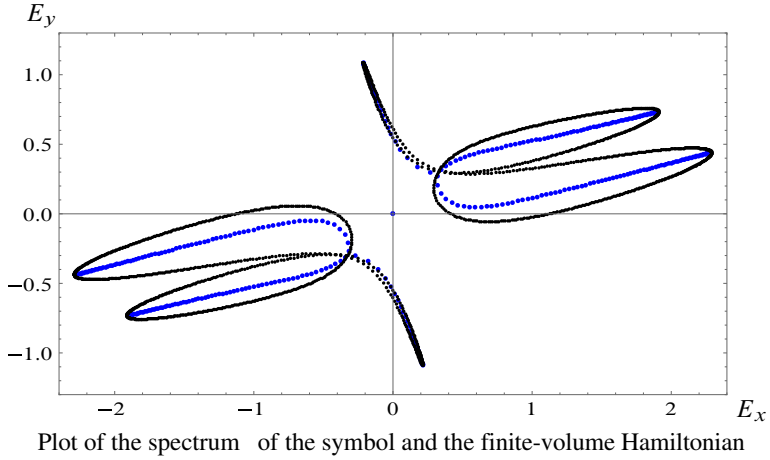
## 5.2. The skin effect

For finite-volume non-Hermitian Hamiltonians  $H_N$ , the eigenstates are generically localized either at the left or right boundary of the system, an effect that is called the *skin effect* in the physics literature (see [48] for a physics review). The aim of this section is to provide a detailed understanding of the skin effect from a mathematical perspective. For the case of compressions of scalar banded Toeplitz matrices, this was already analyzed in [8, Chapter 12], based on [7].



**Figure 4.** Numerical analysis of the chiral Hamiltonian (5.3) with  $v_+ = -0.1 - 0.5i$ ,  $v_- = 0.7$ ,  $t_+ = 0.5 + i$ ,  $t_- = 1 = r_+$  and  $r_- = 1.5 - 0.1i$ . The top-left plot shows the periodic spectrum (outer continuous curve) and the eigenvalues of the finite-volume Hamiltonian of length  $N = 200$ . Note that there is a doubly degenerate (approximate) zero mode. The bottom-left plot (same as in Figure 1) shows the sets  $\Lambda$  and  $\{E \in \Lambda : z_L(E) = z_{L+1}(E)\} \subset \mathcal{F} \cap \Lambda$  (method is described in the text). The right plot shows the non-Hermitian Brillouin zone with holes stemming from numerical approximation and the abrupt changes of the eigenvalues of the transfer matrix along  $\Lambda$ .

Let us start by recalling that Theorem 34 shows that eigenvalues of  $H_N$  accumulate on  $\Lambda$ . Furthermore, by (4.2), one has that  $\Lambda$  is a subset of the union  $\Sigma \cup \{E \in \mathbb{C} : \text{Wind}^E(H) \neq 0\}$ . Hence, considering the decomposition of  $\mathbb{C} \setminus \Sigma$  into connected components,  $\Lambda$  will only lie in those components having non-vanishing winding number. This non-vanishing winding number in turn implies that generalized eigenspace  $\mathcal{E}_1^E$  of all eigenvalues of the transfer matrix  $\mathcal{T}^E$  with modulus less than 1 has a dimension given by  $\dim(\mathcal{E}_1^E) = L + \text{Wind}^E(H)$ , see (3.12). In particular,  $\text{Wind}^E(H) \neq 0$  implies that on  $\Lambda$  one either has at least  $L + 1$  contracting directions of  $\mathcal{T}^E$  (for  $\text{Wind}^E(H) > 0$ ) or  $L + 1$  expanding directions (for  $\text{Wind}^E(H) < 0$ ), so that the middle two eigenvalues  $z_L(E)$ ,  $z_{L+1}(E)$  of same modulus are hence of modulus smaller or larger than 1 respectively. It is shown in Section 4.8 that the eigenfunctions of  $H_N$  are constructed for energies close to  $\Lambda$  by using small perturbations of the two-dimensional subspace spanned by the eigenvectors of  $z_L(E)$ ,  $z_{L+1}(E)$

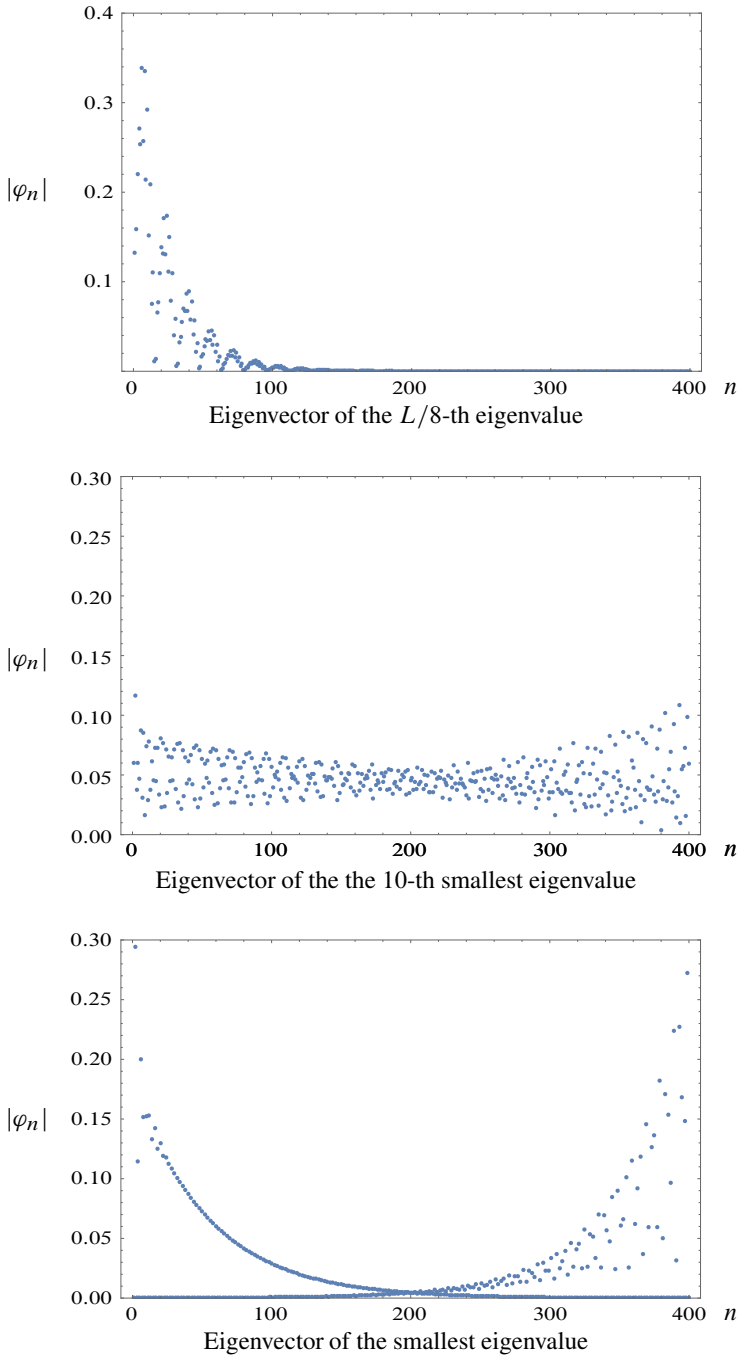


**Figure 5.** Same as first plot in Figure 4, but with scaling  $s = 1.172$ ,  $s = 1.214$ , and  $s = 1.22$ .

(more precisely, the perturbations in energy and the subspace can indeed be done in such a way that the Dirichlet boundary conditions can be satisfied at both sides of the sample, see Section 4.8). On the other hand, for eigenstates of  $H_N$  at energies  $E$  lying in a region of  $\mathbb{C} \setminus \Sigma$  with vanishing winding number (such energies can be either spectral outliers or points converging to  $\Lambda$  in the limit  $N \rightarrow \infty$ , but nevertheless still in a region with vanishing winding number for finite  $N$ ), there are  $L$  eigenvalues of  $\mathcal{T}^E$  less than 1 and  $L$  others larger than 1. Typically, a vector composed of two neighboring entries of an eigenstate is then neither in the increasing nor the decreasing subspace of  $\mathcal{T}^E$ , but rather a linear combination of two vectors from these spaces and hence a linear combination of a left skin and a right skin state. Two examples are given in the middle and bottom plots of Figure 6. Finally, for an eigenvalue  $E$  of  $H_N$  that accidentally falls on  $\Sigma$ , the corresponding eigenstate must have a contribution in the two middle eigenvalues of modulus 1, namely essentially be of plane-wave nature. Let us summarize:

- for  $E \in \text{spec}(H_N)$  lying in a component of  $\mathbb{C} \setminus \Sigma$  with positive/negative winding number  $\text{Wind}^E(H)$ , the eigenstate is localized at the left/right boundary (skin state);
- for  $E \in \text{spec}(H_N)$  lying in a region with vanishing winding number, the eigenstate need not be localized at one side of the sample (but it need not be a plane wave state either);
- for  $E \in \text{spec}(H_N) \cap \Sigma$ , the eigenstate is like a plane wave.

Let us first illustrate these facts by looking at the middle plot in Figure 3, hence corresponding to a model with five scalar diagonals. Clearly, there are three components  $\mathbb{C} \setminus \Sigma$  with non-vanishing winding number and whether it is positive or negative cannot be deduced from the plot which does not show the orientation of  $k \mapsto \det(H(e^{ik}) - E\mathbf{1})$ . The sign of the winding number in each component determines whether the skin states are left or right bound. Moreover, there are two crosses of  $\Sigma$  which also cross  $\Lambda$ , and the corresponding eigenstates of  $H_N$  near these crossings are plane wave like. As one changes the scaling, namely passes from  $H$  to  $H^s$ , the curve  $\Lambda$  remains unchanged (and also the spectrum of  $H_N^s$ ), but the curve  $\Sigma^s = \text{spec}(H^s)$  changes. In particular, the crosses of  $\Sigma^s$  move along  $\Lambda$  as  $s$  changes. Hence, also the left/right nature of the skin states changes as the cross moves along  $\Lambda$ . Generically, eigenstates that are near  $\Sigma$  have a less pronounced skin effect than those lying deep inside the components of  $\mathbb{C} \setminus \Sigma$  with non-vanishing winding number. This is not necessarily true though. For example, for the scalar Hatano–Nelson model (tridiagonal with scalar entries, see first plot in Figure 3), the arguments in Section 5.1 show that the middle two eigenvalues have a constant modulus on  $\Lambda$  so that the skin effect appears with a uniform rate throughout  $\Lambda$ . Furthermore, for  $H^{s_c}$  with critical scaling  $s_c$ , all states are plane-wave states of constant modulus. Further comments on the skin effect in Figures 4 to 6 are given in Section 5.4.



**Figure 6.** Same model as Figure 4, but with scaling parameter  $s = 1.172$  so that  $\Lambda$  is given by first plot in Figure 5. The plots show the absolute value of the normalized eigenvector  $\varphi$  for the 25th largest and the 10th smallest eigenvalues (at  $-2.11 - 0.39i$  and  $0.27 + 0.30i$  respectively) as well as the approximate zero modes.

### 5.3. The generalized Brillouin zone

The generalized Brillouin zone (appearing in so-called *non-Bloch band theory*) is a popular object in the physics literature; see, e.g., [22, 44–46]. Even though this work shows that it is not of relevance for the spectral analysis, let us define it as a mathematical object. Consider the map

$$\Upsilon = (\Upsilon_0, \Upsilon_1) : E \in \Lambda \mapsto (z_L(E), z_{L+1}(E)) \in \mathbb{C}^2, \quad (5.2)$$

where the roots are chosen such that both  $\Upsilon_0$  and  $\Upsilon_1$  are continuous. Then the *non-Hermitian Brillouin zone* is defined as

$$\mathbb{B} = \Upsilon_0(\Lambda) \cup \Upsilon_1(\Lambda).$$

Numerically, this can readily be plotted from the data of  $\Lambda$ . An example is given in Figure 4. In the Hermitian case,  $\Lambda = \text{spec}(H)$  and then  $(z_L(E), z_{L+1}(E)) \in \mathbb{T} \times \mathbb{T}$  for  $E \in \Lambda$ . Hence, the Brillouin zone is given by  $\mathbb{B} = \mathbb{T}$ , as usual. On the other hand, let us stress a difference with the Hermitian case for which the transfer matrix is  $\mathcal{I}$ -unitary. This latter property implies that the eigenvalues of  $\mathcal{T}^E$  lying on the unit circle have stability properties given by Krein theory (see [36] for a review in the context of transfer matrices). In particular, there can be open subsets of energies for which there are more than 2 eigenvalues on  $\mathbb{T}$  (provided that  $L > 1$ , of course). This is non-generic in the class of non-Hermitian systems where there are typically only two eigenvalues of same modulus.

### 5.4. Topological zero modes of finite-volume chiral Hamiltonians

This section considers a tridiagonal block Hamiltonian of the form (1.1) with  $L = 2$  and with symbol given by

$$H(z) = z^{-1} \begin{pmatrix} 0 & r_+ \\ r_- & 0 \end{pmatrix} + \begin{pmatrix} 0 & v_+ \\ v_- & 0 \end{pmatrix} + z \begin{pmatrix} 0 & t_+ \\ t_- & 0 \end{pmatrix}, \quad (5.3)$$

where  $r_{\pm}, v_{\pm}, t_{\pm} \in \mathbb{C}$  are parameters with  $r_{\pm} \neq 0$  and  $t_{\pm} \neq 0$ . The latter two conditions ensure that  $R$  and  $T$  are invertible so that Hypothesis A is satisfied. The model is called the *non-Hermitian SSH model* and is widely studied in the physics literature; see, e.g., [11, 22, 24]. Clearly,  $H(z)$  is off-diagonal and hence satisfies the chiral symmetry relation (3.20). Therefore, all the claims of Sections 3.4 and 4.4 hold for the Hamiltonian (5.3). In particular, the off-diagonal entries define two scalar symbols  $H_{\pm}(z) = r_{\pm}z^{-1} + v_{\pm} + t_{\pm}z$  and one can, for any scaling parameter  $s > 0$ , define winding numbers  $W_{\pm}(s) = \text{Wind}^0(H_{\pm}^s)$  which appear in the criteria in Propositions 15 and 30 (as well as Theorem 2). Furthermore, the sets  $\Lambda$  and  $\Gamma$  satisfy (4.13), namely they are reflection symmetric.

For a numerical illustration, particular values of the parameters are chosen as given in the capture of Figure 4. The first plot shows the periodic spectrum  $\Sigma = \text{spec}(H)$  as well as the spectrum of a finite-volume approximation  $H_N$  for  $N = 200$ . Again, one sees that these spectra differ considerably. However, the finite-volume spectrum is well approximated by the set  $\Lambda$  which can be computed numerically from merely the transfer matrix by the elementary method described at the end of Section 5.1 (compute the spectrum of  $\mathcal{T}^E$  for  $E$  on some grid and check whether the middle two eigenvalues have approximately same modulus). The outcome is shown in the second plot of Figure 4. These and all other plots in this paper were produced with Mathematica. The eigenvalue computations (as in the first plot of Figure 4) seem to become numerically unstable for  $N$  larger than about 300 (the plot of  $\Lambda$ , on the other hand, is very robust). From the numerical data of  $\Lambda$ , one can also readily plot the generalized Brillouin zone simply by implementing (5.2). This is shown in the third plot in Figure 4. The second plot also shows that the set  $\{E \in \Lambda : z_L(E) = z_{L+1}(E)\} \subset \mathcal{F} \cap \Lambda$  consists of merely six points given by the six ends of the graph. The local analysis of these endpoints is sketched in Remark 22 where it is argued that they are linked to the appearance of Jordan blocks in the transfer matrix. In this context, let us state that the branch points of the three legs are approximately at  $E_{\pm} = \pm(0.33 + 0.30i)$  and are *not* in  $\mathcal{F}$ , but that numerics clearly show that there are rather three different eigenvalues of same modulus at these two complex energies (it can be argued that this is a generic scenario, but an analysis of such branch points involves the modulus of the eigenvalues and thus leaves the field of algebraic geometry, other than for the end points).

Next, let us consider varying the scaling parameter  $s > 0$ . Figure 5 shows three plots of the scaled periodic spectrum  $\Sigma^s = \text{spec}(H^s)$ , with a particular focus on a transition point around  $s = 1.214$ . One clearly sees that the finite-volume spectra and thus also the set  $\Lambda$  lie in regions with non-vanishing winding number for all  $s$ , except for the zero mode that will be discussed below. The eigenstates exhibit a skin effect, and as described in Section 5.2 the sign of the winding number determines whether they are localized on the left or right side of the system. Plots of such states are unspectacular because they are typically localized only about 50 off the boundary. If  $E$  is chosen somewhat close to  $\Sigma^s$  though, the states may leak over a 100 sites into the sample (see first plot in Figure 6). Furthermore, for states on  $\Sigma^s$  or even energetic regions with vanishing winding number, the states may be extended throughout the system (as in the second plot in Figure 6).

Finally, let us address the zero modes. First of all, let us stress that in the first plot of Figure 4 as well as all plots in Figure 5 there are 2 eigenvalues close to the origin. A simple state exactly at the origin is not compatible with the spectral symmetry  $\text{spec}(H_N) = -\text{spec}(H_N)$  which follows directly from the chiral symmetry, because the Hilbert space is even dimensional and each eigenvalue  $E$  with eigenstate  $\varphi$  has a partner  $-E$  with eigenstate  $K\varphi$  (on the other hand, for the half-sided operators  $H^R$



and  $H^L$  such exact zero modes of multiplicity 1 are allowed). Whether such a pair of approximate zero modes of  $H_N$  exists can be predicted using Proposition 30. One first determines a scaling parameter  $s$  such that  $\text{Wind}^0(H^s) = 0$  (one can always find such a scaling parameter uncovering the 0). One choice is here given by  $s = 1.172$ , see the first plot of Figure 5. Then one computes numerically the two winding numbers and finds  $W_{\pm}(s) = \text{Wind}^0(H_{\pm}) = \pm 1$ . In conclusion,  $0 \in \Gamma$ . Consequently, Theorem 34 implies that  $H_N^s$  has to have states close to 0 for all  $s > 0$ . The associated eigenstates  $\varphi$  and  $K\varphi$  only differ by signs, and hence their absolute value has the same profile. If  $\text{Wind}^0(H) \neq 0$  as in the Figure 4 and the third plot in Figure 5, the sign of the winding number  $\text{Wind}^0(H)$  determines again whether the zero mode is a left or right skin state. On the other hand, if  $\text{Wind}^0(H) = 0$  as for  $s = 1.172$  in the first plot in Figure 5, the transfer matrix  $\mathcal{T}^E$  has  $L$  eigenvalues of modulus smaller than 1 and  $L$  of modulus larger than 1. Decomposing the eigenvector  $\varphi$  into one vector in the decaying and another in the expanding subspace, the relative proportion then determines skin nature of the approximate zero modes. Typically, one part dominates and the zero mode is a skin state, but in the third plot in Figure 6 the scaling  $s = 1.172$  was chosen such that these zero modes are fairly balanced between left and right edge.

**Acknowledgments.** The authors thank Carolyn Echter, Flore Kunst, and Tom Stoiber for useful and inspiring discussions at an early stage of this work.

**Funding.** This work was partially supported by the DFG grant SCHU 1358/8-1.

## References

- [1] Y. Ashida, Z. Gong, and M. Ueda, [Non-Hermitian physics](#). *Advances Phys.* **69** (2020), 249–435
- [2] E. Basor, A. Böttcher, and T. Ehrhardt, [Harold Widom’s contributions to the spectral theory and asymptotics of toeplitz operators and matrices](#). In *Toeplitz operators and random matrices – In memory of Harold Widom*, pp. 49–74, Oper. Theory Adv. Appl. 289, Birkhäuser/Springer, Cham, 2022 Zbl 07827897 MR 4573943
- [3] G. Baxter and P. Schmidt, [Determinants of a certain class of non-Hermitian Toeplitz matrices](#). *Math. Scand.* **9** (1961), 122–128 Zbl 0119.28603 MR 0126653
- [4] E. J. Bergholtz, J. C. Budich, and F. K. Kunst, [Exceptional topology of non-Hermitian systems](#). *Rev. Modern Phys.* **93** (2021), no. 1, article no. 015005 MR 4290940
- [5] D. S. Borgnia, A. J. Kruchkov, and R.-J. Slager, [Non-Hermitian boundary modes and topology](#). *Phys. Rev. Lett.* **124** (2020), no. 5, article no. 056802 MR 4068223
- [6] A. Böttcher, J. Gasca, S. M. Grudsky, and A. V. Kozak, [Eigenvalue clusters of large tetra-diagonal Toeplitz matrices](#). *Integral Equations Operator Theory* **93** (2021), no. 1, article no. 8 Zbl 1477.47023 MR 4208115

- [7] A. Böttcher, S. Grudsky, and E. Ramírez de Arellano, [On the asymptotic behavior of the eigenvectors of large banded Toeplitz matrices](#). *Math. Nachr.* **279** (2006), no. 1-2, 121–129 Zbl [1101.47015](#) MR [2193611](#)
- [8] A. Böttcher and S. M. Grudsky, *Spectral properties of banded Toeplitz matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005 Zbl [1089.47001](#) MR [2179973](#)
- [9] A. Böttcher and B. Silbermann, *Introduction to large truncated Toeplitz matrices*. Universitext, Springer, New York, 1999 Zbl [0916.15012](#) MR [1724795](#)
- [10] P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*. Progr. Probab. Statist. 8, Birkhäuser, Boston, MA, 1985 Zbl [0572.60001](#) MR [0886674](#)
- [11] J. C. Budich and E. J. Bergholtz, [Non-Hermitian topological sensors](#). *Phys. Rev. Lett.* (2020) **125**, article no. 180403
- [12] D. Damanik, A. Pushnitski, and B. Simon, The analytic theory of matrix orthogonal polynomials. *Surv. Approx. Theory* **4** (2008), 1–85 Zbl [1193.42097](#) MR [2379691](#)
- [13] S. Delvaux, [Equilibrium problem for the eigenvalues of banded block Toeplitz matrices](#). *Math. Nachr.* **285** (2012), no. 16, 1935–1962 Zbl [1276.15008](#) MR [2995424](#)
- [14] V. Dwivedi and V. Chua, [Of bulk and boundaries: Generalized transfer matrices for tight-binding models](#). *Phys. Rev. B* **93** (2016), article no. 134304
- [15] I. C. Gohberg and I. A. Fel'dman, *Convolution equations and projection methods for their solution*. Transl. Math. Monogr. 41, American Mathematical Society, Providence, RI, 1974 MR [0355675](#)
- [16] I. Y. Goldsheid and B. A. Khoruzhenko, [Eigenvalue curves of asymmetric tridiagonal random matrices](#). *Electron. J. Probab.* **5** (2000), article no. 16 Zbl [0983.82006](#) MR [1800072](#)
- [17] R. Hagen, S. Roch, and B. Silbermann, *C\*-algebras and numerical analysis*. Monogr. Textbooks Pure Appl. Math. 236, Marcel Dekker, New York, 2001 Zbl [0964.65055](#) MR [1792428](#)
- [18] N. Hatano and D. R. Nelson, [Localization transitions in non-Hermitian quantum mechanics](#). *Phys. Rev. Lett.* **77** (1996), 570-573
- [19] T. Kato, *Perturbation theory for linear operators*. Grund. Math. Wiss. 132, Springer, New York, 1966 Zbl [0148.12601](#) MR [0203473](#)
- [20] K. Kawabata, K. Shiozaki, M. Ueda, and M. Sato, [Symmetry and topology in non-Hermitian physics](#). *Phys. Rev. X* **9** (2019), article no. 041015
- [21] B. Kramer and A. MacKinnon, [Localization: theory and experiment](#). *Repor. Prog. Phys.* **56** (1993), 1469–1564
- [22] F. K. Kunst and V. Dwivedi, [Non-Hermitian systems and topology: A transfer-matrix perspective](#). *Phys. Rev. B* **99** (2019), article no. 245116
- [23] F. K. Kunst, E. Edvardsson, J. C. Budich, and E. J. Bergholtz, [Biorthogonal bulk-boundary correspondence in non-Hermitian systems](#). *Phys. Rev. Lett.* **121** (2018), article no. 026808
- [24] S. Lieu, [Topological phases in the non-Hermitian Su–Schrieffer–Heeger model](#). *Phys. Rev. B* **97** (2018), article no. 045106

- [25] L. Marin and H. Schulz-Baldes, [Scattering zippers and their spectral theory](#). *J. Spectr. Theory* **3** (2013), no. 1, 47–82 Zbl [1288.47011](#) MR [3031236](#)
- [26] N. Nikolski, [Toeplitz matrices and operators](#). Cambridge Stud. Adv. Math. 182, Cambridge University Press, Cambridge, 2020 Zbl [1466.15001](#) MR [4319036](#)
- [27] N. K. Nikolski, [Operators, functions, and systems: an easy reading](#). Vol. 1. Math. Surveys Monogr. 92, American Mathematical Society, Providence, RI, 2002 Zbl [1007.47001](#) MR [1864396](#)
- [28] N. Okuma, K. Kawabata, K. Shiozaki, and M. Sato, [Topological origin of non-Hermitian skin effects](#). *Phys. Rev. Lett.* **124** (2020), no. 8, article no. 086801 MR [4071786](#)
- [29] T. Ozawa, H. M. Price, A. Amo, N. Goldman, M. Hafezi, L. Lu, M. C. Rechtsman, D. Schuster, J. Simon, O. Zilberberg, and I. Carusotto, [Topological photonics](#). *Rev. Mod. Phys.* **91** (2019), article no. 015006
- [30] V. Peano and H. Schulz-Baldes, [Topological edge states for disordered bosonic systems](#). *J. Math. Phys.* **59** (2018), no. 3, article no. 031901 Zbl [1390.82051](#) MR [3769618](#)
- [31] E. Prodan and H. Schulz-Baldes, [Bulk and boundary invariants for complex topological insulators](#). Math. Phys. Stud., Springer, Cham, 2016 Zbl [1342.82002](#) MR [3468838](#)
- [32] C. Sadel, [Spectral theory of one-channel operators and application to absolutely continuous spectrum for Anderson type models](#). *J. Funct. Anal.* **274** (2018), no. 8, 2205–2244 Zbl [1477.81034](#) MR [3767432](#)
- [33] P. Schmidt and F. Spitzer, [The Toeplitz matrices of an arbitrary Laurent polynomial](#). *Math. Scand.* **8** (1960), 15–38 Zbl [0101.09203](#) MR [0124665](#)
- [34] H. Schulz-Baldes, [Rotation numbers for Jacobi matrices with matrix entries](#). *Math. Phys. Electron. J.* **13** (2007), article no. 5 Zbl [1141.15024](#) MR [2366118](#)
- [35] H. Schulz-Baldes, [Geometry of Weyl theory for Jacobi matrices with matrix entries](#). *J. Anal. Math.* **110** (2010), 129–165 Zbl [1197.47047](#) MR [2753292](#)
- [36] H. Schulz-Baldes, [Signature and spectral flow of  \$J\$ -unitary  \$S^1\$ -Fredholm operators](#). *Integral Equations Operator Theory* **78** (2014), no. 3, 323–374 Zbl [1325.47084](#) MR [3172254](#)
- [37] H. Schulz-Baldes, [Reduced transfer operators for singular difference equations](#). *J. Difference Equ. Appl.* **28** (2022), no. 11–12, 1492–1506 Zbl [07639057](#) MR [4527218](#)
- [38] B. Simon, [Orthogonal polynomials on the unit circle. Part 1](#). Amer. Math. Soc. Colloq. Publ. 54, American Mathematical Society, Providence, RI, 2005 Zbl [1082.42020](#) MR [2105088](#)
- [39] L. N. Trefethen and M. Embree, [Spectra and pseudospectra](#). Princeton University Press, Princeton, NJ, 2005 Zbl [1085.15009](#) MR [2155029](#)
- [40] J. L. Ullman, [A problem of Schmidt and Spitzer](#). *Bull. Amer. Math. Soc.* **73** (1967), 883–885 Zbl [0167.13101](#) MR [0219986](#)
- [41] R. Verresen, N. G. Jones, and F. Pollmann, [Topology and edge modes in quantum critical chains](#). *Phys. Rev. Lett.* **120** (2018), article no. 057001
- [42] H. Widom, [Asymptotic behavior of block Toeplitz matrices and determinants](#). *Advances in Math.* **13** (1974), 284–322 Zbl [0281.47018](#) MR [0409511](#)
- [43] H. Widom, [Asymptotic behavior of block Toeplitz matrices and determinants. II](#). *Advances in Math.* **21** (1976), no. 1, 1–29 Zbl [0344.47016](#) MR [0409512](#)

- [44] Z. Yang, K. Zhang, C. Fang, and J. Hu, [Non-Hermitian bulk-boundary correspondence and auxiliary generalized Brillouin zone theory](#). *Phys. Rev. Lett.* **125** (2020), no. 22, article no. 226402 MR 4188489
- [45] S. Yao and Z. Wang, [Edge states and topological invariants of non-Hermitian systems](#). *Phys. Rev. Lett.* **121** (2018), article no. 086803
- [46] K. Yokomizo and S. Murakami, [Non-Bloch band theory of non-Hermitian systems](#). *Phys. Rev. Lett.* **123** (2019), no. 6, article no. 066404 MR 3993741
- [47] K. Zhang, Z. Yang, and C. Fang, [Correspondence between winding numbers and skin modes in non-Hermitian systems](#). *Phys. Rev. Lett.* **125** (2020), no. 12, article no. 126402 MR 4155144
- [48] X. Zhang, T. Zhang, M.-H. Lu, and Y.-F. Chen, [A review on non-Hermitian skin effect](#). *Adv. Physics: X* **7** (2022), article no. 2109431

Received 10 March 2024; revised 2 July 2024.

**Lars Koekenbier**

Department Mathematik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstraße 11, 91058 Erlangen, Germany; [koekenbier@math.fau.de](mailto:koekenbier@math.fau.de)

**Hermann Schulz-Baldes**

Department Mathematik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstraße 11, 91058 Erlangen, Germany; [schuba@mi.uni-erlangen.de](mailto:schuba@mi.uni-erlangen.de)