UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SOME CLASSES OF JACOBI MATRICES AND SCHRÖDINGER OPERATORS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

INJO HUR
Norman, Oklahoma
2014

SOME CLASSES OF JACOBI MATRICES
AND SCHRÖDINGER OPERATORS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MATHEMATICS

BY

Dr. Christian Remling, Chair

Dr. Kim Milton

Dr. John Albert

Dr. Kevin A. Grasse

Dr. Ralf Schmidt

Acknowledgements

I would like to express my deepest appreciation to my adviser, Professor Christian Remling, who has the attitude and the substance of a genius: he continually conveyed a spirit of adventure in regard to research and scholarship. His enthusiasm and zeal have been major driving forces through my graduate career.

I would like to thank my committee members, Professor Kimball Milton, Professor John Albert, Professor Kevin Grasse and Professor Ralf Schmidt for their guidance.

A very special thank you to Professor Kyungbai Lee for all his encouragement and advice.

Thanks also go out to Professor Jaesung Lee who led me to mathematics and supported me all the time.

I would like to take this opportunity to thank all my friends and all the people in the department of mathematics at the University of Oklahoma for all their support and help over the years.

With all my heart, I would like to thank all my family, my sons, Aiden and Aaron, and, especially, my wife Misun Lee. Without their love, it would not be possible for me to finish my Ph. D. They also make me realize the true happiness in my life.

I also thank all my parents, parents-in-law, and brothers and their families in Korea for their love and support.

Contents

Abstract

This dissertation addresses two classes of Jacobi matrices and Schrödinger operators. First, we consider Jacobi matrices and Schrödinger operators that are reflectionless on an interval. We give a systematic development of a certain parametrization of this class, in terms of suitable spectral data, that is due to Marchenko. Then some applications of these ideas are discussed.

In the second half, we study structural properties of the Lyapunov exponent γ and the density of states k for ergodic (or invariant) Jacobi matrices in a general framework. In this analysis, a central role is played by the function $w = -\gamma + i\pi k$ as a conformal map between certain domains. This idea goes back to Marchenko and Ostrovskii, who used this device in their analysis of the periodic problem.

Chapter 1

Introduction and Background

1.1 Introduction

This dissertation addresses some classes of Jacobi matrices and Schrödinger operators, which are in two recent papers [24, 25].

There are many papers which show the relations between Herglotz functions and these operators. One of the most important connection between them was developed by de Branges [17, 13, 14, 15, 16]: for given any Herlgotz function, there is a (unique) canonical system whose Titchmarsh-Weyl m-function is the given function where

$$Ju'(x,z) = zH(x)u(x,z), \tag{1.1.1}$$

where $H(x)$ is a nonnegative definite 2×2 matrix whose entries are real-valued, locally integrable functions and $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. For the uniquesness, we need the trace-normed condition, $tr(H(x)) = 1$ for all $x \in \mathbb{R}$. See [51] for more details.

It turns out that any eigenvalue equations by Jacobi matrices or Schrödinger operators can be changed to (trace-normed) canonical systems. With keeping this big picture in our mind, our interests are on two subclasses of canonical systems, one of which is the set of reflectionless Jacobi matrices or Schrödinger operators in Chapter 2 and the other of which is the set of (random) Jacobi matrices in Chapter 3. Since we will talk about two completely different classes, we will give an introduction at the beginning of each chapter.

1.2 Herglotz Functions

In this section we have a brief review of Herglotz functions without any proof. These functions will be the basic ingredients later. For more details, see [49].

A holomophic function F from the upper half plane, $\mathbb{C}^+$, to itself is called a *Herglotz* function. Then it is well known that a Herglotz function has the following integral representation:

$$F(z) = a + bz + \int_{\mathbb{R}} \Big(\frac{1}{t-z} - \frac{t}{1+t^2}\Big) d\rho(t) \qquad (1.2.1)$$

where $a \in \mathbb{R}$ (i.e., a is a real number), $b \geq 0$, and ρ is a (nonzero) positive Borel measure on $\mathbb{R}$ which satisfies $\int_{\mathbb{R}} \frac{d\rho(t)}{1+t^2} < \infty$. Moreover, the triple a, b and ρ is determined by F by

$$a = \text{Re } \big(F(i)\big), \quad \lim_{y \to \infty} \frac{F(iy)}{iy} = b \geq 0$$

where $z = x + iy$, and Stieltjes inversion formula

$$\rho\big((t_0, t_1]\big) = \lim_{\delta \downarrow 0} \lim_{y \downarrow 0} \frac{1}{\pi} \int_{t_0+\delta}^{t_1+\delta} \text{Im } \big(F(t+iy)\big) dt$$

where Im F is the imaginary part of F. We may think of Herglotz functions as holomorphic functions on the unit disc of the version of the upper half plane.

We now address some properties of Herglotz functions.

Proposition 1.2.1. *Let F be a Herglotz function with associated measure ρ in (1.2.1). Then we have*

$$\rho(t) = w^* - \lim_{y \downarrow 0} \frac{1}{\pi} Im\ F(t+iy)$$

where $w^* - \lim$ *means the weak-$*$ limit.*

It is well known that these Herglotz functions have boundary values almost everywhere: the normal limit, $F(t) \equiv \lim_{y \downarrow 0} F(t+iy)$, exists almost everywhere on $\mathbb{R}$. We then have a strong property of Herglotz function on the boundary values.

Proposition 1.2.2. *If two Herglotz functions have the same boundary values on any subset of* $\mathbb{R}$ *of positive Lebesgue measure, then they are the same.*

Given any Herglotz function F, $\ln F$ is also a Herglotz function if we choose $\operatorname{Arg} F$ is between $-\pi$ and π. By the above Herglotz representation,

$$F(z) = C \exp \Big(\int_{\mathbb{R}} \Big(\frac{1}{t-z} - \frac{t}{1+t^2} \Big) \xi(t) dt \Big)$$

where $C = \ln |F(i)| \in \mathbb{R}$ and $\xi(t) = \frac{1}{\pi} \lim_{y \downarrow 0} \operatorname{Im} \big(\ln F(t+iy) \big)$ with $\int_{\mathbb{R}} \frac{\xi(t)dt}{1+t^2} < \infty$. Note that $0 \le \xi(t) \le 1$.

1.3 Titchmarsh-Weyl m-Functions

A (right-half-line) *Jacobi matrix* is a difference operator on $u \in \ell^2(\mathbb{N})$ of the form

$$(J_+ u)_n = a_n u_{n+1} + a_{n-1} u_{n-1} + b_n u_n \quad \text{for } n \ge 2 \tag{1.3.1}$$

and

$$(J_+ u)_1 = a_1 u_2 + b_1 u_1 \quad \text{for } n = 1.$$

Alternatively, one can represent J_+ by the following tridiagonal matrix with respect to the standard basis of $\ell^2(\mathbb{N})$:

$$J_+ = \begin{pmatrix} b_1 & a_1 & & & \\ a_1 & b_2 & a_2 & & \\ & a_2 & b_3 & a_3 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

Here, $a_n > 0$ and $b_n \in \mathbb{R}$, and we also assume that $a, b \in \ell^\infty(\mathbb{N})$. Under these assumptions, J_+ is a bounded self-adjoint operator on $\ell^2(\mathbb{N})$. In particular, the spectrum of J_+ belongs to $\mathbb{R}$.

Introduce the difference expression τ by

$$(\tau u)_n = a_n u_{n+1} + a_{n-1} u_{n-1} + b_n u_n.$$

Formally, this looks the same as J_+, but we will apply τ to *arbitrary* sequences u, not necessarily from ℓ^2. To evaluate $(\tau u)_1$, we need a_0, and we can assign an arbitrary (positive) value, say $a_0 = 1$.

Then it turns out that, for any $z \in \mathbb{C}\backslash\mathbb{R}$, $(\tau - z)u = 0$ has exactly one linearly independent solution $u \in \ell^2(\mathbb{N})$. In other words, since J_+ is bounded, a limit-point-case only occurs.

Let $p_n(z)$, $q_n(z)$ be the solutions of $(\tau - z)u = 0$ with the initial values

$$\begin{aligned} a_0 p_0(z) &= 0 \qquad & a_0 q_0(z) &= -1 \\ p_1(z) &= 1 \qquad & q_1(z) &= 0. \end{aligned}$$

By iterating the difference equation, we see that for fixed $n \in \mathbb{N}$, $p_n(z)$ and $q_n(z)$ are polynomials in z of degree $n-1$ and $n-2$, respectively.

Then we define a Titchmarsh-Weyl m-function, $m_+(z)$, by

$$f_n(z) = q_n(z) + m_+(z)p_n(z) \in \ell^2(\mathbb{N}) \tag{1.3.2}$$

for $z \in \mathbb{C}^+$. Note that m_+ is uniquely determined because we have only one linearly independent square-summable solution. By (1.3.2) we have that for $z \in \mathbb{C}^+$

$$m_+(z) = -\frac{f_1(z)}{a_0 f_0(z)}. \tag{1.3.3}$$

Moreover, we can show the following.

Proposition 1.3.1. *For any $z \in \mathbb{C}^+$*

$$m_+(z) = \langle \delta_1, (J_+ - z)^{-1}\delta_1 \rangle \tag{1.3.4}$$

where $\delta_1 = (1, 0, \cdots)$ and $\langle\ ,\ \rangle$ is a sesquilinear form over $\mathbb{C}$.

In other words, m_+ is the (1,1)-entry of the Green function of J_+.

Proof. Let f_n be as in (1.3.2) and let $g = (J_+ - z)^{-1}\delta_1$. Then $(J_+ - z)g = \delta_1$, so $((\tau - z)g)_n = 0$ for $n \geq 0$. Moreover, $g \in \ell^2(\mathbb{N})$. Then g is a constant multiple of f since we have only one linearly independent ℓ^2-solution. By comparing the values at $n = 1, 2$, we have (1.3.4). □

By the functional calculus, (1.3.4) shows that

$$m_+(z) = \int_{\mathbb{R}} \frac{d\mu(t)}{t - z}, \quad d\mu(t) = d||E(t)\delta_1||^2 \tag{1.3.5}$$

where E denotes the spectral resolution of J_+. In particular, m_+ is a Herglotz function.

Similarly, for a (half-line) Schrödinger operator we can define a Titchmarsh-Weyl m-function m_+ by

$$m_+(z) = \frac{f_+'(0,z)}{f_+(0,z)} \tag{1.3.6}$$

where f_+ is a square-integrable solution near $+\infty$. In the Schrödinger case, (1.3.6) may not be uniquely determined because we may have two linearly-independent solutions to the eigenvalue equation for Schrödinger operators. However, in our later setting, we can always determine m uniquely because we are interested in Schrödinger operators that have absolutely continuous spectrum and the existence of absolutely continuous spectrum implies that there is only one square-integrable solution.

We are interested in Titchmarsh-Weyl m-functions because they determine our Jacobi matrices or Schrödinger operators.

Theorem 1.3.2 (Borg[6], Marchenko[30])**.** *m_+ determines a (half-line) Schrödinger operator and the boundary condition at 0.*

In (bounded) Jacobi matrices, this kind of theorem turns out to be much easier.

Chapter 2

Marchenko Representation of Reflectionless Jacobi and Schrödinger operators on one interval

2.1 Introduction

We are interested in one-dimensional (whole-line) Schrödinger operators,

$$(Hy)(x) = -y''(x) + V(x)y(x), \tag{2.1.1}$$

with locally integrable potentials V that are in the limit point case at $\pm\infty$ (i.e., there is only one linearly independent square-integrable solution near $\pm\infty$, respectively) and in (whole-line) Jacobi matrices,

$$(Ju)_n = a_n u_{n+1} + a_{n-1}u_{n-1} + b_n u_n. \tag{2.1.2}$$

Here we assume that $a, b \in \ell^\infty(\mathbb{Z})$, $a_n > 0$, $b_n \in \mathbb{R}$.

The reason why we assume that H is in the limit point case is that we are interested in reflectionless operators. In particular, they have absolutely continuous spectrum. If H is in the limit circle case (or equivalently the eigenvalue equation corresponding to H has two linearly independent square-integrable solutions), then H has purely discrete spectrum.

These operators have associated half line m functions $m_\pm$ by (1.3.6) and a similar way for m_-. In other words, by cutting the whole line $\mathbb{R}$ at 0 and putting the Dirichlet boundary condition at 0, we have two half lines, each of which corresponds to $m_\pm$, respectively. These are *Herglotz functions,* as we discussed in section 1.3.

We call an operator *reflectionless* on a Borel set $S \subset \mathbb{R}$ of positive Lebesgue measure if $m_\pm$ satisfy the following identity

$$m_+(x) = -\overline{m_-(x)} \qquad \text{for (Lebesgue) a.e. } x \in S. \tag{2.1.3}$$

Here $m_\pm(x) \equiv \lim_{y \downarrow 0} m_\pm(x+iy)$. Since the normal limit of Herglotz functions exist almost everywhere, (2.1.3) is well defined. It is well known that (2.1.3) is independent of the boundary condition at 0.

For example, periodic Jacobi and Schrödinger operators (i.e., a_n and b_n are periodic or V is periodic) and finite-gap Jacobi matrices are reflectionless operators on their spectra, which are well studied. Reflectionless operators are important because they can be thought of as the fundamental building blocks of arbitrary operators with some absolutely continuous spectrum. See [27, 41, 43].

Reflectionless operators have remarkable properties. For example, knowledge of the coefficients on any half line is enough to recover all the coefficients, and the reflectionless property is shift-invariant in the sense that we can break down our whole interval at any point to have two half lines. If an operator is reflectionless on an *interval* (rather than a more complicated set), one can say even more. So these operators are of special interest.

Marchenko [31] developed a certain parametrization of the class $\mathcal{M}_R$ of Schrödinger operators H that are reflectionless on $(0, \infty)$ and have spectrum contained in $[-R^2, \infty)$. It is in fact easy in principle to give such a parametrization in terms of certain spectral data, which has been used by many authors [12, 37, 38, 43, 49]. We will briefly review this material in Section 2.2. Marchenko's parametrization is different, and it makes certain properties of reflectionless Schrödinger operators very transparent. Some of these applications will be discussed below.

We have two general goals in this chapter. First, we present a direct and easy approach to Marchenko's parametrization that starts from scratch and does not use any machinery. Marchenko's treatment relies on inverse scattering theory as its main tool and is rather intricate. We hope that our approach will help put things in their proper context; among other things, it will explain the role of the inequalities imposed on the representing measures σ. We will also extend these ideas to the discrete setting; in fact, we will start with this case as some technical issues from the continuous setting are absent here. The second goal is to explore some consequences and applications of Marchenko's parametrization, in the form developed here. We will have more to say about this towards the end of this introduction.

The basic ideas of the Marchenko parametrizations are easy to describe. If S is an interval, then it is well known (compare, for example, [28, Corollary 2]) that (2.1.3) guarantees the existence of a genuine holomorphic continuation of m_+ through S (this is not an immediate consequence of the Schwarz reflection principle because of the possible presence of an exceptional Lebesgue measure-zero set where (2.1.3) fails). More precisely, we have the following (the proof will be reviewed in Section 2.2).

Lemma 2.1.1. *Fix an open interval $S = (a, b)$, and let m_+ be a Herglotz function. Then m_+ satisfies* (2.1.3) *for $S = (a, b)$ (for some Herglotz function m_-) if and only if m_+ has a holomorphic continuation*

$$M : \mathbb{C}^+ \cup S \cup \mathbb{C}^- \to \mathbb{C}^+.$$

Note that there are two conditions really: m_+ must have a continuation M

to $\Omega = \mathbb{C}^+ \cup S \cup \mathbb{C}^-$, and, moreover, M must map all of Ω to $\mathbb{C}^+$. However, these properties are immediate consequences of the fact that if $S = (a, b)$, then the exceptional null set from (2.1.3) is empty, so this is what the Lemma really says.

This continuation M is necessarily given by $M(z) = -\overline{m_-(\overline{z})}$ on the lower half plane $z \in \mathbb{C}^-$. In other words, (2.1.3) for $S = (a, b)$ lets us combine m_+ and m_- into one holomorphic function M on the simply connected domain Ω. We can then introduce a conformal change of variable $z = \varphi(\lambda)$, $\varphi : \mathbb{C}^+ \to \Omega$, to obtain a new Herglotz function $F(\lambda) \equiv M(\varphi(\lambda))$. The measures from the Herglotz representations of these functions F will be the data that we will use to parametrize the operators from the Marchenko class $\mathcal{M}_R$.

Let us now discuss some applications. As an immediate minor pay-off, we obtain a very quick new proof of [42, Theorem 1.2], which is now seen to be an immediate consequence of our Theorem 2.3.1 below. Recall that this result states that if a Jacobi matrix is bounded and reflectionless on $(-2, 2)$, then $a_n \geq 1$ for all $n \in \mathbb{Z}$, and if $a_{n_0} = 1$ for a single $n_0 \in \mathbb{Z}$, then $a_n \equiv 1$, $b_n \equiv 0$. In Proposition 2.3.3 we try to indicate how these ideas could, perhaps, be carried further.

More importantly, the material from Section 2.4 yields continuous analogs of these results. Here are three such consequences of the Marchenko parametrization, combined with the material from [41]. We are now interested in half line Schrödinger operators H_+ on $L^2(0, \infty)$ satisfying the following assumptions:

Hypothesis 2.1.1. $\Sigma_{ac}(H_+) \supset (0, \infty)$ and V is uniformly locally integrable, that is,

$$\sup_{x \geq 0} \int_x^{x+1} |V(t)| \, dt < \infty. \tag{2.1.4}$$

Here, Σ_{ac} denotes an essential support of the absolutely continuous part of the spectral measure of H_+. In other words, we are assuming that $\chi_{(0,\infty)}(E)\, dE \ll d\rho_{ac}(E)$. This implies that, but is not equivalent to $\sigma_{ac}(H_+) \supset [0, \infty)$. An H_+ satisfying Hypothesis 2.1.1 can, of course, have embedded singular spectrum in $(0, \infty)$, and can have arbitrary spectrum outside this set. Notice also that (2.1.4) implies that H_+ is a limit point case and bounded below.

To obtain self-adjoint operators, one has to impose a boundary condition at $x = 0$, but since Σ_{ac} is independent of this boundary condition, we will not make it explicit here.

Let us now state two closely related sample results.

Theorem 2.1.2. *Assume Hypothesis 2.1.1. Then*

$$\limsup_{x\to\infty} \int V(x+t)\varphi(t)\, dt \le 0 \tag{2.1.5}$$

for every compactly supported, continuous function $\varphi \ge 0$.

This says that in the situation described by Hypothesis 2.1.1, the positive part of V will go to zero, in a weak sense.

Theorem 2.1.3. *Assume Hypothesis 2.1.1. If, in addition,* $V \ge 0$ *on* $\bigcup(x_n - d, x_n + d)$ *for some increasing sequence* $x_n \to \infty$ *with bounded gaps (that is,* $\sup(x_{n+1} - x_n) < \infty$*) and some* $d > 0$*, then*

$$\lim_{x\to\infty} \int V(x+t)\varphi(t)\, dt = 0 \tag{2.1.6}$$

for every compactly supported, continuous function φ.

Theorem 2.1.3 is a variation on the (continuous) Denisov-Rakhmanov Theorem [18, 41]. Recall that the DR Theorem asserts that (2.1.6) will follow if, in addition to Hypothesis 2.1.1, we have that $\sigma_{ess}(H_+) = [0, \infty)$, In Theorem 2.1.3, we replace this latter assumption by partial information on V; more precisely, we assume here that V is non-negative every once in a while, with positive frequency.

Theorem 2.1.4. *Assume Hypothesis 2.1.1. We are given $d > 0$ (arbitrarily small) and $\epsilon > 0$ and (arbitrarily many) compactly supported, continuous test functions $\varphi_1, \ldots, \varphi_N$. Then there exist $x_0 > 0$ and $\delta > 0$ so that the following holds: If $x \geq x_0$ and $V(t) \geq -\delta$ for $|t - x| < d$, then*

$$\left| \int V(t)\varphi_j(t-x)\, dt \right| < \epsilon$$

for $j = 1, \ldots, N$.

In particular, this conclusion is obtained if $V \geq 0$ on $|t - x| < d$, in which case δ becomes irrelevant.

This is an Oracle Theorem type statement that, roughly speaking, says that if V is almost non-negative anywhere, then V has to be close to zero on a very long interval centered at that point (not in a pointwise sense, though).

Let us now discuss a completely different application of the Marchenko parametrization. Call a half line operator H_+ or J_+ (on $L^2(0,\infty)$ or $\ell^2(\mathbb{Z}_+)$, respectively) reflectionless on S if the corresponding m function m_+ satisfies (2.1.3) for some (unique, if it exists at all) Herglotz function m_-.

Reflectionless half line operators may, of course, be obtained by restricting reflectionless whole line problems. Since reflectionless operators may be reconstructed from arbitrary half line restrictions, we can actually think of such a half line restriction as just another representation of the original whole line problem. Perhaps somewhat surprisingly, however, there are other examples:

Theorem 2.1.5. *(a) There exists a half line Jacobi matrix J_+ that is reflec-*

tionless on $(-2,2)$*, but is not the restriction of a reflectionless whole line Jacobi matrix.*
(b) There exists a half line Schrödinger operator H_+ *that is reflectionless on* $(0,\infty)$*, but is not the restriction of a reflectionless whole line Schrödinger operator.*

Put differently, the associated m function m_- that is obtained from m_+ via (2.1.3) is not the m function of a Jacobi matrix or Schrödinger operator, respectively. The examples we will construct to prove Theorem 2.1.5 will be quite explicit, especially in the discrete case; they will satisfy $\sigma(J_+) = [-2,2]$, $\sigma(H_+) = [0,\infty)$, so it is not spectrum outside S (there is not any) that produces this effect. We will see below that Theorem 2.1.5 is in fact a rather quick consequence of the Marchenko parametrization.

2.2 Preliminaries

We briefly review some standard material about certain spectral data that are particularly convenient if one wants to discuss reflectionless operators. See [38, 43] for a more comprehensive discussion.

Given a pair of Herglotz functions $m_\pm$ that satisfies (2.1.3), consider $H = m_+ + m_-$. Since this is another Herglotz function, we can take a holomorphic logarithm, which is a Herglotz function itself, if we agree that $\operatorname{Im} \ln H \in (0,\pi)$, say. The *Krein function* of H is then defined (almost everywhere, with respect to Lebesgue measure) by

$$\xi(x) = \frac{1}{\pi} \lim_{y\to 0+} \operatorname{Im} \ln H(x+iy).$$

We have that $0 \le \xi \le 1$, and (2.1.3) implies that $\xi = 1/2$ a.e. on S. Next, if

$$H(z) = A + Bz + \int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) d\rho(t)$$

is the Herglotz representation of H, then it is easy to verify (see, for example, [43, Section 5] for the details) that

$$m_+(z) = A_+ + B_+ z + \int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) f(t)\, d\rho(t), \tag{2.2.1}$$

and here $0 \le B_+ \le B$, $0 \le f \le 1$, $f = 1/2$ Lebesgue-a.e. on S. Here we can think of f as a measurable function to distribute the measure ρ to $m_\pm$.

Conversely, these data determine an m_+ that will satisfy (2.1.3). More explicitly, if measurable functions ξ, f with $0 \le \xi, f \le 1$ and $\xi = f = 1/2$ a.e. on S are given, and if we also choose three constants $C > 0$, $0 \le c \le 1$, $A_+ \in \mathbb{R}$, then ξ and C first of all determine a unique H with $|H(i)| = C$. We in fact have the explicit formula

$$H(z) = C \exp\left[\int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) \xi(t)\, dt \right]. \tag{2.2.2}$$

Then (3.2.8) with $B_+ = cB$ defines an m_+, which will satisfy (2.1.3), with $m_- = H - m_+$. Any m_+ satisfying (2.1.3) is obtained in this way.

Let us now sketch the proof of Lemma 2.1.1.

Proof of Lemma 2.1.1. Obviously, if M is as in the lemma, then (2.1.3) holds, with $m_-(z) := -\overline{M(\overline{z})}$ $(z \in \mathbb{C}^+)$.

Conversely, assume that (2.1.3) holds with $S = (a, b)$. Since it suffices to prove the claim for arbitrary bounded subintervals of S, we may assume that S itself is bounded. Now consider H, defined as above. As observed earlier, its

Krein function satisfies $\xi = 1/2$ a.e. on S. Since

$$\frac{1}{2}\int_a^b \frac{dt}{t-z},$$

originally defined for $z \in \mathbb{C}^+$, has a holomorphic continuation through (a, b) (evaluate the integral!), we see from the exponential Herglotz representation (2.2.2) that H itself has the same property. Now (3.2.8) makes it clear that m_+ has such a holomorphic continuation, too. Here we use the fact that in the situation under consideration, ρ cannot have a singular part on (a, b); this follows immediately from our earlier observation that H can be holomorphically continued through this interval.

By (2.1.3), this continuation of m_+ must be given by $M(z) = -\overline{m_-(\overline{z})}$ for $z = x - iy$, $a < x < b$, $y > 0$ and small, so we can actually continue to all of $\mathbb{C}^-$ and this continuation clearly maps $\mathbb{C}^+ \cup \mathbb{C}^-$ to $\mathbb{C}^+$, and $\operatorname{Im} M(x) \geq 0$ for $x \in S$. The proof is now finished by observing that the open mapping theorem gives us strict inequality here. □

2.3 The Discrete Case: Jacobi Matrices

We are now interested in Jacobi matrices J on $\ell^2(\mathbb{Z})$ that are reflectionless on $S = (-2, 2)$ and satisfy $\|J\| \leq R$ for some $R \geq 2$. We will denote the collection of these Jacobi matrices by $\mathcal{M}_R$.

As we discussed in Chapter 1, we have two m-functions $m_\pm$ by the following: for $z \in \mathbb{C}^+$,

$$m_\pm(z) = \mp \frac{f_\pm(1, z)}{a_0 f_\pm(0, z)}$$

where $f(\cdot, z)$ are the solutions of Jacobi equation that are square summable near $\pm\infty$. We are assuming (2.1.3) on $(-2, 2)$, so by Lemma 2.1.1, we can combine

$m_\pm$ into one function $M : \Omega \to \mathbb{C}^+$, $\Omega = \mathbb{C}^+ \cup (-2,2) \cup \mathbb{C}^-$. Off the interval $(-2,2)$, M is given by

$$M(z) = \begin{cases} m_+(z) & z \in \mathbb{C}^+ \\ -\overline{m_-(\overline{z})} & z \in \mathbb{C}^- \end{cases}. \tag{2.3.1}$$

Following our earlier outline, we now want to introduce a conformal change of variable $\varphi : \mathbb{C}^+ \to \Omega$. We will work with the specific map

$$\varphi(\lambda) = -\lambda - \frac{1}{\lambda}.$$

In the subsequent developments, it is useful to keep in mind that φ maps the upper half of the unit circle onto $(-2,2)$. The upper semi-disk is mapped onto $\mathbb{C}^+$, while the complement (in $\mathbb{C}^+$, of the closed disk) goes to $\mathbb{C}^-$ under φ. (Of course, φ is defined by the formula given for arbitrary $\lambda \neq 0$, and we will frequently make use of this extended map without further comment.)

As anticipated, we now define the new Herglotz function

$$F(\lambda) = M(\varphi(\lambda)) \qquad (\lambda \in \mathbb{C}^+).$$

It will also be convenient to let r denote the solution $r + 1/r = R$ with $0 < r \leq 1$; this is well defined because we are assuming that $R \geq 2$. Also, we will write $\sigma_n \equiv \int t^n \, d\sigma(t)$ for the (generalized) moments of a measure σ, for $n \in \mathbb{Z}$. (These are well-defined because the corresponding spectra are away from 0 and ∞.)

Theorem 2.3.1. *$J \in \mathcal{M}_R$ if and only if the associated F function is of the form*

$$F(\lambda) = -\sigma_{-1} + (1 - \sigma_{-2})\lambda + \int \frac{d\sigma(t)}{t - \lambda}, \tag{2.3.2}$$

for some finite Borel measure σ on $(-1/r,-r)\cup(r,1/r)$ that satisfies

$$1-\sigma_{-2}+\int\frac{d\sigma(t)}{t^2+Et+1}>0 \tag{2.3.3}$$

for all $|E|>R$.

To spell this out even more explicitly, this says that if $J\in\mathcal{M}_R$, then the associated F will have a representation of the form (3.3.3), with a σ that has the stated properties. It is also clear that we have uniqueness: J determines $m_\pm$ and thus F and σ. Conversely, if a measure σ satisfies (3.3.2) (and is supported on the set given), then (3.3.3) defines a function that is the F function of a unique $J\in\mathcal{M}_R$.

In other words, Theorem 2.3.1 sets up a one-to-one correspondence between $J\in\mathcal{M}_R$ and the measures σ on $r<|t|<1/r$ satisfying (3.3.2).

If we are not interested in the specific value of $\|J\|$, then we may interpret Theorem 2.3.1 as setting up a one-to-one correspondence between bounded, reflectionless (on $(-2,2)$) Jacobi matrices and measures σ that are supported by a compact subset of $\mathbb{R}\setminus\{0\}$ and satisfy $\sigma_{-2}<1$. To obtain this version, it suffices to observe that the integral from (3.3.2) goes to zero as $|E|\to\infty$.

The proof will depend on the asymptotic properties of $m_\pm$ for a Jacobi matrix, so we briefly review these first. See, for example, [49, Ch. 2] for this material.

For any J with $\|J\|\le R$, we have that

$$m_+(z)=\int\frac{d\rho_+(t)}{t-z} \tag{2.3.4}$$

$$a_0^2m_-(z)=z-b_0+a_{-1}^2\int\frac{d\rho_-(t)}{t-z}, \tag{2.3.5}$$

and here $\rho_\pm$ are probability (Borel) measures supported by $[-R,R]$. Conversely, if we are given such data (that is, we are given two compactly supported probability measures $\rho_\pm$ and numbers $a_0, a_{-1} > 0$, $b_0 \in \mathbb{R}$), then there will be a bounded whole line Jacobi matrix J with half line m functions given by (2.3.4), (2.3.5). Moreover, if both ρ_+ and ρ_- have infinite supports, then this J will be unique.

Whether or not a given Herglotz function has a representation of this type can be decided by looking at the large z asymptotics:

Lemma 2.3.2. *Let g be a Herglotz function and let $a > 0$. Then*

$$g(z) = \int \frac{d\rho(t)}{t-z}, \qquad \rho(\mathbb{R}) = a$$

for some finite measure ρ if and only if $\lim_{y\to\infty} yg(iy) = ia$.

Proof. If g has such a representation, then $yg(iy) \to ia$ follows immediately from dominated convergence. To prove the converse, write down the (general) Herglotz representation of g:

$$g(z) = A + Bz + \int \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) d\rho(t) \tag{2.3.6}$$

Then

$$y \operatorname{Im} g(iy) = By^2 + \int \frac{y^2}{t^2+y^2}\, d\rho(t).$$

By monotone convergence, the integral converges to $\rho(\mathbb{R})$, so it follows that $\rho(\mathbb{R}) = a$ and $B = 0$. In particular, we know now that ρ is finite, so we may split the integral from (3.7.7) into two parts and, using the hypothesis again, we then conclude that $A - \int t/(t^2+1)\, d\rho(t) = 0$. □

We are now ready for the

Proof of Theorem 2.3.1. We first show that F, σ have the asserted properties if $J \in \mathcal{M}_R$. Recall first of all that $m_\pm$ have holomorphic continuations to a neighborhood of $(-\infty, -R) \cup (R, \infty)$. (This continuation of m_+ will, of course, be different from the continuation M of the same function, where the domains overlap. In particular, the lower half plane will be mapped to itself, rather than the upper half plane.) This follows because $\rho_\pm$ are supported by $[-R, R]$. As a consequence, F can be holomorphically continued through $\mathbb{R} \setminus \{t : r \leq |t| \leq 1/r\}$; indeed, the set removed contains all those $t \in \mathbb{R}$ that get mapped to $[-R, R]$ under the map φ. At $t = 0$, we need to argue slightly differently: F can be holomorphically to a neighborhood of this point because $m_+(z)$ is holomorphic at $z = \infty$. We will discuss this in more detail shortly.

So, if we now write down the Herglotz representation of F, then the representing measure σ will be supported by $\{t : r \leq |t| \leq 1/r\}$. In particular, such a σ is finite, so we may again split off the $t/(t^2 + 1)$ term in (3.7.7) and absorb it by A. We arrive at the following representation:

$$F(\lambda) = A + B\lambda + \int \frac{d\sigma(t)}{t - \lambda} \tag{2.3.7}$$

We can now identify A, B by comparing the asymptotics of this function, as $\lambda \to 0$, with those of m_+. Indeed, if $\lambda \in \mathbb{C}^+$ is close to zero, then $\varphi(\lambda) \in \mathbb{C}^+$, so $F(\lambda) = m_+(\varphi(\lambda))$ for these λ, and (2.3.4) shows that

$$m_+(\varphi(\lambda)) = -\frac{1}{\varphi(\lambda)} + O(\lambda^2) = \lambda + O(\lambda^2).$$

This confirms that $\sigma(\{0\}) = 0$, as claimed earlier. We then see from a Taylor expansion of (2.3.7) that

$$F(\lambda) = A + \sigma_{-1} + (B + \sigma_{-2})\lambda + O(\lambda^2).$$

It follows that $A = -\sigma_{-1}$ and $B = 1 - \sigma_{-2}$, as asserted in (3.3.3).

To obtain (3.3.2), we take a look at the function $H(z) = m_+(z) + m_-(z)$. As observed above, in the proof of Lemma 2.1.1, H has a holomorphic continuation through $(-2, 2)$. Equivalently, the function $h(\lambda) = H(\varphi(\lambda))$, originally defined for $\lambda \in \mathbb{C}^+$, $|\lambda| < 1$, may be holomorphically continued through the upper half of the unit circle. On $|\lambda| = 1$, we can obtain this continuation as

$$h(\lambda) = F(\lambda) - \overline{F(\lambda)}$$

and since $\overline{\lambda} = 1/\lambda$ for these λ, this gives

$$\begin{aligned} h(\lambda) &= B\left(\lambda - \frac{1}{\lambda}\right) + \int \left(\frac{1}{t-\lambda} - \frac{1}{t - 1/\lambda}\right) d\sigma(t) \\ &= \left(\lambda - \frac{1}{\lambda}\right)\left(1 - \sigma_{-2} + \int \frac{d\sigma(t)}{(t-\lambda)(t-1/\lambda)}\right). \end{aligned} \tag{2.3.8}$$

Since the right-hand sides are analytic functions of λ, these formulae hold for all $\lambda \in \mathbb{C}^+$, $|\lambda| \leq 1$. It is useful to observe here that $h_0 = \lambda - 1/\lambda$ is the H function of the free Jacobi matrix $a_n \equiv 1$, $b_n \equiv 0$. Now $a_0^2 H(z) = -1/g(z)$, where $g(z) = \langle \delta_0, (J-z)^{-1}\delta_0 \rangle$ is the Green function of J at $n = 0$. This implies that $H(x) < 0$ for $x < -R$ (to the left of the spectrum) and $H(x) > 0$ for $x > R$. Since h_0 already has the correct signs, this forces the last factor from (2.3.8) to be positive for $|E| > R$. This gives (3.3.2).

Finally, observe that (3.3.2) also prevents point masses at $t = \pm r$, $t = \pm 1/r$, so σ is indeed supported by the (open) set given in the Theorem. For example, if we had $\sigma(\{r\}) > 0$, then the integral from (3.3.2) would diverge to $-\infty$ as $E \to -R$, $E < -R$.

Conversely, assume now that a measure σ on $(-1/r, -r) \cup (r, 1/r)$ satisfying (3.3.2) is given. We want to produce a $J \in \mathcal{M}_R$ so that this σ represents its F

function. It is clear how to proceed: define F by (3.3.3) and let

$$m_+(\varphi(\lambda)) = F(\lambda) \qquad (|\lambda| < 1, \lambda \in \mathbb{C}^+), \tag{2.3.9}$$

$$m_-(\varphi(\lambda)) = -\overline{F(\overline{\lambda})} \qquad (|\lambda| > 1, \lambda \in \mathbb{C}^-). \tag{2.3.10}$$

Since φ maps both of these domains conformally onto $\mathbb{C}^+$, this defines two Herglotz functions $m_\pm$. As the first step, to just obtain a Jacobi matrix J from $m_\pm$, we have to verify that these functions satisfy (2.3.4), (2.3.5).

So let $y > 0$ (typically large), and let $s > 0$ be the unique positive solution of $1/s - s = y$. Then $\varphi(is) = iy$ and $s = 1/y + O(1/y^3)$. Thus a Taylor expansion of (3.3.3) shows that m_+, defined by (2.3.9), satisfies $m_+(iy) = i/y + O(y^{-2})$. Indeed, since

$$\int \frac{d\sigma(t)}{t - is} = \sigma_{-1} + i\sigma_{-2}s + O(s^2),$$

we have that $F(is) = is + O(s^2) = i/y + O(y^{-2}) = m_+(iy)$ by the asymptotic $s = 1/y + O(y^{-3})$. Lemma 2.3.2 implies that m_+ satisfies (2.3.4), with $\rho_+(\mathbb{R}) = 1$. In fact, ρ_+ is supported by $[-R, R]$. This follows because the definition (2.3.9) also makes sure that $m_+(z)$ can be holomorphically continued through the complement (in $\mathbb{R}$) of this interval.

Similarly, for large positive t, we have that

$$-\overline{F(it)} = i(1 - \sigma_{-2})t + \sigma_{-1} + \frac{i\sigma_0}{t} + O(t^{-2}).$$

As before, take $t > 1$ to be the solution of $\varphi(it) = iy$ for (large) $y > 0$. It then follows that m_-, defined by (2.3.10), satisfies

$$m_-(iy) = i(1 - \sigma_{-2})y + \sigma_{-1} + i\frac{1 - \sigma_{-2} + \sigma_0}{y} + O(y^{-2}) \qquad (y \to \infty). \tag{2.3.11}$$

Indeed, since $t = y + 1/y + O(y^{-2})$, $1/t = 1/y + O(y^{-2})$, and $-\overline{F(it)} = i(1 - \sigma_{-2})t + \sigma_{-1} + \frac{i\sigma_0}{t} + O(t^{-2})$, we have (2.3.11). We can now again refer to Lemma 2.3.2 to conclude that m_- satisfies (2.3.5), with

$$a_0 = (1 - \sigma_{-2})^{-1/2}, \quad b_0 = -\frac{\sigma_{-1}}{1 - \sigma_{-2}}. \tag{2.3.12}$$

Note in this context that (3.3.2) implies that $1 - \sigma_{-2} > 0$. So (2.3.12) does define coefficients $a_0 > 0$, $b_0 \in \mathbb{R}$. By suitably defining $a_{-1} > 0$, we can then guarantee that $\rho_-(\mathbb{R}) = 1$. As above, we also see that ρ_- is in fact supported by $[-R, R]$.

By the material reviewed at the beginning of this chapter, we obtain a unique Jacobi matrix J from the pair $m_\pm$. It is indeed unique because $\rho_\pm$ are equivalent to Lebesgue measure on $(-2, 2)$, so are certainly not supported by a finite set. It is immediate from the definition of $m_\pm$ that this J will be reflectionless on $(-2, 2)$, and, by construction, its F function is represented by the measure σ we started out with.

It remains to show that $\|J\| \leq R$. We observed that $\rho_\pm$ are supported by $[-R, R]$, and the essential spectrum can be determined by decomposing into half lines, so if there is spectrum outside $[-R, R]$, it can only consist of discrete eigenvalues. If we had such a discrete eigenvalue at E_0, $|E_0| > R$, then the corresponding eigenfunction u must satisfy $u(0) \neq 0$ because if $u(0) = 0$, then u would be in the domain of the half line problems and thus $\rho_\pm(\{E_0\}) > 0$, contradicting the fact that these measures are supported by $[-R, R]$. However, $u(0) \neq 0$ says that u has non-zero scalar product with δ_0, thus the representing measure of $g(z) = \langle \delta_0, (J - z)^{-1}\delta_0 \rangle$ has a point mass at E_0. This implies that $a_0^2 H(x) = -1/g(x)$ changes its sign at $x = E_0$ (this function is holomorphic near E_0, so this statement makes sense), but we already argued in the first part of this proof that (3.3.2) prevents such a sign change. $\square$

It was proved in [42, Theorem 1.2] that if $J \in \mathcal{M}_R$ for some $R \geq 2$, then $a_n \geq 1$ for all $n \in \mathbb{Z}$. Moreover, if $a_{n_0} = 1$ for a single $n_0 \in \mathbb{Z}$, then $a_n \equiv 1$, $b_n \equiv 0$. This is now an immediate consequence of Theorem 2.3.1. Indeed, (2.3.12) says that $1/a_0^2 = 1 - \sigma_{-2} \leq 1$, and we can only have equality here if $\sigma_{-2} = 0$, which forces σ to be the zero measure. It is easy to check that this makes $m_\pm$ equal to the half line m functions of the free Jacobi matrix. To obtain the full claim, it now suffices to recall that $\mathcal{M}_R$ is shift invariant.

It is tempting to try to obtain more information about the coefficients of a $J \in \mathcal{M}_R$ in this way, by relating them to the moments of σ. The following result is probably unimpressive, but it can serve as an illustration. Also, as we'll discuss after the proof, it is optimal. Recall that we define $r \in (0, 1]$ by the equation $r + 1/r = R$.

Proposition 2.3.3. *If $J \in \mathcal{M}_R$ is not the free Jacobi matrix, then for all $n \in \mathbb{Z}$, we have that $a_n > 1$ and*

$$r^2 < \frac{a_{n+1}^2 - 1}{a_n^2 - 1} < \frac{1}{r^2}. \tag{2.3.13}$$

Note that the reciprocal of the middle has the same bounds.

Proof. The inequality $a_n > 1$ was established above; we only need to prove (2.3.13). By comparing (2.3.11) with (2.3.5), we obtain that

$$\sigma_{-2} = 1 - \frac{1}{a_0^2}, \qquad \sigma_0 = \frac{a_{-1}^2 - 1}{a_0^2}. \tag{2.3.14}$$

Now $r^2 < t^{-2} < 1/r^2$ on the support of σ, hence

$$r^2 \sigma_0 < \sigma_{-2} < \frac{1}{r^2} \sigma_0. \tag{2.3.15}$$

Strict inequality would in fact not follow for the the zero measure $\sigma = 0$, but that would lead us back to the free Jacobi matrix, the case that we explicitly excluded.

Now (2.3.13), for $n = -1$, follows by combining (2.3.15) with (2.3.14). We then obtain (2.3.13) for arbitrary n by shift invariance. □

The inequalities (2.3.13) are indeed sharp, as we pointed out earlier, because they are a rephrasing of (2.3.15), and we can get arbitrarily close to equality here with measures of the form $\sigma = g\delta_{1/r-\epsilon}$ or $\sigma = g\delta_{r+\epsilon}$.

2.4 The Continuous Case: Schrödinger Operators

We consider Schrödinger operators $H = -d^2/dx^2 + V(x)$ on $L^2(\mathbb{R})$, with locally integrable potentials V. We assume limit point case at $\pm\infty$. As we discussed in chapter 1, there are unique (up to a constant factor) solutions $f_\pm$ of $-f'' + Vf = zf$ that are square integrable near $\pm\infty$. The half line m functions may now be defined as follows:

$$m_\pm(z) = \pm\frac{f'_\pm(0,z)}{f_\pm(0,z)} \tag{2.4.1}$$

These obey the asymptotic formulae

$$m_\pm(z) = \sqrt{-z} + o(1) \tag{2.4.2}$$

as $|z| \to \infty$ inside a sector $\delta \leq \arg z \leq \pi - \delta$. See, for example, [1, 19, 22, 23].

We proceed as in the previous chapter. We now say that $H \in \mathcal{M}_R$ if H is reflectionless on $(0,\infty)$ and $\sigma(H) \subset [-R^2,\infty)$. Occasionally, we will abuse terminology and/or notation and instead say that V is in $\mathcal{M}_R$. For $H \in \mathcal{M}_R$, we again obtain a holomorphic function $M : \Omega \to \mathbb{C}^+$ from Lemma 2.1.1, where now $\Omega = \mathbb{C}^+ \cup (0,\infty) \cup \mathbb{C}^-$. Off the real line, M is again given by (2.3.1). We use the conformal map $\varphi : \mathbb{C}^+ \to \Omega$, $\varphi(\lambda) = -\lambda^2$ to introduce the Herglotz function $F(\lambda) = M(\varphi(\lambda))$. We then have the following analog of Theorem 2.3.1.

Theorem 2.4.1. *$H \in \mathcal{M}_R$ if and only if the associated F function is of the form*

$$F(\lambda) = \lambda + \int \frac{d\sigma(t)}{t-\lambda} \tag{2.4.3}$$

for some finite Borel measure σ on $(-R, R)$ that satisfies

$$1 + \int \frac{d\sigma(t)}{t^2 - R^2} \geq 0. \tag{2.4.4}$$

Moreover, if $H \in \mathcal{M}_R$, then V is real analytic. More specifically, $V(x)$ has a holomorphic continuation $V(z)$ to the strip $|\mathrm{Im}\, z| < 1/R$.

As in the discrete case, this establishes a one-to-one correspondence between Schrödinger operators $H \in \mathcal{M}_R$ and measures σ on $(-R, R)$ satisfying (2.4.4). Also as before, if we are not interested in the value of R, then we can say that Theorem 2.4.1 provides us with a one-to-one correspondence between Schrödinger operators H that are reflectionless on $(0, \infty)$ and bounded below and compactly supported measures σ.

Proof. It is again straightforward to check that given an $H \in \mathcal{M}_R$, the corresponding F has such a representation. The general Herglotz representation of F reads

$$F(\lambda) = A + B\lambda + \int \left(\frac{1}{t-\lambda} - \frac{t}{t^2+1} \right) d\sigma(t).$$

Now (2.4.2) immediately shows that $B = 1$ here. Moreover, $m_\pm(z)$ have holomorphic continuations through $(-\infty, -R^2)$. Since $\mathbb{R} \setminus [-R, R]$ gets mapped to this set under φ, it follows that σ is supported by $[-R, R]$, as claimed (point masses at the end points will be prevented by (2.4.4)). We can again split off the second term from the integral and absorb it by A. The redefined A must then satisfy $A = 0$, since (2.4.2) has a little o for the constant term. Thus (2.4.3) holds.

To obtain (2.4.4), we again consider $H = m_+ + m_-$ and $h(\lambda) = H(\varphi(\lambda))$, for $\lambda \in \mathbb{C}^+$, $\operatorname{Re}\lambda < 0$. This function has a holomorphic continuation through the imaginary axis, and for $\lambda = iy$, $y > 0$, we have that $\overline{\lambda} = -\lambda$, thus for these λ, it follows that

$$h(\lambda) = F(\lambda) - \overline{F(\lambda)} = 2\lambda\left(1 + \int \frac{d\sigma(t)}{t^2 - \lambda^2}\right). \tag{2.4.5}$$

We conclude the argument as in the discrete case: By analyticity, (2.4.5) holds for all λ in the second quadrant Q_2. The function $h(\lambda)$ (more precisely: its boundary value as $\varphi(\lambda) \to x \in \mathbb{R}$, $x < -R^2$) must be negative for all $\lambda \in \mathbb{R}$ with $\lambda < -R$, and the factor 2λ already has the correct sign, so the expression in parentheses must be positive. By monotone convergence, when λ increases to $-R$, the integrals $\int \frac{d\sigma}{\lambda^2 - t^2}$ approach $\int \frac{d\sigma}{R^2 - t^2}$ and they increase strictly. Therefore, the condition that the last factor from (2.4.5) is positive for all $\lambda < -R$ is equivalent to (2.4.4).

Conversely, if a measure σ on $(-R, R)$ satisfying (2.4.4) is given, define F by (2.4.3) and then

$$m_+(\varphi(\lambda)) = F(\lambda) \qquad (\lambda \in Q_2) \tag{2.4.6}$$

$$m_-(\varphi(\lambda)) = -\overline{F(\overline{\lambda})} \qquad (\lambda \in Q_4); \tag{2.4.7}$$

here, $Q_j \subset \mathbb{C}$ denotes the (open) jth quadrant. By construction, this pair of Herglotz functions satisfies (2.1.3) on $S = (0, \infty)$. We must show that $m_\pm$ are the half line m functions of a Schrödinger operator H. We thus need an inverse spectral theory result for Schrödinger operators that lets us verify this claim. We will refer to the classical Gelfand-Levitan theory; the version we will use is taken from [40]. Note that since we are dealing with limit point operators here and since it is clear that $m_+(z) = \sqrt{-z} + o(1)$ as $|z| \to \infty$ inside suitable sectors for the m_+ just defined, we may state the results of the discussion of [40, Sect. 19] as follows (for convenience, we focus on the right half line for now): Let $d\rho_0(x) = (1/\pi)\chi_{(0,\infty)}(x)\sqrt{x}\, dx$ be the half line spectral measure for zero potential (with the Dirichlet boundary condition). Consider the signed measure $\nu = \rho_+ - \rho_0$, where ρ_+ is the measure associated with m_+. Then m_+ is the m function of some half line Schrödinger operator (with locally integrable potential) if and only if ρ_+ satisfies the following two conditions:

1. If $f \in L^2(0, L)$ for some $L > 0$ and $\int |F|^2\, d\rho_+ = 0$, with $F(x) = \int f(t) \frac{\sin t\sqrt{x}}{\sqrt{x}}\, dt$, then $f = 0$.

2. It is possible to define a distribution ϕ by

$$\phi(t) = \int \frac{\sin t\sqrt{x}}{\sqrt{x}}\, d\nu(x). \tag{2.4.8}$$

   Moreover, ϕ is a locally integrable function.

More explicitly, what (2) is asking for is the following: If $g \in C_0^\infty(\mathbb{R})$, then

$$\int d|\nu|(x) \left| \int dt\, g(t) \frac{\sin t\sqrt{x}}{\sqrt{x}} \right| < \infty \tag{2.4.9}$$

and there is a locally integrable function ϕ so that for all $g \in C_0^\infty(\mathbb{R})$, we have that

$$\int d\nu(x) \int dt\, g(t) \frac{\sin t\sqrt{x}}{\sqrt{x}} = \int \phi(t) g(t)\, dt. \tag{2.4.10}$$

Let us now check these conditions for the m_+ (or rather, ρ_+) defined above. To learn more about ρ_+, we have to analyze the boundary values of $m_+(z)$ as z approaches the real line; this corresponds to letting $\lambda \in Q_2$ approach either the negative real axis or the positive imaginary axis. We find that

$$d\rho_+(x) = d\mu(x) + \frac{1}{\pi}\chi_{(0,\infty)}(x)\mathrm{Im}\, F(ix^{1/2})\, dx,$$

and here μ is a finite measure, supported by $[-R^2, 0]$. In particular, ρ_+ is equivalent to Lebesgue measure on $(0,\infty)$, so condition (1) holds trivially (in other words, we have an accumulation point of the set of zeros, which means $F \equiv 0$). As for condition (2), this definitely holds for compactly supported ν; the locally integrable function ϕ can then simply be obtained by taking (3.5.1) at face value. Also, to establish (2) for a sum of measures, it clearly suffices to verify this condition for the individual summands separately.

So by splitting off a compactly supported part, we can now focus on

$$d\nu_1(x) = \frac{1}{\pi}\chi_{(1,\infty)}(x)\left(\mathrm{Im}\, F(ix^{1/2}) - x^{1/2}\right)\, dx.$$

Observe that near infinity, $F(\lambda) = \lambda - \sigma_0\lambda^{-1} + O(\lambda^{-2})$ (or Im $F(ix^{\frac{1}{2}}) = \sqrt{x} + \frac{\sigma_0}{\sqrt{x}} + O(\frac{1}{|x|})$), thus

$$d\nu_1(x) = c\chi_{(1,\infty)}(x)x^{-1/2}\, dx + f(x)\, dx$$

where the density $f \in C([1,\infty))$ satisfies $f(x) = O(x^{-1})$. It is clear that this decay is fast enough to give (2) for this part of ν_1; we will again end up interpreting (3.5.1) as a classical integral. By again splitting off a compactly supported part, we thus see that it now suffices to verify (2) for the measure

$$d\nu_2(x) = \chi_{(0,\infty)}(x)x^{-1/2}\, dx$$

Clearly, (2.4.9) holds. It also clear that the left-hand side of (2.4.10) does define a distribution, and in fact a tempered distribution. We now compute its Fourier transform. So apply the left-hand side to the Fourier transform $\widehat{g}$ of a test function g. We obtain that

$$\begin{aligned} \int d\nu_2(x) \int dt\, \widehat{g}(t) \frac{\sin t\sqrt{x}}{\sqrt{x}} &= -i\sqrt{\frac{\pi}{2}} \int_0^\infty \left(g(x^{1/2}) - g(-x^{1/2}) \right) \frac{dx}{x} \\ &= -i\sqrt{2\pi} \int_0^\infty (g(s) - g(-s)) \frac{ds}{s}. \end{aligned}$$

It is easy to verify that this last integral equals $(\mathrm{PV}(1/s), g)$, where the principal value distribution is defined as follows:

$$\left(\mathrm{PV}\left(\frac{1}{s}\right), g \right) = \lim_{\delta \to 0+} \int_{|s|>\delta} \frac{g(s)}{s}\, ds$$

Since $\mathrm{PV}(1/s)$ is the Fourier transform of $i(\pi/2)^{1/2} \operatorname{sgn}(t)$, we now see that $\phi_2(t) = \pi \operatorname{sgn}(t)$, which is a locally integrable function, as claimed.

Of course, one can give an analogous discussion for the left half line and m_-. So, to conclude the proof of the first part of the theorem, we must show that the Schrödinger operator obtained above has spectrum contained in $[-R^2, \infty)$. This can be done by the same arguments as in the discrete case: Clearly, by the decomposition method for σ_{ess}, as $\rho_\pm$ are supported by this set, there is no *essential* spectrum outside $[-R^2, \infty)$. If we had a discrete eigenvalue $E_0 < -R^2$, then the corresponding eigenfunction u must satisfy $u(0) \neq 0$ because otherwise $\rho_\pm(\{E_0\}) > 0$, but we already know that this is not the case. It then follows from the standard construction of a spectral representation of the whole problem (see, for example, [10, Sect. 9.5]) that $\rho(\{E_0\}) > 0$, where ρ denotes the measure associated with the Green function $g = -1/(m_+ + m_-)$. This implies that $H = m_+ + m_-$ changes its sign at E_0, but this is incompatible with (2.4.4): Recall that we in fact specifically formulated (2.4.4) as the condition that would guarantee that H is negative throughout $(-\infty, -R^2)$.

We now move on to the last part of the proof, which discusses the real analyticity of $V \in \mathcal{M}_R$. We will obtain this property from the Riccati equation that is satisfied by m_+, together with a Taylor expansion about infinity. This part of the argument essentially follows the treatment of [31].

We originally define this function for $w \in Q_3$; this choice makes sure that $F(1/w) = m_+(-1/w^2)$. However, it is also clear that p has a holomorphic continuation to a neighborhood of $w = 0$. The corresponding Taylor expansion may be found from (2.4.3):

$$p(w) = \sum_{n=0}^{\infty} \sigma_n w^{n+1}, \tag{2.4.11}$$

where we again write $\sigma_n = \int t^n \, d\sigma(t)$. We now claim that for $n \geq 0$,

$$|\sigma_n| \leq R^{n+2}. \tag{2.4.12}$$

To prove this, observe that obviously $|\sigma_n| \leq \sigma_0 R^n$, since σ is supported by $(-R, R)$. Now condition (2.4.4) implies that $\sigma_0 \leq R^2$, so we obtain (2.4.12). It follows that (2.4.11) converges at least on $|w| < 1/R$.

We now consider the shifted potentials $V_x(t) = V(x + t)$ and the associated data $p(x, w)$, $\sigma_n(x)$. Since $\mathcal{M}_R$ was defined in terms of shift invariant conditions, V_x will also be in $\mathcal{M}_R$ for all x.

By (2.4.1) and the Riccati equations for $m_\pm$,

$$\pm m_\pm'(x, z) = V(x) - z - \big(m_\pm(x, z)\big)^2,$$

, we obtain that (for $w \in Q_3$)

$$\frac{dp}{dx} = -V(x) + p^2(x,w) - \frac{2}{w}\,p(x,w). \tag{2.4.13}$$

We now temporarily work with the integrated form of this equation. We may then

$|w| < 1/R$. The interchange of series and integration in the resulting expressions is easily justified: The coefficients $\sigma_n(x)$ are measurable (they can be obtained as derivatives with respect to w, so are pointwise limits of measurable functions), and (2.4.12) gives uniform (in x) control, so dominated convergence applies. This produces

$$\begin{aligned}\sum_{n\geq 0}(\sigma_n(x) - \sigma_n(0))w^{n+1} = -\int_0^x V(t)\,dt + \sum_{j,k\geq 0} w^{j+k+2}\int_0^x \sigma_j(t)\sigma_k(t)\,dt \\ - 2\sum_{n\geq 0} w^n \int_0^x \sigma_n(t)\,dt.\end{aligned}$$

This was originally derived for $w \in Q_3$, $|w| < 1/R$, but since both sides are holomorphic in w, the equation holds for all $|w| < 1/R$.

We can now compare coefficients in these convergent power series. Starting with the constant terms, this gives that $\int_0^x V\,dt + 2\int_0^x \sigma_0\,dt = 0$ or, by differentiation,

$$V(x) = -2\sigma_0(x) \tag{2.4.14}$$

for almost every x. Since V may be redefined in an arbitrary way on a null set, we can assume that (2.4.14) holds for all $x \in \mathbb{R}$. (Of course, $\sigma_0(x)$ is well defined pointwise, for any given x, independently of the representative of V chosen, as the zeroth moment of the measure $d\sigma(x,\cdot)$ that represents the F function of V_x.)

Next, we obtain that

$$\sigma_0(x) - \sigma_0(0) = -2\int_0^x \sigma_1(t)\,dt.$$

This shows that σ_0 is in fact absolutely continuous, and $\sigma_0' = -2\sigma_1$. Proceeding in this way, we see inductively that $\sigma_n(x)$ is an absolutely continuous function for arbitrary $n \geq 0$. Moreover, since the derivatives σ_n' are built from finitely many other functions σ_j, they are bounded functions by (2.4.12). We have a crude preliminary bound of the form $|\sigma_n'(x)| \leq CnR^n$. This allows us to differentiate the series (2.4.11) (with respect to x) term by term, for $|w| < 1/R$. We then return to the differential version (2.4.13) of the Riccati equation. By again comparing coefficients of power series, we finally arrive at the following recursion formulae:

$$\begin{aligned} V(x) &= -2\sigma_0(x) \\ \sigma_0'(x) &= -2\sigma_1(x) \\ \sigma_n'(x) &= -2\sigma_{n+1}(x) + \sum_{j=0}^{n-1} \sigma_j(x)\sigma_{n-1-j}(x) \qquad (n \geq 1) \end{aligned} \tag{2.4.15}$$

Formally, this could have been obtained very quickly from (2.4.13), but initially we did not know that the $\sigma_n(x)$ are differentiable, so we had to be more circumspect. We now use this recursion to obtain more detailed information about the $\sigma_n(x)$.

Lemma 2.4.2. *The moments $\sigma_n(x)$ satisfy $\sigma_n \in C^\infty(\mathbb{R})$ and*

$$\left|\sigma_n^{(p)}(x)\right| \leq R^{n+p+2}\frac{(n+1+p)!}{(n+1)!}. \tag{2.4.16}$$

Assuming Lemma 2.4.2, we can finish the proof of Theorem 2.4.1 very quickly.

By (2.4.14), the Lemma in particular says that $V \in C^\infty$. Now (2.4.16), for $n = 0$ and general $p \geq 0$, may be used to confirm that the Taylor series of $V(x)$ about an arbitrary $x_0 \in \mathbb{R}$ has radius of convergence $\geq 1/R$. We can then refer to the same estimates and one of the standard bounds on the remainder to see that this Taylor series converges to $V(x)$ on $(x_0 - 1/R, x_0 + 1/R)$. Since the strip $|\operatorname{Im} z| < 1/R$ is simply connected, this shows that V has a holomorphic continuation to the whole strip. $\square$

Proof of Lemma 2.4.2. We already know that $\sigma_n \in C^1$, so the first claim follows from (2.4.15), by an obvious inductive argument. We prove (2.4.16) by induction on p. For $p = 0$, this is just (2.4.12). Now assume that (2.4.16) holds for $0, 1, \ldots, p$ and all $n \geq 0$. We wish to establish the same estimates for $p + 1$ and all $n \geq 0$. We will explicitly discuss only the case $n \geq 1$; $n = 0$ is similar, but much easier. The Leibniz rule says that

$$\frac{d^p}{dx^p}\left(\sigma_j \sigma_{n-1-j}\right) = \sum_{k=0}^{p} \binom{p}{k} \sigma_j^{(k)} \sigma_{n-1-j}^{(p-k)},$$

so from (2.4.15) and the induction hypothesis we obtain that

$$\begin{aligned} \left|\sigma_n^{(p+1)}\right| \leq 2R^{n+p+3}\frac{(n+2+p)!}{(n+2)!} &+ \\ &\sum_{j=0}^{n-1}\sum_{k=0}^{p}\binom{p}{k} R^{n+p+3}\frac{(j+1+k)!}{(j+1)!}\frac{(n-j+p-k)!}{(n-j)!}. \end{aligned}$$

As observed in [31, pg. 293], the sum over k can be evaluated: we have that

$$\sum_{k=0}^{p}\binom{p}{k}\frac{(j+1+k)!}{(j+1)!}\frac{(n-j+p-k)!}{(n-j)!} = \frac{(n+p+2)!}{(n+2)!} \tag{2.4.17}$$

(we'll return to this formula in a moment). Since the answer provided in (2.4.17) is independent of j, we can now also sum over this index. This gives

$$\begin{aligned}\left|\sigma_n^{(p+1)}\right| &\leq 2R^{n+p+3}\frac{(n+2+p)!}{(n+2)!} + nR^{n+p+3}\frac{(n+p+2)!}{(n+2)!} \\ &= R^{n+p+3}\frac{(n+p+2)!}{(n+1)!},\end{aligned}$$

as desired.

It remains to verify (2.4.17). This can be rephrased: we must show that

$$\sum_{k=0}^{p}\binom{N_1+k}{k}\binom{N_2-k}{p-k} = \binom{N_1+N_2+1}{p},$$

for integers $N_1 \geq 1$, $N_2 \geq p$. The left-hand side can be given the same combinatorial interpretation as the right-hand side, so this identity holds. Indeed, choosing p objects from a collection of N_1+N_2+1 is equivalent to the following: denote (N_1+1+N_2)-many balls by $-N_1$, $\cdots$, -1, 0, 1, $\cdots$, N_2, and then choose p-many balls from $\{1, \cdots, N_2\}$ (this is the case when $k=0$ on the sum). Next, we consider two subcollections, $\{-N_1, \cdots, 1, 0\}$ and $\{2, \cdots, N_2\}$ (In other words, remove 1 from the second collection and put 0 on the first collection). Now choose one ball from the first collection and $(p-1)$-many balls from the second collection (this is the case when $k=1$ on the sum). For $k=2$, choose two balls and $(p-2)$-balls from the subcollections $\{-N_1, \cdots, -1, 0, 1\}$ and $\{3, 4, \cdots, N_2\}$, repectively. Observe that there is no repeated choosing in our process. Therefore the identity holds. □

2.5 Proof of Theorems 2.1.2, 2.1.3, 2.1.4

This will depend on material from [41]. We will give a quick review, but will refer the reader to [41] for some of the more technical details.

The key tool is [41, Theorem 3], which says that if V satisfies Hypothesis 2.1.1, then any ω limit point $W = \lim S_{x_n} V$ (that is, any such limit for a sequence $x_n \to \infty$) under the shift map $(S_x V)(t) \equiv V(x+t)$ must be reflectionless on $(0, \infty)$. These limits are taken inside a certain metric space $(\mathcal{V}^C, d)$ of whole line potentials. In fact, $\mathcal{V}^C$ is a space of signed measures on $\mathbb{R}$, and locally integrable potentials U are interpreted as the measures $U(x)\,dx$. However, for our purposes here, measures can be avoided. This is so because the measure analog of the space $\mathcal{M}_R$ contains no new members: all such measures will be (real analytic) functions anyway. The key fact here is the observation that we will still have (2.4.2) for a Schrödinger operator $-d^2/dx^2 + \mu$ with a measure, as long as $\mu(\{0\}) = 0$. This follows from the standard proofs of (2.4.2), suitably adjusted. See also [4, Lemma 5.1]. If $\mu(\{0\}) \neq 0$, then we can shift and instead consider $S_{x_0}\mu$ for an x_0 with $\mu(\{x_0\}) = 0$. With (2.4.2) in place, we can then follow the development given in Sect. 4 to confirm that an operator $-d^2/dx^2 + \mu \in \mathcal{M}_R$ still has an F function of the form described in Theorem 2.4.1, so no new operators are obtained.

The metric d is described in detail in [41]; here, we will be satisfied with a non-technical description. For our purposes, the following properties are important. First of all, convergence to a W with respect to d is equivalent to the condition that

$$\int W(t)\varphi(t)\,dt = \lim_{n\to\infty} \int V(x_n + t)\varphi(t)\,dt \tag{2.5.1}$$

for all continuous, compactly supported test functions φ. (Only limit points $W \in \mathcal{M}_R$ will occur in our situation, so we may assume here that W is continuous, say.) Second, the spaces $(\mathcal{V}^C, d)$ are compact. Since also $\{S_x V\} \subset \mathcal{V}^C$, this means that we can always pass to convergent subsequences of shifted versions of the original potential. Similarly, the spaces $\mathcal{M}_R$ are compact if endowed with the same metric d.

Finally, it is easy to see that limit points W cannot have spectrum outside the (in fact: essential) spectrum of H_+ [41, Proposition 1]. Thus they will lie in $\mathcal{M}_R$ if we take $R \geq 0$ so large that H_+ has no (essential) spectrum below $-R^2$.

The second crucial ingredient to all three proofs is the following immediate consequence of (2.4.14): any $W \in \mathcal{M}_R$ satisfies $W(x) \leq 0$ for all $x \in \mathbb{R}$. Moreover, if $W(x_0) = 0$ for a single $x_0 \in \mathbb{R}$, then $W \equiv 0$. This follows as in the discrete case because $W(x_0) = 0$ forces σ (for x_0) to be the zero measure, and this makes $m_\pm$ equal to the m functions for zero potential.

Proof of Theorem 2.1.2. If the statement of the Theorem did not hold, then we could find a sequence $x_n \to \infty$ so that $S_{x_n} V \to W$ (using compactness) and (2.1.5) along that sequence converges to some $a > 0$, for some test function φ. But then (2.5.1) forces W to be positive somewhere. $\square$

Proof of Theorem 2.1.3. This is similar. The extra assumption on V, if combined with (2.5.1), makes sure that every limit point W is non-negative somewhere. As explained above, this implies that $W \equiv 0$. In other words, the zero potential is the only possible limit point. $\square$

Proof of Theorem 2.1.4. This will again follow from the same ideas. Fix a test function $\psi \geq 0$, $\int \psi = 1$ that is supported by $(-d, d)$. We claim that we can find $\delta > 0$ such that if $W \in \mathcal{M}_R$ satisfies $\int W\psi > -2\delta$ (recall that $W \leq 0$, so $\int W\psi \leq 0$), then

$$\left| \int W(t)\varphi_j(t)\, dt \right| < \epsilon \qquad (j = 1, \ldots, N). \tag{2.5.2}$$

This is a consequence of the compactness of $\mathcal{M}_R$: If our claim was wrong, then we could find a sequence $W_n \to W$, $W_n, W \in \mathcal{M}_R$ so that $\int W_n\psi \to 0$, but (2.5.2) fails for all W_n. But then $\int W\psi = 0$, hence $W = 0$ on the support of ψ, hence $W \equiv 0$. Thus (2.5.2) could not fail for all W_n in this situation. Our claim was correct. We can and will also insist here that $\delta \le \epsilon$.

With this preparation out of the way, use compactness again to find an x_0 with the property that for each $x \ge x_0$, there is a limit point $W \in \mathcal{M}_R$, which will depend on x, so that

$$\left| \int (W(t) - V(x+t))\theta(t)\, dt \right| < \delta$$

for the test functions $\theta = \psi$ and $\theta = \varphi_j$. Now if $V \ge -\delta$ on $(x-d, x+d)$, then $\int V(x+t)\psi(t)\, dt \ge -\delta$, thus $\int W\psi > -2\delta$, so (2.5.2) applies and it follows that

$$\left| \int V(x+t)\varphi_j(t)\, dt \right| < \delta + \epsilon \le 2\epsilon,$$

as desired. □

2.6 Proof of Theorem 2.1.5

(a) Recall how we obtained the conditions on F and σ for a $J \in \mathcal{M}_R$ in the proof of Theorem 2.3.1: Essentially, we had to make sure that the behavior of F as $\lambda \to 0$ and $|\lambda| \to \infty$ is consistent with the known asymptotics of $m_+(z)$ and $m_-(z)$, respectively, as $|z| \to \infty$. If we only want m_+ to be the m function of a (half line) Jacobi matrix, but not m_-, then we only need to make sure that the asymptotics of F as $\lambda \to 0$ come out right.

To obtain such an example, let's just take $\sigma = \delta_1$, so

$$F(\lambda) = -1 + \frac{1}{1-\lambda}. \tag{2.6.1}$$

As F approaches a limit as $|\lambda| \to \infty$, this is clearly not the F function of a whole line Jacobi matrix. (So the point really was to choose a σ with $\sigma_{-2} = 1$, to destroy the required asymptotics at large λ.) However, (2.6.1) will yield an m function m_+ of a (positive) half line Jacobi matrix J_+ via (2.3.9). This follows as in the proof of Theorem 2.3.1; notice that (3.3.2) was not used in this part of the argument. Also, by construction, this m_+ will satisfy (2.1.3) on $(-2, 2)$, for the companion Herglotz function m_- that is also extracted from F, via (2.3.10).

So we have already proved Theorem 2.1.5(a). However, it is also interesting to work things out somewhat more explicitly. We can find $m_+(z)$ most conveniently by using the material from Sect. 2. Notice that (2.3.8) becomes

$$h(\lambda) = h_0(\lambda)\frac{1}{(1-\lambda)(1-1/\lambda)},$$

hence

$$H(z) = \frac{H_0(z)}{z+2}, \tag{2.6.2}$$

where $H_0(z) = \sqrt{z^2-4}$ is the H function of the free Jacobi matrix. Now (3.2.8), specialized to the case at hand, says that $m_+ = A_+ + (1/2)H$. Here we use the fact that the measure ρ associated with H is supported by $(-2, 2)$, as we read off from (2.6.2); there is no point mass at -2 because H_0 contains the factor $(z+2)^{1/2}$. We also know that $m_+(iy) \to 0$ as $y \to \infty$ (because $F(\lambda) \to 0$ as $\lambda \to 0$), and this implies that $A_+ = -1/2$. Thus

$$m_+(z) = \frac{1}{2}\left(\sqrt{\frac{z-2}{z+2}} - 1\right);$$

of course, the square root must be chosen so that m_+ becomes a Herglotz function. With this explicit formula, we can confirm one more time that m_+ is the m function of a Jacobi matrix J_+. The associated measure can also be read off:

$$d\rho_+(x) = \frac{1}{2\pi}\chi_{(-2,2)}(x)\sqrt{\frac{2-x}{2+x}}\, dx$$

In particular, we can now confirm the additional claim that $\sigma(J_+) = [-2,2]$ that was made earlier, in Sect. 1.

It is instructive to obtain this example as a limit of measures $\sigma_\epsilon = (1-\epsilon)\delta_1$. For $\epsilon > 0$ (and small), these measures obey (3.3.2), so are admissible in the sense of Theorem 2.3.1. The F function is given by

$$F_\epsilon(\lambda) = -1 + \epsilon + \epsilon\lambda + \frac{1-\epsilon}{1-\lambda}.$$

A similar analysis can be given. The associated Jacobi matrices J_ϵ have an eigenvalue at $E_\epsilon = -1 - 1/\epsilon$ and no other spectrum outside $[-2,2]$; of course, they are reflectionless on $(-2,2)$. (Operators in $\mathcal{M}_R$ with only discrete spectrum outside $[-2,2]$ are usually called *solitons.*) So our example shows the following: There is a sequence of solitons J_ϵ so that the half line restrictions $(J_\epsilon)_+$ converge, in the strong operator topology, to our J_+ from above. The unrestricted whole line operators J_ϵ do not converge, of course; their operator norms form an unbounded sequence. In fact, (2.3.12) informs us that $a_0 = \epsilon^{-1/2}$, so this is already divergent.

(b) This is very similar, but somewhat more tedious from a technical point of view. Since we already went through similar arguments in the proof of Theorem 2.4.1, we will be satisfied with a sketch. Let

$$d\sigma(t) = \chi_{(1,\infty)}(t)e^{-t}\, dt;$$

as the discussion we are about to give will make clear, only certain general features of this measure matter, not its precise form. Note that a compactly supported σ can not produce an example of the desired type, as observed above, after Theorem 2.4.1. As in part (a), the basic idea is to leave the asymptotics of m_+ essentially untouched while seriously upsetting those of m_-. Indeed, if we now define F by (2.4.3) and then $m_\pm$ by (2.4.6), (2.4.7) and extract the corresponding measures $\rho_\pm$, then we find that

$$d\rho_+(x) = d\rho_0(x) + \chi_{(0,\infty)}(x) f(x)\, dx$$
$$d\rho_-(x) = d\rho_+(x) + \chi_{(-\infty,-1)}(x) e^{-|x|^{1/2}}\, dx,$$

with a density f that again satisfies $f(x) = cx^{-1/2} + O(x^{-1})$ as $x \to \infty$. Exactly this situation was discussed in the proof of Theorem 2.4.1: such a ρ_+ satisfies conditions (1), (2) from the Gelfand-Levitan theory, and since also $m_+(z) = \sqrt{-z} + o(1)$ for large $|z|$, it follows that m_+ is the m function of a half line Schrödinger operator H_+. Notice also that ρ_+ is supported by $(0,\infty)$, so indeed $\sigma(H_+) = [0,\infty)$.

To finish the proof, we show that $\nu = \rho_- - \rho_0$ does not satisfy condition (2). Now we just saw that $\rho_+ - \rho_0$ does define a locally integrable function via (3.5.1) (interpreted in distributional sense), so this will follow if we can show that the formal expression

$$\int_{-\infty}^{-1} \frac{\sin t\sqrt{x}}{\sqrt{x}} e^{-|x|^{1/2}}\, dx \tag{2.6.3}$$

does *not* define a locally integrable function. In fact, it is almost immediate that with the interpretation given above, (2.6.3) does not even define a distribution: Since $x^{-1/2} \sin tx^{1/2} = |x|^{-1/2} \sinh t|x|^{1/2}$ for $x < 0$, it is clear that (2.4.9) diverges for any test function $g \geq 0$, $g \not\equiv 0$ whose support lies to the right of 1.

Chapter 3

Ergodic Jacobi Matrices and Conformal Maps

3.1 Introduction and Basic Setup

In this chapter, we present a general abstract analysis of the basic quantities that are commonly used in the spectral theory of ergodic spaces of Jacobi matrices. Our original inspiration came from the work of Marchenko-Ostrovskii on periodic Schrödinger operators [32], which is perhaps best known (definitely to us) through the reinterpretation of this material that was given in [20, 21]. Marchenko-Ostrovskii use certain conformal maps to parametrize periodic problems, and the same device can be used in a much more general setting. This is one of the main themes of the present paper.

What we do here has some overlap with earlier work on the direct and inverse spectral theory of ergodic and invariant Jacobi matrices, most notably with the by now classical contributions of Kotani [8, 28, 29]. So some parts of this paper are expository in character. Rather than focus exclusively on those parts that (we believe) are new, we have attempted to give a unified, coherent presentation that starts almost from scratch. In those parts where the results are not new, we usually propose alternative arguments.

Recall that a *Jacobi matrix* is a difference operator on $u \in \ell^2$ of the form

$$(Ju)_n = a_n u_{n+1} + a_{n-1} u_{n-1} + b_n u_n.$$

Here, $a_n \geq 0$ and $b_n \in \mathbb{R}$, and we also assume that $a, b \in \ell^\infty(\mathbb{Z})$. Under these assumptions, J is a bounded self-adjoint operator on $\ell^2(\mathbb{Z})$. (One often insists that $a_n > 0$ in Chapter 2, but for what we want to do here, our convention works better.)

We will also impose a uniform bound on the operator norm, and we will in fact work with specifically the space $\mathcal{J}_2$ of all such Jacobi matrices J that satisfy $\|J\| \leq 2$; an arbitrary bounded Jacobi matrix will of course lie in $\mathcal{J}_2$ after multiplication by a suitable constant, and this condition will give us the compactness of our space. It is often useful to make $\mathcal{J}_2$ a compact metric space; one possible choice for such a metric is

$$d(J, J') = \sum_{n\in\mathbb{Z}} 2^{-|n|} \left(|a_n - a'_n| + |b_n - b'_n|\right). \tag{3.1.1}$$

The topology induced by d may be described as the product topology on $\mathcal{J}_2$, now thought of as a subspace of the product of the intervals $[0, 2]$ and $[-2, 2]$ from which we draw the coefficients a_n and b_n, respectively. Alternatively, this topology is also the one induced by both the weak and the strong operator topologies, and we now think of $\mathcal{J}_2$ as a subspace of $B(\ell^2)$, the bounded operators on the Hilbert space $\ell^2(\mathbb{Z})$.

The shift $S(a, b)_n = (a, b)_{n+1}$ acts as a homeomorphism on $(\mathcal{J}_2, d)$. Given an S invariant probability (Borel) measure μ on $\mathcal{J}_2$, we introduce a w function $w = w_\mu$ as follows. We average the spectral measures $d\rho_0(t; J) = d\|E_J(t)\delta_0\|^2$ with respect to μ to obtain the *density of states* measure dk: More precisely, the map $f \mapsto \int d\mu(J) \int d\rho_0(t; J)f(t)$ defines a positive linear functional on the continuous functions f on $[-2, 2]$, so there exists a unique (probability) measure dk on the Borel sets of $[-2, 2]$ so that

$$\int_{\mathcal{J}_2} d\mu(J) \int_{[-2,2]} d\rho_0(t; J)f(t) = \int_{[-2,2]} f(t)\, dk(t) \tag{3.1.2}$$

for all $f \in C[-2,2]$. It's easy to see that $J \mapsto \int f(t)\, d\rho_0(t;J)$ is a continuous map on $\mathcal{J}_2$ for fixed $f \in C[-2,2]$; we will discuss this in more detail in the proof Lemma 3.2.2 below. In particular, this function is measurable and thus the left-hand side of (3.1.2) is well defined.

We also define $A > 0$ by writing

$$\int_{\mathcal{J}_2} \ln a_0(J)\, d\mu(J) = \ln A,$$

at least if $\int \ln a_0\, d\mu > -\infty$. For easier reference, we introduce the notation $\mathcal{M}_0$ for the set of (S invariant, probability) Borel measures μ on $\mathcal{J}_2$ that satisfy this additional condition. We then set

$$w(z) = \ln A - \int_{[-2,2]} \ln(t-z)\, dk(t), \tag{3.1.3}$$

for $z \in \mathbb{C}^+$, the upper half plane in $\mathbb{C}$. Here we take the logarithm with $\operatorname{Im} \ln \zeta \in (-\pi, 0)$ for $\zeta \in \mathbb{C}^-$. So in particular w is a Herglotz function (a holomorphic function $w : \mathbb{C}^+ \to \mathbb{C}^+$). The harmonic (on $\mathbb{C}^+$) function $\gamma(z) = -\text{Re } w(z)$ is called the *Lyapunov exponent.*

These are, of course, well known quantities for ergodic systems of Jacobi matrices, extended here in an obvious way to measures μ that are just invariant. These quantities are often defined in different ways, and indeed there are quite a few well known alternative methods to introduce w. See [9, 34, 49] for (much) more on these topics. Definition (3.1.3) is straightforward and convenient for our purposes.

3.2 Basic Objects

Given $\mu \in \mathcal{M}_0$, define the corresponding w as described above. Write

$$w(z) = -\gamma(z) + i\pi k_0(z);$$

notice that $0 < k_0 < 1$. Also, let

$$k_1(t) = \int_{(-\infty,t]} dk(s)$$

be the increasing function that generates the density of states measure dk.

Proposition 3.2.1. *(a) Let*

$$k(z) = \begin{cases} k_0(z) & z \in \mathbb{C}^+ \\ k_1(z) & z \in \mathbb{R} \end{cases}.$$

Then k is continuous on $\mathbb{C}^+ \cup \mathbb{R}$.

(b) The limit

$$\gamma(x) := \lim_{y \to 0+} \gamma(x + iy)$$

exists for all $x \in \mathbb{R}$. Moreover, $\gamma(z) > 0$ on $z \in \mathbb{C}^+$.

(c) (Thouless formula) For all $z \in \mathbb{C}^+ \cup \mathbb{R}$,

$$\gamma(z) = -\ln A + \int_{[-2,2]} \ln|t - z| \, dk(t)$$

These properties are well known for *ergodic* measures $\mu \in \mathcal{M}_0$. See, for example, [49, Chapter 5]. A discussion of these issues for measures μ that are just invariant may be found in [8].

Sketch of proof. Perhaps the most interesting part of this proof is the one where we establish the inequality $\gamma > 0$ on $\mathbb{C}^+$; once this is available, everything else will then fall into place very quickly or at least follow from routine arguments. Let us first sketch how this can be done, assuming, for the moment, the inequality $\gamma > 0$.

Indeed, part (c) for $z \in \mathbb{C}^+$ is of course an immediate consequence of the definitions of w and γ. Existence of the limit from part (b) can then be deduced from (c) by splitting the region of integration into the two parts $|t - x| \leq 1$ and $|t - x| > 1$ and using monotone and dominated convergence, respectively. These considerations also extend the validity of (c) to $z \in \mathbb{R}$.

Next, we observe that the inequality $\gamma > 0$ together with the Thouless formula force dk to be continuous measure; equivalently, $k_1(t)$ is a continuous function on $\mathbb{R}$. Indeed, for E_0 and E_1 with $E_1 - E_0 < \frac{1}{2}$,

$$\begin{aligned}\int_{[-2,2]} \ln |E - E_0| dk(E) &= \int_{[E_0,E_1]} \ln |E - E_0| dk(E) \\ &+ \int_{|E-E_0|<1, E<E_0 \text{ or } E>E_1} \ln |E - E_0| dk(E) \\ &+ \int_{|E-E_0|\geq 1} |E - E_0| dk(E).\end{aligned}$$

Since the second integral is negative and the third one is bounded (the support of dk is compact), we estimate the left hand side by

$$\int_{[-2,2]} \ln |E - E_0| dk(E) \leq \ln |E_1 - E_0| \int_{[E_0,E_1]} dk(E).$$

By Thouless formula we have

$$\int_{[E_0,E_1]} dk(E) \leq \frac{C}{-\ln |E_1 - E_0|}$$

for some positive number C, and this inequality implies the continuity of the density of states measure dk.

Define

$$k_0(t) = \lim_{y\to 0+} k_0(t+iy); \tag{3.2.1}$$

the limit exists for (Lebesgue) almost every $t \in \mathbb{R}$. Since $k_0(z)$ is bounded, the Herglotz representation of $w(z)$ reads

$$w(z) = C_0 + Dz + \int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) k_0(t)\, dt. \tag{3.2.2}$$

In fact, as Im $w(z) < \pi$ on $\mathbb{C}^+$, we must have $D = 0$ here. By differentiating (3.1.3), we obtain that $w'(z) = \int \frac{dk(t)}{t-z}$, so Im $w'(z) > 0$, or, equivalently, $\partial k_0(x+iy)/\partial x > 0$ on $\mathbb{C}^+$. This implies that $k_0(t)$ is an increasing function on $\mathbb{R}$. Originally, we could only guarantee that $k_0(t)$ was defined off a null set $N \subset \mathbb{R}$, but now we can put $k_0(s) = \lim_{t\to s-; t\notin N} k_0(t)$ for $s \in N$ to obtain an everywhere (on $\mathbb{R}$) defined increasing function k_0. It is also clear, by direct inspection of (3.1.3), that $k_0(t) = 0$ for $t < -2$ and $k_0(t) = 1$ for $t > 2$. Thus k_0 generates a probability measure dk_0 on $[-2, 2]$, and now an integration by parts lets us rewrite (3.2.2) as follows:

$$\begin{aligned} w(z) &= C_0 + \int_{-\infty}^{\infty} \frac{\partial}{\partial t} \left[\ln(t-z) - \frac{1}{2}\ln(t^2+1) \right] k_0(t)\, dt \qquad (3.2.3) \\ &= C_0 + \lim_{R\to\infty} k_0(t) \ln \frac{t-z}{\sqrt{t^2+1}} \bigg|_{t=-R}^{t=R} \\ &\qquad - \int_{-\infty}^{\infty} \left[\ln(t-z) - \frac{1}{2}\ln(t^2+1) \right] dk_0(t) \\ &= C - \int_{-\infty}^{\infty} \ln(t-z)\, dk_0(t). \end{aligned}$$

Measures in Herglotz representations are unique and we can again consider w', so it follows from this that $dk_0 = dk_1$. As already observed above, k_1 is a continuous function on $\mathbb{R}$, and hence so is $k_0(t) = k_1(t)$. Moreover, we defined $k_0(t)$, in (3.2.1), as the boundary value, Lebesgue almost everywhere, of the bounded harmonic function $k_0(z)$, $z \in \mathbb{C}^+$. The Poisson representation formula now shows that $k = k_0$ is continuous on $\mathbb{C}^+ \cup \mathbb{R}$, as claimed.

So, as promised, it only remains to show that $\gamma > 0$. We will in fact assume this inequality for ergodic μ. This is well known; in the ergodic case, γ can be related to the exponential decay rate of certain solutions to the difference equation $Ju = zu$ (thus the term *Lyapunov exponent*). See [49, Chapter 5]. So we will only explain how to generalize the inequality to invariant μ. As mentioned above, this issue is also discussed in [8]; we offer an easy alternative argument here.

Let

$$F_\epsilon(J) = \frac{1}{1+\epsilon}(J + \epsilon J_0), \tag{3.2.4}$$

where J_0 is the Jacobi matrix with $a_n \equiv 1$, $b_n \equiv 0$. In other words, we essentially add ϵ to all a's; the denominator $1+\epsilon$ is not essential and is only introduced to make sure that $F_\epsilon(J) \in \mathcal{J}_2$ again. Given an invariant measure μ, let $\mu_\epsilon = F_\epsilon \mu$ be the corresponding image measure; in other words, $\int f \, d\mu_\epsilon = \int f \circ F_\epsilon \, d\mu$.

Then μ_ϵ is an invariant measure on the compact subspace

$$\mathcal{J}_2^{(\epsilon)} = \{J \in \mathcal{J}_2 : a_n \geq \frac{\epsilon}{1+\epsilon} \text{ for all } n \in \mathbb{Z}\}$$

of $\mathcal{J}_2$. Since the ergodic measures are the extreme points of the set of invariant measures, there are convex combinations $\mu_\epsilon^{(n)}$ of *ergodic* measures $\nu_{j,n,\epsilon}$ on $\mathcal{J}_2^{(\epsilon)}$,

$$\mu_\epsilon^{(n)} = \sum_{j=1}^{N_{n,\epsilon}} c_{j,n,\epsilon} \nu_{j,n,\epsilon},$$

so that $\mu_\epsilon^{(n)} \to \mu_\epsilon$ in weak-$*$ sense as $n \to \infty$. By the result for ergodic measures, we do have that $\gamma_{j,n,\epsilon} > 0$ for the corresponding Lyapunov exponents, and since γ_ν depends linearly on ν, it also follows that $\gamma_\epsilon^{(n)} > 0$. Now on $\mathcal{J}_2^{(\epsilon)}$, the function $J \mapsto \ln a_0(J)$ is continuous (this is the reason why we consider $\gamma_\epsilon^{(n)} > 0$ rather than the original space), so $\ln A_\epsilon^{(n)} \to \ln A_\epsilon$ as $n \to \infty$.

The integrals from the Thouless formula will also converge. To see this, we make use of the following simple fact.

Lemma 3.2.2. *Suppose that $\mu_n \to \mu$ in weak-$*$ sense. Then also $dk_n \to dk$.*

The situation we have in mind here of course includes the assumption that $\mu_n, \mu \in \mathcal{M}_0$, but the Lemma is also valid, with the same proof, for arbitrary finite measures.

Proof of Lemma 3.2.2. Let $f \in C[-2,2]$. Then, from (3.1.2),

$$\begin{aligned}\int f(t)\,dk(t) &= \int_{\mathcal{J}_2} d\mu(J) \int_{[-2,2]} d\rho_0(t;J)\,f(t)\\ &= \lim_{n\to\infty} \int_{\mathcal{J}_2} d\mu_n(J) \int_{[-2,2]} d\rho_0(t;J)\,f(t)\\ &= \lim_{n\to\infty} \int f(t)\,dk_n(t)\end{aligned}$$

because $J \mapsto \int f(t)\,d\rho_0(t;J)$ is a continuous function on $\mathcal{J}_2$. To confirm this last claim, it suffices to observe that convergence with respect to d is equivalent to strong operator convergence and this, in turn, implies weak-$*$ convergence of the spectral measures ρ_0. $\square$

Thus we now know that $\gamma_\epsilon^{(n)}(z) \to \gamma_\epsilon(z)$ on $z \in \mathbb{C}^+$. In particular, it follows that $\gamma_\epsilon \geq 0$ there.

From the definition of μ_ϵ and dominated convergence, it is also clear that $\mu_\epsilon \to \mu$ in weak-$*$ sense as $\epsilon \to 0+$. Hence, as just observed, the integrals from the Thouless formula approach the corresponding limit as $\epsilon \to 0+$. Finally, monotone convergence shows that

$$\ln A_\epsilon = \int \ln a_0(J)\, d\mu_\epsilon(J) = \int \ln a_0(F_\epsilon(J))\, d\mu(J)$$
$$= \int \ln \frac{a_0(J)+\epsilon}{1+\epsilon}\, d\mu(J) \to \int \ln a_0(J)\, d\mu(J) = \ln A.$$

Hence also $\gamma_\epsilon \to \gamma$, so $\gamma \geq 0$. The harmonic function γ is clearly not equal to a constant, hence cannot assume a minimum value, and thus in fact $\gamma > 0$ on $\mathbb{C}^+$. $\square$

It is also useful to notice the following well known consequence of basic potential theory at an early stage:

Lemma 3.2.3. *$A \leq 1$ for any $\mu \in \mathcal{M}_0$.*

Proof. Integrate the Thouless formula with respect to dk. Since $\gamma \geq 0$, we obtain that

$$0 \leq -\ln A + \iint \ln|t-x|\, dk(t)\, dk(x).$$

By the definition of logarithmic capacity [39, Definition 5.1.1], the double integral is $\leq \ln \operatorname{cap}[-2,2] = 0$. $\square$

This argument also shows that if $A = 1$, then $dk = d\omega_{[-2,2]}$, the equilibrium measure of $[-2,2]$. From this one quickly obtains the well known uniqueness result that $\mu = \delta_{J_0}$ if $A = 1$.

We already mentioned the fact that w provides a conformal map from $\mathbb{C}^+$ onto its image. It is advantageous not to work with w itself but with a related function that is obtained by changing variables, as follows. Notice that

$$\zeta \mapsto z = z(\zeta) = -\zeta - \frac{1}{\zeta} \tag{3.2.5}$$

defines a conformal map from the upper semidisk $D^+ = D \cap \mathbb{C}^+$ onto $\mathbb{C}^+$. Here, $D = \{z : |z| < 1\}$ denotes the unit disk. We can therefore introduce

$$F : D^+ \to D^+, \qquad F(\zeta) = e^{w(z(\zeta))}.$$

F indeed maps to the upper unit disk because $\text{Re } w < 0$, $0 < \text{Im } w < \pi$.

Proposition 3.2.4. *F has a holomorphic extension to D, by reflection: $F(\overline{\zeta}) = \overline{F(\zeta)}$. This extended function F is a conformal map from D onto $F(D) \subset D$, with $F(0) = 0$, $F'(0) = A$.*

Since, at least in general, there is some potential for confusion associated with this terminology, we should perhaps clarify our use of language here: by a *conformal map* (also known as a biholomorphic map) we mean a holomorphic bijection between connected open sets (also called *regions* or *domains*); in fact, all domains in this paper will be simply connected.

Proof. It is easy to check that if $\zeta_n \in D^+$, $\zeta_n \to x \in (-1, 1)$, $x \neq 0$, then $\text{Im } F(\zeta_n) \to 0$. Indeed, if $-1 < x < 0$, say, then $z_n = -\zeta_n - 1/\zeta_n \to t > 2$, and thus $k(z_n) \to 1$, by Proposition 3.2.1(a). Since $F = e^{-\gamma} e^{i\pi k}$, this gives the claim in this case. In fact, part (c) of the Proposition shows us that γ is continuous near such a t, so F actually approaches a negative limit. The case $0 < x < 1$ is similar; this time, F converges to a positive limit.

The Schwarz reflection principle therefore provides a holomorphic extension of F to $D \setminus \{0\}$. To define F for $\zeta \in D^-$, we refer to the identity ("reflection") $F(\overline{\zeta}) = \overline{F(\zeta)}$. Moreover, if $\zeta \in D^+$, $\zeta \to 0$, then $z = -\zeta - 1/\zeta$ satisfies $|z| \to \infty$, so

$$w(z) = \ln A - \int \ln(t - z)\, dk(t) = -\ln(-z) + \ln A + O(1/z)$$

as $\zeta \to 0$ and this leads to $F(\zeta) = A\zeta + O(\zeta^2)$. It follows that the singularity at $\zeta = 0$ is removable and $F'(0) = A$, as claimed.

Finally, notice that w is a conformal map from $\mathbb{C}^+$ onto its image. This simply follows from the fact that Im $w'(z) > 0$ on $\mathbb{C}^+$, which we already observed (and used) in the proof of Proposition 3.2.8. It now becomes clear that F also maps D^+ *injectively* onto a subset of D^+ and D^- in the same way onto the corresponding reflected subset of D^-. Moreover, as we observed above, $F(I) \subset I$ for both $I = (-1, 0)$ and $I = (0, 1)$. Thus F could fail to be injective only if $F(x_1) = F(x_2)$ for some points x_1, x_2 that are either both in $(-1, 0)$ or both in $(0, 1)$. However, it's easy to confirm that $\gamma(-x - 1/x)$ is strictly increasing and decreasing, respectively, on these intervals. Indeed, by Thouless formula,

$$\gamma(x_2) - \gamma(x_1) = \int \frac{\ln |t - x_2|}{\ln |t - x_1|} dk(t) > 0$$

for x_1 and x_2 with $2 < x_1 < x_2$ and $t \in E$ =the top supp of dk. Hence F is a conformal map, as claimed. $\square$

We remark in passing that the Schwarz Lemma now provides another simple proof of Lemma 3.2.3.

Proposition 3.2.5. *(a) The domain $\Omega := F(D) \subset D$ is of the following type: If $Re^{i\alpha} \in \Omega$, then $re^{i\alpha} \in \Omega$ for all $r < R$. Also, $re^{i\alpha} \in \Omega$ if and only if $re^{-i\alpha} \in \Omega$.*

(b) A subset $\Omega \subset D$, $\Omega \neq \emptyset$ is open and has the properties stated in part (a) if and only if there exists an upper semicontinuous function $h : S^1 \to [0, 1)$, with $h(e^{-i\alpha}) = h(e^{i\alpha})$, so that

$$\Omega = \Omega_h \equiv \{re^{i\alpha} : 0 \leq r < 1 - h(e^{i\alpha})\}.$$

In other words, Ω is the unit disk with radial slits

$$S_\alpha = \{re^{i\alpha} : 1 - h(e^{i\alpha}) \leq r \leq 1\}$$

removed; the function $h(e^{i\alpha})$ records the height of the slit at angle α.

Proof. (a) We first discuss the corresponding claim about the region $w(\mathbb{C}^+) \subset \{u + iv : u < 0, 0 < v < \pi\}$. Fix v and put

$$L_v = \{u \in \mathbb{R} : u + iv \in w(\mathbb{C}^+)\}.$$

We want to show that $L_v = (-\infty, u_0(v))$. If this were not true, then $(a, b) \subset L_v$, but $a, b \notin L_v$ for some $a < b \leq 0$. This follows because L_v is open, so if it wasn't just a half line, we could take some other component (necessarily bounded).

Take preimages, that is, write $u + iv = w(z(u))$ for $a < u < b$, and with $z(u) \in \mathbb{C}^+$. Clearly, $z(u) \equiv x(u) + iy(u)$ is a continuous function of $u \in (a, b)$. Moreover, $y(u)$ is injective. This follows because Im $w'(z) > 0$, as we observed above, so

$$\frac{\partial k(x + iy)}{\partial x} > 0.$$

Hence it is not possible for two points z_1, z_2 with the same imaginary part to have images $w(z_1), w(z_2)$ whose imaginary parts agree also.

So $y(u)$ must be monotone, and in fact it's not hard to check that $y(u)$ is strictly decreasing on (a, b) (but we don't really need to know this here since an analogous argument would work for strictly increasing $y(u)$). Notice also that the $z(u)$ stay inside a bounded set, because $\gamma(z) \to \infty$ as $|z| \to \infty$. Thus, on a suitable sequence $u_n \to a$, we have that $z(u_n) \to z = x + iy$, and here $y > 0$. It follows that $a + iv = w(z) \in w(\mathbb{C}^+)$, but this contradicts our choice of a.

By transforming back to F and Ω, we now obtain the asserted property of Ω for $0 < \alpha < \pi$, and, by reflection, also for $-\pi < \alpha < 0$. Here, we have already made use of the asserted invariance of Ω under reflection about the real line, but this property is really obvious from the corresponding symmetry of F.

Next, consider $\alpha = 0$. If we recall our discussion of the mapping properties of F from the proof of Proposition 3.2.4, then we see that the positive values of $F(\zeta)$ come from the $\zeta \in (0,1)$. For these ζ, the variable $z = x = -\zeta - 1/\zeta$ varies over $(-\infty, -2)$, so

$$\Omega \cap (0,1) = \{e^{-\gamma(x)} : x < -2\}. \tag{3.2.6}$$

The Thouless formula (Proposition 3.2.1(c)) shows that $\gamma(x)$ is strictly decreasing on $x < -2$, and $\gamma(x) \to \infty$ as $x \to -\infty$, so this set is a ray $(0, R)$, as claimed. The argument for $\alpha = \pi$ is, of course, analogous.

(b) Any domain Ω with the properties just established is equal to a domain Ω_h, if we simply define

$$h(e^{i\alpha}) := \sup\{r \geq 0 : (1-r)e^{i\alpha} \notin \Omega\}. \tag{3.2.7}$$

Furthermore, it is also clear that only this choice of h can possibly work if it is our goal to represent a given Ω as an Ω_h for some h.

Conversely, given any function $h : S^1 \to [0,1)$, we can form the set Ω_h. This set will always contain 0. It is open if and only if h is upper semicontinuous, and it is invariant under reflection about the real line if and only if h is symmetric. Thus, given Ω as described in part (a), h defined by (3.2.7) has these properties. Conversely, if an upper semicontinuous, symmetric h is given, then Ω_h will be as described in (a). $\square$

We are now in a position to appreciate why it was useful to change variables and work with F and $\Omega = F(D) \subset D$ rather than w and $w(\mathbb{C}^+) \subset S = \{x + iy : x < 0, 0 < y < \pi\}$. Since always $F(0) = 0$, $F'(0) > 0$, the conformal map F can be reconstructed, at least in principle, from its image $\Omega = F(D)$. This is not true for w. Indeed, if $\mu = \delta_{AJ_0}$, where J_0 denotes the free Jacobi matrix $a_n \equiv 1$, $b_n \equiv 0$, then $w_A(z) = w_0(z/A)$, and w_0 maps $\mathbb{C}^+$ onto the full strip S. This latter statement follows easily without any calculation from simple properties of dk and γ for the free Jacobi matrix J_0, but one can also use the explicit formula

$$w_0(z) = \ln\left(-\frac{z}{2} + \sqrt{\frac{z^2}{4} - 1}\right)$$

instead. Here, we would have to clarify the precise definitions of the logarithm and the square root, but in fact a much more transparent formulation is obtained if we just say that $F_0(\zeta) = \zeta$.

So $w_A(\mathbb{C}^+) = S$ for all $0 < A \leq 1$, and the image under w does *not* distinguish between these w functions. The domains $\Omega_A \subset D$, on the other hand, have slits at $\alpha = 0, \pi$ of A dependent heights, so are not equal to one another. One can verify directly that these slits become invisible if we transform back to w and z. Theorem 3.6.1 below will throw some additional light on this issue.

The slit height function h is closely related to the Lyapunov exponent γ. In fact, it is essentially γ, plus the change of variables $F = e^w$, $\alpha = \pi k(t)$.

Theorem 3.2.6. *For $0 \leq \alpha \leq \pi$, we have that*

$$h\left(e^{i\alpha}\right) = 1 - e^{-L(\alpha)},$$

where

$$L(\alpha) = \sup\{\gamma(t) : -2 \leq t \leq 2, \pi k(t) = \alpha\}.$$

If $t \in E = \text{top supp } dk$ and t is not an endpoint of a component $(a, b) \subset (-2, 2) \setminus E$, then there is no $s \neq t$ with $k(s) = k(t)$ and thus for these t, the formula above takes the simpler form

$$h\left(e^{i\pi k(t)}\right) = 1 - e^{-\gamma(t)}.$$

Recall in this context that top supp dk, the *topological support* of dk, is defined as the smallest closed subset $E \subset \mathbb{R}$ with $k(E^c) = 0$.

Also, the set $k^{-1}(\{\alpha/\pi\}) \cap [-2, 2]$ is either a single point or a closed interval $[a, b]$, because $k(t)$ is increasing and continuous. In the second case, the interior (a, b) is a component of $(-2, 2) \setminus E$.

Proof. It is again more convenient to discuss the analogous claim about the region $w(\mathbb{C}^+)$. So, for $0 < v < 1$, define

$$H(v) = \sup\{u \geq 0 : -u + i\pi v \notin w(\mathbb{C}^+)\}.$$

We want to show that

$$H(v) = L(v). \tag{3.2.8}$$

Now for any $t \in (-2, 2)$ with $k(t) = v$, we certainly have that $-\gamma(t) + i\pi v \notin w(\mathbb{C}^+)$. Indeed, if $-\gamma(t) + i\pi v = w(z_0)$ for some $z_0 \in \mathbb{C}^+$, then, by open mapping, the image of a small disk $D_r(z_0)$ under w would include a disk about $-\gamma(t) + i\pi v$, but at least some of these points also occur as images of $t + iy$ for small $y > 0$, and this contradicts the fact that w is injective. Thus $H(v) \geq L(v)$.

On the other hand, if we had that $H(v) > L(v)$, say

$$H(v) \geq \gamma(t) + \epsilon \tag{3.2.9}$$

for all $t \in (-2, 2)$ with $k(t) = v$, then we can again look at the preimages of $-u + i\pi v$ for $u > H(v)$. As in the proof of Proposition 3.2.5, write $-u + i\pi v = w(z(u))$. We now let u approach $H(v)$. As above, the $z(u)$ will stay inside a bounded set, so will converge to a limit $t_0 \in \mathbb{C}^+ \cup \mathbb{R}$ along a suitable subsequence. In fact, $t_0 \in \mathbb{C}^+$ is impossible here because then $-H(v) + i\pi v = w(t_0)$ would lie in $w(\mathbb{C}^+)$. Thus $t_0 \in \mathbb{R}$. Since k is continuous on $\mathbb{C}^+ \cup \mathbb{R}$, we can conclude that $k(t_0) = v$, and now (3.2.9) demands that $\gamma(t_0) \leq H(v) - \epsilon$. The function γ is upper semicontinuous, so this inequality would prevent $u = \gamma(z(u))$ from approaching $H(v)$ when we send $u \to H(v)$ along the subsequence chosen above. We can escape this absurd situation only by abandoning (3.2.9). We have established (3.2.8).

This gives the Theorem for $\alpha \neq 0, \pi$. The remaining cases $\alpha = 0, \pi$ do not pose any problems; it suffices to refer to what we discussed already above. See especially (3.2.6). □

3.3 Data Sets

Let us summarize: Starting out from an invariant (probability) measure $\mu \in \mathcal{M}_0$ on $\mathcal{J}_2$, we introduced the density of states dk as the average of the spectral measures ρ_0 and $\ln A = \int \ln a_0 \, d\mu > -\infty$. These have the property that

$$-\ln A + \int_{[-2,2]} \ln |t - z| \, dk(t) \geq 0 \quad \text{for } z \in \mathbb{C}. \tag{3.3.1}$$

We then introduced a variety of additional data, which were computed from (A, dk). We will now show that we can go back and forth between these. More precisely, each of the following is determined by and will determine (A, dk):

- the w function $w(z)$ on $z \in \mathbb{C}^+$;

- the Lyapunov exponent $\gamma(z)$ on $z \in \mathbb{C}^+$;
- the conformal map $F : D \to D$;
- the image domain $\Omega = F(D)$;
- the slit height function h

We will also identify the classes of objects obtained in this way. For easier reference, we give names to the corresponding sets.

Definition 3.3.1. We say that:

(1) $(A, dk) \in \mathcal{D}$ *(density of states)* if $A > 0$ and dk is a probability measure on the Borel sets of $[-2, 2]$ and (3.3.1) holds;

(2) $W \in \mathcal{W}$ *(w function)* if W, W' are Herglotz functions, W maps $\mathbb{C}^+$ to the strip $S = \{x + iy : x < 0, 0 < y < \pi\}$, W' extends holomorphically to $\mathbb{C} \setminus [-2, 2]$ by reflection $W'(\overline{z}) = \overline{W'(z)}$ and $\lim_{y\to\infty} yW'(iy) = i$;

(3) $\Gamma \in \mathcal{L}$ *(Lyapunov exponent)* if $\Gamma, \partial\Gamma/\partial y$ are positive harmonic functions on $\mathbb{C}^+$, Γ extends harmonically to $\mathbb{C} \setminus [-2, 2]$ by reflection $\Gamma(\overline{z}) = \Gamma(z)$, and

$$\lim_{y\to\infty} \frac{\Gamma(iy)}{\ln y} = 1;$$

(4) $G \in \mathcal{C}$ *(conformal map)* if $G : D \to \Omega$ is a conformal map onto a region $\Omega \subset D$ of the type described in Proposition 3.2.5, with $G(0) = 0$, $G'(0) > 0$;

(5) $\Omega \in \mathcal{R}$ *(region)* if $\Omega \subset D$ is a region of the type described in Proposition 3.2.5;

(6) $g \in \mathcal{H}$ *(height function)* if $g : S^1 \to [0, 1)$ is a symmetric $(g(e^{-i\alpha}) = g(e^{i\alpha}))$ upper semicontinuous function.

Theorem 3.3.1. *If $(A, dk) \in \mathcal{D}$ is given, then the associated data w, γ, F, Ω, h have the properties from parts (2)–(6) of Definition 3.3.1. Conversely, if an object of one of these types is given, then there exists a unique pair $(A, dk) \in \mathcal{D}$ that is associated with it.*

At this point, this statement seems to be of conditional type because we have not yet shown that every $(A, dk) \in \mathcal{D}$ is actually obtained from an invariant measure $\mu \in \mathcal{M}_0$, and indeed, we will leave this issue completely aside in this section and the next. However, as we will discuss later, this statement is true; see Theorem 3.5.4 below. For now, it will be important to observe that nowhere in the developments that started with Proposition 3.2.4 did we use the fact that (A, dk) were obtained from a $\mu \in \mathcal{M}_0$; rather, it was only property (3.3.1) that mattered. Similarly, Proposition 3.2.1 continues to hold if we just assume (3.3.1).

We again witness the effect that things become particularly transparent on the level of the conformal maps. Note, for instance, that items (2), (3) from Definition 3.3.1 come with a sizeable amount of fine print, and contrast this with the satisfying fact that all symmetric upper semicontinuous functions occur as slit height functions: it is not easy to see if our function would be a γ or not, but on the image all uppersemicontinuous functions h are all eligible.

In one part of the proof, we will make use of several classical results on conformal maps and their boundary values. This material will also be important in subsequent sections, so let us give a brief review now.

The first tool is the notion of *kernel convergence* for the image domains Ω. For a careful discussion of this topic in a general setting, please see [11, Section 15.4]. We give the basic definition in the form most suitable for our purposes here, and specialized to the case that is of interest to us.

Definition 3.3.2. Let $\Omega_n, \Omega \subset D$ be subdomains of the unit disk of the type discussed in Proposition 3.2.5. We say that $\Omega_n \to \Omega$ in the sense of *kernel convergence* if:

(i) If $z \in \Omega$, then there exist a radius $r = r(z) > 0$ and an index $N = N(z)$ so that $D_r(z) \subset \Omega_n$ for all $n \geq N$.

(ii) If $z \notin \Omega$ and $r > 0$ are given, then there exists $N = N(z, r)$ so that $D_r(z)$ is not contained in Ω_n if $n \geq N$.

To confirm that this is indeed what [11, Definition 15.4.1] says in the present context, observe that the *kernel* with respect to $z_0 = 0$ (as defined in [11]) of a sequence of domains of the type Ω_{h_n}, if it exists, is another domain of the type Ω_h. In particular, there is no need to take a specific connected component of the set introduced in [11]. The general definition of a kernel also demands that $D_r(0) \subset \Omega_n$ for some $r > 0$ and all large n, but this is a consequence of (i) here because we always have that $0 \in \Omega$.

This notion is important for us here because kernel convergence of the image domains is equivalent to the locally uniform convergence of the conformal maps from D onto these domains. We will return to these issues shortly, but let us first give a characterization of kernel convergence in terms of the associated slit height functions h.

Lemma 3.3.2. *Let $\Omega_n, \Omega \subset D$ be domains of the type discussed in Proposition 3.2.5, and let h_n, h be the associated slit height functions. Then the following are equivalent:*

(a) $\Omega_n \to \Omega$ in the sense of kernel convergence;

(b) $\sup \varphi h_n \to \sup \varphi h$ for every $\varphi \in C(S^1)$, $\varphi \geq 0$.

Proof. We first verify that (b) implies (a). Let's start with condition (i) from Definition 3.3.2. Fix an arbitrary point $z \in \Omega_h$, say $z = re^{i\alpha}$. Then $r < 1 - h(e^{i\alpha})$.

The case $r = 0$ is easy: We have that $\sup h < 1$, so condition (b) with $\varphi \equiv 1$ shows that also $\sup h_n \le 1 - \delta$, uniformly in n, for some $\delta > 0$, and thus $D_\delta(0) \subset \Omega_{h_n}$ for all n. So we can now assume that $0 < r < 1 - h(e^{i\alpha})$. Since h is upper semicontinuous, we will have that $h \le 1 - r - 2\epsilon$, say, on a suitable neighborhood of $e^{i\alpha}$, for some $\epsilon > 0$. We can now use (b) with a function φ that is supported by this neighborhood, equal to 1 on a smaller neighborhood of $e^{i\alpha}$, and takes values $0 \le \varphi \le 1$. Assumption (b) then says that for all sufficiently large n, we will also have that $h_n(e^{i\beta}) \le 1 - r - \epsilon$, say, uniformly on some neighborhood $|\beta - \alpha| \le \eta$. In particular, this shows that $D_\delta(z) \subset \Omega_{h_n}$ for all these n, if we take $\delta < \min\{\epsilon, r\eta/100\}$, say.

Let's now move on to condition (ii) from Definition 3.3.2. We are given a $z \notin \Omega_h$ and a radius $\delta > 0$. The assumption that $z = re^{i\alpha} \notin \Omega_h$ means that $r \ge 1 - h(e^{i\alpha})$. Pick a function $0 \le \varphi \le 1$ that is supported by $|\beta - \alpha| \le \delta/10$ and equal to 1 at $e^{i\alpha}$. Condition (b) then provides angles β_n from this neighborhood so that $h_n(e^{i\beta_n}) \ge 1 - r - \delta/2$ for all large n. In particular, this shows that $D_\delta(z)$ is not contained in Ω_{h_n} for these n, as desired. This concludes the proof of the implication (b) $\Longrightarrow$ (a).

We now want to show that, conversely, (a) implies (b). Fix $\varphi \in C(S^1)$, $\varphi \ge 0$. We would first like to show that

$$\liminf_{n\to\infty} (\sup \varphi h_n) \ge \sup \varphi h.$$

The upper semicontinuous function φh assumes a maximum on the compact set S^1, so $\sup \varphi h = \varphi(e^{i\alpha}) h(e^{i\alpha})$ for some $e^{i\alpha} \in S^1$. We may assume here that $\varphi(e^{i\alpha}) h(e^{i\alpha}) > 0$ because otherwise what we're trying to show is trivially true. In fact, for convenience, let's also assume that $\varphi(e^{i\alpha}) = 1$. We have that $(1 - h(e^{i\alpha}))e^{i\alpha} \notin \Omega_h$, and now (ii) from Definition 3.3.2 shows that for any $\delta > 0$

and all large $n \geq N_0 = N_0(\delta)$, we must have $h_n(e^{i\beta}) \geq h(e^{i\alpha}) - \delta$ somewhere on $\alpha - \delta < \beta < \alpha + \delta$, say. Since φ is continuous, it will satisfy $\varphi \geq 1 - \eta$ on this interval, and here $\eta > 0$ can be made arbitrarily small, provided we start the argument with a sufficiently small $\delta > 0$. Putting things together, we conclude that $\sup \varphi h_n \geq \sup \varphi h - \delta - \eta$ for all large n. As discussed, $\delta + \eta$ can be made arbitrarily small here, so this is what we wished to show.

It remains to prove that also

$$\sup \varphi h \geq \limsup_{n \to \infty} (\sup \varphi h_n). \tag{3.3.2}$$

Again, the suprema are really maxima, attained at $e^{i\alpha_n}$, say. We can now pass to a subsequence on which we converge to the lim sup from the right-hand side of (3.3.2), and then pass to a subsequence a second time to make the points converge, say $\alpha_n \to \alpha$. If (3.3.2) were wrong, we would have that $\varphi(e^{i\alpha})h(e^{i\alpha}) \leq \varphi(e^{i\alpha_n})h(e^{i\alpha_n}) - \epsilon_0$, for some $\epsilon_0 > 0$ and all large n from the subsequence that was chosen. Since φ is continuous, it would then also follow that

$$h(e^{i\alpha}) \leq h(e^{i\alpha_n}) - \epsilon, \tag{3.3.3}$$

for these n and some new (possibly smaller) discrepancy $\epsilon > 0$. Now obviously $z_0 := (1 - h(e^{i\alpha}) - \epsilon)e^{i\alpha} \in \Omega_h$, but (3.3.3) says that given any radius $\delta > 0$, no matter how small, the corresponding disk $D_\delta(z_0)$ will not be contained in Ω_{h_n} for infinitely many choices of n. This contradicts condition (i) from Definition 3.3.2. □

The second set of classical results on conformal maps that will play an important role here deals with the boundary values of these functions. The fundamental result in its general form says that a conformal map $F : D \to \Omega$

extends to a homeomorphism $F : \overline{D} \to \widehat{\Omega}$, where $\widehat{\Omega}$ is the union of Ω with the collection of its *prime ends,* endowed with a suitable topology. Please see [11, Sections 14.2, 14.3] for a careful discussion; the result just mentioned is stated as Theorem 3.4 of [11, Section 14.3]. For now, we will need the theory of prime ends only for regions of a relatively simple type; later on, in Section 6, prime ends will make another appearance. In both cases, the material from [11, Sections 14.2, 14.3] will provide more than adequate background.

After these digressions, we now return to Theorem 3.3.1. When we prove this, one assignment will be the task to construct $(A, dk) \in \mathcal{D}$, given a region $\Omega \in \mathcal{R}$. For regions of a certain simple type, this problem admits an explicit solution, and we will base our treatment of the general case on this.

More precisely, call a domain $\Omega \in \mathcal{R}$ a *finite gap domain* if the corresponding slit height function h is non-zero only at finitely many points. So these are regions with finitely many slits; we call them finite gap domains because they correspond to *finite gap Jacobi matrices,* that is, reflectionless Jacobi matrices whose spectrum is a *finite gap set* (a disjoint union of finitely many compact intervals of positive length).

Lemma 3.3.3. *Suppose that $\Omega \in \mathcal{R}$ is a finite gap domain. Then there exists a finite gap set $E \subset [-2, 2]$ so that Ω is the region associated with $A = \operatorname{cap} E$, $dk = d\omega_E$.*

Here, $\operatorname{cap} E$ again denotes the logarithmic capacity of E, and ω_E is the equilibrium measure of E. Please see [39, 44] for background information on potential theory. The proof will show that E can be obtained as the inverse image of ∂D under the (extended) conformal map $z \mapsto w(z) \mapsto F(\zeta)$.

Note also that $\int \ln |t - z| \, d\omega_E(t) \geq \ln \operatorname{cap} E$ for all $z \in \mathbb{C}$ for a finite gap set E, so (3.3.1) holds and thus $(A, d\omega_E)$ is an admissible set of data from the class

$\mathcal{D}$.

Proof. Let $F : D \to \Omega$ the unique conformal map onto Ω with $F(0) = 0$, $F'(0) > 0$. It is easy to find the set of prime ends for a finite gap region Ω. We can conveniently identify this set with a set built from the boundary $\partial\Omega$ as follows. We use two copies of each slit (minus its end point) $\{re^{i\alpha} : 1 - h(e^{i\alpha}) < r \leq 1\}$. Let's call these $S_+(\alpha)$ and $S_-(\alpha)$. Then there is a natural bijection between the prime ends of Ω and the union of these $S_\pm$ with the rest of $\partial\Omega$. Moreover, using this identification, we can also easily describe the topology of $\widehat{\Omega}$, the union of Ω and its prime ends. The topology is in fact the obvious one, if we think of $\widehat{\Omega}$ as the union of Ω and its boundary, but with each slit having two "sides," and points from one side of a slit are not close to those from the other side. More formally, we can say that if $(re^{i\alpha}, +) \in S_+(\alpha)$, say, then a neighborhood base is given by the sets

$$U_\epsilon = \{pe^{i\beta} : |p - r| < \epsilon, \alpha < \beta < \alpha + \epsilon\} \cup \{(pe^{i\alpha}, +) : |p - r| < \epsilon\}$$

for small $\epsilon > 0$. Of course, similar descriptions are available at other points, but we will leave the matter at that.

Recall that we know from [11, Theorem 14.3.4] that F extends to a homeomorphism $F : \overline{D} \to \widehat{\Omega}$. In particular, F maps ∂D homeomorphically onto the prime ends of Ω. By mapping the prime ends back to the correponding points in the complex plane, we also obtain a continuous map F_0 from ∂D onto $\partial\Omega$ (the boundary is now taken as a subset of $\mathbb{C}$). This map is not a homeomorphism; every point on a (half-open) slit has two preimages. The inverse image of ∂D under this map F_0 is a finite disjoint union of subarcs of ∂D; the number of subarcs is equal to the number of slits.

We now transform back to a W function, using the change of variables from

Section 2. Observe that since Ω is invariant under reflection about the real axis, so is F: we have that $F(\overline{\zeta}) = \overline{F(\zeta)}$. This implies that $F(D \cap \mathbb{R}) \subset D \cap \mathbb{R}$, and since $F'(0) > 0$, we also see that $F(D^+) \subset D^+$, $F(D^-) \subset D^-$, where we again abbreviate $D^\pm = D \cap \mathbb{C}^\pm$. Thus we can take a holomorphic logarithm on D^+ and define $W(z) = \ln F(\zeta)$, with $0 < \text{Im}\, W(z) < \pi$ for $z \in \mathbb{C}^+$, and z and ζ are related by (3.2.5). This function W maps $\mathbb{C}^+$ conformally onto the strip $S = \{x + iy : x < 0, 0 < y < \pi\}$ with finitely many horizontal slits of the type

$$S(y, d) = \{x + iy : -d \leq x \leq 0\}$$

removed. What we just said about the boundary behavior of F and F_0 translates into similar statements about W. More precisely, W extends continuously to the boundary $\partial \mathbb{C}^+ = \mathbb{R}$ and maps $\mathbb{R}$ onto the union of ∂S with the slits $S(y_j, d_j)$. Every point of $S(y_j, d_j) \setminus \{-d_j + iy_j\}$ has two preimages, all other boundary points have one preimage.

The points $z \in \mathbb{R} \setminus [-2, 2]$ correspond to $\zeta \in (-1, 1)$, which are not in the boundary of the original domain D, but of course that is no problem at all because F is holomorphic there and thus definitely extends continuously. Somewhat greater care is required to handle possible slits at $\alpha = 0, \pi$. Here, we observe that we obtain precisely one side of such a slit as the image of F, restricted to $\overline{D^+}$. This follows from the reflection symmetry of F.

Let

$$E = W^{-1}\left(\{iy : 0 \leq y \leq \pi\}\right).$$

As explained above, E is a finite gap set; it is the inverse image under (3.2.5) of a finite disjoint union of closed subarcs of ∂D. Since, under (3.2.5), only the $z \in [-2, 2]$ produce values $\zeta \in \partial D$, we also know that $E \subset [-2, 2]$.

Next, write $W = -\gamma + i\pi k$. Since k is continuous up to the real line, the Herglotz representation of this function reads

$$W(z) = C_0 + Bz + \int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) k(t)\, dt.$$

Clearly, the fact that W maps to S forces $B = 0$. Moreover, k is an increasing function. To see this, first recall that k maps the interior of E bijectively onto $\{iy : 0 < y < \pi\} \setminus \{iy_j\}$. Thus k is monotone on each interval from E. On the other hand, if $(a, b) \subset E^c$, then we have to map to the union of the slits with the top and bottom parts of ∂S. As (a, b) is connected, we in fact have to map to a single such horizontal segment, and we now see that k is constant on (a, b). Putting things together, we conclude that k is monotone on $\mathbb{R}$. Finally, arguments $t < -2$ correspond to $\zeta \in (0, 1)$, and since $F'(0) > 0$, these get mapped to positive values again under F, hence $k(t) = 0$ for these t. Similarly, $k(t) = 1$ for $t > 2$.

To summarize: $k(t)$ is strictly increasing on the interior of E and constant on each component of the complement, and k increases from 0 to 1. In particular, k generates a probability measure dk that is supported by E.

We can now run the integration by parts calculation from (3.2.3) again. We obtain that

$$W(z) = C - \int_{[-2,2]} \ln(t-z)\, dk(t). \tag{3.3.4}$$

This formula was derived for $z \in \mathbb{C}^+$, but it remains valid for the continuous extension of W to $z \in \mathbb{R}$ because $\operatorname{Re} W < 0$ on $\mathbb{C}^+$, and now the arguments from the proof of Proposition 3.2.1 yield (3.3.4) on $z \in \mathbb{R}$ also.

Let's take a look at

$$\Phi(z) \equiv -\mathrm{Re}\ W(z) + C = \int_{[-2,2]} \ln|t - z|\, dk(t).$$

From the mapping properties of W, we know that $\Phi = C$ on E, the support of dk, but $\Phi < C$ on $\mathbb{C} \setminus E$. These properties identify Φ as the equilibrium potential of the set E (so $dk = d\omega_E$) and e^C as the logarithmic capacity of E; see [44, Theorem I.3.1] and also Remark 1.5 from Section I.1 of this reference.

So if we use these data $(A, dk) = (\mathrm{cap}\ E, d\omega_E) \in \mathcal{D}$ as our input, then we will obtain the finite gap domain $\Omega \in \mathcal{R}$ we started out with. $\square$

We are now ready for the

Proof of Theorem 3.3.1. We will focus on the existence part exclusively. Indeed, except for small details, our discussion from the previous Sections has already shown that the data we introduced have the stated properties. As mentioned above, it is important to note here that our arguments only used (3.3.1); it was not essential that in the original setting, (A, dk) were obtained from a measure $\mu \in \mathcal{M}_0$. It is also easy to see that each of the data from Definition 3.3.1 determines (A, dk), so we will not spend any time on uniqueness, either.

With these preliminaries out of the way, suppose now that a $W \in \mathcal{W}$ is given. We want to construct $(A, dk) \in \mathcal{D}$ so that

$$W(z) = \ln A - \int \ln(t - z)\, dk(t).$$

The properties of W' in particular ensure that $\overline{W'(x)} = W'(x)$ for $x \in \mathbb{R} \setminus [-2, 2]$, that is, W' is real at these points, and it can have positive imaginary part only in $[-2,2]$. Therefore, the Herglotz representation of W' takes the form

$$W'(z) = C + Dz + \int_{[-2,2]} \frac{dk(t)}{t-z}, \tag{3.3.5}$$

with a finite measure dk and $C \in \mathbb{R}$, $D \geq 0$. In fact, the asymptotics of W' immediately imply that $C = D = 0$, $dk(\mathbb{R}) = 1$. Thus indeed

$$W(z) = B - \int \ln(t-z)\, dk(t).$$

As usual, we take the logarithm with imaginary part in $(0,\pi)$ here. By assumption $0 < \text{Im } W < \pi$ on $\mathbb{C}^+$, and we can now consider $W(Re^{i\alpha})$ with $0 < \alpha < \pi$ and large $R > 0$ to conclude that $\text{Im } B = 0$. In other words, we can indeed write $B = \ln A$ for some $A > 0$, and since also $\text{Re } W < 0$ by assumption, it then follows that (A, dk) satisfy condition (3.3.1), as required.

Assume now that we are given a function $\Gamma \in \mathcal{L}$. The argument, unsurprisingly, will be quite similar to what we just did. Introduce $W(z) = -\Gamma(z) + i\pi K(z)$, where πK is a harmonic conjugate of $-\Gamma$ on $\mathbb{C}^+$. This determines K up to a constant, which will be irrelevant here and can be chosen arbitrarily. The Cauchy-Riemann equations show that

$$\text{Im } W'(x+iy) = \pi \frac{\partial K(x+iy)}{\partial x} = \frac{\partial \Gamma(x+iy)}{\partial y} > 0$$

on $\mathbb{C}^+$. In other words, W' is a Herglotz function.

Consider now the extended function Γ on $\mathbb{C}\setminus[-2,2]$. By assumption, $\Gamma(x+iy)$ is an even function of $y \in \mathbb{R}$ for fixed $|x| > 2$. Thus $\partial\Gamma/\partial y$ is odd, and, in particular,

$$\left. \frac{\partial \Gamma(x+iy)}{\partial y} \right|_{y=0} = 0.$$

In terms of W', this says that the imaginary part of this function is zero on $\mathbb{R} \setminus [-2, 2]$. Thus the associated measure is supported by $[-2, 2]$ and finite, and we again have a representation of the type (3.3.5). Integrate and take real parts to obtain that

$$\Gamma(z) = -Cx - \frac{1}{2} D(x^2 - y^2) + B + \int_{[-2,2]} \ln |t - z| \, dk(t).$$

Since $\Gamma > 0$, we must have that $C = D = 0$ here, and then the information on the asymptotics from Definition 3.3.1 shows that dk is a probability measure. We can again write $B = -\ln A$, with $A > 0$, and (3.3.1) is of course automatic.

In the remaining parts, we will not give a direct construction of (A, dk). Instead, we will approximate and then make use of compactness properties. More specifically, recall that we already discussed the case of a finite gap domain in Lemma 3.3.3, and we will approximate a general domain by these. So assume now that a $G \in \mathcal{C}$ is given, let $\Omega = G(D)$ be the corresponding image domain, and denote the associated slit height function by h.

Let

$$h_n(e^{i\alpha}) = \begin{cases} H_n(j) & \alpha = j\pi/n \quad (j = 0, 1, \ldots, n) \\ 0 & \text{otherwise} \end{cases} ; \qquad (3.3.6)$$

more precisely, we define h_n by such a formula for $0 \leq \alpha \leq \pi$ and then extend symmetrically to the lower semicircle. Here, the $H_n(j)$ are defined as follows:

$$H_n(j) = \sup_{-1/n \leq \delta \leq 1/n} h\left(e^{i\pi(j/n+\delta)}\right). \qquad (3.3.7)$$

Here we use characteristic functions as the cut-off functions. It is then clear that the h_n are slit height functions of finite gap domains Ω_n. We claim that $h_n \to h$ in the sense that the condition from part (b) of Lemma 3.3.2 holds.

The argument is quite similar to what we did in the second part of the proof of this Lemma. Let $\varphi \in C(S^1)$, $\varphi \geq 0$. From the definition of h_n, we have that if $h_n(e^{i\alpha}) > 0$, then $h_n(e^{i\alpha}) = h(e^{i\beta_n})$ for some $\beta_n = \beta_n(\alpha)$ with $|\beta_n - \alpha| \leq \pi/n$. Hence

$$\varphi(e^{i\alpha})h_n(e^{i\alpha}) = \varphi(e^{i\beta_n})h(e^{i\beta_n}) + R_n(\alpha),$$

and here the error R_n may be estimated by the modulus of continuity of φ:

$$|R_n| \leq \omega_{\pi/n}(\varphi) \equiv \sup_{|\delta|\leq\pi/n, \theta\in\mathbb{R}} \left|\varphi(e^{i(\theta+\delta)}) - \varphi(e^{i\theta})\right|.$$

Since φ is uniformly continuous on S^1, we have that $\omega_{\pi/n} \to 0$ as $n \to \infty$, and it follows that $\limsup(\sup \varphi h_n) \leq \sup \varphi h$.

On the other hand, $\sup \varphi h$ is attained at some point $e^{i\alpha} \in S^1$, and, by construction, $h_n(e^{i\beta_n}) = h(e^{i\alpha})$ at some point $|\beta_n - \alpha| \leq \pi/n$. Since φ is continuous, this implies that $\liminf(\sup \varphi h_n) \geq \sup \varphi h$.

Lemma 3.3.2 now informs us that $\Omega_n \to \Omega$ in the sense of kernel convergence. By Carathéodory's Theorem [11, Theorem 15.4.10], the kernel convergence of the image domains is equivalent to the locally uniform convergence of the conformal maps $G_n : D \to \Omega_n$ (normalized, as usual, by agreeing that $G_n(0) = 0$, $G_n'(0) > 0$), to the limit G.

By Lemma 3.3.3, $G_n(\zeta) = e^{w_n(z)}$ and

$$w_n(z) = \ln A_n - \int \ln(t - z)\, dk_n(t) \tag{3.3.8}$$

for certain data $(A_n, dk_n) \in \mathcal{D}$ (we actually have much more explicit information on what these are, but will not use this here). We can now pass to a subsequence (which, for better readability, we will not make explicit in the notation) so that $A_n \to A$ and $dk_n \to dk$ in weak-$*$ sense. Recall in this context that $A_n = G_n'(0)$, and since $G_n'(0) \to G'(0) > 0$, we can be sure that $A > 0$. The measure dk is a probability measure on $[-2, 2]$.

We can now pass to the limit in (3.3.8) to conclude that

$$w_n(z) \to w(z) \equiv \ln A - \int \ln(t - z)\, dk(t)$$

on $z \in \mathbb{C}^+$. Thus $G_n(\zeta) = e^{w_n(z)} \to e^{w(z)}$, and it follows that $G = e^w$. Put differently, G is obtained from (A, dk). Since $G(D) \subset D$, it follows that $\mathrm{Re}\, w < 0$, so (3.3.1) holds and $(A, dk) \in \mathcal{D}$.

If a domain $\Omega \in \mathcal{R}$ or a slit height function $g \in \mathcal{H}$ is given, we can define an associated conformal map $G : D \to \Omega$ (with $\Omega = \Omega_g$ in the latter case) and then use this treatment to again produce a pair $(A, dk) \in \mathcal{D}$ that corresponds to the data that were given. $\square$

The question of whether and how compact subsets of $\mathbb{R}$ can be approximated by periodic spectra (that is, spectra of periodic Jacobi matrices) has received some attention, and completely satisfactory answers were obtained in at least three independent works. These are [5, 35, 50] but see also [47, Sections 5.6, 5.8] for a comprehensive discussion. In all four cases, the effort needed was not inconsiderable. The approximation procedure implemented above, see (3.3.6), (3.3.7), together with material that we will discuss in the following section, could be used to give a tremendously simplified treatment.

3.4 Convergence of Data

Most of the data sets introduced in the previous section come with natural topologies. It seems reasonable to ask what the relations between these are.

Theorem 3.4.1. *Suppose that $(A_n, dk_n), (A, dk) \in \mathcal{D}$, and form the associated* *objects, as above. Then the following are equivalent:*

(a) $A_n \to A$ *and* $dk_n \to dk$ *in weak-$*$ sense;*

(b) $w_n(z) \to w(z)$ *locally uniformly on* $\mathbb{C}^+$*;*

(c) $\gamma_n(z) \to \gamma(z)$ *locally uniformly on* $\mathbb{C}^+$*;*

(d) $F_n(\zeta) \to F(\zeta)$ *locally uniformly on* D*;*

(e) $\Omega_n \to \Omega$ *in the sense of kernel convergence;*

(f) $\sup \varphi h_n \to \sup \varphi h$ *for every* $\varphi \in C(S^1)$*,* $\varphi \geq 0$*.*

Proof. These statements are either obvious or follow from previously discussed material, so we can go through this quickly. Clearly, (a) yields pointwise convergence of the w functions, and a normal families argument then improves this to give the full claim of (b). Obviously, (b) $\Longrightarrow$ (c). If (c) is assumed and an arbitrary subsequence is chosen, then we can make $A_n \to B \geq 0$ and $dk_n \to d\nu$ in weak-$*$ sense on a sub-subsequence (which is not made explicit in the notation) and then pass to the limit in the Thouless formula along this sequence to conclude that

$$\gamma(z) = -\ln B + \int \ln |t - z| \, d\nu(t).$$

We now see, first of all, that $B > 0$ here, and from the uniqueness of such representations we in fact infer that $(B, d\nu) = (A, dk)$. So it turns out that (A, dk) is the only possible limit point of the sequence (A_n, dk_n), and from the compactness property just used we now obtain (a).

Next, if we recall how F was constructed from w, it is also clear that (b) is equivalent to the locally uniform convergence of F_n to F on D^+, which is equivalent to (d), by a normal families argument.

We already observed earlier that the equivalence of (d) and (e) is exactly what Carathéodory's kernel theorem [11, Theorem 15.4.10] has to say in the case at hand. Finally, (e) $\iff$ (f) is Lemma 3.3.2. $\square$

These spaces $\mathcal{D}$, $\mathcal{W}$ etc. from Definition 3.3.1 become compact if we add a degenerate object, which we can think of as corresponding to $\mu \notin \mathcal{M}_0$. We will discuss this in more detail in a moment. We first present the analog of Theorem 3.4.1 for approach to this added object.

Theorem 3.4.2. *Let $(A_n, dk_n) \in \mathcal{D}$, and introduce the corresponding objects, as in Theorem 3.3.1. Then the following are equivalent:*
(a) $A_n \to 0$*;*
(b) $|w_n(z)| \to \infty$ *locally uniformly on* $\mathbb{C}^+$*;*
(c) $\gamma_n(z) \to \infty$ *locally uniformly on* $\mathbb{C}^+$*;*
(d) $F_n \to 0$ *locally uniformly on* D*;*
(e) $\Omega_n \to \{0\}$*;*
(f) $\sup h_n \to 1$.

The condition of part (e) must be interpreted as follows: For every $r > 0$, there exists N so that $D_r(0)$ is not contained in Ω_n for $n \geq N$. For example, $\Omega_n = D \setminus [1/n, 1)$ converges to $\{0\}$ in this sense.

In terms of Carathéodory's concept of kernel convergence, condition (e) states that no subsequence $\{\Omega_{n_j}\}$ has a kernel with respect to $z_0 = 0$; see again [11, Definition 15.4.1] for background information.

Proof. This is similar to the previous proof. It is again easy to see that (a) $\iff$ (b) $\iff$ (c) $\iff$ (d): Indeed, since $0 < k < 1$ on $\mathbb{C}^+$, (b) and (c) are obviously equivalent. It is also clear that (d) implies (c), and conversely, if (c) holds, then at least $F_n \to 0$ locally uniformly on D^+, but that is enough to conclude (d) by a normal families argument again. Since $A_n = F_n'(0)$, (d) implies (a), and (a) clearly implies (b) and (c).

Obviously, (e) and (f) are equivalent, and thus we can finish the proof by relating (f) to one of the first four conditions. If (f) is assumed, then Theorem 3.2.6 shows that also $\sup_{-2\le x\le 2}\gamma_n(x)\to\infty$. Since $\gamma_n(x+iy)>\gamma_n(x)$ for $y>0$, this implies that no subsequence of γ_n can converge locally uniformly to a finite harmonic limit function on $\mathbb{C}^+$. A normal families argument now gives (b).

Conversely, if (f) does not hold, say $\sup h_n \le c < 1$ on a subsequence, then Theorem 3.2.6 shows that there is a corresponding uniform bound, $\gamma_n(x)\le C$, on $x\in[-2,2]$, along the same subsequence. So

$$-\ln A_n + \int \ln|t-x|\,dk_n(t) \le C.$$

Integrate both sides with respect to $d\omega_0$, the equilibrium measure of $[-2,2]$ (this will finish the job in a clean way, but is not really necessary; we could also just integrate with respect to Lebesgue measure on [-2,2]). Since $\text{cap}\,[-2,2]=1$, we know that $\int \ln|t-x|\,d\omega_0(x)=0$ for quasi every (in fact: every) $t\in[-2,2]$. Thus Fubini's Theorem yields the inequality $-\ln A_n\le C$ on the subsequence that was chosen above. This clearly prevents A_n from converging to zero. We have shown that (a) does not hold. □

We would like to emphasize one point here that was already made implicitly in our proof of Theorem 3.4.1. Consider again a sequence $(A_n, dk_n)\in\mathcal{D}$, which converges in the sense that $A_n\to B\ge 0$ and $dk_n\to d\nu$ in weak-$*$ sense. There seem to be three possibilities: (i) $(B,d\nu)\in\mathcal{D}$ also, that is, $B>0$ and (3.3.1) holds; (ii) $B>0$, but (3.3.1) fails; (iii) $B=0$.

It is very easy to see that (ii) does not occur. This will be used several times later on, so we state it separately, for easier reference.

Lemma 3.4.3. *Let $(A_n, dk_n)\in\mathcal{D}$ and suppose that $A_n\to B\ge 0$ and $dk_n\to d\nu$*

in weak-$*$ sense. Then either $(B, d\nu) \in \mathcal{D}$ or $B = 0$.

Proof. Suppose that $B > 0$. Let $\gamma_n(z) \in \mathcal{L}$ be the Lyapunov exponents associated with (A_n, dk_n). Then, by passing to the limit in the Thouless formula,

$$\gamma_n(z) \to \Gamma(z) \equiv -\ln B + \int \ln|t - z|\, d\nu(t)$$

for $z \in \mathbb{C}^+$, and since $\gamma_n(z) > 0$, we also have that $\Gamma(z) \geq 0$. This is what (3.3.1) is asking for, so $(B, d\nu) \in \mathcal{D}$, as claimed. $\square$

Finally, let us return to the topic that was already briefly mentioned above: We can build compact metric spaces starting from the sets $\mathcal{D}$, $\mathcal{W}$ etc. from Definition 3.3.1. We first introduce a metric in such a way that convergence with respect to this metric is equivalent to the conditions discussed in Theorem 3.4.1. These spaces are not yet compact, but we can pass to the one-point compactifications by adding a point at infinity (as the phrase goes); this extended space also admits a compatible metric, and approach to the point at infinity is then equivalent to the conditions from Theorem 3.4.2.

There is, of course, general theory underlying this procedure; see, for example, [33]. However, we can also be explicit here and do things entirely by hand. Let us discuss the space $\mathcal{D}_0 = \mathcal{D} \cup \{0\}$ in this style (we call the added point 0 because it is approached precisely if $A_n \to 0$). We first need a metric on the finite positive Borel measures on $[-2, 2]$ that generates the weak-$*$ topology. Fix such a metric and call it D. Then let

$$d((A, dk), (A', dk')) = |A - A'| + D(A\, dk, A'\, dk')$$

for two points from $\mathcal{D}$ and $d((A, dk), 0) = A + D(A\, dk, 0)$ for the distance to added point 0; here, the second argument in $D(A\, dk, 0)$ denotes the zero measure.

This defines a metric d on $\mathcal{D}_0$ with the desired properties. It follows from Lemmas 3.2.3 and 3.4.3 and the compactness of the space of probability Borel measures ν on $[-2,2]$ that $(\mathcal{D}_0, d)$ is compact. Convergence with respect to d is equivalent to the conditions from Theorems 3.4.1(a) and 3.4.2(a).

We can give similar metrics on the (one-point compactifications of the) other spaces from Definition 3.3.1. Alternatively, we can just use Theorem 3.3.1 and Theorems 3.4.1, 3.4.2 to move things over from $\mathcal{D}_0$ to those spaces. We summarize:

Proposition 3.4.4. *There are metrics on the spaces $\mathcal{D}_0 = \mathcal{D} \cup \{0\}$, $\mathcal{W}_0 = \mathcal{W} \cup \{\infty\}$ etc. so that convergence with respect to the metric is equivalent to the corresponding statements from Theorems 3.4.1 and 3.4.2, respectively. These spaces are compact.*

3.5 Existence of Invariant Measures

We now come to the one of the main points of the whole discussion so far. We also want to show that given data as in Definition 3.3.1, there exists a shift invariant measure on $\mathcal{J}_2$ that produces these data.

For the density of states measure dk, this was already shown in [8]. (Such a result also appears here, as Proposition 3.5.2.) Carmona-Kotani work with an approximation by periodic problems, which is very similar to what we did above in the approximation procedure that was based on (3.3.6), (3.3.7). In fact, these approximating data do come from periodic problems; more generally, finite gap domains yield periodic operators if all slits are located at angles that are rational multiples of π. We cannot guarantee that this method also produces the correct A, and this issue will have to be addressed separately. This difficulty is directly related to the fact that while the density of states depends continuously on μ, the quantity A is, in general, only a semicontinuous function of μ.

Recall that $\mathcal{M}_0$ was defined as the set of invariant probability measures on $\mathcal{J}_2$ with $\ln A_\mu \equiv \int \ln a_0 \, d\mu > -\infty$. If $\mu \notin \mathcal{M}_0$, then we formally set $A_\mu = 0$.

Lemma 3.5.1. *Suppose that $\mu_n \in \mathcal{M}_0$ and $\mu_n \to \mu$ in weak-$*$ sense. Then*

$$A_\mu \geq \limsup_{n\to\infty} A_{\mu_n}. \tag{3.5.1}$$

In particular, $\mu \in \mathcal{M}_0$ if $\limsup A_{\mu_n} > 0$.

The inequality can be strict. For example, if J_0 again denotes the free Jacobi matrix with $a \equiv 1$, $b \equiv 0$ and

$$\mu_n = \left(1 - \frac{1}{n}\right) \delta_{J_0} + \frac{1}{n} \delta_{e^{-n} J_0},$$

then $\mu_n \in \mathcal{M}_0$, $\mu_n \to \mu = \delta_{J_0}$, $\ln A_{\mu_n} = -1$ for all n, but $\ln A_\mu = 0$.

As already mentioned above, we may rephrase by saying that the map $\mu \mapsto A_\mu$ is an upper semicontinuous function on the (compact) set of invariant probability measures on $\mathcal{J}_2$, because it can be thought of as the infimum of continuous functions $\ln\left(a_n(J) + \epsilon\right)$ as $\epsilon \to 0$.

This Lemma is supplemented by Lemma 3.2.2, which says that $dk_{\mu_n} \to dk_\mu$ in the situation under consideration.

Proof. Since a limit of invariant measures is invariant itself, the final claim is an immediate consequence of (3.5.1), so it suffices to prove this inequality. Since $J \mapsto \ln(a_0(J) + \epsilon)$ is a continuous function on $\mathcal{J}_2$ for fixed $\epsilon > 0$, we have that

$$\int \ln(a_0(J) + \epsilon) \, d\mu_n(J) \to \int \ln(a_0(J) + \epsilon) \, d\mu(J). \tag{3.5.2}$$

Moreover, $\int \ln(a_0 + \epsilon) \, d\mu \to \ln A_\mu \in [-\infty, \infty)$ as $\epsilon \to 0+$ by monotone convergence, so if (3.5.1) failed, then we could find a subsequence and $\epsilon > 0$ so that

$$\int \ln(a_0(J) + \epsilon)\, d\mu(J) \leq \int \ln a_0(J)\, d\mu_n(J) - \epsilon$$

along the subsequence chosen. However, the integrals on the right-hand side are clearly dominated by $\int \ln(a_0 + \epsilon)\, d\mu_n$, so this contradicts (3.5.2). $\square$

Proposition 3.5.2. *Suppose that $\Gamma \in \mathcal{L}$. Then there exist $\mu \in \mathcal{M}_0$ and $d \geq 0$ so that*

$$\Gamma(z) = \gamma_\mu(z) + d.$$

Moreover, if $\inf_{z\in\mathbb{C}^+} \Gamma(z) = 0$, then necessarily $d = 0$.

Here, γ_μ of course refers to the Lyapunov exponent that is constructed from $\mu \in \mathcal{M}_0$ as in Section 3.1, via (A_μ, dk_μ) and (3.1.3).

Proof. This is similar to the argument we used in the proof of Theorem 3.3.1 to construct (A, dk), given a conformal map G. First of all, Theorem 3.3.1 provides us with associated data (A, dk), $W(z)$, $G(\zeta)$, Ω, h. Define again approximating finite gap domains as in (3.3.6), (3.3.7), and denote the corresponding data by A_n, dk_n, w_n etc. By Lemma 3.3.3, there are finite gap sets $E_n \subset [-2, 2]$ so that $A_n = \ln \operatorname{cap} E_n$ and $dk_n = d\omega_{E_n}$.

This approximation procedure is exceedingly useful here because if $E \subset [-2, 2]$ is a finite gap set, then we can give a solution to the problem we set out to solve, and a fairly explicit one at that. More precisely, just take any ergodic measure μ that is supported by $\mathcal{R}_0(E)$; here, $\mathcal{R}_0(E)$ denotes the set of Jacobi matrices J with $\sigma(J) = E$ (i.e., no more spectrum!) that are reflectionless on E; these are usually called *finite gap operators,* and they have been studied very heavily. An account of the classical theory may be found in [49, Chapter 9], but see also [38, 43] for much more on the spaces $\mathcal{R}_0(E)$. Note that of course

$\mathcal{R}_0(E) \subset \mathcal{J}_2$ if (and only if) $E \subset [-2, 2]$. Ergodic measures on $\mathcal{R}_0(E)$ exist because these spaces are compact and shift invariant.

We claim that, as desired, $A_\mu = \text{cap } E$ and $dk_\mu = d\omega_E$ for such an ergodic μ on $\mathcal{R}_0(E)$. To prove this, it will suffice to show that $\gamma_\mu = 0$ almost everywhere with respect to ω_E on E and dk_μ is supported by E. Compare the final part of the proof of Lemma 3.3.3 for this step, or, better yet, see Proposition 3.9.2 below.

These two properties are well known standard facts about finite gap operators, so we will be satisfied with just giving a brief review. First of all, the absolutely continuous part of the spectral measure $d\rho_0(J)$ is equivalent to $\chi_E(t)\, dt$ for every $J \in \mathcal{R}_0(E)$, and this is immediate from the definition of the property of being reflectionless. See [49, Chapter 8] or [37, 43] for background. It follows from (the easy Ishii-Pastur part of) Kotani theory [27] that $\gamma_\mu = 0$ (Lebesgue, hence ω_E) almost everywhere on E. Alternative arguments are available, too; for example, [37] has a (sketchy, admittedly) discussion of these issues at the end of the introduction.

Moreover, as $\sigma(J) = E$ for all $J \in \mathcal{R}_0(E)$, the spectral measures $\rho_0(J)$ are supported by E and thus dk_μ, being their average, also has this property. As already explained, it now follows that $dk_\mu = d\omega_E$, $A_\mu = \text{cap } E$.

Returning to the main argument, we now have available invariant measures $\mu_n \in \mathcal{M}_0$ that produce the (finite gap) data constructed above. On a suitable subsequence, which we again assume to be the original sequence for notational convenience, we can make the μ_n converge to a limiting measure μ in weak-$*$ sense. We constructed the approximations so that $h_n \to h$ in the sense of Theorem 3.4.1(f), so we also have part (a) of the Theorem and, in particular, $A_n \to A > 0$. Thus Lemma 3.5.1 guarantees that $\mu \in \mathcal{M}_0$. Lemma 3.2.2 then shows that (on $z \in \mathbb{C}^+$)

$$\gamma_n(z) = -\ln A_n + \int \ln|t-z|\, dk_n(t) \to -\ln A + \int \ln|t-z|\, dk_\mu(t).$$

However, from Theorem 3.4.1(c) we know that γ_n also converges to Γ locally uniformly on $\mathbb{C}^+$. This gives the representation $\Gamma = \gamma_\mu + d$, with $d = \ln(A_\mu/A)$. If we now recall that $A = \lim A_n$, then we can use Lemma 3.5.1 to confirm that $A_\mu \geq A$, so $d \geq 0$, as claimed. The final claim is obvious from this, since $\gamma_\mu \geq 0$. $\square$

This is not completely satisfactory. Of course, we would prefer to be able to represent $\Gamma = \gamma_\mu$, without the shift d. To achieve this, we now show that we can also represent a *larger* function than Γ, and then take a suitable convex combination.

Lemma 3.5.3. *Suppose that $\Gamma \in \mathcal{L}$. Then there exist $D > 0$ and $\mu \in \mathcal{M}_0$ so that*

$$\Gamma(z) = \gamma_\mu(z) - D.$$

Note that the strict positivity of D is crucial.

As an immediate consequence of this, we obtain the desired result.

Theorem 3.5.4. *Suppose that an object as in one of the parts of Definition 3.3.1 is given. Then there exists a $\mu \in \mathcal{M}_0$ that generates this object.*

In other words, if $\Gamma \in \mathcal{L}$ is given, there exists $\mu \in \mathcal{M}_0$ so that $\Gamma = \gamma_\mu$, or if $(A, dk) \in \mathcal{D}$ were given, then we can find $\mu \in \mathcal{M}_0$ so that $A = A_\mu$ and $dk = dk_\mu$ and so forth.

Assuming the Lemma, we can indeed easily establish Theorem 3.5.4, as follows. First of all, by Theorem 3.3.1, it suffices to discuss the case where a

$\Gamma \in \mathcal{L}$ is given. Proposition 3.5.2 now yields a $\mu_1 \in \mathcal{M}_0$ so that $\Gamma = \gamma_{\mu_1} + d_1$. If $d_1 = 0$ here, then we are done. If $d_1 > 0$, use Lemma 3.5.3 to find $\mu_2 \in \mathcal{M}_0$ and $d_2 > 0$ so that $\Gamma = \gamma_{\mu_2} - d_2$. Then

$$\mu = \frac{d_2\mu_1 + d_1\mu_2}{d_1 + d_2}$$

also lies in $\mathcal{M}_0$ and satisfies $\gamma_\mu = \Gamma$, as desired. So it only remains to prove Lemma 3.5.3.

Proof of Lemma 3.5.3. By Theorem 3.3.1, we can write

$$\Gamma(z) = -\ln A + \int_{[-2,2]} \ln|t-z|\, dk(t),$$

for some $(A, dk) \in \mathcal{D}$. Partition $[-2, 2]$ into $2N$ intervals I_j of length $2/N$ each, ignore those I_j with $c_j := \int_{I_j} dk(t) = 0$, and let

$$dk_j(t) = \frac{1}{c_j}\,\chi_{I_j}(t)\, dk(t)$$

for the remaining intervals. Then we can recover dk as the convex combination $dk = \sum c_j\, dk_j$, and the dk_j are themselves admissible density of states measures because the integrals $\int \ln|t-z|\, dk_j(t)$ are still bounded below.

So we can define $A_j > 0$ by writing

$$\ln A_j = \inf_{x\in\mathbb{R}} \int \ln|t-x|\, dk_j(t); \tag{3.5.3}$$

then $(A_j, dk_j) \in \mathcal{D}$, or, equivalently, $\Gamma_j \in \mathcal{L}$, where

$$\Gamma_j(z) = -\ln A_j + \int \ln|t-z|\, dk_j(t).$$

By construction, these new functions all satisfy $\inf \Gamma_j = 0$. Therefore, Proposition 3.5.2 provides us with measures $\mu_j \in \mathcal{M}_0$ so that $\Gamma_j = \gamma_{\mu_j}$. Let $\mu = \sum c_j \mu_j$ (as a convex combination), and also observe that $\ln A_j < -\ln N$; indeed, it suffices to take x as the center of I_j in (3.5.3) to confirm this. We have that

$$\gamma_\mu = \sum c_j \Gamma_j = \Gamma + \ln A - \sum c_j \ln A_j \equiv \Gamma + D,$$

and here we can be sure that

$$D = \ln A - \sum c_j \ln A_j > \ln A + \ln N \cdot \sum c_j = \ln A + \ln N$$

will be indeed positive, provided we took $N \in \mathbb{N}$ large enough. □

3.6 Slits and Gaps

Recall the definitions made in the context of Theorem 3.2.6: Let $E = \text{top supp } dk$ be the topological (= smallest closed) support of dk. E is a compact subset of $[-2, 2]$ (with no isolated points), and thus its complement $(-2, 2) \setminus E$ is a disjoint union of open intervals I_j, which we call *gaps.* On each gap $t \in I_j$, the function $k(t) = \int_{[-2,t]} dk(s)$ has a constant value $k_j \in [0, 1]$, which is unique to this gap. We call k_j the *gap label* of I_j.

It is worth pointing out that $k_0 = 0$ is a gap label in this sense if and only if $\min E > -2$; the corresponding gap is the missing piece $(-2, \min E)$. A similar comment applies to $k_0 = 1$ as a gap label.

We mention in passing that there is an interesting and beautiful theory (the *Gap Labeling Theorem*) that describes the set of possible gap labels in terms of the dynamics of the shift map S on top supp μ. See, for example, [26, 45] for the classical results and [2] for a recent development.

We saw earlier that if E is a finite gap set, then the gap labels correspond exactly to the slits of Ω. More precisely, Ω is the unit disk with finitely many radial slits removed, and these slits are located at the angles $e^{\pm i\pi k_j}$, with k_j being the gap labels. See Lemma 3.3.3 and its proof for these statements.

This correspondence between slits and gaps is valid in general, if we define the notion of a slit for a general region $\Omega \in \mathcal{R}$ appropriately.

Definition 3.6.1. Let $\Omega \in \mathcal{R}$, and let $h \in \mathcal{H}$ be the associated slit height function. We say that Ω has a *slit* at angle $e^{i\alpha}$ if

$$h\left(e^{i\alpha}\right) > \limsup_{t\to 0+} h\left(e^{i(\alpha+\sigma t)}\right)$$

for at least one of $\sigma = 1$ or $\sigma = -1$.

So a *slit,* in this technical sense, corresponds to an at least one-sided jump in the slit height function.

Theorem 3.6.1. *Let $\Omega \in \mathcal{R}$ and $0 \le \alpha \le \pi$. Then Ω has a slit at angle $e^{i\alpha}$ if and only if $k = \alpha/\pi$ is a gap label of $E = \text{top supp } dk$.*

Proof. Suppose first that $k_0 \in [0, 1]$ is the label of some gap (a, b), with $-2 \le a < b \le 2$. This means that $k(t) = k_0$ for $t \in [a, b]$, but also that $k(t) \ne k_0$ if $t \in [-2, 2] \setminus [a, b]$. In this situation, Theorem 3.2.6 says that $h(e^{i\pi k_0}) = 1 - e^{-\Gamma}$, $\Gamma = \sup_{a\le t\le b} \gamma(t)$.

The Thouless formula shows that γ has a harmonic extension to $\mathbb{C} \setminus E$, and

$$\gamma''(t) = -\int_E \frac{dk(s)}{(s-t)^2} < 0$$

for $t \in (a, b)$. It also follows, with the help of monotone convergence, that $\gamma\Big|_{[a,b]}$ is continuous. So, in particular, at least one of the inequalities $\Gamma > \gamma(a)$ or

$\Gamma > \gamma(b)$ holds. Let's assume that $\Gamma > \gamma(a)$ and also that $a > -2$ (if $a = -2$, then $\Gamma > \gamma(b)$ and $b < 2$, and an analogous argument works). Then $\gamma(t) \leq \Gamma - \epsilon$ for all $a - \epsilon \leq t \leq a$ for some small $\epsilon > 0$ because γ is upper semicontinuous. Now $k(a-\epsilon) < k(a) = k_0$, so Theorem 3.2.6 implies that $h(e^{i\pi k}) \leq h(e^{i\pi k_0}) - \delta$ for some $\delta > 0$ and all $k < k_0$ that are sufficiently close to k_0. This is what we wanted to show.

To prove the converse, we again use Carathéodory's theory of the boundary values of conformal maps. Assume that

$$\limsup_{k \to k_0-} h\left(e^{i\pi k}\right) < h\left(e^{i\pi k_0}\right) \equiv h_0, \tag{3.6.1}$$

the other case being analogous, of course. In more geometric terms, assumption (3.6.1) means that $\partial\Omega_h$ contains an exposed line segment

$$S = \left\{ re^{i\pi k_0} : 1 - h_0 + \epsilon < r < 1 - h_0 + 2\epsilon \right\} \tag{3.6.2}$$

that can be accessed from Ω_h through smaller angles. Or, more formally, we can choose $\epsilon > 0$ so small that also $Q \subset \Omega_h$, where

$$Q = \left\{ re^{i\alpha} : 1 - h_0 + \epsilon < r < 1 - h_0 + 2\epsilon, \quad \pi k_0 - \epsilon < \alpha < \pi k_0 \right\}.$$

As a consequence, each point on S from (3.6.2) corresponds to a different prime end. Let us try to say this in more precise language: If z_n is a sequence of points from Q that converges (in traditional sense) to some $z \in S$, then z_n, viewed as a sequence from $\widehat{\Omega}_h$, the union of Ω_h with its prime ends, with the topology discussed in [11, Section 14.3], converges to some prime end. (This is easy to show, but for our purposes here, convergence on a subsequence is enough, and this is automatic because $\widehat{\Omega}_h$ is compact.) Moreover, and this is actually the crucial part, if $z \neq z'$, then the corresponding prime ends are different also. This follows immediately from the way prime ends were defined. Finally, recall again [11, Theorem 14.3.4], which says that F extends to a homeomorphism $F : \overline{D} \to \widehat{\Omega}_h$.

The upshot of all this is the following: We can find two sequences $\zeta_n, \zeta_n' \in D$ which converge to two different boundary points $\zeta, \zeta' \in \partial D$, so that $F(\zeta_n)$, $F(\zeta_n')$ both converge to points on S from (3.6.2). We obtain these sequences by simply picking sequences $z_n, z_n' \in Q$ so that $z_n \to z$, $z_n' \to z'$, and here z, z' are two distinct points from S. We then let $\zeta_n = F^{-1}(z_n)$, $\zeta_n' = F^{-1}(z_n')$.

In fact, we can and must say slightly more here: Since the z_n, z_n' can all be chosen from the same semidisk (either D^+ or D^-), it is also true that ζ, ζ' will either both be on the (closed) upper semicircle, or they will both be on the lower semicircle.

If we now go back to the original variables and recall that $k(z)$ is continuous on $\mathbb{C}^+ \cup \mathbb{R}$ (see Proposition 3.2.1(a)), then this says that there are $t, t' \in [-2, 2]$, $t \neq t'$, with $k(t) = k(t') = k_0$. Thus k_0 is a gap label. $\square$

Tools from the classical theory of conformal maps can be used to analyze other questions, too. For example, [11, Theorem 14.5.5] says that $F : D \to \Omega$ has a continuous extension $F_0 : \overline{D} \to \overline{\Omega}$ if and only if $\partial\Omega$ is locally connected. Note that we are now seeking an extension that takes values in $\mathbb{C}$, so this issue is not directly addressed by the theory of prime ends. This result may be used to establish the following criterion for the continuity of the Lyapunov exponent.

Theorem 3.6.2. *Let $\gamma \in \mathcal{L}$, and let $h \in \mathcal{H}$ be the associated slit height function. Then $\gamma(z)$ is continuous on $\mathbb{C}$ if and only if the following holds: (i) If α/π is not a gap label, then h is continuous at $e^{i\alpha}$; (ii) if α/π is a gap label, then $\lim_{t\to 0+} h(e^{i(\alpha+\sigma t)})$ exists for both $\sigma = 1$ and $\sigma = -1$.*

This can be proved by verifying that $\partial\Omega_h$ is locally connected if and only if (i), (ii) hold. Note that as $k(z)$ is always continuous on $\mathbb{C}^+ \cup \mathbb{R}$, the conformal map w has a continuous extension to this set if and only if γ has this property (and in this case, γ extends continuously to all of $\mathbb{C}$, by the Thouless formula). Also, this condition is of course equivalent to the possibility of extending F continuously to $\overline{D}$. Having made these remarks, we omit the detailed proof of Theorem 3.6.2. An alternative, more direct proof that is based on Theorem 3.2.6 is also possible.

3.7 More on Lyapunov Exponents

In this section, we discuss $\gamma(x)$ as a function on $x \in [-2, 2]$. Potential theory implies that if $\gamma_1(x) = \gamma_2(x)$ for quasi every (that is, off a set of capacity zero) such x, then $\gamma_1 \equiv \gamma_2$. See [44, Section I.3]. So this restriction of γ to $[-2, 2]$ still contains all the information. We do not have a description of the set of all these functions, but we are able to offer the following statements, which supplement Theorems 3.4.1, 3.4.2.

Theorem 3.7.1. *Let $\gamma_n, \gamma \in \mathcal{L}$. Then the following conditions are also equivalent to those from Theorem 3.4.1:*

(a)

$$\sup_{-2\le x\le 2} \varphi(x)\gamma(x) = \lim_{n\to\infty} \sup_{-2\le x\le 2} \varphi(x)\gamma_n(x) \tag{3.7.1}$$

for all $\varphi \in C[-2, 2]$, $\varphi \ge 0$.

(b) The $\gamma_n(x)$ ($n \ge 1$, $-2 \le x \le 2$) are uniformly bounded, and if $\nu \in \mathcal{P}$ (defined below), then

$$\lim_{n\to\infty} \int_{[-2,2]} |\gamma(x) - \gamma_n(x)|\, d\nu(x) = 0. \tag{3.7.2}$$

Here, we let $\mathcal{P}$ be the set of probability measures ν on the Borel sets of

$[-2,2]$ for which the potential

$$\Phi_\nu(x) \equiv \int_{\mathbb{R}} \ln|t-x|\, d\nu(t) \tag{3.7.3}$$

is a continuous function of $x \in \mathbb{R}$. This in particular forces ν to give zero weight to all sets of capacity zero. On the other hand, for any compact $K \subset [-2,2]$ of positive capacity, there exists a $\nu \in \mathcal{P}$ with $\nu(K^c) = 0$. See [44, Corollary I.6.11]. So, in some vague sense, one can perhaps say that the class $\mathcal{P}$ is equivalent to capacity.

There are limits to this, however. More specifically, while the $L^1(\nu)$ convergence from (b) of course implies convergence in measure with respect to every $\nu \in \mathcal{P}$, that is,

$$\nu(|\gamma - \gamma_n| \geq \epsilon) \to 0 \quad \text{for every } \epsilon > 0, \tag{3.7.4}$$

we are *not* claiming that the *capacity* of the set where $|\gamma_n - \gamma| \geq \epsilon$ approaches zero, and indeed this latter statement is false. A counterexample may be constructed by approximating a positive $\gamma \in \mathcal{L}$, say $\gamma(x) \equiv 1$ on $[-2,2]$, by a sequence of γ_n's corresponding to finite gap sets E_n, as in the proof of Theorem 3.3.1 (compare (3.3.6), (3.3.7)). Lemma 3.3.3 then shows that

$$\operatorname{cap}\left(\{x \in [-2,2] : \gamma_n(x) = 0\}\right) = \operatorname{cap} E_n = A_n.$$

Note that this set is contained in the set where $|\gamma_n - \gamma| \geq \epsilon$. By construction, the A_n approach the positive limit $A = F'(0)$, where $F \in \mathcal{C}$ is the conformal map associated with γ (so if $\gamma \equiv 1$, then $F(\zeta) = e^{-1}\zeta$, but we don't need to know this here).

Theorem 3.7.2. *Let $\gamma_n \in \mathcal{L}$. Then the conditions from Theorem 3.4.2 are equivalent to:*

(a)

$$\lim_{n\to\infty} \sup_{-2\le x\le 2} \gamma_n(x) = \infty. \tag{3.7.5}$$

(b) If $\nu \in \mathcal{P}$, then

$$\lim_{n\to\infty} \int_{[-2,2]} \gamma_n(x)\, d\nu(x) = \infty.$$

Since (3.7.1) and (3.7.5) are analogous to conditions (f) from Theorems 3.4.1 and 3.4.2, respectively, and, moreover, γ and h are directly related through changes of variables (and a partial maximization), as spelled out in Theorem 3.2.6, it seems tempting to try to relate these directly. We are going to give a different, more indirect argument, however, which seems easier and more convenient.

Proof of Theorem 3.7.2. We start with this because we will use Theorem 3.7.2 in our proof of Theorem 3.7.1. The equivalence of (a) with the conditions of Theorem 3.4.2 is an immediate consequence of Theorem 3.2.6, which in particular implies that for any $\gamma \in \mathcal{L}$, the associated slit height function satisfies

$$\sup_{0\le\alpha\le\pi} h(e^{i\alpha}) = 1 - \exp\left(- \sup_{-2\le x\le 2} \gamma(x) \right).$$

So (3.7.5) holds if and only if $\sup h_n \to 1$, which is condition (f) from Theorem 3.4.2.

Next, assume that $A_n \to 0$ (this is (a) of Theorem 3.4.2). We want to derive (b) from this. Integrate the Thouless formula with respect to $d\nu$. With the help Fubini's Theorem, this gives

$$\int_{[-2,2]} \gamma_n(x)\, d\nu(x) = -\ln A_n + \int_{[-2,2]} \Phi_\nu(t)\, dk_n(t).$$

Here, Φ_ν is continuous by assumption, hence bounded, and thus the integrals on the right-hand side stay bounded, and (b) follows.

Finally, if (b) is assumed, then (a) follows trivially. □

In the next proof, we will make repeated use of two fundamental potential theoretic results, the *lower envelope theorem* and the *principle of descent.* We will state them here, but please refer to [44, Theorems I.6.8, I.6.9] for a fuller discussion.

Suppose that $dk_n \to d\nu$ in weak-$*$ sense. Then

$$\Phi_\nu(x) = \limsup_{n\to\infty} \Phi_n(x)$$

for quasi every $x \in [-2,2]$ (the *lower envelope theorem*). Here, the logarithmic potential Φ_ν of a measure ν is again defined by (3.7.3), and we of course further abbreviated $\Phi_n \equiv \Phi_{dk_n}$.

This is supplemented by the *principle of descent,* which says that

$$\Phi_\nu(z) \geq \limsup_{n\to\infty} \Phi_n(z)$$

for all $z \in \mathbb{C}$. Again, this is interesting for $z = x \in [-2,2]$. On the complement of this set, the stronger property of locally uniform convergence is obvious.

Proof of Theorem 3.7.1. We first show that the conditions of Theorem 3.4.1 imply (a). Let $\varphi \in C[-2,2]$, $\varphi \geq 0$ be given. As in the proof of Lemma 3.3.2, we will split (3.7.1) into two inequalities. We first show that

$$\sup \varphi\gamma \geq \limsup_{n\to\infty} \left(\sup \varphi\gamma_n\right). \tag{3.7.6}$$

Since the functions $\varphi\gamma_n$ are upper semicontinuous, the suprema are maxima, so if (3.7.6) were wrong, we would find ourselves in the following situation:

$$\sup \varphi\gamma \leq \varphi(x_n)\gamma_n(x_n) - \epsilon, \tag{3.7.7}$$

for all n from a suitable subsequence and certain points $x_n \in [-2, 2]$, and here can also assume that $x_n \to x \in [-2, 2]$ along that same sequence. Let $d\nu_n$ be a shifted version of dk_n; more precisely,

$$\int f(t)\, d\nu_n(t) = \int f(t + x - x_n)\, dk_n(t)$$

for $f \in C(\mathbb{R})$. Notice that $\Phi_{\nu_n}(x) = \Phi_{dk_n}(x_n)$. By (a) of Theorem 3.4.1, $dk_n \to dk$ in weak-$*$ sense and thus also $d\nu_n \to dk$ along the subsequence that was chosen above. Since, furthermore, $A_n \to A$, the principle of descent now says that

$$\gamma(x) \geq \limsup \gamma_n(x_n)$$

(the lim sup is taken along some subsequence, but this is irrelevant here). Since φ is continuous, this contradicts (3.7.7), unless $\varphi(x) = 0$. However, if $\varphi(x) = 0$, then (3.7.7) implies that $\gamma_n(x_n) \to \infty$, and we again obtain a contradiction, this time to Theorem 3.7.2. We have established (3.7.6).

Next, we show that also

$$\sup \varphi\gamma \leq \liminf_{n\to\infty} \left(\sup \varphi\gamma_n\right), \tag{3.7.8}$$

and this together with (3.7.6) will of course establish (3.7.1). Again, we argue by contradiction. If (3.7.8) failed, then we would find a subsequence and $x \in [-2, 2]$ so that

$$\varphi(x)\gamma(x) \geq \varphi(t)\gamma_n(t) + 2\epsilon \tag{3.7.9}$$

for all $t \in [-2, 2]$ and all n from that sequence. We can now use the fact that γ is continuous with respect to the fine topology and slightly change x to obtain another inequality of this type (with 2ϵ replaced by ϵ, say), where we can now also guarantee that x is not from the exceptional capacity zero set from the lower envelope theorem. Thus $\gamma(x) = \limsup \gamma_n(x)$. Here, the lim sup is taken along the same subsequence that was singled out above (this is important); in other words, we applied the lower envelope theorem to this subsequence and not to the original sequence. We obtain a contradiction to (3.7.9) with $t = x$.

To prove that, conversely, (a) above implies part (a) from Theorem 3.4.1, we again exploit the compactness properties that were discussed in Sections 4. Suppose that (3.7.1) holds. We can pass to a subsequence so that $A_n \to B$, $dk_n \to d\nu$. Here, by Lemma 3.4.3, either $B = 0$ or $(B, d\nu) \in \mathcal{D}$. The first case is impossible because then Theorem 3.7.2 would imply that (3.7.5) holds on the subsequence we chose, but this is clearly incompatible with our assumption that we have (3.7.1).

So $(B, d\nu) \in \mathcal{D}$, but then, by what we showed already,

$$\lim \left(\sup \varphi\gamma_n\right) = \sup \varphi\gamma_{(B,d\nu)} \tag{3.7.10}$$

along the subsequence constructed, for all $\varphi \in C[-2, 2]$, $\varphi \geq 0$. However, limits in this sense are unique. In other words, if $\gamma, \tilde{\gamma} \in \mathcal{L}$ are not the same function, then

$$\sup_{-2\leq x\leq 2} \varphi(x)\gamma(x) \neq \sup_{-2\leq x\leq 2} \varphi(x)\tilde{\gamma}(x) \tag{3.7.11}$$

for some nonnegative $\varphi \in C[-2, 2]$. Indeed, if $\gamma(x_0) < \tilde{\gamma}(x_0)$, say, for some $x_0 \in [-2, 2]$, then, as γ is upper semicontinuous, we in fact have that $\gamma(x) \leq \gamma(x_0) - \epsilon$ for all x from some neighborhood of x_0 also, so we can simply take a φ that is supported by this neighborhood, $0 \leq \varphi \leq 1$, and $\varphi(x_0) = 1$, and we are then guaranteed that (3.7.11) holds.

This uniqueness means that (3.7.10) forces $\gamma_{(B,d\nu)}$ to be the function γ from (3.7.1), and thus, by the uniqueness part of Theorem 3.3.1, $(B, d\nu) = (A, dk)$, the data associated with γ. So this is the only possible limit point of the sequence (A_n, dk_n), but any subsequence has a limit point, thus the whole sequence has to approach this limit, and this is condition (a) from Theorem 3.4.1.

Next, we again assume the conditions from Theorem 3.4.1, and we now wish to establish (b). First of all, we certainly have that $\gamma_n(x) \leq C$ for all n, x and some uniform bound C. We have already shown that (3.7.1) holds under the present assumptions, so we can now obtain this uniform bound very conveniently by just taking $\varphi \equiv 1$ in this condition.

So we can focus on (3.7.2). Fix a $\nu \in \mathcal{P}$. We will show that $\gamma_n \to \gamma$ in measure, that is, (3.7.4) holds. This is sufficient because, as just discussed, $0 \leq \gamma_n, \gamma \leq C$, so $L^1(\nu)$ convergence will follow from this.

We will argue by contradiction and thus assume hypothetically that (3.7.4) fails. Then there exists $\epsilon > 0$ so that

$$\nu(|\gamma - \gamma_n| \geq \epsilon) \geq \epsilon \tag{3.7.12}$$

for all n taken from some subsequence.

Recall that $\gamma(x) \geq \limsup \gamma_n(x)$ for all x by the principle of descent. So if we are given an $\eta > 0$, we can find an integer $N = N(x, \eta)$ so that

$$\gamma_n(x) \leq \gamma(x) + \eta \qquad \text{for all } n \geq N.$$

We can also choose these integers $N(x, \eta)$ as a measurable function of $x \in [-2, 2]$.

Then $\nu(N > N_0) \to 0$ as $N_0 \to \infty$ by monotone convergence, so we can in fact find a (constant) integer N_0 and an exceptional set $\mathcal{E} \subset [-2,2]$ with $\nu(\mathcal{E}) < \eta$ so that

$$\gamma_n(x) \le \gamma(x) + \eta$$

whenever $n \ge N_0$ and $x \notin \mathcal{E}$. If we take $\eta < \epsilon/2$, say, then (3.7.12) now has the more specific consequence that

$$\nu(\gamma - \gamma_n \ge \epsilon) \ge \frac{\epsilon}{2}$$

for all $n \ge N_0$ from the sequence that was determined earlier. Abbreviate

$$S_n = \{x \in [-2,2] : \gamma_n(x) \le \gamma(x) - \epsilon\};$$

then, as just observed, $\nu(S_n) \ge \epsilon/2$ for these n. It follows that

$$\begin{aligned}
\int_{[-2,2]} \gamma_n(x)\, d\nu(x) &= \int_{S_n} \gamma_n(x)\, d\nu(x) + \int_{S_n^c} \gamma_n(x)\, d\nu(x) \\
&\le \int_{S_n} \gamma(x)\, d\nu(x) - \frac{\epsilon^2}{2} + \int_{S_n^c} \gamma_n(x)\, d\nu(x) \\
&\le \int_{[-2,2]} \gamma(x)\, d\nu(x) + (C+1)\eta - \frac{\epsilon^2}{2}.
\end{aligned}$$

To obtain the last line, we further split S_n^c into two parts. On $S_n^c \cap \mathcal{E}^c$, we have the inequality $\gamma_n \le \gamma + \eta$, so this part of the integral may be estimated by $\int_{S_n^c} \gamma\, d\nu + \eta$, and on $S_n^c \cap \mathcal{E}$, we just use that $\gamma_n \le C$ and $\nu(\mathcal{E}) < \eta$.

If we took $\eta > 0$ so small that $(C+1)\eta < \epsilon^2/2$, then this says that $\int \gamma_n\, d\nu \le \int \gamma\, d\nu - \delta$ for some $\delta > 0$ and all n from a certain subsequence. This is impossible because we can also show that $\int \gamma_n\, d\nu \to \int \gamma\, d\nu$. This is done as above, by integrating the Thouless formula and using Fubini's Theorem:

$$\begin{aligned}\int_{[-2,2]} \gamma_n(x)\, d\nu(x) &= -\ln A_n + \int_{[-2,2]} \Phi_\nu(t)\, dk_n(t)\\ &\to -\ln A + \int_{[-2,2]} \Phi_\nu(t)\, dk(t)\\ &= \int_{[-2,2]} \gamma(x)\, d\nu(x),\end{aligned}$$

because $A_n \to A > 0$ and $dk_n \to dk$ in weak-$*$ sense by assumption, and, also by assumption, Φ_ν is a continuous function. This contradiction proves (3.7.4).

Conversely, if (b) is assumed, we repeat the argument from above: Consider any subsequence on which $A_n \to B \geq 0$, $dk_n \to d\rho$. We want to show that then necessarily $B = A > 0$, $d\rho = dk$, where $(A, dk) \in \mathcal{D}$ are the data of γ. As above, $B = 0$ is impossible because then Theorem 3.7.2(b) would apply on the corresponding subsequence, and this is incompatible with our assumption that (3.7.2) holds. So $(B, d\rho) \in \mathcal{D}$ by Lemma 3.4.3. As a consequence, by what we showed already, $\gamma_n \to \gamma_{(B,d\rho)}$ in $L^1(\nu)$ along the corresponding subsequence. Thus

$gamma_{(B,d\rho)}(x) = \gamma(x)$ almost everywhere with respect to ν for all $\nu \in \mathcal{P}$. This implies that $\gamma_{(B,d\rho)}(x) = \gamma(x)$ for quasi every $x \in [-2, 2]$ because, as we reviewed above, any positive capacity set admits a measure $\nu \in \mathcal{P}$ that is supported by it. We conclude that $\gamma_{(B,d\rho)} = \gamma$ are the same function, thus $(B, d\rho) = (A, dk)$ by the uniqueness part of Theorem 3.3.1. □

3.8 Positive Lyapunov Exponents

In this section, we present a variation on a theme composed by Avila and Damanik [3]. These authors show that if an ergodic system is fixed and factors (= homomorphic images) are considered, then generically the Lyapunov exponent is positive Lebesgue almost everywhere, with respect to a natural topology.

The material discussed in this paper provides a very natural approach to these issues. The key fact is the following consequence of Theorem 3.7.1(b).

Lemma 3.8.1. *Let $\nu \in \mathcal{P}$. For any $a, b \geq 0$, the set*

$$S(a,b) = \{\gamma \in \mathcal{L} : \nu(\gamma \leq a) \geq b\}$$

is a closed subset of (the metric space) $\mathcal{L}$.

Here, we again use the customary self-explanatory notation where a condition is used to denote the set it defines.

Proof. Let $\nu \in \mathcal{P}$. Suppose that $\gamma_n \in S(a,b)$, $\gamma \in \mathcal{L}$, $\gamma_n \to \gamma$ in the sense of Theorem 3.4.1(c) or one of the equivalent descriptions of this mode of convergence.

Given $\epsilon > 0$, no matter how small, Theorem 3.7.1(b), or rather its consequence (3.7.4), lets us find an integer N and an exceptional set $\mathcal{E} \subset [-2, 2]$, such that $\nu(\mathcal{E}) < \epsilon$ and $|\gamma_N(x) - \gamma(x)| < \epsilon$ if $-2 \leq x \leq 2$, $x \notin \mathcal{E}$. Since $\gamma_N \in S(a,b)$ by assumption, this implies that

$$\nu(\gamma \leq a + \epsilon) \geq b - \epsilon.$$

With the help of the monotone convergence theorem, one can now check that this condition for arbitrary $\epsilon > 0$ implies that $\gamma \in S(a,b)$, as desired. $\square$

The Lemma can be rephrased, as follows: The function $\gamma \mapsto \nu(\gamma \leq a)$ is upper semicontinuous. Compare this formulation with [3, Lemma 1].

Corollary 3.8.2. *Let $\nu \in \mathcal{P}$. Then the set*

$$\{\gamma \in \mathcal{L} : \gamma(x) > 0 \quad \text{for } \nu\text{-almost every } x\}$$

is a dense G_δ subset of the compact metric space $\mathcal{L}_0$.

Recall that $\mathcal{L}_0$ was defined as the one-point compactification of $\mathcal{L}$; please review Proposition 3.4.4 and its discussion in this context.

The Corollary has further implications because, by the classical Kotani theory [27], absolutely continuous spectrum for ergodic systems corresponds to zero Lyapunov exponents. See [3] for these aspects of the Corollary.

Proof. By Lemma 3.8.1, the sets

$$U(a,b) = S(a,b)^c = \{\gamma \in \mathcal{L} : \nu(\gamma > a) > 1 - b\}$$

are open in $\mathcal{L}$ and thus also in $\mathcal{L}_0$. Monotone convergence shows that $\nu(\gamma > 0) = \lim_{a\to 0+} \nu(\gamma > a)$, so the set from the Corollary may be represented as follows

$$\bigcap_{n\geq 1} \bigcup_{a>0} U(a, 1/n);$$

it is a countable intersection of open sets, as claimed. It is also dense because for any $\gamma(z) \in \mathcal{L}$, we have that $\gamma(z) + 1/n \in \mathcal{L}$ also, and this sequence converges to $\gamma(z)$ in $\mathcal{L}$. (Approximation of $\gamma = \infty$ by members of the set from the Corollary is of course a trivial assignment.) $\square$

3.9 Ergodic Measures

Return to the discussion of Section 3.5. We are given a $\Gamma \in \mathcal{L}$ (or other data with the properties from Definition 3.3.1), and we constructed an invariant measure $\mu \in \mathcal{M}_0$ so that $\Gamma = \gamma_\mu$. We cannot guarantee that μ will be ergodic here (even if Proposition 3.5.2 already provides the correct μ and we choose the approximating measures μ_n as ergodic measures really nothing has been achieved because a limit of ergodic measures need not be ergodic itself). It is natural to ask if it is also possible to find an ergodic μ so that $\Gamma = \gamma_\mu$.

Unfortunately, we don't have anything substantially new to say on this interesting question. Basically, we will review and put into context some observations made by Kotani in [28], and then point out some obvious open questions.

Proposition 3.9.1. *Suppose that $\Gamma \in \mathcal{L}$ is an extreme point of the convex set $\mathcal{L}$. Then there exists an ergodic measure $\mu \in \mathcal{M}_0$ so that $\Gamma = \gamma_\mu$.*

This does not come as a big suprise. Ergodic measures are precisely the extreme points of the set of invariant measures, so one would expect extreme points to play a role here. The converse of Proposition 3.9.1 is false, however. A counterexample is provided by any ergodic model whose Lyapunov exponent satisfies $\gamma \geq c > 0$. This behavior has been established for the Lyapunov exponent of the Almost Mathieu operator for large coupling [7] (in fact, Bourgain-Jitomirskaya compute the Lyapunov exponent exactly). Such a Lyapunov exponent is not an extreme point of $\mathcal{L}$, for the simple reason that $\gamma \pm c \in \mathcal{L}$ also, and of course $\gamma = \frac{1}{2}(\gamma + c + \gamma - c)$.

Proof. Suppose that $\Gamma \in \mathcal{L}$ is an extreme point, and let $\mu \in \mathcal{M}_0$ be an invariant measure so that $\Gamma = \gamma_\mu$. We now use Choquet theory (see [36], especially Sections 3 and 12 of this reference) to decompose $\mu = \int \nu \, d\sigma(\nu)$ into ergodic measures ν on the Borel sets of $\mathcal{J}_2$. This means that

$$\int_{\mathcal{J}_2} f(J)\, d\mu(J) = \int_{\mathcal{M}} d\sigma(\nu) \int_{\mathcal{J}_2} d\nu(J)\, f(J)$$

for all bounded Borel functions f. Choquet's Theorem says that there is such a measure $d\sigma$, with the following additional properties: it is a probability measure on the Borel sets of the space $\mathcal{M}$ of invariant probability measures on (the Borel sets of) $\mathcal{J}_2$ (with the topology induced by the weak-$*$ topology of the regular Borel measures on $\mathcal{J}_2$, viewed as the dual of $C(\mathcal{J}_2)$). Moreover, and this is crucial, $d\sigma$ is supported by the subset of *ergodic* measures.

We claim that we then also have that

$$\gamma_\mu(z) = \int_{\mathcal{M}} \gamma_\nu(z)\, d\sigma(\nu) \tag{3.9.1}$$

for $z \in \mathbb{C}^+$. Indeed, if we set

$$L_n(J) = \max\{\ln a_0(J), -n\},$$

say, then monotone convergence, applied a total of three times, shows that

$$\begin{aligned} \ln A_\mu &= \int_{\mathcal{J}_2} \ln a_0(J)\, d\mu(J) = \lim_{n\to\infty} \int_{\mathcal{J}_2} L_n(J)\, d\mu(J) \\ &= \lim_{n\to\infty} \int_{\mathcal{M}} d\sigma(\nu) \int_{\mathcal{J}_2} d\nu(J)\, L_n(J) \\ &= \int_{\mathcal{M}} d\sigma(\nu) \lim_{n\to\infty} \int_{\mathcal{J}_2} d\nu(J)\, L_n(J) = \int_{\mathcal{M}} \ln A_\nu\, d\sigma(\nu). \end{aligned}$$

(This also shows that $d\sigma$ is supported by $\mathcal{M}_0$.) Furthermore, by just chasing definitions, we can also easily confirm that $\int f\, dk_\mu = \int d\sigma(\nu) \int dk_\nu\, f$ for continuous f, so we do obtain (3.9.1) by integrating the Thouless formula for γ_ν with respect to $d\sigma$.

Now γ_μ is an extreme point by assumption, so if $M \subset \mathcal{M}$ is any Borel subset, then necessarily $\int_M \gamma_\nu\, d\sigma = \sigma(M)\gamma_\mu$ also. In particular, sets of the type

$$M_{z,\epsilon} = \{\nu \in \mathcal{M}_0 : \gamma_\nu(z) \geq \gamma_\mu(z) + \epsilon\},$$

with $z \in \mathbb{C}^+$, $\epsilon > 0$ all satisfy $\sigma(M_{z,\epsilon}) = 0$, and of course the same goes for sets defined by an inequality of the form $\gamma_\nu(z) \leq \gamma_\mu(z) - \epsilon$. Thus, by taking a suitable countable union, we see that $\gamma_\nu \equiv \gamma_\mu$ for σ-almost every $\nu \in \mathcal{M}_0$. As pointed out above, almost all of these measures ν are also ergodic. □

So it would be interesting to know what the extreme points of $\mathcal{L}$ are. As observed above, γ is not an extreme point if $\inf \gamma > 0$. At the other end of the spectrum, we have the following statement, which we adapted from [28, Theorem 6.3] and its proof.

Proposition 3.9.2. *Let $(A, dk) \in \mathcal{D}$, and let $\gamma \in \mathcal{L}$ be the corresponding Lyapunov exponent. Write $E = \text{top supp}\, dk \subset [-2, 2]$. Suppose that one of the following equivalent conditions holds:*

(a) $A = \text{cap}\ E$, $dk = d\omega_E$;

(b) $\gamma(t) = 0$ for quasi every $t \in E$;

(c) $\gamma(t) = 0$ for ω_E-almost every $t \in E$.

Then γ is an extreme point of $\mathcal{L}$.

So here we assume that $\gamma = 0$ essentially everywhere where this function can be equal to zero. Thus there is a huge gap between the Proposition and our first observation that γ is not an extreme point if $\gamma \geq c > 0$ everywhere.

Proof. The equivalence of (a)–(c) follows from a routine application of potential theoretic tools; compare, for example, [46]. We sketch the argument here for the reader's convenience. First of all, if (a) is assumed, then what (b) asserts is known as *Frostman's Theorem* [39, Theorem 3.3.4]. Next, (b) clearly implies (c) since ω_E gives zero weight to all sets of capacity zero. If (c) holds, then we can integrate the Thouless formula with respect to ω_E and use Fubini's theorem to obtain that

$$0 = -\ln A + \int_{[-2,2]} \Phi_{\omega_E}(t)\, dk(t) = \ln(\text{cap}\ E/A).$$

The last step again depends on Frostman's Theorem. So we indeed have that $A = \text{cap } E$. On the other hand, we may also integrate with respect to dk, and we then obtain that

$$I(dk) \equiv \int_{[-2,2]} dk(t) \int_{[-2,2]} dk(x) \, \ln|t-x| \geq \ln A.$$

The equilibrium measure ω_E may be characterized as the measure that maximizes I among all probability measures supported by E, and this maximum value equals $I(\omega_E) = \ln \text{cap } E$. Thus it now follows that $dk = d\omega_E$, and we have obtained (a).

Such a γ clearly is an extreme point. Indeed, if $\gamma = \frac{1}{2}(\gamma_1 + \gamma_2)$, then, by Theorem 3.3.1, we must also have that $dk = \frac{1}{2}(dk_1 + dk_2)$, so, in particular, $E_1, E_2 \subset E$ and hence $\gamma_j = 0$ quasi everywhere on E_j also. As we just saw, this property identifies $dk_j = d\omega_{E_j}$ as the corresponding equilibrium measures. As $\gamma_j > 0$ on E_j^c, it in fact follows that $E_1 = E_2 = E$ and thus $\gamma_1 = \gamma_2 = \gamma$. $\square$

This provides a class of examples where ergodic measures can always be found. We do not know if there are any $\Gamma \in \mathcal{L}$ that do not admit ergodic measures for their representation. Note also that a certain subclass of the examples discussed in Proposition 3.9.2 has the much stronger property that *every* $\mu \in \mathcal{M}_0$ with $\Gamma = \gamma_\mu$ is ergodic (which also means that there is only one such μ because otherwise we could take convex combinations to obtain non-ergodic μ's). This happens when E is a finite gap set with rationally independent gap labels (this is classical and follows from an analysis of the shift on these spaces; see [49, Chapter 9]), but also for certain sets E with infinitely many gaps and this property (we know this thanks to work of Sodin-Yuditskii [48]). It is not clear if there are other examples of Lyapunov exponents Γ with this property that there is only one (ergodic) μ with $\Gamma = \gamma_\mu$.

Bibliography

[1] F.V. Atkinson, On the location of the Weyl circles, *Proc. Roy. Soc. Edinburgh A* **88** (1981), 345–356.

[2] A. Avila, J. Bochi, and D. Damanik, Opening gaps in the spectrum of strictly ergodic Schrödinger operators, *J. Eur. Math. Soc.* **14** (2012), 61–106.

[3] A. Avila and D. Damanik, Generic singular spectrum for ergodic Schrödinger operators, *Duke Math. J.* **130** (2005), 393–400.

[4] A. Ben Amor and C. Remling, Direct and inverse spectral theory of one-dimensional Schrödinger operators with measures, *Int. Eq. Op. Theory* **52** (2005), 395–417.

[5] A.B. Bogatyrev, On the efficient computation of Chebyshev polynomials for several intervals, *Sb. Math.* **190** (1999), 1571–1605.

[6] G. Borg, Uniqueness theorems in the spectral theory of $y'' + (\lambda - q(x))y = 0$, *Proc. 11th Scandinavian Congress of Mathematicians*, Johan Grundt Tanums Forlag, Oslo, 1952, 276–287.

[7] J. Bourgain and S. Jitomirskaya, Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential, *J. Statist. Phys.* **108** (2002), 1203-1218.

[8] R. Carmona and S. Kotani, Inverse spectral theory for random Jacobi matrices, *J. Stat. Phys.* **46** (1987), 1091–1114.

[9] R. Carmona and J. Lacroix, Spectral Theory of Random Schrödinger Operators, Birkhäuser, Boston, MA (1990).

[10] E.A. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

[11] J.B. Conway, Functions of One Complex Variable II, Graduate Texts in Mathematics 159, Springer, New York, 1995.

[12] W. Craig, The trace formula for Schrödinger operators on the line, *Comm. Math. Phys.* **126** (1989), 379–407.

[13] L. de Branges, Some Hilbert spaces of entire functions I, *Trans. Amer. Math. Soc.***96**, (1960), 259–295.

[14] L. de Branges, Some Hilbert spaces of entire functions II, *Trans. Amer. Math. Soc.***99**, (1961), 118–152.

[15] L. de Branges, Some Hilbert spaces of entire functions III, *Trans. Amer. Math. Soc.***100**, (1961), 73–115.

[16] L. de Branges, Some Hilbert spaces of entire functions IV, *Trans. Amer. Math. Soc.***105**, (1962), 43–83.

[17] L. de Branges, Hilbert Spaces of Entire Functions, Prentice-Hall, Englewood Cliffs, 1968.

[18] S. Denisov, On the continuous analog of Rakhmanov's theorem for orthogonal polynomials, *J. Funct. Anal.* **198** (2003), 465–480.

[19] W.N. Everitt, On a property of the m-coefficient of a second-order linear differential equation, *J. London Math. Soc.* **4** (1972), 443–457.

[20] J. Garnett and E. Trubowitz, Gaps and bands of one-dimensional periodic Schrödinger operators, *Comment. Math. Helv.* **59** (1984), 258-312.

[21] J. Garnett and E. Trubowitz, Gaps and bands of one-dimensional periodic Schrödinger operators II, *Comment. Math. Helv.* **62** (1987), 18–37.

[22] F. Gesztesy and B. Simon, A new approach to inverse spectral theory II: General real potentials and the connection to the spectral measure, *Annals of Math. (2)* **152** (2000), 593–643.

[23] B. J. Harris, The asymptotic form of the Titchmarsh-Weyl m function associated with a second order differential equation with locally integrable coefficient, *Proc. Roy. Soc. Edinburgh A* **102** (1986), 243–251.

[24] I. Hur, M. McBride and C. Remling, The Marchenko representation of reflectionless Jacobi and Schrödinger operators, Preprint, arXiv:1401.7704.

[25] I. Hur and C. Remling, Ergodic Jacobi matrices and conformal maps, *Math. Phys. Anal. Geom* **15** (2012), 121–162.

[26] R. Johnson, Exponential dichotomy, rotation number, and linear differential operators with bounded coefficients, *J. Diff. Equations* **61** (1986), 54-78.

[27] S. Kotani, Ljapunov indices determine absolutely continuous spectra of stationary random one-dimensional Schrödinger operators, Stochastic analysis (Katata/Kyoto, 1982), 225–247, North-Holland Math. Library, 32, North-Holland, Amsterdam, 1984.

[28] S. Kotani, One-dimensional random Schrödinger operators and Herglotz functions, Probabilistic methods in mathematical physics (Katata/Kyoto, 1985), 219–250, Academic Press, Boston, MA, 1987.

[29] S. Kotani, Generalized Floquet theory for stationary Schrödinger operators in one dimension, *Chaos Sol. Fract.* **8** (1997), 1817–1854.

[30] V. Marchenko, Some questions in the theory of one-dimensional linear differential operators of the second order, I, Trudy Moskov. Mat. Obšč. **1** (1952), 327–420 (Russian); English transl. in Amer. Math. soc. Transl. **101** (1973), 1–104.

[31] V.A. Marchenko, The Cauchy problem for the KdV equation with non-decreasing initial data, in *What is integrability?,* 273–318, Springer, Berlin, 1991.

[32] V.A. Marchenko and I.V. Ostrovskii, A characterization of the spectrum of the Hill operator (Russian), *Mat. Sb. (N.S.)* **97** (1975), 540–606.

[33] J.R. Munkres, Topology, Prentice Hall, Englewood Cliffs, NJ, 1975.

[34] L. Pastur and A. Figotin, Spectra of Random and Almost Periodic Operators, Grundlehren der mathematischen Wissenschaften 297, Springer-Verlag, Berlin, 1992.

[35] F. Peherstorfer, Deformation of minimal polynomials and approximation of several intervals by an inverse polynomial mapping, *J. Approx. Theory* **111** (2001), 180–195.

[36] R.R. Phelps, Lectures on Choquet's Theorem, 2nd edition, Lecture Notes in Mathematics, 1757, Springer-Verlag, Berlin, 2001.

[37] A. Poltoratski and C. Remling, Reflectionless Herglotz functions and Jacobi matrices, *Comm. Math. Phys.* **288** (2009), 1007–1021.

[38] A. Poltoratski and C. Remling, Approximation results for reflectionless Jacobi matrices, *Int. Math. Res. Not.* **16** (2011), 3575–3617.

[39] T. Ransford, Potential Theory in the Complex Plane, London Mathematical Society Students Texts 28, Cambridge University Press, Cambridge, 1995.

[40] C. Remling, Schrödinger operators and de Branges spaces, *J. Funct. Anal.* **196** (2002), 323–394.

[41] C. Remling, The absolutely continuous spectrum of one-dimensional Schrödinger operators, *Math. Phys. Anal. Geom.* **10** (2007), 359–373.

[42] C. Remling, Uniqueness of reflectionless Jacobi matrices and the Denisov-Rakhmanov theorem, *Proc. Amer. Math. Soc.* **139** (2011), 2175–2182.

[43] C. Remling, The absolutely continuous spectrum of Jacobi matrices, *Annals of Math. (2)* **174** (2011), 125–171.

[44] E.B. Saff and V. Totik, Logarithmic Potentials with External Fields, Grundlehren der mathematischen Wissenschaften 316, Springer, Berlin, 1997.

[45] S. Schwartzman, Asymptotic cycles, *Annals of Math. (2)* **66** (1957), 270-284.

[46] B. Simon, Equilibrium measures and capacities in spectral theory, *Inv. Problem Imaging* **1** (2007), 713–772.

[47] B. Simon, Szegő's Theorem and Its Descendants, Princeton University Press, Princeton, NJ, 2011.

[48] M. Sodin and P. Yuditskii, Almost periodic Jacobi matrices with homogeneous spectrum, infinite-dimensional Jacobi inversion, and Hardy spaces of character-automorphic functions, *J. Geom. Anal.* **7** (1997), 387–435.

[49] G. Teschl, Jacobi Operators and Completely Integrable Nonlinear Lattices, Mathematical Monographs and Surveys, 72, American Mathematical Society, Providence, 2000.

[50] V. Totik, Polynomial inverse images and polynomial inequalities, *Acta Math.* **187** (2001), 139–160.

[51] H. Winkler, The inverse spectral problem for canonical systems, Integral Equations Operator Theory 22 (1995), no. 3, 360–374.