ERGODIC JACOBI MATRICES AND CONFORMAL MAPS

INJO HUR AND CHRISTIAN REMLING

Abstract. We study structural properties of the Lyapunov exponent γ and the density of states k for ergodic (or just 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.

1. Introduction and basic setup

1.1. **Introduction.** In this paper, 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 [15], which is perhaps best known (definitely to us) through the reinterpretation of this material that was given in [8, 9]. 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 [5, 13, 14]. 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.

Date: October 31, 2011.

Key words and phrases. Jacobi matrix, density of states, Lyapunov exponent.

2000 *Mathematics Subject Classification.* Primary 47B36 81Q10; Secondary 30C20.

CR's work supported by NSF grant DMS 0758594.

1.2. **Basic setup.** Let us now try to give a somewhat more detailed description of what we will do here. Recall that a *Jacobi matrix* is a difference operator on $u \in \ell^2$ of the form

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

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

$$J = \begin{pmatrix} \ddots & \ddots & \ddots & & & & \\ & a_{-2} & b_{-1} & a_{-1} & & & \\ & & a_{-1} & b_0 & a_0 & & \\ & & & a_0 & b_1 & a_1 & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

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 $a_n > 0$, 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. It is often useful to make $\mathcal{J}_2$ a compact metric space; one possible choice for such a metric is

$$d(J, J') = \sum_{n \in \mathbb{Z}} 2^{-|n|} \left(|a_n - a_n'| + |b_n - b_n'| \right) . \tag{1.1}$$

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

$$\int_{\mathcal{J}_2} d\mu(J) \int_{[-2,2]} d\rho_0(t;J) f(t) = \int_{[-2,2]} f(t)\, dk(t) \tag{1.2}$$

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 2.2 below. In particular, this function is measurable and thus the left-hand side of (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

$$w(z) = \ln A - \int_{[-2,2]} \ln(t-z)\, dk(t), \tag{1.3}$$

for $z \in \mathbb{C}^+$, the upper half plane in $\mathbb{C}$. Here we take the logarithm with $\text{Im } \ln\zeta \in (-\pi, 0)$ 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 [6, 17, 31] for (much) more on these topics. Definition (1.3) is straightforward and convenient for our purposes.

1.3. **Overview of main themes.** As already announced, one of the recurring themes of this paper will be the generalized version of the observation of Marchenko and Ostrovskii that w maps $\mathbb{C}^+$ conformally onto an image domain $w(\mathbb{C}^+)$ of a certain type, and, conversely, these domains can be used to reconstruct w, A, and dk (in fact, this is not literally true; it becomes true after a suitable change of variables, as we'll discuss below). See the discussion of Section 2, especially Proposition 2.4. A variety of other data are available, and we study the relations between these in some detail in Sections 3 and 4. Section 5 contains one of the main results of this paper, Theorem 5.4: Given suitable data (for example, given a w function), we can find an invariant measure $\mu \in \mathcal{M}_0$ that will produce these data. A less complete result of this type was proved earlier in [5]. Other main results include Theorem 4.2, especially part (f), and the results from Sections 6 and 7.

Section 6 introduces a new topic. Here we show that the correspondence between the gaps of the spectrum and slits of the image domain that is one of the cornerstones of the Marchenko-Ostrovskii method (and obvious in the original setting) extends to the general case, if suitable definitions are made. In Section 7, we study the Lyapunov exponent as a function on $[-2,2]$ (rather than as a harmonic function on $\mathbb{C}^+$).

Sections 2–7 form the main part of this paper. The final three sections are lighter in tone. In Section 8, we revisit work of Avila-Damanik [2] on the positivity of generic Lyapunov exponents from the point of view suggested by the material of this paper. Section 9 offers a brief discussion of the possibility of finding *ergodic* (and not just invariant) measures μ, but what we have to say here does not really go beyond the work of Kotani [13], and we present more questions than answers. In the final section, we give an easy argument for the invariance of $w(z)$ under a class of transformations that includes all Toda flows.

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 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, [31, Chapter 5]. A discussion of these issues for measures μ that are just invariant may be found in [5].

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 $\mathbb{R}$.

Define

$$k_0(t) = \lim_{y \to 0+} k_0(t+iy); \tag{2.1}$$

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

$$w(z) = C_0 + Dz + \int_{-\infty}^{\infty} \left(\frac{1}{t-z} - \frac{t}{t^2+1} \right) k_0(t)\, dt. \tag{2.2}$$

In fact, as $\operatorname{Im} w(z) < \pi$ on $\mathbb{C}^+$, we must have $D = 0$ here. By differentiating (1.3), we obtain that $w'(z) = \int \frac{dk(t)}{t-z}$, so $\operatorname{Im} w'(z) > 0$, 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 (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 (2.2) as follows:

$$w(z) = C_0 + \int_{-\infty}^{\infty} \frac{\partial}{\partial t}\left[\ln(t-z) - \frac{1}{2}\ln(t^2+1)\right] k_0(t)\, dt \tag{2.3}$$
$$= C_0 + \lim_{R\to\infty} k_0(t) \ln \frac{t-z}{\sqrt{t^2+1}}\Big|_{t=-R}^{t=R}$$
$$- \int_{-\infty}^{\infty} \left[\ln(t-z) - \frac{1}{2}\ln(t^2+1)\right]\, dk_0(t)$$
$$= C - \int_{-\infty}^{\infty} \ln(t-z)\, dk_0(t).$$

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 (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 [31, Chapter 5]. So we will only explain how to generalize the inequality to invariant μ. As mentioned above, this issue is also discussed in [5]; we offer an easy alternative argument here.

Let

$$F_\epsilon(J) = \frac{1}{1+\epsilon}(J + \epsilon J_0), \tag{2.4}$$

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, 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 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 2.2. Let $f \in C[-2,2]$. Then, from (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

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

Hence also $\gamma_\epsilon \to \gamma$, so $\gamma \ge 0$. 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 2.3. *$A \le 1$ for any $\mu \in \mathcal{M}_0$.*

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

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

By the definition of logarithmic capacity [22, Definition 5.1.1], the double integral is $\le \ln \operatorname{cap}\,[-2,2] = 0$. $\square$

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 don't want to give any details here, but see Proposition 9.2 below and its discussion for a more general argument of this type. The material from [23] is also closely related to these issues.

We already mentioned several times 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

$$\zeta \mapsto z = z(\zeta) = -\zeta - \frac{1}{\zeta} \tag{2.5}$$

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 Re $w < 0$, $0 <$ Im $w < \pi$.

Proposition 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's 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 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 $\text{Im } w'(z) > 0$ on $\mathbb{C}^+$, which we already observed (and used) in the proof of Proposition 2.1. 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. Hence F is a conformal map, as claimed. $\square$

We remark in passing that the Schwarz Lemma now provides another simple proof of Lemma 2.3.

Proposition 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 either $L_v = \emptyset$, or there is an interval $(a, b) \subset L_v$, with $a, b \notin L_v$ and $a < b \leq 0$. This follows because L_v is open, so if this set is non-empty and not just a half-line, then we can take some other component, which will necessarily be bounded. Now $L_v = \emptyset$ is clearly impossible because k takes values arbitrarily close to 0 and also other values that come arbitrarily close to 1, and $w(\mathbb{C}^+)$ is connected.

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 $\text{Im}\, w'(z) > 0$, as we observed above, so

$$\frac{\partial k(x + iy)}{\partial x} > 0. \tag{2.6}$$

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 first property of Ω for $0 < \alpha < \pi$, and, by reflection, also for $-\pi < \alpha < 0$. Here, we have already made use of the invariance of Ω under reflection about the real line, but this second 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 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

$$\Omega \cap (0, 1) = \{e^{-\gamma(x)} : x < -2\}. \tag{2.7}$$

The Thouless formula (Proposition 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

$$h(e^{i\alpha}) := \sup\{r \geq 0 : (1-r)e^{i\alpha} \notin \Omega\}. \tag{2.8}$$

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 (2.8) has these properties. Conversely, if an upper semicontinuous, symmetric h is given, then Ω_h will be as described in (a). □

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

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

where

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

If $t \in E =$ top supp dk and t is not an endpoint of a component $(a, b) \subset (-2, 2) \setminus E$, then there is no $s \ne t$ 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

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

We want to show that

$$H(v) = L(v). \tag{2.9}$$

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 $H(v) \ge L(v)$.

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

$$H(v) \ge \gamma(t) + \epsilon \tag{2.10}$$

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 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 (2.10) demands that $\gamma(t_0) \le H(v) - \epsilon$. 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 (2.10). We have established (2.9).

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 (2.7). $\square$

3. Data sets

Let us summarize: Starting out from an invariant 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

$$-\ln A + \int_{[-2,2]} \ln |t - z| \, dk(t) \geq 0 \quad \text{for } z \in \mathbb{C}. \tag{3.1}$$

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 3.1. 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 (3.1) 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 $\mathbb{C} \backslash [-2, 2]$ 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 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 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 3.1. *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 3.1. 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 5.4 below. For now, it will be important to observe that nowhere in the developments that started with Proposition 2.4 did we use the fact that (A, dk) were obtained from a $\mu \in \mathcal{M}_0$; rather, it was only property (3.1) that mattered. Similarly, Proposition 2.1 continues to hold if we just assume (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 3.1 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.

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 [7, 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 3.2. Let $\Omega_n, \Omega \subset D$ be subdomains of the unit disk of the type discussed in Proposition 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 [7, Definition 15.4.1] says in the present context, observe that the *kernel* with respect to $z_0 = 0$ (as defined in [7]) 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 [7]. 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 3.2. *Let $\Omega_n, \Omega \subset D$ be domains of the type discussed in Proposition 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 3.2. 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 $\sup h_n \leq 1 - \delta$, 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 $h \leq 1 - r - 2\epsilon$, 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 $0 \leq \varphi \leq 1$. Assumption (b) then says that for all sufficiently large n, we will also have that $h_n(e^{i\beta}) \leq 1 - r - \epsilon$, say, uniformly on some neighborhood $|\beta - \alpha| \leq \eta$. 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 3.2. 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 $r \geq 1 - h(e^{i\alpha})$. Pick a function $0 \leq \varphi \leq 1$ that is supported by $|\beta - \alpha| \leq \delta/10$ and equal to 1 at $e^{i\alpha}$. Condition (b) then provides angles β_n from this neighborhood so that $h_n(e^{i\beta_n}) \geq 1 - r - \delta/2$ 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 $\varphi \in C(S^1)$, $\varphi \geq 0$. We would first like to show that

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

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 3.2 shows that for any $\delta > 0$ and all large $n \ge N_0 = N_0(\delta)$, we must have $h_n(e^{i\beta}) \ge h(e^{i\alpha}) - \delta$ somewhere on $\alpha - \delta < \beta < \alpha + \delta$, say. Since φ is continuous, it will satisfy $\varphi \ge 1 - \eta$ 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 $\sup \varphi h_n \ge \sup \varphi h - \delta - \eta$ 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

$$\sup \varphi h \ge \limsup_{n\to\infty} (\sup \varphi h_n). \tag{3.2}$$

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 (3.2), and then pass to a subsequence a second time to make the points converge, say $\alpha_n \to \alpha$. If (3.2) were wrong, we would have that $\varphi(e^{i\alpha})h(e^{i\alpha}) \le \varphi(e^{i\alpha_n})h_n(e^{i\alpha_n}) - \epsilon_0$, 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

$$h(e^{i\alpha}) \le h_n(e^{i\alpha_n}) - \epsilon, \tag{3.3}$$

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) 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 3.2. $\square$

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 [7, Sections 14.2, 14.3] for a careful discussion; the result just mentioned is stated as Theorem 3.4 of [7, 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 [7, Sections 14.2, 14.3] will provide more than adequate background.

After these digressions, we now return to Theorem 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. *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 $A = \text{cap } E$, $dk = d\omega_E$.*

Here, cap E again denotes the logarithmic capacity of E, and ω_E is the equilibrium measure of E. Please see [22, 26] 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 $z \mapsto F = e^{w(z)}$.

Note also that $\int \ln |t - z|\, d\omega_E(t) \geq \ln \text{cap } E$ for all $z \in \mathbb{C}$ for a finite gap set E, so (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 [7, 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 putative 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 (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

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

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

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

As explained above, E is a finite gap set; it is the inverse image under (2.5) of a finite disjoint union of closed subarcs of ∂D. Since, under (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 W 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 this set under W to the union of the slits and 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 (2.3) again. We obtain that

$$W(z) = C - \int_{[-2,2]} \ln(t - z)\, dk(t). \tag{3.4}$$

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 $\text{Re } W < 0$ on $\mathbb{C}^+$, and now the arguments from the proof of Proposition 2.1 yield (3.4) on $z \in \mathbb{R}$ also.

Let's take a look at

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

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 [26, Theorem I.3.1] and also Remark 1.5 from Section I.1 of this reference.

So if we use these data $(A, dk) = (\text{cap } E, d\omega_E) \in \mathcal{D}$ 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 3.1. We will focus on the existence part ("Conversely, ...") exclusively. Indeed, except for small details, which we leave to the reader to fill in, our discussion from Section 2 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 (3.1); 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 3.1 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. Therefore, the Herglotz representation of W' takes the form

$$W'(z) = C + Dz + \int_{[-2,2]} \frac{dk(t)}{t - z}, \tag{3.5}$$

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 (3.1), 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 (3.5). 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 3.1 shows that dk is a probability measure. We can again write $B = -\ln A$, with $A > 0$, and (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, 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

$$h_n(e^{i\alpha}) = \begin{cases} H_n(j) & \alpha = j\pi/n \quad (j = 0, 1, \ldots, n) \\ 0 & \text{otherwise} \end{cases}; \tag{3.6}$$

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

$$H_n(j) = \sup_{-1/n \le \delta \le 1/n} h\left(e^{i\pi(j/n+\delta)}\right). \tag{3.7}$$

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 3.2 holds. The argument is quite similar to what we did in the second part of the proof of this Lemma. Let $\varphi \in C(S^1)$, $\varphi \ge 0$. 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 $|\beta_n - \alpha| \le \pi/n$. 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:

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

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

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

Lemma 3.2 now informs us that $\Omega_n \to \Omega$ in the sense of kernel convergence. By Carathéodory's Theorem [7, 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, $G_n(\zeta) = e^{w_n(z)}$ and

$$w_n(z) = \ln A_n - \int \ln(t - z)\, dk_n(t) \tag{3.8}$$

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 $[-2, 2]$.

Taking limits in (3.8), we 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 $\text{Re}\, w < 0$, so (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 [3, 18, 32] but see also [29, 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 (3.6), (3.7), together with material that we will discuss in the following section, could be used to give a tremendously simplified treatment.

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 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 [7, Theorem 15.4.10] has to say in the case at hand. Finally, (e) $\iff$ (f) is Lemma 3.2. $\square$

These spaces $\mathcal{D}$, $\mathcal{W}$ etc. from Definition 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 4.1 for approach to this added object.

Theorem 4.2. *Let* $(A_n, dk_n) \in \mathcal{D}$*, and introduce the corresponding objects, as in Theorem 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 $n \ge N$. 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 [7, Definition 15.4.1] for background information.

Proof. This is similar to the previous proof. It's 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 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 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 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. $\square$

We would like to emphasize one point here that was already made implicitly in our proof of Theorem 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 (3.1) holds; (ii) $B > 0$, but (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 4.3. *Let $(A_n, dk_n) \in \mathcal{D}$ and suppose that $A_n \to B \geq 0$ 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 (3.1) is asking for, so $(B, d\nu) \in \mathcal{D}$, as claimed. □

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 3.1. We first introduce a metric in such a way that convergence with respect to this metric is equivalent to the conditions discussed in Theorem 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 4.2.

There is, of course, general theory underlying this procedure; see, for example, [16]. 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 2.3 and 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 4.1(a) and 4.2(a).

We can give similar metrics on the (one-point compactifications of the) other spaces from Definition 3.1. Alternatively, we can just use Theorem 3.1 and Theorems 4.1, 4.2 to move things over from $\mathcal{D}_0$ to those spaces. We summarize:

Proposition 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 4.1 and 4.2, respectively. These spaces are compact.*

5. Existence of invariant measures

We now come to one of the main points of the whole discussion so far. We also want to show that given data as in Definition 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 [5]. (Such a result also appears here, as Proposition 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 (3.6), (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 5.1. *Suppose that $\mu_n \in \mathcal{M}_0$ and $\mu_n \to \mu$ in weak-$*$ sense. Then*

$$A_\mu \geq \limsup_{n\to\infty} A_{\mu_n}. \tag{5.1}$$

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 $\mathcal{J}_2$.

This Lemma is supplemented by Lemma 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 (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

$$\int \ln(a_0(J) + \epsilon)\, d\mu_n(J) \to \int \ln(a_0(J) + \epsilon)\, d\mu(J). \tag{5.2}$$

Moreover, $\int \ln(a_0 + \epsilon)\, d\mu \to \ln A_\mu \in [-\infty, \infty)$ as $\epsilon \to 0+$ by monotone convergence, so if (5.1) failed, then we could find a subsequence and $\epsilon > 0$ so that

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

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 (5.2). □

Proposition 5.2. *Suppose that $\Gamma \in \mathcal{L}$. Then there exist $\mu \in \mathcal{M}_0$ and $d \ge 0$ 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 2, via (A_μ, dk_μ) and (1.3).

Proof. This is similar to the argument we used in the proof of Theorem 3.1 to construct (A, dk), given a conformal map G. First of all, Theorem 3.1 provides us with associated data (A, dk), $W(z)$, $G(\zeta)$, Ω, h. Define again approximating finite gap domains as in (3.6), (3.7), and denote the corresponding data by A_n, dk_n, w_n etc. By Lemma 3.3, there are finite gap sets $E_n \subset [-2, 2]$ so that $A_n = \text{cap } E_n$ 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 $\sigma(J) = E$ 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 [31, Chapter 9], but see also [21, 24] 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: (i)

$\gamma_\mu = 0$ almost everywhere with respect to ω_E; (ii) dk_μ is supported by E. Compare the final part of the proof of Lemma 3.3 for this step, or, better yet, see Proposition 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 [31, Chapter 8] or [20, 24] for background. It follows from (the easy Ishii-Pastur part of) Kotani theory [12] that $\gamma_\mu = 0$ (Lebesgue, hence ω_E) almost everywhere on E. Alternative arguments are available, too; for example, [20] 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.

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 4.1(f), so we also have part (a) of the Theorem and, in particular, $A_n \to A > 0$. Thus Lemma 5.1 guarantees that $\mu \in \mathcal{M}_0$. Lemma 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 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 5.1 to confirm that $A_\mu \ge A$, so $d \ge 0$, as claimed. The final claim is obvious from this, since $\gamma_\mu \ge 0$. $\square$

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

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

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

Theorem 5.4. *Suppose that an object as in one of the parts of Definition 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 5.4, as follows. First of all, by Theorem 3.1, it suffices to discuss the case where a $\Gamma \in \mathcal{L}$ is given. Proposition 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 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 5.3.

Proof of Lemma 5.3. By Theorem 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

$$\ln A_j = \inf_{x\in\mathbb{R}} \int \ln|t - x|\, dk_j(t); \tag{5.3}$$

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 5.2 provides us with measures $\mu_j \in \mathcal{M}_0$ so that $\Gamma_j = \gamma_{\mu_j}$. Let $\mu = \sum c_j \mu_j$, and also observe that $\ln A_j < -\ln N$; indeed, it suffices to take x as the center of I_j in (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. □

6. Slits and gaps

Recall the definitions made in the context of Theorem 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, [10, 27] for the classical results and [1] 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 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 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 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 $E =$ 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 $k(t) \neq k_0$ if $t \in [-2,2] \setminus [a,b]$. In this situation, Theorem 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 $\gamma(t) \le \Gamma - \epsilon$ for all $a - \epsilon \le t \le a$ for some small $\epsilon > 0$ because γ is upper semicontinuous. Now $k(a-\epsilon) < k(a) = k_0$, so Theorem 2.6 implies that $h(e^{i\pi k}) \le h(e^{i\pi k_0}) - \delta$ 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

$$\limsup_{k \to k_0-} h\left(e^{i\pi k}\right) < h\left(e^{i\pi k_0}\right) \equiv h_0, \tag{6.1}$$

the other case being analogous, of course. In more geometric terms, assumption (6.1) means that $\partial\Omega_h$ contains an exposed line segment

$$S = \left\{ re^{i\pi k_0} : 1 - h_0 + \epsilon < r < 1 - h_0 + 2\epsilon \right\} \tag{6.2}$$

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 (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 [7, 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 [7, 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 (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 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, [7, 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 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 6.2. An alternative, more direct proof that is based on Theorem 2.6 is also possible.

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 [26, 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 4.1, 4.2.

Theorem 7.1. *Let $\gamma_n, \gamma \in \mathcal{L}$. Then the following conditions are also equivalent to those from Theorem 4.1:*
(a)

$$\sup_{-2\le x\le 2} \varphi(x)\gamma(x) = \lim_{n\to\infty} \sup_{-2\le x\le 2} \varphi(x)\gamma_n(x) \tag{7.1}$$

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

$$\lim_{n\to\infty} \int_{[-2,2]} |\gamma(x) - \gamma_n(x)|\, d\nu(x) = 0. \tag{7.2}$$

Here, we let $\mathcal{P}$ be the set of probability measures ν on the Borel sets of $[-2,2]$ for which the potential

$$\Phi_\nu(x) \equiv \int_{\mathbb{R}} \ln|t-x|\, d\nu(t) \tag{7.3}$$

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 [26, 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,

$$\nu(|\gamma - \gamma_n| \ge \epsilon) \to 0 \quad \text{for every } \epsilon > 0, \tag{7.4}$$

we are *not* claiming that the *capacity* of the set where $|\gamma_n - \gamma| \ge \epsilon$ 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 3.1 (compare (3.6), (3.7)). Lemma 3.3 then shows that

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

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 7.2. *Let $\gamma_n \in \mathcal{L}$. Then the conditions from Theorem 4.2 are equivalent to:*
(a)

$$\lim_{n\to\infty} \sup_{-2\le x\le 2} \gamma_n(x) = \infty. \tag{7.5}$$

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

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

Since (7.1) and (7.5) are analogous to conditions (f) from Theorems 4.1 and 4.2, respectively, and, moreover, γ and h are directly related through changes of variables (and a partial maximization), as spelled out in Theorem 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 7.2. We start with this because we will use Theorem 7.2 in our proof of Theorem 7.1. The equivalence of (a) with the conditions of Theorem 4.2 is an immediate consequence of Theorem 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 (7.5) holds if and only if $\sup h_n \to 1$, which is condition (f) from Theorem 4.2.

Next, assume that $A_n \to 0$ (this is (a) of Theorem 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 [26, 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 (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 7.1. We first show that the conditions of Theorem 4.1 imply (a). Let $\varphi \in C[-2,2]$, $\varphi \geq 0$ be given. As in the proof of Lemma 3.2, we will split (7.1) into two inequalities. We first show that

$$\sup \varphi\gamma \geq \limsup_{n\to\infty} \left(\sup \varphi\gamma_n\right). \tag{7.6}$$

Since the functions $\varphi\gamma_n$ are upper semicontinuous, the suprema are maxima, so if (7.6) were wrong, we would find ourselves in the following situation:

$$\sup \varphi\gamma \leq \varphi(x_n)\gamma_n(x_n) - \epsilon, \tag{7.7}$$

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 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 $\limsup$ is taken along some subsequence, but this is irrelevant here). Since φ is continuous, this contradicts (7.7), unless $\varphi(x) = 0$. However, if $\varphi(x) = 0$, then (7.7) implies that $\gamma_n(x_n) \to \infty$, and we again obtain a contradiction, this time to Theorem 7.2. We have established (7.6).

Next, we show that also

$$\sup \varphi\gamma \leq \liminf_{n\to\infty} \left(\sup \varphi\gamma_n\right), \tag{7.8}$$

and this together with (7.6) will of course establish (7.1). Again, we argue by contradiction. If (7.8) failed, then we would find a subsequence and $x \in [-2,2]$ so that

$$\varphi(x)\gamma(x) \geq \varphi(t)\gamma_n(t) + 2\epsilon \tag{7.9}$$

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 $\limsup$ 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 (7.9) with $t = x$.

To prove that, conversely, (a) above implies part (a) from Theorem 4.1, we again exploit the compactness properties that were discussed in Sections 4. Suppose that (7.1) holds. We can pass to a subsequence so that $A_n \to B$, $dk_n \to d\nu$. Here, by Lemma 4.3, either $B = 0$ or $(B, d\nu) \in \mathcal{D}$. The first case is impossible because then Theorem 7.2 would imply that (7.5) holds on the subsequence we chose, but this is clearly incompatible with our assumption that we have (7.1).

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

$$\lim\left(\sup \varphi\gamma_n\right) = \sup \varphi\gamma_{(B,d\nu)} \tag{7.10}$$

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 $\gamma, \widetilde{\gamma} \in \mathcal{L}$ are not the same function, then

$$\sup_{-2\leq x\leq 2} \varphi(x)\gamma(x) \neq \sup_{-2\leq x\leq 2} \varphi(x)\widetilde{\gamma}(x) \tag{7.11}$$

for some nonnegative $\varphi \in C[-2,2]$. Indeed, if $\gamma(x_0) < \widetilde{\gamma}(x_0)$, 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 (7.11) holds.

This uniqueness means that (7.10) forces $\gamma_{(B,d\nu)}$ to be the function γ from (7.1), and thus, by the uniqueness part of Theorem 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 4.1.

Next, we again assume the conditions from Theorem 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

(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 (7.2). Fix a $\nu \in \mathcal{P}$. We will show that $\gamma_n \to \gamma$ in measure, that is, (7.4) holds. This is sufficient because, as just discussed, $0 \le \gamma_n, \gamma \le C$, so $L^1(\nu)$ convergence will follow from this.

We will argue by contradiction and thus assume hypothetically that (7.4) fails. Then there exists $\epsilon > 0$ so that

$$\nu(|\gamma - \gamma_n| \ge \epsilon) \ge \epsilon \tag{7.12}$$

for all n taken from some subsequence.

Recall that $\gamma(x) \ge \limsup \gamma_n(x)$ 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

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

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 (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 (7.4).

Conversely, if (b) is assumed, we repeat the argument from above: Consider any subsequence on which $A_n \to B \ge 0$, $dk_n \to d\rho$. 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 7.2(b) would apply on the corresponding subsequence, and this is incompatible with our assumption that (7.2) holds. So $(B, d\rho) \in \mathcal{D}$ by Lemma 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 3.1. □

8. Positive Lyapunov exponents

In this section, we present a variation on a theme composed by Avila and Damanik [2]. 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 7.1(b).

Lemma 8.1. *Let $\nu \in \mathcal{P}$. For any $a, b \ge 0$, the set*

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

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 4.1(c) or one of the equivalent descriptions of this mode of convergence.

Given $\epsilon > 0$, no matter how small, Theorem 7.1(b), or rather its consequence (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 $-2 \le x \le 2$, $x \notin \mathcal{E}$. Since $\gamma_N \in S(a,b)$ by assumption, this implies that

$$\nu(\gamma \le a + \epsilon) \ge b - \epsilon.$$

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 $\gamma \mapsto \nu(\gamma \le a)$ is upper semicontinuous. Compare this formulation with [2, Lemma 1].

Corollary 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 4.4 and its discussion in this context.

The Corollary has further implications because, by the classical Kotani theory [12], absolutely continuous spectrum for ergodic systems corresponds to zero Lyapunov exponents. See [2] for these aspects of the Corollary.

Proof. By Lemma 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

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

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.) $\square$

9. Ergodic measures

Return to the discussion of Section 5. We are given a $\Gamma \in \mathcal{L}$ (or other data with the properties from Definition 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 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 [13], and then point out some obvious open questions.

Proposition 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 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 [4] (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 [19], 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

$$\gamma_\mu(z) = \int_{\mathcal{M}} \gamma_\nu(z)\, d\sigma(\nu) \tag{9.1}$$

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 (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

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

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 $\gamma_\nu(z) \le \gamma_\mu(z) - \epsilon$. 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 [13, Theorem 6.3] and its proof.

Proposition 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, [28]. 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* [22, 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 $I(\omega_E) = \ln \operatorname{cap} E$. 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 3.1, 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 $\gamma_1 = \gamma_2 = \gamma$. □

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 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 [31, Chapter 9]), but also for certain sets E with infinitely many gaps and this property (we know this thanks to work of Sodin-Yuditskii [30]). It is not clear if there are other examples of Lyapunov exponents Γ with this property that there is only one (ergodic) μ with $\Gamma = \gamma_\mu$.

10. Invariance under Toda maps

In this final section, we show that w is invariant under maps of Toda type. We will give a simple abstract version of this result, which, at the same time, will also be more general. It is not necessary here to be familiar with the theory of Toda flows. We consider continuous maps $\varphi : \mathcal{J}_2 \to \mathcal{J}_2$ that preserve the shift dynamics in the sense that $S\varphi = \varphi S$. This also makes sure that the induced map $\mu \mapsto \varphi\mu$ on the probability measures on the Borel sets of $\mathcal{J}_2$ preserves the property of being an invariant measure; recall in this context that the image measure $\varphi\mu$ is defined by the condition that $\int f\, d(\varphi\mu) = \int f \circ \varphi\, d\mu$ for continuous functions f. The invariance of $\varphi\mu$ is most elegantly established by observing that a measure is S invariant precisely if it coincides with its image measure under S. Now the fact that S and φ commute implies that similarly $S\varphi\mu = \varphi S\mu$, and this latter measure equals $\varphi\mu$ by the invariance of μ.

Theorem 10.1. *Suppose that $\varphi : \mathcal{J}_2 \to \mathcal{J}_2$ is a bijective continuous transformation that commutes with the shift, $S\varphi = \varphi S$, and preserves spectra: $\sigma(\varphi(J)) = \sigma(J)$. Then $\varphi\mu \in \mathcal{M}_0$ for every $\mu \in \mathcal{M}_0$ and $w_{\varphi\mu} = w_\mu$.*

As alluded to above, the time one map of any Toda flow has these properties; please see [31, Chapter 12] and [25] for background. In particular, there is a reasonably large supply of such maps. Note, however, that while (classical) Toda flows act by unitary conjugation, we are *not* assuming here that $\varphi(J)$ and J are unitarily equivalent; the spectra are only conserved as sets.

The invariance of w under (genuine) Toda flows was established earlier by Knill [11]; see especially Theorem 5.1 of [11].

Proof. First of all, it suffices to prove this for ergodic measures $\mu \in \mathcal{M}_0$. To see this, we again use the ergodic decomposition of an invariant measure μ that was discussed in the previous section (see the proof of Proposition 9.1). So write $\mu = \int \nu\, d\sigma(\nu)$. It follows directly from the definitions that then similarly $\varphi\mu = \int \varphi\nu\, d\sigma(\nu)$, and the measures $\varphi\nu$, as well as the measures ν themselves, are ergodic σ-almost everywhere. So if we can show the invariance of w_ν under φ for ergodic ν, then we will obtain the general case from (9.1). A similar argument is possible concerning the claim that $\varphi\mu \in \mathcal{M}_0$.

Given an ergodic $\mu \in \mathcal{M}_0$, we will first approximate it by the measures $\mu_\epsilon = F_\epsilon \mu$. Here, we use the same notation as in the proof of Proposition 2.1; see (2.4) and the discussion that follows. Note that the μ_ϵ are also ergodic, and recall that $\gamma_{\mu_\epsilon}(z) \to \gamma_\mu(z)$ as $\epsilon \to 0+$, for $z \in \mathbb{C}^+$.

We will then approximate these measures μ_ϵ by periodic measures; here, we call a probability measure ρ on $\mathcal{J}_2$ *periodic* if it is of the form

$$\rho = \frac{1}{p} \sum_{j=1}^{p} \delta_{S^j J}$$

for some $J \in \mathcal{J}_2$ with $S^p J = J$. We formulate this step as a separate Lemma.

Lemma 10.2. *Let μ be an ergodic measure on $\mathcal{J}_2$. Then there are periodic measures ρ_n so that $\rho_n \to \mu$ in weak-$*$ sense. Moreover, if μ is supported by $\mathcal{J}_2^{(\epsilon)}$, then the ρ_n can be chosen so that they also have this property.*

Proof of Lemma 10.2. The ergodic theorem says that if $f \in C(\mathcal{J}_2)$ is given, then

$$\lim_{p\to\infty} \frac{1}{p} \sum_{j=1}^{p} f(S^j J) = \int_{\mathcal{J}_2} f(J)\, d\mu(J)$$

for μ-almost every choice of J. Since $C(\mathcal{J}_2)$ is separable, this implies that also

$$\mu = \lim_{p\to\infty} \frac{1}{p} \sum_{j=1}^{p} \delta_{S^j J} \tag{10.1}$$

in weak-$*$ sense for μ-almost every J. Fix such a J, and consider periodic modifications J' of J. More precisely, we let J' have the same coefficients as J on $n = 1, 2, \ldots, p$ for some $p \geq 1$, and then continue periodically. In other words, $(a', b')_n = (a, b)_n$ for $n = 1, 2, \ldots, p$, and the remaining coefficients are obtained from the condition that $(a', b')_{n+p} = (a', b')_n$ for all $n \in \mathbb{Z}$.

Recall now how the metric d on $\mathcal{J}_2$ was defined; see (1.1). Since $a, b \in \ell^\infty$, it follows that we can find a constant C so that

$$d(S^j J, S^j J') \leq C 2^{-p^{1/2}} \quad \text{for } p^{1/2} \leq j \leq p - p^{1/2}.$$

Indeed, if j is from this range, then the coefficients of $S^j J$ and $S^j J'$ agree on an interval centered at 0 of size at least $p^{1/2}$, and the estimate follows at once from this. If we now replace J with $J' = J'_p$ in (10.1), then the periodic measures obtained in this way will still converge to μ because $S^j J'$ will be uniformly close to $S^j J$ for the lion's share of the sum, and because of the factor $1/p$, the remaining $\approx p^{1/2}$ summands cannot make an appreciable contribution.

This procedure also establishes the final claim because all J'_p will be in $\mathcal{J}_2^{(\epsilon)}$ if J was from this subspace. □

Apply this to the measures μ_ϵ. We obtain periodic measures $\mu_{n,\epsilon} \to \mu_\epsilon$. Now for a periodic measure ρ, we certainly have that $\gamma_{\varphi\rho} = \gamma_\rho$. This follows because if $\rho = (1/p)\sum \delta_{S^j J}$, then $\varphi\rho = (1/p)\sum \delta_{\varphi S^j J}$, but since φ commutes with S, this is again a periodic measure, and it is formed with the periodic Jacobi matrix $\varphi(J)$. For a periodic operator J and the associated measure ρ, the corresponding data are $(A_\rho, dk_\rho) = (\text{cap } E, d\omega_E)$, where $E = \sigma(J)$ (compare also our discussion of finite gap domains in this context). So they only depend on the spectrum of J, but we assumed that φ preserves this.

Thus $\gamma_{\varphi\mu_{n,\epsilon}} = \gamma_{\mu_{n,\epsilon}}$. We now send $n \to \infty$. Since the measures $\mu_{n,\epsilon}$ are all supported by $\mathcal{J}_2^{(\epsilon)}$, we can be sure that $\gamma_{\mu_{n,\epsilon}} \to \gamma_{\mu_\epsilon}$; compare again the proof of Proposition 2.1 for this step. There is no such additional information available for the measures $\varphi\mu_{n,\epsilon}$, so we will just use Lemmas 2.2 and 5.1 here. Notice that we do know that $\varphi\mu_{n,\epsilon} \to \varphi\mu_\epsilon$ as $n \to \infty$. It follows that

$$\gamma_{\mu_\epsilon}(z) = \gamma_{\varphi\mu_\epsilon}(z) + c_\epsilon,$$

with $c_\epsilon \geq 0$. If we now also take $\epsilon \to 0+$, then, as just explained, the left-hand side will converge to γ_μ. On the right-hand side, we again refer to Lemmas 2.2 and 5.1 to conclude that

$$\gamma_\mu(z) = \gamma_{\varphi\mu}(z) + c \qquad (c \geq 0). \tag{10.2}$$

In particular, we have learnt from this argument that $-\ln A_{\varphi\mu_\epsilon}$ stays bounded as $\epsilon \to 0+$, so Lemma 5.1 does make sure that $\varphi\mu \in \mathcal{M}_0$, as claimed.

To obtain the full assertion of the Theorem, all that remains to be done is to let $\varphi\mu$ and $\mu = \varphi^{-1}(\varphi\mu)$ swap roles. So only $c = 0$ is possible in (10.2). Since γ and w determine each other, the claim may be phrased in terms of w, which is what we did in the formulation of the theorem. □

References

[1] A. Avila, J. Bochi, and D. Damanik, Opening gaps in the spectrum of strictly ergodic Schrödinger operators, preprint.

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

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

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

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

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

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

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

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

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

[11] O. Knill, Isospectral deformations of random Jacobi operators, *Comm. Math. Phys.* **151** (1993), 403-426.

[12] S. Kotani, Lyapunov indices determine absolutely continuous spectra of stationary one-dimensional Schrödinger operators, in Stochastic Analysis (Katata/Kyoto, 1982), 225-247, North-Holland Math. Library **32**, North-Holland, Amsterdam, 1984.

[13] S. Kotani, One-dimensional random Schrödinger operators and Herglotz functions, in Probabilistic Methods in Mathematical Physics (Katata/Kyoto, 1985), 219-250, Academic Press, Boston, MA, 1987.

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

[15] V.A. Marcenko and I.V. Ostrovskii, A characterization of the spectrum of the Hill operator (Russian), *Mat. Sb. (N.S.)* **97** (1975), 540–606.

[16] J.R. Munkres, Topology, Prentice Hall, Englewood Cliffs, NJ, 1975.

[17] L. Pastur and A. Figotin, Spectra of Random and Almost Periodic Operators, Grundlehren der mathematischen Wissenschaften 297, Springer-Verlag, Berlin, 1992.

[18] F. Peherstorfer, Deformation of minimal polynomials and approximation of several intervals by an inverse polynomial mapping, *J. Approx. Theory* **111** (2001), 180–195.

[19] R.R. Phelps, Lectures on Choquet's Theorem, 2nd edition, Lecture Notes in Mathematics, 1757, Springer-Verlag, Berlin, 2001.

[20] A. Poltoratski and C. Remling, Reflectionless Herglotz functions and Jacobi matrices, *Comm. Math. Phys.* **288** (2009), 1007-1021.

[21] A. Poltoratski and C. Remling, Approximation results for reflectionless Jacobi matrices, *Int. Math. Res. Not.* **16** (2011), 3575–3617.

[22] T. Ransford, Potential Theory in the Complex Plane, London Mathematical Society Students Texts 28, Cambridge University Press, Cambridge, 1995.

[23] C. Remling, Uniqueness of reflectionless Jacobi matrices and the Denisov-Rakhmanov Theorem, *Proc. Amer. Math. Soc.* **139** (2011), 2175–2182.

[24] C. Remling, The absolutely continuous spectrum of Jacobi matrices, *Annals of Math. (2)* **174** (2011), 125–171.

[25] C. Remling, General Toda flows, in preparation.

[26] E.B. Saff and V. Totik, Logarithmic Potentials with External Fields, Grundlehren der mathematischen Wissenschaften 316, Springer, Berlin, 1997.
[27] S. Schwartzman, Asymptotic cycles, *Annals of Math. (2)* **66** (1957), 270-284.
[28] B. Simon, Equilibrium measures and capacities in spectral theory, *Inv. Problem Imaging* **1** (2007), 713–772.
[29] B. Simon, Szegő's Theorem and Its Descendants, Princeton University Press, Princeton, NJ, 2011.
[30] 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.
[31] G. Teschl, Jacobi Operators and Completely Integrable Nonlinear Lattices, Mathematical Monographs and Surveys, 72, American Mathematical Society, Providence, 2000.
[32] V. Totik, Polynomial inverse images and polynomial inequalities, *Acta Math.* **187** (2001), 139–160.

Mathematics Department, University of Oklahoma, Norman, OK 73019

E-mail address: `ihur@math.ou.edu`

Mathematics Department, University of Oklahoma, Norman, OK 73019

E-mail address: `cremling@math.ou.edu`

URL: `www.math.ou.edu/~cremling`