ON A DENSITY PROBLEM RELATED TO A THEOREM OF SZEGŐ

CHIARA PAULSEN

Abstract. A classical theorem of Szegő states that for any probability measure $\mu = w\frac{d\theta}{2\pi} + \mu_s$ on the unit circle the polynomials are dense in $L^2(\mathbb{T}, \mu)$ if and only if $\log(w) \notin L^1(\mathbb{T})$. A related question asks whether the monomials with exponents in some subset $\Lambda \subseteq \mathbb{N}_0$ already span $L^2(\mathbb{T}, \mu)$ if $\log(w) \notin L^1(\mathbb{T})$. A result by Olevskii and Ulanovskii gives an answer if μ belongs to a class of absolutely continuous measures. We investigate the same question for Markoff measures.

1. Introduction

Let μ be a probability measure on the complex unit circle $\mathbb{T}$, the latter of which we shall identify with the interval $[0, 2\pi)$ in the usual way. Denote by

$$dm = \frac{d\theta}{2\pi}$$

the normalized Lebesgue measure on $\mathbb{T}$. By the Lebesgue decomposition theorem, one can decompose μ with respect to m

$$d\mu = w\, dm + d\mu_s$$

where μ_s and m are mutually singular and w is the Radon-Nikodym derivative of μ and m.

We call μ *non-degenerate* if $|\text{supp}(\mu)| = \infty$. In the following, we denote by $\mathcal{P}$ the set of all non-degenerate probability measures on $\mathbb{T}$.

An important theorem by Szegő [GS58] in the theory of orthogonal polynomials on the unit circle (OPUC) implies that for any $\mu \in \mathcal{P}$ the polynomials are dense in $L^2(\mathbb{T}, \mu)$ if and only if

$$\int_0^{2\pi} \log(w(\theta))\, d\theta = -\infty. \tag{1}$$

In fact, this equivalence already appeared in the work of Kolmogorov [Kol41] in the context of prediction theory. The connection between Kolmogorov's work on prediction theory and Szegő's work on OPUC was made by Krein [Kre45]. A contemporary discussion of the connection between Szegő's theorem and prediction theory can be found in [Bin12]. Let us briefly outline this connection. For every Gaussian stationary stochastic process $(X_n)_{n\in\mathbb{Z}}$ with zero mean, there exists $\mu \in \mathcal{P}$ such that

$$E[X_n\overline{X_0}] = \int_{\mathbb{T}} z^{-n}\, d\mu.$$

By Kolmogorov [Kol41, Lemma 4] there is an isomorphism between the Hilbert space $\mathcal{H}$ spanned by $\{X_n \mid n \in \mathbb{Z}\}$ and $L^2(\mathbb{T}, \mu)$ mapping X_n to z^n. Via this

2020 *Mathematics Subject Classification.* 42C05, 60G25.

isomorphism, the error of predicting X_0 with the knowledge of $\{X_n \mid n \leq -1\}$ is given by [Bin12, Theorem 3], [Kol41, Chapter 9]

$$E\Big[(X_0 - P_{(-\infty,-1]}X_0)^2\Big] = \exp\Big(\int_0^{2\pi} \log w \, \mathrm{d}m\Big)$$

where $P_{(-\infty,-1]}$ is the projection onto the subspace of $\mathcal{H}$ spanned by $\{X_n \mid n \leq -1\}$. Thus, $(X_n)_n$ is a deterministic process (that is, the past predicts the future with zero error) if and only if the associated measure μ fulfills (1), which, in turn, is equivalent to the polynomials being dense in $L^2(\mathbb{T}, \mu)$.

A measure $\mu \in \mathcal{P}$ that fulfills (1) is called *non-Szegő*. If the integral in (1) is finite then μ is called *Szegő*. We denote the class of all Szegő measures by Sz and put $Sz^c := \mathcal{P} \backslash Sz$.

The connection to prediction theory motivates the following questions: Is it already sufficient to know only parts of the past to determine the future events of $(X_n)_n$? If so, how large a part of the past can be 'forgotten'? In order to make the question more precise, consider the family of exponentials

$$E(\Lambda) := \operatorname{span}\{z^\lambda \mid z \in \mathbb{T}, \lambda \in \Lambda\}$$

for $\Lambda \subseteq \mathbb{N}_0$. Furthermore, define

$$\mathscr{A} := \{(\Lambda, \mathcal{C}) \mid \Lambda \subseteq \mathbb{N}_0, \mathcal{C} \subseteq \mathcal{P}, E(\Lambda) \text{ dense in } L^2(\mathbb{T}, \mu) \text{ for all } \mu \in \mathcal{C}\}.$$

Question 1. *For which $\Lambda \subseteq \mathbb{N}_0$ and $\mathcal{C} \subseteq \mathcal{P}$ is $(\Lambda, \mathcal{C}) \in \mathscr{A}$?*

Remark 1.1. We observe the following obvious facts.

1.) By Szegő's theorem, for any $\mu \in Sz$ and $\Lambda \subseteq \mathbb{N}_0$, $E(\Lambda)$ is not dense in $L^2(\mathbb{T}, \mu)$. Thus, for every pair $(\Lambda, \mathcal{C}) \in \mathscr{A}$, it follows that $\mathcal{C} \subseteq Sz^c$.
2.) One has the following implications

$$\Lambda' \subseteq \Lambda, (\Lambda', \mathcal{C}) \in \mathscr{A} \Rightarrow (\Lambda, \mathcal{C}) \in \mathscr{A},$$
$$\mathcal{C}' \subseteq \mathcal{C}, (\Lambda, \mathcal{C}) \in \mathscr{A} \Rightarrow (\Lambda, \mathcal{C}') \in \mathscr{A}.$$

Thus, the most difficult part about Question 1 is to make Λ as small and $\mathcal{C}$ as large as possible.
3.) It is easy to show that $(\mathbb{N}_0 \backslash \Gamma, Sz^c) \in \mathscr{A}$ for all finite $\Gamma \subseteq \mathbb{N}_0$ (see Corollary 3.5).

If one studies Question 1 for a proper subclass $\mathcal{C}$, instead of considering the maximal class Sz^c, Olevskii and Ulanovskii [OU21, Theorem 1] have obtained a result for the class

$$\mathcal{W} := \Big\{\mu = w \, \mathrm{d}m \mid w > 0 \ \mathrm{d}m\text{-a.e.}, w \text{ bounded}, w \text{ increasing on } (0, 2\pi)\Big\}.$$

Theorem (Olevskii, Ulanovskii)**.** *Let $\Gamma \subseteq \mathbb{N}$ with*

$$\sum_{\gamma \in \Gamma} \frac{1}{\sqrt{\gamma}} < \infty.$$

Then $(\mathbb{N}_0 \backslash \Gamma, \mathcal{W} \cap Sz^c) \in \mathscr{A}$.

In this paper, we will add an answer to Question 1 for sets Λ of the form

$$\Lambda(\mathbf{k}, \boldsymbol{\ell}) := \bigcup_{j \in \mathbb{N}} [\![k_j, k_j + \ell_j]\!], \qquad (\mathbf{k} \in \mathbb{N}^{\mathbb{N}}, \boldsymbol{\ell} \in \mathbb{N}_0^{\mathbb{N}})$$

where $\mathbf{k}$ is strictly increasing and $[\![n,m]\!] := [n,m] \cap \mathbb{Z}$ for $n, m \in \mathbb{Z}, n \leq m$ and where $\mathcal{C}$ is the class of Markoff measures $\mathrm{Mar}(\mathbb{T})$. This class was introduced by Khrushchev in [Khr02]. Informally, μ belongs to $\mathrm{Mar}(\mathbb{T})$ if the sequence of Verblunsky coefficients $(\alpha_n(\mu))_n$ contains a sufficiently dense subsequence that remains bounded away from zero. We will give a formal definition of Markoff measures and Verblunsky coefficients in Section 2. In Section 4 we will show the following.

Theorem 1.2. *Let $s > 1$ and let $\mathbf{k} := (k_j)_j \in \mathbb{N}^{\mathbb{N}}$ be strictly increasing. Set $\lfloor \mathbf{k}^s \rfloor := (\lfloor k_j^s \rfloor)_j \in \mathbb{N}^{\mathbb{N}}$. Then,*

$$\Big(\Lambda\big(\mathbf{k}, \lfloor \mathbf{k}^s \rfloor\big), \, \mathrm{Mar}(\mathbb{T})\Big) \in \mathscr{A}.$$

In particular, there exists a subset $\Lambda \subseteq \mathbb{N}$ with lower density $\underline{\mathrm{d}}\big(\Lambda\big) = 0$ such that $(\Lambda, \mathrm{Mar}(\mathbb{T})) \in \mathscr{A}$.

Theorem 1.2 will follow from the main result of this article, Theorem 3.3, which states that for every $\mu \in Sz^c$ and every strictly increasing $\mathbf{k} \in \mathbb{N}^{\mathbb{N}}$ there exists $\boldsymbol{\ell} \in \mathbb{N}_0^{\mathbb{N}}$ such that $E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))$ is dense in $L^2(\mathbb{T}, \mu)$. In Corollary 3.4 we further show that $\mathbf{k}$ and $\boldsymbol{\ell}$ can be chosen such that $\underline{\mathrm{d}}(E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))) = 0$.

It should be noted that [OU21] and this paper explore very different situations. For every $\mu \in \mathcal{W}$ one has $\alpha_n \to 0$ by Rakhmanov's theorem [Rak83], thus $\mathrm{Mar}(\mathbb{T})$ and $\mathcal{W}$ are disjoint. Moreover, every measure contained in $\mathcal{W}$ is absolutely continuous with full support. In contrast, if $\mu \in \mathrm{Mar}(\mathbb{T})$ then $\mathrm{supp}(\mu_{ac}) \neq \mathbb{T}$ and also, $\mathrm{Mar}(\mathbb{T})$ contains measures with $\mu_s \neq 0$ and pure point measures with full support which are important in physics because of the Anderson localization phenomenon (see Lemma 2.5 and Remark 2.6).

The structure of this paper is as follows. In Section 2 we briefly review some notions from OPUC theory. In particular, we state Szegő's theorem (Theorem 2.1) and give a definition and some examples of Markoff measures. In Section 3 we prove the main theorem, Theorem 3.3. In Section 4 we prove Theorem 1.2 and show other applications of Theorem 3.3.

Notation.

- We use the conventions $\mathbb{N} := \{n \in \mathbb{Z} \mid n \geq 1\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.
- Interval of integers: For $n, m \in \mathbb{Z}$ with $n \leq m$ define
$$[\![n,m]\!] := [n,m] \cap \mathbb{Z}.$$
- Upper and lower density: For $\Lambda \subseteq \mathbb{N}$ define the lower and upper density by
$$\underline{\mathrm{d}}(\Lambda) := \liminf_{N \to \infty} \frac{|\Lambda \cap [\![1,N]\!]|}{N}, \quad \overline{\mathrm{d}}(\Lambda) := \limsup_{N \to \infty} \frac{|\Lambda \cap [\![1,N]\!]|}{N}$$
respectively.
- $\mathcal{P}$ is the set of all non-degenerate probability measures on $\mathbb{T}$.
- Closure of a subset: For $\mu \in \mathcal{P}$ and a subset $M \subseteq L^2(\mathbb{T}, \mu)$ we write
$$\overline{M}^{L^2(\mathbb{T},\mu)}$$
for the closure of M in $L^2(\mathbb{T}, \mu)$.
- Sz is the set of all measures $\mathrm{d}\mu = w \, \mathrm{d}m + \mathrm{d}\mu_s \in \mathcal{P}$ for which
$$\int_0^{2\pi} \log(w(\theta)) \, \mathrm{d}\theta > -\infty.$$

- $\operatorname{Mar}(\mathbb{T})$ denotes the set of all Markoff measures (see definition on page 5).
- We already defined earlier
$$\begin{aligned} E(\Lambda) &= \operatorname{span}\{z^\lambda \mid z \in \mathbb{T}, \lambda \in \Lambda\}, \\ \mathscr{A} &= \{(\Lambda, \mathcal{C}) \mid \Lambda \subseteq \mathbb{N}_0, \mathcal{C} \subseteq \mathcal{P}, E(\Lambda) \text{ dense in } L^2(\mathbb{T},\mu) \text{ for all } \mu \in \mathcal{C}\}, \\ \Lambda(\mathbf{k}, \boldsymbol{\ell}) &= \bigcup_{j\in\mathbb{N}} [\![k_j, k_j + \ell_j]\!], \qquad (\mathbf{k} \in \mathbb{N}^{\mathbb{N}}, \boldsymbol{\ell} \in \mathbb{N}_0^{\mathbb{N}}) \end{aligned}$$
where $\mathbf{k}$ is strictly increasing.
- Let $\mathbb{P}$ be the space of all polynomials. For $n \in \mathbb{N}_0$, let $\mathbb{P}_{\leq n}$ be the space of all polynomials with degree less than or equal to n.
- We will also need the following definitions later
$$\begin{aligned} \beta_\mu(k,n) &:= \min_{\pi\in\mathbb{P}_{\leq n}} \|z^{-k} - \pi(z)\|_{L^2(\mathbb{T},\mu)}, \qquad (k,n \in \mathbb{N}_0, \mu \in \mathcal{P}), \\ \beta_\mu(k,\infty) &:= \lim_{n\to\infty} \beta_\mu(k,n) \\ &= \inf_{\pi\in\mathbb{P}} \|z^{-k} - \pi(z)\|_{L^2(\mathbb{T},\mu)}, \qquad (k \in \mathbb{N}_0, \mu \in \mathcal{P}). \end{aligned}$$

Acknowledgment. I am very grateful to my advisor, Professor Sergey Denisov, for suggesting the problem formulated as Question 1, and for his countless helpful discussions and comments.

2. Facts from OPUC

Theorem 2.1 (Szegő)**.** *Let $\mu \in \mathcal{P}$ with $\mathrm{d}\mu = w\,\mathrm{d}m + \mathrm{d}\mu_s$. Then*
$$\exp\Big(\int_0^{2\pi} \log w \,\mathrm{d}m\Big) = \beta_\mu(1,\infty).$$

This appears first in [GS58, Chapter 3.1], see also [Sim05a, Chapter 2]. We have the following corollary [Sim05a, Theorem 1.5.7].

Corollary 2.2. *Let $\mu \in \mathcal{P}$ with $\mathrm{d}\mu = w\,\mathrm{d}m + \mathrm{d}\mu_s$. Then $\mu \in Sz^c$ if and only if the polynomials are dense in $L^2(\mathbb{T},\mu)$, i.e. $\beta_\mu(k,\infty) = 0$ for all $k \in \mathbb{N}$.*

Let $\mu \in \mathcal{P}$ with $\mathrm{d}\mu = w\,\mathrm{d}m + \mathrm{d}\mu_s$. Because of the non-degeneracy of μ, the monomials z^n are linearly independent in $L^2(\mathbb{T},\mu)$. Thus, for every $n \in \mathbb{N}_0$ there is a uniquely determined polynomial
$$\Phi_n(\mu;z) = \sum_{i=0}^n b_{n,i}(\mu) z^i \quad (z \in \mathbb{T})$$
such that $b_{n,i} \in \mathbb{C}$ for every $i \in [\![0,n]\!]$, $b_{n,n} = 1$ and, for every $k \in [\![0, n-1]\!]$,
$$\big\langle \Phi_n(\mu;z), \Phi_k(\mu;z)\big\rangle_{L^2(\mathbb{T},\mu)} = 0.$$
The polynomials $\Phi_n(\mu;z)$ are called *monic orthogonal polynomials of* μ. The monic orthogonal polynomials satisfy a recursion relation, the *Szegő recurrence*,
$$\Phi_{n+1}(\mu;z) = z\Phi_n(\mu;z) - \overline{\alpha_n(\mu)}\Phi_n^*(\mu;z)$$
for every $n \in \mathbb{N}_0$, where, for $z \in \mathbb{T}$,
$$\Phi_n^*(\mu;z) = z^n \overline{\Phi_n(\mu;z)} = \sum_{i=0}^n \overline{b_{n,n-i}(\mu)} z^i.$$

For $n \in \mathbb{N}_0$ the recursion coefficient $\alpha_n(\mu)$ is called the n-th *Verblunsky coefficient* of μ. One has $\alpha_n(\mu) \in \mathbb{D}$ for every $n \in \mathbb{N}_0$. In fact, the mapping

$$\mathcal{P} \to \mathbb{D}^{\mathbb{N}_0}, \quad \mu \mapsto (\alpha_n(\mu))_n$$

is a bijection. This is known as *Verblunksy's theorem* [Sim05a, Theorem 1.7.11]. One can also express the L^2-norms of the monic orthogonal polynomials and the integral of $\log(w)$ by the Verblunsky coefficients [Sim05a, Theorems 1.5.2, 2.3.1]

$$\|\Phi_{n+1}(\mu;z)\|^2_{L^2(\mathbb{T},\mu)} = \prod_{i=0}^{n}(1-|\alpha_i(\mu)|^2), \tag{2}$$

$$\exp\Big(\int_0^{2\pi} \log w \,\mathrm{d}m\Big) = \lim_{n\to\infty} \|\Phi_n(\mu;z)\|^2_{L^2(\mathbb{T},\mu)} = \prod_{i=0}^{\infty}(1-|\alpha_i(\mu)|^2).$$

The second equation proves the following corollary.

Corollary 2.3. *Let $\mu \in \mathcal{P}$. Then $\mu \in Sz$ if and only if $(\alpha_n(\mu))_n \in \ell^2$.*

In the following, we will write $\Phi_n, b_{n,i}$ and α_n instead of $\Phi_n(\mu;z), b_{n,i}(\mu)$ and $\alpha_n(\mu)$ whenever the measure μ is clear from the context.

Markoff measures. Let $\mu \in \mathcal{P}$. For $\ell \in \mathbb{N}_0$ and $\varepsilon \in (0,1)$ define the class $\mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T}) \subseteq \mathcal{P}$ via

$$\mu \in \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T}) \Leftrightarrow \inf_{n\in\mathbb{N}_0} \max_{n\le j\le n+\ell} |\alpha_j| \ge \varepsilon. \tag{3}$$

The class $\mathrm{Mar}(\mathbb{T})$ of *Markoff measures*, which was introduced by Khrushchev in [Khr02], is defined as follows

$$\mathrm{Mar}(\mathbb{T}) := \bigcup_{\ell\in\mathbb{N}_0} \bigcup_{\varepsilon>0} \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T}).$$

Remark 2.4. This definition is in fact a characterization of the original definition [Khr02, Theorem 1.8]. Khrushchev defines $\mathrm{Mar}(\mathbb{T})$ as every $\mu \in \mathcal{P}$ such that $\mathrm{d}m$ is not contained in the derived set of

$$\{|\varphi_n|^2 \mathrm{d}\mu \mid n \in \mathbb{N}_0\}$$

where φ_n is the n-th orthogonal polynomial in $L^2(\mathbb{T},\mu)$ obtained by using the Gram-Schmidt algorithm on the monomials z^n with $n \in \mathbb{N}_0$.

Let us collect a few properties and examples of Markoff measures.

Lemma 2.5 (Properties of Markoff measures)**.** *Let $\mu \in \mathcal{P}$.*

a) $\mathrm{Mar}(\mathbb{T}) \subseteq Sz^c$.

b) $\mu \in \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T}) \Rightarrow \|\Phi_n\|^2_{L^2(\mathbb{T},\mu)} \le (1-\varepsilon^2)^{\frac{n-1}{\ell+1}-1}$.

c) $\mu \in \mathrm{Mar}(\mathbb{T}) \Rightarrow \mathrm{supp}(\mu_{ac}) \ne \mathbb{T}$.

d) $\mathrm{supp}(\mu) \ne \mathbb{T} \Rightarrow \mu \in \mathrm{Mar}(\mathbb{T})$.

The converse implications in c) and d) are not true.

Proof. a) Let $\mu \in \mathrm{Mar}(\mathbb{T})$. Then, $(\alpha_n)_n$ does not converge to zero by definition of $\mathrm{Mar}(\mathbb{T})$. In particular, $(\alpha_n)_n \notin \ell^2$. Thus, by Corollary 2.3, $\mu \in Sz^c$.

b) Let $\mu \in \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T})$. From (2) it follows that

$$\|\Phi_n\|^2_{L^2(\mathbb{T},\mu)} = \prod_{i=0}^{n-1}(1-|\alpha_i|^2) \leq (1-\varepsilon^2)^{\lfloor \frac{n-1}{\ell+1} \rfloor} \leq (1-\varepsilon^2)^{\frac{n-1}{\ell+1}-1}.$$

c) By Rakhmanov's theorem [Rak83, §3], [Sim05b, Corollary 9.1.11], for any $\mu \in \mathcal{P}$ with $\mathrm{supp}(\mu_{ac}) = \mathbb{T}$ one has

$$\lim_{n\to\infty} \alpha_n(\mu) = 0.$$

Let $\mu \in \mathrm{Mar}(\mathbb{T})$. Then $(\alpha_n(\mu))_n$ does not converge to 0. Thus, by Rakhmanov's theorem, $\mathrm{supp}(\mu_{ac}) \neq \mathbb{T}$. To see that the converse implication of c) doesn't hold true, consider a measure $\mu \in \mathcal{P}$ such that for all $n \in \mathbb{N}_0$

$$\alpha_n(\mu) = \begin{cases} \frac{1}{\sqrt{k}}, & n = k!, k \in \mathbb{N}_0 \\ 0, & \text{else.} \end{cases}$$

By Corollary 2.3 $\mu \in Sz^c$. However, $\mu \notin \mathrm{Mar}(\mathbb{T})$ since $a_n \to 0$. Furthermore, μ is purely singular continuous (see [Sim05b, Theorem 12.5.2]). In particular, $\mu_{ac} = 0$.

d) For a proof see [Khr02, Corollary 1.9]. We briefly present a counterexample to the converse implication due to Zhedanov [Zhe20]. For $q \in \mathbb{T}$ which is not a root of unity and $0 < p < 1$ set

$$\mu := (1-p)\sum_{n=0}^{\infty} p^n \delta_{q^n}$$

where δ_w is the Dirac measure with $\mathrm{supp}(\delta_w) = \{w\}$. Clearly, $\mu \in \mathcal{P}$. There is also a simple formula for $|\alpha_n|^2$ in terms of p and q (see [Zhe20, p.5]). For all $n \in \mathbb{N}_0$

$$|\alpha_n|^2 = \frac{(1-p)^2}{1+p^2-2p\,\mathrm{Re}(q^{n+1})}.$$

Thus,

$$\frac{1-p}{1+p} < |\alpha_n| < 1$$

for all $n \in \mathbb{N}_0$ which implies that $\mu \in \mathrm{Mar}(\mathbb{T})$. Furthermore, since q is not a root of unity,

$$\mathrm{supp}(\mu) = \overline{\{q^n \mid n \in \mathbb{N}_0\}} = \mathbb{T}. \qquad \square$$

Remark 2.6. In some sense, measures which are pure point with full support are generic examples. Namely, for any rotation invariant probability measure β_0 on $\mathbb{D}$ with $\beta_0 \neq \delta_0$ and

$$\int_{\mathbb{D}} \log(1-|\omega|)\,\mathrm{d}\beta_0(\omega) > -\infty$$

one has that almost every sequence $(\omega_n)_n \in \mathbb{D}^{\mathbb{N}_0}$ with respect to the product measure

$$\beta := \bigotimes_{i=0}^{\infty} \beta_0$$

generates a measure μ, by choosing μ with $\alpha_n(\mu) = \omega_n$ for every $n \in \mathbb{N}_0$, that is pure point and has full support [Sim05b, Theorem 12.6.2]. One can also view the Verblunsky coefficients α_n as being values of an i.i.d. process $(\omega_n)_n$ with common distribution β_0. Such measures are important in physics because, going back to the work of Anderson [And58], certain random Hamiltonians on lattices produce dense point spectra - a phenomenon called *Anderson localization*. Mathematical references include [CL90] and [PF92].

3. Main Theorem

Recall that for $k, n \in \mathbb{N}_0$

$$\beta_\mu(k,n) = \min_{\pi \in \mathbb{P}_{\le n}} \|z^{-k} - \pi(z)\|_{L^2(\mathbb{T},\mu)}.$$

Definition 3.1. Let $\mu \in \mathcal{P}$. A strictly increasing function $f : \mathbb{N} \to \mathbb{N}$ is called *β-approximating* for μ if

$$\lim_{k\to\infty} \beta_\mu(k, f(k)) = 0.$$

Corollary 3.2. *Let $\mu \in \mathcal{P}$. Then there exists a β-approximating f if and only if $\mu \in Sz^c$.*

Proof. Let $\mu \in Sz^c$. Then by Corollary 2.2

$$\lim_{n\to\infty} \beta_\mu(k,n) = \beta_\mu(k,\infty) = 0$$

for every $k \in \mathbb{N}$. Thus there exists an $f : \mathbb{N} \to \mathbb{N}$ that is β-approximating for μ.

For the converse implication, let $f : \mathbb{N} \to \mathbb{N}$ be β-approximating for μ. Since $(\beta_\mu(k,n))_n$ is a decreasing sequence for every $k \in \mathbb{N}$ and f is increasing

$$\beta_\mu(k,\infty) = \inf_{k\in\mathbb{N}} \beta_\mu(k, f(k)) = \lim_{k\to\infty} \beta_\mu(k, f(k)) = 0.$$

Thus, by Corollary 2.2, $\mu \in Sz^c$. □

Recall also, that for two sequences $\boldsymbol{\ell} := (\ell_j)_j \in \mathbb{N}_0^{\mathbb{N}}, \mathbf{k} := (k_j)_j \in \mathbb{N}^{\mathbb{N}}$, where $\mathbf{k}$ is strictly increasing,

$$\Lambda(\mathbf{k}, \boldsymbol{\ell}) = \bigcup_{j\in\mathbb{N}} [\![k_j, k_j + \ell_j]\!].$$

Note that the definition of $\Lambda(\mathbf{k}, \boldsymbol{\ell})$ does not require the intervals $[\![k_j, k_j + \ell_j]\!]$ to be disjoint.

Now we are ready to state the main theorem.

Theorem 3.3. *Let $\mu \in Sz^c$ and f be any β-approximating function for μ. Let $\mathbf{k} := (k_j)_j \in \mathbb{N}^{\mathbb{N}}$ be strictly increasing and $\boldsymbol{\ell} := (\ell_j)_j \in \mathbb{N}_0^{\mathbb{N}}$ such that*

$$\lim_{j\to\infty} \ell_j - f(k_j) = \infty.$$

Then $E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))$ is dense in $L^2(\mathbb{T},\mu)$.

Let us first prove the following two corollaries.

Corollary 3.4. *Let $\mu \in Sz^c$. Then there exists a set $\Lambda \subseteq \mathbb{N}$ with $\underline{\mathrm{d}}(\Lambda) = 0$ such that $E(\Lambda)$ is dense in $L^2(\mathbb{T},\mu)$.*

Proof. By Corollary 3.2 there exists $f : \mathbb{N} \to \mathbb{N}$ which is β-approximating for μ. Set $k_1 := 1$ and, for $j \in \mathbb{N}$, define k_{j+1} recursively such that

$$(k_j + f(k_j) + j)j + 1 \le k_{j+1}.$$

Put $\ell_j := f(k_j) + j$ for every $j \in \mathbb{N}$, $\boldsymbol{\ell} := (\ell_j)_j$ and $\mathbf{k} := (k_j)_j$. By Theorem 3.3, $E(\Lambda(\mathbf{k},\boldsymbol{\ell}))$ is dense in $L^2(\mathbb{T},\mu)$. Furthermore, we get the following estimate for $\Lambda(\mathbf{k},\boldsymbol{\ell})$.

$$\underline{\mathrm{d}}(\Lambda(\mathbf{k},\boldsymbol{\ell})) \leq \liminf_{j\to\infty} \frac{\left|\Lambda(\mathbf{k},\boldsymbol{\ell}) \cap [\![1, k_{j+1}-1]\!]\,\right|}{k_{j+1}-1}$$

$$\leq \liminf_{j\to\infty} \frac{k_j+\ell_j}{k_{j+1}-1} \leq \liminf_{j\to\infty} \frac{1}{j} = 0. \qquad \square$$

Corollary 3.5. $(\mathbb{N}_0\backslash\Gamma, Sz^c) \in \mathscr{A}$ *for every finite* $\Gamma \subseteq \mathbb{N}_0$.

Proof. For $\mu \in Sz^c$ choose f which is β-approximating for μ. Set

$$k_1 := \max_{\gamma\in\Gamma} \gamma + 1, \quad k_{j+1} := k_j + f(k_j) + j$$

for all $j \in \mathbb{N}$. Set furthermore $\ell_j := f(k_j) + j$ for all $j \in \mathbb{N}$. Then

$$\Lambda(\mathbf{k},\boldsymbol{\ell}) = \mathbb{N}_0\backslash[\![0, \max_{\gamma\in\Gamma}\gamma]\!] \subseteq \mathbb{N}_0\backslash\Gamma.$$

By Theorem 3.3, $E(\Lambda(\mathbf{k},\boldsymbol{\ell}))$ is dense in $L^2(\mathbb{T},\mu)$. Thus, also $E(\mathbb{N}_0\backslash\Gamma)$ is dense in $L^2(\mathbb{T},\mu)$. $\square$

Remark 3.6. Corollary 3.5 is an extension of Szegő's classical theorem. Szegő's theorem is the case $\Gamma = \emptyset$.

Before we can prove the main theorem, we first prove two lemmas about approximating negative powers of z with polynomials in $L^2(\mathbb{T},\mu)$. Throughout, we will use the convention $\sum_{j\in J} a_j := 0$ if $J = \emptyset$.

Lemma 3.7. *Let* $\mu \in \mathcal{P}$. *For all* $k, n \in \mathbb{N}$ *we have the estimate*

$$\beta_\mu(k,n) \leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + \sum_{i=1}^{k-1} \sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!] \\ j_1+\dots+j_i\leq k-1}} |b_{n+1,n+1-j_1}| \cdot \ldots \cdot |b_{n+1,n+1-j_i}|\right). \tag{4}$$

Lemma 3.8. *Let* $\mu \in \mathcal{P}$. *Then, for all* $k, n \in \mathbb{N}$

$$\beta_\mu(k,n) \leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} (2n+2)^{k-1}.$$

Proof of Lemma 3.7. The proof is by induction. For $k = 1$ and $n \in \mathbb{N}$ we show

$$\beta_\mu(1,n) = \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}.$$

This is a standard result in OPUC theory but the proof is included here for completeness.

For a subspace $V \subseteq L^2(\mathbb{T},\mu)$ denote by P_V the orthogonal projection onto V.

$$\begin{aligned}\beta_\mu(1,n) &= \min\left\{\|z^{-1} - \pi(z)\|_{L^2(\mathbb{T},\mu)} \mid \pi \in \mathbb{P}_{\leq n}\right\} \\ &= \min\left\{\|1 - \pi(z)\|_{L^2(\mathbb{T},\mu)} \mid \pi \in z\cdot\mathbb{P}_{\leq n}\right\} \\ &= \|P_{(z\cdot\mathbb{P}_{\leq n})^\perp}(1)\|_{L^2(\mathbb{T},\mu)}.\end{aligned}$$

Note that $\Phi^*_{n+1}(0) = b_{n+1,n+1} = 1$ and thus

$$\Phi^*_{n+1}(z) - 1 = \sum_{i=1}^{n+1} \overline{b_{n+1,n+1-i}} z^i \in z\cdot\mathbb{P}_{\leq n}.$$

Furthermore, for all $\pi \in \mathbb{P}_{\leq n}$,

$$\langle z\pi, \Phi^*_{n+1}\rangle_{L^2(\mathbb{T},\mu)} = \langle z\pi, z^{n+1}\overline{\Phi_{n+1}}\rangle_{L^2(\mathbb{T},\mu)} = \langle \Phi_{n+1}, z^n\overline{\pi}\rangle_{L^2(\mathbb{T},\mu)} = 0$$

since $\Phi_{n+1} \perp \mathbb{P}_{\leq n}$. Thus $P_{(z\cdot\mathbb{P}_{\leq n})^\perp}(1) = \Phi^*_{n+1}$ and therefore

$$\beta_\mu(1,n) = \|\Phi^*_{n+1}\|_{L^2(\mathbb{T},\mu)} = \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}.$$

Now, we make an induction in k. Assume that for all $n \in \mathbb{N}$ one can estimate $\beta_\mu(1,n),\ldots,\beta_\mu(k,n)$ from above by the right-hand side in (4). By applying the triangle inequality one gets for all $n \in \mathbb{N}$

$$\begin{aligned}
\beta_\mu(k+1,n) &\leq \|z^{-(k+1)} + \sum_{m=1}^{n+1} \overline{b_{n+1,n+1-m}}\, P_{(\mathbb{P}_{\leq n})}(z^{m-(k+1)})\|_{L^2(\mathbb{T},\mu)} \\
&\leq \|z^{-(k+1)} + \sum_{m=1}^{n+1} \overline{b_{n+1,n+1-m}}\, z^{m-(k+1)}\|_{L^2(\mathbb{T},\mu)} \\
&\qquad + \sum_{m=1}^{\min(k,n+1)} |b_{n+1,n+1-m}| \cdot \|z^{m-(k+1)} - P_{(\mathbb{P}_{\leq n})}(z^{m-(k+1)})\|_{L^2(\mathbb{T},\mu)} \\
&= \|z^{-(k+1)} + \sum_{m=1}^{n+1} \overline{b_{n+1,n+1-m}}\, z^{m-(k+1)}\|_{L^2(\mathbb{T},\mu)} \\
&\qquad + \sum_{m=1}^{\min(k,n+1)} |b_{n+1,n+1-m}|\, \beta_\mu(k+1-m,n). \qquad (5)
\end{aligned}$$

The first summand on the right-hand side of (5) equals

$$\|z^{-(k+1)} + z^{-(k+1)}(\Phi^*_{n+1}(z) - 1)\|_{L^2(\mathbb{T},\mu)} = \|\Phi^*_{n+1}(z)\|_{L^2(\mathbb{T},\mu)} = \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}. \quad (6)$$

By the induction hypothesis, each $\beta_\mu(k+1-m,n)$ in the second summand on the right-hand side of (5) can be bounded from above by the following expression

$$\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\Big(1 + \sum_{i=1}^{k-1} \sum_{\substack{j_1,\ldots,j_i \in [\![1,n+1]\!],\\ j_1+\ldots+j_i \leq k-m}} |b_{n+1,n+1-j_1}| \cdot \ldots \cdot |b_{n+1,n+1-j_i}|\Big). \quad (7)$$

Note that when applying the induction hypothesis to $\beta_\mu(k+1-m,n)$ we would a priori get a sum in i that goes from 1 to $k-m$. The reason why the sum in (7) goes from 1 to $k-1$ instead is that for $i > k-m$ the index set

$$\{j_1,\ldots,j_i \in [\![1,n+1]\!] \mid j_1+\ldots+j_i \leq k-m\}$$

is empty and thus the sum over this index set equals zero. By (7) and (5), the second summand on the right-hand side of (5) is bounded from above by

$$\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\Bigg(\sum_{m=1}^{\min(k,n+1)}|b_{n+1,n+1-m}| \\ +\sum_{m=1}^{\min(k,n+1)}\sum_{i=1}^{k-1}\sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!],\\ j_1+\dots+j_i\le k-m}}|b_{n+1,n+1-m}|\cdot|b_{n+1,n+1-j_1}|\cdot\ldots\cdot|b_{n+1,n+1-j_i}|\Bigg). \tag{8}$$

By rearranging the summands in (8) we get

$$\begin{aligned}&\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\Bigg(\sum_{m=1}^{\min(k,n+1)}|b_{n+1,n+1-m}| \\ &\qquad+\sum_{i=1}^{k-1}\sum_{m=1}^{\min(k,n+1)}\sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!],\\ j_1+\dots+j_i\le k-m}}|b_{n+1,n+1-m}|\cdot|b_{n+1,n+1-j_1}|\cdot\ldots\cdot|b_{n+1,n+1-j_i}|\Bigg)\\ &=\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\Bigg(\sum_{m=1}^{\min(k,n+1)}|b_{n+1,n+1-m}| \\ &\qquad+\sum_{i=2}^{k}\sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!],\\ j_1+\dots+j_i\le k}}|b_{n+1,n+1-j_1}|\cdot\ldots\cdot|b_{n+1,n+1-j_i}|\Bigg).\end{aligned} \tag{9}$$

In (9), the sum in m is equal to the sum in $j_1,\dots,j_i$ in the case $i=1$. Thus, the right-hand side in (9) equals

$$\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\sum_{i=1}^{k}\sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!],\\ j_1+\dots+j_i\le k}}|b_{n+1,n+1-j_1}|\cdot\ldots\cdot|b_{n+1,n+1-j_i}|.$$

Now we have an upper bound for the second summand in the right-hand side of (5). In (6) we asserted that the first summand in the right-hand side of (5) equals $\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}$. Putting both together we obtain the desired upper bound

$$\begin{aligned}&\beta_\mu(k+1,n)\\ &\le\|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)}\Bigg(1+\sum_{i=1}^{k}\sum_{\substack{j_1,\dots,j_i\in[\![1,n+1]\!],\\ j_1+\dots+j_i\le k}}|b_{n+1,n+1-j_1}|\cdot\ldots\cdot|b_{n+1,n+1-j_i}|\Bigg).\ \square\end{aligned}$$

Proof of Lemma 3.8. Let $w_1,\dots,w_n$ be the zeros of Φ_n. Since the zeros of monic orthogonal polynomials on the unit circle are contained in the unit disk [Sim05a, Theorem 1.7.1] we have

$$|b_{n,k}|\le\sum_{J\subseteq[\![1,n]\!],\ |J|=n-k}\ \prod_{j\in J}|w_j|\le\binom{n}{k}. \tag{10}$$

Together with Lemma 3.7 we get a new estimate for $\beta_\mu(k,n)$ by using (10) to bound each coefficient $b_{n+1,n+1-j_i}$ in the right-hand side of (4)

$$\beta_\mu(k,n) \leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + \sum_{i=1}^{k-1} \sum_{\substack{j_1,\ldots,j_i \in [\![1,n+1]\!] \\ j_1+\ldots+j_i \leq k-1}} \binom{n+1}{j_1} \cdot \ldots \cdot \binom{n+1}{j_i}\right).$$

For each of the binomial coefficients, we now use the bound $\binom{n}{k} \leq n^k$. In general, this bound is very wasteful, but in this case, one always gets the summand $(n+1)^{k-1}$ from the tuple $(j_1,\ldots,j_{k-1})$ with entries $j_i = 1$ for each $i \in [\![1,k-1]\!]$. Hence,

$$\begin{aligned}
\beta_\mu(k,n) &\leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + \sum_{i=1}^{k-1} \sum_{\substack{j_1,\ldots,j_i \in [\![1,n+1]\!] \\ j_1+\ldots+j_i \leq k-1}} (n+1)^{j_1} \cdot \ldots \cdot (n+1)^{j_i}\right) \\
&\leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + \sum_{i=1}^{k-1} \sum_{\substack{j_1,\ldots,j_i \in [\![1,n+1]\!] \\ j_1+\ldots+j_i \leq k-1}} (n+1)^{k-1}\right) \\
&= \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + (n+1)^{k-1} \sum_{i=1}^{k-1} \sum_{\ell=i}^{k-1} \sum_{\substack{j_1,\ldots,j_i \in [\![1,\ell]\!] \\ j_1+\ldots+j_i = \ell}} 1\right). \qquad (11)
\end{aligned}$$

The last sum in (11) now counts the number of tuples $(j_1,\ldots,j_i)$ with entries in $\mathbb{N}$ such that $\sum_{s=1}^{i} j_s = \ell$. This number is $\binom{\ell-1}{i-1}$ since one has to count how many ways there are to draw $i-1$ separating lines between ℓ 1's. Thus, from (11) we get the following estimate for $\beta_\mu(k,n)$

$$\beta_\mu(k,n) \leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + (n+1)^{k-1} \sum_{i=1}^{k-1} \sum_{\ell=i}^{k-1} \binom{\ell-1}{i-1}\right). \qquad (12)$$

To simplify the right-hand side, we use the following combinatorial identity

$$\sum_{\ell=i}^{k-1} \binom{\ell-1}{i-1} = \binom{k-1}{i} \qquad (13)$$

for $1 \leq i \leq k-1$. The identity follows by induction in k from Pascal's rule for binomial coefficients. Indeed, if we assume (13) holds true for some $k \geq 2$, then

$$\sum_{\ell=i}^{k} \binom{\ell-1}{i-1} = \sum_{\ell=i}^{k-1} \binom{\ell-1}{i-1} + \binom{k-1}{i-1} = \binom{k-1}{i} + \binom{k-1}{i-1} = \binom{k}{i}$$

where the last equation is Pascal's rule. Combining (13) and (12) yields

$$\begin{aligned}
\beta_\mu(k,n) &\leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + (n+1)^{k-1} \sum_{i=1}^{k-1} \binom{k-1}{i}\right) \\
&= \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} \left(1 + (n+1)^{k-1}(2^{k-1}-1)\right) \\
&\leq \|\Phi_{n+1}\|_{L^2(\mathbb{T},\mu)} (2n+2)^{k-1}
\end{aligned}$$

which is the estimate we wanted to show. $\square$

Proof of Theorem 3.3. Without loss of generality
$$k_j + \ell_j < k_{j+1} \tag{14}$$
for every $j \in \mathbb{N}$. Otherwise one passes to subsequences $\mathbf{k}' := (k_{n_j})_j$ and $\boldsymbol{\ell}' := (\ell_{n_j})_j$ that fulfill condition (14). Since $\Lambda(\mathbf{k}', \boldsymbol{\ell}') \subseteq \Lambda(\mathbf{k}, \boldsymbol{\ell})$, the density of $E(\Lambda(\mathbf{k}', \boldsymbol{\ell}'))$ implies the density of $E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))$ in $L^2(\mathbb{T}, \mu)$. Thus, we can assume (14).

For $k \in \mathbb{N}$ let $\pi_k \in \mathbb{P}_{\leq f(k)}$ such that
$$\|z^{-k} - \pi_k\|_{L^2(\mathbb{T},\mu)} = \min_{\pi \in \mathbb{P}_{\leq f(k)}} \|z^{-k} - \pi(z)\|_{L^2(\mathbb{T},\mu)} = \beta_\mu(k, f(k)).$$
By Theorem 2.1 it suffices to show that
$$\overline{E(\mathbb{N}_0)}^{L^2(\mathbb{T},\mu)} = \overline{E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))}^{L^2(\mathbb{T},\mu)}.$$
To this end, let $k \in \mathbb{N}_0$. We need to show that $z^k \in \overline{E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))}^{L^2(\mathbb{T},\mu)}$.

Let $J \in \mathbb{N}$ such that
$$k \leq \ell_j - f(k_j)$$
for every $j \geq J$. Then, for every $j \geq J$,
$$k_j \leq k + k_j + \deg(\pi_{k_j}) \leq k + k_j + f(k_j) \leq k_j + \ell_j.$$
It follows that
$$z^{k+k_j}\pi_{k_j} \in \operatorname{span}\{z^m \mid k_j \leq m \leq k_j + \ell_j\} \subseteq E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))$$
for every $j \geq J$. Furthermore, for every $j \in \mathbb{N}$,
$$\|z^k - z^{k+k_j}\pi_{k_j}\|_{L^2(\mathbb{T},\mu)} = \|z^{-k_j} - \pi_{k_j}\|_{L^2(\mathbb{T},\mu)} = \beta_\mu(k_j, f(k_j)).$$
Since f is β-approximating for μ,
$$\lim_{j\to\infty} \|z^k - z^{k+k_j}\pi_{k_j}\|_{L^2(\mathbb{T},\mu)} = 0.$$
Thus, $z^k \in \overline{E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))}^{L^2(\mathbb{T},\mu)}$ which implies that $E(\Lambda(\mathbf{k}, \boldsymbol{\ell}))$ is dense in $L^2(\mathbb{T}, \mu)$. □

4. Applications

We now want to prove Theorem 1.2.

Proof of Theorem 1.2. Let $\mu \in \mathrm{Mar}(\mathbb{T})$ and let $1 < t < s$. Put
$$f : \mathbb{N} \to \mathbb{N}, \quad f(n) := \lfloor n^t \rfloor.$$
We first show that f is β-approximating for μ. By Lemma 2.5 there exist $\varepsilon > 0$ and $\ell \in \mathbb{N}$ such that
$$\|\Phi_n\|^2_{L^2(\mathbb{T},\mu)} \leq (1 - \varepsilon^2)^{\frac{n-1}{\ell} - 1}$$
for all $n \in \mathbb{N}$. Thus, for every $k \in \mathbb{N}$,
$$\begin{aligned}
\|\Phi_{\lfloor k^t \rfloor + 1}\|_{L^2(\mathbb{T},\mu)}(2\lfloor k^t \rfloor + 2)^{k-1} &\leq (1 - \varepsilon^2)^{\frac{\lfloor k^t \rfloor}{2\ell} - \frac{1}{2}}(2\lfloor k^t \rfloor + 2)^{k-1} \\
&\leq (1 - \varepsilon^2)^{\frac{k^t}{2\ell} - \frac{1}{2} - \frac{1}{2\ell}}(2k^t + 2)^{k-1} \\
&\leq \frac{(1 - \varepsilon^2)^{-\frac{1}{2} - \frac{1}{2\ell}}}{2k^t + 2}\Big((1 - \varepsilon^2)^{\frac{k^{t-1}}{2\ell}}(2k^t + 2)\Big)^k.
\end{aligned}$$
The right-hand side converges to 0 when $k \to \infty$. Thus, by Lemma 3.8, f is β-approximating for μ. Furthermore,
$$\lim_{j\to\infty} \lfloor k_j^s \rfloor - f(k_j) = \lim_{j\to\infty} \lfloor k_j^s \rfloor - \lfloor k_j^t \rfloor = \infty$$

since $t < s$. Hence, by Theorem 3.3, $E(\Lambda(\lfloor \mathbf{k}, \lfloor \mathbf{k}^s \rfloor))$ is dense in $L^2(\mathbb{T}, \mu)$. Finally, if one chooses $\mathbf{k}$ such that for every $j \in \mathbb{N}$

$$(k_j + \lfloor k_j^s \rfloor) j + 1 \leq k_{j+1}$$

then

$$\underline{\mathrm{d}}(\Lambda(\mathbf{k}, \lfloor \mathbf{k}^s \rfloor)) \leq \liminf_{j\to\infty} \frac{\left|\Lambda(\mathbf{k}, \lfloor \mathbf{k}^s \rfloor) \cap [\![1, k_{j+1}-1]\!]\right|}{k_{j+1}-1}$$

$$\leq \liminf_{j\to\infty} \frac{k_j + \lfloor k_j^s \rfloor}{k_{j+1}-1} \leq \liminf_{j\to\infty} \frac{1}{j} = 0. \qquad \square$$

Remark 4.1. Comparing the results from [OU21] with ours, one can have $(\Lambda, \mathrm{Mar}(\mathbb{T})) \in \mathscr{A}$ for sets Λ that are much sparser than any known set Λ for which one has $(\Lambda, \mathcal{W} \cap Sz^c) \in \mathscr{A}$. To be more precise, in [OU21] it was proven that $(\mathbb{N}_0 \setminus \Gamma, \mathcal{W} \cap Sz^c) \in \mathscr{A}$ if

$$\sum_{\gamma\in\Gamma} \frac{1}{\sqrt{\gamma}} < \infty.$$

This implies $\underline{\mathrm{d}}(\mathbb{N}_0 \backslash \Gamma) = 1$ since for every $M \in \mathbb{N}$ one has

$$\infty > \sum_{\gamma\in\Gamma} \frac{1}{\sqrt{\gamma}} \geq \limsup_{N\to\infty} |\Gamma \cap [\![1, N]\!]| \cdot \frac{1}{\sqrt{N}} \geq \overline{\mathrm{d}}(\Gamma)\sqrt{M}.$$

Thus, by letting $M \to \infty$, it follows that $\overline{\mathrm{d}}(\Gamma) = 1 - \underline{\mathrm{d}}(\mathbb{N}_0 \backslash \Gamma) = 0$.

For the class $\mathrm{Mar}(\mathbb{T})$ by contrast, there exists $\Lambda \subseteq \mathbb{N}$ with $\underline{\mathrm{d}}(\Lambda) = 0$ such that $(\Lambda, \mathrm{Mar}(\mathbb{T})) \in \mathscr{A}$ by Theorem 1.2.

We now look at the class $\mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T})$ introduced in (3). In this class one can find an even 'thinner' set $\Lambda(\mathbf{k}, \boldsymbol{\ell})$ than the one in Theorem 1.2 such that

$$\big(\Lambda(\mathbf{k}, \boldsymbol{\ell}), \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T})\big) \in \mathscr{A}.$$

To be more precise, in the situation when $\mu \in \mathrm{Mar}(\mathbb{T})$ (Theorem 1.2), there are β-approximating functions f with

$$f(k) = O(k^s), \quad (k \to \infty)$$

for any $s > 1$. However, when $\mu \in \mathrm{Mar}_{\varepsilon,\ell}(\mathbb{T})$ (Corollary 4.2), there exist f which are β-approximating for μ with

$$f(k) = O(\log(k)k), \quad (k \to \infty)$$

where the implicit constant depends on ε and ℓ.

Corollary 4.2. *Let $\varepsilon > 0, \ell \in \mathbb{N}$, $t > 2$ and let $\mathbf{k} := (k_j)_{j\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ be strictly increasing. Put*

$$C_{\varepsilon,\ell} := \frac{-\ell}{\log(1-\varepsilon^2)}$$

and $\lfloor t\, C_{\varepsilon,\ell} \log(\mathbf{k})\, \mathbf{k} \rfloor := \big(\lfloor t\, C_{\varepsilon,\ell} \log(k_j)\, k_j \rfloor\big)_j$. Then,

$$\Big(\Lambda(\mathbf{k}, \lfloor t\, C_{\varepsilon,\ell} \log(\mathbf{k})\, \mathbf{k} \rfloor),\ \mathrm{Mar}_{\varepsilon,\ell-1}(\mathbb{T})\Big) \in \mathscr{A}.$$

Proof. Let $\mu \in \mathrm{Mar}_{\varepsilon,\ell-1}(\mathbb{T})$ and let $2 < \tau < t$. First, we show that

$$f : \mathbb{N} \to \mathbb{N}, \quad f(k) := \lfloor \tau\, C_{\varepsilon,\ell} \log(k)\, k \rfloor$$

is β-approximating for μ. By Lemma 2.5 b),

$$\|\Phi_n\|^2_{L^2(\mathbb{T},\mu)} \le (1-\varepsilon^2)^{\frac{n-1}{\ell}-1}.$$

Thus, we can estimate

$$\begin{aligned}
&\|\Phi_{\tau\lfloor C_{\varepsilon,\ell}\log(k)k\rfloor+1}\|_{L^2(\mathbb{T},\mu)}(2\lfloor \tau C_{\varepsilon,\ell}\log(k)k\rfloor+2)^{k-1}\\
&\quad\le (1-\varepsilon^2)^{\frac{\lfloor \tau C_{\varepsilon,\ell}\log(k)k\rfloor}{2\ell}-\frac{1}{2}}(2\lfloor \tau C_{\varepsilon,\ell}\log(k)k\rfloor+2)^{k-1}\\
&\quad\le (1-\varepsilon^2)^{\frac{\tau C_{\varepsilon,\ell}\log(k)k}{2\ell}-\frac{1}{2}-\frac{1}{2\ell}}(2\tau C_{\varepsilon,\ell}\log(k)k+2)^{k-1}\\
&\quad\le \frac{(1-\varepsilon^2)^{-\frac{1}{2}-\frac{1}{2\ell}}}{2\tau\, C_{\varepsilon,\ell}\log(k)k+2}\cdot\Big((1-\varepsilon^2)^{\frac{\tau\, C_{\varepsilon,\ell}}{2\ell}\log(k)}\big(2\tau\, C_{\varepsilon,\ell}\log(k)k+2\big)\Big)^k\\
&\quad\le \frac{(1-\varepsilon^2)^{-\frac{1}{2}-\frac{1}{2\ell}}}{2\tau\, C_{\varepsilon,\ell}\log(k)k+2}\cdot\Big(k^{-\frac{\tau}{2}}\big(2\tau\, C_{\varepsilon,\ell}\log(k)k+2\big)\Big)^k.
\end{aligned}$$

The right-hand side converges to 0 when k goes to infinity. Thus, by Lemma 3.8, f is β-approximating for μ. Furthermore,

$$\lim_{j\to\infty}\lfloor t\, C_{\varepsilon,\ell}\log(k_j)\,k_j\rfloor - f(k_j) = \lim_{j\to\infty}\lfloor t\, C_{\varepsilon,\ell}\log(k_j)\,k_j\rfloor - \lfloor \tau\, C_{\varepsilon,\ell}\log(k_j)\,k_j\rfloor = \infty$$

since $\tau < t$. Thus, by Theorem 3.3, $E(\Lambda(\mathbf{k}, \lfloor t\, C_{\varepsilon,\ell}\log(\mathbf{k})\,\mathbf{k}\rfloor))$ is dense in $L^2(\mathbb{T},\mu)$. □

As a consequence of Theorem 1.2, we show that if $\Lambda = \mathbb{N}_0 \setminus \Gamma$ with

$$\Gamma = \{\lfloor e^{t^n}\rfloor \mid n \in \mathbb{N}\}$$

for some $t > 1$, then $(\Lambda, \mathrm{Mar}(\mathbb{T})) \in \mathscr{A}$. One can in fact even choose Γ to be a union of large intervals, in the sense that the intervals have double exponential length.

Corollary 4.3. *Let $t \ge \tilde{t} > 1$ and $C > 0$. Let*

$$\Gamma := \bigcup_{j\in\mathbb{N}} [\![\lfloor e^{t^j}\rfloor, \lfloor e^{t^j}\rfloor + \lfloor Ce^{\tilde{t}^j}\rfloor]\!].$$

Then $(N_0\backslash\Gamma, \mathrm{Mar}(\mathbb{T})) \in \mathscr{A}$.

Proof. Let $1 < s < t$. Then

$$\limsup_{j\to\infty}\frac{\lfloor e^{t^{j+1}}\rfloor}{(\lfloor e^{t^j}\rfloor + \lfloor Ce^{\tilde{t}^j}\rfloor+1)^s} \ge \limsup_{j\to\infty}\frac{e^{t^{j+1}}}{2(e^{t^j}+Ce^{\tilde{t}^j}+1)^s} \ge \limsup_{j\to\infty}\frac{e^{t^{j+1}}}{4e^{st^j}} = \infty.$$

Thus, there exists $N \in \mathbb{N}$ such that for every $j \ge N$

$$\lfloor e^{t^{j+1}}\rfloor \ge 2\Big(\lfloor e^{t^j}\rfloor + \lfloor Ce^{\tilde{t}^j}\rfloor + 1\Big)^s.$$

Put $k_j := \lfloor e^{t^j}\rfloor + \lfloor Ce^{\tilde{t}^j}\rfloor + 1$ for every $j \in \mathbb{N}$. It follows that

$$\mathbb{N}_0\backslash\Gamma \supseteq \bigcup_{j\ge N}[\![k_j, \lfloor e^{t^{j+1}}\rfloor]\!] \supseteq \bigcup_{j\ge N}[\![k_j, k_j + \lfloor k_j\rfloor^s]\!] = \Lambda(\mathbf{k}, \lfloor \mathbf{k}^s\rfloor)$$

where $\mathbf{k} := (k_j)_{j\geq N}$ and $\lfloor \mathbf{k}^s \rfloor := (\lfloor k_j^s \rfloor)_{j\geq N}$. Since $E(\Lambda(\mathbf{k}, \lfloor \mathbf{k}^s \rfloor))$ is dense in $L^2(\mathbb{T}, \mu)$ for every $\mu \in \mathrm{Mar}(\mathbb{T})$ by Theorem 1.2, also $E(\mathbb{N}_0 \setminus \Gamma)$ is dense in $L^2(\mathbb{T}, \mu)$ for every $\mu \in \mathrm{Mar}(\mathbb{T})$. $\square$

References

[And58] P. W. Anderson, *Absence of diffusion in certain random lattices*, Phys. Rev. **109** (1958), 1492–1505.

[Bin12] N. H. Bingham, *Szegő's theorem and its probabilistic descendants*, Probab. Surv. **9** (2012), 287–324.

[CL90] R. Carmona and J. Lacroix, *Spectral theory of random Schrödinger operators*, Probab. Appl., Basel etc.: Birkhäuser Verlag, 1990.

[GS58] U. Grenander and G. Szegő, *Toeplitz forms and their applications*, 1958.

[Khr02] S. V. Khrushchev, *Classification theorems for general orthogonal polynomials on the unit circle*, J. Approx. Theory **116** (2002), no. 2, 268–342.

[Kol41] A. N. Kolmogorov, *Stationary sequences in Hilbert's space*, Bolletin Moskovskogo Gosudarstvenogo Universiteta. Matematika **2** (1941), 1–40. (in Russian; reprinted, *Selected works of A. N. Kolmogorov, Vol. 2: Theory of probability and mathematical statistics*, Nauka, Moskva, 1986, 215-255).

[Kre45] M. Krein, *On a generalization of some investigations of G. Szegő, V. Smirnoff and A. N. Kolmogoroff*, C. R. (Dokl.) Acad. Sci. URSS, n. Ser. **46** (1945), 91–94.

[OU21] A. Olevskii and A. Ulanovskii, *On the Szegő-Kolmogorov prediction theorem*, Isr. J. Math. **246** (2021), no. 1, 335–351.

[PF92] L. Pastur and A. Figotin, *Spectra of random and almost-periodic operators*, Grundlehren Math. Wiss., vol. 297, Berlin etc.: Springer-Verlag, 1992 (English).

[Rak83] E. A. Rakhmanov, *On the asymptotics of the ratio of orthogonal polynomials. II*, Math. USSR, Sb. **46** (1983), 105–117.

[Sim05a] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory*, Colloq. Publ., Am. Math. Soc., vol. 54, Providence, RI: American Mathematical Society (AMS), 2005.

[Sim05b] ———, *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*, Colloq. Publ., Am. Math. Soc., vol. 54, Providence, RI: American Mathematical Society, 2005.

[Zhe20] A. Zhedanov, *An explicit example of polynomials orthogonal on the unit circle with a dense point spectrum generated by a geometric distribution*, SIGMA, Symmetry Integrability Geom. Methods Appl. **16** (2020), paper 140, 9.

Chiara Paulsen, University of Wisconsin-Madison, Department of Mathematics, 480 Lincoln Drive, Madison, WI 53706, USA

Email address: cpaulsen3@wisc.edu