Probability Surveys
Vol. 20 (2023) 664–721
ISSN: 1549-5787
https://doi.org/10.1214/23-PS21

Asymptotic behavior of the prediction error for stationary sequences*

Nikolay M. Babayan and Mamikon S. Ginovyan†

Russian-Armenian University, Yerevan, Armenia
Boston University, Boston, USA
e-mail: nmbabayan@gmail.com; ginovyan@math.bu.edu

Abstract: One of the main problem in prediction theory of discrete-time second-order stationary processes $X(t)$ is to describe the asymptotic behavior of the best linear mean squared prediction error in predicting $X(0)$ given $X(t)$, $-n \leq t \leq -1$, as n goes to infinity. This behavior depends on the regularity (deterministic or nondeterministic) and on the dependence structure of the underlying observed process $X(t)$. In this paper we consider this problem both for deterministic and nondeterministic processes and survey some recent results. We focus on the less investigated case – deterministic processes. It turns out that for nondeterministic processes the asymptotic behavior of the prediction error is determined by the dependence structure of the observed process $X(t)$ and the differential properties of its spectral density f, while for deterministic processes it is determined by the geometric properties of the spectrum of $X(t)$ and singularities of its spectral density f.

MSC2020 subject classifications: Primary 60G10, 60G25, 62M15, 62M20; secondary 15A18, 30C10.
Keywords and phrases: Prediction error, deterministic and nondeterministic process, spectral density, Ibragimov's theorems, Rosenblatt's theorems, Szegő's condition, transfinite diameter, eigenvalues of truncated Toeplitz matrices.

Received October 2022.

Contents

arXiv: 2210.06603
*The work is dedicated to our teacher Academician, Professor Il'dar Abdullovich Ibragimov on the occasion of his 90th birthday.
†Corresponding author.

1. Introduction

1.1. The finite prediction problem

Let $X(t)$, $t \in \mathbb{Z} := \{0, \pm 1, \ldots\}$, be a centered discrete-time second-order stationary process. The process is assumed to have an absolutely continuous spectrum with spectral density function $f(\lambda)$, $\lambda \in [-\pi, \pi]$. The *'finite' linear prediction problem* is as follows.

Suppose we observe a finite realization of the process $X(t)$:

$$\{X(t),\ -n \le t \le -1\}, \quad n \in \mathbb{N} := \{1, 2, \ldots\}.$$

We want to make an one-step ahead prediction, that is, to predict the unobserved random variable $X(0)$, using the *linear predictor*

$$Y = \sum_{k=1}^{n} c_k X(-k).$$

The coefficients c_k, $k = 1, 2, \ldots, n$, are chosen so as to minimize *the mean-squared error*: $\mathbb{E}\,|\mathrm{X}(0) - \mathrm{Y}|^{\,2}$, where $\mathbb{E}[\cdot]$ stands for the expectation operator. If such minimizing constants $\widehat{c}_k := \widehat{c}_{k,n}$ can be found, then the random variable

$$\widehat{X}_n(0) := \sum_{k=1}^{n} \widehat{c}_k X(-k)$$

is called *the best linear one-step ahead predictor* of $X(0)$ based on the observed finite past: $X(-n), \ldots, X(-1)$. The minimum mean-squared error:

$$\sigma_n^2(f) := \mathbb{E}\left|\mathrm{X}(0) - \widehat{\mathrm{X}}_\mathrm{n}(0)\right|^{\,2} \ge 0$$

is called *the best linear one-step ahead prediction error* of $X(t)$ based on the past of length n.

One of the main problem in prediction theory of second-order stationary processes, called *the'direct' prediction problem* is to describe the asymptotic behavior of the prediction error $\sigma_n^2(f)$ as $n \to \infty$. This behavior depends on the regularity nature (deterministic or nondeterministic) of the observed process $X(t)$.

Observe that $\sigma^2_{n+1}(f) \leq \sigma^2_n(f)$, $n \in \mathbb{N}$, and hence the limit of $\sigma^2_n(f)$ as $n \to \infty$ exists. Denote by $\sigma^2(f) := \sigma^2_\infty(f)$ the prediction error of $X(0)$ by the entire infinite past: $\{X(t),\, t \leq -1\}$.

From the prediction point of view it is natural to distinguish the class of processes for which we have *error-free prediction* by the entire infinite past, that is, $\sigma^2(f) = 0$. Such processes are called *deterministic* or *singular*. Processes for which $\sigma^2(f) > 0$ are called *nondeterministic*.

Note. The term'deterministic' here is not used in the usual sense of absence of randomness. Instead determinism of a process means that there is an extremely strong dependence between the successive random variables forming the process, yielding error-free prediction when using the entire infinite past (for more about this term see Section 2.4, and also Bingham [10], and Grenander and Szegő [31], p.176.)

Define the *'relative' prediction error*

$$\delta_n(f) := \sigma^2_n(f) - \sigma^2(f),$$

and observe that $\delta_n(f)$ is non-negative and tends to zero as $n \to \infty$. But what about the speed of convergence of $\delta_n(f)$ to zero as $n \to \infty$? The paper deals with this question. Specifically, the prediction problem we are interested in is *to describe the rate of decrease of $\delta_n(f)$ to zero as $n \to \infty$*, depending on the regularity nature of the observed process $X(t)$.

We consider the problem both for deterministic and nondeterministic processes and survey some recent results. We focus on the less investigated case – deterministic processes. It turns out that for nondeterministic processes the asymptotic behavior of the prediction error is determined by the dependence structure of the observed process $X(t)$ and the differential properties of its spectral density f, while for deterministic processes it is determined by the geometric properties of the spectrum of $X(t)$ and singularities of its spectral density f.

1.2. A brief history

The prediction problem stated above goes back to classical works of A. N. Kolmogorov [44, 45], G. Szegő [67, 70] and N. Wiener [75]. It was then considered by many authors for different classes of nondeterministic processes (see, e.g., Baxter [1], Devinatz [17], Geronimus [22, 23], Golinski [26], Golinski and Ibragimov [27], Grenander and Rosenblatt [29, 30], Grenander and Szegő [31], Helson and Szegő [33], Hirshman [34], Ibragimov [36, 37], Ibragimov and Rozanov [39], Ibragimov and Solev [40], Inoue [41], Pourahmadi [53], Rozanov [60], and reference therein). More references can be found in the survey papers Bingham [10] and Ginovyan [25]. In Section 4 of the paper we state some important known results for nondeterministic processes.

We focus in this paper on deterministic processes, that is, on the class of processes for which $\sigma^2(f) = 0$. This case is not only of theoretical interest, but is also important from the point of view of applications. For example, as pointed out by Rosenblatt [59] (see also Pierson [50]), situations of this type arise in Neumann's theoretical model of storm-generated ocean waves. Such models are also of interest in meteorology (see, e.g., Fortus [20]).

Only few works are devoted to the study of the speed of convergence of $\delta_n(f) = \sigma_n^2(f)$ to zero as $n \to \infty$, that is, the asymptotic behavior of the prediction error for deterministic processes. One needs to go back to the classical work of M. Rosenblatt [59]. Using the technique of orthogonal polynomials on the unit circle (OPUC), M. Rosenblatt investigated the asymptotic behavior of the prediction error $\sigma_n^2(f)$ for deterministic processes in the following two cases:

(a) the spectral density $f(\lambda)$ is continuous and positive on a segment of $[-\pi, \pi]$ and is zero elsewhere,
(b) the spectral density $f(\lambda)$ has a very high order of contact with zero at points $\lambda = 0, \pm\pi$, and is strictly positive otherwise.

Later the problems (a) and (b) were studied by Babayan [2, 3], Babayan and Ginovyan [4, 5, 6], Babayan et al. [7] (see also Davisson [15] and Fortus [20]), where some generalizations and extensions of Rosenblatt's results have been obtained.

1.3. Notation and conventions

Throughout the paper we will use the following notation and conventions.

The standard symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of natural, integer, real and complex numbers, respectively. Also, we denote $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$, $\Lambda := [-\pi, \pi]$, $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$. For a point $\lambda_0 \in \Lambda$ and a number $\delta > 0$ by $O_\delta(\lambda_0)$ we denote a δ-neighborhood of λ_0, that is, $O_\delta(\lambda_0) := \{\lambda \in \Lambda : |\lambda - \lambda_0| < \delta\}$. By $L^p(\mu) := L^p(\mathbb{T}, \mu)$ $(p \geq 1)$ we denote the weighted Lebesgue space with respect to the measure μ, and by $\|\cdot\|_{p,\mu}$ we denote the norm in $L^p(\mu)$. In the special case where μ is the Lebesgue measure, we will use the notation L^p and $||\cdot||_p$, respectively. For a function $h \geq 0$ by $G(h)$ we denote the geometric mean of h. For two functions $f(\lambda) \geq 0$ and $g(\lambda) \geq 0$ we will write $f(\lambda) \sim g(\lambda)$ as $\lambda \to \lambda_0$ if $\lim_{\lambda \to \lambda_0} f(\lambda)/g(\lambda) = 1$; $f(\lambda) \simeq g(\lambda)$ as $\lambda \to \lambda_0$ if $\lim_{\lambda \to \lambda_0} f(\lambda)/g(\lambda) = c > 0$, and $f(\lambda) \asymp g(\lambda)$ if there are constants c_1, c_2 $(0 < c_1 \leq c_2 < \infty)$ such that $0 < c_1 \leq f(\lambda)/g(\lambda) \leq c_2 < \infty$ for all $\lambda \in \Lambda$. We will use similar notation for sequences: for two sequences $\{a_n \geq 0, n \in \mathbb{N}\}$ and $\{b_n > 0, n \in \mathbb{N}\}$, we will write $a_n \sim b_n$ if $\lim_{n \to \infty} a_n/b_n = 1$, $a_n \simeq b_n$ if $\lim_{n \to \infty} a_n/b_n = c > 0$, $a_n \asymp b_n$ if $c_1 \leq a_n/b_n \leq c_2$ for all $\lambda \in \mathbb{N}$, $a_n = O(b_n)$ if a_n/b_n is bounded, and $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$. For a set E by $\overline{E}$ we denote the closure of E. The letters C, c, M and m with or without indices are used to denote positive constants, the values of which can vary from line to line.

We will use the abbreviations: OPUC for 'orthogonal polynomials on the unit circle', PACF for 'partial autocorrelation function', and 'a.e.' for 'almost everywhere' (with respect to the Lebesgue measure). We will assume that all the relevant objects are defined in terms of Lebesgue integrals, and so are invariant under change of the integrand on a null set.

1.4. The structure of the paper

The paper is structured as follows. In Section 2 we describe the model of interest – a stationary process, and recall some key notions and results from the theory of stationary processes. In Section 3 we present formulas for the finite prediction error $\sigma_n^2(f)$, and state some preliminary results. In Section 4 we state some well known results on asymptotic behavior of the prediction error for nondeterministic processes. Asymptotic behavior of the finite prediction error $\sigma_n^2(f)$ for deterministic processes is discussed in Section 5. Here we state extensions of Rosenblatt's and Davisson's results, and discuss a number of examples. In Section 6 we analyze the relationship between the rate of convergence to zero of the prediction error $\sigma_n^2(f)$ and the minimal eigenvalue of a truncated Toeplitz matrix generated by the spectral density f. In Section 7 we briefly discuss the tools, used to prove the theorems stated in the paper.

2. The model. Key notions and some basic results

In this section we introduce the model of interest – a second-order stationary process, and recall some key notions and results from the theory of stationary processes.

2.1. Second-order (wide-sense) stationary processes

Let $\{X(t),\ t \in \mathbb{Z}\}$ be a centered real-valued second-order (wide-sense) stationary process defined on a probability space $(\Omega, \mathcal{F}, P)$ with covariance function $r(t)$, that is,

$$\mathbb{E}[X(t)] = 0, \quad r(t) = \mathbb{E}[X(t+s)X(s)], \quad s, t \in \mathbb{Z},$$

where $\mathbb{E}[\cdot]$ stands for the expectation operator with respect to the measure P.

By the Herglotz theorem (see, e.g., Brockwell and Davis [12], p. 117-118), there is a finite measure μ on Λ such that the covariance function $r(t)$ admits the following *spectral representation*:

$$r(t) = \int_{-\pi}^{\pi} e^{-it\lambda} d\mu(\lambda), \quad t \in \mathbb{Z}. \tag{2.1}$$

The measure μ in (2.1) is called the *spectral measure* of the process $X(t)$. If μ is absolutely continuous (with respect to the Lebesgue measure), then the function $f(\lambda) := d\mu(\lambda)/d\lambda$ is called the *spectral density* of $X(t)$. We assume that $X(t)$ is a *non-degenerate* process, that is, $\mathrm{Var}[X(0)] := \mathbb{E}|X(0)|^2 = r(0) > 0$ and, without loss of generality, we may take $r(0) = 1$. Also, to avoid the trivial cases, we assume that the spectral measure μ is *non-trivial*, that is, μ has infinite support.

Notice that if the spectral density $f(\lambda)$ exists, then $f(\lambda) \geq 0$, $f(\lambda) \in L^1(\Lambda)$, and (2.1) becomes

$$r(t) = \int_{-\pi}^{\pi} e^{-it\lambda} f(\lambda) d\lambda, \quad t \in \mathbb{Z}. \tag{2.2}$$

Thus, the covariance function $r(t)$ and the spectral function $F(\lambda)$ (resp. the spectral density function $f(\lambda)$) are equivalent specifications of the second order properties for a stationary process $\{X(t),\ t \in \mathbb{Z}\}$.

Remark 2.1. The parametrization of the unit circle $\mathbb{T}$ by the formula $z = e^{i\lambda}$ establishes a bijection between $\mathbb{T}$ and the interval $[-\pi, \pi)$. By means of this bijection the measure μ on Λ generates the corresponding measure on the unit circle $\mathbb{T}$, which we also denote by μ. Thus, depending on the context, the measure μ will be supported either on Λ or on $\mathbb{T}$. We use the standard Lebesgue decomposition of the measure μ:

$$d\mu(\lambda) = d\mu_a(\lambda) + d\mu_s(\lambda) = f(\lambda)d\lambda + d\mu_s(\lambda), \tag{2.3}$$

where μ_a is the absolutely continuous part of μ (with respect to the Lebesgue measure) and μ_s is the singular part of μ, which is the sum of the discrete and continuous singular components of μ.

By the well-known Cramér theorem (see, e.g., Cramér and Leadbetter [14], Sec. 7.5, Doob [18], p. 481, Shiryaev [66], p. 430), for any stationary process $\{X(t), t \in \mathbb{Z}\}$ with spectral measure μ there exists an orthogonal stochastic measure $Z = Z(B)$, $B \in \mathfrak{B}(\Lambda)$, such that for every $t \in \mathbb{Z}$ the process $X(t)$ admits the following *spectral representation*:

$$X(t) = \int_{\Lambda} e^{-it\lambda} dZ(\lambda), \quad t \in \mathbb{Z}. \tag{2.4}$$

Moreover, $\mathbb{E}\left[|Z(B)|^2\right] = \mu(B)$ for every $B \in \mathfrak{B}(\Lambda)$. Here $\mathfrak{B}(\Lambda)$ stands for the Borel σ-algebra of the sets of Λ. For definition and properties of orthogonal stochastic measures and stochastic integral in (2.4) we refer, e.g., Cramér and Leadbetter [14], Ibragimov and Linnik [38], and Shiryaev [66].

2.2. Linear processes. Existence of spectral density functions

We will consider here stationary processes possessing spectral density functions. For the following results we refer to Ibragimov and Linnik [38], Sect. 16.7, Theorem 16.7.1.

Theorem 2.1. *The following assertions hold.*

(a) The spectral function $F(\lambda)$ of a stationary process $\{X(t), t \in \mathbb{Z}\}$ is absolutely continuous (with respect to the Lebesgue measure), that is, $F(\lambda) = \int_{-\pi}^{\lambda} f(x)dx$, if and only if it can be represented as an infinite moving average:

$$X(t) = \sum_{k=-\infty}^{\infty} a(t-k)\xi(k), \qquad \sum_{k=-\infty}^{\infty} |a(k)|^2 < \infty, \tag{2.5}$$

where $\{\xi(k), k \in \mathbb{Z}\} \sim$ WN(0,1) is a standard white-noise, that is, a sequence of orthonormal random variables.

(b) The covariance function $r(t)$ and the spectral density $f(\lambda)$ of $X(t)$ are given by formulas:

$$r(t) = \mathbb{E}X(t)X(0) = \sum_{k=-\infty}^{\infty} a(t+k)a(k), \tag{2.6}$$

and

$$f(\lambda) = \frac{1}{2\pi}\left|\sum_{k=-\infty}^{\infty} a(k)e^{-ik\lambda}\right|^2 = \frac{1}{2\pi}|\widehat{a}(\lambda)|^2, \quad \lambda \in \Lambda. \tag{2.7}$$

(c) In the case where $\{\xi(k), k \in \mathbb{Z}\}$ is a sequence of Gaussian random variables, the process $\{X(t), t \in \mathbb{Z}\}$ is Gaussian.

2.3. Dependence (memory) structure of the model

Depending on the memory (dependence) structure, we will distinguish the following types of stationary models:

(a) short memory (or short-range dependent),
(b) long memory (or long-range dependent),
(c) intermediate memory (or anti-persistent).

The memory structure of a stationary process is essentially a measure of the dependence between all the variables in the process, considering the effect of all correlations simultaneously. Traditionally memory structure has been defined in the time domain in terms of decay rates of the autocorrelations, or in the frequency domain in terms of rates of explosion of low frequency spectra (see, e.g., Beran et al. [9], and references therein).

It is convenient to characterize the memory structure in terms of the spectral density function.

2.3.1. Short memory models

A stationary process $\{X(t), t \in \mathbb{Z}\}$ with spectral density function $f(\lambda)$ is said to be a *short memory* process if the spectral density $f(\lambda)$ is bounded away from zero and infinity, that is, there are constants C_1 and C_2 such that

$$0 < C_1 \le f(\lambda) \le C_2 < \infty.$$

A typical short memory model example is the stationary Autoregressive Moving Average (ARMA)(p,q) process $X(t)$ defined to be a stationary solution of the difference equation:

$$\psi_p(B)X(t) = \theta_q(B)\varepsilon(t), \quad t \in \mathbb{Z},$$

where ψ_p and θ_q are polynomials respectively of degrees p and q having no zeros on the unit circle $\mathbb{T}$, B is the backshift operator defined by $BX(t) = X(t-1)$, and $\{\varepsilon(t), t \in \mathbb{Z}\}$ is a WN(0,σ^2) white noise, that is, a sequence of zero-mean, uncorrelated random variables with variance σ^2. The covariance function $r(t)$ of (ARMA)(p,q) process is exponentially bounded:

$$|r(t)| \leq Cr^{-t}, \quad t = 1, 2, \ldots; \quad 0 < C < \infty;\ 0 < r < 1,$$

and the spectral density $f(\lambda)$ is a rational function (see, e.g., Brockwell and Davis [12], Section 3.1):

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta_q(e^{-i\lambda})|^2}{|\psi_p(e^{-i\lambda})|^2}. \tag{2.8}$$

2.3.2. Long-memory and anti-persistent models

A *long-memory* model is defined to be a stationary process with *unbounded* spectral density, and an *anti-persistent* model – a stationary process with *vanishing* (at some fixed points) spectral density (see, e.g., Beran et al. [9], Brockwell and Davis [12], and references therein).

A typical model example that displays long-memory and intermediate memory (anti-persistent) is the Autoregressive Fractionally Integrated Moving Average (ARFIMA)(p,d,q) process $X(t)$ defined to be a stationary solution of the difference equation (see, e.g., Brockwell and Davis [12], Section 13.2):

$$\psi_p(B)(1-B)^d X(t) = \theta_q(B)\varepsilon(t), \quad d < 1/2,$$

where B is the backshift operator, $\varepsilon(t)$ is a WN(0,σ^2) white noise, and ψ_p and θ_q are polynomials of degrees p and q, respectively. The spectral density $f_X(\lambda)$ of $X(t)$ is given by

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f(\lambda) = (2\sin(\lambda/2))^{-2d} f(\lambda), \quad d < 1/2, \tag{2.9}$$

where $f(\lambda)$ is the spectral density of an ARMA(p,q) process, given by (2.8). The condition $d < 1/2$ ensures that $\int_{-\pi}^{\pi} f(\lambda)d\lambda < \infty$, implying that the process $X(t)$ is well defined because $\mathbb{E}|X(t)|^2 = \int_{-\pi}^{\pi} f(\lambda)d\lambda$.

Observe that for $0 < d < 1/2$ the model $X(t)$ specified by the spectral density (2.9) displays long-memory. In this case we have $f(\lambda) \sim c\,|\lambda|^{-2d}$ as $\lambda \to 0$, that is, $f(\lambda)$ blows up at $\lambda = 0$ like a power function, which is the typical behavior of a long memory model. For $d < 0$, the model $X(t)$ displays intermediate-memory, and in this case, the spectral density in (2.9) vanishes at $\lambda = 0$. For $d = 0$ the model $X(t)$ displays short-memory. For $d \geq 1/2$ the function $f_X(\lambda)$ in (2.9) is not integrable, and thus it cannot represent a spectral density of a stationary process.

2.4. Deterministic and nondeterministic processes

In this section we state Kolmogorov's isometric isomorphism theorem and the infinite prediction problem. We give time-domain (Wold's theorem) and frequency-domain (Kolmogorov-Szegő theorem) characterizations of deterministic and nondeterministic processes.

2.4.1. Kolmogorov's isometric isomorphism theorem

Given a probability space $(\Omega, \mathcal{F}, P)$, define the L^2-space of real-valued random variables $\xi = \xi(\omega)$ with $\mathbb{E}[\xi] = 0$:

$$L^2(\Omega) := \left\{ \xi : \, ||\xi||^2 := \int_\Omega |\xi(\omega)|^2 dP(\omega) < \infty \right\}. \tag{2.10}$$

Then $L^2(\Omega)$ becomes a Hilbert space with the following inner product: for $\xi, \eta \in L^2(\Omega)$

$$(\xi, \eta) = \mathbb{E}[\xi\eta] = \int_\Omega \xi(\omega)\eta(\omega) \mathrm{dP}(\omega). \tag{2.11}$$

For $a, b \in \mathbb{Z}$, $-\infty \leq a \leq b \leq \infty$, we define the space $H_a^b(X)$ to be the closed linear subspace of the space $L^2(\Omega)$ spanned by the random variables $X(t) = X(t, \omega)$, $t \in [a, b]$, $\omega \in \Omega$:

$$H_a^b(X) := \overline{sp}\{X(t), \; a \leq t \leq b\}_{L^2(\Omega)}. \tag{2.12}$$

Observe that the space $H_a^b(X)$ consists of all finite linear combinations of the form $\sum_{k=a}^b c_k X(k)$, as well as, their $L^2(\Omega)$-limits.

The space $H(X) := H_{-\infty}^{\infty}(X)$ is called the *Hilbert space generated by the process* $X(t)$, or the *time-domain* of $X(t)$.

Let μ be the spectral measure of the process $\{X(t), \, t \in \mathbb{Z}\}$. Consider the weighted L^2-space $L^2(\mu) := L^2(\mu, \Lambda)$ of complex-valued functions $\varphi(\lambda)$, $\lambda \in \Lambda$, defined by

$$L^2(\mu) := \left\{ \varphi(\lambda) : \, ||\varphi||_\mu^2 := \int_\Lambda |\varphi(\lambda)|^2 d\mu(\lambda) < \infty \right\}. \tag{2.13}$$

Then $L^2(\mu)$ becomes a Hilbert space with the following inner product: for $\varphi, \psi \in L^2(\mu)$

$$(\varphi, \psi)_\mu = \int_\Lambda \varphi(\lambda)\overline{\psi}(\lambda) d\mu(\lambda). \tag{2.14}$$

The Hilbert space $L^2(\mu, \Lambda)$ is called the *frequency-domain* of the process $X(t)$.

Theorem 2.2 (Kolmogorov's isometric isomorphism theorem)**.** *For any stationary process $X(t)$, $t \in \mathbb{Z}$, with spectral measure μ there exists a unique isometric isomorphism V between the time-domain $H(X)$ and the frequency-domain $L^2(\mu)$, such that $V[X(t)] = e^{it}$ for any $t \in \mathbb{Z}$.*

Thus, in view of Theorem 2.2, any linear problem in the time-domain $H(X)$ can be translated into one in the frequency-domain $L^2(\mu)$, and vice versa. This fact allows to study stationary processes using analytic methods.

2.4.2. The infinite least-squares prediction problem

Observe first that since by assumption $X(t)$ is a non-degenerate process, the time-domain $H(X)$ of $X(t)$ is non-trivial, that is, $H(X)$ contains elements different from zero.

Definition 2.1. The space $H^t_{-T}(X)$ is called the *finite history*, or *past of length T and present* of the process $X(u)$ up to time t. The space $H_t(X) := H^t_{-\infty}(X)$ is called the *entire history*, or *infinite past and present* of the process $X(u)$ up to time t. The space

$$H_{-\infty}(X) := \cap_t H^t_{-\infty}(X) \tag{2.15}$$

is called the *remote past* of the process $X(u)$.

It is clear that

$$H_{-\infty}(X) \subset \cdots \subset H^t_{-\infty}(X) \subset H^{t+\tau}_{-\infty}(X) \subset \cdots \subset H(X), \quad \tau \in \mathbb{N}. \tag{2.16}$$

The Hilbert space setting provides a natural framework for stating and solving the problem of predicting future values of the process $X(u)$ from the observed past values. Assume that a realization of the process $X(u)$ for times $u \le t$ is observed and we want to predict the value $X(t+\tau)$ for some $\tau \ge 1$ from the observed values. Since we will never know what particular realization is being observed, it is reasonable to consider as a predictor $\widehat{X}(t,\tau)$ for $X(t+\tau)$ a function of the observed values, $g(\{X(u), u \le t\})$, which is good "on the average". So, as an optimality criterion for our predictor we take the L^2-distance, that is, the mean squared error, and consider only the linear predictors. With these restrictions, the infinite linear prediction problem can be stated as follows.

The infinite linear least-squares prediction problem. Given a'parameter' of the process $X(u)$ (e.g., the covariance function $r(t)$ or the spectral function $F(\lambda)$), the entire history $H^t_{-\infty}(X)$ of $X(u)$, and a number $\tau \in \mathbb{N}$, find a random variable $\widehat{X}(t,\tau)$ such that

a) $\widehat{X}(t,\tau)$ is *linear*, that is, $\widehat{X}(t,\tau) \in H^t_{-\infty}(X)$,
b) $\widehat{X}(t,\tau)$ is *mean-square optimal (best)* among all elements $Y \in H^t_{-\infty}(X)$, that is, $\widehat{X}(t,\tau)$ minimizes the mean-squared error $||X(t+\tau) - Y||^2_{L^2(\Omega)}$:

$$||X(t+\tau) - \widehat{X}(t,\tau)||^2_{L^2(\Omega)} = \min_{Y \in H^t_{-\infty}(X)} ||X(t+\tau) - Y||^2_{L^2(\Omega)}. \tag{2.17}$$

The solution – the random variable $\widehat{X}(t,\tau)$ satisfying a) and b), is called the *best linear τ-step ahead predictor* for an element $X(t+\tau) \in H(X)$. The quantity

$$\sigma^2(\tau) := ||X(t+\tau) - \widehat{X}(t,\tau)||^2_{L^2(\Omega)} = ||X(t+\tau)||^2_{L^2(\Omega)} - ||\widehat{X}(t,\tau)||^2_{L^2(\Omega)}, \tag{2.18}$$

which is independent of t, is called the *prediction error (variance)*.

The advantage of the Hilbert space setting now becomes apparent. Namely, by the *projection theorem* in Hilbert spaces (see, e.g., Pourahmadi [53], p. 312), such a predictor $\widehat{X}(t,\tau)$ exists, is unique, and is given by

$$\widehat{X}(t,\tau) = P_t X(t+\tau), \tag{2.19}$$

where $P_t := P^t_{-\infty}$ is the orthogonal projection operator in $H(X)$ onto $H^t_{-\infty}(X)$.

Remark 2.2. The reason for restricting attention to linear predictors is that the best linear predictor $\widehat{X}(t,\tau)$, in this case, depends only on knowledge of the covariance function $r(t)$ or the spectral function $F(\lambda)$. The prediction problem becomes much more difficult when nonlinear predictors are allowed (see, e.g., Hannan [32], Koopmans [46]).

2.4.3. Deterministic (singular) and nondeterministic processes. Characterizations

From prediction point of view it is natural to distinguish the class of processes for which we have *error-free prediction*, that is, $\sigma(\tau) = 0$ for all $\tau \geq 1$, or equivalently, $\widehat{X}(t,\tau) = X(t+\tau)$ for all $t \in \mathbb{Z}$ and $\tau \geq 1$. In this case, the prediction is called *perfect*. It is clear that a process $X(t)$ possessing perfect prediction represents a singular case of *extremely strong dependence* between the random variables forming the process. Such a process $X(t)$ is called *deterministic* or *singular*. From the physical point of view, singular processes are exceptional. From application point of view, of considerable interest is the class of processes for which we have $\sigma(\tau) > 0$ for all $\tau \geq 1$. In this case the prediction is called *imperfect*, and the process $X(t)$ is called *nondeterministic*.

Observe that the time-domain $H(X)$ of any non-degenerate stationary process $\{X(t),\, t \in \mathbb{Z}\}$ can be represented as the orthogonal sum $H(X) = H_1(X) \oplus H_{-\infty}(X)$, where $H_{-\infty}(X)$ is the remote past of $X(t)$ defined by (2.15), and $H_1(X)$ is the orthogonal complement of $H_{-\infty}(X)$. So, we can give the following geometric definition of the deterministic (singular), nondeterministic and purely nondeterministic (regular) processes.

Definition 2.2. A stationary process $\{X(t),\, t \in \mathbb{Z}\}$ is called

- *deterministic or singular* if $H_{-\infty}(X) = H(X)$, i.e., $H^t_{-\infty}(X) = H^s_{-\infty}(X)$ for all $t, s \in \mathbb{Z}$*,*
- *nondeterministic* if $H_{-\infty}(X)$ is a proper subspace of $H(X)$, i.e., $H_{-\infty}(X) \subset H(X)$*,*
- *purely nondeterministic (PND) or regular* if $H_{-\infty}(X) = \{0\}$, that is, the remote past $H_{-\infty}(X)$ of $X(t)$ is the trivial subspace, consisting of the singleton zero.

The next theorem contains a characterization of deterministic and purely nondeterministic processes in terms of prediction error.

Theorem 2.3. *A stationary process* $\{X(t),\, t \in \mathbb{Z}\}$ *is*

(a) deterministic if and only if $\sigma(\tau_0) = 0$ *for some* $\tau_0 \geq 1$, $\tau_0 \in \mathbb{N}$ *(then* $\sigma(\tau) = 0$ *for all* $\tau \in \mathbb{N}$*).*
(b) purely nondeterministic if and only if $\lim_{\tau\to\infty} \sigma^2(\tau) = E|X(t)|^2 = r(0)$.

Remark 2.3. Every purely nondeterministic process $X(t)$ is nondeterministic, but the converse is generally not true. An example of such process provides the process $X(t) = \varepsilon(t) + \xi$, where $\{\varepsilon(t),\, t \in \mathbb{Z}\}$ is a WN$(0, \sigma^2)$ (a centered white noise with variance σ^2), and ξ is a random variable such that $Var(\xi) = \sigma_0^2$ and $(\varepsilon(t), \xi) = 0$ for all $t \in \mathbb{Z}$ (see Pourahmadi [53], p. 163). Observe also that the only process $X(t)$ that is both deterministic and purely nondeterministic is the degenerate process. Assuming that $X(t)$ is a non-degenerate process, we exclude this trivial case.

The next result, known as Wold's decomposition theorem (see, e.g., Brockwell and Davis [12], Sec. 5.7, Shiryaev [66], Sec. VI.5), provides a key step for solution of the infinite prediction problem in the time-domain setting, and essentially says that any stationary process can be represented in the form of a sum of two orthogonal stationary components, one of which is perfectly predictable (singular component), while for the other (regular component) an explicit formula for the predictor can be obtained.

Theorem 2.4 (Wold's decomposition)**.** *Every centered non-degenerate discrete-time stationary process $X(t)$ admits a unique decomposition:*

$$X(t) = X_S(t) + X_R(t),$$

where

- *(a) the processes $X_R(t)$ and $X_S(t)$ are stationary, centered, mutually uncorrelated (orthogonal), and completely subordinated to $X(t)$, i.e., $H_{-\infty}^t(X_R) \subseteq H_{-\infty}^t(X)$ and $H_{-\infty}^t(X_S) \subseteq H_{-\infty}^t(X)$ for all $t \in \mathbb{Z}$.*
- *(b) the process $X_S(t)$ is deterministic (singular),*
- *(c) the process $X_R(t)$ is purely nondeterministic (regular) and has the infinite moving-average representation:*

$$X_R(t) = \sum_{k=0}^{\infty} a_k \varepsilon_0(t-k), \quad \sum_{k=0}^{\infty} |a_k|^2 < \infty, \tag{2.20}$$

where $\varepsilon_0(t)$ is an innovation of $X_R(t)$, that is, $\varepsilon_0(t)$ is a standard white-noise process, such that $H_{-\infty}^t(X_R) = H_{-\infty}^t(\varepsilon_0)$ for all $t \in \mathbb{Z}$.

The next result describes the asymptotic behavior of the prediction error $\sigma_n^2(\mu)$ for a stationary process $X(t)$ with spectral measure μ of the form (2.3) and gives spectral characterizations of deterministic, nondeterministic and purely nondeterministic processes (see, e.g., Grenander and Szegő [31], p. 44, and Ibragimov and Rozanov [39], p. 35–36).

Theorem 2.5. *Let $X(t)$ be a non-degenerate stationary process with spectral measure μ of the form (2.3). The following assertions hold.*

(a) (Kolmogorov-Szegő Theorem). *The following relations hold.*

$$\lim_{n\to\infty} \sigma_n^2(\mu) = \lim_{n\to\infty} \sigma_n^2(f) = \sigma^2(f) = 2\pi G(f), \tag{2.21}$$

where $G(f)$ is the geometric mean of f, namely

$$G(f) := \begin{cases} \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda\right\} & \text{if} \quad \ln f \in L^1(\Lambda) \\ 0, & \text{otherwise,} \end{cases} \tag{2.22}$$

(b) $H^0_{-\infty}(\mu_s) = H(\mu_s) \Leftrightarrow \sigma^2(\mu) = 0 \Leftrightarrow X(t)$ *is deterministic,*

(c) (Kolmogorov-Szegő alternative). *Either*

$$H^0_{-\infty}(\mu_a) = H(\mu_a) \Leftrightarrow \int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda = -\infty \Leftrightarrow \sigma^2(f) = 0,$$

or else

$$H^0_{-\infty}(\mu_a) \neq H(\mu_a) \Leftrightarrow \int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda > -\infty \Leftrightarrow \sigma^2(f) > 0.$$

(d) The process $X(t)$ *is regular (PND) if and only if it is nondeterministic and* $\mu_s \equiv 0$.

Remark 2.4. The second equality in (2.21) was proved by Szegő in 1920, while the first equality was proved by Kolmogorov in 1941 (see, e.g., Hoffman [35], p. 49). *It is remarkable that* (2.21) *is independent of the singular part* μ_s.

The condition $\ln f \in L^1(\Lambda)$ in (2.22) is equivalent to the *Szegő condition*:

$$\int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda > -\infty \tag{2.23}$$

(this equivalence follows because $\ln f(\lambda) \leq f(\lambda)$). The Szegő condition (2.23) is also called the *non-determinism condition*.

In this paper we consider the class of deterministic processes with absolutely continuous spectra.

We will say that the spectral density $f(\lambda)$ has a *very high order of contact with zero at a point* λ_0 if $f(\lambda)$ is positive everywhere except for the point λ_0, due to which the Szegő condition (2.23) is violated. Observe that the Szegő condition is related to the character of the singularities (zeroes and poles) of the spectral density f, and does not depend on the differential properties of f. For example, for any $a > 0$, the function $\widehat{f}_a(\lambda) = \exp\{-|\lambda|^{-a}\}$ is infinitely differentiable. In addition, for $a < 1$ Szegő's condition is satisfied, and hence the corresponding process $X(t)$ is nondeterministic, while for $a \geq 1$ Szegő's condition is violated, and $X(t)$ is deterministic (see, e.g., Pourahmadi [53], p.68, Rakhmanov [55]). Thus, according to the above definition, for $a \geq 1$ this function has a very high order of contact with zero at the point $\lambda = 0$.

3. Formulas for the finite prediction error $\sigma_n^2(f)$ and some properties

In this section, we provide various formulas for the prediction error $\sigma_n^2(f)$ in terms of orthogonal polynomials and their parameters (Verblunsky's coefficients), Toeplitz determinants, state Szegő's, Verblunsky's and Rakhmanov's theorems, and list a number of properties of $\sigma_n^2(f)$.

3.1. Formulas for the prediction error $\sigma_n^2(f)$

We present here formulas for the finite prediction error $\sigma_n^2(f)$ and state some preliminary results, which will be used in the sequel.

Suppose we have observed the values $X(-n),\dots,X(-1)$ of a centered, real-valued stationary process $X(t)$ with spectral measure μ of the form (2.3). The *one-step ahead linear prediction problem* in predicting a random variable $X(0)$ based on the observed values $X(-n),\dots,X(-1)$ involves finding constants $\widehat{c}_k := \widehat{c}_{k,n}$, $k=1,2,\dots,n$, that minimize the one-step ahead prediction error:

$$\sigma_n^2(\mu) := \min_{\{c_k\}} \mathbb{E}\left|X(0)-\sum_{k=1}^{n} c_k X(-k)\right|^2 = \mathbb{E}\left|X(0)-\sum_{k=1}^{n} \widehat{c}_k X(-k)\right|^2. \tag{3.1}$$

Using Kolmogorov's isometric isomorphism $V: X(t)\leftrightarrow e^{it\lambda}$, in view of (3.1), for the prediction error $\sigma_n^2(\mu)$ we can write

$$\sigma_n^2(\mu) = \min_{\{c_k\}} \left\|1-\sum_{k=1}^{n} c_k e^{-ik\lambda}\right\|_{2,\mu}^2 = \min_{\{q_n\in\mathcal{Q}_n\}} \|q_n\|_{2,\mu}^2 , \tag{3.2}$$

where $||\cdot||_{2,\mu}$ is the norm in $L^2(\mathbb{T},\mu)$, and

$$\mathcal{Q}_n := \left\{q_n : q_n(z) = z^n + c_1 z^{n-1} + \cdots c_n\right\} \tag{3.3}$$

is the class of monic polynomials (i.e. with $c_0=1$) of degree n. Thus, the problem of finding $\sigma_n^2(\mu)$ becomes to the problem of finding the solution of the minimum problem (3.2)–(3.3).

The polynomial $p_n(z) := p_n(z,\mu)$ which solves the minimum problem (3.2)–(3.3) is called the *optimal polynomial* for μ in the class $\mathcal{Q}_n$. This minimum problem was solved by G. Szegő by showing that the optimal polynomial $p_n(z)$ exists, is unique and can be expressed in terms of *orthogonal polynomials on the unit circle* with respect to the measure μ (see Theorem 3.2 below).

To state Szegő's solution of the minimum problem (3.2)–(3.3), we first recall some facts from the theory of orthogonal polynomials on the unit circle (OPUC).

The system of orthogonal polynomials on the unit circle associated with the measure μ:

$$\{\varphi_n(z)=\varphi_n(z;f), \quad z=e^{i\lambda}, \quad n\in\mathbb{Z}_+\}$$

is uniquely determined by the following two conditions:

(i) $\varphi_n(z)=\kappa_n z^n+\cdots+l_n$ is a polynomial of degree n, in which the leading coefficient κ_n is positive;

(ii) $(\varphi_k,\varphi_j)_\mu=\delta_{kj}$ for arbitrary $k,j\in\mathbb{Z}_+$, where δ_{kj} is the Kronecker delta.

Define the monic ($p_n(z)$) and the reciprocal ($p_n^*(z)$) polynomials (see, e.g., Simon [65], p. 2):

$$p_n(z) := p_n(z,\mu) = \kappa_n^{-1}\varphi_n(z) = z^n+\cdots+l_n\kappa_n^{-1}, \tag{3.4}$$

$$p_n^*(z) := p_n^*(z,\mu) = z^n\overline{p_n(1/\overline{z})} = \bar{l}_n\kappa_n^{-1}z^n + \cdots + 1. \tag{3.5}$$

We have

$$||p_n||_{2,\mu} = ||p_n^*||_{2,\mu} = \kappa_n^{-2}. \tag{3.6}$$

The polynomials $p_n(z)$ and $p_n^*(z)$ satisfy *Szegő's recursion relation* (see Simon [65], p. 56):

$$p_{n+1}(z) = zp_n(z) - \overline{v}_{n+1}p_n^*(z), \quad n \in \mathbb{Z}_+ \tag{3.7}$$

where

$$v_{n+1} = -\overline{p_{n+1}(0)} = \bar{l}_{n+1}\kappa_{n+1}^{-1}, \quad |v_{n+1}| < 1. \tag{3.8}$$

In view of (3.7) we have (see Simon [65], p. 56)

$$||p_n||_{2,\mu}^2 = (1-|v_n|^2)||p_{n-1}||_{2,\mu}^2 = \prod_{j=1}^{n}(1-|v_j|^2), \quad n \in \mathbb{N}. \tag{3.9}$$

From (3.6) and (3.9) we obtain

$$\kappa_n^2\kappa_{n+1}^{-2} = 1 - |v_{n+1}|^2. \tag{3.10}$$

The parameters $v_n := v_n(\mu)$ ($n \in \mathbb{N}$), which play an important role in the theory of OPUC, are called *Verblunsky coefficients* (also known as the Szegő, Schur, and canonical moments; see Simon [65], Sect. 1.1, and Dette and Studden [16], Sect. 9.4).

Note. The term 'Verblunsky coefficient' is from Simon [65]. Observe that we write v_{n+1} for Simon's α_n, and so one has $n \in \mathbb{N}$ for Simon's $n \in \mathbb{Z}_+$. Our notational convention is already established in the time-series literature and is more convenient in our context of the PACF (defined below), where $n \in \mathbb{N}$ (see Bingham [10], Brockwell and Davis [12], Sec. 5.2, Inoue [41], Pourahmadi [53], Sec. 7.3).

The following result shows that Verblunsky coefficients provide a convenient way for the parametrization of probability measures on the unit circle $\mathbb{T}$ (see, e.g., Verblunsky [72, 73], Ramsey [56], Simon [65], p. 2).

Theorem 3.1 (Verblunsky [72]). *Let $\mathbb{D}^\infty := \times_{k=0}^{\infty}\mathbb{D}$ be the set of complex sequences $v := (v_n, n \in \mathbb{N})$ with $v_n \in \mathbb{D}$. The map $\mathcal{S} : \mu \longmapsto v$ is a bijection between the set of nontrivial probability measures $\{\mu\}$ on $\mathbb{T}$ and $\mathbb{D}^\infty$.*

This result was established by Verblunsky [72] in 1935, in connection with OPUC. It was re-discovered by Ramsey [56] in 1974, in connection with parametrization of time-series models.

Partial autocorrelation function(PACF). For a stationary process $X(t)$ with a non-trivial spectral measure μ the *partial autocorrelation function* (PACF) of $X(t)$, denoted by $\pi_n = \pi_n(\mu)$ ($n \in \mathbb{N}$), is defined to be the correlation coefficient between the forward and backward residuals in the linear prediction of the variables $X(n)$ and $X(0)$ on the basis of the intermediate observations $X(1), \ldots, X(n-1)$, that is,

$$\pi_n := \mathrm{corr}(X(n) - \widehat{X}(n), X(0) - \widehat{X}(0)).$$

It turns out that the Verblunsky coefficients v_n and the PACF π_n coincide, that is, $v_n = \pi_n$ for all $n \in \mathbb{N}$ (see Dette and Studden [16], Sect. 9.6). Thus, the Verblunsky sequence $v := (v_n, n \in \mathbb{N})$ provides a link between OPUC and time-series analysis, and, in view of the equality $v_n = \pi_n$, the Verblunsky bijection gives a *simple and unconstrained parametrization* of stationary processes, in contrast to using the covariance function, which has to be positive-definite.

The next result by Szegő solves the minimum problem (3.2)–(3.3) (see, e.g., Grenander and Szegő [31], p. 38).

Theorem 3.2 (Szegő)**.** *The unique solution of the minimum problem* (3.2)–(3.3) *is the monic polynomial* $p_n(\mu) := p_n(z,\mu)$ *given by formula* (3.4)*, and the minimum in* (3.2) *is equal to* $||p_n||_{2,\mu} = \kappa_n^{-2}$ *(see* (3.6)*).*

Thus, for the prediction error $\sigma_n^2(\mu)$ we have the following formula:

$$\sigma_n^2(\mu) = \min_{\{q_n \in \mathcal{Q}_n\}} \|q_n\|_{2,\mu}^2 = \|p_n(\mu)\|_{2,\mu}^2 = \kappa_n^{-2}. \tag{3.11}$$

Remark 3.1. Define

$$\mathcal{Q}_n^* := \left\{ q_n : q_n(z) = c_0 z^n + c_1 z^{n-1} + \cdots c_n,\ c_n = 1 \right\}, \tag{3.12}$$

and observe that the classes of polynomials $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ defined in (3.3) and (3.12), respectively, differ by normalization: in (3.12) we have $c_n = 1$, while in (3.3) we have $c_0 = 1$. Also, the optimal polynomial for μ in the class $\mathcal{Q}_n^*$ is the reciprocal polynomial $p_n^*(z)$ (see (3.5)). Taking into account (3.6), we have the following formula for the prediction error $\sigma_n^2(\mu)$ in terms of the optimal polynomial $p_n^*(z)$:

$$\sigma_n^2(\mu) = \min_{\{q_n \in \mathcal{Q}_n^*\}} \|q_n\|_{2,\mu}^2 = \|p_n^*(\mu)\|_{2,\mu}^2. \tag{3.13}$$

Remark 3.2. Denote by $D_n(\mu)$ the n^{th} Toeplitz determinant generated by the measure μ:

$$D_n(\mu) := \det[r(t-s),\ t,s = 0,1,\dots n],$$

where $r(t)$ is the covariance function given by (2.1). Taking into account that $\kappa_n^2 = D_{n-1}(\mu)/D_n(\mu)$ (see, e.g., Grenander and Szegő [31], p. 38), in view of (3.11) we obtain the following formula for the prediction error $\sigma_n^2(\mu)$ in terms of $D_n(\mu)$:

$$\sigma_n^2(\mu) = \frac{D_n(\mu)}{D_{n-1}(\mu)}. \tag{3.14}$$

Remark 3.3. In view of (3.11), the formulas (3.9) and (3.10) can be written as follows

$$\sigma_n^2(\mu) = \prod_{j=1}^{n} (1 - |v_j|^2) \quad \text{and} \quad \frac{\sigma_{n+1}^2(\mu)}{\sigma_n^2(\mu)} = 1 - |v_n|^2. \tag{3.15}$$

From the second formula in (3.15), it follows that the convergence of the sequences $|v_n|$ and $\sigma_{n+1}(\mu)/\sigma_n(\mu)$ are equivalent, and, the greater the limiting value of $|v_n|$, the faster the rate of decrease of $\sigma_n(\mu)$.

For a general measure μ of the form (2.3) the asymptotic relation

$$\lim_{n\to\infty} v_n(\mu) = 0 \tag{3.16}$$

is of special interest in the theory of OPUC. In this respect the following question arises naturally: what is the'minimal' sufficient condition on the measure μ ensuring the relation (3.16)? The next result of Rakhmanov [54, 55] shows that for (3.16) (or equivalently, for $\lim_{n\to\infty} \sigma_{n+1}(\mu)/\sigma_n(\mu) = 1$) it is enough only to have a.e. positiveness on $\mathbb{T}$ of the spectral density f (see also Babayan et al. [7] and Simon [65], p. 5).

Theorem 3.3 (Rakhmanov [54])**.** *Let the measure μ have the form (2.3) with $f > 0$ a.e. on $\mathbb{T}$. Then the asymptotic relation (3.16) is satisfied.*

Note that the converse of Rakhmanov's theorem, in general, is not true. A partial converse of Rakhmanov's theorem is stated in Theorem 5.7.

Bello and López [8] proved the following extension of Rakhmanov's theorem: *Let Γ_δ be a closed arc of the unit circle of length 2δ $(0 < \delta \le \pi)$, and let μ and (v_n) be as in Rakhmanov's Theorem. Assume that the measure μ is supported on the arc Γ_δ with $f > 0$ a.e. on Γ_δ. Then* $\lim_{n\to\infty} |v_n| = \cos(\delta/2)$. The case $\delta = \pi$ corresponds to Rakhmanov's theorem.

3.2. Properties of the prediction error $\sigma_n^2(f)$

In what follows we assume that the spectral measure μ is absolutely continuous with spectral density f, and instead of $\sigma_n^2(\mu)$, $p_n(\mu)$ and $D_n(\mu)$ we use the notation $\sigma_n^2(f)$, $p_n(f)$ and $D_n(f)$, respectively.

In the next proposition we list a number of properties of the prediction error $\sigma_n^2(f)$. The proof can be found in Babayan and Ginovyan [5].

Proposition 3.1. *The prediction error $\sigma_n^2(f)$ possesses the following properties.*

(a) $\sigma_n^2(f)$ *is a non-decreasing functional of f: $\sigma_n^2(f_1) \le \sigma_n^2(f_2)$ when $f_1(\lambda) \le f_2(\lambda)$, $\lambda \in [-\pi, \pi]$.*
(b) If $f(\lambda) = g(\lambda)$ almost everywhere on $[-\pi, \pi]$, then $\sigma_n^2(f) = \sigma_n^2(g)$.
(c) For any positive constant c we have $\sigma_n^2(cf) = c\sigma_n^2(f)$.
(d) If $\bar{f}(\lambda) = f(\lambda - \lambda_0)$, $\lambda_0 \in [-\pi, \pi]$, then $\sigma_n^2(\bar{f}) = \sigma_n^2(f)$.

4. Asymptotic behavior of the prediction error $\delta_n(f)$ for nondeterministic processes

In this section we study the asymptotic behavior of the the relative prediction error $\delta_n(f) = \sigma_n^2(f) - \sigma^2(f)$ for nondeterministic processes, and review some important known results.

4.1. Asymptotic behavior of $\delta_n(f)$ for short-memory processes

Recall that a *short memory* processes is a second order stationary processes possessing a spectral density f which is bounded away from zero and infinity.

4.1.1. Exponential rate of decrease of $\delta_n(f)$

The first result of this type goes back to the Grenander and Rosenblatt [29]. The next theorem was proved by Ibragimov [36].

Theorem 4.1 (Ibragimov [36]). *A necessary and sufficient condition for*

$$\delta_n(f) = O(q^n), \quad q = e^{-c}, \quad c > 0, \quad n \to \infty \tag{4.1}$$

is that $f(\lambda)$ is a spectral density of a short-memory process, and $1/f(\lambda) \in A_c$, where A_c is the class of 2π–periodic continuous functions $\varphi(\lambda)$, $\lambda \in \mathbb{R}$, admitting an analytic continuation into the strip $z = \lambda + i\mu$, $-\infty < \lambda < \infty$, $|\mu| \le c$.

Observe that (4.1) will be true for all $c > 0$ if and only if the analytic continuation of $f(\lambda)$ is an entire function of $z = \lambda + i\mu$.

Thus, to have exponential rate of decrease to zero for $\delta_n(f)$ the underlying model should be a short-memory process with sufficiently smooth spectral density.

4.1.2. Hyperbolic rate of decrease of $\delta_n(f)$

Here we are interested in estimates for $\delta_n(f)$ of type

$$\delta_n(f) = O(n^{-\gamma}), \quad \gamma > 0, \quad n \to \infty. \tag{4.2}$$

$$\delta_n(f) = o(n^{-\gamma}), \quad \gamma > 0, \quad n \to \infty. \tag{4.3}$$

Bounds of type (4.2) with $\gamma > 1$ for different classes of spectral densities were obtained by Baxter [1], Devinatz [17], Geronimus [23, 24], Grenander and Rosenblatt [29], Grenander and Szegő [31], and others (see, e.g., Ginovyan [25], and references therein). The most general result in this direction has been obtained by Ibragimov [36]. To state Ibragimov's theorem, we first introduce Hölder classes of functions.

For a function $\varphi(\lambda) \in L^p(\mathbb{T})$, we define its L^p-modulus of continuity by

$$\omega_p(\varphi;\delta) = \sup_{0<|t|\le\delta} ||\varphi(\cdot + t) - \varphi(\cdot)||_p, \quad \delta > 0. \tag{4.4}$$

Given numbers $0 < \alpha < 1$, $r \in \mathbb{Z}_0 := \{0, 1, 2 \ldots\}$, and $p \ge 1$, we put $\gamma := r + \alpha$. A Hölder class of functions, denoted by $H_p(\gamma)$, is defined to be the set of those functions $\varphi(\lambda) \in L^p(\mathbb{T})$ that have r-th derivative $\varphi^{(r)}(\lambda)$, such that $\varphi^{(r)}(\lambda) \in L^p(\mathbb{T})$ and $\omega_p(\varphi^{(r)};\delta) = O(\delta^\alpha)$ as $\delta \to 0$.

Theorem 4.2 (Ibragimov [36]). *A necessary and sufficient condition for*

$$\delta_n(f) = O(n^{-\gamma}), \quad \gamma = 2(r+\alpha) > 1;\ 0 < \alpha < 1,\ r \in \mathbb{Z}_0, \quad \text{as} \quad n \to \infty \tag{4.5}$$

is that $f(\lambda)$ is a spectral density of a short-memory process belonging to $H_2(\gamma)$.

Bounds of type (4.3) for short memory models have been obtained by Baxter [1], Devinatz [17], Hirshman [34] and Golinskii [26]. For 'large' values of γ ($\gamma > 1$), Hirshman has obtained the following necessary and sufficient condition for (4.3) (see Hirshman [34], p. 314).

Theorem 4.3 (Hirshman [34])**.** *If $X(t)$ is a short-memory process, then $\delta_n(f) = o(n^{-\gamma})$ with $\gamma > 1$ as $n \to \infty$ if and only if $\sum_{|t|\geq n} |r(t)|^2 = o(n^{-\gamma})$ as $n \to \infty$, where $r(t)$ is the covariance function of $X(t)$.*

The next theorem was proved by G. Baxter (see Baxter [1], Theorem 3.1).

Theorem 4.4 (Baxter [1])**.** *If $X(t)$ is a short-memory process with covariance function $r(t)$ satisfying $\sum_{t=1}^{\infty} t^{\gamma}|r(t)| < \infty$, $\gamma > 0$, then $\delta_n(f) = o(n^{-2\gamma})$ as $n \to \infty$.*

Remark 4.1. It follows from Theorem 4.2 that if $\delta_n(f) = O(n^{-\gamma})$ with $\gamma > 1$, then the underlying model $X(t)$ is necessarily a short-memory process. Moreover, as it was pointed out by Grenander and Rosenblatt [29] (see, also, Devinatz [17], p. 118), if the model is not a short-memory process, that is, the spectral density f has zeros or is unbounded, then, in general, we cannot expect $\delta_n(f)$ to go to zero faster than $1/n$ as $n \to \infty$. This question we discuss in the next section.

4.2. Asymptotic behavior of $\delta_n(f)$ for anti-persistent and long-memory processes

In this section we describe the rate of decrease of the relative prediction error $\delta_n(f)$ to zero as $n \to \infty$, in the case where the underlying process $X(t)$ is nondeterministic and is anti-persistent or has long-memory, that is, the spectral density $f(\lambda)$ of $X(t)$ has zeros or is unbounded at a finite number of points, such that $\ln f(\lambda) \in L^1(\mathbb{T})$. This case is of great interest because in many applications the model is described by processes of a such type.

4.2.1. An example

We start with an example, which shows that the asymptotic behavior of the prediction error $\delta_n(f)$ essentially depends on the dependence (memory) structure of the underlying model $X(t)$ (see, e.g., Grenander and Szegő [31], p. 191).

Example 4.1. Consider a first-order moving-average $MA(1)$ process $X(t)$:

$$X(t) = \varepsilon(t) - b \cdot \varepsilon(t-1), \quad \varepsilon(t) \sim WN(0, \sigma_\varepsilon^2), \quad 0 \leq b \leq 1.$$

The spectral density is $f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot |1 - be^{i\lambda}|^2$ (see formula (2.8)).

a) First assume that $X(t)$ has short-memory, that is, $0 \leq b < 1$. It is easy to check that

$$\delta_n(f) = \sigma_n^2(f) - \sigma^2(f) = \frac{b^{2n}(b^2 - b^4)}{1 - b^{2n+2}} \sim b^{2n} \quad \text{as} \quad n \to \infty,$$

showing that in this case $\delta_n(f)$ goes to zero with exponential rate.

b) Now let $b=1$. We have $f(\lambda)=\frac{\sigma_\varepsilon^2}{2\pi}\cdot|1-e^{i\lambda}|^2$, that is, the process $X(t)$ is anti-persistent. In this case we have

$$\delta_n(f)=\sigma_n^2(f)-\sigma^2(f)=\frac{1}{n+2}\sim\frac{1}{n}\quad\text{as}\quad n\to\infty,$$

showing that $\delta_n(f)$ goes to zero at precisely the rate $1/n$. The slow rate is due to the presence of a zero of $f(\lambda)$ at $\lambda=0$ (see Grenander and Szegő [31], p. 191).

It can be shown that for models with spectral densities of the form $f(\lambda)=\frac{\sigma_\varepsilon^2}{2\pi}\cdot|q(e^{i\lambda})|^2$, where $q(e^{i\lambda})$ is a polynomial with at least one root on the unit circle $\mathbb{T}$, we have $\delta_n(f)\sim\frac{1}{n}$ as $n\to\infty$.

4.2.2. The ARFIMA(p,d,q) model

As was mentioned in Section 2.3 a well-known example of processes that displays long memory or is anti-persistent is an ARFIMA(p,d,q) process $X(t)$ with spectral density $f_X(\lambda)$ given by (see (2.9)):

$$f_X(\lambda)=|1-e^{-i\lambda}|^{-2d}f(\lambda),\quad d<1/2,\tag{4.6}$$

where $f(\lambda)$ is the spectral density of an ARMA(p,q) process, given by (2.8). Recall that for $0<d<1/2$ the model $X(t)$ specified by spectral density (4.6) displays long-memory, for $d<0$ it is anti-persistent, and for $d=0$ it displays short-memory.

The following theorem was proved by A. Inoue (see Inoue [41], Theorem 4.3).

Theorem 4.5 (Inoue [41])**.** *Let $f_X(\lambda)$ have the form (4.6) with $0<d<1/2$, where $f(\lambda)$ is the spectral density of an ARMA(p,q) process. Then*

$$\delta_n(f)\sim\frac{d^2}{n}\quad\text{as}\quad n\to\infty.\tag{4.7}$$

Remark 4.2. Note that for ARFIMA$(0,d,0)$ model the asymptotic relation (4.7) remains valid for all $d<1/2$ ($d\neq 0$). In this case, for the Verblunsky coefficients (parameters) v_n we have $v_n=\frac{d}{n-d+1}$ (see Golinskii [26], p. 703).

4.2.3. The Jacobian model

Another well-known example of processes that display long memory or is anti-persistent is the Jacobian model. We say that a stationary process $X(t)$ is a Jacobian process, and the corresponding model is a Jacobian model, if its spectral density $f(\lambda)$ has the following form:

$$f(\lambda)=f_1(\lambda)\prod_{k=1}^{m}|e^{i\lambda}-e^{i\lambda_k}|^{-2d_k},\tag{4.8}$$

where $f_1(\lambda)$ is the spectral density of a short-memory process, the points $\lambda_k \in [-\pi, \pi]$ are distinct, and $d_k \le 1/2$, $k = 1, \ldots, m$.

The asymptotic behavior of $\delta_n(f)$ as $n \to \infty$ for Jacobian model (4.8) has been considered in a number of papers (see, e.g., Golinskii [26], Grenander and Rosenblatt [29], Ibragimov [36], Ibragimov and Solev [40].)

The following theorem was proved in Ibragimov and Solev [40].

Theorem 4.6 (Ibragimov and Solev [40])**.** *Let $f(\lambda)$ have the form (4.8), where $f_1(\lambda)$ is the spectral density of a short-memory process, the points $\lambda_k \in [-\pi, \pi]$ are distinct, and $d_k \le 1/2$, $k = 1, \ldots, m$. If the function $f_1(\lambda)$ satisfies a Lipschitz condition with exponent $\alpha \ge 1/2$, then*

$$\delta_n(f) \sim 1/n \quad \text{as} \quad n \to \infty. \tag{4.9}$$

5. Asymptotic behavior of the finite prediction error $\sigma_n^2(f)$ for deterministic (singular) processes

5.1. Background

The linear prediction problem has been studied most intensively for nondeterministic processes, that is, in the case where the prediction error is known to be positive ($\sigma^2(f) > 0$) (see Section 4).

In this section we focus on the less investigated case – deterministic processes, that is, when $\sigma^2(f) = 0$. This case is not only of theoretical interest, but is also important from the point of view of applications. For example, as pointed out by Rosenblatt [59] (see also Pierson [50]), situations of this type arise in Neumann's theoretical model of storm-generated ocean waves. Such models are also of interest in meteorology (see, e.g., Fortus [20]).

Only few works are devoted to the study of the speed of convergence of $\delta_n(f) = \sigma_n^2(f)$ to zero as $n \to \infty$, that is, the asymptotic behavior of the prediction error for deterministic processes. One needs to go back to the classical work of M. Rosenblatt [59]. Using the technique of orthogonal polynomials on the unit circle, M. Rosenblatt investigated the asymptotic behavior of the prediction error $\sigma_n^2(f)$ for deterministic processes in the following two cases:

(a) the spectral density $f(\lambda)$ is continuous and positive on a segment of $[-\pi, \pi]$ and zero elsewhere.

(b) the spectral density $f(\lambda)$ has a very high order of contact with zero at points $\lambda = 0, \pm\pi$, and is strictly positive otherwise.

Later the problems (a) and (b) were studied by Babayan [2, 3], Babayan and Ginovyan [4, 5, 6], and Babayan et al. [7] (see also Davisson [15] and Fortus [20]), where some generalizations and extensions of Rosenblatt's results were obtained.

We start by describing Rosenblatt's results concerning the asymptotic behavior of the prediction error $\sigma_n^2(f)$, obtained in Rosenblatt [59] for the above stated cases (a) and (b).

5.2. Rosenblatt's results about speed of convergence

For the case (a) above, M. Rosenblatt proved in [59] that the prediction error $\sigma_n^2(f)$ decreases to zero exponentially as $n \to \infty$. More precisely, M. Rosenblatt proved the following theorem.

Theorem 5.1 (Rosenblatt's first theorem)**.** *Let the spectral density f of a discrete-time stationary process $X(t)$ be positive and continuous on the segment $[\pi/2 - \alpha, \pi/2 + \alpha]$, $0 < \alpha < \pi$, and zero elsewhere. Then the prediction error $\sigma_n^2(f)$ approaches zero exponentially as $n \to \infty$. More precisely, the following asymptotic relation holds:*

$$\sigma_n^2(f) \simeq (\sin(\alpha/2))^{2n+1} \quad \text{as} \quad n \to \infty. \tag{5.1}$$

Thus, when the spectral density f is continuous and vanishes on an entire segment, then the prediction error $\sigma_n^2(f)$ approaches zero with a sufficiently high speed, namely as a geometric progression with common ratio $\sin^2(\alpha/2) < 1$. Notice that (5.1) implies that

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \sin(\alpha/2). \tag{5.2}$$

Concerning the case (b) above, for a specific deterministic process $X(t)$, Rosenblatt proved in [59] that the prediction error $\sigma_n^2(f)$ decreases to zero *like a power* as $n \to \infty$. More precisely, the deterministic process $X(t)$ considered in Rosenblatt [59] has the spectral density

$$f_a(\lambda) := \frac{e^{(2\lambda-\pi)\varphi(\lambda)}}{\cosh(\pi\varphi(\lambda))}, \quad f_a(-\lambda) = f_a(\lambda), \quad 0 \le \lambda \le \pi, \tag{5.3}$$

where $\varphi(\lambda) = (a/2)\cot\lambda$ and a is a positive parameter.

Using the technique of orthogonal polynomials on the unit circle and Szegő's results, Rosenblatt [59] proved the following theorem.

Theorem 5.2 (Rosenblatt's second theorem)**.** *Suppose that the process $X(t)$ has spectral density f_a given by (5.3). Then the following asymptotic relation for the prediction error $\sigma_n^2(f_a)$ holds:*

$$\sigma_n^2(f_a) \sim \frac{\Gamma^2\left((a+1)/2\right)}{\pi 2^{2-a}}\, n^{-a} \quad \text{as} \quad n \to \infty. \tag{5.4}$$

Note that the function in (5.3) was first considered by Pollaczek [51], and then by Szegő [69], as a weight-function of a class of orthogonal polynomials that serve as illustrations for certain'irregular' phenomena in the theory of orthogonal polynomials. For the function f_a in (5.3), we have the following asymptotic relation (for details see Szegő [69], Babayan and Ginovyan [5], and Section 5.5.4):

$$f_a(\lambda) \sim \begin{cases} 2e^a \exp\{-a\pi/|\lambda|\} & \text{as } \lambda \to 0, \\ 2\exp\{-a\pi/(\pi-|\lambda|)\} & \text{as } \lambda \to \pm\pi. \end{cases} \tag{5.5}$$

Thus, the function f_a in (5.3) has a very high order of contact with zero at points $\lambda = 0, \pm\pi$, due to which the process with spectral density f_a is deterministic and the prediction error $\sigma_n^2(f_a)$ in (5.4) decreases to zero like a power as $n \to \infty$.

Remark 5.1. In view of formulas in (3.15), under the conditions of Theorem 5.1, we have

$$\lim_{n\to\infty} \sigma^2_{n+1}(f)/\sigma^2_n(f) = \sin^2(\alpha/2) \quad \text{and} \quad \lim_{n\to\infty} |v_n(f)| = \cos(\alpha/2).$$

Similarly, under the conditions of Theorem 5.2, we have

$$\lim_{n\to\infty} \sigma_{n+1}(f_a)/\sigma_n(f_a) = 1 \quad \text{and} \quad \lim_{n\to\infty} v_n(f_a) = 0,$$

where $v_n(f)$ and $v_n(f_a)$ are the Verblunsky coefficients corresponding to functions f and f_a, respectively.

In the papers Babayan [2, 3], Babayan and Ginovyan [4, 5, 6], and Babayan et al. [7], the above stated Theorems 5.1 and 5.2 were extended to broader classes of spectral densities.

Concerning Theorem 5.1, in Babayan et al. [7] (see also Babayan [2, 3]) there was described an extension of the asymptotic relation (5.2) to the case of several arcs, without having to stipulate continuity of the spectral density f (see Theorem 5.3 in Section 5.3.1).

As for the extension of Theorem 5.2, in Babayan et al. [7] (see also Babayan and Ginovyan [4, 5]) it was proved that if the spectral density f is such that the sequence $\sigma_n(f)$ is weakly varying (a term defined in Section 7.2) and if, in addition, g is a nonnegative function that can have arbitrary power type singularities, then the sequences $\{\sigma_n(fg)\}$ and $\{\sigma_n(f)\}$ have the same asymptotic behavior as $n \to \infty$, up to some positive multiplicative factor. This allows to derive the asymptotic behavior of $\sigma_n(fg)$ from that of $\sigma_n(f)$.

Using this result, Rosenblatt's Theorem 5.2 was extended in Babayan and Ginovyan [5] and in Babayan et al. [7] to a class of spectral densities of the form $f = f_a g$, where f_a is as in (5.3) and g is a nonnegative function that can have arbitrary power type singularities (see Corollary 5.5 in Section 5.5.2).

5.3. Extensions of Rosenblatt's first theorem

In this section we extend Rosenblatt's first theorem (Theorem 5.1) to a broader class of deterministic processes, possessing spectral densities. More precisely, we extend the asymptotic relation (5.2) to the case of several arcs, without having to stipulate continuity of the spectral density f. Besides, we state necessary as well as sufficient conditions for the exponential decay of the prediction error $\sigma_n(f)$. Also, we calculate the transfinite diameters of some subsets of the unit circle, and thus, obtain explicit asymptotic relations for $\sigma_n(f)$ similar to Rosenblatt's relation (5.2).

5.3.1. Extensions of Theorem 5.1

In what follows, by E_f we denote the spectrum of the process $X(t)$, that is,

$$E_f := \{e^{i\lambda} : f(\lambda) > 0\}. \tag{5.6}$$

Thus, the closure $\overline{E}_f$ of E_f is the support of the spectral density f.

The next theorem extends Rosenblatt's first theorem (Theorem 5.1). More precisely, the result that follows extends the asymptotic relation (5.2) to the case of several arcs, without having to stipulate continuity of the spectral density f.

Theorem 5.3 (Babayan et al. [7]). *Let the support $\overline{E}_f$ of the spectral density f of the process $X(t)$ consists of a finite number of closed arcs of the unit circle $\mathbb{T}$, and let $f > 0$ a.e. on $\overline{E}_f$. Then the sequence $\sqrt[n]{\sigma_n(f)}$ converges, and*

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \tau_f, \tag{5.7}$$

where $\tau_f := \tau(\overline{E}_f)$ is the transfinite diameter of $\overline{E}_f$.

The definition and properties of the transfinite diameter and related metric characteristics (the Chebyshev constant and the capacity) are discussed later in the paper (see Section 7.1).

Remark 5.2. In Theorem 5.1, $\overline{E}_f = \{e^{i\lambda} : \lambda \in [\pi/2 - \alpha, \pi/2 + \alpha]\}$, which represents a closed arc of length 2α, and, according to Proposition 7.2(d), we have $\tau(\overline{E}_f) = \sin(2\alpha/4) = \sin(\alpha/2)$. Thus, the asymptotic relation (5.2) is a special case of (5.7).

We will need the following definition, which characterizes the rate of variation of a sequence of non-negative numbers compared with a geometric progression (see also Simon [65], p. 91).

Definition 5.1. *(a) A sequence $\{a_n \ge 0,\, n \in \mathbb{N}\}$ is called exponentially neutral if*

$$\lim_{n\to\infty} \sqrt[n]{a_n} = 1.$$

(b) A sequence $\{b_n \ge 0,\, n \in \mathbb{N}\}$ is called exponentially decreasing if

$$\limsup_{n\to\infty} \sqrt[n]{b_n} < 1.$$

For instance, the sequence $\{a_n = n^\alpha,\, \alpha \in \mathbb{R},\, n \in \mathbb{N}\}$ is exponentially neutral because $\ln \sqrt[n]{n^\alpha} = (\alpha/n) \ln n \to 0$ as $n \to \infty$. The geometric progression $\{b_n = q^n,\, 0 < q < 1,\, n \in \mathbb{N}\}$ is exponentially decreasing because $\sqrt[n]{b_n} = q^{n/n} = q < 1$. The sequence $\{b_n = n^\alpha q^n,\, \alpha \in \mathbb{R},\, 0 < q < 1,\, n \in \mathbb{N}\}$ is also exponentially decreasing because $\sqrt[n]{b_n} = n^{\alpha/n} q \to q < 1$. In fact, it can easily be shown that a sequence $\{c_n \ge 0,\, n \in \mathbb{N}\}$ is exponentially decreasing if and only if there exists a number q $(0 < q < 1)$ such that $c_n = O(q^n)$ as $n \to \infty$.

Remark 5.3. It follows from relation (5.7) that the question of exponential decay of the prediction error $\sigma_n(f)$ as $n \to \infty$ is determined solely by the value of the transfinite diameter of the support $\overline{E}_f$ of the spectral density f, and does not depend on the values of f on $\overline{E}_f$. Denoting $\gamma_n := \sigma_n(f)/\tau_f^n$, from (5.7) we infer that $\lim_{n\to\infty} \sqrt[n]{\gamma_n} = 1$ and

$$\sigma_n(f) = \tau_f^n \cdot \gamma_n. \tag{5.8}$$

Thus, in the case where $\tau_f < 1$, the prediction error $\sigma_n(f)$ is decomposed into a product of two factors, one of which (τ_f^n) is a geometric progression, and the second (γ_n) is an exponentially neutral sequence. Also, if g is another spectral density satisfying the conditions of Theorem 5.3, then in view of (5.8), we have

$$\frac{\sigma_n(g)}{\sigma_n(f)} = \left(\frac{\tau_g}{\tau_f}\right)^n \cdot \gamma_n',$$

where γ_n' is an exponentially neutral sequence. It is worth noting that the last relation does not depend on the structures of the supports $\overline{E}_f$ and $\overline{E}_g$ (viz., the number and the lengths of arcs constituting these sets, as well as, their location on the unit circle $\mathbb{T}$).

The following result provides a sufficient condition for the exponential decay of the prediction error $\sigma_n(f)$.

Theorem 5.4 (Babayan et al. [7]). *If the spectral density f of the process $X(t)$ vanishes on an arc, then the prediction error $\sigma_n(f)$ decreases to zero exponentially. More precisely, if f vanishes on an arc $\Gamma_\delta \subset \mathbb{T}$ of length 2δ $(0 < \delta < \pi)$, then*

$$\limsup_{n\to\infty} \sqrt[n]{\sigma_n(f)} \le \cos(\delta/2) < 1. \tag{5.9}$$

The next result gives a necessary condition for the exponential decay of $\sigma_n(f)$.

Theorem 5.5 (Babayan et al. [7]). *A necessary condition for the prediction error $\sigma_n(f)$ to tend to zero exponentially is that the spectral density f should vanish on a set of positive Lebesgue measure.*

Remark 5.4. Theorem 5.5 shows that if the spectral density f is almost everywhere positive, then it is impossible to obtain exponential decay of the prediction error $\sigma_n(f)$, no matter how high the orders of the zeros of f.

In view of (3.15), as a consequence of Theorem 5.3, we obtain the following result.

Theorem 5.6 (Babayan et al. [7]). *Let the support $\overline{E}_f$ and the spectral density f satisfy the conditions of Theorem 5.3. If the sequence of Verblunsky coefficients $v_n(f)$ converges in modulus, then*

$$\lim_{n\to\infty} |v_n(f)| = \sqrt{1-\tau_f^2}. \tag{5.10}$$

Remark 5.5. It is well-known that for an arbitrary sequence of positive numbers a_n the convergence $a_{n+1}/a_n \to a$ implies the convergence $\sqrt[n]{a_n} \to a$. The converse, in general, is not true, that is, the sequence $\sqrt[n]{a_n}$ can be convergent, while a_{n+1}/a_n divergent. Indeed, for the sequence a_n:

$$a_n := \begin{cases} 2^{-3k} & \text{if} \quad n = 2k-1 \\ 2^{-(3k+1)} & \text{if} \quad n = 2k, \end{cases} \qquad k \in \mathbb{N},$$

we have $\lim_{n\to\infty} \sqrt[n]{a_n} = 2^{-3/2}$, while the limit $\lim_{n\to\infty} a_{n+1}/a_n$ does not exists.

In the context of the considered sequences, $|v_n(f)|$ and $\sqrt[n]{\sigma_n(f)}$, in view of (3.15), we can assert that the convergence of $|v_n(f)|$ (or equivalently, the convergence of $\sigma_{n+1}(f)/\sigma_n(f)$) implies the convergence of $\sqrt[n]{\sigma_n(f)}$, but not the converse. Hence, the condition of convergence (in modulus) of Verblunsky sequence in Theorem 5.6 is essential.

As a consequence of Theorem 5.3 we obtain the following result (see Geronimus [21]), which is a partial converse of Rakhmanov's theorem:

Theorem 5.7. *If the sequence $\sigma_n(f)$ satisfies the following condition:*

$$\limsup_{n\to\infty} \sqrt[n]{\sigma_n(f)} = 1 \tag{5.11}$$

(in particular, if $\lim_{n\to\infty} v_n(f) = 0$), then $\overline{E}_f = \mathbb{T}$, i.e. the spectrum of the process is dense in $\mathbb{T}$.

The next theorem extends Theorem 5.3.

Theorem 5.8 (Babayan [2])**.** *Let E_f be the spectrum of a stationary process $X(t)$ possessing a spectral density $f(\lambda)$, that is, $E_f = \{\lambda : f(\lambda) > 0\}$, and let $\tau(E_f)$, $\tau_*(E_f)$ and $\tau^*(E_f)$ be the transfinite diameter and the inner and the outer transfinite diameters of E_f, respectively (for definition of $\tau_*(E_f)$ and $\tau^*(E_f)$ see formula (7.8)). Then the following assertions hold.*

(a) The following inequalities hold:

$$\limsup_{n\to\infty} (\sigma_n(f))^{1/n} \le \tau^*(E_f), \tag{5.12}$$

$$\liminf_{n\to\infty} (\sigma_n(f))^{1/n} \ge \tau_*(E_f). \tag{5.13}$$

(b) If the spectrum E_f consists of a countable number of open arcs of the unit circle $\mathbb{T}$ and is τ-measurable, that is, $\tau_*(E_f) = \tau^*(E_f) = \tau(E_f)$, then

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \tau(E_f). \tag{5.14}$$

As an immediate corollary of Theorem 5.8 we have the following result.

Corollary 5.1. *A sufficient condition for the prediction error $\sigma_n(f)$ of a deterministic stationary sequence to decrease to zero at least exponentially as $n \to \infty$, that is, $\sigma_n(f) = O(e^{-bn})$ for some $b > 0$, is that the outer transfinite diameter of the spectrum E_f should be less than unity.*

5.3.2. Examples. Calculation of transfinite diameters of some special sets

Motivated by Theorems 5.1 and 5.3 and Remark 5.2, the following question arises naturally: calculate the transfinite diameter $\tau(\overline{E}_f)$ of the set $\overline{E}_f$ consisting of several closed arcs of the unit circle $\mathbb{T}$, and thus, obtain an explicit asymptotic relation for the prediction error $\sigma_n(f)$ similar to Rosenblatt's relation (5.2).

The calculation of the transfinite diameter (and hence, the capacity and the Chebyshev constant) is a challenging problem (for details see Section 7.1), and in only very few cases has the transfinite diameter been exactly calculated (see, e.g., Landkof [48], p. 172–173, Ransford [57], p.135, and Proposition 7.2). One such example provides Theorem 5.1, in which case the transfinite diameter of the set $\overline{E}_f := \{e^{i\lambda} : \lambda \in [\pi/2-\alpha, \pi/2+\alpha]\}$ is $\sin(\alpha/2)$. Below we give some other examples, where we can explicitly calculate the Chebyshev constant (and hence the transfinite diameter and the capacity) by using some properties of the transfinite diameter, stated in Proposition 7.2, and results due to Fekete [19] and Robinson [58] concerning the relation between the transfinite diameters of related sets (see Propositions 7.3 and 7.5).

The examples given below show that Fekete's formula (7.9) and Robinson's formula (7.12) give a simple way to calculate the transfinite diameters of some subsets of the unit circle, based only on the formula of the transfinite diameter of a line segment (see Proposition 7.2(e)).

We will use the following notation: given $0 < \beta < 2\pi$ and $z_0 = e^{i\theta_0}$, $\theta_0 \in [-\pi, \pi)$, we denote by $\Gamma_\beta(\theta_0)$ an arc of the unit circle of length β which is symmetric with respect to the point $z_0 = e^{i\theta_0}$:

$$\Gamma_\beta(\theta_0) := \{e^{i\theta} : |\theta-\theta_0| \le \beta/2\} = \{e^{i\theta} : \theta \in [\theta_0-\beta/2, \theta_0+\beta/2]\}. \quad (5.15)$$

Example 5.1. Let $\Gamma_{2\alpha} := \Gamma_{2\alpha}(0)$. Then the projection $\Gamma^x_{2\alpha}$ of $\Gamma_{2\alpha}$ onto the real axis is the segment $[\cos\alpha, 1]$ (see Figure 1a)), and by Proposition 7.2(e) for the transfinite diameter $\tau(\Gamma^x_{2\alpha})$ we have

$$\tau(\Gamma^x_{2\alpha}) = \frac{1-\cos\alpha}{4} = \frac{\sin^2(\alpha/2)}{2}.$$

Hence, according to Robinson's formula (7.12), we obtain

$$\tau(\Gamma_{2\alpha}) = [2\tau(\Gamma^x_{2\alpha})]^{1/2} = \left[2\frac{\sin^2(\alpha/2)}{2}\right]^{1/2} = \sin(\alpha/2). \quad (5.16)$$

Taking into account that the transfinite diameter is invariant with respect to rotation (see Proposition 7.2(b)), from (5.16) for any $\theta_0 \in [-\pi, \pi)$ we have

$$\tau(\Gamma_{2\alpha}(\theta_0)) = \sin(\alpha/2). \quad (5.17)$$

Remark 5.6. Notice that the expression $\sin(\alpha/2)$ in (5.16) was first obtained by Szegő [68], where he calculated it as the Chebyshev constant of the arc $\Gamma_{2\alpha}(\pi/2)$, then it was deduced by Rosenblatt [59], as the capacity of $\Gamma_{2\alpha}(\pi/2)$.

Example 5.2. Let $\Gamma_{2\alpha}(\alpha)$ be an arc of length 2α, defined by (5.15): $\Gamma_{2\alpha}(\alpha) = \{e^{i\theta} : \theta \in [0, 2\alpha]\}$, and let $\Gamma(2,\alpha)$ be the preimage of the arc $\Gamma_{2\alpha}(\alpha)$ under the mapping $p(z) = z^2$. It can be shown (see Babayan et al. [7] for details) that the set $\Gamma(2,\alpha)$ is the union of two closed arcs of equal lengths α, symmetrically located with respect to the center of the unit circle (see Figure 1b):

$$\Gamma(2,\alpha) = \{e^{i\omega} : \omega \in [-\pi, -\pi+\alpha] \cup [0, \alpha]\}. \quad (5.18)$$

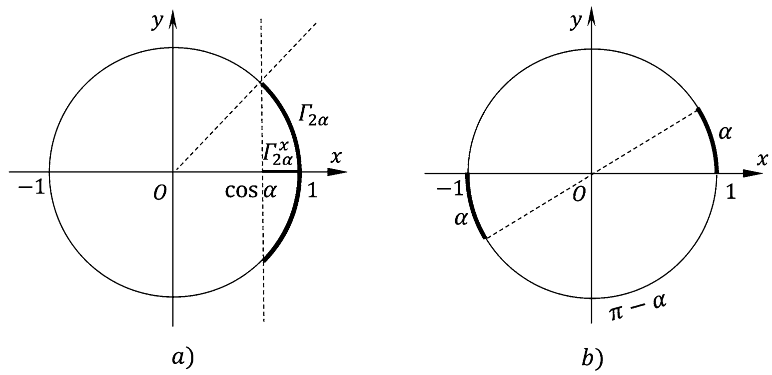

FIG 1. *a) The sets* $\Gamma_{2\alpha}$ *and* $\Gamma^x_{2\alpha}$. *b) The set* $\Gamma(k,\alpha)$ *with* $k=2$.

Then, by the Fekete theorem (see Proposition 7.3) and formula (5.17), for the transfinite diameter $\tau(\Gamma(2,\alpha))$ we have

$$\tau(\Gamma(2,\alpha)) = [\tau(\Gamma_{2\alpha}(\alpha))]^{1/2} = (\sin(\alpha/2))^{1/2}\,.$$

The above result can easily be extended to the case of k ($k>2$) arcs. Let $\Gamma(k,\alpha)$ be the union of k ($k\in\mathbb{N}$, $k\geq 2$) closed arcs of equal lengths α, which are symmetrically located on the unit circle (the arcs are assumed to be equidistant). It can be shown that the set $\Gamma(k,\alpha)$ is the preimage (to within rotation) under the mapping $p(z)=z^k$ of the arc $\Gamma_{k\alpha}(k\alpha/2)$ of length $k\alpha$ defined by (5.15). Therefore, by Fekete's formula (7.9) and the invariance property of the transfinite diameter with respect to rotation (see Proposition 7.2(b)), for the transfinite diameter $\tau(\Gamma(k,\alpha))$, we have

$$\tau(\Gamma(k,\alpha)) = (\sin(k\alpha/4))^{1/k}\,. \tag{5.19}$$

Example 5.3. Let $\alpha>0$, $\delta\geq 0$ and $\alpha+\delta\leq\pi$. Consider the set

$$\Gamma_{\alpha,\delta}(\theta_0) := \Gamma_{\alpha+\delta}(\theta_0)\setminus\Gamma_\delta(\theta_0) \tag{5.20}$$

consisting of the union of two arcs of the unit circle of lengths α, the distance of which (over the circle) is equal to 2δ. Define (see Fig. 2a)):

$$\Gamma_{\alpha,\delta} := \Gamma_{\alpha,\delta}(0) = \{e^{i\theta} : \theta\in[-(\delta+\alpha),-\delta]\cup[\delta,\delta+\alpha]\}. \tag{5.21}$$

Then the projection $\Gamma^x_{\alpha,\delta}$ of $\Gamma_{\alpha,\delta}$ onto the real axis is the segment $\Gamma^x_{\alpha,\delta}=[\cos(\alpha+\delta),\cos\delta]$, and by Proposition 7.2(e) for the transfinite diameter $\tau(\Gamma^x_{\alpha,\delta})$ we have

$$\tau(\Gamma^x_{\alpha,\delta}) = \frac{\cos\delta-\cos(\alpha+\delta)}{4} = \frac{\sin(\alpha/2)\sin(\alpha/2+\delta)}{2}.$$

Hence, according to Robinson's formula (7.12), for the transfinite diameter $\tau(\Gamma_{\alpha,\delta})$, we obtain

$$\tau(\Gamma_{\alpha,\delta}) = [2\tau(\Gamma^x_{\alpha,\delta})]^{1/2} = (\sin(\alpha/2)\sin(\alpha/2+\delta))^{1/2}\,. \tag{5.22}$$

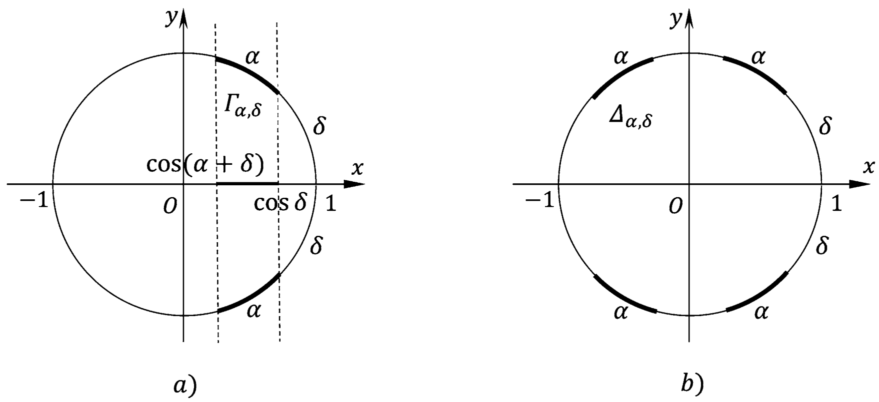

FIG 2. *a) The set* $\Gamma_{\alpha,\delta}$. *b) The set* $\Delta_{\alpha,\delta}$.

In view of Proposition 7.2(b), from (5.22) for any $\theta_0 \in [-\pi, \pi)$ we have

$$\tau(\Gamma_{\alpha,\delta}(\theta_0)) = (\sin(\alpha/2)\sin(\alpha/2+\delta))^{1/2}. \tag{5.23}$$

Observe that for $\delta = 0$ we have $\Gamma_{\alpha,\delta}(\theta_0) = \Gamma_{2\alpha}(\theta_0)$ (see (5.15) and (5.20)), and formula (5.23) becomes (5.17).

Example 5.4. Let the arc $\Gamma_{\alpha,\delta}$ be as in Example 5.3 (see (5.21)) with α, δ satisfying $\alpha + \delta \leq \pi/2$, that is, $\Gamma_{\alpha,\delta}$ is a subset of the right semicircle $\mathbb{T}$. Denote by $\Gamma'_{\alpha,\delta}$ the symmetric to $\Gamma_{\alpha,\delta}$ set with respect to y-axis, that is,

$$\Gamma'_{\alpha,\delta} := \{e^{i\theta} : \theta \in [-\pi+\delta, -\pi+(\delta+\alpha)] \cup [\pi-(\delta+\alpha), \pi-\delta]\}.$$

Define $\Delta_{\alpha,\delta} := \Gamma_{\alpha,\delta} \cup \Gamma'_{\alpha,\delta}$, and observe that the set $\Delta_{\alpha,\delta}$ consists of four arcs of equal lengths α, which are symmetrically located with respect to both axes (see Figure 2b)). It is easy to see that the set $\Delta_{\alpha,\delta}$ is the preimage (to within rotation) of the set $\Gamma_{2\alpha,2\delta}$ under the mapping $p(z) = z^2$. Hence, according to Fekete's formula (7.9) and (5.22), for the transfinite diameter $\tau(\Delta_{\alpha,\delta})$, we obtain

$$\tau(\Delta_{\alpha,\delta}) = (\tau(\Gamma_{2\alpha,2\delta}))^{1/2} = (\sin\alpha\sin(\alpha+2\delta))^{1/4}. \tag{5.24}$$

Denote by $\Delta_{\alpha,\delta}(\theta_0)$ the image of the set $\Delta_{\alpha,\delta}$ under mapping $q(z) = e^{i\theta_0}z$, that is, under the rotation by the central angle θ_0 around the origin. Then, in view of Proposition 7.2(b)), from (5.24) for any $\theta_0 \in [-\pi, \pi)$ we have

$$\tau(\Delta_{\alpha,\delta}(\theta_0)) = (\sin\alpha\sin(\alpha+2\delta))^{1/4}. \tag{5.25}$$

5.3.3. A consequence of Theorem 5.3

Now we apply Theorem 5.3 to obtain the asymptotic behavior of the prediction error $\sigma_n(f)$ in the cases where the spectrum of a stationary process $X(t)$ is as in Examples 5.1–5.4. Specifically, putting together Theorem 5.3 and Examples 5.1–5.4, we obtain the following result.

Theorem 5.9 (Babayan et al. [7]). *Let $\overline{E}_f$ be the support of the spectral density f of a stationary process $X(t)$, and let $f > 0$ a.e. on $\overline{E}_f$. Then for the prediction error $\sigma_n(f)$ the following assertions hold.*

(a) *If $\overline{E}_f = \Gamma_{2\alpha}(\theta_0)$, where $\Gamma_{2\alpha}(\theta_0)$ is as in Example 5.1, then*

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \sin(\alpha/2).$$

(b) *If $\overline{E}_f = \Gamma(k,\alpha)$, where $\Gamma(k,\alpha)$ is as in Example 5.2, then*

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \left(\sin(k\alpha/4)\right)^{1/k}.$$

(c) *If $\overline{E}_f = \Gamma_{\alpha,\delta}(\theta_0)$, where $\Gamma_{\alpha,\delta}(\theta_0)$ is as in Example 5.3, then*

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \left(\sin(\alpha/2)\sin(\alpha/2+\delta)\right)^{1/2}.$$

(d) *If $\overline{E}_f = \Delta_{\alpha,\delta}(\theta_0)$, where $\Delta_{\alpha,\delta}(\theta_0)$, is as in Example 5.4, then*

$$\lim_{n\to\infty} \sqrt[n]{\sigma_n(f)} = \left(\sin\alpha\sin(\alpha+2\delta)\right)^{1/4}.$$

Remark 5.7. The assertion (a) is a slight extension of the Rosenblatt relation (5.2). The assertion (c) is an extension of assertion (a), which reduces to assertion (a) if $\delta = 0$.

5.4. Davisson's theorem and its extension

In this section, we consider a question of bounding the prediction error $\sigma_n^2(f)$. Using constructive methods, Davisson [15] obtained an upper bound (rather than an asymptote) for the prediction error $\sigma_n^2(f)$ without imposing a continuity requirement on the spectral density $f(\lambda)$. Specifically, in Davisson [15] was proved the following result:

Theorem 5.10 (Davisson [15]). *Let the spectral density $f(\lambda)$, $\lambda \in [-\pi,\pi]$ of the process $X(t)$ be identically zero on a closed interval of length $2\pi - 2\alpha$, $0 < \alpha < \pi$. Then for the prediction error $\sigma_n^2(f)$ the following inequality holds:*

$$\sigma_n^2(f) \le 4c\left(\sin(\alpha/2)\right)^{2n-2}, \tag{5.26}$$

where $c = r(0)$ and $r(\cdot)$ is the covariance function of $X(t)$ (see formula (2.2)).

The theorem that follows, proved in Babayan and Ginovyan [6], extends Davisson's theorem to the case where the spectrum of the process $X(t)$ consists of a union of two equal arcs.

Let $\alpha > 0$, $\delta \ge 0$ and $\alpha + \delta \le \pi$, and let $\Gamma_{\alpha,\delta}$ be the set defined by (5.21). Recall that $\Gamma_{\alpha,\delta}$ is the union of two arcs of the unit circle of lengths α, the distance between which (over the circle) is equal to 2δ (see Example 5.3 and Figure 2a)).

Theorem 5.11 (Babayan and Ginovyan [6]). *Let the spectral density $f(\lambda)$, $\lambda \in [-\pi, \pi]$ of the process $X(t)$ vanish outside the set $\Gamma_{\alpha,\delta}$. Then for the prediction error $\sigma_n^2(f)$ the following inequality holds:*

$$\sigma_n^2(f) \leq 4c\,(\sin(\alpha/2))^{n-1}\,(\sin(\alpha/2+\delta))^{n-1}\,, \tag{5.27}$$

where c is as in Theorem 5.10.

Remark 5.8. For $\delta = 0$ the set $\Gamma_{\alpha,\delta}$ defined by (5.21) is an arc of length 2α, and, in this case, the inequality (5.27) becomes Davisson's inequality (5.26).

5.5. Extensions of Rosenblatt's second theorem

In this section, we analyze the asymptotic behavior of the prediction error in the case where the spectral density $f(\lambda)$, $\lambda \in [-\pi, \pi]$ of the model is strictly positive except one or several points at which it has a very high order contact with zero, so that the Szegő condition (2.23) is violated.

Based on Rosenblatt's result for this case, namely Theorem 5.2, we can expect that for any deterministic process with spectral density possessing a singularity of the type (5.5), the rate of the prediction error $\sigma_n^2(f)$ should be the same as in (5.4). However, the method applied in Rosenblatt [59] does not allow us to prove this assertion. In Babayan and Ginovyan [4, 5] and in Babayan et al. [7], using a different approach, Rosenblatt's second theorem was extended to broader classes of spectral densities. To state the corresponding results, we first examine the asymptotic behavior as $n \to \infty$ of the ratio:

$$\frac{\sigma_n^2(fg)}{\sigma_n^2(f)},$$

where g is a non-negative function.

To clarify the approach, we first assume that f is the spectral density of a nondeterministic process, in which case the geometric mean $G(f)$ is positive (see (2.21) and (2.22)). We can then write

$$\lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = \frac{\sigma_\infty^2(fg)}{\sigma_\infty^2(f)} = \frac{2\pi G(fg)}{2\pi G(f)} = \frac{G(f)G(g)}{G(f)} = G(g). \tag{5.28}$$

It turns out that under some additional assumptions imposed on functions f and g, the asymptotic relation (5.28) remains also valid in the case of deterministic processes, that is, when $G(f) = 0$.

5.5.1. Preliminaries

In what follows we consider the class of *deterministic* processes possessing spectral densities f for which the sequence of prediction errors $\{\sigma_n(f)\}$ is weakly varying (see Definition 7.1 in Section 7.2), and denote by $\mathcal{F}$ the class of the corresponding spectral densities:

$$\mathcal{F} := \left\{ f \in L^1(\Lambda) : \ f \geq 0, \ G(f) = 0, \ \lim_{n\to\infty} \frac{\sigma_{n+1}(f)}{\sigma_n(f)} = 1 \right\}. \tag{5.29}$$

Remark 5.9. According to Rakhmanov's theorem (Theorem 3.3), a sufficient condition for $f \in \mathcal{F}$ is that $f > 0$ almost everywhere on Λ and $G(f) = 0$. Thus, the considered class $\mathcal{F}$ includes all deterministic processes ($G(f) = 0$) with almost everywhere positive spectral densities ($f > 0$ a.e.). On the other hand, according to Theorem 3.2 and Remark 5.5, the class $\mathcal{F}$ does not contain spectral densities which vanish on an entire segment of Λ (or on an arc of the unit circle $\mathbb{T}$). Also, from Theorem 5.7 and Remark 5.5 we infer that a necessary condition for $f \in \mathcal{F}$ is that the spectrum E_f is dense in Λ.

Definition 5.2. *Let $\mathcal{F}$ be the class of spectral densities defined by (5.29). For $f \in \mathcal{F}$ denote by $\mathcal{M}_f$ the class of nonnegative functions $g(\lambda)$ ($\lambda \in \Lambda$) satisfying the conditions: $G(g) > 0$, $fg \in L^1(\Lambda)$, and*

$$\lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = G(g), \tag{5.30}$$

that is,

$$\mathcal{M}_f := \left\{ g \geq 0,\ G(g) > 0,\ fg \in L^1(\Lambda),\ \lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = G(g) \right\}. \tag{5.31}$$

The next proposition shows that the class $\mathcal{F}$ is closed under multiplication by functions from the class $\mathcal{M}_f$.

Proposition 5.1 (Babayan and Ginovyan [5])**.** *If $f \in \mathcal{F}$ and $g \in \mathcal{M}_f$, then $fg \in \mathcal{F}$.*

The next result shows that the class $\mathcal{M}_f$ in a certain sense is closed under multiplication.

Proposition 5.2 (Babayan and Ginovyan [5])**.** *Let $f \in \mathcal{F}$. If $g_1 \in \mathcal{M}_f$ and $g_2 \in \mathcal{M}_{fg_1}$, then $g := g_1 g_2 \in \mathcal{M}_f$ and $fg \in \mathcal{F}$. In particular, if $g \in \mathcal{M}_f \cap \mathcal{M}_{fg}$, then $g^2 \in \mathcal{M}_f$.*

In the next definition we introduce certain classes of bounded functions.

Definition 5.3. *We define the class B to be the set of all nonnegative, Riemann integrable on $\Lambda = [-\pi, \pi]$ functions $h(\lambda)$. Also, we define the following subclasses:*

$$B_+ := \{h \in B : h(\lambda) \geqslant m\},\ B^- := \{h \in B : h(\lambda) \leqslant M\},\ B_+^- := B_+ \cap B^-, \tag{5.32}$$

where m and M are some positive constants.

In the next proposition we list some obvious properties of the classes B_+, B^- and B_+^-.

Proposition 5.3. *The following assertions hold.*

a) If $h \in B_+(B^-)$, then $1/h \in B^-(B_+)$.
b) If $h_1, h_2 \in B_+(B^-)$, then $h_1 + h_2 \in B_+(B^-)$ and $h_1 h_2 \in B_+(B^-)$.
c) If $h_1, h_2 \in B^-$ and h_1/h_2 is bounded, then $h_1/h_2 \in B^-$.

d) If $h_1, h_2 \in B_+^-$, then $h_1 + h_2 \in B_+^-$, $h_1 h_2 \in B_+^-$ and $h_1/h_2 \in B_+^-$.

In the next proposition we list some properties of weakly varying sequences for functions from the above defined classes B_+, B^- and B_+^- (see Babayan et al. [7]).

Proposition 5.4. *Let the spectral density f be such that the sequence $\sigma_n(f)$ is weakly varying. The following assertions hold.*

a) If $g \in B_+^-$, then the sequence $\sigma_n(fg)$ is also weakly varying.
b) If $g \in B^-$ with $G(g) = 0$, then $\sigma_n(fg) = o(\sigma_n(f))$ as $n \to \infty$. Thus, multiplying singular spectral densities we obtain a spectral density with higher'order of singularity'.
c) If $g \in B_+$ with $G(g) = \infty$, and $fg \in B$, then $\sigma_n(f) = o(\sigma_n(fg))$ as $n \to \infty$.

5.5.2. Extensions of Rosenblatt's second theorem

The following theorem, proved in Babayan et al. [7], describes the asymptotic behavior of the ratio $\sigma_n^2(fg)/\sigma_n^2(f)$ as $n \to \infty$, and essentially states that if the spectral density f is from the class $\mathcal{F}$ (see (5.29)), and g is a nonnegative function, which can have *polynomial* type singularities, then the sequences $\{\sigma_n(fg)\}$ and $\{\sigma_n(f)\}$ have the same asymptotic behavior as $n \to \infty$ up to a positive numerical factor.

Theorem 5.12 (Babayan et al. [7])**.** *Let f be an arbitrary function from the class $\mathcal{F}$, and let g be a function of the form:*

$$g(\lambda) = h(\lambda) \cdot \frac{t_1(\lambda)}{t_2(\lambda)}, \quad \lambda \in \Lambda, \tag{5.33}$$

where $h \in B_+^-$, t_1 and t_2 are nonnegative trigonometric polynomials, such that $fg \in L^1(\Lambda)$. Then $g \in \mathcal{M}_f$ and $fg \in \mathcal{F}$, that is, fg is the spectral density of a deterministic process with weakly varying prediction error, and the relation (5.30) holds.

In view of Remark 5.9, as a consequence of Theorem 5.12 we obtain the following result.

Corollary 5.2. *Let the spectral density f of a deterministic process $X(t)$ be a.e. positive, and let g be as in Theorem 5.12. Then g is the spectral density of a nondeterministic process and the relation (5.30) holds.*

As an immediate consequence of Theorem 5.12 and Proposition 7.7(d), we have the following result.

Corollary 5.3. *Let the functions f and g be as in Theorem 5.12. Then the sequence $\sigma_n(fg)$ is also weakly varying.*

The theorems that follow extend the above stated Theorem 5.12 to a broader class of spectral densities, for which the function g can have *arbitrary power type singularities.*

Theorem 5.13 (Babayan and Ginovyan [5]). *Let f be an arbitrary function from the class $\mathcal{F}$, and let g be a function of the form:*

$$g(\lambda) = h(\lambda) \cdot |t(\lambda)|^{\alpha}, \quad \alpha > 0, \ \lambda \in \Lambda, \tag{5.34}$$

where $h \in B_+^-$ and t is an arbitrary trigonometric polynomial. Then $g \in \mathcal{M}_f$ and $fg \in \mathcal{F}$, that is, fg is the spectral density of a deterministic process with weakly varying prediction error, and the relation (5.30) holds.

Using inductive arguments and Theorem 5.13 we can state the following result.

Corollary 5.4. *The conclusion of Theorem 5.13 remains valid if the function g has the following form:*

$$g(\lambda) = h(\lambda) \cdot |t_1(\lambda)|^{\alpha_1} \cdot |t_2(\lambda)|^{\alpha_2} \cdot \dots \cdot |t_m(\lambda)|^{\alpha_m}, \quad \lambda \in \Lambda,$$

where $h \in B_+^-$, $t_1, t_2, \dots, t_m$ are arbitrary trigonometric polynomials, $\alpha_1, \alpha_2, \dots,$ α_m are arbitrary positive numbers, and $m \in \mathbb{N}$.

Theorem 5.14 (Babayan and Ginovyan [5]). *Let f be an arbitrary function from the class $\mathcal{F}$, and let g be a function of the form:*

$$g(\lambda) = h(\lambda) \cdot t^{-\alpha}(\lambda), \quad \alpha > 0, \ \lambda \in \Lambda, \tag{5.35}$$

where $h \in B_+^-$ and t is a nonnegative trigonometric polynomial. Then the following assertions hold.

(a) $g \in \mathcal{M}_f$ and $fg \in \mathcal{F}$ provided that $\alpha \in \mathbb{Z}$ and $ft^{-\alpha} \in L^1(\Lambda)$.
(b) $g \in \mathcal{M}_f$ and $fg \in \mathcal{F}$ provided that $\alpha \notin \mathbb{Z}$ and $ft^{-(k+1)} \in L^1(\Lambda)$, where $k := [\alpha]$ is the integer part of α.

To state the next result we need the following definition.

Definition 5.4. *Let E_1 and E_2 be two numerical sets such that for any $x \in E_1$ and $y \in E_2$ we have $x < y$. We say that the sets E_1 and E_2 are separated from each other if* $\sup E_1 < \inf E_2$. *Also, we say that a numerical set E is separated from infinity if it is bounded from above.*

Theorem 5.15 (Babayan and Ginovyan [5]). *Let $f(\lambda)$ and $\hat{f}(\lambda)$ $(\lambda \in \Lambda)$ be spectral densities of stationary processes satisfying the following conditions:*

1) $f, \hat{f} \in B^-$;
2) the functions $f(\lambda)$ and $\hat{f}(\lambda)$ have k common essential zeros

$$\lambda_1, \lambda_2, \dots, \lambda_k \in \Lambda \quad (-\pi < \lambda_1 < \lambda_2 < \dots < \lambda_k \leq \pi, \ k \in \mathbb{N}),$$

that is,

$$\lim_{\lambda \to \lambda_j} f(\lambda) = \lim_{\lambda \to \lambda_j} \hat{f}(\lambda) = 0, \quad j = 1, 2, \dots, k; \tag{5.36}$$

3) the functions $f(\lambda)$ and $\hat{f}(\lambda)$ are infinitesimal of the same order in a neighborhood of each point λ_j $(j = 1, 2, \ldots, k)$, that is,

$$\lim_{\lambda\to\lambda_j} \frac{\hat{f}(\lambda)}{f(\lambda)} = c_j > 0, \quad j = 1, 2, \ldots, k; \tag{5.37}$$

4) the functions $f(\lambda)$ and $\hat{f}(\lambda)$ are bounded away from zero outside any neighborhood $O_\delta(\lambda_j)$ $(j = 1, 2, \ldots, k)$, which is separated from the neighboring zeros λ_{j-1} and λ_{j+1} of λ_j, that is, there is a number $m := m_\delta > 0$ such that $f(\lambda) \geq m$ and $\hat{f}(\lambda) \geq m$ for almost all $\lambda \notin \cup_{j=1}^k O_\delta(\lambda_j)$. Then the following assertions hold:

a)

$$h(\lambda) := \frac{\hat{f}(\lambda)}{f(\lambda)} \in B_+^-;$$

b) the processes with spectral densities f and $\hat{f}$ either both are deterministic or both are nondeterministic;

c) if one of the functions f and $\hat{f}$ is from the class $\mathcal{F}$, then so is the other, and the following relation holds:

$$\lim_{n\to\infty} \frac{\sigma_n^2(\hat{f})}{\sigma_n^2(f)} = G(h) > 0. \tag{5.38}$$

Remark 5.10. The conditions of Theorem 5.15 mean that the points λ_j $(j = 1, 2, \ldots, k)$ are the only common zeros of functions $f(\lambda)$ and $\hat{f}(\lambda)$. Besides, in the case of deterministic processes, at least one of these zeros should be of sufficiently high order. Also, notice that the conditions 1) and 4) of Theorem 5.15 will be satisfied if the functions $f(\lambda)$ and $\hat{f}(\lambda)$ are continuous on Λ.

Theorem 5.16 (Babayan and Ginovyan [5])**.** *Let f be an arbitrary function from the class $\mathcal{F}$, and let g be a function of the form:*

$$g(\lambda) = h(\lambda) \cdot |q(\lambda)|^\alpha, \quad \alpha \in \mathbb{R},\ \lambda \in \Lambda, \tag{5.39}$$

where $h \in B_+^-$, q is an arbitrary algebraic polynomial with real coefficients, and $fg \in L^1(\Lambda)$. Then $fg \in \mathcal{F}$ and $g \in \mathcal{M}_f$.

Taking into account that the sequence $\{n^{-\alpha},\ n \in \mathbb{N},\ \alpha > 0\}$ is weakly varying, as an immediate consequence of Theorems 5.13, 5.14, 5.16 or Corollary 5.4, we have the following result (see Babayan and Ginovyan [5], and Babayan et al. [7]).

Corollary 5.5. *Let the functions f and g satisfy the conditions of one of Theorems 5.13, 5.14, 5.16 or Corollary 5.4, and let $\sigma_n(f) \sim cn^{-\alpha}$ $(c > 0, \alpha > 0)$ as $n \to \infty$. Then*

$$\sigma_n(fg) \sim cG(g)n^{-\alpha} \quad \text{as} \quad n \to \infty,$$

where $G(g)$ is the geometric mean of g.

The next result, which immediately follows from Theorem 5.2 and Corollary 5.5, extends Rosenblatt's second theorem (Theorem 5.2) (see Babayan and Ginovyan [5], and Babayan et al. [7]).

Theorem 5.17 (Babayan and Ginovyan [5])**.** *Let $f = f_a g$, where f_a is defined by (5.3), and let g be a function satisfying the conditions of one of Theorems 5.13, 5.14, 5.16 or Corollary 5.4. Then*

$$\delta_n(f) = \sigma_n^2(f) \sim \frac{\Gamma^2\left(\frac{a+1}{2}\right) G(g)}{\pi 2^{2-a}}\; n^{-a} \quad \text{as} \quad n \to \infty,$$

where $G(g)$ is the geometric mean of g.

We thus have the same limiting behavior for $\sigma_n^2(f)$ as in the Rosenblatt's relation (5.4) up to an additional positive factor $G(g)$.

Remark 5.11. In view of Remark 5.9 it follows that all the above stated results remain true if the condition $f \in \mathcal{F}$ is replaced by the following slightly strong but more constructive condition: '*the spectral density f is positive ($f > 0$) almost everywhere on Λ and $G(f) = 0$*'.

5.5.3. Examples

In this section we discuss examples demonstrating the result stated in Section 5.5.2. In these examples we assume that $\{X(t),\ t \in \mathbb{Z}\}$ is a stationary deterministic process with a spectral density f satisfying the conditions of Theorem 5.12, and the function g is given by formula (5.33). To compute the geometric means we use the properties stated in Proposition 7.8(a).

Example 5.5. Let the function $g(\lambda)$ be as in (5.33) with $h(\lambda) = c > 0$ and $t_1(\lambda) = t_2(\lambda) = 1$, that is, $g(\lambda) = c > 0$. Then for the geometric mean $G(g)$ we have

$$G(g) = G(c) = c, \tag{5.40}$$

and in view of (5.30), we get

$$\lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = G(g) = c.$$

Thus, multiplying the spectral density f by a constant $c > 0$ multiplies the prediction error by c.

Example 5.6. Let the function g be as in (5.33) with $h(\lambda) = e^{\varphi(\lambda)}$, where $\varphi(\lambda)$ is an arbitrary odd function, and let $t_1(\lambda) = t_2(\lambda) = 1$, that is, $g(\lambda) = e^{\varphi(\lambda)}$. Then for the geometric mean $G(g)$ we have

$$\begin{aligned} G(g) = G(e^{\varphi(\lambda)}) &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln g(\lambda)\, d\lambda\right\} \\ &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi(\lambda)\, d\lambda\right\} = e^0 = 1, \end{aligned} \tag{5.41}$$

and in view of (5.30), we get

$$\lim_{n\to\infty}\frac{\sigma_n^2(fg)}{\sigma_n^2(f)}=G(g)=1.$$

Thus, multiplying the spectral density f by the function $e^{\varphi(\lambda)}$ with odd $\varphi(\lambda)$ does not change the asymptotic behavior of the prediction error.

Example 5.7. Let the function g be as in (5.33) with $h(\lambda)=\lambda^2+1$ and $t_1(\lambda)=t_2(\lambda)=1$, that is, $g(\lambda)=\lambda^2+1$. Then for the geometric mean $G(g)$ by direct calculation we obtain

$$\begin{aligned}G(g)&=\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln(\lambda^2+1)\,d\lambda\right\}\\&=\exp\{\ln(1+\pi^2)-2+\frac{2}{\pi}\arctan\pi\}\approx 3.3,\end{aligned}\tag{5.42}$$

and in view of (5.30), we get

$$\lim_{n\to\infty}\frac{\sigma_n^2(fg)}{\sigma_n^2(f)}=G(g)=\exp\{\ln(1+\pi^2)-2+\frac{2}{\pi}\arctan\pi\}\approx 3.3.$$

Thus, multiplying the spectral density f by the function λ^2+1 multiplies the prediction error approximately by 3.3.

Example 5.8. Let the function g be as in (5.33) with $h(\lambda)=t_2(\lambda)=1$, and $t_1(\lambda)=\sin^{2k}(\lambda-\lambda_0)$, where $k\in\mathbb{N}$ and λ_0 is an arbitrary point from $[-\pi,\pi]$, that is, $g(\lambda)=\sin^{2k}(\lambda-\lambda_0)$. To compute the geometric mean $G(g)$, we first find the algebraic polynomial $s_2(z)$ in the Fejér-Riesz representation (7.15) of the non-negative trigonometric polynomial $\sin^2(\lambda-\lambda_0)$ of degree 2. For any $\lambda_0\in[-\pi,\pi]$ we have

$$\sin^2(\lambda-\lambda_0)=|\sin(\lambda-\lambda_0)|^2=\left|\frac{1}{2}(e^{2i(\lambda-\lambda_0)}-1)\right|^2=\left|s_2(e^{i\lambda})\right|^2,$$

where

$$s_2(z)=\frac{1}{2}(e^{-2i\lambda_0}z^2-1).\tag{5.43}$$

Therefore, by Proposition 7.8(d) and (5.43), we have

$$G(\sin^2(\lambda-\lambda_0))=|s_2(0)|^2=\left(\frac{1}{2}\right)^2=\frac{1}{4}.\tag{5.44}$$

Now, in view of Proposition 7.8(a) and (5.44), for the geometric mean of $g(\lambda)=t_1(\lambda)=\sin^{2k}(\lambda-\lambda_0)$ $(k\in\mathbb{N})$, we obtain

$$G(g)=G(\sin^{2k}(\lambda-\lambda_0))=G^k(\sin^2(\lambda-\lambda_0))=4^{-k},\tag{5.45}$$

and in view of (5.30), we get

$$\lim_{n\to\infty}\frac{\sigma_n^2(fg)}{\sigma_n^2(f)}=G(g)=\frac{1}{4^k}.$$

Thus, multiplying the spectral density f by the non-negative trigonometric polynomial $\sin^{2k}(\lambda-\lambda_0)$ of degree $2k$ ($k\in\mathbb{N}$), yields a 4^k-fold asymptotic reduction of the prediction error.

Example 5.9. Let the function g be as in (5.33) with $h(\lambda)=t_1(\lambda)=1$, and $t_2(\lambda)=\sin^{2l}(\lambda-\lambda_0)$, where $l\in\mathbb{N}$ and λ_0 is an arbitrary point from $[-\pi,\pi]$, that is, $g(\lambda)=\sin^{-2l}(\lambda-\lambda_0)$. Then, in view of the third equality in (7.14) and (5.45) for the geometric mean $G(g)$ we have

$$G(g)=G(\sin^{-2l}(\lambda-\lambda_0))=G^{-1}(\sin^{2l}(\lambda-\lambda_0))=4^l, \tag{5.46}$$

and in view of (5.30), we get

$$\lim_{n\to\infty}\frac{\sigma_n^2(fg)}{\sigma_n^2(f)}=G(g)=4^l.$$

Thus, dividing the spectral density f by the non-negative trigonometric polynomial $\sin^{2l}(\lambda-\lambda_0)$ of degree $2l$ ($l\in\mathbb{N}$), yields a 4^l-fold asymptotic increase of the prediction error.

Notice that the values of the geometric mean $G(g)$ obtained in (5.45) and (5.46) do not depend on the choice of the point $\lambda_0\in[-\pi,\pi]$.

Putting together Examples 5.5–5.9 and using Proposition 7.8(a) we have the following summary example.

Example 5.10. Let $\{X(t),\ t\in\mathbb{Z}\}$ be a stationary deterministic process with a spectral density f satisfying the conditions of Theorem 5.12. Let $h(\lambda)=ce^{\varphi(\lambda)}(\lambda^2+1)$, $t_1(\lambda)=\sin^{2k}(\lambda-\lambda_1)$ and $t_2(\lambda)=\sin^{2l}(\lambda-\lambda_2)$, where c is an arbitrary positive constant, $\varphi(\lambda)$ is an arbitrary odd function and λ_1,λ_2 are arbitrary points from $[-\pi,\pi]$. Let the function g be defined as in (5.33), that is,

$$g(\lambda)=h(\lambda)\cdot\frac{t_1(\lambda)}{t_2(\lambda)}=ce^{\varphi(\lambda)}(\lambda^2+1)\frac{\sin^{2k}(\lambda-\lambda_1)}{\sin^{2l}(\lambda-\lambda_2)}. \tag{5.47}$$

Then, in view of Proposition 7.8(a) and relations (5.40)–(5.42) and (5.45)–(5.47), we have

$$\begin{aligned}G(g)&=G(h)\frac{G(t_1)}{G(t_2)}=G(c)G(e^{\varphi})G(\lambda^2+1)G(\sin^{2k}(\lambda-\lambda_1))G(\sin^{-2l}(\lambda-\lambda_2))\\&=(c)(1)\exp\{\ln(1+\pi^2)-2+\frac{2}{\pi}\arctan\pi\}(4^{-k})(4^l)\approx 3.3c4^{l-k},\end{aligned} \tag{5.48}$$

and in view of (5.30) and (5.48), we get

$$\lim_{n\to\infty}\frac{\sigma_n^2(fg)}{\sigma_n^2(f)}=G(g)\approx 3.3c4^{l-k}.$$

Example 5.11. Let the function $g(\lambda)$ ($\lambda\in\Lambda$) be as in (5.34) with $h(\lambda)=1$ and $t(\lambda)=\sin(\lambda-\lambda_0)$, where λ_0 is an arbitrary point from $[-\pi,\pi]$, that is,

$g(\lambda) = |\sin(\lambda - \lambda_0)|^\alpha$, $\alpha > 0$. Then, according to Example 5.8, for the geometric mean of $\sin^2(\lambda - \lambda_0)$ we have

$$G(\sin^2(\lambda - \lambda_0)) = \frac{1}{4}. \tag{5.49}$$

According to Proposition 7.8(a) and (5.49), for the geometric mean of $g(\lambda)$, we obtain

$$\begin{aligned} G(g) &= G(|\sin(\lambda - \lambda_0)|^\alpha) = G\left(\left(\sin^2(\lambda - \lambda_0)\right)^{\alpha/2}\right) \\ &= G^{\alpha/2}(\sin^2(\lambda - \lambda_0)) = \frac{1}{2^\alpha}, \end{aligned} \tag{5.50}$$

and in view of (5.30), we get

$$\lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = G(g) = \frac{1}{2^\alpha}.$$

Thus, multiplying the spectral density $f(\lambda)$ by the function $g(\lambda) = |\sin(\lambda - \lambda_0)|^\alpha$ yields a 2^α-fold asymptotic reduction of the prediction error.

Example 5.12. Let the function $g(\lambda)$ be as in (5.39) with $h(\lambda) = 1$ and $q(\lambda) = \lambda$, that is, $g(\lambda) = |\lambda|^\alpha$, $\alpha \in \mathbb{R}$. By direct calculation we obtain

$$\ln G(g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |\lambda|^\alpha \, d\lambda = \frac{\alpha}{\pi} \int_0^\pi \ln \lambda \, d\lambda = \alpha \ln(\pi/e).$$

Therefore

$$G(g) = (\pi/e)^\alpha \approx (1.156)^\alpha,$$

and in view of (5.30), we get

$$\lim_{n\to\infty} \frac{\sigma_n^2(fg)}{\sigma_n^2(f)} = G(g) = \left(\frac{\pi}{e}\right)^\alpha \approx (1.156)^\alpha.$$

Thus, multiplying the spectral density $f(\lambda)$ by the function $g(\lambda) = |\lambda|^\alpha$ multiplies the prediction error asymptotically by $(\pi/e)^\alpha \approx (1.156)^\alpha$.

It follows from Proposition 3.1(d) that the same asymptotic is true for the prediction error with spectral density $\bar{g}(\lambda) = |\lambda - \lambda_0|^\alpha$, $\lambda_0 \in [-\pi, \pi]$.

5.5.4. Rosenblatt's second theorem revisited

We first analyze the Pollaczek-Szegő function $f_a(\lambda)$ given by (5.3) (cf. Pollaczek [51] and Szegő [69]). We have

$$f_a(\lambda) = \frac{2e^{2\lambda\varphi(\lambda)}e^{-\pi\varphi(\lambda)}}{e^{\pi\varphi(\lambda)} + e^{-\pi\varphi(\lambda)}} = \frac{2e^{2\lambda\varphi(\lambda)}}{e^{2\pi\varphi(\lambda)} + 1}, \quad 0 \le \lambda \le \pi, \tag{5.51}$$

where $\varphi(\lambda) := \varphi_a(\lambda) = (a/2)\cot\lambda$. Observe that $\varphi(\lambda) \to +\infty$ as $\lambda \to 0^+$, and we have

$$\varphi(\lambda) \sim a/(2\lambda), \quad e^{2\lambda\varphi(\lambda)} \sim e^a, \quad e^{2\pi\varphi(\lambda)} + 1 \sim e^{a\pi/\lambda} \quad \text{as} \quad \lambda \to 0^+. \tag{5.52}$$

Taking into account that $f_a(\lambda)$ is an even function, from (5.51) and (5.52) we obtain the following asymptotic relation for $f_a(\lambda)$ in a vicinity of the point $\lambda = 0$.

$$f_a(\lambda) \sim 2e^a \exp\{-a\pi/|\lambda|\} \quad \text{as} \quad \lambda \to 0. \tag{5.53}$$

Next, observe that $\varphi(\lambda) \to -\infty$ as $\lambda \to \pi$, and we have as $\lambda \to \pi$

$$\varphi(\lambda) = -\varphi(\pi - \lambda) \sim (-a/2)(\pi - \lambda), \quad 2\lambda\varphi(\lambda) \sim -a\pi/(\pi - \lambda). \tag{5.54}$$

In view of (5.51) and (5.54) we obtain the following asymptotic of the function $f_a(\lambda)$ in a vicinity of the point $\lambda = \pi$.

$$f_a(\lambda) \sim 2e^{2\lambda\varphi(\lambda)} \sim 2\exp\{-a\pi/(\pi - \lambda)\} \quad \text{as} \quad \lambda \to \pi. \tag{5.55}$$

Putting together (5.53) and (5.55), and taking into account evenness of $f_a(\lambda)$, we conclude that

$$f_a(\lambda) \sim \begin{cases} 2e^a \exp\{-a\pi/|\lambda|\} & \text{as } \lambda \to 0, \\ 2\exp\{-a\pi/(\pi - |\lambda|)\} & \text{as } \lambda \to \pm\pi, \end{cases} \tag{5.56}$$

Thus, the function $f_a(\lambda)$ is positive everywhere except for points $\lambda = 0, \pm\pi$, and has a very high order of contact with zero at these points, so that Szegő's condition (2.23) is violated implying that $G(f_a) = 0$. Also, observe that $f_a(\lambda)$ is infinitely differentiable at all points of the segment $[-\pi, \pi]$ including the points $\lambda = 0, \pm\pi$, and attains it maximum value of 1 at the points $\pm\pi/2$. For some specific values of the parameter a the graph of the function $f_a(\lambda)$ is represented in Figure 3a).

For $a > 0$ and $\lambda \in [-\pi, \pi]$, consider the pair of functions $\hat{f}_1(\lambda)$ and $\hat{f}_2(\lambda)$ defined by formulas:

$$\hat{f}_1(\lambda) := \exp\{-a\pi/|\lambda|\}, \quad \hat{f}_2(\lambda) := \exp\{-a\pi/(\pi - |\lambda|)\}. \tag{5.57}$$

Observe that the function $\hat{f}_1(\lambda)$ is positive everywhere except for point $\lambda = 0$ at which it has the same order of contact with zero as $f_a(\lambda)$, and hence $G(\hat{f}_1) = 0$. Also, $\hat{f}_1(\lambda)$ is infinitely differentiable at all points of the segment $[-\pi, \pi]$ except for the points $\lambda = \pm\pi$, where it attains its maximum value equal to e^{-a}. As for the function $\hat{f}_2(\lambda)$, it is positive everywhere except for points $\lambda = \pm\pi$, at which it has the same order of contact with zero as $f_a(\lambda)$, and hence $G(\hat{f}_2) = 0$. Also, $\hat{f}_2(\lambda)$ is infinitely differentiable at all points of the segment $[-\pi, \pi]$ except for the point $\lambda = 0$, where it attains its maximum value equal to e^{-a}. For some specific values of the parameter a the graphs of functions $\hat{f}_1(\lambda)$ and $\hat{f}_2(\lambda)$ are represented in Figure 4.

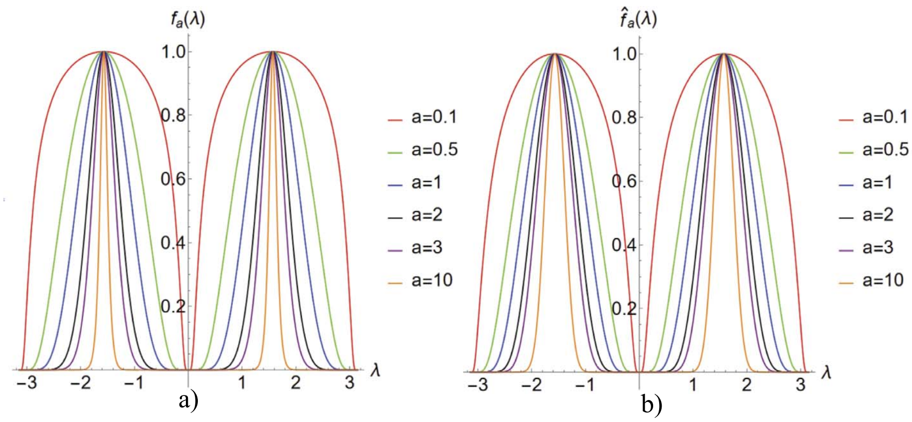

FIG 3. *a) Graph of the function $f_a(\lambda)$. b) Graph of the function $\hat{f}_a(\lambda)$.*

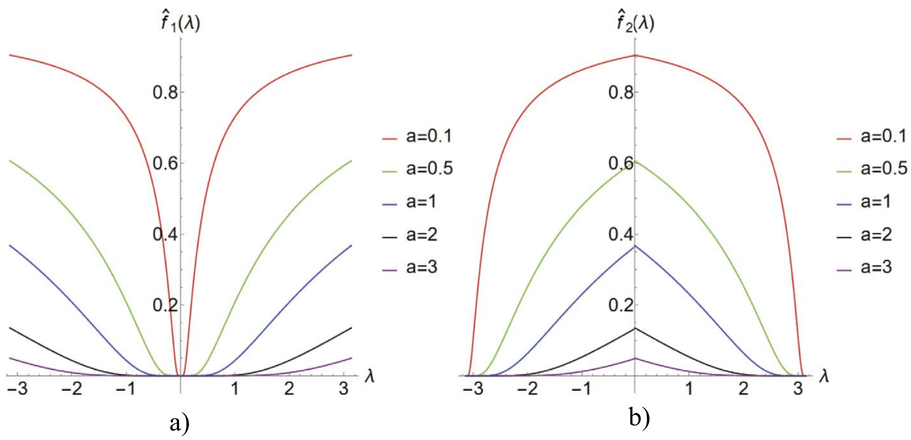

FIG 4. *a) Graph of the function $\hat{f}_1(\lambda)$. b) Graph of the function $\hat{f}_2(\lambda)$.*

Denote by $\hat{f}_a(\lambda)$ the product of functions $\hat{f}_1(\lambda)$ and $\hat{f}_2(\lambda)$ defined in (5.57) and normalized by the factor e^{4a}:

$$\hat{f}_a(\lambda) := e^{4a}\hat{f}_1(\lambda)\hat{f}_2(\lambda) = e^{4a}\exp\left\{-a\pi^2/(|\lambda|(\pi - |\lambda|))\right\}, \tag{5.58}$$

and observe that $\hat{f}_a(\lambda)$ behaves similar to $f_a(\lambda)$. Indeed, the function $\hat{f}_a(\lambda)$ also is positive everywhere except for points $\lambda = 0, \pm\pi$, it is infinitely differentiable at all points of the segment $[-\pi, \pi]$ including the points $\lambda = 0, \pm\pi$, and attains it maximum value of 1 at the points $\pm\pi/2$. Also, in view of (5.56) and (5.58), at points $\lambda = 0, \pm\pi$ the function $\hat{f}_a(\lambda)$ has the same order of zeros as $f_a(\lambda)$, and hence $G(\hat{f}_a) = 0$. Thus, the process $X(t)$ with spectral density $\hat{f}_a(\lambda)$ is deterministic. For some specific values of the parameter a the graph of the function $\hat{f}_a(\lambda)$ is represented in Figure 3b).

TABLE 1
The values of constants $\hat{C}(a)$ and $C(a)$

a	$\frac{\Gamma^2((a+1)/2)}{\pi 2^{2-a}}$	$\hat{C}(a)$	$C(a)$
0.1	0.223	0.797	0.178
0.5	0.169	1.113	0.188
1.0	0.159	2.545	0.406
1.5	0.185	6.446	1.193
2.0	0.250	16.830	4.214
3.0	0.637	119.220	76.379
3.3	0.902	215.715	194.656
3.4	1.020	263.173	268.375
5.0	10.186	6128.990	62429.000
10.0	223256	$1.104 \cdot 10^8$	$2.428 \cdot 10^{13}$

The functions $f_a(\lambda)$ and $\hat{f}_a(\lambda)$ defined by (5.3) and (5.58), respectively, satisfy the conditions of Theorem 5.15. Therefore, we have (see (5.38))

$$\lim_{n\to\infty} \frac{\sigma_n^2(\hat{f}_a)}{\sigma_n^2(f_a)} = G(\hat{f}_a/f_a) := \hat{C}(a) > 0. \tag{5.59}$$

In view of (5.4) and (5.59) we have

$$\sigma_n^2(\hat{f}_a) \sim C(a) \cdot n^{-a} \quad \text{as} \quad n \to \infty. \tag{5.60}$$

where

$$C(a) := \frac{\Gamma^2\left((a+1)/2\right) \hat{C}(a)}{\pi 2^{2-a}}. \tag{5.61}$$

The values of the constants $\hat{C}(a)$ and $C(a)$ for some specific values of the parameter a are given in Table 1.

Now we compare the prediction errors $\sigma_n^2(\hat{f}_1)$ and $\sigma_n^2(\hat{f}_2)$ with $\sigma_n^2(f_a)$. To this end, observe first that the function $g_1(\lambda) := f_a(\lambda)/\hat{f}_1(\lambda)$ has a very high order of contact with zero at points $\lambda = \pm\pi$, so that Szegő's condition (2.23) is violated, implying that $G(g_1) = 0$. Besides, the function $g_1(\lambda)$ is continuous on $[-\pi, \pi]$, and hence $g_1 \in B^-$. Therefore, according to Proposition 5.4 b), we have

$$\sigma_n^2(f_a) = o\left(\sigma_n^2(\hat{f}_1)\right) \quad \text{as} \quad n \to \infty. \tag{5.62}$$

Similar arguments applied to the function $f_2(\lambda)$ yield

$$\sigma_n^2(f_a) = o\left(\sigma_n^2(\hat{f}_2)\right) \quad \text{as} \quad n \to \infty. \tag{5.63}$$

The relations (5.62) and (5.63) show that the rate of convergence to zero of the prediction errors $\sigma_n^2(\hat{f}_1)$ and $\sigma_n^2(\hat{f}_2)$ is less than the one for $\sigma_n^2(f_a)$, that is, the power rate of convergence n^{-a} (see (5.60)). Thus, the rate of convergence n^{-a} is due to the joint contribution of all zeros $\lambda = 0, \pm\pi$ of the function $f_a(\lambda)$, whereas none of these zeros separately guarantees the rate of convergence n^{-a}.

6. An Application. Asymptotic behavior of the extreme eigenvalues of truncated Toeplitz matrices

In this section we analyze the relationship between the rate of convergence to zero of the prediction error $\sigma_n^2(f)$ and the minimal eigenvalue of a truncated Toeplitz matrix generated by the spectral density f, by showing how it is possible to obtain information in both directions.

The problem of asymptotic behavior of the extreme eigenvalues of truncated (finite sections) Toeplitz matrices goes back to the classical works by Kac, Murdoch and Szegő [42], Parter [49], Widom [74], and Chan [13], where the problem was studied for truncated Toeplitz matrices generated by continuous and continuously differentiable functions (symbols). Since then the problem for various classes of symbols was studied by many authors. For instance, Pourahmadi [52], Serra [62, 63], and Babayan and Ginovyan [6] considered the problem in the case where the symbol of Toeplitz matrix is not (necessarily) continuous nor differentiable (see also Böttcher and Grudsky [11]). In this section, we review and summarize some known results from the above cited references and state some new results.

6.1. Extreme eigenvalues of truncated Toeplitz matrices

Let $f(\lambda)$ be a real-valued Lebesgue integrable function defined on $\Lambda := [-\pi, \pi]$, $T_n(f) := ||r_{k-j}||$, $j, k = 0, 1, ..., n$, be the truncated Toeplitz matrix generated by the Fourier coefficients of f, and let $\lambda_{1,n}(f) \leq \lambda_{2,n}(f) \leq \cdots \lambda_{n+1,n}(f)$ be the eigenvalues of $T_n(f)$. We denote by $m_f := \text{ess inf} f$ and $M_f := \text{ess sup} f$ the essential minimum and the essential maximum of f, respectively. We will refer to $f(\lambda)$ as a *symbol* for the Toeplitz matrix $T_n(f)$.

We first recall Szegő's distribution theorem (see, e.g., Grenander and Szegő [31], p. 64–65).

Theorem 6.1. *For every continuous function F defined in $[m_f, M_f]$ the following asymptotic relation holds:*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=1}^{n+1} F(\lambda_{k,n}(f)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(f(u))\, du. \tag{6.1}$$

Moreover, the spectrum of $T_n(f)$ is contained in (m_f, M_f), and

$$\lim_{n\to\infty} \lambda_{1,n}(f) = m_f \quad \text{and} \quad \lim_{n\to\infty} \lambda_{n+1,n}(f) = M_f. \tag{6.2}$$

The problem of interest is to describe the rate of convergence in (6.2), depending on the properties of the symbol f. In the following, without loss of generality, we assume that $m_f := \text{ess inf} f = 0$. Also, we study the asymptotic behavior of the minimum eigenvalue of $T_n(f)$, for the maximum eigenvalue it is sufficient to consider the minimum eigenvalue of the matrix $T_n(-f)$.

The rate of convergence of extreme eigenvalues has been studied by Kac, Murdoch and Szegő [42], Parter [49], Widom [74] and Chan [13], under the following regularity condition on f (see, e.g., Grenander and Szegő [31], Section 5.4(a), p. 72).

Condition A. Let $f(\lambda)$, $\lambda \in [-\pi, \pi]$ be real and periodic with period 2π, satisfying the conditions:

(**A1**) f has a zero at $\lambda = \lambda_0$ of order $2k$ ($k \in \mathbb{N}$), that is, $f(\lambda) \asymp (\lambda - \lambda_0)^{2k}$;

(**A2**) f is continuous and has continuous derivatives of order $2k$ in a neighborhood of $\lambda = \lambda_0$ with $f^{(2k)}(\lambda_0) \neq 0$;

(**A3**) $\lambda = \lambda_0$ is the unique zero of f in $[-\pi, \pi]$.

Theorem 6.2 (Kac, Murdoch and Szegő [42])**.** *Under Condition A the following asymptotic relation holds:*

$$\lambda_{1,n}(f) \simeq n^{-2k} \quad \text{as} \quad n \to \infty. \tag{6.3}$$

Observe that conditions (**A2**) and (**A3**) are too restrictive which may be hard to verify or may even be unsatisfied in some areas of application such as prediction theory of stationary processes and signal processing, where f is viewed as a spectral density of a stationary process.

By using new linear algebra tools, Serra [62] has extended Theorem 6.2, by proving that the rate of convergence of $\lambda_{1,n}(f)$ depends only on the order of the zero of f, but not (necessarily) on the smoothness of f (conditions (**A2**)) as required in Theorem 6.2. In particular, in Serra [62] was proved the following result.

Proposition 6.1 (Serra [62], Corollary 2.1)**.** *Let f be a nonnegative integrable function on $[-\pi, \pi]$. If f has a unique zero of order $2k$ at a point $\lambda = \lambda_0$ (conditions (**A1**) and (**A3**)), then $\lambda_{1,n}(f) \asymp n^{-2k}$.*

Moreover, in Serra [64] this theory was further extended to the case of a function $f \in L^1[-\pi, \pi]$ having several global minima (zeros) by suppressing the condition (**A3**) as well, by showing that the maximal order of the zeros of function f is the only parameter which characterizes the rate of convergence of $\lambda_{1,n}(f)$ (see Serra [64], Theorem 4.2).

6.2. The relationship between the prediction error and the minimal eigenvalue

Let $X(t)$, $t = 0, \pm 1, \ldots$, be a stationary sequence possessing a spectral density function $f(\lambda)$, $\lambda \in [-\pi, \pi]$, and let $\sigma_n^2(f)$ be the prediction error in predicting $X(0)$ by the past of $X(t)$ of length n (see formula (3.1)).

The next proposition provides a relationship between the minimal eigenvalue $\lambda_{1,n}(f)$ of a truncated Toeplitz matrix $T_n(f)$ generated by spectral density f and the prediction error $\sigma_n^2(f)$ (see Pourahmadi [52] and Serra [62]).

Proposition 6.2. *Let f, $\lambda_{1,n}(f)$ and $\sigma_n^2(f)$ be as above. Then for any $n \in \mathbb{N}$ the following inequalities hold:*

$$\lambda_{1,n}(f) \leq \sigma_n^2(f) \leq M_f \frac{\lambda_{1,n}(f)}{\lambda_{1,n-1}(f)}. \tag{6.4}$$

The first inequality in (6.4) was proved in Pourahmadi [52], while the proof of the second inequality in (6.4) can be found in Serra [62].

Recall that for a stationary process $X(t)$ with spectral density $f(\lambda)$ by E_f we denote the spectrum of $X(t)$, that is, $E_f := \{\lambda : f(\lambda) > 0\}$ (see (5.6)). Thus, the closure $\overline{E}_f$ of E_f is the support of the spectral density f. Also, by $\tau^*(E_f)$ we denote the outer transfinite diameter of the set E_f (defined by (7.8)).

The following theorem is an immediate consequence of Theorem 5.8(b) and Proposition 6.2 (cf. Pourahmadi [52].)

Theorem 6.3. *Let f, $\lambda_{1,n}(f)$ and $\tau^*(E_f)$ be as above. Then the following inequality holds:*

$$\limsup_{n\to\infty} \sqrt[2n]{\lambda_{1,n}(f)} \leq \tau^*(E_f). \tag{6.5}$$

Thus, in order that the minimal eigenvalue $\lambda_{1,n}(f)$ should decrease to zero at least exponentially as $n \to \infty$, it is sufficient that the outer transfinite diameter of the spectrum of the process $X(t)$ be less than 1. As such the continuity and differentiability of spectral density f are not required for the exponential rate of convergence of the minimal eigenvalue $\lambda_{1,n}(f)$ to zero.

Now we proceed to discuss two specific models of deterministic processes, for which one can obtain more information on the rate of convergence of the minimal eigenvalue $\lambda_{1,n}(f)$ to zero as $n \to \infty$ from that of prediction error $\sigma_n^2(f)$. Notice that, for the first model, the spectral density f of the process is discontinuous and has uncountably many zeros, while, for the second model, the function f has a zero of exponential order at points $0, \pm\pi$. Therefore, in both cases, Condition A of Kac, Murdoch and Szegő is violated.

6.2.1. A model with spectral density f which is discontinuous, zero on an interval, and positive elsewhere

Let $X(t)$ be a stationary process for which the support $\overline{E}_f$ of the spectral density $f(\lambda)$ is as in Examples 5.1–5.4. Then, we can apply Theorems 5.9 and 6.3 to obtain asymptotic estimates for the minimal eigenvalue $\lambda_{1,n}(f)$.

In the next theorem we state the result in the cases where the support $\overline{E}_f$ of f is as in Examples 5.1 and 5.3, similar estimates can be stated in the cases where $\overline{E}_f$ is as in Examples 5.2 and 5.4.

Theorem 6.4 (Babayan and Ginovyan [6])**.** *Let $\overline{E}_f$ be the support of the spectral density f of a stationary process $X(t)$. Then for the the minimal eigenvalue $\lambda_{1,n}(f)$ of $T_n(f)$ the following asymptotic estimates hold.*

(a) *If* $\overline{E}_f = \Gamma_{2\alpha}(\theta_0)$*, where* $\Gamma_{2\alpha}(\theta_0)$ *is as in Example 5.1, then*

$$\limsup_{n\to\infty} \sqrt[n]{\lambda_{1,n}(f)} \le \sin^2(\alpha/2). \tag{6.6}$$

(b) *If* $\overline{E}_f = \Gamma_{\alpha,\delta}(\theta_0)$*, where* $\Gamma_{\alpha,\delta}(\theta_0)$ *is as in Example 5.3, then*

$$\limsup_{n\to\infty} \sqrt[n]{\lambda_{1,n}(f)} \le \sin(\alpha/2)\sin(\alpha/2+\delta). \tag{6.7}$$

It is important to note that in Theorem 6.4 the essential infimum $m_f = 0$ is attained at uncountably many points, and the spectral density f, in general, is not continuous on E_f or at the endpoints of E_f. Thus, f does not satisfy Condition A and yet the rate of convergence of $\lambda_{1,n}(f)$ to zero is much faster than in (6.3).

Using Davisson's theorem (Theorem 5.10) and its extension (Theorem 5.11) we obtain exact upper bounds for the minimal eigenvalue $\lambda_{1,n}(f)$ rather than the asymptotic estimates (6.6) and (6.7).

Theorem 6.5 (Babayan and Ginovyan [6])**.** *Let* $\overline{E}_f$ *be the support of the spectral density* f *of a stationary process* $X(t)$*. Then for the minimal eigenvalue* $\lambda_{1,n}(f)$ *the following inequalities hold.*

(a) *If* $\overline{E}_f = \Gamma_{2\alpha}(\theta_0)$*, where* $\Gamma_{2\alpha}(\theta_0)$ *is as in Example 5.1, then in view of* (5.26) *and the first inequality in* (6.4) *we have*

$$\lambda_{1,n}(f) \le 4c\,(\sin(\alpha/2))^{2n-2}, \tag{6.8}$$

where $c = r(0)$ *and* $r(\cdot)$ *is the covariance function of* $X(t)$*.*

(b) *If* $\overline{E}_f = \Gamma_{\alpha,\delta}(\theta_0)$*, where* $\Gamma_{\alpha,\delta}(\theta_0)$ *is as in Example 5.3, then in view of* (5.27) *and the first inequality in* (6.4) *we have*

$$\lambda_{1,n}(f) \le 4c\,(\sin(\alpha/2))^{n-1}\,(\sin(\alpha/2+\delta))^{n-1}. \tag{6.9}$$

6.2.2. A model with spectral density f possessing exponential order zero

Let $X(t)$ be a stationary process with spectral density f_a given by formula (5.3), that is, f_a is the Pollaczek-Szegő function. As it was observed (see (5.5))

$$f_a(\lambda) \sim \begin{cases} 2e^a \exp\{-a\pi/|\lambda|\} & \text{as } \lambda \to 0, \\ 2\exp\{-a\pi/(\pi-|\lambda|)\} & \text{as } \lambda \to \pm\pi. \end{cases}$$

Thus, the function f_a in (6.10) has a zero at points $x = 0, \pm\pi$ of exponential order and is positive elsewhere ($m_f = 0 < M_f$). Then by Theorem 5.2 we have

$$\delta_n(f_a) = \sigma_n^2(f_a) \sim n^{-a} \quad \text{as} \quad n \to \infty. \tag{6.10}$$

Now, by using the first inequality in (6.4), from (6.10) we conclude that

$$\lambda_{1,n}(f_a) = O\left(n^{-a}\right) \quad \text{as} \quad n \to \infty. \tag{6.11}$$

Thus, by choosing a large enough one can obtain a very fast rate of convergence of $\lambda_{1,n}(f_a)$ to zero as $n \to \infty$.

Remark 6.1. The asymptotic relation (6.11) remains valid for more general models. Indeed, let $X(t)$ be a stationary process with spectral density given by $f(\lambda) = f_a(\lambda)g(\lambda)$, where $f_a(\lambda)$ is as in (5.3), and $g(\lambda)$ is a function satisfying the conditions of one of Theorems 5.12, 5.13, 5.14, 5.16 or Corollary 5.4. Then for the minimal eigenvalue $\lambda_{1,n}(f)$ of a truncated Toeplitz matrix $T_n(f)$ generated by spectral density f, we have

$$\lambda_{1,n}(f) = O\left(n^{-a}\right) \quad \text{as} \quad n \to \infty.$$

More results concerning asymptotic behavior of the extreme eigenvalues of truncated Toeplitz matrices can be found in Böttcher and Grudsky [11], and Serra [62]–[64]). Observe that in the above models information from the theory of stationary processes was used to find linear-algebra results.

7. Tools

In this section we briefly discuss the tools, used to prove the results stated in Sections 4 and 5 (see Babayan and Ginovyan [4]–[6], and Babayan et al. [7].

7.1. Some metric characteristics of bounded closed sets in the plane

We introduce here some metric characteristics of bounded closed sets in the plane, such as, the transfinite diameter, the Chebyshev constant, and the capacity, and discuss some properties of these characteristics. Then, we state the theorems of Fekete and Robinson on the transfinite diameters of related sets as well as an extension of Robinson's theorem.

7.1.1. The transfinite diameter, the Chebyshev constant and the capacity

One of the fundamental results of geometric complex analysis is the classical theorem by Fekete and Szegő, stating that for any compact set F in the complex plane $\mathbb{C}$ the transfinite diameter, the Chebyshev constant and the capacity of F *coincide*, although they are defined from very different points of view. Namely, the transfinite diameter of the set F characterizes the asymptotic size of F, the Chebyshev constant of F characterizes the minimal uniform deviation of a monic polynomial on F, and the capacity of F describes the asymptotic behavior of the Green function at infinity. For the definitions and results stated in this subsection we refer the reader to the following references: Fekete [19], Goluzin [28], Chapter 7, Kirsch [43], Landkof [48], Chapter II, Ransford [57], Chapter 5, Saff [61], Szegő [70], Chapter 16, and Tsuji [71], Chapter III.

Transfinite diameter. Let F be an infinite bounded closed (compact) set in the complex plane $\mathbb{C}$. Given a natural number $n \geq 2$ and points $z_1, \ldots, z_n \in F$, we define

$$d_n(F) := \max_{z_1,\ldots,z_n \in F} \left[\prod_{1 \leq j < k \leq n}^{n} |z_j - z_k| \right]^{2/[n(n-1)]}, \tag{7.1}$$

which is the maximum of products of distances between the $\binom{n}{2} = n(n-1)/2$ pairs of points z_k, $k = 1, \dots, n$, as the points z_k range over the set F. The quantity $d_n(F)$ is called the nth *transfinite diameter* of the set F. Note that $d_2(F)$ is the diameter of F, and $d_n(F) \le d_2(F)$. Observe that the sequence $d_n(F)$ is non-increasing and non-negative, so has a finite limit as $n \to \infty$ which does not exceed the diameter $d_2(F)$ of F (see, e.g., Goluzin [28], p. 294). This limit, denoted by $d_\infty(F)$, is called the *transfinite diameter* of F. Thus, we have

$$d_\infty(F) := \lim_{n\to\infty} d_n(F). \tag{7.2}$$

If F is empty or consists of a finite number of points, we put $d_\infty(F) = 0$.

Chebyshev constant. Let F be as before, we put $m_n(F) := \inf \max_{z\in F} |q_n(z)|$, where the infimum is taken over all monic polynomials $q_n(z)$ from the class $\mathcal{Q}_n$, where $\mathcal{Q}_n$ is as in (3.3). Then there exists a unique monic polynomial $T_n(z, F)$ from the class $\mathcal{Q}_n$, called the *Chebyshev polynomial* of F of order n, such that

$$m_n(F) = \max_{z\in F} |T_n(z, F)|. \tag{7.3}$$

Fekete [19] proved that $\lim_{n\to\infty}(m_n(F))^{1/n}$ exists. This limit, denoted by $\tau(F)$, is called the *Chebyshev constant* for the set F. Thus,

$$\tau(F) := \lim_{n\to\infty} (m_n(F))^{1/n}. \tag{7.4}$$

Capacity (logarithmic). Let F be as above, and let D_F denote the complementary domain to F, containing the point $z = \infty$. If the boundary $\Gamma := \partial D_F$ of the domain D_F consists of a finite number of rectifiable Jordan curves, then for the domain D_F one can construct a Green function $G_F(z, \infty) := G_{D_F}(z, \infty)$ with a pole at infinity. This function is harmonic everywhere in D_F, except at the point $z = \infty$, is continuous including the boundary Γ and vanishes on Γ. It is known that in a vicinity of the point $z = \infty$ the function $G_F(z, \infty)$ admits the representation (see, e.g., Goluzin [28]), p. 309–310):

$$G_F(z, \infty) = \ln|z| + \gamma + O(z^{-1}) \quad \text{as} \quad z \to \infty. \tag{7.5}$$

The number γ in (7.5) is called the *Robin constant* of the domain D_F, and the number

$$C(F) := e^{-\gamma} \tag{7.6}$$

is called the *capacity* (or the *logarithmic capacity*) of the set F.

Now we are in position to state the above mentioned fundamental result of geometric complex analysis, due to M. Fekete and G. Szegő (see, e.g., Goluzin [28], p. 197 and Tsuji [71], p. 73).

Proposition 7.1 (Fekete-Szegő theorem)**.** *For any compact set $F \subset \mathbb{C}$, the transfinite diameter $d_\infty(F)$ defined by (7.2), the Chebyshev constant $\tau(F)$ defined by (7.4), and the capacity $C(F)$ defined by (7.6) coincide, that is,*

$$d_\infty(F) = C(F) = \tau(F). \tag{7.7}$$

It what follows, we will use the term 'transfinite diameter' and the notation $\tau(F)$ for (7.7).

The calculation of the transfinite diameter (and hence, the capacity and the Chebyshev constant) is a challenging problem, and in only very few cases has the transfinite diameter been exactly calculated (see, e.g., Landkof [48], p. 172–173, Ransford [57], p.135, and also Examples 5.1–5.4).

In the next proposition we list a number of properties of the transfinite diameter (and hence, of the capacity and the Chebyshev constant), which were used to prove the results stated in Section 5.3.

Proposition 7.2. *The transfinite diameter possesses the following properties.*

(a) The transfinite diameter is monotone, that is, for any closed sets F_1 and F_2 with $F_1 \subset F_2$, we have $\tau(F_1) \le \tau(F_2)$ (see, e.g., Saff [61], p. 169, Tsuji [71], p. 56).

(b) If a set F_1 is obtained from a compact set $F \subset \mathbb{C}$ by a linear transformation, that is, $F_1 := aF + b = \{az + b : z \in F\}$, then $\tau(F_1) = |a|\tau(F)$. In particular, the transfinite diameter $\tau(F)$ is invariant with respect to parallel translation and rotation of F (see, e.g., Goluzin [28], p. 298, Saff [61], p. 169, Tsuji [71], p. 56).

(c) The transfinite diameter of an arbitrary circle of radius R is equal to its radius R. In particular, the transfinite diameter of the unit circle $\mathbb{T}$ is equal to 1 (Tsuji [71], p. 84).

(d) The transfinite diameter of an arc Γ_α of a circle of radius R with central angle α is equal to $R\sin(\alpha/4)$. In particular, for the unit circle $\mathbb{T}$, we have $\tau(\Gamma_\alpha) = \sin(\alpha/4)$ (Tsuji [71], p. 84).

(e) The transfinite diameter of an arbitrary line segment F is equal to one-fourth its length, that is, if $F := [a, b]$, then $\tau(F) = \tau([a, b]) = (b - a)/4$. (see, e.g., Tsuji [71], p. 84).

The inner and outer transfinite diameters. τ-measurable sets. For an arbitrary (not necessarily closed) bounded set $E \subset \mathbb{C}$, we also define the *inner* and the *outer transfinite diameters*, denoted by $\tau_*(E)$ and $\tau^*(E)$, respectively, as follows (see, e.g., Babayan [2], Korovkin [47]):

$$\tau_*(E) := \sup_{F \subset E} \tau(F) \quad \text{and} \quad \tau^*(E) := \tau(\overline{E}), \tag{7.8}$$

where the supremum is taken over all compact subsets F of the set E, and $\overline{E}$ stands for the closure of E. Observe that $\tau_*(E) \le \tau^*(E)$. The set E for which $\tau_*(E) = \tau^*(E)$ is said to be τ-*measurable*, and in this case, we write $\tau(E)$ for the common value: $\tau(E) = \tau_*(E) = \tau^*(E)$.

7.1.2. Transfinite diameters of related sets

We state here the theorems of Fekete [19] and Robinson [58] on the transfinite diameters of related sets, as well as, an extension of Robinson's theorem.

The following classical theorem about the transfinite diameter of related sets was proved by Fekete [19] (see also Goluzin [28], pp. 299–300).

Proposition 7.3 (Fekete's theorem, Fekete [19]). *Let F be a bounded closed set in the complex w-plane $\mathbb{C}_w$, and let $p(z) := p_n(z) = z^n + c_1 z^{n-1} \cdots + c_n$ be an arbitrary monic polynomial of degree n. Let F^* be the preimage of F in the z-plane $\mathbb{C}_z$ under the mapping $w = p(z)$, that is, F^* is the set of all points $z \in \mathbb{C}_z$ such that $w := p(z) \in F$. Then*

$$\tau(F^*) = [\tau(F)]^{1/n}, \tag{7.9}$$

where $\tau(F)$ and $\tau(F^*)$ stand for the transfinite diameters of the sets F and F^*, respectively.

Observe that if in Fekete's theorem the mapping is carried out by an arbitrary (not necessarily monic) polynomial of degree n: $p(z) := p_n(z) = az^n + a_1 z^{n-1} \cdots + a_n$ $(a \neq 0)$, then we have

$$\tau(F^*) = [\tau(F)/|a|]^{1/n}. \tag{7.10}$$

In Robinson [58], Fekete's theorem was extended to the case where the mapping is carried out by a rational function instead of a polynomial. More precisely, in Robinson [58] was proved the following theorem.

Proposition 7.4 (Robinson's theorem, Robinson [58]). *Let $p(z) := p_n(z) = z^n + a_1 z^{n-1} \cdots + a_n$ and $q(z) := q_k(z)$ be arbitrary relatively prime polynomials of degrees n and k, respectively, with $k < n$. Let F be a bounded closed set in the complex w-plane $\mathbb{C}_w$, and let F^* be the preimage of F in the z-plane $\mathbb{C}_z$ under the mapping $w = \varphi(z) := p(z)/q(z)$. Assume that $|q(z)| = 1$ for all $z \in F^*$. Then*

$$\tau(F^*) = [\tau(F)]^{1/n}. \tag{7.11}$$

Remark 7.1. It is clear that the condition $|q(z)| = 1$ for all $z \in F^*$ in Robinson's theorem can be replaced by the condition $|q(z)| = C$ for all $z \in F^*$ with an arbitrary positive constant C, and, in this case, the relation (7.11) becomes $\tau(F^*) = [C\tau(F)]^{1/n}$.

Observe that in the special case where $q(z) \equiv 1$, Robinson's theorem reduces to the Fekete theorem (Proposition 7.3). Another case of special interest is $p(z) = z^2 + 1$ and $q(z) = 2z$. In this case, the mapping given by the rational function

$$\varphi(z) := \frac{p(z)}{q(z)} = \frac{1}{2}\left(z + \frac{1}{z}\right)$$

projects the subsets of the unit circle $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ onto the real axis $\mathbb{R}$, and, in view of Remark 7.1, Robinson's theorem (Proposition 7.4) reads as follows.

Proposition 7.5 (Robinson [58]). *Let F be a bounded closed subset of the complex plane $\mathbb{C}$ lying on the unit circle $\mathbb{T}$ and symmetric with respect to real axis, and let F^x be the projection of F onto the real axis. Then*

$$\tau(F) = [2\tau(F^x)]^{1/2}. \tag{7.12}$$

Remark 7.2. The examples given in Section 5.3.2 show that the formula (7.12) gives a simple way to calculate the transfinite diameters of some subsets of the unit circle $\mathbb{T}$, based only on the formula of the transfinite diameter of a line segment (see Proposition 7.2(e)).

7.1.3. An extension of Robinson's theorem

As observed above, the condition $|q(z)| = C$ for all $z \in F^*$ in Robinson's theorem (Proposition 7.4) is too restrictive, and it essentially reduces the range of applicability of the theorem into the following two cases:

(a) $q(z) \equiv 1$, and Robinson's theorem reduces to the Fekete theorem.

(b) $p(z) = z^2 + 1$ and $q(z) = 2z$, and, in this case, the rational function $\varphi(z) = (z^2+1)/(2z)$ projects the subsets of the unit circle $\mathbb{T}$ onto the real axis $\mathbb{R}$.

Therefore, the question of extending Robinson's theorem to the case where the condition $|q(z)| = M$ is replaced by a weaker condition becomes topical. The next result, which was proved in Babayan and Ginovyan [6], provides such an extension.

Proposition 7.6 (Babayan and Ginovyan [6])**.** *Let the polynomials $p(z)$, $q(z)$, the sets F, F^*, and the mapping $w = \varphi(z) := p(z)/q(z)$ be as in Proposition 7.4, and let $m := \min_{z\in F^*} |q(z)|$ and $M := \max_{z\in F^*} |q(z)|$. Then the following inequalities hold:*

$$[m\tau(F)]^{1/n} \le \tau(F^*) \le [M\tau(F)]^{1/n}. \tag{7.13}$$

Remark 7.3. If the condition $|q(z)| = C$ is satisfied for all $z \in F^*$, then we have $m = M = C$, and Proposition 7.6 reduces to Robinson's theorem (Proposition 7.4).

Remark 7.4. Proposition 7.6 can easily be extended to the more general case where $p(z) := p_n(z)$ is an arbitrary (not necessarily monic) polynomial of degree n: $p(z) = az^n + a_1 z^{n-1} \cdots + a_n$, $a \neq 0$. Indeed, in this case, canceling the fraction $\varphi(z) := p(z)/q(z)$ by a, we get $\varphi(z) := p_1(z)/q_1(z)$, where now $p_1(z) = p(z)/a = z^n +$ lower order terms, is a monic polynomial. Also, we have $\min_{z\in F^*} |q_1(z)| = m/|a|$ and $\max_{z\in F^*} |q_1(z)| = M/|a|$, where m and M are as in Proposition 7.6. Hence we can apply the inequality (7.13) to obtain

$$\left[\frac{m}{|a|}\tau(F)\right]^{1/n} \le \tau(F^*) \le \left[\frac{M}{|a|}\tau(F)\right]^{1/n}.$$

7.2. Weakly varying sequences

We introduce here the notion of weakly varying sequences and state some of their properties (see Babayan et al. [7]). This notion was used in the specification of the class $\mathcal{F}$ of deterministic processes considered in Section 5.5 (see formula (5.29)).

Definition 7.1. *A sequence of non-zero numbers* $\{a_n,\ n \in \mathbb{N}\}$ *is said to be weakly varying if*

$$\lim_{n\to\infty} a_{n+1}/a_n = 1.$$

For example, the sequence $\{n^\alpha,\ \alpha \in \mathbb{R},\ n \in \mathbb{N}\}$ is weakly varying (for $\alpha < 0$ it is weakly decreasing and for $\alpha > 0$ it is weakly increasing), while the geometric progression $\{q^n,\ 0 < q < 1,\ n \in \mathbb{N}\}$ is not weakly varying.

In the next proposition we list some simple properties of the weakly varying sequences, which can easily be verified.

Proposition 7.7. *The following assertions hold.*

(a) If a_n *is a weakly varying sequence, then* $\lim_{n\to\infty} a_{n+\nu}/a_n = 1$ *for any* $\nu \in \mathbb{N}$.
(b) If a_n *is such that* $\lim_{n\to\infty} a_n = a \neq 0$*, then* a_n *is a weakly varying sequence.*
(c) If a_n *and* b_n *are weakly varying sequences, then* ca_n $(c \neq 0)$*,* $a_n^\alpha\,(\alpha \in \mathbb{R}, a_n > 0)$*,* a_nb_n *and* a_n/b_n *also are weakly varying sequences.*
(d) If a_n *is a weakly varying sequence, and* b_n *is a sequence of non-zero numbers such that* $\lim_{n\to\infty} b_n/a_n = c \neq 0$*, then* b_n *is also a weakly varying sequence.*
(e) If a_n *is a weakly varying sequence of positive numbers, then it is exponentially neutral (see Definition 5.1(a) and Remark 5.5).*

7.3. Some properties of the geometric mean and trigonometric polynomials

Recall that a trigonometric polynomial $t(\lambda)$ of degree ν is a function of the form:

$$t(\lambda) = a_0 + \sum_{k=1}^{\nu}(a_k \cos k\lambda + b_k \sin k\lambda) = \sum_{k=-\nu}^{\nu} c_k e^{ik\lambda}, \quad \lambda \in \mathbb{R},$$

where $a_0, a_k, b_k \in \mathbb{R}$, $c_0 = a_0$, $c_k = 1/2(a_k - ib_k)$, $c_{-k} = \overline{c}_k = 1/2(a_k + ib_k)$, $k = 1, 2, \ldots, \nu$.

Recall that for a function $h \geq 0$ by $G(h)$ we denote the geometric mean of h (see formula (2.22)). In the next proposition we list some properties of the geometric mean $G(h)$ and trigonometric polynomials (see Babayan et al. [7]).

Proposition 7.8. *The following assertions hold.*

(a) Let $c > 0$*,* $\alpha \in \mathbb{R}$*,* $f \geq 0$ *and* $g \geq 0$*. Then*

$$G(c) = c; \quad G(fg) = G(f)G(g); \quad G(f^\alpha) = G^\alpha(f)\ (G(f) > 0). \tag{7.14}$$

(b) $G(f)$ *is a non-decreasing functional of* f*: if* $0 \leq f(\lambda) \leq g(\lambda)$*, then* $0 \leq G(f) \leq G(g)$*. In particular, if* $0 \leq f(\lambda) \leq 1$*, then* $0 \leq G(f) \leq 1$*.*

(c) (Fejér-Riesz theorem, see. e.g., Grenander and Szegő, Sec. 1.12). *Let* $t(\lambda)$ *be a non-negative trigonometric polynomial of degree* ν. *Then there exists an algebraic polynomial* $s_\nu(z)$ $(z \in \mathbb{C})$ *of the same degree* ν, *such that* $s_\nu(z) \neq 0$ *for* $|z| < 1$, *and*

$$t(\lambda) = |s_\nu(e^{i\lambda})|^2. \tag{7.15}$$

Under the additional condition $s_\nu(0) > 0$ *the polynomial* $s_\nu(z)$ *is determined uniquely.*

(d) *Let* $t(\lambda)$ *and* $s_\nu(z)$ *be as in Assertion (c). Then* $G(t) = |s_\nu(0)|^2 > 0$.

Acknowledgments

The authors are grateful to their supervisor Academician, Professor Il'dar Abdullovich Ibragimov for introducing them to this research area and permanent support.

The authors would like to thank the referees and Editor in Chief Mikhail Lifshits for their constructive comments and suggestions.

References

[1] Baxter, G. (1962). An Asymptotic Result for the Finite Predictor. *Math. Scand.* **10**, 137–144. MR0149584

[2] Babayan, N. M. (1984). On the asymptotic behavior of prediction error. *J. of Soviet Mathematics* **27**(6), 3170–3181. MR0772073

[3] Babayan, N. M. (1985). On asymptotic behavior of the prediction error in the singular case. *Theory Probab. Appl.* **29**(1), 147–150. MR0739513

[4] Babayan, N. M. and Ginovyan, M. S. (2020). On hyperbolic decay of prediction error variance for deterministic stationary sequences. *J. Cont. Math Anal.* **55**(2), 76–95. MR4107999

[5] Babayan, N. M. and Ginovyan, M. S. (2022). On asymptotic behavior of the prediction error for a class of deterministic stationary sequences. *Acta Math. Hungar.* **167** (2), 501–528. MR4487622

[6] Babayan, N. M. and Ginovyan, M. S. (2022). On the prediction error for singular stationary processes and transfinite diameters of related sets. *Zapiski Nauchn. Semin. POMI* (accepted). MR4503187

[7] Babayan, N. M., Ginovyan, M. S. and Taqqu, M.S. (2021). Extensions of Rosenblatt's results on the asymptotic behavior of the prediction error for deterministic stationary sequences. *J. Time Ser. Anal.*, **42**, 622–652. MR4325668

[8] Bello, M. H. and López G. L. (1998). Ratio and Relative Asymptotics of Polynomials Orthogonal on an Arc of the Unit Circle, *J. Approx. Theory* **92**, 216–244. MR1604927

[9] Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods.* Springer, Berlin. MR3075595

[10] Bingham, N.H. (2012). Szegő's theorem and its probabilistic descendants. *Probability Surveys* **9**, 287–324. MR2956573

[11] Böttcher, A. and Grudsky, S. (1998). On the condition numbers of large semidefinite Toeplitz matrices. *Linear Algebra Appl.* **279**, 285–301. MR1637913

[12] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Second Edition. Springer, New York. MR1093459

[13] Chan, R. H. (1991). Toeplitz preconditioners for Toeplitz systems with nonnegative generating functions. *IMA J. Numer. Anal.* **11**, 333–345. MR1118960

[14] Cramér, H. and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications.* Wiley, New York. MR0217860

[15] Davisson, L. D. (1965). Prediction of time series from finite past. *J. Soc. Indust. Appl. Math.* **13**(3), 819–826. MR0183580

[16] Dette, H. and Studden, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis.* Wiley, New York. MR1468473

[17] Devinatz, A. (1964). Asymptotic estimates for the finite predictor, *Math. Scand.* **15**, 111–120. MR0188708

[18] Doob, J.L. (1953). *Stochastic Processes*, Wiley, New York. MR0058896

[19] Fekete, M. (1930). Über den transfiniten Durchmesser ebener Punktmengen. Zweite Mitteilung. *Math. Z.* **32**, 215–221. MR1545162

[20] Fortus, M. I.(1990). Prediction of a stationary time series with the spectrum vanishing on an interval. *Akademiia Nauk SSSR, Izvestiia, Fizika Atmosfery i Okeana* **26**, 1267–1274.

[21] Geronimus, Ya. L. (1948). Polynomials orthogonal on circle and their applications. *Zapiski Nauchno-Issled. Inst. Mat. Mech. Harkov Mat Obsh* **4**(19), 35–120. MR0036872

[22] Geronimus, Ya. L. (1948). On certain asymptotic properties of polynomials. *Mat. Sb.* **23**(1), 77–88. MR0026715

[23] Geronimus, Ya. L. (1958). *Polynomials orthogonal on ciscle and segment.* Fiz-Mat, Moscow. MR0133641

[24] Geronimus, Ya. L. (1967). On a problem of G. Szegő, M. Kac, G. Baxter and I. Hirshman, *Izv. AN SSSR, ser. Matematika* **31**, 289–304. MR0213801

[25] Ginovian, M. S. (1999). Asymptotic behavior of the prediction error for stationary random sequences. *J. Cont. Math. Anal.* **4**(1), 14–33. MR1854056

[26] Golinskii, B. L. (1974). On asymptotic behavior of the prediction error. *Theory Probab. Appl.* **19**(4), 724–739. MR0420815

[27] Golinskii, B. L. and Ibragimov, I. A. (1971). On G. Szegő limit theorem. *Izv. AN SSSR, Ser. Matematika*, **35**, 408–427. MR0291713

[28] Goluzin, G. M. (1969). *Geometric Theory of Functions of a Complex Variable.* Amer. Math. Soc., Providence. MR0247039

[29] Grenander, U. and Rosenblatt, M. (1954). An Extension of a Theorem of G. Szegő and its Application to the Study of Stochastic Processes. *Trans.*

Amer. Math. Soc. **76**, 112–126. MR0058902

[30] Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series.* Wiley, New York (2nd corr. ed. AMS Chelsea publishing). MR0890514

[31] Grenander, U. and Szegő, G. (1958). *Toeplitz Forms and Their Applications.* University of California Press, Berkeley and Los Angeles. MR0094840

[32] Hannan, E. J. (1970). *Multiple Time Series.* Wiley, New York. MR0279952

[33] Helson, H. and Szegő, G. (1960). A problem in prediction theory. *Acta Mat. Pura Appl.* **51**, 107–138. MR0121608

[34] Hirschman, I. I. (1965). Finite sections of Wiener-Hopf equations and Szegő polynomials. *J. Math. Anal. Appl.* **11**, 290–320. MR0181907

[35] Hoffman, K.. (1958). *Banach Spaces of Analytic Functions.* Prentice-Hall, Englewood Cliffs, N.J. MR0133008

[36] Ibragimov, I. A. (1964). On asymptotic behavior of the prediction error. *Theory Probab. Appl.*, **9**(4), 695–703. MR0172340

[37] Ibragimov, I. A. (1968). On a Theorem of G. Szegő. *Mat. Zametki* **3**(6), 693–703. MR0231114

[38] Ibragimov, I.A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff Publishing Groningen, The Netherlands. MR0322926

[39] Ibragimov, I.A. and Rozanov, Yu. A. (1978). *Gaussian Random Processes.* Springer, New York. MR0543837

[40] Ibragimov, I.A. and Solev, V. N. (1968). The asymptotic behavior of the prediction error of a stationary sequence with the spectral density function of a special form. *Theory Probab. Appl.* **13**(4), 746–750. MR0242242

[41] Inoue, A. (2002). Asymptotic behavior for partial autocorrelation functions of fractional ARIMA processes. *Ann. Appl. Probab.* **12**(4), 1471–1491. MR1936600

[42] Kac, M., Murdoch, W. L. and Szegő, G. (1953). On the eigenvalues of certain Hermitian forms, *Rational Mech. Anal.* **9**, 767–800. MR0059482

[43] Kirsch, S. (2005). Transfinite diameter, Chebyshev constant and capacity. *Handbook of Complex Analysis: Geometric Function Theory. Vol. 2* (Kühnau R., ed.), Elsevier, Amsterdam, 243–308. MR2121861

[44] Kolmogorov, A. N. (1941). Stationary sequences in a Hilbert space. *Bull. Moscow State University*, **2**(6), 1–40. MR0009098

[45] Kolmogorov, A. N. (1941). Interpolation and Extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR. Ser. Mat.* **5**, 3–14.

[46] Koopmans, L.H. (1974). *The spectral analysis of time series.* Academic Press, New York. MR0348953

[47] Korovkin, P. P. (1958). The capacity of sets and polynomials minimizing an integral. *Uch. Zap. Kaliningr. Pedinst.*, **5**, 34–52.

[48] Landkof, N. S. (1972). *Fundamentals of Modern Potential Theory.* Springer, New York. MR0350027

[49] Parter, S. V. (1966). Extreme eigenvalues of Toeplitz forms and applications to elliptic difference equations. *Trans. Amer. Math. Soc.* **99**, 153–192.

MR0120492
[50] Pierson, W. J. JR. (1955). Wind Generated Gravity Waves. *Advances in Geophysics 2*, 93–178. MR0134316
[51] Pollaczek, F. (1949). Sur une généralisation des polynômes de Legendre, *C. R. Acad. Sci. Paris* **228**, 1363–1365. MR0030037
[52] Pourahmadi, M. (1988). Remarks on extreme eigenvalues of Toeplitz matrices. *Internat. J. Math. & Math. Sci.*, **11**(1), 23–26. MR0918214
[53] Pourahmadi, M. (2001). *Fundamentals of Time Series Analysis and Prediction Theory.* Wiley, New York. MR1849562
[54] Rakhmanov, E. A. (1983). On the asymptotics of the ratio of orthogonal polynomials. II. *Math. USSR Sb.* **46**(1), 105–117. MR0654647
[55] Rakhmanov, E. A. (1987). On asymptotic properties of polynomials orthogonal on the circle with weights not satisfying Szegő's condition. *Math. USSR Sb.* **58**, 149–167. MR0854969
[56] Ramsey, F. L. (1974). Characterization of the partial autocorrelation function. *Ann. Statist.* **2**, 1296–1301. MR0359219
[57] Ransford, T. (1995). *Potential Theory in the Complex Plane.* Cambridge University Press, Cambridge. MR1334766
[58] Robinson, R. M. (1969). On the transfinite diameters of some related sets. *Math. Z.* **108**, 377–380. MR0240294
[59] Rosenblatt, M. (1957). Some Purely Deterministic Processes. *J. of Math. and Mech.* **6**(6), 801–810. (Reprinted in: *Selected works of Murray Rosenblatt*, Davis R.A, Lii K.-S., Politis D.N. eds., Springer, New York, 124–133, (2011). MR0093827
[60] Rozanov, Yu. A. (1967). *Stationary random processes.* Holden-Day, San Francisco. MR0214134
[61] Saff, E.B. (2010). Logarithmic Potential Theory with Applications to Approximation Theory. *Surveys in Approximation Theory* **5**, 165–200. MR2734198
[62] Serra, S. (1998). On the Extreme Eigenvalues of Hermitian (Block) Toeplitz Matrices. *Linear Algebra and its Applications* **270**, 109–129. MR1484077
[63] Serra, S. (2000). How bad can positive definite Toeplitz matrices be? *Numer. Func. Anal. and Optimiz.* **21**, 255–261. MR1760001
[64] Serra, S. (1996). On the extreme spectral properties of Toeplitz generated by L_1 functions with several minima/maxima. *BIT*, **36**, 135–142. MR1431579
[65] Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle. Part 1: Classical Theory, Part 2: Spectral Theory.* AMS Colloqium Publications, Vol. **54**, Parts 1,2, Providence, Rhode Island. MR2105088
[66] Shiryaev, A. N. (1984). *Probability.* Springer, New York. MR0737192
[67] Szegő, G. (1915). Ein Grenzwertsatz über die Toeplitzschen Determinanten einer reellen positiven Funktion. *Math. Ann.* **76**, 490–503. MR1511838
[68] Szegő, G. (1921). Über orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehören. *Math. Zeitschrift.* **9**, 218–270. MR1544465

[69] Szegő, G. (1950). On certain special sets of orthogonal polynomials. *Proc. Amer. Math. Soc.* **1**, 731–737. MR0042546

[70] Szegő, G. (1939). *Orthogonal Polynomials.* Amer. Math. Soc. Colloq. Publ., 23, American Mathematical Society, Providence (3rd edition, 1967). MR0310533

[71] Tsuji, M. (1975). *Potential theory in modern function theory*, 2nd edition, Chelsea Pub. Co, New York. MR0414898

[72] Verblunsky, S. (1935). On positive harmonic functions: A contribution to the algebra of Fourier series. *Proc. London Math. Soc.* (2) **38**, 125–157. MR1576309

[73] Verblunsky, S. (1936). On positive harmonic functions (second paper). *Proc. London Math. Soc.* (2) **40**, 290–320. MR1575824

[74] Widom, H. (1958). On the eigenvalues of certain Hermitian operators. *Trans. Amer. Math. Soc.* **88**, 491–522. MR0098321

[75] Wiener, N. (1949). *Extrapolation, interpolation and smoothing of stationary time series. With engineering applications.* MIT Press/Wiley (reprinted, Martino Publishing, 2013). MR0031213