Transactions of A. Razmadze
Mathematical Institute
Vol. **176** (2022), issue 3, 323–344

RECENT PROGRESS IN COMPUTABILITY FOR PREDICTION AND WIENER FILTER THEORY

HOLGER BOCHE[1], VOLKER POHL[1] AND H. VINCENT POOR[2]

Dedicated to the Memory of Edem Lagvilava

Abstract. This paper reviews some fundamental results in prediction theory, spectral factorization and Wiener filtering with a particular focus on questions of computability. Since the mathematical theory of prediction and estimation of stationary time series was established by Kolmogorov, Paley, Szegö, Wiener, and many others, it seems to be an open question whether it is possible to effectively compute optimal filter coefficients and important performance measures on digital computers, i.e. on Turing machines.

In this paper, we show that the optimum mean squared error (MSE) for predicting a stationary time series from its past observations is generally not Turing computable. However, under an additional condition on the stochastic process, namely, for strictly positive spectral densities, Turing computability of the corresponding optimal MSE can be guaranteed. Nevertheless, even if the MSE is Turing computable, there always exist spectral densities that are polynomial-time computable on a Turing machine, but such that the corresponding optimal MSE is not polynomial-time computable. This observation proves a complexity blowup for the computation of the MSE on digital computers. Finally, we show that the spectral factorization and the calculation of the optimal prediction filter are generally not Turing computable even under additional strong assumptions on the smoothness of the spectral density.

1. Introduction

The theory of interpolation, prediction and estimation of time-series has a central role in many different areas of science and engineering such as communications, control, signal processing, econometrics, statistics, and many more. The general problem is to predict or estimate stochastic time-series from observations of the same time-series or from observations of a correlated time-series.

From the engineering side, the main part of the theory was developed during the 1940's and is particular related to the works of Wiener [46] and Kolmogorov [25]. The theory was then further developed in many different directions by many other researches [8, 9, 23, 48]. This paper presents some recent results in three major areas of this theory, namely, in 1) the prediction of stochastic processes, 2) the computation of the spectral factorization, and 3) the computation of the Wiener filter.

In prediction theory, we consider the problem of predicting the value x_0 of a discrete stochastic time series $\{x_n\}_{n\in\mathbb{Z}}$ from the past values $\{x_{-1}, x_{-2}, x_{-3}, \dots\}$. If this prediction should be done by a linear filter H, it will have the form $\widehat{x}_0 = \mathrm{H}(\boldsymbol{x}) = \sum_{n=1}^{\infty} h_n x_{-n}$ and the problem is to find the filter coefficients $\{h_n\}_{n=1}^{\infty}$ such that the mean squared error (MSE) $\sigma^2 = \mathrm{E}\big[\,|\widehat{x}_0 - x_0|^2\,\big]$ is minimized. As mentioned before, the solution of this problem is well known and there exist many different algorithms for determining the filter coefficients $\{h_n\}_{n=1}^{\infty}$ based on the spectral density φ of $\boldsymbol{x}$. In general, the optimal filter H will be an infinite impulse response (IIR) filter, i.e. all coefficients $\{h_n\}_{n=1}^{\infty}$ might be non-zero. Then for practical implementations, the IIR filter is usually approximated by a finite impulse response (FIR) filter of length N, obtained, for example, by simply truncating the IIR filter. This gives

2020 *Mathematics Subject Classification.* Primary 68Q25, 60G25; Secondary 68Q04, 47A68.

Key words and phrases. Computability; Complexity blowup; Estimation; Prediction Spectral factorization; Turing machine; Wiener Filter.

a prediction filter $\widehat{x}_0 = \mathrm{H}_N(\boldsymbol{x}) = \sum_{n=1}^{N} h_n x_{-n}$. It is clear that the FIR filter has a slightly larger MSE $\sigma_N^2 = \mathrm{E}\big[\,|\mathrm{H}_N(\boldsymbol{x}) - x_0|^2\,\big] \geq \sigma^2$. So, the obvious practical question is then: How to choose the filter length N? The answer to this questions seems to be almost trivial: Choose N so that $|\sigma_N^2 - \sigma^2| < 2^{-M}$ for a required precession $M \in \mathbb{N}$, which depends on the requirements of the actual application. There are several algorithms that are especially suited to implement this idea. One example is the Durbin–Levinson algorithm (cf., Section 3.5). It calculates successively the filter coefficients $h_1, h_2, h_3, \ldots$ and, at the same time, the corresponding MSEs $\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots$. However, in order to terminate the algorithm, one has to decide algorithmically in every step whether $|\sigma_N^2 - \sigma^2| < 2^{-M}$ or not. So, the question arises whether it is possible to implement an algorithm on a digital computer whose input is a spectral density φ and a natural number $M > 0$ and that is able to calculate an index $N \in \mathbb{N}$ so that $|\sigma_N^2 - \sigma^2| < 2^{-M}$. We show in this paper that this question has, generally, a negative answer. Even more, our result does not only hold for the Durbin–Levinson algorithm, but it implies that such an algorithm never exists for *any* possible method for calculating FIR approximations of the optimal filter H. Nevertheless, in a second step, we will derive sufficient conditions on the spectral density φ of the stochastic process such that a procedure for approximating the optimal filter H by an FIR filter exists for which it is possible to control the approximation error, i.e. for which it will be possible to choose algorithmically the approximation degree N so that $|\sigma_N^2 - \sigma^2| < 2^{-M}$. However, even if the necessary approximation degree N can be determined algorithmically from the given spectral density φ, we may ask for the computational complexity of this algorithm. We are going to show that even for low-complexity[1] spectral densities the algorithmic computation of the necessary FIR approximation length N will be of high complexity which means that the computational complexity of determining N grows faster than any polynomial in the desired precision M.

Apart from the Durbin–Levinson algorithm mentioned above, there are many other algorithms to determine the optimal prediction filter. In fact, it is known that there even exists a closed form expression for the transfer function of the optimal prediction filter H. This may be used to determine the optimal filter coefficients and an FIR approximation [4]. The closed form solution for H relies on the so-called *spectral factorization* (also known as *Wiener–Hopf factorization* [45]) of the spectral density φ, an operation that is also used in many other applications and which is of great importance in engineering and signal processing. Consequently, many algorithms were developed [26, 35] for computing the spectral factorization or at least to approximate the spectral factors.

The more demanding case of matrix-valued spectral factorization [47] was investigated and advanced especially by *Edem Lagvilava* and co-authors [13–16, 21, 22] and also, for this case there exist many different algorithms.

In this paper, we consider only the scalar case and ask whether for a given spectral density φ the spectral factor φ_+ can be computed on a digital computer, i.e. whether it is possible to have an algorithm that computes an approximation of φ_+ and that stops automatically if a predefined error bound is reached. Of course, the answer depends again on the spectral density and we will provide a sharp characterization (in terms of the smoothness of the spectral density) of classes of spectral densities for which the spectral factor can be computed on a digital computer and also of classes for which this is not possible.

The organization of this paper is as follows. Section 2 introduces the main notation and function spaces that are used throughout this paper. It also gives a short review of the main results and a notion from the computability theory. In Section 3, we give a short summary on techniques, algorithms and results in prediction theory, which are later investigated from a computational point of view. The computability of the optimal mean squared error (MSE) is investigated in Section 4, and Section 5 will depict that the computation of the optimal MSE shows complexity blowup. Section 6 will finally discuss the computability of the spectral factorization with an application to Wiener filtering. The paper is concluded with a short summary in Section 7.

[1]We will precisely define what this means in Section 2.2.

2. Notation and Preliminaries

This section introduces the main notion and function spaces that will be used throughout the remainder of this paper. The first subsection gives definitions of several Banach spaces of functions on the unit circle and of analytic functions inside the unit disk. In the second subsection, we shortly introduce the main concepts and notion from computability theory.

2.1. **Function spaces on the unit circle and unit disk.** Throughout this paper, $\mathbb{D} = \{z \in \mathbb{C} \ : \ |z| < 1\}$ stands for the open *unit disk* in the complex plane $\mathbb{C}$, whereas its boundary, the *unit circle*, is denoted by $\mathbb{T} = \{z \in \mathbb{C} \ : \ |z| = 1\}$.

Let μ be a finite positive measure on $\mathbb{T}$. Then for any $1 \leq p < \infty$, we write $L^p(\mu)$ for the usual Banach spaces of integrable functions on $\mathbb{T}$ with

$$\|f\|_p = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(\mathrm{e}^{\mathrm{i}\theta}) \right|^p \mathrm{d}\mu(\mathrm{e}^{\mathrm{i}\theta}) \right)^{1/p} < \infty \, ,$$

and $L^\infty(\mu)$ is the Banach space of essentially bounded (with respect to μ) functions on $\mathbb{T}$, i.e. the functions for which

$$\|f\|_\infty = \operatorname*{ess\,sup}_{\zeta \in \mathbb{T}} |f(\zeta)| < \infty \, .$$

If μ is the Lebesgue measure, we simply write $L^p(\mathbb{T})$. Moreover, we notice that $L^2(\mu)$ is a Hilbert space with the inner product $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\mathrm{e}^{\mathrm{i}\theta}) \overline{g(\mathrm{e}^{\mathrm{i}\theta})} \, \mathrm{d}\mu(\mathrm{e}^{\mathrm{i}\theta})$.

As usual, $\mathcal{C}(\mathbb{T})$ stands for the Banach space of continuous functions on $\mathbb{T}$ equipped with the maximum norm $\|f\|_\infty = \max_{\zeta \in \mathbb{T}} |f(\zeta)|$. A function f on $\mathbb{T}$ is said to be *absolute continuous* if there exists a $g \in L^1(\mathbb{T})$, so that $f(\mathrm{e}^{\mathrm{i}\theta}) = f(1) + \int_0^\theta g(\mathrm{e}^{\mathrm{i}\tau}) \, \mathrm{d}\tau$ for all $\theta \in [-\pi, \pi)$. We write $\mathcal{C}_{\mathrm{ac}}(\mathbb{T})$ for the set of all absolute continuous functions on $\mathbb{T}$. For any $f \in L^1(\mathbb{T})$, the *Fourier coefficients* are given by

$$c_n(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{e}^{-\mathrm{i}n\theta} \, \mathrm{d}\theta \, , \qquad n \in \mathbb{Z} \, , \tag{1}$$

and $f(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{n \in \mathbb{Z}} c_n(f) \, \mathrm{e}^{\mathrm{i}n\theta}$ is the corresponding *Fourier series* which converges in the $L^p(\mathbb{T})$-norm for all $1 < p < \infty$. The *Wiener algebra* $\mathcal{W}$ is the set of all $f \in L^1(\mathbb{T})$ with an absolutely converging Fourier series, i.e.

$$\mathcal{W} = \left\{ f \in L^1(\mathbb{T}) \ : \ \|f\|_{\mathcal{W}} = \sum_{n \in \mathbb{Z}} |c_n(f)| < \infty \right\} ,$$

and one has $\mathcal{W} \subsetneq \mathcal{C}(\mathbb{T})$ with $\|f\|_\infty \leq \|f\|_{\mathcal{W}}$ for all $f \in \mathcal{W}$.

The set of all polynomials of degree $N \in \mathbb{N}$, i.e. the set of all functions of the form $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots a_N z^N$ with $z \in \mathbb{C}$ and complex coefficients $\{a_n\}_{n=0}^N \subset \mathbb{C}$ is denoted by $\mathcal{P}_N$. Moreover, we write $H(\mathbb{D})$ for the set of all functions that are holomorphic (i.e. analytic) in the unit disk $\mathbb{D}$. For $1 \leq p < \infty$, $H^p(\mathbb{D})$ denotes the usual *Hardy space* (see, e.g., [12, 19, 20]) of all $f \in H(\mathbb{D})$ for which

$$\|f\|_p = \sup_{r<1} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| f(r\mathrm{e}^{\mathrm{i}\theta}) \right|^p \, \mathrm{d}\theta \right)^{1/p} < \infty \, ,$$

and $H^\infty(\mathbb{D})$ is the Banach space of all bounded analytic functions, i.e. the set of all $f \in H(\mathbb{D})$ with $\|f\|_\infty = \sup_{|z|<1} |f(z)| < \infty$. For any $f \in H^p(\mathbb{D})$, the radial limit $\lim_{r \to 1} f(r\mathrm{e}^{\mathrm{i}\theta}) = f(\mathrm{e}^{\mathrm{i}\theta})$ exists for almost every $\theta \in [-\pi, \pi)$ and this boundary function belongs to $L^p(\mathbb{T})$. Therewith, $H^p(\mathbb{D})$ can be characterized as a closed subspaces of $L^p(\mathbb{T})$, namely, $H^p(\mathbb{D}) = \{f \in L^p(\mathbb{T}) \ : \ c_n(f) = 0 \text{ for all } n < 0\}$. So, any $f \in H^p(\mathbb{D})$ can be written as a power series $f(z) = \sum_{n=0}^{\infty} c_n(f) z^n$. The *disk algebra* $\mathcal{A}(\mathbb{D})$ is

the Banach algebra of all holomorphic functions in $\mathbb{D}$ that are continuous in the closed units disk $\overline{\mathbb{D}} = \mathbb{D} \cup \mathbb{T}$ with the norm $\|f\|_\infty = \max_{z \in \overline{\mathbb{D}}} |f(z)|$ and with a pointwise multiplication.

Finally, $H_0^p(\mathbb{D}) = \{f \in H^p(\mathbb{D}) : f(0) = c_0(f) = 0\}$ and the subspaces $L_0^p(\mathbb{T}) \subset L^p(\mathbb{T})$, $\mathcal{C}_0(\mathbb{T}) \subset \mathcal{C}(\mathbb{T})$, and $\mathcal{A}_0(\mathbb{D}) \subset \mathcal{A}(\mathbb{D})$ are defined in the same way.

Remark 2.1. In this paper, we always work with spaces of functions that are analytic inside the unit disk $\mathbb{D}$. Nevertheless, especially in engineering and signal processing (cf., e.g., [29,32]), one often works with functions that are analytic outside the unit disk $\mathbb{D}$. As an example, consider a time-discrete signal $\boldsymbol{x} = \{x_n\}_{n \in \mathbb{Z}}$ with $\|\boldsymbol{x}\|_{\ell^1} = \sum_{n \in \mathbb{Z}} |x_n| < \infty$. Then its *z-transform* is defined to be

$$X(z) = \sum_{n \in \mathbb{Z}} x_n \, z^{-n} \,, \qquad |z| = 1 \,. \tag{2}$$

For sequences $\boldsymbol{x} \in \ell^1$ with $x_n = 0$ for all $n < 1$, the z-transform (2) is analytic for $|z| > 1$ and continuous for all $|z| \geq 1$. Then the function $X_1(z) := X(z^{-1})$ is analytic in $\mathbb{D}$ and continuous for all $|z| \leq 1$. So, our decision to work with analytic functions in $\mathbb{D}$ is mainly a question of notation. Using the mapping $z \mapsto z^{-1}$, all of our results can be translated to results on functions, analytic outside the unit disk.

2.2. **Computability analysis.** We investigate the computability of different important performance measures in prediction theory. So, we ask whether several quantities in this theory can be computed on a digital computer. To this end, we use the standard model of a *Turing machine* [41–43]. This is an abstract device which provides a theoretical model describing the fundamental limits of any realizable digital computer.

Definition 2.1. A number $t \in \mathbb{R}$ is said to be *computable* if there exists a Turing machine TM with input $n \in \mathbb{N}$ and output $\gamma(n) = \mathrm{TM}(n) \in \mathbb{Q}$ such that

$$|t - \gamma(n)| \leq 2^{-n} \,, \qquad \text{for all } n \in \mathbb{N} \,. \tag{3}$$

In this case, we say that $\gamma(n)$ *binary converges* to t, and we write $\mathbb{R}_{\mathrm{c}} \subsetneq \mathbb{R}$ for the set of all computable real numbers, and $\mathbb{C}_{\mathrm{c}} = \{z = x + \mathrm{i}y : x, y \in \mathbb{R}_{\mathrm{c}}\}$ is the set of all computable complex numbers.

For a given $t \in \mathbb{R}$, the Turing machine TM in Definition 2.1 will generally need several iterations to calculate $\gamma(n) \in \mathbb{Q}$, and it will usually require more iterations (i.e. computation time) to determine $\gamma(n)$ if n increases. To quantify this behavior, one defines the *complexity* for computing the number $t \in \mathbb{R}$.

Definition 2.2. Let $t \in \mathbb{R}_{\mathrm{c}}$ be a computable number. We say that the *computational complexity* of t is bounded by a function $q : \mathbb{N} \to \mathbb{N}$ if there exists a Turing machine TM such that (3) is satisfied after at most $q(n)$ iterations. The number $t \in \mathbb{R}_{\mathrm{c}}$ is said to be *polynomial-time computable* if its computational complexity is bounded by a polynomial q.

We also need the notion of *computable functions* and some measures to assess the complexity of computing functions on digital computers. To this end, we use two different machine models which both give a slightly different view on the problem. Both models are equivalent, however, each model has different advantages for the understanding of computability in prediction theory and Wiener filtering. These two computation models are shortly introduced and discussed in the following paragraphs. Thereby, we mainly follow the presentations in [7, 18, 24].

Function-oracle Turing machines. A function-oracle Turing machines is an ordinary Turing machine TM, but with an additional *function-oracle* γ that is able to calculate the function value γ in a single operation. The computation of the function-oracle machine TM on input t with oracle γ is written as $\mathrm{TM}_\gamma(t)$.

Definition 2.3. A function $f : [a, b] \to \mathbb{R}$ is said to be *computable* on the interval $[a, b] \subseteq \mathbb{R}$ if there exists a function-oracle Turing machine TM so that for each $t \in [a, b]$ and for each γ that binary

converges to t, the function $\widetilde{f}(n) = \mathrm{TM}_\gamma(n)$ computed by TM with oracle γ binary converges to $f(t)$, i.e.,

$$|f(t) - \mathrm{TM}_\gamma(n)| < 2^{-n}\,, \qquad \text{for all } n \in \mathbb{N}\,. \tag{4}$$

So, any computable function is defined by a function-oracle Turing machine in which the real numbers at the input and output are replaced by rational approximations that binary converge to the exact values. Note that any computable function $f : [a,b] \to \mathbb{R}$ is necessarily continuous (see, e.g., [31]), i.e., $f \in \mathcal{C}([a,b])$.

There exist many different but equivalent definitions for computable functions. However, based on Definition 2.3, one can also conveniently characterize the complexity for computing a function.

Definition 2.4. Let $f : [a,b] \to \mathbb{R}$ be computable. We say that the *complexity of f is bounded by a function $q : \mathbb{N} \to \mathbb{N}$* if there exists a function–oracle Turing machine TM that computes f so that for all γ that binary converge to a $t \in [0,1]$, $\mathrm{TM}_\gamma(n)$ satisfies (4) after a computation time of at most $q(n)$. We say that $f : [0,1] \to \mathbb{R}$ is *polynomial-time computable* if its complexity is bounded by a polynomial q.

In principle, a Turing machine can exactly compute only with rational numbers. However, later we consider Turing machines TM whose inputs are not just rational numbers but computable numbers x or even computable functions f. It should keep clearly in mind that these inputs are given to a Turing machine in the form of a program (i.e. in the form of a *description*) that allows TM to compute the input effectively. So, if we want to give a computable function f to a Turing machine TM, then we hand over a description of f, i.e. a program which can be executed on TM and which is able to *effectively compute* for every computable t in the domain of f a rational approximation of $f(t)$ up to any necessary precision.

Computation on dyadic grids. We also consider another model to characterize the complexity of computing functions f on an interval $[a,b]$. For simplicity and without loss of generality, we discuss here only functions on $[0,1]$. In this approach, we restrict *a priori* the points t for which $f(t)$ is computed to a discrete set $\mathcal{D}_1 \subset [0,1] \cap \mathbb{R}_\mathrm{c}$. Namely, $\mathcal{D}_1$ will be the set of *dyadic decimals*, i.e. rational numbers in the interval $[0,1]$ of the form

$$t_{j,n} = j \cdot 2^{-n}\,, \qquad 0 \le j \le 2^n\,,$$

for some $n \in \mathbb{N}$. Next, we give the formal definitions for this model of computation. All of them are basically taken from [18]. We start with the definition of $\mathcal{D}_1$.

Definition 2.5. A *dyadic decimal* consists of the symbol '+' or '−' followed by a (possibly empty) string of 0's and 1's which starts (if it is nonempty) with 1, followed by a decimal point, followed by a second (nonempty) string of 0's and 1's.
The set of all dyadic decimals is denoted by $\mathcal{D}$ and $\mathcal{D}_1$ denotes all $d \in \mathcal{D}$ which begin with '+' and for which the string to the left of the decimal point is empty.

If $d \in \mathcal{D}$, then $\mathrm{tnd}(d)$ stands for the total number of digits in d, and $\mathrm{pcs}(d)$ is the total number of digits to the right of the decimal point in d. In this computing model, dyadic decimals d are used to approximate arbitrary real numbers $t \in \mathbb{R}$. Then $\mathrm{pcs}(d)$ says something about the precision with that t is approximated by d, and for any $t \in \mathbb{R}$ we write $d \sim t$ if $|d - t| \le 2^{-\mathrm{pcs}(d)}$.

Definition 2.6. A *dyadic approximation function* is a pair (g,i) with $i \in \mathbb{N}$ and with a function $g : \mathcal{D}_1 \to \mathcal{D}$, so that for all $n \in \mathbb{N}$, there exists an $m \in \mathbb{N}$ such that for every $d \in \mathcal{D}_1$ with $\mathrm{tnd}(d) \ge m$, one has $\mathrm{tnd}(g(d)) \ge n$.
We say that a dyadic approximation function (g,i) approximates a function $f : [0,1] \to \mathbb{R}$ if and only if for all $t \in [0,1]$ and all $d \in \mathcal{D}_1$ with $\mathrm{tnd}(d) \ge i$ and $d \sim t$, we have $g(d) \sim f(t)$. In this case, we write $(g,i) \sim f$.

Remark 2.2. If for a function $f : [0,1] \to \mathbb{R}$ there exists a dyadic approximation function, then there exist a $g : \mathcal{D}_1 \to \mathcal{D}$ and an $i \in \mathbb{N}$ so that $(g,i) \sim f$, according to Definition 2.6. The function g has the property that for a required precision n for calculating the values of g, one can always find

a sufficiently fine dyadic grid (of width 2^{-m}) in the domain $[0,1]$ so that evaluating g on this grid determines g up to a precision of at least n. The positive integer $i \in \mathbb{N}$ determines for which $d \in \mathcal{D}_1$ the function g is an approximation of f. For every $t \in [0,1]$, this has to be the case only for such $d \in \mathcal{D}_1$ which satisfy $d \sim t$ and $\operatorname{tnd}(d) \geq i$, i.e. for $d \in \mathcal{D}_1$ satisfying $|d-t| \leq 2^{-i}$.

Definition 2.7. Let h be a function on $\mathbb{N}$ and let (g,i) be a dyadic approximation function of f. We call h a *modulus* for (g,i) if for every $n \in \mathbb{N}$ and all $d \in \mathcal{D}_1$ with $\operatorname{tnd}(d) \geq h(n)$ holds $\operatorname{tnd}(f(d)) \geq n$. If h is a polynomial, then we say that (g,i) has a *polynomial modulus*.

So, the modulus of a dyadic approximation function characterizes the necessary precision in the domain of f to guarantee a certain precision in the range of g.

We say that a function $g : \mathcal{D}_1 \to \mathcal{D}$ is *polynomial-time computable* if there exist a polynomial q and a Turing machine TM with the input $d \in \mathcal{D}_1$ and output $\mathrm{TM}(d) = g(d)$ such that the computation time is at most $q(\operatorname{tnd}(d))$ for every $d \in \mathcal{D}_1$. This definition of "polynomial-time computable" now incorporates the computational complexity for approximating the number $t \in [0,1]$ by a dyadic decimal $d \in \mathcal{D}_1$. If a high precision for d is necessary (this depends on g), then $\operatorname{tnd}(d)$ and, consequently, the maximal computation time $q(\operatorname{tnd}(d))$ is large. Based on this definition of polynomial-time complexity for dyadic approximation functions, one can now define polynomial-time complexity for arbitrary functions on $[0,1]$.

Definition 2.8. A function $f : [0,1] \to \mathbb{R}$ is said to be *polynomial-time computable on dyadic grids* if there exists a polynomial-time computable function $g : \mathcal{D}_1 \to \mathcal{D}$ and an $i \in \mathbb{N}$ so that (g,i) has a polynomial modulus, and $(g,i) \sim f$.

We see that the complexity of approximating $t \in [0,1] \cap \mathbb{R}_{\mathrm{c}}$ by a dyadic decimal $d \in \mathcal{D}_1$ is included in the definition of a polynomial-time computable function. However, if the dyadic approximation function f is polynomial-time computable *and* if the necessary precision for the argument d grows at most polynomially with the required precision for determining g, the function g is polynomial-time computable.

3. Prediction Theory

We recall shortly the main concepts and notation from prediction theory and refer to many excellent textbooks (see, e.g., [10, 27, 30, 34, 37]) for more details. In particular, we follow the recent overview paper [3] to present the main results in the following subsections.

3.1. Stationary stochastic processes.

Let $(\Omega, \mathscr{F}, \nu)$ be a probability space, and let $\mathcal{R} = \mathcal{R}(\Omega, \mathscr{F}, \nu)$ be the space of all (complex) random variables (rvs) x with zero mean $\mathrm{E}[x] = \int_\Omega x(\omega)\,\mathrm{d}\nu(\omega) = 0$ and finite second moments $\mathrm{E}\big[\,|x|^2\,\big] < \infty$. This space $\mathcal{R}$ becomes a Hilbert space if the inner product is defined as the *covariance* of two rvs, i.e.,

$$\langle x, y\rangle_{\mathcal{R}} = \operatorname{cov}(x,y) = \mathrm{E}[x\overline{y}] = \int_\Omega x(\omega)\,\overline{y(\omega)}\,\mathrm{d}\nu(\omega)\ ,$$

with the corresponding norm $\|x\|_{\mathcal{R}} = \sqrt{\mathrm{E}[|x|^2]}$. The *correlation* of two rvs $x, y \in \mathcal{R}$ is defined to be

$$\operatorname{corr}(x,y) = \frac{\langle x, y\rangle_{\mathcal{R}}}{\|x\|_{\mathcal{R}} \cdot \|y\|_{\mathcal{R}}} = \frac{\operatorname{cov}(x,y)}{\left(E[|x|^2] \cdot E[|y|^2]\right)^{1/2}}\ .$$

A sequence $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}} \subset \mathcal{R}$ is said to be a *wide-sense stationary (wss) stochastic process* if $\langle x_{n+k}, x_k\rangle_{\mathcal{R}} = \langle x_n, x_0\rangle_{\mathcal{R}}$ for all $n, k \in \mathbb{Z}$, and the corresponding function

$$\gamma_{\boldsymbol{x}}(n) = \langle x_n, x_0\rangle_{\mathcal{R}}\ , \qquad n \in \mathbb{Z}\ ,$$

is said to be the *auto-covariance function* of $\boldsymbol{x}$. If $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}} \subset \mathcal{R}$ is a wss stochastic process, then there exists an orthogonal stochastic measure $Z_{\boldsymbol{x}} = Z_{\boldsymbol{x}}(\omega)$, $\omega \in \mathscr{B}(\mathbb{T})$, on the Borel sets of $\mathbb{T}$ such

that

$$x_n = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta} \mathrm{d}Z_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) \,, \qquad \text{for all } n \in \mathbb{Z} \,,$$

and the auto-covariance has the *spectral representation*

$$\gamma_{\boldsymbol{x}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta} \mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) \,, \qquad n \in \mathbb{Z} \,, \tag{5}$$

with the finite *spectral measure* $\mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = 2\pi \left\| \mathrm{d}Z_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) \right\|_{\mathcal{R}}^2$. As usual, this measure can be decomposed as

$$\mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = \varphi(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta + \mathrm{d}\mu_{\mathrm{s}}(\mathrm{e}^{\mathrm{i}\theta}) \tag{6}$$

with the so-called *spectral density* $\varphi \in L^1(\mathbb{T})$ of $\boldsymbol{x}$ and where μ_{s} is the singular part of $\mu_{\boldsymbol{x}}$ (with respect to the Lebesgue measure).

For any wss stochastic process $\boldsymbol{x}$, we write $\mathcal{X} = \overline{\mathrm{span}}\{x_n : n \in \mathbb{Z}\} \subset \mathcal{R}$ for the Hilbert space spanned by the whole stochastic process $\boldsymbol{x}$. For any two $n, m \in \mathbb{Z}$ with $n \leq m$, we write $\mathcal{X}_{[n,m]} = \mathrm{span}\{x_n, x_{n+1}, \ldots, x_{m-1}, x_m\}$ for the subspace spanned by $\{x_n, \ldots, x_m\}$. In particular, we write

$$\mathcal{X}_n = \mathcal{X}_{[-\infty,n]} = \overline{\mathrm{span}}\,\{x_k : k \leq n\} \qquad \text{and} \qquad \mathcal{X}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{X}_n$$

for the subspace spanned by the past up to time n and their intersection (called the *remote past*) respectively. A wss stochastic process $\boldsymbol{x}$ is called *singular* if $\mathcal{X} = \mathcal{X}_{-\infty}$, and *regular* if $\mathcal{X}_{-\infty} = \{0\}$.

3.2. **Linear prediction.** A classical problem is to find the best linear predictor $\widehat{x}_n$ of x_n from finite (or infinite) many observations of the sequence $\boldsymbol{x}$. Without loss of generality, we discuss only the prediction of x_0. Moreover, we consider only the prediction of x_0 from the past observations of $\boldsymbol{x}$, i.e. from the observations of $\{\ldots, x_{-3}, x_{-2}, x_{-1}\}$. If, for example, one wants to predict x_0 from all past observations, then the optimal linear prediction is given by

$$\widehat{x}_0 = \operatorname*{arg\,min}_{x \in \mathcal{X}_{[-\infty,-1]}} \|x - x_0\|_{\mathcal{R}}^2 = \mathrm{P}_{[-\infty,-1]}(x_0) \,, \tag{7}$$

where $\mathrm{P}_{[-\infty,-1]} : \mathcal{X} \to \mathcal{X}_{[-\infty,-1]}$ denotes the orthogonal projection onto $\mathcal{X}_{[-\infty,-1]} \subset \mathcal{X}$. The resulting *mean squared error (MSE)* is then given by

$$\sigma^2 = \|x_0 - \widehat{x}_0\|_{\mathcal{R}}^2 = \mathrm{E}\big[\, |x_0 - \widehat{x}_0|^2 \,\big] = \mathrm{E}\Big[\, \big|x_0 - \mathrm{P}_{[-\infty,-1]}(x_0)\big|^2 \,\Big] \,.$$

The stochastic process $\boldsymbol{x}$ is called *deterministic* if $\sigma^2 = 0$, because then x_0 can be perfectly predicted from the past observation. Otherwise, if $\sigma^2 > 0$, the process $\boldsymbol{x}$ is said to be *non-deterministic*. We will consider only non-deterministic stochastic processes and the following result gives the necessary and sufficient conditions on the spectral measure for $\boldsymbol{x}$ to be non-deterministic.

Theorem 3.1. *Let $\boldsymbol{x}$ be a wss stochastic sequence with spectral measure* (6)*. Then $\boldsymbol{x}$ is non-deterministic if and only if $\log \varphi \in L^1(\mathbb{T})$, i.e., if and only if*

$$\int_{-\pi}^{\pi} \log \varphi(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta > -\infty \,. \tag{8}$$

In this case, the minimum mean squared error is given by

$$\sigma^2 = \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log \varphi(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta \right) > 0 \,. \tag{9}$$

Remark 3.1. Condition 8, is also known as *Szegö's condition* [40], whereas (9) is known as *Kolmogorov's formula* [38].

Theorem 3.1 implies in particular that the spectral measure of a non-deterministic wss stochastic process has necessarily a non-vanishing spectral density. The so-called *Wold decomposition* shows that any wss stochastic sequence $\boldsymbol{x}$ has a unique decomposition $\{x_n\} = \{x_n^{\mathrm{r}}\} + \{x_n^{\mathrm{s}}\}$ into a non-deterministic (or regular) sequence $\boldsymbol{x}^{\mathrm{r}} = \{x_n^{\mathrm{r}}\}$ and a deterministic (or singular) sequence $\boldsymbol{x}^{\mathrm{s}} = \{x_n^{\mathrm{s}}\}$. Then the spectral measure of $\boldsymbol{x}^{\mathrm{r}}$ is the absolute continuous part of $\mu_{\boldsymbol{x}}$, whereas the spectral measure of $\boldsymbol{x}^{\mathrm{s}}$ is the singular measure μ_{s} of $\mu_{\boldsymbol{x}}$. Here, we consider only non-deterministic sequences for which the singular (i.e. the deterministic) part is identical with zero. Such sequences are called *purely non-deterministic*. In particular, we always subsequently assume that the spectral measure $\mu_{\boldsymbol{x}}$ is completely determined by the spectral density φ and that φ satisfies Szegö's condition (8).

3.3. **Prediction filters.** We now discuss shortly the relation to filter design in signal processing. To this end, we always assume in this section that the spectral density φ of a purely non-deterministic wss stochastic process $\boldsymbol{x}$ is a continuous function on $\mathbb{T}$ and satisfies $\min\limits_{\zeta\in\mathbb{T}} |\varphi(\zeta)| > 0$.

From a signal processing point of view, the linear predictor (7) has the form of a causal linear filter

$$\widehat{x}_0 = \mathrm{H}(\boldsymbol{x}) = \sum_{k=1}^{\infty} h_k x_{-k}, \tag{10}$$

where $\{h_k\}_{k=1}^{\infty}$ is the *impulse response* of H and $h(\mathrm{e}^{\mathrm{i}\theta}) = \sum\limits_{k=1}^{\infty} h_k\, \mathrm{e}^{\mathrm{i}k\theta}$ is said to be the *transfer function* of H. Since the input of the filter is a stationary stochastic process with auto-covariance $\gamma_{\boldsymbol{x}}$ and spectral density $\varphi_{\boldsymbol{x}}$, the sum in (10) converges in mean squared (i.e. in the norm of $\mathcal{R}$) if the impulse response $\{h_k\}_{k\in\mathbb{N}}$ satisfies

$$\|\widehat{x}_0\|_{\mathcal{R}}^2 = \mathrm{E}\big[\,|\widehat{x}_0|^2\,\big] = \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} h_k \overline{h_\ell}\, \gamma_{\boldsymbol{x}}(\ell - k) < \infty\,, \tag{11}$$

which translates into the condition $h \in L^2(\mu_{\boldsymbol{x}})$, i.e.,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \big|h(\mathrm{e}^{\mathrm{i}\theta})\big|^2\, \mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) < \infty\,,$$

for the transfer function of the filter H. Moreover, since H is a causal filter, all filter coefficients h_k with $k \le 0$ are equal to zero, and so, the transfer function h can be extended to an analytic function $h(z) = \sum_{k=1}^{\infty} h_k z^k$ in the unit disk that satisfies $h(0) = 0$. Consequently, we have to require that the transfer function h of the linear filter (10) belongs at least to $H_0(\mathbb{D}) \cap L^2(\mu_{\boldsymbol{x}})$.

Now, with exactly the same arguments as before, we see that for a fixed transfer function $h \in H_0(\mathbb{D}) \cap L^2(\mu_{\boldsymbol{x}})$ the *mean squared error (MSE)* is given by

$$\begin{aligned}\sigma^2(h) &= \|x_0 - \widehat{x}_0\|_{\mathcal{R}}^2 = \mathrm{E}\big[\,|x_0 - \widehat{x}_0(h)|^2\,\big] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \big|1 - h(\mathrm{e}^{\mathrm{i}\theta})\big|^2\, \mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = \|1 - h\|_{L^2(\mu_{\boldsymbol{x}})}^2\,. \end{aligned} \tag{12}$$

Then one can always find a unique $h_{\mathrm{opt}} \in H_0^\infty(\mathbb{D})$ that minimizes the MSE, i.e. so that

$$\begin{aligned}\sigma^2 = \sigma^2(h_{\mathrm{opt}}) &= \mathrm{E}\left[|x_0 - \widehat{x}_0(h_{\mathrm{opt}})|^2\right] \\ &= \inf_{h\in H_0(\mathbb{D})} \mathrm{E}\left[|x_0 - \widehat{x}_0(h)|^2\right] = \inf_{h\in \mathcal{A}_0(\mathbb{D})} \mathrm{E}\left[|x_0 - \widehat{x}_0(h)|^2\right]\,. \end{aligned} \tag{13}$$

For the last equation, we have used the fact that $\mathcal{A}_0(\mathbb{D})$ is a dense subset of $H_0(\mathbb{D})$, so that it is sufficient to optimize over $\mathcal{A}_0(\mathbb{D})$ which is a much more suitable space for practical applications.

However, it should be noted that the minimizer h_{opt} is generally not in $\mathcal{A}_0(\mathbb{D})$ but only in $H_0^\infty(\mathbb{D})$. So, (12) shows that the minimal MSE is given by

$$\sigma^2 = \inf_{h\in H_0(\mathbb{D})} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|1-h(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = \inf_{h\in\mathcal{A}_0(\mathbb{D})} \|1-h\|^2_{L^2(\mu_{\boldsymbol{x}})} \ . \tag{14}$$

From a practical point of view, the optimal filter $h_{\text{opt}} \in H_0^\infty(\mathbb{D})$ is usually approximated by a certain FIR filter of a certain length N. This means that $h_k = 0$ for all $k > N$ in (10). Then the optimal MSE for all FIR filters of length N is obtained by optimizing in (14) over $h \in \mathcal{P}_{N,0}$. However, it should be noted that (14) is usually not very useful for finding the optimal filter h_{opt} (or a certain FIR approximation), since it is not clear how the minimizer in (14) can actually be calculated. Therefore the following subsection shortly reviews the methods for the determination of h_{opt} and corresponding FIR approximations.

3.4. **Spectral factorization.** One method to derive a closed form expression for the transfer function h_{opt} of the optimal prediction filter is based on the spectral factorization of the density φ of the regular stochastic process $\boldsymbol{x}$.

Definition 3.1. A non-negative function $\varphi \in L^1(\mathbb{T})$ is said to possess a *spectral factorization* if there exists a $\varphi_+ \in H^1(\mathbb{D})$ with $\varphi_+(z) \neq 0$ for all $z \in \mathbb{D}$ and so,

$$\varphi(\mathrm{e}^{\mathrm{i}\theta}) = \left|\varphi_+(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \qquad \text{for almost all } \theta \in [-\pi,\pi) \ . \tag{15}$$

The function φ_+ is said to be the *spectral factor* of φ.

Remark 3.2. The spectral factor φ_+ is a so-called *outer function*, i.e. it is analytic and non-zero in $\mathbb{D}$, and it can be written as a power series as $\varphi_+(z) = \sum_{n=0}^{\infty} a_n z^n$. Moreover, since φ_+ is an outer function, also $1/\varphi_+$ is an outer function.

Remark 3.3. It is often useful to define $\varphi_-(z) = \overline{\varphi_+(\overline{z}^{-1})} = \sum_{n=0}^{\infty} \overline{a}_n z^{-n}$. This function is analytic and non-zero for all $|z| > 1$. Therewith (15) can be written as $\varphi(\mathrm{e}^{\mathrm{i}\theta}) = \varphi_+(\mathrm{e}^{\mathrm{i}\theta})\varphi_-(\mathrm{e}^{\mathrm{i}\theta})$.

The following well known statement provides a necessary and sufficient condition such that a spectral density possesses a spectral factorization and it gives a closed form expression for the corresponding spectral factor.

Theorem 3.2. *A spectral density $\varphi \in L^1(\mathbb{T})$ possesses a spectral factorization if and only if φ satisfies Szegö's condition* (8)*. Then its spectral factor is given by*

$$\varphi_+(z) = \exp\left(\frac{1}{4\pi}\int_{-\pi}^{\pi} \log\varphi(\mathrm{e}^{\mathrm{i}\theta})\frac{\mathrm{e}^{\mathrm{i}\theta}+z}{\mathrm{e}^{\mathrm{i}\theta}-z}\,\mathrm{d}\theta\right), \qquad z\in\mathbb{D}\,,$$

where the integral is a Cauchy principal value integral. The spectral factor φ_+ is unique up to a unitary factor.

Remark 3.4. Note that in connection with the design of optimal causal linear filters, condition (8), required in Theorem 3.2, is often called *Paley–Wiener condition*. Moreover, Kolmogorov's formula (9) may be expressed in terms of the spectral factor as $\sigma^2 = |\varphi_+(0)|^2$.

So, if $\boldsymbol{x}$ is a non-deterministic wss stochastic process, then its spectral density φ satisfies Szegö's condition (cf., Theorem 3.1) and therefore the corresponding spectral factor φ_+ always exists. Based on this spectral factor, one can now give a simple expression for the optimal prediction filter h_{opt}.

Theorem 3.3. *Let $\boldsymbol{x}$ be a purely non-deterministic wss stochastic process with spectral density φ, then the optimal prediction filter for estimating x_0 from the past is given as*

$$h_{\text{opt}}(z) = \frac{\varphi_+(z)-\varphi_+(0)}{\varphi_+(z)} = 1 - \frac{\varphi_+(0)}{\varphi_+(z)}\,, \tag{16}$$

where φ_+ is the spectral factor of φ.

Remark 3.5. In view of (14) and by the orthogonality principle in Hilbert spaces, one has only to show that $1 - h_{\mathrm{opt}}$ is orthogonal (in $L^2(\varphi)$) to all monomials $e_m(z) = z^m$, with $m \geq 1$, to prove this theorem.

Since φ_+ is an outer function, i.e. analytic and non-zero in $\mathbb{D}$, its inverse φ_+^{-1} is again analytic in $\mathbb{D}$ and so, (16) shows that h_{opt} is an analytic function in $\mathbb{D}$, satisfying $h_{\mathrm{opt}}(0) = 0$. Thus $h_{\mathrm{opt}} \in H_0(\mathbb{D})$ and so, it generally defines an *infinite impulse response* (IIR) prediction filter (10). If in addition $\min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0$, as we will suppose most of the time, then $h_{\mathrm{opt}} \in H_0^\infty(\mathbb{D})$ and so, the filter coefficients in (10) are given as the Fourier coefficients of h_{opt}, i.e. $h_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_{\mathrm{opt}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{e}^{-\mathrm{i}k\theta}\,\mathrm{d}\theta$. However, from a practical point of view, only the FIR filters of the form

$$\widehat{x}_0 = \mathrm{H}_N(\boldsymbol{x}) = \sum_{k=1}^{N} h_k x_{-k}$$

can be implemented. Such an FIR approximation might be obtained from the optimal IIR filter (10) by truncating the infinite sum at a certain degree N. Then there arises the question whether it is possible to determine algorithmically the necessary filter length N in order to guarantee a certain predefined error 2^{-M}. So, the problem that we will study in detail in Section 4 is as follows: *Given a wss stochastic process $\boldsymbol{x}$ with spectral density φ and a precision $M \in \mathbb{N}$, is it then possible to find an algorithm that determines $N \in \mathbb{N}$ so that $\|\mathrm{H}\boldsymbol{x} - \mathrm{H}_N\boldsymbol{x}\|_{\mathcal{R}}^2 < 2^{-M}$?*

3.5. **Partial auto-correlations and the Durbin–Levinson algorithm.** Another approach to determine the optimal prediction filter is based on the so-called *partial autocorrelation function*. For a wss stochastic process $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}} \subset \mathcal{R}$, it is define by

$$\alpha_{\boldsymbol{x}}(n) = \mathrm{corr}\left(x_n - \mathrm{P}_{[1,n-1]}x_n\ ,\ x_0 - \mathrm{P}_{[1,n-1]}x_0\right)\ , \qquad n = 1, 2, 3, \ldots\ . \tag{17}$$

So, $\alpha_{\boldsymbol{x}}(n)$ is the correlation between the prediction error at time n and 0 resulting from the optimal prediction of these values from the intermediate values $x_1, \ldots, x_{n-1}$. These values are also known as *Verblunsky coefficients* of $\boldsymbol{x}$. Their importance streams to a large extend from *Verblunsky's Theorem*.

Theorem 3.4. *There is a bijective relation between the sequences $\boldsymbol{\alpha} = \{\alpha(n)\}_{n\in\mathbb{N}}$ with $\alpha(n) \in \mathbb{D}$ and the probability measures μ on $\mathbb{T}$.*

Remark 3.6. So, any arbitrary sequence $\{\alpha_{\boldsymbol{x}}(n)\}_{n\in\mathbb{N}}$, which is only restricted by the obvious condition $|\alpha_{\boldsymbol{x}}(n)| < 1$ for all $n \in \mathbb{N}$, is related to a probability measure $\mu_{\boldsymbol{x}}$ on $\mathbb{T}$, which in turn characterizes a wss stochastic process. In this sense, one has an *unrestricted parametrization* for the probability measures on $\mathbb{T}$.

In contrast to this, we notice that also the spectral representation (5) of the auto-covariance function $\boldsymbol{\gamma}_{\boldsymbol{x}} = \{\gamma_{\boldsymbol{x}}(n)\}_{n\in\mathbb{Z}}$ gives a bijection between sequences $\boldsymbol{\gamma}_{\boldsymbol{x}}$ and probabilities measures $\mu_{\boldsymbol{x}}$ on $\mathbb{T}$. However, in this case, the sequence $\boldsymbol{\gamma}_{\boldsymbol{x}}$ is restricted by the condition to be positive definite (cf., (11)). So, the auto-covariance functions provide only a *restricted parametrization* for the probability measures on $\mathbb{T}$.

The partial auto-correlation function plays a fundamental role in the so-called Durbin–Levinson algorithm for determining the best causal prediction. To explain this algorithm, let $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}}$ be the wss stochastic process and let

$$\widehat{x}_{n+1} = h_{n,1}x_n + \cdots + h_{n,n}x_1 = \sum_{k=1}^{n} h_{n,k}x_{n-k+1}$$

be the best linear predictor of x_{n+1} based on $x_n, \cdots, x_1$, we write $\mathbf{h}_n = (h_{n,1}, \cdots, h_{n,n})^{\mathrm{T}}$ for the vector containing n coefficients of this predictor and denote by $\sigma_n^2 = \mathrm{E}\big[\,|x_{n+1} - \widehat{x}_{n+1}|^2\,\big]$ the corresponding MSE at time n. The *Durbin-Levinson algorithm* [11,28] determines now the coefficients $\mathbf{h}_{n+1}$ and the MSE σ_{n+1}^2 at time $n+1$ recursively based on the data at time n as follows:

$$h_{n+1,n+1} = \frac{1}{\sigma_n^2}\left[\gamma_{\boldsymbol{x}}(n+1) - \sum_{j=1}^{n} h_{n,j}\gamma_{\boldsymbol{x}}(n-j)\right]$$

and

$$\begin{pmatrix} h_{n+1,1} \\ \vdots \\ h_{n+1,n} \end{pmatrix} = \begin{pmatrix} h_{n,1} \\ \vdots \\ h_{n,n} \end{pmatrix} - h_{n+1,n+1} \begin{pmatrix} h_{n,n} \\ \vdots \\ h_{n,1} \end{pmatrix} . \tag{18}$$

Moreover, starting with $\sigma_0^2 = 1$, the MSE is determined recursively by

$$\sigma_{n+1}^2 = \sigma_n^2 \left(1 - |h_{n+1,n+1}|^2\right) . \tag{19}$$

We also remark that the coefficients $h_{n,n}$ are equal to the Verblunsky coefficients $\alpha_{\boldsymbol{x}}(n)$ (17), i.e. $h_{n,n} = \alpha_{\boldsymbol{x}}(n)$ for all $n \in \mathbb{N}$. So, in particular, (18) gives

$$h_{n,k} - h_{n+1,k} = \alpha_{\boldsymbol{x}}(n+1)\, h_{n,n-k+1} , \quad k = 1, 2, \ldots, n ,$$

and (19) implies that the optimal MSE at step n is given by

$$\sigma_n^2 = \prod_{k=1}^{n} \left(1 - |\alpha_{\boldsymbol{x}}(k)|^2\right) . \tag{20}$$

This shows, in particular, that the sequence $\left\{\sigma_n^2\right\}_{n\in\mathbb{N}}$ is monotonically decreasing and so it converges as $n \to \infty$ and one can show that it converges to the minimal MSE (13), i.e., $\lim\limits_{n\to\infty} \sigma_n^2 = \sigma^2$.

The Durbin–Levinson algorithm is a very efficient method to determine recursively the filter coefficients of the optimal prediction filter. However, here also arises the question whether it is possible to have an algorithmic procedure to stop the iteration of the Durbin–Levinson algorithm at an appropriate index N if for a predefined precision $M \in \mathbb{N}$ the approximation error for the MSE (20) satisfies

$$\left|\sigma_N^2 - \sigma^2\right| < 2^{-M} .$$

In other words: *Is it possible to find an algorithmic stopping criterion for the Durbin–Levinson algorithm that is able to control at each step the quality of the actual approximation?* We are going to show in Section 4.1 that, in general, no such stopping criterion exists, neither for the Durbin–Levinson algorithm, nor for any other possible algorithm that approximates the optimal MSE σ^2.

4. Computation of the Mean Squared Prediction Error

Let $\boldsymbol{x}$ be a purely non-deterministic wss stationary process with spectral density φ. According to Theorem 3.1, φ satisfies Szegö's condition (8). However, in the following we restrict our considerations to stochastic processes with very well behaving spectral density. Concretely, we consider the set

$$\mathcal{M}_{\mathrm{D}} = \left\{\varphi \in \mathcal{C}_{\mathrm{c}}(\mathbb{T}) : \varphi' \in \mathcal{C}_{\mathrm{c}}(\mathbb{T}) \text{ and } \log\varphi \in L^1(\mathbb{T})\right\} ,$$

of spectral densities φ that are computable continuous functions on $\mathbb{T}$ with a first derivative φ' that is also a computable continuous function on $\mathbb{T}$. Then we define the functional $E_\infty : \mathcal{M}_{\mathrm{D}} \to \mathbb{R}_+$ given by

$$\sigma^2 = E_\infty(\varphi) = \inf_{h\in\mathcal{A}_0(\mathbb{D})} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - h(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \varphi(\mathrm{e}^{\mathrm{i}\theta})\, \mathrm{d}\theta . \tag{21}$$

As discussed in Section 3.3, $\sigma^2 = E_\infty(\varphi)$ is the optimal MSE that can be obtained by predicting x_0 of a regular wss stochastic process $\boldsymbol{x}$ with spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ from its past $\{\ldots, x_{-3}, x_{-2}, x_{-1}\}$ using a causal linear filter.

4.1. Non-computability of the optimal MSE.

It is a problem of considerable practical importance to calculate the optimal MSE $\sigma^2 = E_\infty(\varphi)$ given by (21) for arbitrary $\varphi \in \mathcal{M}_{\mathrm{D}}$. Even though $\mathcal{M}_{\mathrm{D}}$ contains already only spectral densities with very nice analytic properties, it is clear that even for fairly simple densities $\varphi \in \mathcal{M}_{\mathrm{D}}$, (21) cannot be calculated in a closed form and so, one generally needs numerical algorithms to calculate $E_\infty(\varphi)$. This brings us to the following problem.

Question 1. Does there exist a Turing machine TM whose input is a description of the spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ and which calculates a description of the number $E_\infty(\varphi) \in \mathbb{R}_+$?

In (21), we optimize over all linear filters $h \in \mathcal{A}_0(\mathbb{D})$. From a practical point of view, it might be sufficient to ask for an FIR prediction filter. This yields the simpler optimization problem

$$\sigma_n^2 = E_n(\varphi) = \inf_{p \in \mathcal{P}_{n,0}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - p(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \varphi(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta \,, \tag{22}$$

where we minimize only over all polynomials of maximum degree n. Then $E_n(\varphi)$ is the minimal MSE that can be achieved by predicting x_0 of a regular wss stochastic process with spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ from its past $\{x_{-n}, \ldots, x_{-3}, x_{-2}, x_{-1}\}$ by using a causal linear FIR filter of length n. It is clear from the definition that $E_{n+1}(\varphi) \leq E_n(\varphi)$ for all $n \in \mathbb{N}$ and that

$$\lim_{n \to \infty} E_n(\varphi) = E_\infty(\varphi) \,, \qquad \text{for all } \varphi \in \mathcal{M}_{\mathrm{D}} \,.$$

It seems an obvious idea, to use the sequence $\{E_n(\varphi)\}_{n \in \mathbb{N}}$ of simpler functionals to approximate the value $E_\infty(\varphi)$. However, to make this idea effective, one needs an algorithmic stopping criterion, i.e. one needs a Turing machine that is able to find algorithmically for any $M \in \mathbb{N}$ and any given $\varphi \in \mathcal{M}_{\mathrm{D}}$ an index $N \in \mathbb{N}$ such that $|E_N(\varphi) - E_\infty(\varphi)| < 2^{-M}$. This brings us to our second problem.

Question 2. Does there exist a Turing machine TM whose two inputs are an arbitrary description of $\varphi \in \mathcal{M}_{\mathrm{D}}$ and a number $M \in \mathbb{N}$ and that is able to calculate an index $N = N(\varphi, M) \in \mathbb{N}$ such that $\left|E_{N(\varphi,M)}(\varphi) - E_\infty(\varphi)\right| < 2^{-M}$?

We are going to show that both, Question 1 and Question 2, have a negative answer. This will follow from the following

Theorem 4.1. *There exists a $\varphi_* \in \mathcal{M}_{\mathrm{D}}$ so that $E_\infty(\varphi_*) \notin \mathbb{R}_{\mathrm{c}}$.*

Remark 4.1. The proof of this result will be given in a forthcoming publication.

Theorem 4.1 immediately implies the following practical consequences concerning our two questions.

Corollary 4.2. *Let $\varphi_* \in \mathcal{M}_{\mathrm{D}}$ be the spectral density as in Theorem* 4.1*. There exists no Turing machine that is able to compute $E_\infty(\varphi_*)$. In particular, Question* 1 *has a negative answer.*

Note that the negative answer to Question 1 is the strongest possible. There not only exists Turing machine $\mathrm{TM}(\varphi)$ whose input is a (description of a) spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ and that is able to calculate (a description of) $E_\infty(\varphi)$ for all possible spectral densities $\varphi \in \mathcal{M}_{\mathrm{D}}$, it is even impossible to have a particular Turing machine TM_* that is able to compute $E_\infty(\varphi_*)$ for only one spectral density $\varphi_* \in \mathcal{M}_{\mathrm{D}}$.

Corollary 4.3. *For the spectral density $\varphi_* \in \mathcal{M}_{\mathrm{D}}$ as in Theorem* 4.1 *there does not exist a Turing machine* TM *whose input is a number $M \in \mathbb{N}$ and that is able to determine $N \in \mathbb{N}$ so that $\left|\sigma_N^2 - \sigma^2\right| = |E_N(\varphi_*) - E_\infty(\varphi_*)| < 2^{-M}$. In particular, Question* 2 *has a negative answer.*

So, even although $E_n(\varphi)$ monotonically converges to $E_\infty(\varphi)$ as $n \to \infty$, Corollary 4.3 shows that the convergence speed is generally not algorithmically computable. Similarly, as for Question 1, this negative answer to Question 2 is the strongest possible in the sense that there is not only a general Turing machine for all $\varphi \in \mathcal{M}_{\mathrm{D}}$ but it is even impossible to design a particular Turing machine $\mathrm{TM}_*(M)$ for the spectral density φ_* and with input $M \in \mathbb{N}$ that is able to compute $N = N(M)$ for which $|E_N(\varphi_*) - E_\infty(\varphi_*)| < 2^{-M}$.

After the joint works of Kolmogorov, Paley, Szegö, Wiener, many other researchers tried to find algorithms to determine FIR approximations for the optimal prediction filter by controlling the performance of this FIR approximation. One example is the Durbin–Levinson algorithm, shortly discussed in Section 3.5. This recursive algorithm determines in each iteration the MSE $\sigma_n^2 = E_n(\varphi)$. However, Corollary 4.3 shows that it is generally impossible to decide algorithmically whether σ_n^2 is sufficiently close to the optimal MSE σ^2, i.e. there is generally no way to terminate the iteration if certain error bound is achieved, simply because it is impossible to determine whether the error bound is already achieved.

Moreover, it should be emphasized that the above results (in particular, Corollary 4.3) are independent of any particular method that tries to approximate the minimal MSE $\sigma^2 = E_\infty(\varphi)$. So, these negative results hold not only for the particular sequence $\{\sigma_n^2\}_{n\in\mathbb{N}}$ given by (22), but for any other sequence $\{\widetilde{\sigma}_n^2\}_{n\in\mathbb{N}}$ that converges to σ^2. In other words, there exists, in general, no algorithm on a digital computer that is able to effectively compute an approximation of the minimum mean squared error $E_\infty(\varphi)$ in the sense that for every $M \in \mathbb{N}$, the algorithm is able to compute an $\widetilde{\sigma}_n^2$ that is guaranteed to satisfy $\left|\sigma^2 - \widetilde{\sigma}_n^2\right| < 2^{-M}$.

4.2. **Sufficient condition for computability.** We have seen in Subsection 4.1 that for some spectral densities $\varphi \in \mathcal{M}_{\mathrm{D}}$, the corresponding optimal prediction MSE $\sigma^2 = E_\infty(\varphi)$ given by (21), is not a computable number. So, it is natural to ask for conditions on the spectral density φ so that $E_\infty(\varphi)$ is a computable number.

Question 3. For which $\varphi \in \mathcal{M}_{\mathrm{D}}$ do we have $\sigma^2 = E_\infty(\varphi) \in \mathbb{R}_{\mathrm{c}}$?

The following theorem gives an answer, namely, it provides a sufficient condition on the spectral density φ for $\sigma^2 = E_\infty(\varphi)$ to be computable. A proof of this theorem will be presented in a forthcoming publication.

Theorem 4.4. *If $\varphi \in \mathcal{M}_{\mathrm{D}}$ satisfies $\varphi(\zeta) > 0$ for all $\zeta \in \mathbb{T}$, then $\sigma^2 = E_\infty(\varphi) \in \mathbb{R}_{\mathrm{c}}$.*

Remark 4.2. Theorem 4.4 only shows that if φ is strictly positive, then the optimal MSE is a computable number. However, as we will see later, this does, generally, *not* imply that also the spectral factor φ_+ or the corresponding optimal prediction filter h_{opt} is computable.

So, Theorem 4.4 shows that to any given $\varphi \in \mathcal{M}_{\mathrm{D}}$ with $\min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0$ there exists a computable (i.e., recursive) function $e : \mathbb{N} \to \mathbb{N}$ such that for all $M \in \mathbb{N}$, one has

$$|E_\infty(\varphi) - E_N(\varphi)| < 2^{-M} \quad \text{provided that } N \geq e(M)\ .$$

However, knowing that $\sigma^2 = E_\infty(\varphi)$ is computable, we may ask for the complexity of computing $E_\infty(\varphi)$. This will be done in more detail in Section 5.

5. Complexity Blowup of Computing the Optimal MSE

The previous section gave the sufficient conditions on the spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ such that the corresponding optimal MSE $\sigma^2 = E_\infty(\varphi)$ is computable. This section now analyzes the computational complexity for computing σ^2. To this end, we assume that $\varphi \in \mathcal{M}_{\mathrm{D}}$ is a spectral density that satisfies $\min\limits_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0$. Then, according to Theorem 4.4, $\sigma^2 = E_\infty(\varphi)$ is a computable number. This means that there is a Turing machine TM_{σ^2} with inputs φ and $n \in \mathbb{N}$ and whose output satisfies

$$\left|\sigma^2 - \mathrm{TM}_{\sigma^2}(\varphi, n)\right| < 2^{-n}\ . \tag{23}$$

If σ^2 would even be polynomial-time computable, then there would exist a polynomial q so that the computation time for achieving the error bound (23) is at most $q(n)$. Our question is now, whether σ^2 is indeed polynomial-time computable.

To this end, it is important to notice first that the computational complexity of $\mathrm{TM}_{\sigma^2}(\varphi, n)$ depends strongly on the computational complexity of φ. In fact, in order that TM_{σ^2} is able to process the input φ, one needs to prepare a description of φ which could be understood by TM_{σ^2}. If, for example, φ is polynomial-time computable, then there exists a function-oracle Turing machine TM_γ (cf., Def. 2.4) such that for every $\theta \in [-\pi, \pi)$, TM_γ obtains an output that satisfies

$$\left|\varphi(\mathrm{e}^{\mathrm{i}\theta}) - \mathrm{TM}_\gamma(n)\right| < 2^{-n}$$

in a computational time which grows at most polynomial in n. In other words, TM_γ is a Turing machine that is able to determine the desired description of φ in polynomial time. So, it is clear that the complexity for computing σ^2 is at least as large as the computational complexity of the spectral density φ. Nevertheless, we may hope that for a polynomial-time computable φ, the optimal MSE σ^2 is also polynomial time computable. Otherwise, if the computation of $\sigma^2 = E_\infty(\varphi)$ is much more complex, than the computation of φ, we speak of *a complexity blowup*.

Definition 5.1. We say that the calculation of the minimum mean squared error shows *complexity blowup*, if there exists a polynomial-time computable spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ with $\min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0$ so that $\sigma^2 = E_\infty(\varphi)$ is not polynomial-time computable.

We are going to show that the calculation of the optimal MSE shows indeed complexity blowup. In order to formulate this statement precisely, we first need some more notion from the complexity theory which is shortly reviewed in the following subsection.

5.1. **Complexity classes.** Classifying the complexity of numerical problems is based on *complexity classes*. The best known are the classes P and NP which are related to the *decision problems*. P is the class of all questions (problems) that can be answered (solved) by a deterministic Turing machine in polynomial time, and NP is the class of all questions (problems) for which a given answer (solution) can be verified in polynomial time by a deterministic Turing machine. From the definition, it is clear that $P \subseteq NP$ but it is still an open question whether $P = NP$ or $P \subsetneq NP$. It is widely assumed that P is a proper subset of NP.

This paper considers also the function evaluation problems. Therefore, the complexity classes FP and $\#P$ are needed. Let $\{0,1\}^n$ be the set of all words of length $n \in \mathbb{N}$ in the alphabet $\{0,1\}$ and let $\{0,1\}^*$ be the set of all finite words in the alphabet $\{0,1\}$. The length of a word $x \in \{0,1\}^*$ will be denoted by $\mathrm{len}[x]$.

Definition 5.2. A function $f : \{0,1\}^* \to \mathbb{N}$ is in FP if it can be computed by a deterministic Turing machine in polynomial time. A function $f : \{0,1\}^* \to \mathbb{N}$ is in $\#P$ if there exist a polynomial $p : \mathbb{N} \to \mathbb{N}$ and a polynomial-time Turing machine M, so that for every string $x \in \{0,1\}^*$,

$$f(x) = \left|\left\{y \in \{0,1\}^{p(\mathrm{len}[x])} : \mathrm{M}(x,y) = 1\right\}\right| .$$

Remark 5.1. $\mathrm{M} : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ might be seen as a Turing machine that checks in polynomial time a certificate y for a problem x. Then $f(x)$ is the number of valid certificates for the problem x. So whereas a problem in NP only asks whether there exists a polynomial-time certificate for the input, a problem in $\#P$ asks for the number of such certificates.

Again, it follows from the definition that $FP \subseteq \#P$ but it is an open problem whether $FP = \#P$, i.e. whether any problem in $\#P$ can efficiently (in polynomial time) be solved by a Turing machine. It is commonly assumed that $FP \subsetneq \#P$. Moreover, if $FP = \#P$, then this would imply that $P = NP$. Conversely, $P \neq NP$ implies $FP \neq \#P$. Finally, a function $f \in \#P$ is said to be *complete in* $\#P$ if any other $g \in \#P$ can be reduced to f by a polynomial-time Turing machine.

We will also need sub-classes of FP and $\#P$ containing all functions in FP and $\#P$, respectively, that are defined only on $\{0\}^* \subset \{0,1\}^*$, i.e. on the set of all finite words in the alphabet $\{0\}$, i.e. the set of all 0-sequences: $\{0\}^* = \{\{0\}, \{0,0\}, \{0,0,0\}, \dots\}$.

Definition 5.3. A function $f : \{0\}^* \to \mathbb{N}$ is said to be in FP_1 if it can be computed by a deterministic Turing machine in polynomial time. A function $f : \{0\}^* \to \mathbb{N}$ is said to be in $\#P_1$ if there exist a polynomial $p : \mathbb{N} \to \mathbb{N}$ and a polynomial-time Turing machine M, so that for every string $x \in \{0\}^*$,

$$f(x) = \left|\left\{y \in \{0\}^{p(\mathrm{len}[x])} : \mathrm{M}(x,y) = 1\right\}\right| .$$

As in the previous cases, we have $FP_1 \subseteq \#P_1$ but it is open whether $FP_1 = \#P_1$. However, it is widely assumed that $FP_1 \subsetneq \#P_1$. Similarly as above, an $f \in \#P_1$ is said to be *complete in* $\#P_1$ if any other $g \in \#P_1$ can be reduced to f by a polynomial-time Turing machine.

5.2. **Complexity blowup for computing the MSE prediction error.** After the necessary preparations in the previous subsection, we are now able to give our main results regarding the computational complexity of the optimal mean squared prediction error $\sigma^2_{\boldsymbol{x}}$ for a wss stochastic process with spectral density $\varphi_{\boldsymbol{x}}$.

Theorem 5.1.

1. *For any polynomial-time computable* $\varphi \in \mathcal{M}_{\mathrm{D}}$ *that satisfies* $\min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0$*, the computation of* $E_\infty(\varphi)$ *is in* $\#P_1$*.*

2. *There exists a* $\varphi_* \in \mathcal{M}_{\mathrm{D}}$ *with* $\min_{\zeta\in\mathbb{T}} \varphi_*(\zeta) > 0$ *which is polynomial-time computable but such that the computation of* $E_\infty(\varphi_*)$ *is* $\#P_1$ *complete.*

Theorem 5.1 shows that the computation of the optimal MSE $E_\infty(\varphi)$ is generally a problem in $\#P_1$. So, if the conjecture $FP_1 \neq \#P_1$ is indeed true, then $E_\infty(\varphi)$ might not be polynomial-time computable. Indeed, the second part of Theorem 5.1 shows that there exists a polynomial-time computable spectral density φ_* such that the computation of $E_\infty(\varphi_*)$ is $\#P_1$ complete, i.e. it is at least as complex as any other problem in $\#P_1$. Consequently, if $FP_1 \neq \#P_1$, then $E_\infty(\varphi_*)$ is not polynomial-time computable, even though φ_* is. This verifies that the computation of the optimal MSE shows a complexity blowup (provided $FP_1 \neq \#P_1$).

Similarly, as in Section 4.1, we emphasis that the statement of Theorem 5.1 is independent of any specific algorithm but holds for any possible algorithm which computes $\sigma^2 = E_\infty(\varphi)$. For example, one may use the Durbin–Levinson algorithm as discussed in Section 3.5 to calculate a sequence σ_n^2, $n \in \mathbb{N}$ that converges to $\sigma^2 = E_\infty(\varphi)$ as $n \to \infty$. This algorithm needs as input a description (approximation) $\widetilde{\varphi}$ of the spectral density φ. If one requires that $|\varphi(\zeta) - \widetilde{\varphi}(\zeta)| < 2^{-M}$ for all $\zeta \in \mathbb{T}$ and for some precision $M \in \mathbb{N}$, then the preparation of $\widetilde{\varphi}$ needs a computation time that grows polynomially in M, because φ is assumed to be polynomial-time computable. However, Theorem 5.1 shows that the calculation of a σ_n^2 and the verification that σ_n^2 satisfies $\left|\sigma_n^2 - \sigma^2\right| < 2^{-M}$ needs generally (and in particular for the spectral density φ_*) a computation time that grows faster than any polynomial in M, provided $FP_1 \neq \#P_1$. The same is true for any other algorithm that is designed to compute the minimum mean squared error σ^2.

It is also interesting to note that computer science basically applies two different approaches to characterize the "non-computability" of computational problems. From a mathematical point of view, one has a very precise definition of "non-computability" (see, e.g., [31, 43]), namely, every computational problem for which the solution does not depend recursively on the parameters and inputs is *not computable* (on a digital computer). However, in practical and applied investigations, one uses a more heuristic definition to characterize whether a computational problem is "(practically) computable" or not. Namely, a problem is considered to be "(practically) non-computable" if the complexity for calculating the solution (using the best possible algorithm) grows faster than any polynomial in the number of parameters of the problem. Conversely, if the computational complexity (i.e. the number of calculation steps) grows at most polynomially in the number of parameters, the problem is considered to be "(practically) computable". This second approach of "practical non-computability" lies at the heart of modern cryptography and it is therefore the basis of present-day technology for secure communication. In this respect, Theorem 5.1 shows that the computation of the optimal MSE is in general "practically non-computable".

Up to this point, we have used the notion of polynomial-time computability as given in Definition 2.4. This definition is based on a function oracle Turing machine, which seems to be somewhat artificial at the beginning. Nevertheless, we have also introduced the notion of computability on dyadic-grids in Definition 2.8 which seems to be a very natural model for algorithms on digital computers. As mentioned earlier, both computation models are equivalent. Therefore, our previous results on the complexity blowup of computing $E_\infty(\varphi)$ can be reformulated in the computation model on dyadic grids. This follows from the next theorem due to Friedman. It shows that both computation models are equivalent.

Theorem 5.2 (Friedman [18])**.** *Let* $f : [0,1] \to \mathbb{R}$ *be a given function. Then* f *is polynomial-time computable on dyadic grids* (*cf., Definition* 2.8) *if and only if* f *is polynomial-computable* (*in the sense of Definition* 2.4)*.*

Using this theorem, we easily get the following reformulation of our blowup result in Theorem 5.1

Corollary 5.3. *Assume* $FP_1 \neq \#P_1$*, then there always exists a spectral density* $\varphi_* \in \mathcal{M}_{\mathrm{D}}$ *with* $\min_{\zeta\in\mathbb{T}} |\varphi_*(\zeta)| > 0$ *with the following to properties:*

1. *For every $n \in \mathbb{N}$, one can compute $\varphi_*(\zeta)$ on the grid*

$$\zeta_{k,n} = \exp\left(\mathrm{i}2\pi \cdot k \cdot 2^{-n}\right) , \qquad 0 \le k \le 2^n ,$$

exactly up to an error of at most 2^{-n} in polynomial time, i.e. one can compute $\widetilde{\varphi}_*(\zeta_{k,n})$ so that

$$|\widetilde{\varphi}_*(\zeta_{k,n}) - \varphi_*(\zeta_{k,n})| < 2^{-n} , \quad \text{for all} \quad 0 \le k \le 2^n ,$$

and in a computation times that grows at most polynomial in n.

2. *The optimal MSE $\sigma^2 = E_\infty(\varphi_*)$ cannot be computed in polynomial time, i.e. the time for computing an approximation $\widetilde{\sigma}^2$ so that*

$$\left|\sigma^2 - \widetilde{\sigma}^2\right| < 2^{-n}$$

grows faster than any polynomial in the required precision n.

6. Computability of the Spectral Factorization

This section investigates under which conditions on a computable spectral density φ, the spectral factor φ_+ will be again a computable function. The previous section already showed that the spectral density φ needs to be strictly positive for the minimal mean squared error $E_\infty(\varphi)$ to be a computable number (cf., Theorem 4.4). We will show in this section that generally much stronger conditions on φ are necessary in order for the spectral factor φ_+ to be a computable function.

To formalize these additional conditions on φ, we first need some more notation and function spaces which are introduced in the first subsection. Here, we follow mainly the notion in [6].

6.1. **Functions of finite Dirichlet energy.** Let $f \in \mathcal{C}(\mathbb{T})$ be a continuous function on $\mathbb{T}$ with Fourier coefficients (1). The *Dirichlet energy* of f is given by the square of the extended (cf. [2]) seminorm

$$\|f\|_{\mathrm{E}} = \left(\sum_{n\in\mathbb{Z}} |n| \ |c_n(f)|^2 \right)^{1/2} . \tag{24}$$

We notice that the seminorm (24) has an important interpretation as physical energy in many applications (see, e.g., [5, 33]). For $f \in \mathcal{A}$, the Dirichlet energy can be expressed in terms of the *Dirichlet integral* as

$$\|f\|_{\mathrm{E}}^2 = \frac{1}{\pi} \iint\limits_{\mathbb{D}} |f'(z)|^2 \, \mathrm{d}z .$$

We write $H^{1/2}(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) \ : \ \|f\|_{\mathrm{E}} < +\infty \right\}$ for the Sobolev space of all functions of finite Dirichlet energy with the norm $\|f\|_{H^{1/2}(\mathbb{T})} = \max\left(\|f\|_\infty , \|f\|_{\mathrm{E}}\right)$, and $\mathcal{B} = \{f \in \mathcal{C}(\mathbb{T}) \ : \ \|f\|_{\mathrm{E}} < \infty\} \subset H^{1/2}(\mathbb{T})$ stands for the subspace of continuous functions of finite Dirichlet energy with the norm $\|f\|_{\mathcal{B}} = \max\left(\|f\|_\infty , \|f\|_{\mathrm{E}} \right)$. Then $\mathcal{B}_{\mathrm{c}} = H^{1/2}(\mathbb{T}) \cap \mathcal{C}_{\mathrm{c}}(\mathbb{T})$ will be the set of all $f \in \mathcal{B}$ that are computable as the functions in $\mathcal{B}$.

Since we study spectral factorization, we need spaces of positive functions. Therefore we define $\mathcal{D}$ as the set of all spectral densities $\varphi \in \mathcal{B}$ that are strictly positive, i.e.,

$$\mathcal{D} = \left\{ \varphi \in \mathcal{B} \ : \ \min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0 \right\} .$$

We emphasize that $\mathcal{D}$ is not a linear space because of the required positivity of $\varphi \in \mathcal{D}$. For this reason, it turns out to be useful to consider the sets

$$\log(\mathcal{D}) = \{ f = \log\varphi \ : \ \varphi \in \mathcal{D} \} \quad \text{and} \quad \exp(\mathcal{B}) = \{ \varphi = \exp f \ : \ f \in \mathcal{B} \}$$

of all functions defined by $f(\zeta) = \log\varphi(\zeta)$, $\zeta \in \mathbb{T}$ with $\varphi \in \mathcal{D}$, and the set of all positive functions φ defined by $\varphi(\zeta) = \exp\left(f(\zeta)\right)$ for $\zeta \in \mathbb{T}$ with $f \in \mathcal{B}$, respectively.

Since we are interested whether a given density $\varphi \in \mathcal{D}$ has a computable spectral factor φ_+, we require *a priori* that φ itself is computable. Therefore we consider

$$\begin{aligned}\mathcal{D}_{\mathrm{c}} &= \Big\{ \varphi \in \mathcal{D} \,:\, \varphi \in \mathcal{C}_{\mathrm{c}}(\mathbb{T}) \Big\} \\ &= \Big\{ \varphi \in \mathcal{C}_{\mathrm{c}}(\mathbb{T}) \,:\, \|\varphi\|_{\mathrm{E}} < \infty \text{ and } \min_{\zeta\in\mathbb{T}} \varphi(\zeta) > 0 \Big\} ,\end{aligned}$$

i.e. the set of spectral densities $\varphi \in \mathcal{D}$ that are computable as a continuous function. Subsequently, we consider only spectral densities that belong to $\mathcal{D}_{\mathrm{c}}$ or to certain subsets of $\mathcal{D}_{\mathrm{c}}$.

6.2. **Non-computability and computability of the spectral factor.** Our first result taken from [6], shows that even for spectral densities with fairly nice analytic properties, the spectral factor will usually not be computable.

Theorem 6.1. *There exists a spectral density $\varphi \in \mathcal{D}_{\mathrm{c}}$ with the following properties:*

(a) $\varphi \in \mathcal{C}_{\mathrm{ac}}(\mathbb{T})$ *and* $\varphi \in \mathcal{W}$*;*

(b) $\varphi_+ \in \mathcal{B}$ *and* $\varphi_+ \in \mathcal{C}_{\mathrm{ac}}(\mathbb{T})$ *and* $\varphi_+ \in \mathcal{W}$*,*

but such that $\varphi_+(1) \notin \mathbb{C}_{\mathrm{c}}$.

So, there are spectral densities φ with very favorable analytic properties but with a spectral factor φ_+ that is not computable at certain points on the unit circle. In fact, Theorem 6.1 shows that there is a spectral density φ that is strictly positive, continuous, computable, that has finite Dirichlet energy, and for which we know that its spectral factor φ_+ is absolutely continuous with an absolute converging Fourier series but such that $\varphi_+(1)$ is not a computable number. This shows, in particular, that for such spectral density φ, the spectral factor φ_+ does not satisfy the weakest necessary condition for a computable function. and so these spectral factors are not computable in any stronger notion [1, 43] of computability. In particular, φ_+ is not a computable continuous function.

Since, according to Theorem 6.1, not all computable spectral densities $\varphi \in \mathcal{D}_{\mathrm{c}}$ have a computable spectral factor φ_+, we may ask for subsets of $\mathcal{D}_{\mathrm{c}}$ so that every spectral density in this subset possesses a computable spectral factor. To characterize such subsets, we introduce for every $\alpha \geq 0$, the usual Sobolev space $H^\alpha(\mathbb{T})$ of functions in $L^2(\mathbb{T})$ with

$$\|f\|_{H^\alpha(\mathbb{T})} = \Big(\sum_{n\in\mathbb{Z}} (1+|n|^\alpha)^2 \, |c_n(f)|^2 \Big)^{1/2} < \infty \, .$$

For $\alpha = 1/2$, the so-defined space coincides with $H^{1/2}(\mathbb{T})$ of all functions of finite Dirichlet energy as defined previously. For $\alpha > 1/2$, we have $H^\alpha(\mathbb{T}) \subset H^{1/2}(\mathbb{T})$, i.e. all functions in $H^\alpha(\mathbb{T})$ have finite Dirichlet energy. Moreover, we notice that for $\alpha > 1/2$, we have $H^\alpha(\mathbb{T}) \subset \mathcal{W} \subset \mathcal{C}(\mathbb{T})$. Indeed, let $f \in H^\alpha(\mathbb{T})$ and $m \in \mathbb{N}$ be arbitrary, then the Cauchy–Schwarz inequality gives

$$\begin{aligned}\big|(\mathrm{S}_m f)\,(\mathrm{e}^{\mathrm{i}\theta})\big| &= \Big| \sum_{n=-m}^{m} c_n(f)\, \mathrm{e}^{\mathrm{i}n\theta} \Big| \leq \sum_{n=-m}^{m} |c_n(f)| \\ &\leq \Big(\sum_{n=-m}^{m} (1+|n|^\alpha)^2 \, |c_n(f)|^2 \Big)^{1/2} \Big(\sum_{n=-m}^{m} \frac{1}{(1+|n|^\alpha)^2} \Big)^{1/2} \\ &\leq K(\alpha)\, \|f\|_{H^{1/2}(\mathbb{T})} \end{aligned} \tag{25}$$

with the constant

$$K(\alpha) = \Big(\sum_{n=-\infty}^{\infty} \frac{1}{(1+|n|^\alpha)^2} \Big)^{1/2} \leq \Big(1 + 2 \sum_{n=1}^{\infty} \frac{1}{n^{2\alpha}} \Big)^{1/2} ,$$

which is a computable real number for every $\alpha \in \mathbb{R}_{\mathrm{c}}$ with $\alpha > 1/2$. Since the right-hand side of (25) is independent of m, this inequality also holds for $m \to \infty$, and so it shows that for every $f \in H^\alpha(\mathbb{T})$, the partial Fourier series $\mathrm{S}_m f$ converges to f uniformly on $\mathbb{T}$. In particular, we have $H^\alpha(\mathbb{T}) \subset \mathcal{B}$ for $\alpha > 1/2$, and it is clear that

$$\|f - \mathrm{S}_{m_2} f\|_{H^\alpha(\mathbb{T})} \leq \|f - \mathrm{S}_{m_1} f\|_{H^\alpha(\mathbb{T})} \ , \qquad \text{for } m_2 \geq m_1 \, .$$

Based on these preparations, we are now able to characterize computable spectral densities $\varphi \in \mathcal{D}_{\mathrm{c}}$ that possess a computable spectral factor φ_+.

Theorem 6.2. *Let $\alpha \in \mathbb{R}_{\mathrm{c}}$, $\alpha > 1/2$ be fixed, and let $\varphi \in \mathcal{D}_{\mathrm{c}}$ be a spectral density such that $\log \varphi \in H_{\mathrm{c}}^{\alpha}(\mathbb{T})$. Then $\varphi_+ \in \mathcal{C}_{\mathrm{c}}(\mathbb{T})$.*

Note that Theorem 6.2 is sharp with respect to the parameter α. Therewith we have a complete characterization of sets of spectral densities such that for every φ in these sets, the spectral factor φ_+ is guaranteed to be a computable continuous function. Indeed, for $\alpha = 1/2$ and $\varphi \in \mathcal{D}_{\mathrm{c}}$, one can show that $\log \varphi \in H_{\mathrm{c}}^{1/2}(\mathbb{T})$. However, according to Theorem 6.1, there are the densities φ such that $\varphi_+(1) \notin \mathbb{C}_{\mathrm{c}}$, i.e. such that φ_+ is not a function that maps computable numbers in $\mathbb{T} \cap \mathbb{C}_{\mathrm{c}}$ onto computable numbers in $\mathbb{C}_{\mathrm{c}}$. So, Theorem 6.2 does indeed not hold for $\alpha = 1/2$. Only if $\log \varphi \in H_{\mathrm{c}}^{\alpha}(\mathbb{T})$ with $\alpha > 1/2$, $\alpha \in \mathbb{R}_{\mathrm{c}}$, then φ_+ is always a computable continuous function, according to Theorem 6.2.

6.3. **Spectral factorization for polynomial densities.** One particular but important case of spectral factorization is the situation where the spectral density φ is a trigonometric polynomial, i.e. a function of the form $\varphi(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{n=-N}^{N} c_n \mathrm{e}^{\mathrm{i}n\theta}$. For all polynomial spectral densities φ that satisfy $\min_{\zeta \in \mathbb{T}} |\varphi(\zeta)| > 0$, the spectral factor φ_+ is always computable. Indeed, for polynomials, the spectral factorization is equivalent to the Fejér–Riesz theorem [17]. It states that if φ is a function of the form $\varphi(z) = \sum_{n=-N}^{N} c_n z^n$ that is non-negative for $|z| = 1$, then the roots of φ occur in para-conjugate pairs ξ and $1/\overline{\xi}$. So, either a root of φ lies on the unit circle (in that case, it will have even multiplicity), or one root of the pair ξ, $1/\overline{\xi}$ lies inside $\mathbb{D}$, whereas the other lies outside $\overline{\mathbb{D}}$. Therefore, any polynomial density φ can be written as

$$\varphi(z) = c_0 \prod_{n=1}^{N} (z - \xi_n) \left(z^{-1} - \overline{\xi_n}\right) , \quad z \in \mathbb{C} ,$$

with a constant $c_0 > 0$ and with $\xi_n \in \mathbb{C}$ satisfying $|\xi_n| > 1$ for all $n = 1, 2, \ldots, N$. Consequently, the spectral factor of φ is given by $\varphi_+(z) = \sqrt{c_0} \prod_{n=1}^{N} (z - \xi_n)$ and the spectral factorization problem reduces to the problem of subdividing the roots of φ into those roots which lie inside and outside the unit disk, respectively. In other words, for determining the spectral factor of a positive polynomial φ, one has to compute the roots of φ. This is a classical problem in mathematics and there exist many different proofs of the fundamental theorem of algebra since the first proof of Gauss in 1799. However, it was not before 1925 that Hermann Weyl [44] gave a first constructive proof of this theorem (cf. also discussion in [36]).

The main problem is the fact that there exists, in general, no simple function that is able to determine the zeros of a polynomial of degree larger than 4 from its coefficients[2]. Nevertheless, since the field $\mathbb{C}_{\mathrm{c}}$ of computable complex numbers is algebraically closed, the roots of polynomials with computable coefficients are again computable complex numbers. Even though this observation does not necessarily imply that it is always possible to determine the roots of a polynomial *effectively* from its coefficients, E. Specker proved [39] that the relation between the coefficients and roots of a computable polynomial is, in fact, effective. So, let $\mathcal{P}_{N,\mathrm{c}}$ be the set of all computable polynomials of degree N (i.e. polynomials of degree N with computable coefficients). Then for every $N \in \mathbb{N}$ there exists a recursive function on $\mathcal{P}_{N,\mathrm{c}}$ that associates the roots of a polynomial $p \in \mathcal{P}_{N,\mathrm{c}}$ with the polynomial p [39]. So, starting with a computable polynomial density $\varphi \in \mathcal{P}_{N,\mathrm{c}} \cap \mathcal{D}_{\mathrm{c}}$, it is, in principle, possible to effectively determine its roots $\{\xi_n\}_{n=1}^{2N} \subset \mathbb{C}_{\mathrm{c}}$. Since $\varphi \in \mathcal{D}_{\mathrm{c}}$ is strictly positive on $\mathbb{T}$, it follows that $|\xi_n| \neq 1$ for all $n = 1, 2 \ldots, N$. Then one can determine all those ξ_n with $|\xi_n| > 1$. This gives finally the spectral factor φ_+ as described above.

[2]In other words and according to Galois theory only polynomial equations of degree $N \leq 4$ can generally be solved by radicals.

As in Section 5, we may ask for the computational complexity of determining the spectral factor φ_+ of a polynomial density $\varphi_N \in \mathcal{P}_{N,\mathrm{c}}$. In [36], computational complexity was precisely investigated and it was shown that the complexity grows at most polynomial with the degree N and in the precision M.

6.4. **Application to prediction and Wiener filtering.** We already discussed in Section 3.4 that the spectral factorization is closely related to the problem of finding the optimal (minimal MSE) causal filter for predicting the value x_0 of a wss stochastic process $\boldsymbol{x}$ from it past values $\{x_{-1}, x_{-2}, x_{-3}, \dots\}$. If the spectral density of $\boldsymbol{x}$ is φ, then the transfer function h_{opt} of the optimal prediction filter is given by (16). Form this formula, it is immediately clear that Theorem 6.1 and Theorem 6.2 hold verbatim also for h_{opt}.

Corollary 6.3. *There exists a wss stochastic process $\boldsymbol{x}$ with a spectral density $\varphi \in \mathcal{D}_{\mathrm{c}}$ such that $\varphi \in \mathcal{C}_{\mathrm{ac}}(\mathbb{T}) \cap \mathcal{W}$ and $\varphi_+ \in \mathcal{B} \cap \mathcal{C}_{\mathrm{ac}}(\mathbb{T}) \cap \mathcal{W}$ but such that the optimal prediction filter* (16) *satisfies $h_{\mathrm{opt}}(1) \notin \mathbb{C}_{\mathrm{c}}$.*

If, on the other side, the spectral density $\varphi \in \mathcal{D}_{\mathrm{c}}$ of $\boldsymbol{x}$ satisfies $\log \varphi \in H^{\alpha}_{\mathrm{c}}(\mathbb{T})$ for some $\alpha \in \mathbb{R}_{\mathrm{c}}$, $\alpha > 1/2$, then $h_{\mathrm{opt}} \in \mathcal{C}_{\mathrm{c}}(\mathbb{T})$.

There is a slightly but practically important generalization of the previously discussed prediction problem. To explain this, assume $\boldsymbol{y} = \{y_n\}_{n\in\mathbb{Z}} \subset \mathcal{R}$ is a second wss stationary stochastic process that is stationary correlated with $\boldsymbol{x}$. This means that $\mathrm{E}[x_{n+k}\overline{y_k}] = \langle x_{n+k}, y_k\rangle_{\mathcal{R}} = \langle x_n, y_0\rangle_{\mathcal{R}} = \mathrm{E}[x_n\overline{y_0}]$ for all $n, k \in \mathbb{Z}$. Then the *cross-covariance function* is defined to be $\gamma_{\boldsymbol{x},\boldsymbol{y}}(n) = \langle x_n, y_0\rangle_{\mathcal{R}}$ with a spectral representation similarly to (5) but with a spectral measure $\mu_{\boldsymbol{x},\boldsymbol{y}}$. As before, we assume that both stochastic processes are purely non-deterministic and that the spectral measure has no singular part. Then the *cross-covariance function* is completely determined by the cross-spectral density $\psi \in L^1(\mathbb{T})$, so that

$$\gamma_{\boldsymbol{x},\boldsymbol{y}}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}n\theta}\psi(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta\,, \qquad n \in \mathbb{Z}\,.$$

Now our aim is to find an estimate $\widehat{x}_0$ of the random variable x_0 based on the past $\{y_n : n \le 0\}$ of the stochastic process $\boldsymbol{y}$ such that the mean squared error $\mathrm{E}[|x_0 - \widehat{x}_0|^2]$ is minimized. If we require that the estimator is a linear filter, it will have the form

$$\widehat{x}_0 = \mathrm{H}(\boldsymbol{y}) = \sum_{k=0}^{\infty} h_k\, y_{-k}\,,$$

with the impulse response $\{h_k\}_{k=0}^{\infty}$ and transfer function $h(\mathrm{e}^{\mathrm{i}\theta}) = \sum\limits_{k=0}^{\infty} h_k\,\mathrm{e}^{\mathrm{i}k\theta}$. The optimal filter is known as a (causal) *Wiener filter* [25, 46] given by

$$\begin{aligned} h_{\mathrm{WF}}(\mathrm{e}^{\mathrm{i}\omega}) &= \frac{1}{\varphi_+(\mathrm{e}^{\mathrm{i}\omega})}\left(\mathrm{P}_+\Big[\frac{\psi}{\varphi_-}\Big]\right)(\mathrm{e}^{\mathrm{i}\omega}) \\ &= \frac{1}{\varphi_+(\mathrm{e}^{\mathrm{i}\omega})}\,\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\psi(\mathrm{e}^{\mathrm{i}\theta})}{\varphi_-(\mathrm{e}^{\mathrm{i}\theta})}\,\frac{\mathrm{e}^{\mathrm{i}\theta}}{\mathrm{e}^{\mathrm{i}\theta}-\mathrm{e}^{\mathrm{i}\omega}}\,\mathrm{d}\theta\,. \end{aligned} \tag{26}$$

Therein $\mathrm{P}_+ : L^2(\mathbb{T}) \to H^2(\mathbb{D})$ is the orthogonal projection from $L^2(\mathbb{T})$ onto $H^2(\mathbb{D})$ given by

$$\mathrm{P}_+ \;:\; f(\mathrm{e}^{\mathrm{i}\omega}) = \sum_{n=-\infty}^{\infty} c_n(f)\,\mathrm{e}^{\mathrm{i}n\omega} \mapsto \sum_{n=0}^{\infty} c_n(f) z^n\,.$$

Similarly, as for the prediction filter, we may ask whether the optimal Wiener filter (26) is computable. The following statement, which easily follows from Theorem 6.1 (cf., [6]), shows that h_{WF} is generally not computable even under strong assumptions on the spectral densities φ and ψ.

Corollary 6.4. *There exist spectral densities $\varphi, \psi \in \mathcal{B}_{\mathrm{c}}$ with $\varphi, \psi \in \mathcal{C}_{\mathrm{ac}}(\mathbb{T})$ and $\varphi, \psi_{\boldsymbol{x},\boldsymbol{y}} \in \mathcal{W}$ such that the transfer function of the causal Wiener filter* (26) *satisfies $h_{\mathrm{WF}} \in \mathcal{B}$, $h_{\mathrm{WF}} \in \mathcal{C}_{\mathrm{ac}}(\mathbb{T})$, and $h_{\mathrm{WF}} \in \mathcal{W}$ but such that $h_{\mathrm{WF}}(1)$ is not computable.*

So, similarly as for the prediction filter discussed in Section 6.2, the Wiener filter is, generally, not computable, even if the two spectral densities φ and ψ are computable functions with very nice analytic properties, the transfer function of the corresponding Wiener filter (26) is usually not a computable function. In fact, h_{WF} is even not computable at certain points on the unit circle and, in particular, h_{WF} is not a computable continuous function.

7. Summary

This paper has considered computational aspects in some important areas of prediction and Wiener filter theory. It was shown that for the problem of linear causal prediction, the important performance measure of the minimum mean squared error (MMSE) is, generally not a computable number, even for spectral densities with fairly nice analytic properties. Then, based on the smoothness of the spectral density, we have been able to characterize precisely subsets of spectral densities for which the MMSE is a computable number. Nevertheless, it turned out that even for these subsets the computation of the MMSE shows complexity blowup, i.e. even if the spectral density is low-complexity (i.e. polynomial-time computable), the calculation of the MMSE is generally of high complexity (i.e. not polynomial-time computable) under the complexity theoretical condition that $FP_1 \neq \#P_1$.

As a second problem, we investigated the computability of the spectral factorization. There we could show that even under strong conditions on the smoothness of the spectral density, the spectral factor is generally not computable. This non-computability then easily translates into the non-computability of the Wiener filter. On the other hand, we have given a sharp characterization of sets of spectral densities φ that possess a computable spectral factor. It was shown that if $\log \varphi$ is sufficiently smooth, then spectral factor will always be computable.

Acknowledgement

The work of H. Boche was supported in part by the German Federal Ministry of Education and Research (BMBF) within the national initiative on 6G Communication Systems through the research hub 6G – life under Grant 16KISK002, within the national initiative for Post Shannon Communication (NewCom) under Grant 16KIS1003K, and the project Hardware Platforms and Computing Models for Neuromorphic Computing (NeuroCM) under Grant 16ME0442. It has further received funding by the Bavarian Ministry of Economic Affairs, Regional Development and Energy as part of the project 6G Future Lab Bavaria as well as in part by the German Research Foundation (DFG) within Germany's Excellence Strategy EXC–2092–390781972.

V. Pohl acknowledges the support of the DFG under Grant PO 1347/3–2.

The work of H. V. Poor was supported by the U.S. National Science Foundation under Grants CCF-1908308 and CNS-2128448.

References

1. J. Avigad, B. Vasco, Computability and analysis: the legacy of Alan Turing. In: *Turing's legacy: developments from Turing's ideas in logic*, pp. 1–47, Lect. Notes Log., 42, Assoc. Symbol. Logic, La Jolla, CA, 2014.
2. G. Beer, Norms with infinite values. *J. Convex Anal.* **22** (2015), no. 1, 37–60.
3. N. H. Bingham, Szegö's theorem and its probabilistic descendants. *Probab. Surv.* **9** (2012), 287–324.
4. H. Boche, V. Pohl, Structural properties of the Wiener filter—stability, smoothness properties, and FIR approximation behavior. *IEEE Trans. Inform. Theory* **52** (2006), no. 9, 4272–4282.
5. H. Boche, V. Pohl, Calculating the Hilbert transform on spaces with energy concentration: convergence and divergence regions. *IEEE Trans. Inform. Theory* **65** (2019), no. 1, 586–603.
6. H. Boche, V. Pohl, On the algorithmic solvability of spectral factorization and applications. *IEEE Trans. Inform. Theory* **66** (2020), no. 7, 4574–4592.
7. H. Boche, V. Pohl, Complexity blowup in simulating analog linear time-invariant systems on digital computers. *IEEE Transactions on Signal Processing* **69** (2021), 5005–5020.
8. H. W. Bode, C. E. Shannon, A simplified derivation of linear least square smoothing and prediction theory. *Proc. I.R.E.* **38** (1950), 417–425.
9. S. Darlington, Linear least-squares smoothing and prediction, with applications. *Bell System Tech. J.* **37** (1958), 1221–1294.
10. J. L. Doob, *Stochastic Processes*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953.
11. J. Durbin, The fitting of time-series models. *Rev. Int. Statist. Inst* **28** (1960), no. 3, 233–244.

12. P. L. Duren, *Theory of H^p Spaces*. Pure and Applied Mathematics, Vol. 38 Academic Press, New York-London 1970.
13. L. Ephremidze, G. Janashia, E. Lagvilava, An analytic proof of the matrix spectral factorization theorem. *Georgian Math. J.* **15** (2008), no. 2, 241–249.
14. L. Ephremidze, G. Janashia, E. Lagvilava, A simple proof of the matrix-valued Fejér-Riesz theorem. J. Fourier Anal. Appl. 15 (2009), no. 1, 124–127.
15. L. Ephremidze, G. Janashia, E. Lagvilava, On approximate spectral factorization of matrix functions. *J. Fourier Anal. Appl.* **17** (2011), no. 5, 976–990.
16. L. Ephremidze, F. Saied, I. M. Spitkovsky, On the algorithmization of Janashia-Lagvilava matrix spectral factorization method. *IEEE Trans. Inform. Theory* **64** (2018), no. 2, 728–737.
17. L. Fejér, Über trigonometrische Polynome. (German) *J. Reine Angew. Math.* **146** (1916), 53–82.
18. H. Friedman, The computational complexity of maximization and integration. *Adv. in Math.* **53** (1984), no. 1, 80–98.
19. J. B. Garnett, *Bounded Analytic Functions*. Pure and Applied Mathematics, 96. Academic Press, Inc., New York-London, 1981.
20. K. Hoffman, *Banach Spaces of Analytic Functions*. Prentice-Hall Series in Modern Analysis Prentice-Hall, Inc., Englewood Cliffs, N.J. 1962
21. G. Janashia, E. Lagvilava, A method of approximate factorization of positive definite matrix functions. *Studia Math.* **137** (1999), no. 1, 93–100.
22. G. Janashia, E. Lagvilava, L. Ephremidze, A new method of matrix spectral factorization. *IEEE Trans. Inform. Theory* **57** (2011), no. 4, 2318–2326.
23. R. E. Kalman, A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D. J. Basic Engrg.* **82** (1960), no. 1, 35–45.
24. K. Ko, *Complexity Theory of Real Functions*. Progress in Theoretical Computer Science. Birkhäuser Boston, Inc., Boston, MA, 1991.
25. A. Kolmogoroff, Interpolation und Extrapolation von stationären zufälligen Folgen. (Russian) *Bull. Acad. Sci. URSS Sér. Math.* **5** (1941), 3–14.
26. V. Kucera, Factorization of rational spectral matrices: A survey of methods. In: *International Conference on Control* pp. 1074–1078, Control'91, IET, 1991.
27. J. Lamperti, *Stochastic Processes*. A survey of the mathematical theory. Applied Mathematical Sciences, Vol. 23. Springer-Verlag, New York-Heidelberg, 1977.
28. N. Levinson, The Wiener (root mean square) error criterion in filter design and prediction. *J. Math. Phys.* **25** (1946), no. 1–4, 261–278.
29. A. Oppenheim, R. Schafer, *Discrete-time Signal Processing*. Vol. 2. Prentice-Hall, Upper Saddle River, NJ, 1999.
30. H. V. Poor, *An Introduction to Signal Detection and Estimation*. Second edition. Springer Texts in Electrical Engineering. A Dowden & Culver Book. Springer-Verlag, New York, 1994.
31. M. B. Pour-El, J. I. Richards, *Computability in Analysis and Physics*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1989.
32. J. G. Proakis, *Digital Communications*. McGraw-Hill, New York, 1989.
33. W. T. Ross, The classical Dirichlet space. In: *Recent advances in operator-related function theory*, pp. 171–197, Contemp. Math., 393, Amer. Math. Soc., Providence, RI, 2006.
34. Yu. A. Rozanov, *Stationary Random Processes*. Translated from the Russian by A. Feinstein Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam 1967.
35. A. H. Sayed, T. Kailath, A survey of spectral factorization methods. Numerical linear algebra techniques for control and signal processing. *Numer. Linear Algebra Appl.* **8** (2001), no. 6-7, 467–496.
36. A. Schönhage, *The Fundamental Theorem of Algebra in Terms of Computational Complexity*. Manuscript. Univ. of Tübingen, Germany, 1982.
37. A. N. Shiryaev, *Probability*. Translated from the first (1980) Russian edition by R. P. Boas. Second edition. Graduate Texts in Mathematics, 95. Springer-Verlag, New York, 1996.
38. A. N. Shiryayev, Stationary Sequences in Hubert Space. In: *Selected Works of AN Kolmogorov*, pp. 228–271. Springer, Dordrecht, 1992.
39. E. Specker, The fundamental theorem of algebra in recursive analysis. In: *Ernst Specker Selecta*, pp. 264–272. Birkhäuser, Basel, 1990.
40. G. Szegö, Beiträge zur Theorie der Toeplitzschen Formen. (German) *Math. Z.* **6** (1920), no. 3-4, 167–202.
41. A. M. Turing, On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc. (2)* **42** (1936), no. 3, 230–265.
42. A. M. Turing, On Computable Numbers, with an Application to the Entscheidungsproblem. A Correction. *Proc. London Math. Soc. (2)* **43** (1937), no. 7, 544–546.
43. K. Weihrauch, *Computable Analysis*. An introduction. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2000.
44. H. Weyl, Randbemerkungen zu Hauptproblemen der Mathematik. (German) *Math. Z.* **20** (1924), no. 1, 131–150.
45. N. Wiener, Über eine klasse singulärer integralgleichungen. *Sitz. Ber. Preuss. Akad. Wiss., Phys.-Math.* **1** (1931), 696–706.

46. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. With Engineering Applications. Technology Press of The Massachusetts Institute of Technology, Cambridge, Mass.; John Wiley & Sons, Inc., New York, N.Y.; Chapman & Hall, Ltd., London, 1949.
47. N. Wiener, P. Masani, The prediction theory of multivariate stochastic processes. I. The regularity condition. *Acta Math.* **98** (1957), 111–150.
48. L. A. Zadeh, J. R. Ragazzini, An extension of Wiener's theory of prediction. *J. Appl. Phys.* **21** (1950), 645–655.

(Received 01.08.2022)

[1]Institute of Theoretical Information Technology, Technical University of Munich, 21 Arcisstrasse, 80290 München, Germany

[2]Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA
Email address: boche@tum.de
Email address: volker.pohl@tum.de
Email address: poor@princeton.edu