2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

The Wiener Theory of Causal Linear Prediction Is Not Effective

Holger Boche, Volker Pohl and H. Vincent Poor

Abstract— **In this paper, it will be shown that the minimum mean square error (MMSE) for predicting a stationary stochastic time series from its past observations is not generally Turing computable, even if the spectral density of the stochastic process is differentiable with a computable first derivative. This implies that for any approximation sequence that converges to the MMSE there does not exist an algorithmic stopping criterion that guarantees that the computed approximation is sufficiently close to the true value of the MMSE. Furthermore, it will be shown that under the same conditions on the spectral density, it is also the case that coefficients of the optimal prediction filter are not generally Turing computable.**

I. Introduction

A classical problem in many different areas of engineering is to predict the value x_0 of a discrete stochastic time series $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}}$ from observations of its past values $\{x_{-1}, x_{-2}, x_{-3}, \dots\}$. A corresponding linear prediction filter H has the form

$$\widehat{x}_0 = \mathrm{H}(\boldsymbol{x}) = \textstyle\sum_{n=1}^{\infty} h_n x_{-n} \tag{1}$$

and the problem is to determine the filter coefficients $\{h_n\}_{n=1}^{\infty}$ in such a way that the mean square error (MSE) $\sigma^2 = \mathrm{E}[|\widehat{x}_0 - x_0|^2]$ is minimized. Moreover, the minimal possible MSE, denoted by $\sigma^2_{\min}$, is an important performance measure and therefore it is often necessary to compute $\sigma^2_{\min}$ for a given $\boldsymbol{x}$.

This problem is well studied. Starting in the 1940's with the seminal works of Kolmogorov [1] and Wiener [2], the theory was later further developed in different directions by many other researchers [3]–[6]. Nowadays, this problem can be found in numerous fields of science and engineering [7], such as control [8]–[12], communication [13], [14], signal processing [15]–[18], to mention only very few. Now, closed form expressions for the optimal filter and $\sigma^2_{\min}$ are known and there are many different algorithms to determine approximations of the optimal impulse response $\{h_n\}_{n=1}^{\infty}$ by a finite impulse response (FIR) filter $\{h_n^{(N)}\}_{n=1}^{N}$ of order N. Predicting x_0 with such an approximate FIR filter

$$\widehat{x}_0 = \mathrm{H}_N(\boldsymbol{x}) = \textstyle\sum_{n=1}^{N} h_n^{(N)} x_{-n} \tag{2}$$

yields an MSE $\sigma^2_N \geq \sigma^2_{\min}$ that monotonically decreases as N increases and which eventually converges to $\sigma^2_{\min}$ as $N \to \infty$. From a practical point of view, the question is then how to choose the filter length N. It seems to be natural to chose N such that $\left|\sigma^2_N - \sigma^2_{\min}\right|$ is sufficiently small. However, to make this procedure effective, we still need an *algorithmic stopping criterion*, i.e. we need an algorithm whose input is the desired precision $M \in \mathbb{N}$ and whose output will be a filter length N_0 that guarantees that $\left|\sigma^2_N - \sigma^2_{\min}\right| < 2^{-M}$ as long as $N \geq N_0$.

It is a remarkable observation that such an algorithmic stopping criterion is *not* known in general, but only for very special stochastic processes. Therefore, we first investigate
Problem 1: Is it possible to find an algorithmic stopping criterion for the computation of $\sigma^2_{\min}$ for stochastic processes with smooth, computable spectral densities?
The answer generally depends on the actual FIR approximation algorithm and on the actual stochastic process. Nevertheless, we are going to show that there exist (infinitely many) stochastic processes with a smooth and computable spectral density such that for any arbitrary FIR approximation algorithm no such stopping criterion can exist on any digital computer. In fact, we are going to show that the minimum mean square error (MMSE) $\sigma^2_{\min}$ is not generally a computable number.

However, even if for a stochastic process with smooth and computable spectral density, $\sigma^2_{\min}$ is not computable, we may ask whether its is possible to effectively compute the optimal coefficients h_n in (1). Assume $\{h_n^{(N)} : n = 1, \cdots, N\}_{N\in\mathbb{N}}$ is a sequence of FIR approximations of the optimal impulse response $\{h_n : n \in \mathbb{N}\}$, i.e. assume $\lim_{N\to\infty} h_n^{(N)} = h_n$ for every $n \in \mathbb{N}$. To make such an approximation effective, one needs again an algorithmic stopping criterion for each $n \in \mathbb{N}$, i.e. one needs an algorithm with input $M \in \mathbb{N}$ and output $N_0 \in \mathbb{N}$ so that $\left|h_n^{(N)} - h_n\right| < 2^{-M}$ provided $N \geq N_0$. Similarly to the above, it is also the case that no stopping criterion is known, and therefore we consider
Problem 2: Is it possible to find an algorithmic stopping criterion for the computation of the optimal filter coefficients h_n for stochastic processes with smooth and computable spectral density?
It will be shown that no such general stopping criterion can exist for any possible algorithm for determining FIR approximations $\{h_n^{(N)} : n = 1, \cdots, N\}_{N\in\mathbb{N}}$.

H. Boche was supported by the German Ministry of Education and Research (BMBF) within the program "Souverän. Digital. Vernetzt" in the joint project 6G-life under Grant 16KISK002 and within the joint project Post Shannon Communication (NewCom) under Grant 16KIS1003K. He was also supported by the German Research Foundation (DFG) under Grant BO 1734/37–1. The work of H. V. Poor was partly supported by the U.S National Science Foundation under Grants CNS-2128448 and ECCS-2335876.

H. Boche and V. Pohl are with the School of Computation, Information and Technology, Technical University of Munich, 80333 München, Germany, {boche, volker.pohl}@tum.de. H. Boche is also with the Munich Center for Quantum Science and Technology (MCQST), 80799 München, Germany. H. V. Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA poor@princeton.edu.

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

This paper investigates Problems 1 and 2 for non-deterministic stationary stochastic time series $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}}$ that are characterized by smooth (i.e. differentiable) and computable spectral densities $\varphi_{\boldsymbol{x}}$. We will show that in the set of stochastic processes with smooth and computable spectral densities there exist infinitely many processes for which $\sigma^2_{\min}$ is not Turing computable. This implies that there exist no algorithm for the computation of FIR approximations (2) so that σ^2_N effectively converges to $\sigma^2_{\min}$ and such that the filter coefficients $h^{(N)}_n$ effectively converge to the filter coefficient h_n of the optimal Wiener prediction filter. It will be shown that for this class of smooth and computable spectral densities there does *not* exist a computable stopping criterion for calculating the minimum mean square prediction error $\sigma^2_{\min}$, and there exist no stopping criterion for computing the impulse response of the optimal causal Wiener prediction filter.

The organization of this paper is as follows. Section II introduces our notation and gives a very short review of the concepts of computability analysis that we use. Then Section III gives a more detailed problem formulation from the point of view of prediction theory for stochastic processes. Subsequently, Sections IV and V present our main results with some discussion. The paper closes with a short outlook in Section VI. Because of space constraints, this paper contains no proofs. The technical and long proofs of our main results will be presented in an extended journal publication [19].

II. Notation and Preliminaries

Subsection II-A defines our main notation whereas Subsection II-B briefly recalls basic concepts from computability analysis [20]–[22] as far as needed in this paper.

A. General notation

$\mathbb{D} = \{z \in \mathbb{C} \,:\, |z| < 1\}$ stands for the open *unit disk* and $\mathbb{T} = \{z \in \mathbb{C} \,:\, |z| = 1\}$ for the *unit circle* in the complex plane $\mathbb{C}$. For an arbitrary finite positive measure μ on $\mathbb{T}$ and for any $1 \leq p < \infty$, we write $L^p(\mu)$ for the Banach spaces of integrable functions on $\mathbb{T}$ with norm

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

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

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

If μ is the Lebesgue measure, we simply write $L^p(\mathbb{T})$, and we notice that $L^2(\mu)$ is a Hilbert space with 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, $\mathcal{C}(\mathbb{T})$ denotes the Banach space of continuous functions on $\mathbb{T}$ with maximum norm $\|f\|_\infty = \max_{\zeta\in\mathbb{T}}|f(\zeta)|$.

We write $H(\mathbb{D})$ for the set of all functions that are holomorphic (i.e. analytic) in the unit disk $\mathbb{D}$, and $H^\infty(\mathbb{D})$ is the Banach space of all bounded analytic functions, i.e. the set of all functions $f \in H(\mathbb{D})$ with $\|f\|_\infty = \sup_{|z|<1}|f(z)| < \infty$. For any $f \in H^\infty(\mathbb{D})$ 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 $L^\infty(\mathbb{T})$. Therewith, $H^\infty(\mathbb{D})$ is the closed subspace of all $f \in L^\infty(\mathbb{D})$ that can be written as a power series $f(z) = \sum_{n=0}^{\infty} c_n(f) z^n$. Finally, $H^\infty_0(\mathbb{D}) = \{f \in H^\infty(\mathbb{D}) : f(0) = 0\}$.

B. Computability analysis

We work with the standard model of a *Turing machine* [21], [23], [24], which is an abstract device that provides a theoretical model describing the fundamental limits of any realizable digital computer.

Definition II.1 (Computable number): *A number $x \in \mathbb{R}$ is said to be* computable *if there exists a Turing machine* TM *with input $n \in \mathbb{N}$ and output $\xi(n) = \mathrm{TM}(n) \in \mathbb{Q}$, such that*

$$|x - \xi(n)| \leq 2^{-n}\,, \qquad \textit{for all } n \in \mathbb{N}\,. \tag{3}$$

The set of all computable real numbers is denoted by $\mathbb{R}_{\mathrm{c}}$.

Remark: If (3) is satisfied, one says that the sequence $\{\xi(n)\}_{n\in\mathbb{N}}$ *effectively* converges to x.

Note that $\mathbb{R}_{\mathrm{c}}$ is a proper subfield of $\mathbb{R}$. The following simple characterization of $\mathbb{R}_{\mathrm{c}}$ is of particular relevance for this paper.

Lemma II.1: *A real number $x \in \mathbb{R}$ is computable if and only if there exist two sequences $\{\xi_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ and $\{\zeta_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ such that $\xi_{n+1} \leq \xi_n$ and $\zeta_{n+1} \geq \zeta_n$ for all $n \in \mathbb{N}$, and such that $\lim_{n\to\infty}\xi_n = \lim_{n\to\infty}\zeta_n = x$.*

So $x \in \mathbb{R}$ is computable if and only if there exists a monotonically decreasing and a monotonically increasing sequence that both converges to x. Our main result will be based on the fact that for the computability of x it is not sufficient that only one of these sequences exist. Therefore, we will need the set $\Pi_{\mathrm{u}} \subset \mathbb{R}$ of all $x \in \mathbb{R}$ for which there exists a sequence $\{\xi_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ such that

$$\xi_{n+1} \leq \xi_n \text{ for all } n \in \mathbb{N} \qquad \text{and} \qquad \lim_{n\to\infty}\xi_n = x\,.$$

Then we will use the fact that $\mathbb{R}_{\mathrm{c}} \subsetneq \Pi_{\mathrm{u}}$, i.e. there exist numbers $x \in \Pi_{\mathrm{u}}$ that are not computable.

Later, we need to make the assumption that our given spectral density can be processed by a digital computer in order to compute the corresponding optimal prediction filter and the corresponding MMSE. To his end, the spectral density needs to be a computable function.

Definition II.2 (Computable function): *A function $f : \mathbb{T} \to \mathbb{R}$ is said to be* computable *if there exists a computable sequence of real trigonometric polynomials $\{p_m\}_{m\in\mathbb{N}}$ that effectively and uniformly converges to f on $\mathbb{T}$, i.e. if there exists a Turing machine* TM $: \mathbb{N} \to \mathbb{N}$ *such that $m \geq \mathrm{TM}(N)$ implies $|f(\zeta) - p_m(\zeta)| \leq 2^{-N}$ for all $\zeta \in \mathbb{T}$.*

Remark: Note that the previous definition implies that any computable function $f : \mathbb{T} \to \mathbb{R}$ is necessarily continuous (see, e.g., [20]), and we will write $\mathcal{C}_{\mathrm{c}}(\mathbb{T})$ for the set of all continuous computable functions on $\mathbb{T}$.

III. Prediction Theory and Problem Statements

This section briefly recalls the main concepts and notation from prediction theory as far as they are needed in this paper.

We refer to standard textbooks and recent overview articles (e.g., [25]–[29]) for details. At the same time, we give a more detailed problem statement.

A. Stationary stochastic processes

If $(\Omega, \mathscr{F}, \nu)$ is a probability space, then $\mathcal{R} = \mathcal{R}(\Omega, \mathscr{F}, \nu)$ denotes 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 $\mathrm{E}[|x|^2] < \infty$. This space is a Hilbert space if the inner product is defined by the *covariance* of two rvs, i.e.

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

with the corresponding norm $\|x\|_{\mathcal{R}} = \sqrt{\mathrm{E}[|x|^2]}$. 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 $n, k \in \mathbb{Z}$. The corresponding function $\gamma_{\boldsymbol{x}}(n) = \langle x_n, x_0\rangle_{\mathcal{R}}$, $n \in \mathbb{Z}$ is said to be the *auto-covariance function* of $\boldsymbol{x}$. To every wss stochastic process $\boldsymbol{x}$ 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 $x_n = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\theta}\mathrm{d}Z_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})$ for all $n \in \mathbb{Z}$, and the auto-covariance has the *spectral representation*

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

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

$$\mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = \varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta + \mathrm{d}\mu_{\mathrm{s}}(\mathrm{e}^{\mathrm{i}\theta}) \tag{4}$$

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

According to the *Wold decomposition*, 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 $\boldsymbol{x}^{\mathrm{s}} = \{x_n^{\mathrm{s}}\}$. The spectral measure of $\boldsymbol{x}^{\mathrm{r}}$ is the absolutely 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 part is identical to zero. Such sequences are called *purely non-deterministic*.

B. The minimum MSE of linear prediction

An important practical problem is to find the best linear predictor $\widehat{x}_n$ of x_n from finitely (or infinitely) many observations of the sequence $\boldsymbol{x}$. Without loss of generality, we only discuss the prediction of x_0 from past observations of $\boldsymbol{x}$, i.e. from observations of $\{x_{-1}, x_{-2}, x_{-3,\dots}\}$. Then the optimal linear prediction is given by

$$\widehat{x}_0 = \underset{x\in\mathcal{X}_{[-\infty,-1]}}{\arg\min}\ \|x - x_0\|_{\mathcal{R}}^2 = \mathrm{P}_{[-\infty,-1]}(x_0)\ , \tag{5}$$

wherein $\mathcal{X}_{[-\infty,-1]} = \overline{\operatorname{span}}\{x_n : n \leq -1\} \subset \mathcal{R}$ stands for the closed subspace spanned by $\{x_{-1}, x_{-2}, \dots\}$ and where $\mathrm{P}_{[-\infty,-1]} : \mathcal{X} \to \mathcal{X}_{[-\infty,-1]}$ denotes the orthogonal projection from $\mathcal{X} = \overline{\operatorname{span}}\{x_n : n \in \mathbb{Z}\}$ onto $\mathcal{X}_{[-\infty,-1]}$. The resulting MMSE is then

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

If $\sigma_{\min}^2 = 0$ then x_0 can be perfectly predicted from the past observations and so $\boldsymbol{x}$ is called *deterministic*. If, on the other hand, $\sigma_{\min}^2 > 0$, the process $\boldsymbol{x}$ is said to be *non-deterministic*. We only consider non-deterministic stochastic processes and the following theorem characterizes such stochastic processes $\boldsymbol{x}$ in terms of their spectral measure.

Theorem III.1: *Let $\boldsymbol{x}$ be a wss stochastic sequence with spectral measure* (4)*. Then $\boldsymbol{x}$ is non-deterministic if and only if $\log\varphi_{\boldsymbol{x}} \in L^1(\mathbb{T})$, i.e. if and only if*

$$\int_{-\pi}^{\pi} \log\varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta > -\infty\ . \tag{6}$$

In this case the minimum mean square error is given by

$$\sigma_{\min}^2(\varphi_{\boldsymbol{x}}) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta\right) > 0\ . \tag{7}$$

Remark: Condition (6), is also known as *Szegö's condition* [30], whereas (7) is known as *Kolmogorov's formula* [31].

Theorem III.1 implies that the spectral measure of a non-deterministic wss stochastic process $\boldsymbol{x}$ has necessarily a non-vanishing spectral density $\varphi_{\boldsymbol{x}}$, and (7) shows that the MMSE $\sigma_{\min}^2$ depends only on $\varphi_{\boldsymbol{x}}$.

The MMSE, given in (7), is an important performance measure since it gives a lower bound on the achievable error for linear prediction. Therefore, it is of practical importance to compute this value for a given spectral density $\varphi_{\boldsymbol{x}}$. However, (7) already indicates that $\sigma_{\min}^2$ is generally not a simple rational number and so the question arises whether $\sigma_{\min}^2$ is generally a computable number:

Question 1: Let $\mathcal{M}_{\mathrm{D}}$ be a set of smooth and computable spectral densities. Does there exist a Turing machine TM with two inputs $\varphi \in \mathcal{M}_{\mathrm{D}}$ and $M \in \mathbb{N}$ and with output $\sigma_{\varphi,M}^2 = \mathrm{TM}(\varphi, M)$ such that for all $\varphi \in \mathcal{M}_{\mathrm{D}}$ and every $M \in \mathbb{N}$, we have $\left|\mathrm{TM}(\varphi, M) - \sigma_{\min}^2(\varphi)\right| < 2^{-M}$.

The set $\mathcal{M}_{\mathrm{D}}$ will be defined precisely in Section IV and it will be shown that the answer to this question is negative. In fact, we will prove an even stronger statement, namely that in $\mathcal{M}_{\mathrm{D}}$ there exist infinitely many spectral densities such that for any such φ there exists *no* Turing machine TM_φ (which is designed precisely for this particular φ) with input $M \in \mathbb{N}$ and output $\sigma_M^2 = \mathrm{TM}_\varphi(M)$ so that $\left|\sigma_M^2 - \sigma_{\min}^2\right| < 2^{-M}$.

C. Optimal linear prediction filters

A linear predictor (5) has the form of a causal linear filter (1) with *impulse response* $\{h_k\}_{k=1}^\infty$ and *transfer function* $h(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{k=1}^\infty h_k\,\mathrm{e}^{\mathrm{i}k\theta}$. The input of H is a stationary stochastic process with auto-covariance $\gamma_{\boldsymbol{x}}$ and spectral measure $\mu_{\boldsymbol{x}}$. Therefore (1) converges in mean square (i.e. in the norm of $\mathcal{R}$) if the impulse response $\{h_k\}_{k\in\mathbb{N}}$ satisfies

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

i.e. if $h \in L^2(\mu_{\boldsymbol{x}})$. The filter (1) is a causal filter, i.e. $h_k = 0$ for all $k \leq 0$. Therefore, the transfer function h can be extended to a function $h(z) = \sum_{k=1}^\infty h_k z^k$ which is analytic in $\mathbb{D}$ and satisfies $h(0) = 0$. So we have to require that h belongs at least to $H_0(\mathbb{D}) \cap L^2(\mu_{\boldsymbol{x}})$.

Let $h \in H_0(\mathbb{D}) \cap L^2(\mu_{\boldsymbol{x}})$ be an arbitrary but fixed transfer function of a prediction filter (1). Then it is not hard to see, that the corresponding (MSE) is

$$\sigma_h^2 = \|x_0 - \widehat{x}_0\|_{\mathcal{R}}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - h(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) \ . \quad (8)$$

This MSE depends on the chosen transfer function h. Optimizing (8) over all $h \in H_0^\infty(\mathbb{D})$ gives the MMSE (7), i.e.

$$\sigma_{\min}^2 = \inf_{h \in H_0^\infty(\mathbb{D})} \mathrm{E}\left[\left|x_0 - \mathrm{H}(\boldsymbol{x})\right|^2\right] = \inf_{h \in H_0^\infty(\mathbb{D})} \|1 - h\|_{L^2(\mu_{\boldsymbol{x}})}^2 \ , \quad (9)$$

and the unique minimizer will be denoted by h_{opt}. A closed form expression for h_{opt} can be obtained by means of the spectral factorization of the spectral density $\varphi_{\boldsymbol{x}}$ of $\boldsymbol{x}$.

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

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

The function φ_+ is called the spectral factor *of φ.*

The following well known statement gives a necessary and sufficient condition on a spectral density $\varphi \in L^1(\mathbb{T})$ so that it possesses a spectral factorization and it provides a closed form expression for the corresponding spectral factor φ_+.

Theorem III.2: *A function $\varphi \in L^1(\mathbb{T})$ possesses a spectral factorization if and only if φ satisfies Szegö's condition (6). Then its spectral factor is given by*

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

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

If $\boldsymbol{x}$ is a non-deterministic wss stochastic process then its spectral density φ satisfies Szegö's condition (cf. Theorem III.1) and therefore φ_+ always exists. Therewith, one can express the transfer function h_{opt} of the optimal predictor.

Theorem III.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 by*

$$h_{\mathrm{opt}}(z) = \frac{\varphi_+(z) - \varphi_+(0)}{\varphi_+(z)} = 1 - \frac{\varphi_+(0)}{\varphi_+(z)} \ , \quad (10)$$

wherein φ_+ is the spectral factor of φ.

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 (10) shows that h_{opt} is an analytic function in $\mathbb{D}$ with $h_{\mathrm{opt}}(0) = 0$. Thus $h_{\mathrm{opt}} \in H_0(\mathbb{D})$ and so it generally defines an *infinite impulse response* (IIR) prediction filter (1). However, from a practical point of view, only FIR filters of the form (2) can be implemented. Such an FIR approximation might, for example, be obtained from the optimal IIR filter (1) by truncating the infinite sum at a certain degree N.

Similarly as in the previous subsection, (10) indicates that the filter coefficients of h_{opt} are not generally rational numbers. So the question arises whether it is always possible to find a Turing machine that is able to effectively compute an FIR approximation of the optimal prediction filter.

Question 2: Let $\mathcal{M}_{\mathrm{D}}$ be a set of smooth and computable spectral densities. Does there exists an algorithm that is able to determine for any $\varphi \in \mathcal{M}_{\mathrm{D}}$ a computable sequence $\{\boldsymbol{h}^{(N)}\}_{N \in \mathbb{N}} = \{h_n^{(N)} : n = 1, 2, \dots\}_{N \in \mathbb{N}}$ of FIR approximations of the impulse response $\{h_n : n \in \mathbb{N}\}$ of the optimal causal Wiener filter (10) such that $\{h_n^{(N)}\}_{N \in \mathbb{N}}$ effectively converges to h_n as $N \to \infty$ for all $n = 1, 2, \dots$.
In other words, does there exist for any $n \in \mathbb{N}$ a Turing machine TM_n with inputs $\varphi \in \mathcal{M}_{\mathrm{D}}$ and $N \in \mathbb{N}$ and with output $h_n^{(N)} = \mathrm{TM}_n(\varphi, N)$ such that for all $\varphi \in \mathcal{M}_{\mathrm{D}}$ and every $M \in \mathbb{N}$, we have $\left|h_n^{(N)} - h_n\right| < 2^{-N}$.

There exist many known algorithms that determine FIR approximations that converge to the optimal Wiener filter. So the important part of the previous question is whether these algorithms *effectively* converge, i.e. whether it is possible to algorithmically control the approximation error. We are going to show that in the set $\mathcal{M}_{\mathrm{D}}$ (which is precisely defined below), there exist infinitely many spectral densities for which this is not the case. This is even shown for $n = 1$.

IV. The Non-Computability of the Minimum MSE

Let $\boldsymbol{x}$ be a purely non-deterministic wss stationary process with spectral density φ. According to Theorem III.1, φ satisfies Szegö's condition (6) and so the MMSE (7) and the optimal linear prediction filter (10) are well defined. Nevertheless, in order to obtain stronger results, we restrict our considerations to spectral densities that belong to 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 computable continuous functions on $\mathbb{T}$ with first derivatives φ' that are also computable continuous functions on $\mathbb{T}$ and that satisfy Szegö's condition (6).

In practical applications, it is often important to compute the MMSE $\sigma_{\min}^2$. However, even through $\mathcal{M}_{\mathrm{D}}$ contains only spectral densities with very nice analytic properties, it is clear that even for fairly simple densities $\varphi \in \mathcal{M}_{\mathrm{D}}$, (7) cannot be calculated in closed form and so one generally needs numerical algorithms to compute $\sigma_{\min}^2$. This brings us to Question 1 concerning whether $\sigma_{\min}^2$ is a computable number for any arbitrary $\varphi \in \mathcal{M}_{\mathrm{D}}$.

Before we answer this question, we give a formal algorithm for the calculation of $\sigma_{\min}^2$ which also helps to illustrate the proofs of the subsequent results.

Algorithm 1: As discussed in Section III-C, the minimal MSE $\sigma_{\min}^2$ can be obtained by the optimization problem (9). Therefore, we define for $n \in \mathbb{N}$ the simpler problem

$$\sigma_n^2(\varphi) = \inf_{p \in \mathcal{P}_{n,0}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - p(\mathrm{e}^{\mathrm{i}\theta})\right|^2 \varphi(\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta \ , \quad (11)$$

where the minimization is only over the set $\mathcal{P}_{n,0}$ of all polynomials p of degree n with $p(0) = 0$. Thus σ_n^2 is the MMSE that can be achieved by predicting x_0 from its n past values $\{x_{-1}, x_{-2}, x_{-3}, \cdots, x_{-n}\}$ using a linear filter of length n. It is clear from the definition that $\sigma_{n+1}^2 \leq \sigma_n^2$

for all $n \in \mathbb{N}$ and that $\lim_{n\to\infty} \sigma_n^2 = \sigma_{\min}^2$. Therefore, one may use (11) to compute an approximation of $\sigma_{\min}^2$.

To make Algorithm 1 effective, one needs an algorithmic *stopping criterion*, i.e. one needs a Turing machine that is able to find algorithmically for any $M \in \mathbb{N}$ an index $N \in \mathbb{N}$ so that $\left|\sigma_N^2 - \sigma_{\min}^2\right| < 2^{-M}$. This brings us to

Question 3: Does there exist a Turing machine TM with two inputs $\varphi \in \mathcal{M}_{\mathrm{D}}$ and $M \in \mathbb{N}$ and with output $N = \mathrm{TM}(\varphi, M) \in \mathbb{N}$ such that for all $\varphi \in \mathcal{M}_{\mathrm{D}}$ and every $M \in \mathbb{N}$, we have $\left|\sigma_N^2(\varphi) - \sigma_{\min}^2(\varphi)\right| < 2^{-M}$?

We are going to show that both Question 1 and Question 3 have a negative answers. This will easily follow from the following theorem. Its proof will be published in [19].

Theorem IV.1: *For every $\varphi \in \mathcal{M}_{\mathrm{D}}$, the corresponding MMSE* (7) *satisfies $\sigma_{\min}^2(\varphi) \in \Pi_{\mathrm{u}}$. Conversely, to every $s \in \Pi_{\mathrm{u}}$, $s > 0$ there exists a $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ such that*

$$\sigma_{\min}^2(\varphi_s) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \varphi_s(\mathrm{e}^{\mathrm{i}\theta})\, \mathrm{d}\theta\right) = s\ .$$

Recall that the set $\mathbb{R}_{\mathrm{c}}$ of all computable numbers is a proper subset of Π_{u}. So in the second statement of Theorem IV.1, we may choose a positive number $s \in \Pi_{\mathrm{u}}$ that is not computable. This gives the following important observation.

Corollary IV.2: *To every $s \in \Pi_{\mathrm{u}}$, $s > 0$, $s \notin \mathbb{R}_{\mathrm{c}}$ there exists a spectral density $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ such that $\sigma_{\min}^2(\varphi_s) = s \notin \mathbb{R}_{\mathrm{c}}$.*

In other words there exist spectral densities $\varphi \in \mathcal{M}_{\mathrm{D}}$ for which the corresponding MMSE $\sigma_{\min}^2(\varphi)$ is not a computable number. As an immediate consequence, we have the following result concerning the approximation method of Algorithm 1.

Corollary IV.3: *Let $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ be constructed as in Corollary IV.2, and let $\sigma_{\min}^2(\varphi_s)$ and $\sigma_N^2(\varphi_s)$ be given by* (7) *and* (11)*, respectively. There exists no Turing machine TM_{φ_s} with input $M \in \mathbb{N}$ and that is able to compute a stopping index $N = \mathrm{TM}_{\varphi_s}(M) \in \mathbb{N}$ so that $\left|\sigma_N^2(\varphi_s) - \sigma_{\min}^2(\varphi_s)\right| < 2^{-M}$.*

However, Corollary IV.2 does not only imply that the particular way to approximate $\sigma_{\min}^2(\varphi)$, as described by Algorithm 1, is not effective, but that any method to approximate $\sigma_{\min}^2(\varphi)$ is generally not effective for infinitely many spectral densities $\varphi \in \mathcal{M}_{\mathrm{D}}$.

Corollary IV.4: *Let $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ be constructed as in Corollary IV.2, and let $\sigma_{\min}^2(\varphi_s)$ be given by* (7)*. There exists no Turing machine TM_{φ_s} with input $M \in \mathbb{N}$ and that is able to compute an approximation $\widetilde{\sigma}^2 = \mathrm{TM}_{\varphi_s}(M)$ so that $\left|\widetilde{\sigma}^2 - \sigma_{\min}^2(\varphi_s)\right| < 2^{-M}$.*

It is important to note that the Turing machines in the statements of Corollaries IV.3 and IV.4 are designed for the chosen spectral density φ_s. So in these corollaries it is assumed that for any chosen spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$, we are allowed to construct a particular Turing machine TM_φ that computes the desired approximation index and the desired approximation, respectively. Then the previous corollaries show that in the set $\mathcal{M}_{\mathrm{D}}$, there always exist (infinitely many) spectral densities for which such a Turing machine can not be constructed. But of course, $\mathcal{M}_{\mathrm{D}}$ contains also spectral densities for which such a Turing machine exists.

In practical applications, however, we usually look for a universal Turing machine for the whole set $\mathcal{M}_{\mathrm{D}}$. The input of such a universal Turing machine would be (a description of) an arbitrary spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ and the precision M. The output would be the desired approximation index or the desired approximation (see Questions 1 and 3). However, it is clear that if it is not possible to design for each individual $\varphi \in \mathcal{M}_{\mathrm{D}}$ a particular Turing machine with the desired properties, then it is *a fortiori* not possible to have a universal Turing machine for the wole set $\mathcal{M}_{\mathrm{D}}$.

Corollary IV.5: *The answer to Questions 1 and 3 is negative.*

Corollary IV.3 states that there is no Turing machine that is able to compute a stopping index for the approximation sequence $\left\{\sigma_n^2\right\}_{n\in\mathbb{N}}$, defined in (11). However, we would like to emphasis that Corollary IV.4 implies that the same result holds for any sequence that approximates the minimum MSE. In fact, there exist many different algorithms that aim to determine FIR approximations for the optimal prediction filter by controlling the corresponding MSE. One well known example is the Durbin–Levinson algorithm [32], [33]. Similarly as in (11), this recursive algorithm determines at each iteration an FIR approximation of h_{opt} and the corresponding MSE σ_n^2. Corollary IV.4 shows that also for this algorithm (and for any other such algorithm) it is generally impossible to decide algorithmically whether σ_n^2 is sufficiently close to the MMSE $\sigma_{\min}^2$.

V. Computability of the Optimal Wiener Filter

Let $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}} \subset \mathcal{R}$ be a wss stochastic sequence with spectral density $\varphi \in \mathcal{M}_{\mathrm{D}}$ and let (1) be the optimal linear prediction filter for estimating x_0 from the past. Then the transfer function $h_{\mathrm{opt}}(z) = \sum_{k=1}^{\infty} h_k z^k$, $|z| < 1$ of this optimal filter is given by (10). We are interested in the question as to whether it is possible to compute the corresponding filter coefficients $\{h_k\}_{k\in\mathbb{N}}$ on a digital computer. For simplicity and clarity of the presentation, we only ask whether the first coefficient h_1 is computable.

Theorem V.1: *Let $s \in \Pi_{\mathrm{u}}$, $s > 0$, $s \notin \mathbb{R}_{\mathrm{c}}$ be arbitrary and let $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ be the spectral density associated with s as in Corollary IV.2. If $\{h_n\}_{n=1}^{\infty}$ is the impulse response of the optimal prediction filter associated with φ_s then $h_1 \notin \mathbb{R}_{\mathrm{c}}$.*

So there are always exist wss stochastic processes $\boldsymbol{x}$ (whose spectral density $\varphi_{\boldsymbol{x}}$ can be constructed as in Theorem IV.1) such that the first filter coefficient h_1 of the optimal prediction filter is not a computable number. As a consequence of Theorem V.1, we obtain that the answer to Question 2 is negative.

Corollary V.2: *Let $s \in \Pi_{\mathrm{u}}$, $s > 0$, $s \notin \mathbb{R}_{\mathrm{c}}$ be arbitrary and let $\varphi_s \in \mathcal{M}_{\mathrm{D}}$ be the spectral density associated with s as in Corollary IV.2. Then there exists* no *computable sequence $\left\{\boldsymbol{h}^{(N)}\right\}_{N\in\mathbb{N}} = \left\{h_n^{(N)} : n = 1, 2, \dots\right\}_{N\in\mathbb{N}}$ of FIR approximations of the optimal causal Wiener filter such that $\left\{h_1^{(N)}\right\}_{N\in\mathbb{N}}$ effectively converges to h_1 as $N \to \infty$.*

Remark: The same statement can be proved for IIR approximations.

As mentioned before, there exist many different algorithms that are designed to give an good approximation of the optimal (IIR) prediction filter. In particular, all these algorithms can determine a computable sequence $\{h_1^{(N)}\}_{N\in\mathbb{N}}$ of computable numbers that converge to the first filter coefficient of h_{opt} in (10) as $N \to \infty$. Then, from a practical point of view, one needs a stopping criterion that tells us at which $N \in \mathbb{N}$ we can stop such that $h_1^{(N)}$ is sufficiently close to h_1, i.e. we need an algorithm that computes for any given $M \in \mathbb{N}$ an $N \in \mathbb{N}$ so that $\left|h_1^{(N)} - h_1\right| < 2^{-M}$. Theorem V.1 and Corollary V.2 show that such a general stopping criterion can not exist.

VI. Conclusions and Future Works

It has been shown that for wss stochastic processes $\boldsymbol{x}$ with spectral densities $\varphi_{\boldsymbol{x}}$ in the set $\mathcal{M}_{\text{D}}$, the impulse response of the optimal Wiener prediction filter and the minimum mean square error of this prediction is not generally Turing computable. In particular, there is no algorithmic stopping criterion for calculating these quantities on digital computers. So all of our Questions 1–3 have a negative answer for the set $\mathcal{M}_{\text{D}}$ of differentiable and computable spectral densities that satisfy Szegö's condition.

However, it is clear that for sufficiently small sets of spectral densities, the answers to our three questions might be positive. So it is an interesting question for future research to characterize conditions on the spectral densities such that the MMSE and the optimal impulse response are guaranteed to be Turing computable.

References

[1] A. N. Kolmogorov, "Interpolation und Extrapolation von stationären zufälligen Folgen," *Bull. Acad. Sci. (URSS), Ser. Math.*, vol. 5, pp. 3–14, 1941.

[2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. Cambridge: The MIT Press and John Wiley & Sons, Inc., 1949.

[3] H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," *Proc. IRE*, vol. 38, pp. 417–425, Apr. 1950.

[4] S. Darlington, "Linear least-squares smoothing and prediction, with applications," *Bell Sys. Techn. J.*, vol. 37, no. 5, pp. 1221–1294, Sep. 1958.

[5] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, no. 7, pp. 645–655, 1950.

[6] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME (Serie D), J. Basic Engineering*, vol. 82, pp. 35–55, 1960.

[7] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press, 2011.

[8] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Autom. Control*, vol. AC-25, no. 3, pp. 531–536, Jun. 1980.

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, New Jersey: John Wiley & Sons Inc., 2015.

[10] F. L. Lewis, L. Xie, and D. Popa, *Optimal and Robust Estimation: With an Introduction to Stochastic Control Theory*. Boca Raton: CRC Press, 2008.

[11] K. Nar and T. Başar, "Sampling multidimensional Wiener processes," in *Proc. 53rd IEEE Conf. on Decision and Control (CDC)*, Los Angeles, CA, USA, Dec. 2014, pp. 3426–3431.

[12] F. Oldewurtel, C. N. Jones, A. Parisio, and M. Morari, "Stochastic model predictive control for building climate control," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 3, pp. 1198–1205, May 2014.

[13] J. P. González-Coma, J. Rodríguez-Fernández, N. González-Prelcic, L. Castedo, and R. W. Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[14] E. Aslan, M. E. Çelebi, and F. Pekergin, "Wiener and Kalman detection methods for molecular communications," *IEEE Trans. NanoBioscience*, vol. 21, no. 2, pp. 256–264, Apr. 2022.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, etc.: Springer, 1994.

[16] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.

[17] J. Dong, S. Roth, and B. Schiele, "Deep Wiener deconvolution: Wiener meets deep learning for image deblurring," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 1048–1059.

[18] M. Z. A. Bhotto and A. Antoniou, "New improved recursive least-squares adaptive-filtering algorithms," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 6, pp. 1548–1558, Jun. 2013.

[19] H. Boche, V. Pohl, and H. V. Poor, "The optimal prediction is not computable," 2023, in preparation.

[20] M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics*. Berlin: Springer-Verlag, 1989.

[21] K. Weihrauch, *Computable Analysis*. Berlin: Springer-Verlag, 2000.

[22] J. Avigad and V. Brattka, "Computability and analysis: The legacy of Alan Turing," in *Turing's Legacy: Developments from Turing's Ideas in Logic*, ser. Lecture Notes in Logic, Bd. 42. New York: Cambridge University Press, 2014, pp. 1–47.

[23] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, vol. s2-42, no. 1, pp. 230–265, 1937.

[24] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem. A correction," *Proc. London Math. Soc.*, vol. s2-43, no. 1, pp. 544–546, 1938.

[25] J. L. Doob, *Stochastic Processes*. New York: John Wiley, 1953.

[26] Y. A. Rozanov, *Stationary Random Processes*. San Francisco: Holden-Day, 1967.

[27] A. N. Shiryaev, *Probability*, 2nd ed. New York: Springer-Verlag, 1996.

[28] N. Bingham, "Szegö's theorem and its probabilistic descendants," *Probab. Surv.*, vol. 9, pp. 287–324, 2012.

[29] H. Boche, V. Pohl, and H. V. Poor, "Recent progress in computability for prediction and Wiener filter theory," *Trans. A. Razmadze Math. Inst.*, vol. 176, no. 3, pp. 323–344, Dec. 2022.

[30] G. Szegö, "Beiträge zur Theorie the Toeplitzen Formen (Erste Mitteilung)," *Math. Z.*, vol. 6, pp. 167–202, 1920.

[31] A. N. Kolmogorov, "Stationary sequences in Hilbert spaces," in *Selected Works of A.N. Kolmogorov, Vol. 2: Probability Theory and Mathematical Statistics*, A. N. Shiryayev, Ed. Dordrecht: Springer, 1992, pp. 228–271.

[32] N. Levinson, "The Wiener (root mean square) error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, no. 1-4, pp. 261–278, Apr. 1946.

[33] J. Durbin, "The fitting of time-series models," *Rev. Int. Statist. Inst.*, vol. 28, no. 3, pp. 233–244, 1960.