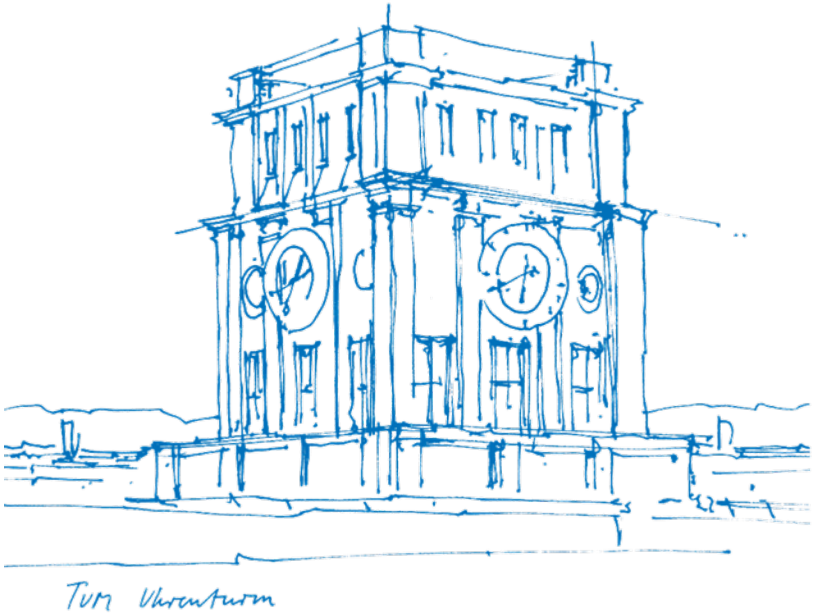
The Wiener Theory of Causal Linear Prediction Is Not Effective

Holger Boche[†] **Volker Pohl**[†] **H. Vincent Poor**[*]

[†]Technical University of Munich
Department of Electrical and Computer Engineering

[*]Princeton University
Department of Electrical an Computer Engineering

62nd IEEE Conference on Decision and Control (CDC)
Dec. 13.–15, 2023, Singapore



Content & Outline of the talk

- ▷ We consider the problem of predicting wss stochastic sequences.
- ▷ There exists well established and beautiful theory due to Kolmogorov, Wiener, ...
  - „Closed form“ solution for the minimum mean square error (MMSE).
  - „Closed form“ solution for the optimal (in the MSE sense) linear prediction filter.
- ▷ There are many (iterative) algorithms for computing these quantities. However there seems to be no stopping rules.
- ▷ **Question**: Can we effectively compute these „closed form“ expressions?
- ▷ **Answer**: No, not in general.

Classical problem of causal prediction

- Let $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{Z}}$ be a wide-sense stationary (wss) stochastic sequences.
- Predict x_0 from past observations of $\boldsymbol{x}$ using a linear prediction filter

$$\widehat{x}_0 = \mathrm{H}(\boldsymbol{x}) = \sum_{n=1}^{\infty} h_n x_{-n} \;=\; h_1 x_{-1} + h_2 x_{-2} + h_3 x_{-3} + \dots .$$

- Determine the impulse response $\boldsymbol{h} = \{h_n\}_{n=1}^{\infty}$ of the filter such that the mean square error (MSE)

$$\sigma_{\boldsymbol{h}}^2 \;=\; \mathrm{E}[|\widehat{x}_0 - x_0|^2] \;=\; \mathrm{E}[|\mathrm{H}(\boldsymbol{x}) - x_0|^2]$$

  is minimized.
- The minimum mean square error (MMSE)

$$\sigma_{\min}^2 = \min_{\boldsymbol{h}} \sigma_{\boldsymbol{h}}^2 = \min_{\boldsymbol{h}} \mathrm{E}[|\mathrm{H}(\boldsymbol{x}) - x_0|^2]$$

  is an importance performance measure.

Classical and beautiful theory (Kolmogorov, Wiener, ...)

- Every wss stochastic sequence is characterized by its spectral measure $\mu_{\boldsymbol{x}}$.
- One is basically interested in non-deterministic stochastic sequences where the future can not perfectly be predicted from the past.

Theorem: Let $\boldsymbol{x}$ be a wss stochastic sequence with spectral measure $\mathrm{d}\mu_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta}) = \varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta + \mathrm{d}\mu_{\mathrm{s}}(\mathrm{e}^{\mathrm{i}\theta})$. 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

$$\int_{-\pi}^{\pi} \log \varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta > -\infty\,. \qquad \text{(Szegö's condition)}$$

In this case the minimum mean square error is given by

$$\sigma^2_{\min}(\varphi_{\boldsymbol{x}}) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \varphi_{\boldsymbol{x}}(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta\right) > 0\,. \qquad \text{(Kolmogorov's formula)}$$

Optimal prediction filter

Definition: 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

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

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

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

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

Theorem: Let $\boldsymbol{x}$ be a purely non-deterministic wss stochastic process with spectral density φ. Then the optimal prediction filter for estimating x_0 is given by

$$h_{\mathrm{opt}}(z) = \sum_{n=1}^{\infty} h_n z^n = \frac{\varphi_+(z) - \varphi_+(0)}{\varphi_+(z)} = 1 - \frac{\varphi_+(0)}{\varphi_+(z)} , \qquad |z| < 1 . \tag{1}$$

wherein φ_+ is the spectral factor of φ.

Algorithms for determine the MMSE and the optimal filter

- For practical implementation, the optimal prediction filter (1) is approximated by an finite impulese response (FIR) filter of finite degree $N \in \mathbb{N}$

$$\widehat{x}_{0,N} = \mathrm{H}_N(\boldsymbol{x}) = \textstyle\sum_{n=1}^{N} h_n^{(N)} x_{-n} \tag{2}$$

- This yields an MSE

$$\sigma_N^2 = \mathrm{E}[|\widehat{x}_{0,N} - x_0|^2] = \mathrm{E}[|\mathrm{H}_N(\boldsymbol{x}) - x_0|^2]$$

- This MSE is monotonically decreasing in N and converges to the MMSE:

$$\sigma_{N_1}^2 \geq \sigma_{N_1}^2 \geq \sigma_{\min}^2 \quad \text{for all } N_2 \leq N_1 \qquad \text{and} \qquad \lim_{N\to\infty} \sigma_N^2 = \sigma_{\min}^2 \,.$$

- There exists many different algorithms which determine recursively

  the impulse response of the FIR filter $\left\{h_n^{(N)}\right\}_{n=1}^{N}$ together with the corresponding MSE σ_N^2

  Example: Durbin–Levinson algorithm

- How to choose the degree N of the FIR approximation?

Problem 1: Existence of a specific stopping rule

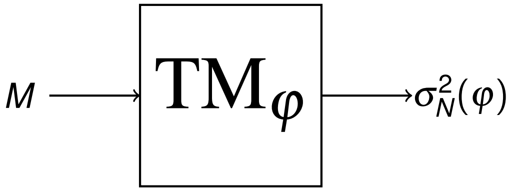
Given an arbitrary precision $M \in \mathbb{N}$, choose $N \in \mathbb{N}$ such that

$$\left|\sigma_N^2 - \sigma_{\min}^2(\varphi_{\boldsymbol{x}})\right| < 2^{-M}\,.$$

Problem 1: Existence of a specific Turing machine

Given a spectral density φ of a wss stochastic sequence $\boldsymbol{x}$. Is it possible to find a specific Turing machine TM_φ with input $M \in \mathbb{N}$ and and output $N_0 = \mathrm{TM}_\varphi(M)$ such that

$$N \geq N_0 \qquad \text{implies} \qquad \left|\sigma_N^2 - \sigma_{\min}^2(\varphi_{\boldsymbol{x}})\right| < 2^{-M}?$$



Problem 2: Existence of a universal algebraic stopping rule

Problem 2: Existence of a universal Turing machine

Given a set $\mathscr{M}_{\mathrm{D}}$ of spectral densities. Is it possible to find a universal Turing machine $\mathrm{TM}_{\mathscr{M}_{\mathrm{D}}}$ with two inputs $\varphi \in \mathscr{M}_{\mathrm{D}}$ and $M \in \mathbb{N}$ and and output $N_0 = \mathrm{TM}_{\varphi}(\varphi, M)$ such that

$$N \geq N_0 \qquad \text{implies} \qquad \left|\sigma_N^2 - \sigma_{\min}^2(\varphi_{\boldsymbol{x}})\right| < 2^{-M}?$$

Computable Analysis – Computable numbers and functions

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

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

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

Definition (Computable Function)

A function $\varphi : \mathbb{T} \to \mathbb{R}$ is said to be *computable* if there exists a computable sequence of real trigonometric polynomials $\{p_N\}_{N \in \mathbb{N}}$ that effectively and uniformly converges to f on $\mathbb{T}$,
i.e. there exists a Turing machine $\mathrm{TM} : \mathbb{N} \to \mathbb{N}$ such that

$$N \geq \mathrm{TM}(M) \qquad \text{implies} \qquad |\varphi(\zeta) - p_N(\zeta)| \leq 2^{-M} \quad \text{for all } \zeta \in \mathbb{T}.$$

Set of stochastic processes

We consider stochastic processes with spectral densities which are computable continuous functions on the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ with first derivatives φ' that are also computable continuous functions on $\mathbb{T}$ and that satisfy Szegö's condition, i.e.

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

Generally, the MMSE is not a computable number

Theorem 1: There exist spectral densities $\varphi \in \mathscr{M}_{\mathrm{D}}$ such that the MMSE $\sigma^2_{\min}(\varphi)$ is not a computable number.

Remark: In fact, to every $s \in \mathbb{R}$, $s > 0$ which is not computable, there exists a spectral density $\varphi_s \in \mathscr{M}_{\mathrm{D}}$ such that $\sigma^2_{\min}(\varphi_s) = s \notin \mathbb{R}_{\mathrm{c}}$

No specific algorithmic stopping criterion for the MMSE

Theorem 2: For every spectral density $\varphi \in \mathscr{M}_{\mathrm{D}}$ from Theorem 1, and for every monotonically decreasing approximation sequence $\{\sigma_N^2\}_{N\in\mathbb{N}}$ with

$$\lim_{N\to\infty} \sigma_N^2 = \sigma_{\min}^2(\varphi)$$

there exists no specific Turing machine TM_φ with input $M \in \mathbb{N}$ that is able to compute a stopping index $N_0 = \mathrm{TM}_\varphi(M)$ such that

$$N \geq N_0 \qquad \text{implies} \qquad \left|\sigma_N^2 - \sigma_{\min}^2(\varphi)\right| < 2^{-M}\,.$$

Corollary: Problem 1 (existence of a specific stopping rule) is generally not solvabel.

No universal stopping criterion for the MMSE

- In practice, we look for a universal stopping criterion which is able to determine the stopping index for all spectral densities in a certain set, e.g. for all $\varphi \in \mathscr{M}_{\mathrm{D}}$.
- Since there exists no specific stopping criterion for some $\varphi \in \mathscr{M}_{\mathrm{D}}$, it is clear that there exists no universal stopping criterion for $\mathscr{M}_{\mathrm{D}}$.

Theorem 3: There exists no universal Turing machine TM with two inputs $\varphi \in \mathscr{M}_{\mathrm{D}}$ and $M \in \mathbb{N}$ and with output $N_0 = \mathrm{TM}(\varphi, M) \in \mathbb{N}$ such that for all $\varphi \in \mathscr{M}_{\mathrm{D}}$ and for every monotonically decreasing approximation sequence $\{\sigma_N^2\}_{N\in\mathbb{N}}$ with

$$\lim_{N\to\infty} \sigma_N^2 = \sigma_{\min}^2(\varphi)$$

one has

$$N \geq N_0 = \mathrm{TM}(\varphi, M) \qquad \text{implies} \qquad \left|\sigma_N^2(\varphi) - \sigma_{\min}^2(\varphi)\right| < 2^{-M}.$$

Corollary: Problem 2 (existence of a universal stopping rule for $\mathscr{M}_{\mathrm{D}}$) is generally not solvabel.

The optimal Wiener prediction filter is not effectively computable

- ▷ So there is no stopping criterion for the MMSE.
- ▷ It might still be possible to compute the prediction optimal filter.

Theorem 4: Let $\varphi \in \mathscr{M}_{\mathrm{D}}$ be a spectral density as in Theorem 1, and let $\{h_n\}_{n=1}^{\infty}$ be the impulse response of the corresponding optimal prediction filter. Then

- h_1 is not a computable number, i.e. there is no stopping criterion for any sequence which approximates h_1.
- there exists *no* computable sequence $\{\boldsymbol{h}^{(N)}\}_{N\in\mathbb{N}} = \{h_n^{(N)} : n = 1, 2, \ldots, N\}_{N\in\mathbb{N}}$ of FIR approximations that effectively converges to the optimal impulse response.

Summary, Outlook

- We showed that for spectral densities in $\mathscr{M}_{\mathrm{D}}$ the MMSE and the optimal prediction filter is not effectively computable, in general, i.e.
  - $\mathscr{M}_{\mathrm{D}}$ contains spectral density φ so that no specific stopping rule for the computation of $\sigma^2_{\min}(\varphi)$ exists.
  - $\mathscr{M}_{\mathrm{D}}$ contains spectral densities for which the MMSE is effectively computable -> specific stopping criterion.
  - there is no universal stopping rule for $\mathscr{M}_{\mathrm{D}}$.

Further Problems:

- Find subsets of $\mathscr{M} \subset \mathscr{M}_{\mathrm{D}}$ such that
  - for every $\varphi \in \mathscr{M}$ the MMSE $\sigma^2_{\min}(\varphi)$ is computable -> there exists at least specific stopping criteria.
  - there exists a universal stopping rule for $\mathscr{M}$