Sub-ideal causal smoothing filters for real sequences

Nikolai Dokuchaev
Department of Mathematics & Statistics, Curtin University,
GPO Box U1987, Perth, 6845 Western Australia
e-mail N.Dokuchaev@curtin.edu.au. Tel.: 61 8 92663144.

December 1, 2014

Abstract

The paper considers causal smoothing of the real sequences, i.e., discrete time processes in a setting without a probability measure. A family of causal linear time-invariant filters are suggested. These filters are sub-ideal meaning that they approximate the gain decay for some family of ideal smoothing filters that transfer a sequence into a predictable one, i.e., into a non-random sequence such that its the future values are uniquely defined by the past values. In this sense, the suggested filters are sub-ideal; a faster gain decay would lead to the loss of causality. Application to predicting algorithms are discussed.

Key words: LTI filters, smoothing filters, casual filters, predicting, sub-ideal filters, Hardy spaces.

AMS 2010 classification: 42A38, 93E11, 93E10, 42B30

1 Introduction

The paper studies causal smoothing of the real sequences. The consideration is restricted by the causal continuous time linear time-invariant filters (LTI filters), i.e. linear filters represented as convolution integrals over the historical data. These filters are used in dynamic smoothing, when the future values of the process are not available.

For many applications, it is preferable to replace a process by a more smooth process. In the frequency domain, smoothing means reduction of the energy on the higher frequencies. In particular, an ideal low-pass filter is a smoothing filter. However, this filter is not causal, i.e., it requires the future value of the process. Moreover, a filter with too high rate of decay of the frequency response at a certain point of the unit circle also cannot be causal, since a filter with this feature could transform a general kind of a process into a predictable process; clearly, this would be impossible for a causal filter. It follows from the fact that a sufficiently high rate of decay of the energy at a point of the unit circle implies predictability of the processes [11].
In continuous time setting, analytic functions are predictable, and an ideal smoothing filter converts a function into an analytic one. For discrete time processes or sequences, it is not obvious how to define an analog of the continuous time analyticity. A classical approach is to consider predicability instead of analyticity. This problem is related to the concept of the randomness for the real sequences in the pathwise setting without a probability measure. There are many classical works devoted to this extremely important concept, starting from Mises (1919) [24], Church (1940) [6], Kolmogorov (1965) [21], Loveland (1966) [23]; see the references in [22]. So far, the predicability criterion and the related restrictions on the gain decay for causal filters for stochastic Gaussian stationary discrete time processes in the frequency domain setting are given by the classical Szegő-Kolmogorov Theorem. This theorem says that the optimal prediction error is zero if

\[ \int_{-\pi}^{\pi} \log(\phi(e^{i\omega})) \, d\omega = -\infty, \]  

(1)

where \( \phi \) is the spectral density \( \phi \); see Kolmogorov [21], Szegő [27, 28], Verblunsky [29], and more recent literature reviews in [3, 26]. This means that a stationary Gaussian process is predictable if its spectral density is vanishing on a part of the unit circle \( \{ z \in \mathbb{C} : |z| = 1 \} \), i.e., if the process is “band-limited” in this sense. This result was expanded on more general stable stochastic processes allowing spectral representations with spectral density via processes with independent increments; see, e.g., [7]. The stochastic setting is the most common in causal smoothing and sampling; see, e.g., [1, 2, 4, 5, 8, 9, 11, 12, 14, 15, 16, 17, 19, 21, 25, 24].

In the present paper, the setting does not involve stochastic processes and probability measure; it is oriented on smoothing the real sequences. We assume that the smoothness of a real sequence is defined by its predicability. We readdress the problem of developing causal smoothing filters that can be arbitrarily close to some ideal non-causal smoothing filters transforming sequences of a general type into predicable ones. We consider this problem for real sequences, i.e., in deterministic pathwise setting, without any probabilistic assumptions. We use the criteria of predictability for real valued sequences obtained in [11, 12], where some predictors were developed. It can be noted that these predictors can be as well applied as well in stochastic setting for classes of processes with unknown shape of the spectral representation.

We suggest a family of causal smoothing filters with "almost” ideal rate of damping the energy on the higher frequencies and with the frequency response that can be selected to approximate the real unity uniformly on an arbitrarily large part of the unit circle. These filters are sub-ideal in the sense that their effectiveness in the damping of higher frequencies cannot be exceeded; a faster decay of the frequency response is not possible for causal filters. This is because this family of causal filters approximates the decay rate of a reference set of non-causal ”ideal” filters [2]; these ideal filters transform a non-predicable processes into predictable ones such that the predictor from
can be used. This follows the approach developed in [10] for the continuous time setting. We investigate a possibility to use the suggested filters to improve the performance of the predictors from [11] in the presence of noise contamination.

Some definitions and notations

We denote by $\mathbb{Z}$ the set of all integers.

We denote by $L_2(S)$ the usual Hilbert space of complex valued square integrable functions $x : S \to \mathbb{C}$, where $S$ is an interval in $\mathbb{R}$. We denote by $\ell_r$ the set of all sequences $x = \{x(t)\}_{t \in \mathbb{Z}} \subset \mathbb{R}$, such that $\|x\|_{\ell_r} = \left(\sum_{t=-\infty}^{\infty} |x(t)|^r\right)^{1/r} < +\infty$ for $r \in [1, \infty)$ or $\|x\|_{\ell_\infty} = \sup_{t \in \mathbb{Z}} |x(t)| < +\infty$ for $r = +\infty$.

Let $\ell_2^+$ be the set of all sequences $x \in \ell_2$ such that $x(t) = 0$ for $t = -1, -2, -3, \ldots$.

Let $D \overset{\Delta}{=} \{z \in \mathbb{C} : |z| \leq 1\}$, let $D^c \overset{\Delta}{=} \mathbb{C} \setminus D$, and let $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

For $x \in \ell_1$ or $x \in \ell_2$, we denote by $X = Zx$ the Z-transform

$$X(z) = \sum_{t=-\infty}^{\infty} x(t)z^{-t}, \quad z \in \mathbb{C}.$$  

Respectively, the inverse Z-transform $x = Z^{-1}X$ is defined as

$$x(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{i\omega})e^{i\omega t}d\omega, \quad t = 0, \pm 1, \pm 2, \ldots$$

If $x \in \ell_1$, then $X|_{\mathbb{T}}$ is defined as an element of $L_2(\mathbb{T})$, i.e., $X(e^{i\omega}) \in L_2(-\pi, \pi)$. If $X(e^{i\omega}) \in L_1(-\pi, \pi)$, then $x = Z^{-1}X$ is defined as an element of $\ell_1$.

For $r \in [1, +\infty)$, let $H^r$ be the Hardy space of functions that are holomorphic on $D^c$ including the point at infinity with finite norm $\|h\|_{H^r} = \sup_{\rho > 1} \|h(\rho e^{i\omega})\|_{L_r(-\pi, \pi)}, \quad r \in [1, +\infty)$ (see, e.g., [13]). Note that Z-transform defines a bijection between the sequences from $\ell_2^+$ and the restrictions (i.e., traces) $X|_{\mathbb{T}}$ of the functions from $H^2$ on $\mathbb{T}$ such that $\overline{X(e^{i\omega})} = X(e^{-i\omega})$. If $X(e^{i\omega}) \in L_1(-\pi, \pi)$ and $X(e^{i\omega}) = X(e^{-i\omega})$, then $x = Z^{-1}X$ is defined as an element of $\ell_\infty^+$.

2 Problem setting

Let $x(t)$ be a real sequence, $t \in \mathbb{Z}$. The output of a linear filter is the process

$$y(t) = \sum_{s=-\infty}^{\infty} h(t-s)x(s),$$

where $h : \mathbb{Z} \to \mathbb{R}$ is a given impulse response function.

If $h(t) = 0$ for $t < 0$, then the output of the corresponding filter is

$$y(t) = \sum_{s=-\infty}^{t} h(t-s)x(s).$$
In this case, the filter and the impulse response function are said to be causal. The output of a causal filter at time $t$ can be calculated using only past historical values $x(s)|_{s \leq t}$ of the currently observable continuous-time input process.

The goal is to approximate $x$ by a “smooth” filtered process $y$ via selection of an appropriate causal impulse response function $h$.

We are looking for families of the causal smoothing impulse response functions $h$ satisfying the following conditions.

(A) The outputs $y$ approximate inputs $x$; an arbitrarily close approximation can be achieved via selection of a filter from this family.

(B) The outputs $y$ have low energy on the higher frequencies.

(C) The effectiveness of the damping on the energy on the higher frequencies approximates the effectiveness of some reference family of non-causal smoothing filters with a reasonably fast decay of the frequency response.

Note that it is not a trivial task to satisfy Conditions (C)-(D). In theory, there are sets of ideal low-pass filters such that the distance of these sets from the set of all causal filters is zero. It was shown in [2] that this is true for the set of low-pass filters with increasing pass interval $[-\Delta, \Delta]$, where $\Delta < \pi$. However, it is unknown yet how to construct the corresponding sequence of causal filter.

The targeted properties of the subideal filters

Our purpose is to construct a family of causal filters such that the Conditions (A)–(C) are satisfied. We will be using a reference family of “ideal” smoothing filters with the frequency response

$$M_{\theta,q}(e^{i\omega}) = \exp \left( -\frac{\theta}{|1+e^{i\omega}|^q} \right), \quad \theta > 0, \quad q > 1. \quad (2)$$

For these filters, Condition (A) is satisfied for all $q > 1$ as $\theta \to 0$, and Conditions (B) is satisfied for all $\theta > 0, q > 1$. However, these filters are non-causal: for any $x(\cdot) \in \ell_2$, the values $x(t+1)$ of the output processes of these filters are weakly predictable at time $t$ [11]. This is because $M_{\theta,q}(-1) = 0$ and $M_{\theta,q}(e^{i\omega}) \to 0$ as $\omega \to -\pi$ fast enough.

Let an integer $m \geq 1$ be given. We will construct a family of causal filters with impulse responses $\{h_{\nu}\}_{\nu=1}^\infty \subset \ell^+_\infty$ and with the corresponding $Z$-transforms $H_{\nu}$ such that the following more special Conditions (a)-(c) are satisfied.

(a) Approximation of the identity operator:

- (a1) $\sup_{\omega \in [0,\pi], \alpha} |H_{\nu}(e^{i\omega})| < +\infty$. 
• (a2) For any $\Omega > 0$, $H_\nu(e^{i\omega}) \to 1$ as $\nu \to +\infty$ uniformly in $\omega \in [-\Omega, \Omega]$.

• (a3) For any $X(e^{i\omega}) \in L_1(-\pi, \pi)$ and any $x = Z^{-1}X$,

$$\|y_\nu(\cdot) - x(\cdot)\|_{\ell_{\infty}} \to 0 \quad \text{as} \quad \nu \to \infty,$$

where $y_\nu$ is the output process

$$y_\nu(t) = \sum_{t=-\infty}^{t} h_\nu(t - \tau)x(\tau).$$

(b) The spectrum is vanishing at a point at $T$:

• (b1) For all $\nu$, $H_\nu(e^{i\omega})$ is $m$ times differentiable at $\omega = \pi$, and

$$H_\nu(-1) = 0, \quad \left. \frac{d^k H(e^{i\omega})}{d\omega^k} \right|_{\omega=\pi} = 0, \quad k = 1, \ldots, m.$$

• (b2) For any $\varepsilon > 0$, there exists $\delta > 0$ and $\nu_0$ such that, for all $\nu \geq \nu_0$,

$$\sup_{\omega \in [\pi - \delta, \pi + \delta]} |H_\nu(e^{i\omega})| < \varepsilon.$$

(c) Approximation of non-causal filters with respect to the effectiveness in damping: For any $\varepsilon > 0$ and any $\Omega \in (0, \pi)$, $\Omega_0 \in (\Omega_0, \pi)$, $\Omega_1 \in (\Omega_1, \pi)$ there exists $\theta > 0, q > 1, \nu > 0$ such that

$$|H_\nu(e^{i\omega}) - 1| \leq \varepsilon, \quad \omega \in [-\Omega, \Omega],$$

$$|H_\nu(e^{i\omega})| \leq |M_{\theta,q}(e^{i\omega})|, \quad \omega \in [-\Omega_1, -\Omega_0] \cup [\Omega_0, \Omega_1]. \quad (3)$$

$$|H_\nu(e^{i\omega})| \leq |M_{\theta,q}(e^{i\omega})|, \quad \omega \in [-\Omega_1, -\Omega_0] \cup [\Omega_0, \Omega_1]. \quad (4)$$

Conditions (a)-(c) ensure that Conditions (A)-(C) are satisfied, in a certain sense. In particular, Condition (c) ensures that Condition (C) is satisfied, since the effectiveness of smoothing is defined by the rate of damping of the higher frequencies.

3 A family of sub-ideal smoothing filters

For real numbers $a \in (0, 1)$, $p \in (1/2, 1]$, and for integers $N \geq 1$ and $m \geq 1$, we consider transfer functions

$$H_\nu(z) = \left( \exp \left( \frac{(1-a)^p}{z + a} + G_\nu(z) \right) \right)^m, \quad z \in \mathbb{C}, \quad (5)$$

where

$$G_\nu(z) = -\xi(a, p) + \frac{\gamma(a, p)}{N} \left[ (-1)^N z^{-N} - 1 \right],$$
and where
\[ \xi(a, p) = \exp[-(1 - a)^{p-1}], \quad \gamma(a, p) = |1 - a|^{p-2} \xi(a, p). \]

Let us consider the set of all transfer functions (5) with some fixed integer \( m \geq 0 \) and some fixed real \( p \in (1/2, 1] \), with a sequence of rational numbers \( a \in (0, 1) \), and with a set of all integers \( N > 0 \). We assume that this countable set is counted as a sequence \( \{H_\nu\}_{\nu=1}^\infty \) such that \( a \to 1 - 0 \) and \( N \to +\infty \) as \( \nu \to +\infty \).

**Theorem 1** Conditions (a)-(c) are satisfied for the family of filters defined by the transfer functions \( \{H_\nu\}_{\nu=1}^\infty \). (Therefore, Conditions (A)-(C) are satisfied for this family).

**Proof of Theorem 1** Let us assume first that \( m = 1 \).

Clearly, the functions \( H_\nu \) are holomorphic in \( D^c \). Hence the inverse Z-transforms \( h_\nu = Z^{-1}H_\nu \) are causal impulse responses, i.e., \( h_\nu(t) = 0 \) for \( t < 0 \) (see, e.g., [13]).

Let \( f(a) = (1 - a)^p \) and \( \Psi_\nu(z) = f(a)(z + a)^{-1} \). By the definitions, \( H_\nu(z) = \exp \Psi_\nu(z) + G_\nu(z) \), and

\[ \Psi_\nu(e^{i\omega}) = f(a) \frac{\cos(\omega) + a - i \sin(\omega)}{\cos(\omega) + a + i \sin(\omega)^2}. \]

Let us prove that Condition (a) holds.

Clearly, \( |G_\nu(e^{i\omega})| \to 0 \) as \( \nu \to +\infty \) uniformly in \( \omega \in (-\pi, \pi] \).

Let \( \omega_a \in (\pi/2, \pi) \) be such that \( \cos(\omega_a) + a = 0 \). We have that \( \text{Re} \Psi_\nu(e^{i\omega}) > 0 \) for all \( \omega \in [-\omega_a, \omega_a] \) and \( \text{Re} \Psi_\nu(e^{i\omega}) < 0 \) for all \( \omega \in [-\pi, \omega_a) \cup (\omega_a, \pi] \).

Further, we have that
\[ \inf_{\omega \in [-\omega_a, \omega_a]} |e^{i\omega} + a| \geq \sqrt{1 - a^2}. \]

Hence
\[ \sup_{\omega \in [-\omega_a, \omega_a]} |\Psi_\nu(e^{i\omega})| \leq \frac{f(a)}{\sqrt{1-a^2}} = \frac{(1-a)^{p-1/2}}{(1+a)^{1/2}}. \]

Therefore, the value \( |H(e^{i\omega})| \) is uniformly bounded in \( a, \omega \), Hence Condition (a1) holds.

Further, we have that
\[ \omega_a \to \pi - 0 \quad \text{as} \quad a \to 1. \]

Hence, for any \( \Omega \in [0, \pi) \), we have that
\[ \sup_{\omega \in [-\Omega, \Omega]} |\Psi_\nu(e^{i\omega})| \to 0 \quad \text{as} \quad a \to 1. \]
We have that

\[ \sup_{\omega \in [-\Omega, \Omega]} |H(e^{i\omega}) - 1| \to 0 \quad \text{as} \quad a \to 1. \]

Hence Condition (a2) holds.

Let us show that Condition (a3) holds. Let \( x \in \ell_2, X(e^{i\omega}) = \mathbb{Z} x \in L_2(-\pi, \pi) \), and let \( Y_\nu = H_\nu X \). By Condition (a2), \( Y_\nu(e^{i\omega}) \to X(e^{i\omega}) \) as \( \nu \to +\infty \) for all \( \omega \in \mathbb{R} \). Clearly, there exists \( \nu_0 > 0 \) such that \( \sup_{\nu \geq \nu_0} |H_\nu(e^{i\omega})| \leq 2 \). Hence \( |Y_\nu(e^{i\omega}) - X(e^{i\omega})| \leq 3|X(e^{i\omega})| \).

We have that \( \|X(e^{i\omega})\|_{L_1(-\pi, \pi)} \leq (2\pi)^{1/2}\|X(e^{i\omega})\|_{L_2(-\pi, \pi)} = \|x\|_{\ell_2} < +\infty \). By Lebesgue Dominance Theorem, it follows that

\[ \|Y_\nu(e^{i\omega}) - X(e^{i\omega})\|_{L_1(-\pi, \pi)} \to 0 \quad \text{as} \quad \nu \to +\infty. \]

Therefore, Condition (a3) holds and Condition (a) holds.

Let us show that Condition (b) holds. We have that

\[ \exp(\Psi_\nu(e^{i\pi})) = \exp\left(f(a) - \frac{1 + a}{(1 - a)^2}\right) = \exp\left(-\frac{(1 - a)^p}{1 - a}\right) = -\xi(a, p) = -G_\nu(-1). \]

Hence \( H_\nu(-1) = \exp(\Psi_\nu(-1)) + G_\nu(-1) = 0 \). We have that

\[ G_\nu(e^{i\omega}) = -\xi(a, p) + \frac{\gamma(a, p)}{N}(e^{-iN(\omega - \pi)} - 1). \]

Clearly, the function \( \Psi_\nu(e^{i\omega}) \) is differentiable in \( \omega \in \mathbb{R} \) for any \( \nu \). In addition, we have that \( \Re \Psi(e^{i\omega}) = \Re \Psi_\nu(e^{-i\omega}) \). Hence

\[ \frac{d}{d\omega} \Re \exp(\Psi_\nu(e^{i\pi})) \bigg|_{\omega=\pi} = 0. \]

Let

\[ r(\omega) = \Im \exp(\Psi_\nu(e^{i\pi})), \quad q(\omega) = \Im \left(\frac{1}{N} (e^{-iN(\omega - \pi)} - 1)^{-1}\right). \]

We have that \( r(\pi) = q(\pi) = 0 \). By the definitions, \( \exp \Re \Psi_\nu(e^{i\omega}) \to \xi(a, p) \) as \( \omega \to \pi \). The L'Hôpital's rule gives have that

\[ \lim_{\omega \to \pi} \frac{dr(\omega)}{d\omega} = \lim_{\omega \to \pi} \frac{r(\omega)}{q(\omega)} = \lim_{\omega \to \pi} \frac{\xi(a, p) \sin\left(\frac{-(1-a)^p \sin(\omega)}{(a+\cos(\omega))^2+\sin(\omega)^2}\right)}{N \sin(N(\omega - \pi))} \]

\[ = -\xi(a, p) \left(\frac{(1 - a)^p}{(a - 1)^2}\right) = -\gamma(a, p). \]

Clearly, \( \frac{dq(\omega)}{d\omega} \bigg|_{\omega=\pi} = -1 \). Hence

\[ \frac{d}{d\omega} \exp(\Psi_\nu(e^{i\pi})) \bigg|_{\omega=\pi} = i\gamma(a, p). \]
Further, by the definitions, we have that
\[ G_\nu(e^{i\omega}) = -\xi(a, p) + \frac{\gamma(a, p)}{N} \left( e^{-iN(\omega - \pi)} - 1 \right) \]
and
\[ \frac{dG_\nu(e^{i\omega})}{d\omega} \bigg|_{\omega=\pi} = -i\gamma(a, p). \]

Hence
\[ \frac{d}{d\omega} H_\nu(e^{i\pi}) \bigg|_{\omega=\pi} = \frac{d}{d\omega} \exp \left( \Psi_\nu(e^{i\pi}) \right) \bigg|_{\omega=\pi} - \frac{dG_\nu(e^{i\omega})}{d\omega} \bigg|_{\omega=\pi} = 0. \]

Therefore, Condition (b1) holds.

Let us show that Condition (b2) holds. We have that
\[ \Re \Psi_\nu(e^{i\omega}) = \frac{f(a)(\cos(\omega) + a)}{(\cos(\omega) + a)^2 + \sin(\omega)^2}. \]

For all \( \omega \in (-\pi, \pi] \), we have
\[ f(a)^{-1}\Re \Psi_\nu(e^{i\omega}) \to \frac{\cos(\omega) + 1}{(\cos(\omega) + 1)^2 + \sin(\omega)^2} \quad \text{as} \quad a \to 1. \]

We have that \( \Re(e^{i\omega} - a)/|e^{i\omega} - a| \) is non-increasing in \( \omega \in [\omega_\nu, \pi] \) and \( 1/|e^{i\omega} - a| \) non-decreasing in \( \omega \in [\omega_\nu, \pi] \) as well, and that \( \Re \Psi_\nu(e^{i\omega}) \) is non-increasing in \( \omega \in [\omega_\nu, \pi] \); in addition, \( \Re \Psi_\nu(e^{i\omega}) = \Re \Psi_\nu(e^{-i\omega}) \). Hence we can select \( \tilde{\omega}_\nu \in [\omega_\nu, \pi] \) such that \( \Re \Psi_{\tilde{a}_\nu}(e^{i\omega}) < \Re \Psi_\nu(e^{i\pi})/2 \) for all \( \omega \in [-\tilde{\omega}_\nu, \tilde{\omega}_\nu] \). Let us select \( \tilde{\nu} \) such that \( \Re \Psi_{\tilde{\nu}}(e^{i\pi})/2 < \log(\varepsilon/2) \) for all \( \nu \geq \tilde{\nu} \). Let \( \delta = \pi - \tilde{\omega}_\nu \). Then Condition (b2) is satisfied for \( H_\nu \) replaced by \( \exp(\Psi_{\tilde{\nu}}(z)) \) with these \( \tilde{\nu} \) and \( \delta \). In addition, we can select \( \tilde{\nu} \geq \tilde{\nu} \) such that \( |G_\nu(e^{i\omega})| \leq \varepsilon/2 \) for all \( \nu \geq \tilde{\nu} \). Therefore, Condition (b) holds.

Let us show that Condition (c) holds. As was shown in the proof of (a) above, for a given \( \varepsilon > 0 \) and \( \Omega \), we can select \( \tilde{\nu} \) such that (3) holds for \( \nu \geq \tilde{\nu} \). Further, let us fix some \( q > 1 \). For any \( \Omega_0 > \Omega \) and \( \Omega_1 > \Omega_0 \),
\[ \sup_{\omega \in [-\Omega_1, -\Omega_0] \cup \Omega_0, \Omega_1]} |M_{\theta,q}(e^{i\omega}) - 1| \to 0 \quad \text{as} \quad \theta \to 0. \]

Clearly, (4) holds for small enough \( \theta \). Hence Condition (c) holds.

We have proved the theorem for the case where \( m = 1 \). The extension on the case where \( m > 1 \) is straightforward. This completes the proof of Theorem 1. \( \square \)

**Illustrative examples**

Figures 1-3 illustrates the behaviors of the frequency responses and the impulse functions for the suggested filters and for the reference family of the filters.
Figure 1 shows the shapes of gain curves $|M_{\theta,q}(e^{i\omega})|$ for reference non-causal filter (2) with $\theta = 0.2$, $q = 1.01$, and $|H_\nu(e^{i\omega})|$ for sub-ideal causal filters (5) with $a = 0.99$, $p = 0.6$, $N = 50$, $m = 2$.

Figure 2 shows the shapes of error curves for approximation of identity operator on low frequencies. More precisely, it shows $|M_{\theta,q}(e^{i\omega}) - 1|$ for reference non-causal filter (2) and $|H_\nu(e^{i\omega}) - 1|$ for sub-ideal causal filters (5), with the same parameters as for Figure 1.

Figure 3 shows an example of impulse response $h = Z^{-1}H_\nu$ calculated as the inverse Z-transform for causal filter (5) with $a = 0.8$, $p = 0.6$, $N = 10$, $m = 1$. Figure 4 shows an example of impulse response $h = Z^{-1}H_\nu$ calculated as the inverse Z-transform for causal filter (5) with $a = 0.99$, $p = 0.6$, $N = 10$, $m = 2$. Since the properties of $H_\nu$ guarantee that $\text{Im} h_\nu(t) = 0$ for all $t$ and that $h_\nu(t) = 0$ for all $t < 0$, we show the values for $t \geq 0$ only.

It is interesting that the impulse response functions are not non-negative functions; we have that $h_\nu(t) > 0$ for some $t > 0$. Therefore, the suggested filters do not represent an averaging with a positive kernels.

Figure 1: Gain decay: the values $|M_{\theta,q}(e^{i\omega})|$ for a non-causal filter (2) with $\theta = 0.2$ and $q = 1.01$, and $|H_\nu(e^{i\omega})|$ for a causal filter (5) with $a = 0.99$, $p = 0.6$, $N = 50$, $m = 2$.

4 Applications to the forecasting

The possible application of suggested filters is preliminary smoothing of the input signals for the predicting algorithms. In this setting, the causality is crucial. It is known that the band-limited sequences are predicable, i.e., the sequences are predicable if with the spectrum vanishing on a interval in $T$. In addition, there are predicable sequences such that the spectrum is vanishing in a single point of $T$. These two classes of sequences require different predicting algorithms; some
Figure 2: Approximation of identity operator: shapes of the distances from 1, i.e., for the values $|M_{\theta,q}(e^{i\omega}) - 1|$ and $|H_\nu(e^{i\omega}) - 1|$, for a non-causal filter (2) with $\theta = 0.2$ and $q = 1.01$, and $|H_\nu(e^{i\omega})|$ for a causal filter (3) with $\alpha = 0.99$, $p = 0.6$, $N = 50$, $m = 2$.

algorithms were suggested in [11, 12].

It can be noted that, for $\alpha$ close to 1, the predictability of real sequences the suggested filters do not change the input sequences significantly; the energy of the input is not damped on some given interval on $\mathbb{T}$. In fact, the energy is damped on a small neighborhood of the point $e^{i\omega} = 1$, and the size of this neighborhood converges to zero as $\alpha \to 1$. Therefore, the filters introduced here will not provide sufficient smoothing for the predicting algorithms that require that there is a fixed arc on the unit circle where the spectrum is vanishing; an example of such algorithm can be found in [12].

However, the application of these filters improves the effectiveness of the predictors oriented on processes $x$ with spectrum vanishing in a single point such as predictors introduced in [11]; these predictor require that $X(e^{i\omega}) = 0$ at $\omega = \pi$ and that the value $|X(e^{i\omega})|$ is vanishing fast enough as $\omega \to \pi$, where $X = \mathcal{Z}_x$. More precisely, it was required that

$$\inf_{c_1 > 0, c_2 > 0} \sup_{\omega \in [-\pi, \pi]} \log |X(e^{i\omega})| - c_1 + \frac{c_2}{1 + e^{i\omega}q} \leq 0. \quad (6)$$

It can be noted that non-causal filters (2) transfer sequences of a general type into sequences such that (6) holds. For an input process $x(t)$, predictor [11] produces the process

$$y(t) = \sum_{d = -\infty}^{t} k(t - d)x(t), \quad k = \mathcal{Z}_K.$$
The predicting kernel was defined \( k = k(\cdot, \gamma) = Z^{-1} K \), where

\[
K(z) \triangleq z \left( 1 - \exp \left[ -\frac{\gamma}{z + 1 - \gamma^2 \mu/(1-q)} \right] \right).
\] (7)

Here \( \mu > 1 \) is a given parameter, \( \gamma > 0 \) is a parameter. This function approximates the function \( e^{i\omega} \) representing forward one-step shift; the value \( |K(e^{i\omega}) - e^{i\omega}| \) is small everywhere but in a small neighborhood of \( \omega = \pi \), where the value of \( |K(e^{i\omega})| \) can be quite large.

The process \( y(t) \) represents an one step prediction \( \hat{x}(t) \) of \( x(t) \), i.e., \( \hat{x}(t) = y(t - 1) \). It was shown in \([11]\)

\[
\sup_t |x(t + 1) - y(t)| \to 0 \quad \text{as} \quad \gamma \to 0,
\]

for real sequences \( x \) such that (6) holds, i.e., that the prediction error vanishes as \( \gamma \to +\infty \). Moreover, the error vanishes uniformly over classes of processes \( x \) from some bounded sets from \( \ell_{\infty} \), such that (6) for bounded sets of \( c_i \).

In theory, predictors (7) are robust with respect to some small noise contamination, meaning that the prediction error depends continuously from the intensity of the contaminating noise. However, for large \( \gamma \), the values of \( K(e^{i\omega}) \) can be very large in a neighborhood of \( \omega = \pi \); in this case, the error can be the error large even if the noise on these frequencies is small.

It could be interesting if we can apply filter (5) to compensate the presence of large values of \( K(e^{i\omega}) \) in a small neighborhood of \( \omega = \pi \) and therefore to reduce the impact of the presence of the noise.

Our setting does not involve stochastic processes and probability measure; it is oriented on smoothing the real sequences. However, to provide an example of the application of our smoothing filters, we considered a toy example with prediction of a stochastic Gaussian stationary process \( x(t) \) evolving as an autoregression of AR(2) type

\[
x(t) = \beta_1 x(t-1) + \beta_2 x(t-2) + \sigma \eta(t), \quad t \in \mathbb{Z}.
\] (8)

Here \( \eta(t) \) is a stochastic discrete time Gaussian white noise, \( \mathbb{E} \eta(t) = 0 \), \( \mathbb{E} \eta(t)^2 = 1 \). The coefficient \( \sigma \) describes the intensity of the noise inputs.

For the estimation of the effectiveness of predictors, we will used the ratio

\[
e(b_1, b_2) = \frac{(\sum_{t=1}^{n} |y(t-1) - x(t)|^2)^{1/2}}{(\sum_{t=1}^{n} (b_1 x(t-1) + b_2 x(t-2) - x(t))^2)^{1/2}}.
\] (9)

The values \( b_k \in \mathbb{R} \) here are parameters. The values \( y(t) \) are supposed to be the predictions of \( x(t + 1) \). Ratio (9) allows to compare the error of a predicting algorithm generating \( y(t) \) and the error generated by a linear predictor. More precisely, the value \( e(b_1, b_2) \) represent the ratio of the error generated by the predictor producing \( y \) and the error generated with the error of the linear predictor based on the hypothesis that \( \beta_1 = b_1 \) and \( \beta_2 = b_2 \).
If the vector \((\beta_1, \beta_2)\) is known, then the optimal one step predictor of \(x(t + 1)\) is

\[
y(t) = \beta_1 x(t) + \beta_2 x(t - 1),
\]

(10)
The value \(\sum_{t=1}^{n} |\beta_1 x(t - 1) + \beta_2 x(t - 2) - x(t)|^2\) represents the average squared error of this optimal predictor, with a known values of \((\beta_1, \beta_2)\). Therefore, \(e(\beta_1, \beta_2) = 1\) for optimal predictor (10). It follows that a value \(e(\beta_1, \beta_2) < 1\) is not feasible for predictor (7), for a large sample. This can be explained as the following: since we the sequence \(x(t)\) does not satisfy (6) due to the presence of noise, a forecasting error is inevitable. Since predictor (10) is optimal, the value \(e(\beta_1, \beta_2)\) for (7) cannot be less than the the value \(e(\beta_1, \beta_2) = 1\) for \(y(t)\) defined by (10).

However, in many practical situations, the value of \((\beta_1, \beta_2)\) is unknown, and hence predictor (10) cannot be used. Predictor (7) does not require to know \((\beta_1, \beta_2)\) and can be applied in models with unknown or random and time variable \((\beta_1, \beta_2)\) where predictors (10) is not applicable. In other words, predictor (7) can be as well applied for processes with unknown shape of the spectral representation. Therefore, it is reasonable to estimate the performance of a predictor using \(e(\beta_1, \beta_2)\) for (7) cannot be less than the the value \(e(\beta_1, \beta_2) = 1\) for \(y(t)\) defined by (10).

We will use value (9) to estimate the performance of predicting algorithms for the following two cases:

- The case without filtering, where \(y = Z^{-1}(KX)\); we denote by \(e_K(b_1, b_2)\) the corresponding values (9);
- The case with filtering, where \(y = Z^{-1}(KH_{\nu}X)\), with \(K\) defined by (7) and \(H_{\nu}\) defined by (5); we denote by \(e_{KH}(b_1, b_2)\) the corresponding values (9).

In both cases, \(y(t)\) can be calculated using currently available historical data \(\{x(s)\}_{s \leq t-1}\).

In our experiments, we used equations (5) and (7) with

\[
\gamma = 1.02, \quad q = 1.2, \quad \mu = 1.01, \quad a = 0.5, \quad p = 0.7, \quad N = 100, \quad m = 3.
\]

(11)

Note that selection of too large \(\gamma\) makes calculation of \(h\) challenging, since it involves precise integration of fast growing \(K(\epsilon^{i\omega})\). The choice of parameters in (11) ensures that the values of \(|K(\epsilon^{i\omega})|\) are not large.

Figure 5 shows the shapes of error curves for approximation of the forward one step shift operator. More precisely, it shows \(|K(\epsilon^{i\omega}) - e^{i\omega}|\) for the transfer function of the predictor (7) and \(|K(\epsilon^{i\omega}) H_{\nu}(\epsilon^{i\omega}) - e^{i\omega}|\) for the product of the transfer functions (5) and (7), which corresponds to preliminary smoothing of the input process by filters (5). It can be seen that application of the filter improves the approximation of the one-step forward shift operator.
Figure 6 shows the corresponding impulse responses $Z^{-1}K$ and $Z^{-1}(KH)\nu$.

In our experiment with AR(2) process, we used 10,000 Monte-Carlo trials with $n = 100$ and $\sigma_1 = 0.3$. For each trial, we selected $(\beta_1, \beta_2)$ randomly and independently. The distribution of $(\beta_1, \beta_2)$ at each trial was the following: $\beta_1$ has the uniform distribution on the interval $(0, 1)$, and $\beta_2 = \xi \sqrt{1 - \beta_1^2}$, where $\xi$ is a random variable independent on $\beta_1$ uniformly distributed on the interval $(-1, 1)$. This choice ensures that the eigenvalues of the autoregression stays inside of the unit circle $D$ almost surely. In these experiments, we obtained that the mean value over all trials of $e_K(\beta_1, \beta_2)$ is 1.6058 and the mean value of $e_{KH}(\beta_1, \beta_2)$ is 1.1730. This indicates that the use of the filter improves the performance of the predictor. In this setting, we compare the performance of the predictor with the performance of the optimal predictor given that the values $\beta_1$ and $\beta_2$ are known. As was mentioned above, the mean values of $e_K(\beta_1, \beta_2)$ and $e_{KH}(\beta_1, \beta_2)$ cannot be less than one, so the performance of the predictor (7) combined with filters (5) is quite good.

In addition, we calculated the values $e_K(b_1, b_2)$ and $e_{KH}(b_1, b_2)$ with $b_k = E\beta_k$, $k = 1, 2$, where $E\beta_k$ is the population mean for $\beta_k$, i.e., $E\beta_1 = 0.5$, $E\beta_2 = 0$. In this setting, we compare the performance of the predictor with the performance of the predictor where the population means of $\beta_1$ and $\beta_2$ are known and the one step predictor of $x(t+1)$ is

$$y(t) = (E\beta_1)x(t) + (E\beta_2)x(t - 1).$$

In this experiment, we obtained $e_K(0.5, 0) = 1.2390$ and $e_{KH}(0.5, 0) = 0.9543$.

Figure 7 shows a sample path of AR(2) process $x(t)$ and a filtered process obtained using filter (5) with the parameters defined by (11). Figure 8 shows sample paths of AR(2) process $x(t)$ and outputs $y(t)$ of predictor (11) without preliminary filtering and with preliminary filtering using filter (5) with the parameters defined by (11). It shows the values $y(t)$ being the predictions of $\hat{x}(t+1)$ versus the values of $x(t+1)$.

In addition, we considered a modification of process (8) with $\beta_2 = 0$, i.e., AR(1) process. We set 10,000 Monte-Carlo trials with $n = 100$, with $\sigma_1 = 0.3$, and with random selected $\beta_1$ such that $\beta_1$ was distributed uniformly on the interval $(0, 1)$, and with $\beta_2 = 0$. In these experiments, we obtained that the mean value of $e_K(\beta_1, 0)$ is 1.1075 and the mean value of $e_{KH}(\beta_1, \beta_2)$ is 1.2805. This indicates again that the use of the filter improves the performance of the predictor. In this setting, we compare the performance of the predictor with the performance of the optimal predictor given that the value $\beta_1$ is known. Again, the mean values of $e_K(\beta_1, 0)$ and $e_{KH}(\beta_1, 0)$ cannot be less than one, so the performance of the predictor (7) combined with filters (5) is quite good. In addition, we calculated the values $e_K(b_1, 0)$ and $e_{KH}(b_1, 0)$ with $b_1 = E\beta_1 = 0.5$, where again $E\beta_1$ is the population mean for $\beta_1$. In this setting, we compare the performance of the predictor with the performance of the predictor where the population means of $\beta_1$ is known and the one step predictor of $x(t+1)$ is $y(t) = (E\beta_1)x(t)$. We obtained $e_K(0.5, 0) = 1.1792$ and $e_{KH}(0.5, 0) = 1.02$. 

13
5 Conclusion

The paper proposes a family of causal smoothing filters with a high rate of damping of the energy. In addition, the frequency response that can be selected to be arbitrarily close to the real unity uniformly on an arbitrarily large part of the unit circle. These filters are sub-ideal meaning that a higher rate of damping would lead to the loss of causality; this is because they approximate non-causal filters. A possible application is preliminary smoothing of the inputs for predicting algorithms. The transfer functions obtained are not rational functions; it would be interesting to consider their approximation by the rational functions. We leave it for the future research.

Acknowledgment

This work was supported by ARC grant of Australia DP120100928 to the author.

References

[1] Aldroubi, A., and Unser, M. A general sampling theory for nonideal acquisition devices. (1994). *IEEE Trans. Signal Process.*, Vol.42, No. 11 2915–2995.

[2] J.M. Almira and A.E. Romero. How distant is the ideal filter of being a causal one? *Atlantic Electronic Journal of Mathematics* 3 (1) (2008) 46–55.

[3] Bingham, N. H. (2012). Szegö’s theorem and its probabilistic descendants. *Probability Surveys* 9, 287-324.

[4] F.G. Beutler. Error-free recovery of signals from irregularly spaced samples, *SIAM Review* 8 (3) (1966) 328–335.

[5] J.R. Brown Jr. Bounds for truncation error in sampling expansion of band-limited signals, *IEEE Transactions on Information Theory* 15 (4) (1969) 440-444.

[6] Church, A. (1940). On the Concept of Random Sequence. *Bull.Amer.Math.Soc.* 46, 254–260.

[7] Cambanis, S., and Soltani, A.R. (1984). Prediction of stable processes: spectral and moving average representations. *Z. Wahrsch. Verw. Gebiete* 66, no. 4, 593–612.

[8] N. Dokuchaev. The predictability of band-limited, high-frequency, and mixed processes in the presence of ideal low-pass filters, *Journal of Physics A: Mathematical and Theoretical* 41 (2008) 382002 (7pp).

[9] N. Dokuchaev. Predictability on finite horizon for processes with exponential decrease of energy on higher frequencies, *Signal processing* 90 (2) (2010) 696–701.
[10] Dokuchaev, N. (2012). On sub-ideal causal smoothing filters. *Signal Processing* **92**, iss. 1, 219-223.

[11] N. Dokuchaev, ”Predictors for discrete time processes with energy decay on higher frequencies”, *IEEE Transactions on Signal Processing* **60**, No. 11, 2012, 6027-6030.

[12] Dokuchaev, N. (2012). On predictors for band-limited and high-frequency time series. *Signal Processing* **92**, iss. 10, 2571-2575.

[13] P. Duren. *Theory of $H^p$-Spaces*. Academic Press, New York, 1970.

[14] Ferreira P. G. S. G. (1994). Interpolation and the discrete Papoulis-Gerchberg algorithm. *IEEE Transactions on Signal Processing*, **42** (10), 2596–2606.

[15] Ferreira P. G. S. G. (1995a). Nonuniform sampling of nonbandlimited signals. *IEEE Signal Processing Letters* **2**, Iss. 5, 89–91.

[16] Ferreira P. G. S. G. (1995b). Approximating non-band-limited functions by nonuniform sampling series. In: *SampTA’95, 1995 Workshop on Sampling Theory and Applications*, 276–281.

[17] Ferreira P. J. S. G., Kempf A., and Reis M. J. C. S. (2007). Construction of Aharonov-Berry's superoscillations. *J. Phys. A, Math. Gen.*, vol. 40, pp. 5141–5147.

[18] Frank, M., and Klotz, L. (2002). A duality method in prediction theory of multivariate stationary sequences. *Math. Nachr*. 244, 64–77.

[19] J.R. Higgins. *Sampling Theory in Fourier and Signal Analysis*. Oxford University Press, NY, 1996.

[20] Kolmogorov, A.N. (1941). Interpolation and extrapolation of stationary stochastic series. *Izv. Akad. Nauk SSSR Ser. Mat.,* 5:1, 3–14.

[21] Kolmogorov, A.N. (1965). Three approaches to the quantitative definition of information. *Problems of Information and Transmission*, 1:1, 1–7.

[22] Li, M., Vitanyi, P.B.M. (1993). *An introduction to Kolmogorov complexity and its applications*, Springer.

[23] Loveland, D.M. (1966). A new interpretation of von Mises’ concept of random sequence. *Z. Math. Logik Grundlagen Math* **12**, 279–294.

[24] von Mises, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.* 5, 52–99.
[25] Miamee, A.G., and Pourahmadi, M. (1988). Best approximations in $L^p(d\mu)$ and prediction problems of Szegö, Kolmogorov, Yaglom, and Nakazi. *J. London Math. Soc.* 38, no. 1, 133–145.

[26] Simon, B. (2011). Szegö’s Theorem and Its Descendants. Spectral Theory for $L^2$ Perturbations of Orthogonal Polynomials. *M.B. Porter Lectures*. Princeton University Press, Princeton.

[27] Szegö, G. (1920). Beiträge zur Theorie der Toeplitzschen Formen. *Math. Z.* 6, 167–202.

[28] Szegö, G. (1921). Beiträge zur Theorie der Toeplitzschen Formen, II. *Math. Z.* 9, 167-190.

[29] Verblunsky, S. (1936). On positive harmonic functions (second paper). *Proc. London Math. Soc.* 40, 290–320.
Figure 3: Impulse response $h_a(t) = (\mathcal{F}^{-1} H_a)(t)$ for causal filter (5) with $a = 0.8$, $p = 0.6$, $N = 10$.

Figure 4: Impulse response $h_a(t) = (\mathcal{F}^{-1} H_a)(t)$ for causal filter (5) with $a = 0.99$, $p = 0.6$, $N = 10$. 
Figure 5: Approximation of the one-step forward shift operator: the values $|K(e^{i\omega}) - e^{i\omega}|$ for the transfer function of the predictor (7) and $|K(e^{i\omega})H_{\nu}(e^{i\omega}) - e^{i\omega}|$, i.e., with smoothing of the input process by filters (5) with the parameters given in (11).

Figure 6: The impulse response $Z^{-1}K$ of predictor (7) and the impulse response $Z^{-1}(KH_{\nu})$ of the predictor combined with filters (5) with the parameters given in (11).
Figure 7: A path of AR(2) process $x(t)$ versus the output of filter (5) with the parameters given in (11).

Figure 8: A path of AR(2) process $x(t + 1)$ versus two predictions $y(t)$ of $x(t + 1)$: one was calculated without filtering, and another was calculated after application of filter (5) with the parameters given in (11).