Sub-ideal causal smoothing filters for the real sequences

Nikolai Dokuchaev

Department of Mathematics & Statistics, Curtin University,

GPO Box U1987, Perth, 6845 Western Australia

Abstract

The paper considers causal smoothing of the real sequences, i.e., discrete time processes in a deterministic setting. A family of causal linear time-invariant filters is suggested. These filters approximate the gain decay for some non-causal ideal smoothing filters with transfer functions vanishing at a point of the unit circle and such that they transfer processes into predictable ones. In this sense, the suggested filters are near-ideal; a faster gain decay would lead to the loss of causality. Applications to predicting algorithms are discussed and illustrated by experiments with forecasting of autoregressions with the coefficients that are deemed to be untraceable.

Key words: smoothing filters, casual filters, predicting, near-ideal filters, LTI filters.

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

1 Introduction

The paper studies causal smoothing of the discrete time processes. For many applications, it is preferable to replace a process by a more smooth process. In continuous time setting, smoothness is associated with predictability. Smooth analytic functions are predictable, i.e., their values on any interval define uniquely their values outside of this interval, and an ideal low-pass filter converts a function into an analytic one. For discrete time processes, it is not obvious how to define an analog of the continuous time analyticity and smoothness. A classical approach is to consider predictability instead of analyticity. So far, the predictability criterion for stochastic Gaussian stationary discrete time processes in the frequency domain setting are given by the classical Szegö-Kolmogorov Theorem. This theorem says that the optimal

1

prediction error is zero if

$$\int_{-\pi}^{\pi} \log \phi \left(e^{i\omega} \right) d\omega = -\infty, \tag{1}$$

where ϕ is the spectral density; see Kolmogorov [20], Szegö [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 $\{z \in \mathbf{C} : |z| = 1\}$, 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, 3, 8, 9, 12, 11, 13, 14, 15, 16, 18, 20, 25, 24]. A transition to the predicability of discrete time processes in deterministic setting is non-trivial and 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 important concept, starting from Mises [24], Church [6], Kolmogorov [20], Loveland [23]; see the references in [21].

It was found that real sequences are predictable if their Z-transform vanishes on an arc of the unit circle [11] on the complex plane or at a point $z = -1$ of the unit circle [12]. Therefore, smoothing can be interpreted as reduction of the energy on the higher frequencies. In particular, an ideal low-pass filter is a smoothing filter. This filter is non-causal, i.e., it requires the future value of the process. Similarly, 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 causality is inconsistent with predictability of outputs described in [12].

The present paper readdresses the problem of causal smoothing of the discrete time processes in the deterministic pathwise setting, without probabilistic assumptions. We suggest a family of causal smoothing filters that can be arbitrarily close to some ideal non-causal smoothing filters defined by equation (2) below. The suggested filters are near-ideal in the sense that they ensure "almost" ideal rate of damping the energy at the point $z = -1$; a faster decay of the frequency response is impossible for causal filters. This follows from predictability criterion [11]. In fact, the particular reference family of non-causal ideal filters (2) was selected because these filters transfer non-predictable processes into predictable ones satisfying the criterion from [12]. Similar approach was used in [10] for the continuous time setting.

The suggested near-ideal filters are discrete time causal linear time-invariant filters (LTI filters); they

2

are represented as convolution integrals over the historical data and approximate the real unity uniformly on an arbitrarily large part of the unit circle.

It appears that these causal filters can be used to improve the performance of predictors suggested in [12]. These are robust with respect to the noise contamination, and the error caused by a high-frequency noise depends on the intensity of this noise. A near-ideal smoothing filter cannot remove the high-frequency noise entirely but still can reduce it. This approach is discussed in Section 4, where we present some numerical experiments with the suggested near-ideal filters applied to forecasting of autoregressions in a setting where the autoregression coefficients are deemed to be untraceable.

Some definitions and notations

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

For $r \in [1, +\infty]$, we denote by ℓ_r the set of all sequences $x = \{x(t)\}_{t \in \mathbb{Z}} \subset \mathbf{R}$, such that $\|x\|_{\ell_r} = \left(\sum_{t=-\infty}^{\infty} |x(t)|^r \right)^{1/r} < +\infty$ for $r \in [1, \infty)$ or $\|x\|_{\ell_\infty} = \sup_t |x(t)| < +\infty$ for $r = +\infty$. We denote by ℓ_r^+ the set of all sequences $x \in \ell_r$ such that $x(t) = 0$ for $t < 0$.

We denote by $L_r(-\pi, \pi)$ the usual Banach space of complex valued L_r-integrable functions $x : [-\pi, \pi] \to \mathbf{C}$.

Let $D^c \triangleq \{z \in \mathbf{C} : |z| > 1\}$, and let $\mathbb{T} = \{z \in \mathbf{C} : |z| = 1\}$.

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

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

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

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

If $x \in \ell_2$, then $X|_{\mathbb{T}}$ is defined as an element of $L_2(\mathbb{T})$, i.e., $X\left(e^{i\omega}\right) \in L_2(-\pi, \pi)$. If $X\left(e^{i\omega}\right) \in L_1(-\pi, \pi)$, then $x = \mathcal{Z}^{-1}X$ is defined as an element of ℓ_∞.

Let $\mathrm{H}^2(D^c)$ be the Hardy space of functions that are holomorphic on D^c including the point at infinity with finite norm $\|h\|_{\mathrm{H}^2(D^c)} = \sup_{\rho > 1} \|h(\rho e^{i\omega}))\|_{L_2(-\pi, \pi)}$. Note that Z-transform defines a bijection between the sequences from ℓ_2^+ and the restrictions (i.e., traces) $X|_{\mathbb{T}}$ of the functions from $H^2(D^c)$ such that $\overline{X\left(e^{i\omega}\right)} = X\left(e^{-i\omega}\right)$ for $\omega \in \mathbf{R}$; see, e.g., [22], Section 4.3. If $X\left(e^{i\omega}\right) \in L_1(-\pi, \pi)$ and $\overline{X\left(e^{i\omega}\right)} = X\left(e^{-i\omega}\right)$, then $x = \mathcal{Z}^{-1}X$ is defined as an element of ℓ_∞^+.

2 Problem setting

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

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

where $h : \mathbb{Z} \to \mathbf{R}$ 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 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 spectrum of the output y vanishes on 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 that transfer processes into predictable ones, i.e., such that the future values are uniquely defined by the past values.

Note that it is not a trivial task to satisfy Conditions (A)-(C) simultaneously. For example, there are sets of ideal low-pass filters such that the distance of these sets from the set of all causal filters is zero. In [2], this was shown for the set of low-pass filters with increasing pass interval $[-\Delta, \Delta]$, where $\Delta \in (0, \pi)$. However, Condition (B) is not satisfied for the corresponding causal approximations. Moreover, all known causal approximations of ideal filters do not feature zero values for the transfer functions. The present paper suggests transfer functions vanishing at $z = -1$ together with their derivatives.

The targeted properties of the near ideal 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

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

where $q > 1$ and $\theta > 0$ are parameters. For these filters, Condition (A) is satisfied as $\theta \to 0$, and Conditions (B) is satisfied for all $\theta > 0$. However, these filters are non-causal because, for any $x \in l_2$, the values $x(t+1)$ of the output processes of these filters are weakly predictable at time t [12]. This is since $M_{\theta,q}\left(e^{i\omega}\right) \to 0$ fast enough as $\omega \to -\pi$.

For a given integer $m \geq 1$, we will construct a family of causal filters with impulse responses $h_a \in \ell_\infty^+$ and with the corresponding Z-transforms $H_a = \mathcal{Z}h_a$, where $a \in (0,1)$ is a parameter. For this family, the following more special Conditions (a)-(c) will be satisfied (the number m is used in Condition (b1) below).

(a) *Approximation of the identity operator:*

- (a1) $\sup_{\omega\in[0,\pi],a}|H_a\left(e^{i\omega}\right)| < +\infty$.

- (a2) For any $\Omega > 0$, $H_a\left(e^{i\omega}\right) \to 1$ as $a \to 1 - 0$ uniformly in $\omega \in [-\Omega, \Omega]$.

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

$$\|y_a(\cdot) - x(\cdot)\|_{\ell_\infty} \to 0 \quad \text{as} \quad a \to 1 - 0,$$

where y_a is the output process

$$y_a(t) = \sum_{\tau=-\infty}^{t} h_a(t-\tau)x(\tau).$$

(b) *The spectrum is vanishing at a point at $\mathbb{T}$:*

- (b1) For all a, $H_a\left(e^{i\omega}\right)$ is m times differentiable at $\omega = \pi$, and

$$H_a(-1) = 0, \qquad \frac{d^k H_a}{d\omega^k}\left(e^{i\omega}\right)\Big|_{\omega=\pi} = 0, \quad k = 1,...,m.$$

5

- (b2) For any $\varepsilon > 0$, there exists $\delta > 0$ and $a \in (0,1)$ such that

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

(c) *Approximation of non-causal filters (2) with respect to the effectiveness in damping:* For any $\varepsilon > 0$ and any $\Omega \in (0, \pi)$, $\Omega_0 \in (\Omega, \pi)$, $\Omega_1 \in (\Omega_0, \pi)$ there exists $\theta > 0$, $q > 1$, $a > 0$ such that

$$|H_a\left(e^{i\omega}\right) - 1| \leq \varepsilon, \quad \omega \in [-\Omega, \Omega], \tag{3}$$

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

Conditions (a)-(c) represent particular versions of less specific Conditions (A)-(C). In particular, estimate (4) ensures that Condition (C) is satisfied.

3   A family of near-ideal smoothing filters

Let a real number $p \in (1/2, 1)$ and integers $N \geq 1$ and $m \geq 1$ be given. For the real numbers $a \in (0,1)$, we define transfer functions

$$H_a(z) = \left(\exp \frac{(1-a)^p}{z+a} + G_a(z) \right)^m, \quad z \in \mathbf{C}, \tag{5}$$

where

$$G_a(z) = -\xi(a,p) + \frac{\gamma(a,p)}{N} \left((-1)^N z^{-N} - 1 \right),$$

and where

$$\xi(a,p) = \exp[-(1-a)^{p-1}], \quad \gamma(a,p) = |1-a|^{p-2}\xi(a,p).$$

We consider the set $\{H_a\}_{a \in (0,1)}$ of transfer functions (5) with a fixed triplet (m, N, p).

Theorem 1 *Conditions (a)-(c) are satisfied for the family of filters defined by the transfer functions* $\{H_a\}_{a \in (0,1)}$. *(Therefore, Conditions (A)-(C) are satisfied for this family).*

Proof of Theorem 1. Let us assume first that $m = 1$.

Clearly, the functions H_a are holomorphic in D^c and bounded in $D^c \cup \mathbb{T}$ for any $a \in (0,1)$. Hence the inverse Z-transforms $h_a = \mathcal{Z}^{-1}H_a$ are causal impulse responses, i.e., $h_a(t) = 0$ for $t < 0$; see, e.g., [22], Theorem 4.3.2.

Let $f(a) = (1-a)^p$ and $\Psi_a(z) = f(a)(z+a)^{-1}$. By the definitions, $H_a(z) = \exp \Psi_a(z) + G_a(z)$, and

$$\Psi_a\left(e^{i\omega}\right) = f(a)\frac{\cos(\omega) + a - i\sin(\omega)}{(\cos(\omega) + a)^2 + \sin(\omega)^2}.$$

Let us prove that Condition (a) holds.

Clearly, $\left|G_a\left(e^{i\omega}\right)\right| \to 0$ as $a \to 1-0$ uniformly in $\omega \in (-\pi, \pi]$.

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

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

Therefore, the value $|H_a\left(e^{i\omega}\right)|$ is uniformly bounded in a, ω. 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

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

and

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

Hence Condition (a2) holds.

Let as show that Condition (a3) holds. Let $Y_a = H_a X$. By Condition (a2), $Y_a\left(e^{i\omega}\right) \to X\left(e^{i\omega}\right)$ as $a \to 1-0$ for all $\omega \in \mathbf{R}$. Clearly, there exists $a_0 \in (0,1)$ and $c_0 > 0$ such that $\sup_{\omega, a \geq a_0} |H_a\left(e^{i\omega}\right)| \leq c_0$. Hence $|Y_a\left(e^{i\omega}\right) - X\left(e^{i\omega}\right)| \leq (c_0 + 1)|X\left(e^{i\omega}\right)|$. By the assumptions, $X\left(e^{i\omega}\right) = \mathcal{Z}x \in L_1(-\pi, \pi)$. By the Lebesgue Dominance Theorem, it follows that

$$\left\|Y_a\left(e^{i\omega}\right) - X\left(e^{i\omega}\right)\right\|_{L_1(-\pi, \pi)} \to 0 \quad \text{as} \quad a \to 1 - 0.$$

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

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

$$\exp\left(\Psi_a\left(e^{i\pi}\right)\right) = \exp\left(f(a)\frac{-1+a}{(1-a)^2}\right) = \exp\left(-\frac{(1-a)^p}{1-a}\right)$$
$$= \xi(a,p) = -G_a(-1).$$

Hence $H_a(-1) = \exp\left(\Psi_a\left(-1\right)\right) + G_a(-1) = 0$.

Let us show that $\frac{dH_a}{d\omega}\left(e^{i\omega}\right)\Big|_{\omega=\pi} = 0$. Let

$$r(\omega) = \mathrm{Re}\,\exp\left(\Psi_a\left(e^{i\omega}\right)\right), \quad s(\omega) = \mathrm{Im}\,\exp\left(\Psi_a\left(e^{i\omega}\right)\right),$$
$$q(\omega) = \frac{1}{N}\mathrm{Im}\left(e^{-iN(\omega-\pi)} - 1\right).$$

Clearly, the function $\Psi_a\left(e^{i\omega}\right)$ is differentiable in $\omega \in \mathbf{R}$ for any a, as well as functions $r(\omega)$, $s(\omega)$, and $q(\omega)$. In addition, we have that $r(\omega) = r(\pi - \omega)$. Hence $r(\omega)$ is even about the point $\omega = \pi$ and differentiable. This implies that

$$\frac{dr}{d\omega}(\omega)\Big|_{\omega=\pi} = 0.$$

By the definitions, $s(\omega) = \exp(\mathrm{Re}\,\Psi_a)\sin(\mathrm{Im}\,\mathrm{Re}\,\Psi_a)$ and $\exp(\mathrm{Re}\,\Psi_a)\left(e^{i\omega}\right) \to \xi(a,p)$ as $\omega \to \pi$. We have that $s(\pi) = q(\pi) = 0$. The L'Hôpital's rule gives that

$$\lim_{\omega\to\pi}\frac{\frac{ds(\omega)}{d\omega}}{\frac{dq(\omega)}{d\omega}} = \lim_{\omega\to\pi}\frac{s(\omega)}{q(\omega)} = \lim_{\omega\to\pi}\frac{\xi(a,p)\sin\left(\frac{-(1-a)^p\sin(\omega)}{(a+\cos(\omega))^2+\sin(\omega)^2}\right)}{\frac{1}{N}\sin(N(\omega-\pi))}$$
$$= -\xi(a,p)\frac{(1-a)^p}{(a-1)^2} = -\gamma(a,p).$$

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

$$\frac{d}{d\omega}\exp\Psi_a\left(e^{i\omega}\right)\Big|_{\omega=\pi} = i\frac{ds(\omega)}{d\omega}\big|_{\omega=\pi} = i\gamma(a,p).$$

On the other hand,

$$\frac{dG_a\left(e^{i\omega}\right)}{d\omega}\Big|_{\omega=\pi} = \frac{d}{d\omega}\left(-\xi(a,p) + \frac{\gamma(a,p)}{N}\left(e^{-iN(\omega-\pi)} - 1\right)\right)\Big|_{\omega=\pi}$$
$$= -i\gamma(a,p).$$

Hence

$$\frac{d}{d\omega}H_a\left(e^{i\pi}\right)\Big|_{\omega=\pi} = \frac{d}{d\omega}\exp\left(\Psi_a\left(e^{i\pi}\right)\right)\Big|_{\omega=\pi} + \frac{dG_a\left(e^{i\omega}\right)}{d\omega}\Big|_{\omega=\pi} = 0.$$

8

Therefore, Condition (b1) holds.

Let us show that Condition (b2) holds. We have that

$$\operatorname{Re} \Psi_a\left(e^{i\omega}\right) = \frac{f(a)(\cos(\omega)+a)}{(\cos(\omega)+a)^2+\sin(\omega)^2} = f(a)\frac{\operatorname{Re}\left(e^{i\omega}+a\right)}{\left|e^{i\omega}+a\right|^2}.$$

We have that $-\operatorname{Re}\left(e^{i\omega}+a\right)/\left|e^{i\omega}+a\right|$ is non-decreasing in $\omega \in [\omega_a, \pi]$ and converges to 1 as $\omega \to \pi - 0$, and $1/\left|e^{i\omega}+a\right|$ is non-decreasing in $\omega \in [\omega_a, \pi]$ and converges to $(1-a)^{-1}$ as $\omega \to \pi - 0$. Hence the product of these functions, $-f(a)^{-1}\operatorname{Re}\Psi_a\left(e^{i\omega}\right)$, is non-decreasing in $\omega \in [\omega_a, \pi]$. Hence we can select $\widehat{\omega}_a \in [\omega_a, \pi]$ such that $-\operatorname{Re}\Psi_a\left(e^{i\omega}\right) \geq -\operatorname{Re}\Psi_a\left(e^{i\pi}\right)/2$ for all $\omega \in [\widehat{\omega}_a, \pi]$, i.e., $\operatorname{Re}\Psi_a\left(e^{i\omega}\right) \leq \operatorname{Re}\Psi_a\left(e^{i\pi}\right)/2$ for all $\omega \in [\widehat{\omega}_a, \pi]$. In addition, $\operatorname{Re}\Psi_a\left(e^{i\omega}\right) = \operatorname{Re}\Psi_a\left(e^{-i\omega}\right)$. Hence $\operatorname{Re}\Psi_a\left(e^{i\omega}\right) \leq \operatorname{Re}\Psi_a\left(e^{i\pi}\right)/2$ for all $\omega \in [\widehat{\omega}_a, 2\pi - \widehat{\omega}_a] = [\pi - \delta_a, \pi + \delta_a]$, where $\delta_a = \pi - \widehat{\omega}_a$.

Further, we have that

$$\operatorname{Re}\Psi_a\left(e^{i\pi}\right) = (1-a)^p\frac{-1+a}{(-1+a)^2} \to -\infty \quad \text{as} \quad a \to 1.$$

For a given $\varepsilon > 0$, let us select $\bar{a}$ such that $\operatorname{Re}\Psi_a\left(e^{i\pi}\right)/2 < \log(\varepsilon/2)$ for all $a \geq \bar{a}$. In addition, we can select $\widetilde{a} \geq \bar{a}$ such that $\left|G_a\left(e^{i\omega}\right)\right| \leq \varepsilon/2$ for all $a \geq \widetilde{a}$ and all ω. Then Condition (b2) holds with $a = \widetilde{a}$ and $\delta = \delta_{\widetilde{a}}$ selected for given ε. Therefore, Condition (b) holds.

Let us show that Condition (c) holds. It follows from the proof of (a) above, that, for a given $\varepsilon > 0$ and Ω, we can select $\bar{a}$ such that (3) holds for $a \geq \bar{a}$. Further, let $q > 1$ be any. For any $\Omega_0 > \Omega$ and $\Omega_1 > \Omega_0$,

$$\sup_{\omega \in [-\Omega_1, -\Omega_0] \cup [\Omega_0, \Omega_1]} \left|M_{\theta,q}\left(e^{i\omega}\right) - 1\right| \to 0 \quad \text{as} \quad \theta \to 0.$$

Clearly, (4) holds for small enough θ. 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 shows examples of the frequency responses and the impulse functions for the filters described above.

Figure 1 shows the shapes of gain curves $\left|M_{\theta,q}\left(e^{i\omega}\right)\right|$ for reference non-causal filter (2) with $\theta = 0.02$, $q = 1.01$, and $\left|H_a\left(e^{i\omega}\right)\right|$ for near-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 $|H_a(e^{i\omega}) - 1|$ for near-ideal causal filters (5), with the same parameters as for Figure 1.

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

In particular, these examples show that the impulse response functions h_a can take negative values; i.e., these filters do not represent an averaging with a positive kernels.

4 Applications to the forecasting

A possible application of suggested above filters is preliminary smoothing of the input signals for the predicting algorithms. For this task, the causality is crucial. It is known that the band-limited sequences are predictable, i.e., the sequences are predictable if with the spectrum vanishing on a interval in $\mathbb{T}$. In addition, there are predictable sequences such that the spectrum is vanishing in a single point of $\mathbb{T}$; see [11, 12], where some predicable algorithms were suggested.

It can be noted that, the suggested above filters do not change the input sequences significantly if a is close to 1; the energy of the input is not damped on a given arc of $\mathbb{T}$. In fact, the energy is damped on a small neighborhood of the point $e^{i\pi} = -1$, and the size of this neighborhood converges to zero as $a \to 1$. Therefore, one cannot expect that the filters introduced above will help to improve the performance of the predicting algorithms [11] requiring that the spectrum is vanishing on a fixed arc on the unit circle.

However, it appears that these filters can help to improve the performance of predictors [12] oriented on processes x with spectrum vanishing in a single point. More precisely, predictors [12] are applicable for discrete time processes x such that, for some $\theta > 0$, $q > 1$, $c > 0$,

$$\sup_{\omega \in [-\pi, \pi]} |X(e^{i\omega})| \leq c M_{\theta,q}(e^{i\omega}), \quad X = \mathcal{Z} x. \tag{6}$$

In particular, it follows that filters (2) transfer sequences of a general type into predictable sequences such that (6) holds; respectively, filters (2) cannot be causal.

The predicting kernel [12] was defined as $k = k(\cdot, \gamma) = \mathcal{Z}^{-1} K$, where

$$K(z) \triangleq z \left(1 - \exp \left[-\frac{\gamma}{z + 1 - \gamma^{-r}} \right] \right), \tag{7}$$

10

and where $r > 0$ and $\gamma > 0$ are parameters. This predictor produces the process

$$y(t) = \sum_{d=-\infty}^{t} k(t-d)x(t)$$

approximating $x(t+1)$ for $\gamma \to +\infty$ for all inputs x satisfying (6) with $q > 1 + 2/r$. (In the notations from [12], $r = 2\mu/(q-1)$, where $\mu > 1$, $q > 1$ are the parameters). The function $K\left(e^{i\omega}\right)$ approximates the function $e^{i\omega}$ representing the forward one-step shift in the time domain; the value $|K\left(e^{i\omega}\right) - e^{i\omega}|$ is small everywhere but in a small neighborhood of $\omega = \pi$. Therefore, the process $y(t)$ represents an one step prediction of $x(t+1)$ if $X\left(e^{i\omega}\right)$ vanishes with a certain rate at $\omega = -\pi$. It was shown in [12] that

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

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 ℓ_∞, such that (6) holds with a given c.

Predictors (7) are robust with respect to some small noise contamination, meaning that the prediction error depends continuously on the intensity of the contaminating noise. However, for large γ, the values of $K\left(e^{i\omega}\right)$ can be very large in a neighborhood of $\omega = \pi$; in this case, the error can be large even for a small noise.

We suggest to apply filter (5) to compensate the presence of large values of $K\left(e^{i\omega}\right)$ in a small neighborhood of $\omega = \pi$ and therefore to reduce the impact of the presence of the high-frequency noise. This is illustrated by Figure 4 showing the shapes of error curves for approximation of the forward one step shift operator. More precisely, it shows the shape of $|K\left(e^{i\omega}\right) - e^{i\omega}|$ for the predictor (7) and the shape of $|K\left(e^{i\omega}\right) H_a\left(e^{i\omega}\right) - e^{i\omega}|$ for the transfer functions (5) and (7), which corresponds to preliminary smoothing of the input process by filters (5). These shapes characterize imperfection of the predictors, since the transfer function $e^{i\omega}$ corresponds to the one-step forward shift operator in time domain, i.e., $e^{i\omega}$ represents an ideal non-causal error-free one-step ahead predictor. It appears that the application of the filter improves the approximation of $e^{i\omega}$.

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

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

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

For the estimation of the effectiveness of predictors, we use the ratio

$$e(b_1, b_2) = \frac{\left(\sum_{t=1}^n |y(t-1) - x(t)|^2\right)^{1/2}}{\left(\sum_{t=1}^n |b_1 x(t-1) + b_2 x(t-2) - x(t)|^2\right)^{1/2}}, \tag{9}$$

where $b_k \in \mathbf{R}$ are parameters. The values $y(t-1)$ are supposed to be the predictions, at time $t-1$, of the future values $x(t)$. Since the sequence $\{x(t)\}$ does not satisfy (6) due to the presence of noise, a forecasting error is inevitable. Ratio (9) allows to compare the error of a predicting algorithm generating y and the error generated by a linear predictor with the coefficients b_1 and b_2. 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 (β_1, β_2) is known, then the optimal one step predictor of $x(t)$ is

$$y(t-1) = \beta_1 x(t-1) + \beta_2 x(t-2). \tag{10}$$

In this case, the value $n^{-1} \sum_{t=1}^n |\beta_1 x(t-1) + \beta_2 x(t-2) - x(t)|^2$ represents the sample mean of the squared error of this optimal predictor with known values of (β_1, β_2). Therefore, the optimal predictor (10) ensures that $e(\beta_1, \beta_2) \approx 1$ for a large enough n. Similarly, for any given n, an average value for $e(\beta_1, \beta_2)$ is also close to one for a sufficiently large number of Monte-Carlo trials, for optimal predictor (10). Respectively, any other predictor besides (10), including predictor (7), cannot achieve a lesser average value of $e(\beta_1, \beta_2)$ for a sufficiently large n or as an average value for a sufficiently large number of Monte-Carlo trials.

However, in many practical situations, the value of (β_1, β_2) is unknown, and, respectively, predictor (10) cannot be used. On the other hand, predictor (7) does not require to know (β_1, β_2) and can be applied in models with unknown or random and time variable (β_1, β_2) where predictors (10) is not applicable. In other words, predictor (7) can be applied for processes with unknown shape of the spectral representation. Therefore, it is reasonable to estimate the performance of a predictor using $e(b_1, b_2)$ with $b_k \neq \beta_k$, for instance, with b_k selected as the expected value, or the median, or upper or lower boundaries of unknown β_k.

12

Since it is impossible to implement convolution with infinitely supported kernels and inputs, one has to use truncated kernels and inputs for calculations. In the experiments described below, we replaced k and $h_a = \mathcal{Z}^{-1}H_a$ by the truncated kernels

$$k_d(t) = \mathbb{I}_{\{t \le d\}}k(t), \quad h_{a,d}(t) = \mathbb{I}_{\{t \le d\}}h_a(t), \quad d > 0. \tag{11}$$

In other words, the original $k = \mathcal{Z}^{-1}K$ and $h_a = \mathcal{Z}^{-1}H_a$ were used as some benchmarks; only their truncated versions were actually implemented.

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

- The algorithm is applied without filtering and produces $y = k_d \circ x$; we denote by $e_K(b_1, b_2)$ the corresponding values (9).

- The algorithm is applied with filtering and produces $y = (k_d \circ h_{a,d}) \circ x$; we denote by $e_{KH}(b_1, b_2)$ the corresponding values (9).

In both cases, $y(t)$ is calculated using historical data $\{x(s)\}_{t-d \le s \le t}$.

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

$$\gamma = 1.1, \quad r = 1.1, \quad a = 0.6, \quad p = 0.7, \quad N = 100, \quad m = 2. \tag{12}$$

Note that selection of too large γ makes calculation of k challenging, since it involves precise integration of fast growing $K\left(e^{i\omega}\right)$. The choice of parameters in (12) ensures that the values of $|K\left(e^{i\omega}\right)|$ are not large. Figure 5 shows the corresponding impulse response $\mathcal{Z}^{-1}H_a$. Figure 6 shows the corresponding impulse responses $\mathcal{Z}^{-1}K$ and $\mathcal{Z}^{-1}(KH_a)$.

In our experiment with AR(2) process, we used 10,000 Monte-Carlo trials with $n = d = 100$ and $\sigma = 0.3$. For each trial, we selected (β_1, β_2) randomly and independently. The distribution of (β_1, β_2) at each trial was the following: β_1 has the uniform distribution on the interval $(0, 1)$, and $\beta_2 = \xi\sqrt{1 - \beta_1^2}$, where ξ is a random variable independent on β_1 and uniformly distributed on the interval $(-1, 1)$. This choice ensures that the eigenvalues of the autoregression stay inside of the unit circle D almost surely.

We used MATLAB and standard personal computers; an experiment with 10,000 Monte-Carlo trials would take about five minutes of calculation time.

First, we compared the relative performance of our predictors with respect to the performance of the optimal predictor that requires that the values β_1 and β_2 are known. We obtained that the mean value over all Monte-Carlo trials of $e_K(\beta_1, \beta_2)$ is 1.5019 and the mean value of $e_{KH}(\beta_1, \beta_2)$ is 1.1177. This indicates that application of filter (5) improves the performance of the predictor. As was mentioned above, we cannot expect, for sufficiently many Monte-Carlo trials, the mean values of $e_K(\beta_1, \beta_2)$ and $e_{KH}(\beta_1, \beta_2)$ to be less than one, so the performance of the predictor (7) combined with filters (5) is reasonably good.

Second, we calculated the mean values $e_K(b_1, b_2)$ and $e_{KH}(b_1, b_2)$ with $b_k = \mathbf{E}\beta_k$, where $\mathbf{E}\beta_k$ is the population mean for β_k, $k = 1, 2$. For our parameters of the Monte-Carlo trials, we have $\mathbf{E}\beta_1 = 0.5$ and $\mathbf{E}\beta_2 = 0$. The corresponding values $e_K(0.5, 0)$ and $e_{KH}(0.5, 0)$ represent comparison of the performance of our predictor (without or with preliminary filtering) with an one-step predictor of $x(t)$ given by

$$y(t - 1) = (\mathbf{E}\beta_1)x(t - 1) + (\mathbf{E}\beta_2)x(t - 2).$$

Note that this predictor requires to know the population means of β_1 and β_2. We obtained that the mean value for $e_K(0.5, 0)$ is 1.2292 and the mean value for $e_{KH}(0.5, 0)$ is 0.9545. These numbers indicate a good performance of our filter/predictor system, especially if we take into account that our system does not require to know the values $\mathbf{E}\beta_k$.

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 (12). Figure 8 shows sample paths of AR(2) process $x(t)$ and outputs $y(t)$ of predictor [12] without preliminary filtering and with preliminary filtering using filter (5) with the parameters defined by (12). It shows the values $y(t - 1)$, i.e., predictions of $x(t)$, versus the values of $x(t)$.

In addition, we considered a modification of process (8) with $\beta_2 \equiv 0$, i.e., AR(1) process. We used the same predictors and filters as for the experiments with AR(2) process described above.

We set again $10,000$ Monte-Carlo trials with $n = d = 100$, with $\sigma = 0.3$, and with randomly selected β_1 such that β_1 was distributed uniformly on the interval $(0, 1)$. We compared the relative performance of our predictors with respect to the performance of the optimal predictor that requires that the value β_1 is known. We obtained that the mean value of $e_K(\beta_1, 0)$ is 1.2830 and the mean value of $e_{KH}(\beta_1, 0)$ is 1.1023. The numbers indicate again that the use of the filter improves the performance of the predictor. 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 reasonably good.

Finally, we calculated the values $e_K(b_1, 0)$ and $e_{KH}(b_1, 0)$ with $b_1 = \mathbf{E}\beta_1 = 0.5$ for AR(1) process defined by (8) with $\beta_2 \equiv 0$. In other words, we compared the relative performance of the one step predictor (5) with or without filtering with respect to the predictor $y(t) = (\mathbf{E}\beta_1)x(t)$ that requires to known the population mean for β_1. We obtained that $e_K(0.5, 0) = 1.1824$ and $e_{KH}(0.5, 0) = 1.0155$. This shows again that application of the filter reduces the forecasting error. Given that our system does not require to know $\mathbf{E}\beta_1$, the performance is reasonably good.

The results of these experiments appear to be consistent and stable with respect to the variations of the parameters.

On the impact of truncation

Since it is impossible to implement convolution with infinitely supported kernels and inputs, we have to run numerical calculations with truncated processes. Let us show that the described above filtering and forecasting are robust with respect to the truncation, i.e., that the truncation in (11) has a vanishing impact for large d. Since $H_a\left(e^{i\omega}\right) \in L_\infty(-\pi, \pi)$, we have that $h_a \in \ell_2$. Hence $h_{a,d} \circ x \to h_a \circ x$ in ℓ_∞ as $d \to +\infty$ for $x \in \ell_2$; in practice, only truncated inputs $x \in \ell_2$ are available. It can be also noted that the predicting kernel $k = \mathcal{Z}^{-1}K$ defined by (7) belongs to ℓ_2. Therefore, the kernel $k \circ h_{a,d}$ converges to $k \circ h_a$ in ℓ_∞ as $d \to +\infty$.

Our numerical experiments for autoregressions with the coefficients deemed to untraceable demonstrated that truncation with relatively small $d = 100$ does not diminish a good forecasting performance.

5 Conclusion

The paper proposes a family of causal smoothing filters. These filters are near-ideal meaning that a higher rate of damping of the energy on the high frequencies would lead to the loss of causality; this is because they approximate non-causal filters transferring non-predicable processes into predictable ones. A possible application is preliminary smoothing of the inputs for predicting algorithms. Certain mild but stable improvement of forecasting accuracy is demonstrated in experiments with simple autoregressions in a setting where theirs coefficients are deemed to be untraceable.

It could be interesting to investigate the computational limits of the algorithms described above, for

instance, for long sequences, for higher order autoregressions, or for other types of input processes. It could be interesting to find other filters with similar properties. It could be useful to represent the filtering algorithm in the terms of the discrete Fourier transform. We leave this for future work.

Acknowledgment

References

References

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

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

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

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

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

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

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

[8] N. Dokuchaev. (2008). 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**, 382002 (7pp).

[9] N. Dokuchaev. (2010). Predictability on finite horizon for processes with exponential decrease of energy on higher frequencies, *Signal processing* **90** (2), 696–701.

[10] N. Dokuchaev. (2012). On sub-ideal causal smoothing filters. *Signal Processing* **92**, iss. 1, 219-223.

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

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

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

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

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

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

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

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

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

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

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

[22] A. Lindquist, G. Picci. (2015). Linear Stochastic Systems *A Geometric Approach to Modeling, Estimation and Identification.* Springer-Verlag, Berlin Heidelberg.

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

[24] R. von Mises. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Z.* 5, 52–99.

[25] A.G. Miamee, and M. Pourahmadi. (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] B. Simon. (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] G. Szegö. (1920). Beiträge zur Theorie der Toeplitzschen Formen. *Math. Z.* 6, 167–202.

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

[29] S. Verblunsky. (1936). On positive harmonic functions (second paper). *Proc. London Math. Soc.* 40, 290–320.
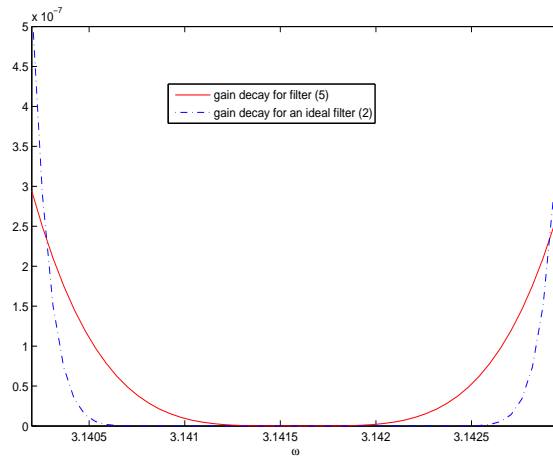
Figure 1: Gain decay: the values $\left|M_{\theta,q}\left(e^{i\omega}\right)\right|$ for a non-causal filter (2) with $\theta = 0.02$ and $q = 1.01$, and $\left|H_a\left(e^{i\omega}\right)\right|$ for a causal filter (5) with $a = 0.99$, $p = 0.6$, $N = 50$, $m = 2$.
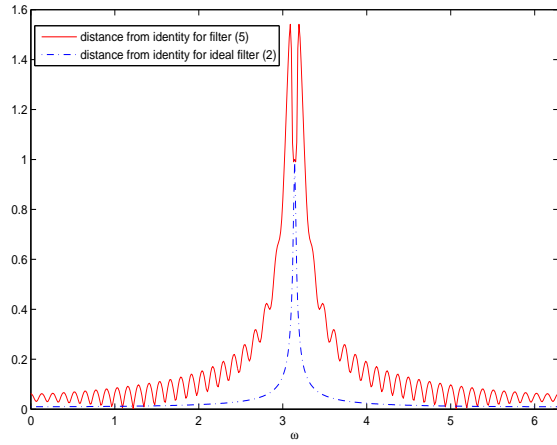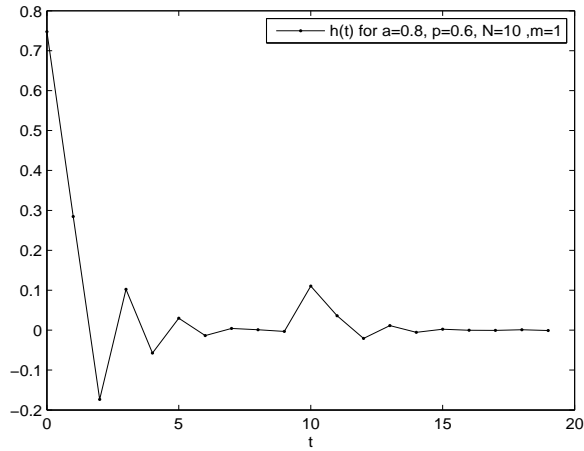
Figure 2: Approximation of identity operator: shapes of the distances from 1, i.e, for the values $|M_{\theta,q}\left(e^{i\omega}\right) - 1|$ and $|H_a\left(e^{i\omega}\right) - 1|$, for a non-causal filter (2) with $\theta = 0.02$ and $q = 1.01$, and $|H_a\left(e^{i\omega}\right)|$ for a causal filter (5) with $a = 0.99$, $p = 0.6$, $N = 50$, $m = 2$.

Figure 3: Impulse response $h_a(t) = (\mathcal{F}^{-1}H_a)(t)$ for causal filter (5) with $a = 0.8$, $p = 0.6$, $N = 10$, $m = 1$.
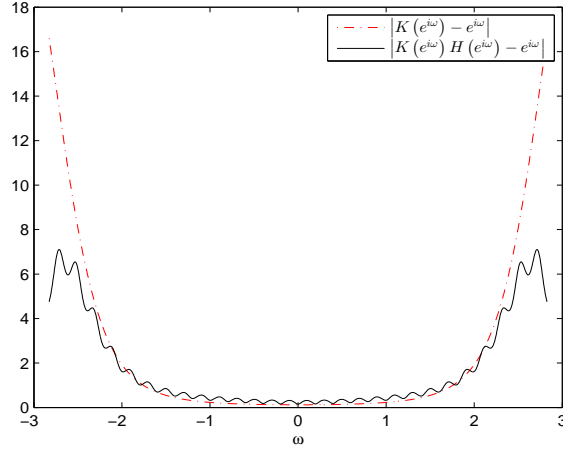
Figure 4: Approximation of the one-step forward shift operator: the values $\left|K\left(e^{i\omega}\right)-e^{i\omega}\right|$ for the transfer function of the predictor (7) and $\left|K\left(e^{i\omega}\right)H_\nu\left(e^{i\omega}\right)-e^{i\omega}\right|$, i.e., with smoothing of the input process by filters (5) with $\gamma=2.5$, $r=0.225$, $a=0.8$, $p=0.7$, $N=30$, $m=3$.

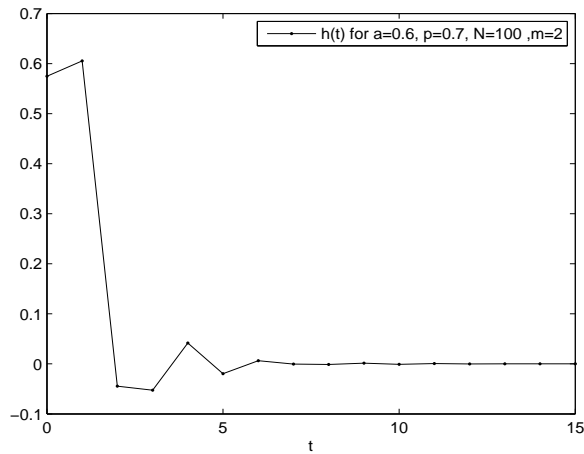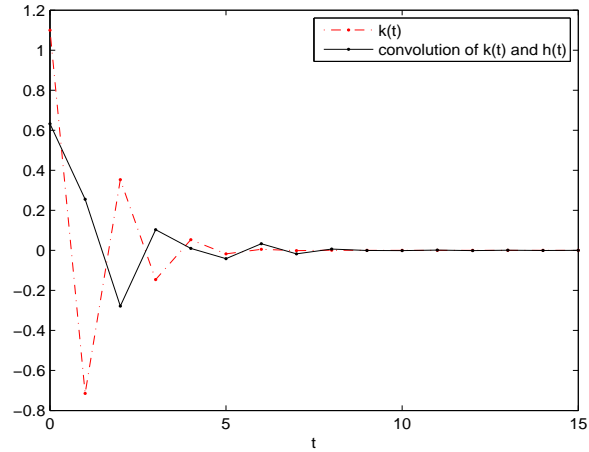**Figure 5:** Impulse response $h_a(t)=(\mathcal{F}^{-1}H_a)(t)$ for causal filter (5) with the parameters given in (12).

Figure 6: The impulse response $\mathcal{Z}^{-1}K$ of predictor (7) and the impulse response $\mathcal{Z}^{-1}(KH_a)$ of the predictor combined with filters (5) with the parameters given in (12).
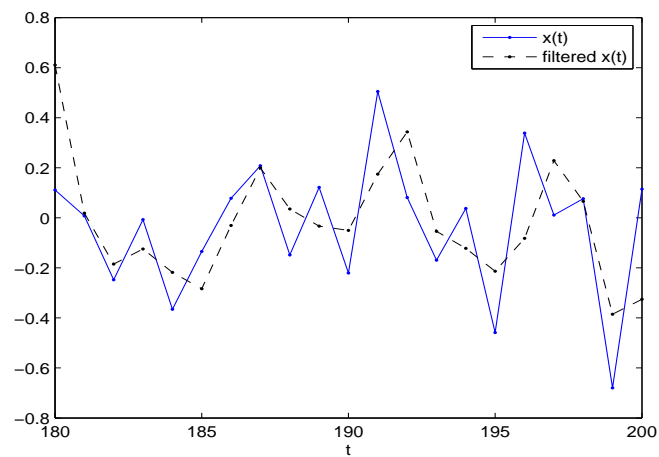

Figure 7: A path of AR(2) process $x(t)$ versus the output of filter (5) with the parameters given in (12).
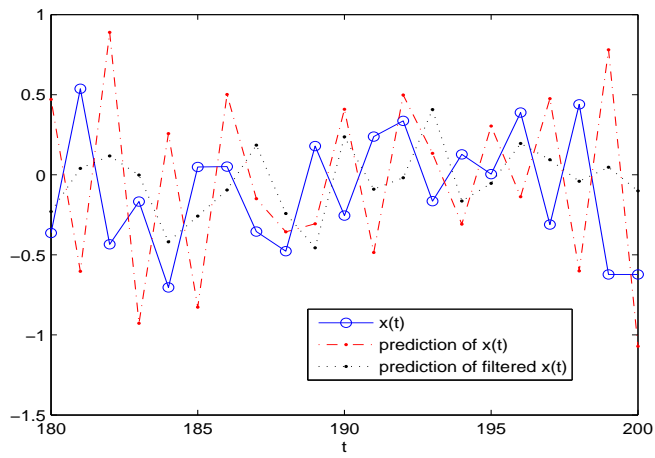
Figure 8: A path of AR(2) process $x(t)$ versus two predictions $y(t-1)$ of $x(t)$; one was calculated without filtering, and another was calculated after application of filter (5) with the parameters given in (12).