Research Article

Mathematical Methods in the Applied Sciences

Received 27 October 2015 Published online 16 March 2016 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/mma.3795
MOS subject classification: 62E17, 6207, 41Axx

Adaptive Fourier tester for statistical estimation

Qiuhui Chen[a*†], Tao Qian[b], Yuan Li[c], Weixiong Mai[d] and Xingfa Zhang[c]

Communicated by W. Sprößig

Based on Takenaka–Malmquist (TM) system, a new nonparametric estimator for probability density function is proposed. The TM estimation method is completely different from the existent density estimation methods in that the estimator depends on an approximate system with poles in a complex plane. Compared with the classic Fourier estimator, the TM estimator will offer more flexibility and adaptivity for real data due to the poles and nonlinearity of the phase of TM system. We compare the TM estimator with kernel, wavelet, and spline estimators by simulations. It shows that the introduced TM estimator is a more promising method than the existing and commonly used methods. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: TM estimator; Hardy space; Cayley transform; adaptive; approximation order

1. Introduction

There have been many discussions about probability density estimations. M. Rosenblatt [1] applied the kernel method to estimate the density function. After Rosenblatt, many researchers studied this problem by the kernel method, such as E. Parzen [2] and S. Kumar [3]. Other estimation methods of density functions were also developed, for example, orthogonal series methods, interpolation methods, and the approach of characteristic functions. For details, see S. C. Schwartz [4], G. Walter [5], and J. Blum and V. Sursala [6]. G. Walter and J. Blum [7] studied the density function by delta sequences. N. Hjort and M. Jones [8] proposed a local kernel-smoothed likelihood function to estimate the density function.

In this paper, we study the probability density estimation by using the Takenaka–Malmquist (TM) systems. The TM systems, including the Laguerre and the two-parameter Kautz systems, are natural generalizations of the half-Fourier system. They have enjoyed a long-term development with ample applications in applied mathematics [9–12] and engineering, including control theory and signal processing [13–15]. From the point of view of approximation, the TM systems compare favorably with Fourier system, spline systems, and wavelet system because they can offer more flexibility and adaptivity due to the parameters in structure of the TM systems. Compared with other systems, the biggest advantage of the TM systems lies in the nonlinearity of phase functions; that is, when we write the atoms of the TM systems in the polar form, the corresponding phase functions are nonlinear functions. In language of signal processing, the TM systems have the form of amplitude modulation and frequency modulation and can offer adaptive approximation for transient signals. Then based on the TM systems, we construct the probability density estimators in the contexts of both unit disk $\mathbb{D} = \{|z| < 1\}$ and the upper half plane $\mathbb{C}^+ = \{z : \operatorname{Im} z > 0\}$. Such estimators are more appropriate and effective for nonlinear and non-stationary process.

This paper is organized as follows. Section 2 introduces the TM systems and investigates their approximation. Section 3 focuses on the design and approximation of probability density estimators from the TM systems. Based on observation data from a density, an algorithm for adaptive choices of parameters in the TM estimators is proposed. Section 4 is devoted to numerical simulation about our estimators. The Appendix contributes to the case of the upper half plane.

[a] *College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China*
[b] *Department of Mathematics, Faculty of Science and Technology, University of Macau, Macao, China*
[c] *Department of Statistics, Guangzhou University, Guangzhou, China*
[d] *Department of Mathematics, Faculty of Science and Technology, University of Macau, Macao, China*
* *Correspondence to: Qiuhui Chen, College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China.*
† *E-mail: chenqiuhui@hotmail.com*

2. Takenaka–Malmquist system and approximation

The TM systems are closely related to the Hardy space $H^p(\mathbb{D})$ on the unit disk $\mathbb{D}$. Here, we say $f \in H^p(\mathbb{D})$, $0 < p < \infty$, if

$$\sup_{r\in(0,1)} \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(re^{i\theta})|^p d\theta = \|f\|^p_{H^p(\mathbb{D})} < \infty.$$

For $p = \infty$, we say $f \in H^\infty(\mathbb{D})$ if f is a bounded analytic function on $\mathbb{D}$ and we write

$$\|f\|_{H^\infty(\mathbb{D})} = \sup_{z\in\mathbb{D}} |f(z)|.$$

The Hardy space $H^p(\partial\mathbb{D})$ on the unit circle $\partial\mathbb{D} = \{z : |z| = 1\}$ consists of the non-tangential boundary limits of the related holomorphic Hardy space functions inside the open unit disk. For the same p, the two types of Hardy spaces are isometrically isomorphism. In the case $p = 2$, as a reproducing kernel Hilbert space, $H^2(\mathbb{D})$ is equipped with the inner product

$$\langle f, g\rangle_{\mathbb{D}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})\overline{g(e^{i\theta})}d\theta, \ \ f, g \in H^2(\mathbb{D}).$$

The TM systems are natural generalizations of the half-Fourier system. There are two parallel contexts for the TM system theory, namely the unit disk and the upper-half complex plane. We first introduce the case of the unit disk. The upper half plane case will be discussed in the Appendix. For a given parameter sequence $\{a_n\}_{n=0}^{\infty} \subset \mathbb{D}$, the corresponding TM system $\{e_n\}_{n=0}^{\infty}$ is

$$e_n(z) = \frac{\sqrt{1-|a_n|^2}}{1-\overline{a_n}z} B_{n,\mathbb{D}}(z), \ \ z \in \mathbb{D}, \ \ n \in \mathbb{Z}_+, \tag{2.1}$$

where $B_{n,\mathbb{D}}$ denotes the nth Blaschke product on the unit disk

$$B_{n,\mathbb{D}}(z) = \prod_{j=0}^{n-1} \frac{z-a_j}{1-\overline{a_j}z}.$$

Here, for $n = 0$, we adopt the convention $\prod_{j=0}^{-1} = 1$. Throughout this paper, we set $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$.

The case $a_n = 0, n \in \mathbb{N}$ corresponds to the half-Fourier system that gives rise to classical Fourier analysis. It is an important result in classical Fourier analysis that in all $L^p(\partial\mathbb{D})$ (or $H^p(\partial\mathbb{D})$), $1 < p < \infty$, the Fourier system (or the half-Fourier system) is a Schauder basis (see, for instance, [16]). It is well known that, under the condition $\sum_{n\in\mathbb{Z}}(1 - |a_n|) = \infty$, $\{e_n\}_{n=0}^{\infty}$ is an orthonormal base of the Hilbert space $H^2(\partial D)$. Furthermore, in [17], it is proven that $\{e_n\}_{n=0}^{\infty}$ is a Schauder base for the Banach space $H^p(\partial D)$ for $p \in (1, \infty)$.

Denote by $\mathcal{H}$ the circular Hilbert transform for $f \in L^p(\partial\mathbb{D})$. Then for any $f \in L^p(\partial\mathbb{D})$, one obtains the decomposition $f = f^+ + f^-$ with $f^+ = \frac{1}{2}(f + i\mathcal{H}f) \in H^p(\partial\mathbb{D})$ and $f^- = \frac{1}{2}(f - i\mathcal{H}f) \in H^p(\partial(\mathbb{C} \setminus \mathbb{D}))$. We call f^+ circular analytic signal and f^- circular conjugate analytic signal. Both of them can be extended to be analytic functions in $H^p(\mathbb{D})$ and $H^p(\mathbb{C} \setminus \mathbb{D})$, respectively. For simplicity, we will use the same notation f^+ or f^- for the analytic functions in the respective regions and their corresponding boundary values.

We need to revisit some results about approximation order of the partial sum sequence based on the TM systems, the importance of which lies in that the mean square error of the TM system estimator (see next section) is dominated by the variance of the estimator and the approximate rate of the partial sum sequence.

For any function f analytic in the unit disk, set

$$L_{n,\mathbb{D}}f(z) = \sum_{k=0}^{n} c_k e_k(z), \ z = re^{ix} \in \mathbb{D} \tag{2.2}$$

with $c_k = \int_{-\pi}^{\pi} f(e^{it})\overline{e_k(e^{it})}dt$. The operator $L_{n,\mathbb{D}}$ has the integral representation

$$L_{n,\mathbb{D}}f(z) = \int_{|\xi|=1} f(\xi)K_{n,\mathbb{D}}(z,\xi)\frac{d\xi}{i\xi}, \ z = re^{ix}$$

with the kernel $K_{n,\mathbb{D}}(z,\xi) = \sum_{k=0}^{n} \overline{e_k(\xi)}e_k(z) = \frac{1}{1-\bar{\xi}z}\left(1 - B_{n,\mathbb{D}}(z)\overline{B_{n,\mathbb{D}}(\xi)}\right)$ by applying the Christoffel–Darboux formula.

For real variable- and complex-valued function having Fourier expansion $f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$, the corresponding partial sum operator $S_{n,\mathbb{D}}$ is defined by the symmetric sum

$$S_{n,\mathbb{D}}f(x) = \sum_{k=-n}^{n} c_k e_k(x), \ x \in [-\pi, \pi]. \tag{2.3}$$

For simplicity, we use $e_k(x)$ instead of $e_k(e^{ix})$, the non-tangential limit of the TM atom defined in (2.1). For $k > 0$, set $e_{-k} = \overline{e_k(x)}$. The integral representation of the partial sum operators is in the succeeding text:

$$S_{n,\mathbb{D}}f(x) = \int_{-\pi}^{\pi} f(t)D_{n,\mathbb{D}}(x,t)dt, \tag{2.4}$$

where $D_{n,\mathbb{D}}(x,t) = K_{n,\mathbb{D}}(e^{ix}, e^{it}) + \overline{K_{n,\mathbb{D}}(e^{ix}, e^{it})} - \overline{K_{0,\mathbb{D}}(e^{ix}, e^{it})}$. It can be checked that

$$D_{n,\mathbb{D}}(x,t) = \frac{\sin\left(\frac{x-t}{2} + \theta_n(x,t)\right)}{2\pi \sin\left(\frac{x-t}{2}\right)},$$

where $\theta_n(x,t) = \sum_{k=1}^{n} (\theta_{a_n}(x) - \theta_{a_n}(t))$ and θ_{a_n} is the indefinite integral of the periodic Poisson kernel $p_{a_n}(t) = \frac{1-|a_n|^2}{1-2|a_n|cos(t-t_{a_n})+|a_n|^2}$ with the parameter $a_n = |a_n|\exp(it_{a_n})$. In particular, when all parameters are zero, $D_n(x,t)$ is just the usual Dirichlet kernel

$$D_n(x-t) = \frac{\sin\left(n+\frac{1}{2}\right)(x-t)}{2\pi\left(\frac{1}{2}(x-t)\right)} = \frac{1}{\pi}\left(\frac{1}{2} + \sum_{k=1}^{n}\cos k(x-t)\right).$$

Now, we turn to the approximation behavior of the operator $S_{n,\mathbb{D}}$. The approximation of $S_{n,\mathbb{D}}$ is closely related to that of $L_{n,\mathbb{D}}$, which was investigated in [18, 19].

Define by $\mathbb{B}_0\mathbb{H}_1$ the space that any element is analytic in $\mathbb{D}$, is continuous in $\overline{\mathbb{D}}$, and has finite variation $V(f) = 1$ at the boundary $\partial\mathbb{D}$. Generally, $f \in \mathbb{B}_r\mathbb{H}_1$ means that f is analytic in $\mathbb{D}$, $f^{(p)}, 1 \le p \le r$, is continuous on $\partial\mathbb{D}$, and $f^{(r)} \in B_0H_1$. Russian mathematician Pycak established the following proposition:

Proposition 2.1
Suppose that $f \in \mathbb{B}_r\mathbb{H}_1$ for some nonnegative integer r and $\omega(f^{(r)}, t) = O(t^\alpha)$ for some $\alpha \in (0,1)$. Then there exists some suitable parameter vector $a = (a_0, \dots, a_{n-1})$ such that

$$\max_{|z|\le 1} |f(z) - L_{n,\mathbb{D}}f(z)| \le C\frac{\ln^3 n}{n^{r+1}}, \tag{2.5}$$

where C is only dependent on α and $\omega(f,t)$ is the continuous modulus of f defined by $\omega(f,t) = \sup_{h\in[0,t],\theta\in[-\pi,\pi]} |f(e^{i(\theta+h)}) - f(e^{i\theta})|$.

To apply the complex approximation (2.5) to that of a real variable signal, we need the following lemma.

Lemma 2.2
Set $a_0 = 0$. Suppose that $f \in L^2(\partial\mathbb{D})$. Then

$$S_{n,\mathbb{D}}f(x) - f(x) = 2\mathrm{Re}\left[L_{n,\mathbb{D}}f^+(z) - f^+(z)\right]_{z=e^{ix}}. \tag{2.6}$$

Proof
This identity can be obtained by involving Cauchy integral equation and the identities $\langle f^+, e_k\rangle = \overline{\langle f^-, e_{-k}\rangle}$ for any integer k and $\langle f^+, e_k\rangle = 0$ for $k < 0$. □

Combing this lemma with Pycak's proposition, an estimation for pointwise approximation of $S_{n,\mathbb{D}}$ is established in the succeeding text.

Theorem 2.3
Suppose that $a_0 = 0$, $f \in L^2(-\pi,\pi)$, $f^+ \in \mathbb{B}_r\mathbb{H}_1$, and $\omega((f^+)^{(r)}, t) = O(t^\mu)$ for some $\mu \in (0,1)$. Then there exists some suitable parameter vector $a = (a_0, \dots, a_{n-1})$ such that

$$\max_{x\in[-\pi,\pi]} |S_{n,\mathbb{D}}f(x) - f(x)| \le C\frac{\ln^3 n}{n^{r+1}} \tag{2.7}$$

where C is only dependent of μ.

As for the approximation order of $S_{n,\mathbb{D}}$ in the sense of L^2-norm, Plancherel theorem indicates that it depends on the decaying rate of the Fourier coefficients. Recall that the order of the nth Fourier coefficient of any twice continuously differentiable function in $L^2(-\pi,\pi)$ is $O\left(\frac{1}{|n|^2}\right)$ (see corollary 2.4, p. 42 of [20]). This result was generalized by A. Bultheel and P. Carrette in 2003 as follows (see [21]).

Proposition 2.4
Assume that $a_0 = 0$ and f is Kth continuously differentiable function with periodic 2π with $K \ge 3$, then decaying order of the generalized Fourier coefficients $\langle f, e_n\rangle$ is $O(n^{-(K-1)})$.

Applying Proposition 2.4 to the operator $S_{n,\mathbb{D}}$, the following result holds.

Theorem 2.5
Assume that $a_0 = 0$ and suppose that f is third continuously differentiable. Then

$$\|S_{n,D}f - f\|_2^2 = O\left(n^{-1}\right). \tag{2.8}$$

Proof
This is a direct consequence from Proposition 2.4 and the identity $\|S_{n,\mathbb{D}}f - f\|_2^2 = \sum_{|k|\geq n+1} |c_k|^2$. □

3. Takenaka–Malmquist estimators

Suppose that X is a random variable having density $q \in L^2(\mathbb{R})$ with compact support $[-\pi, \pi]$. Let $X_1, X_2, \cdots, X_N$ be the i.i.d. samples from X. Define the TM estimators by

$$\hat{q}_{n,N}(x) := \sum_{k=-n}^{n} \hat{c}_k e_k(x),\ x \in [-\pi, \pi], n \in \mathbb{Z}_+, \tag{3.9}$$

where

$$\hat{c}_k = \frac{1}{2\pi N} \sum_{j=1}^{N} e_k(X_j),\ k \in \mathbb{Z}. \tag{3.10}$$

We give a brief explanation why we prefer the estimators (3.9). Any L^2-function f can be approximated by the sequence $\sum_{k=-n}^{n} c_k e_k(\cdot)$, where c_k is the generalized Fourier coefficient, that is, the inner product of f and e_k. In our case, the density function is unknown and only finite observation data are given. For retrieval of the density, we design the empirical inner $\hat{c}_k$ instead of c_k and then construct the specific estimators (3.9).

3.1. *Properties of Takenaka–Malmquist estimators*

This section investigates some statistical properties of TM estimators. The following theorem states that the estimator $\hat{q}_{n,N}$ is an asymptotically unbiased estimator of the density function q.

Theorem 3.1
For the estimator $\hat{q}_{n,N}$, as $n, N \to \infty$ and $\frac{n}{N} \to 0$, we have

1. $\lim_{n\to\infty} \hat{q}_{n,N}(x) = q(x), \quad \forall x \in [-\pi, \pi]$; and
2. $Var(\hat{q}_{n,N}(x)) = O\left(\frac{n}{N}\right) = o(1)$.

The next theorem describes the approximation order of $\hat{q}_{n,N}$ in the sense of L^2-norm convergence.

Theorem 3.2
Suppose that $a_0 = 0$ and $\sum_{k=-n}^{n} (1 - |a_k|)^{-1} = O(n)$. If the density function q is compactly supported in $[-\pi, \pi]$, $q(-\pi) = q(\pi)$ and third continuously differentiable, then

$$E\left(\|\hat{q}_{n,N} - q\|_2^2\right) = O\left(\frac{1}{\sqrt{N}}\right) \quad \text{with} \quad n = \sqrt{N}. \tag{3.11}$$

The following two theorems discuss some auxiliary statistic errors of $\hat{q}_{n,N}$ instead of norm error, that is, $E\left((\hat{q}_{n,N}(x) - q(x))^2\right)$ and $E\left(\mu\left\{x : |\hat{q}_{n,N}(x) - q(x)| > \epsilon\right\}\right)$. Here, μ is the usual Lebesgue measure. About the mean square error, we have the following theorem.

Theorem 3.3
Assume that $a_0 = 0$. Suppose that the density function q satisfies the conditions in Theorem 2.3 with $r > 0$. Then

$$E\left(\hat{q}_{n,N}(x) - q(x)\right)^2 = O\left(\frac{1}{\sqrt{N}}\right), \quad n = \sqrt{N}.$$

About the quantity $E\left(\mu\left\{x : |\hat{q}_{n,N}(x) - q(x)| > \epsilon\right\}\right)$, we have the following result.

Theorem 3.4
Suppose that $a_0 = 0$ and $\sum_{k=-n}^{n} (1 - |a_k|)^{-1} = O(n)$. If the density function q is compactly supported in $[-\pi, \pi]$, $q(-\pi) = q(\pi)$ and third continuously differentiable, then

$$E\left(\mu\left\{x : |\hat{q}_{n,N}(x) - q(x)| > \epsilon\right\}\right) = O\left(\frac{1}{\sqrt{N}}\right) \quad \text{with} \quad n = \sqrt{N}. \tag{3.12}$$

Proof of Theorem 3.1

Proof

(1) Note that

$$E(\hat{c}_k) = \frac{1}{2\pi}E(e_k(X_1)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e_k(x)q(x)dx = c_k =< q, e_k >$$

and

$$E(\hat{q}_{n,N}(x)) = \sum_{k=-n}^{n} E(\hat{c}_k)e_k(x) = \sum_{k=-n}^{n} c_k e_k(x),\ x \in [-\pi,\pi].$$

It follows from Theorem 2.3 that as $n \to \infty$,

$$\lim_{n\to\infty} S_{n,\mathbb{D}}\, q(x) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} c_k e_k(x) = q(x),\ x \in [-\pi,\pi].$$

Then

$$\lim_{n\to\infty} E(\hat{q}_{n,N}(x)) = q(x),\ x \in [-\pi,\pi].$$

(2) By direct calculation, we obtain

$$\begin{aligned} Var(\hat{q}_{n,\mathbb{N}}(x)) &= Var\left[\sum_{k=-n}^{n} \hat{c}_k e_k(x)\right] = Var\left[\sum_{j=1}^{N}\left(\sum_{k=-n}^{n} \frac{1}{2\pi N} e_k(x)e_k(X_j)\right)\right] \\ &= NVar\left(\sum_{k=-n}^{n} \frac{1}{2\pi N} e_k(x)e_k(X_1)\right) \\ &\le NE\left(\sum_{k=-n}^{n} \frac{1}{2\pi N} e_k(x)e_k(X_1)\right)^2 \\ &= \frac{1}{4\pi^2 N}\int_{-\pi}^{\pi}\left(\sum_{k=-n}^{n} e_k(x)e_k(t)\right)^2 q(t)dt \\ &= \frac{1}{4\pi^2 N}\int_{-\pi}^{\pi} D_n^2(x,t)q(t)dt, \end{aligned}$$

where $D_n(x,t) = \sum_{k=-n}^{n} e_k(x)e_k(t)$. Then we have

$$Var(\hat{q}_{n,N}(x)) \le \frac{1}{4\pi^2 N}\|q\|_\infty \|D_n(x,\cdot)\|_2^2.$$

It is easily shown that $\|D_n(x,\cdot)\|_2^2 \le n$. Then it holds that

$$Var(\hat{q}_{n,N}(x)) = O\left(\frac{n}{N}\right) = o(1),\ x \in [-\pi,\pi].$$

□

Proof of Theorem 3.2

Proof

The outline of our proof is as follows. By Cauchy–Schwartz inequality, we have

$$E\left(\|\hat{q}_{n,N} - q\|_2^2\right) \le 2E\left(\|\hat{q}_{n,N} - q_n\|_2^2\right) + 2\|q_n - q\|_2^2,$$

where $q_n = S_{n,\mathbb{D}}\, q$. On one hand, by noting that the second term $\|q_n - q\|_2^2$ in the previous inequality is just the L^2-norm error of the partial sum operator $S_{n,\mathbb{D}}$ and using Theorem 2.5, we know that $\|q_n - q\|_2^2 = O\left(\frac{1}{n}\right)$. On the other hand, it suffices to conclude (3.11) if we prove that

$$E\left(\|\hat{q}_{n,N} - q_n\|_2^2\right) = O\left(\frac{n}{N}\right). \tag{3.13}$$

The proof about the estimation for the quantity $E\left(\|\hat{q}_{n,N} - q_n\|_2^2\right)$ is a little complicated. We offer the detail in the succeeding text.

By Fubini theorem and the finite additivity of integral, we have

$$
\begin{aligned}
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right) &= E\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}|\sum_{k=-n}^{n}(\hat{c}_k-c_k)e_k(x)|^2dx\right)\\
&= \frac{1}{2\pi}E\left(\int_{-\pi}^{\pi}\sum_{k=-n}^{n}\sum_{j=-n}^{n}e_k(x)\overline{e_j(x)}\left(\hat{c}_k\bar{\hat{c}}_j-c_k\bar{\hat{c}}_j-\hat{c}_k\overline{c_j}+c_k\hat{c}_j\right)dx\right)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k=-n}^{n}\sum_{j=-n}^{n}e_k(x)\overline{e_j(x)}E\left(\hat{c}_k\bar{\hat{c}}_j-c_k\bar{\hat{c}}_j-\hat{c}_k\overline{c_j}+c_k\hat{c}_j\right)dx\\
&= \frac{1}{2\pi}\sum_{k=-n}^{n}\sum_{j=-n}^{n}E\left(\hat{c}_k\bar{\hat{c}}_j-c_k\bar{\hat{c}}_j-\hat{c}_k\overline{c_j}+c_k\hat{c}_j\right)\int_{-\pi}^{\pi}e_k(x)\overline{e_j(x)}dx.
\end{aligned}
$$

Applying the orthonormality of the system $\{e_k\}$ and the fact $E(\hat{c}_k)=c_k$, it follows

$$
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right)=\sum_{k=-n}^{n}\left(E\left(|\hat{c}_k|^2\right)-|c_k|^2\right). \tag{3.14}
$$

Now, we calculate $E\left(|\hat{c}_k|^2\right)$. By the definition of $\hat{c}_k$, we have

$$
\begin{aligned}
&E\left(|\hat{c}_k|^2\right)\\
&= E\left(\frac{1}{4\pi^2N^2}\sum_{j=1}^{N}\sum_{l=1}^{N}e_k(X_j)\overline{e_k(X_l)}\right)\\
&= \frac{1}{4\pi^2N^2}\sum_{j=1}^{N}\sum_{l=1}^{N}E\left(e_k(X_j)\overline{e_k(X_l)}\right)\\
&= \frac{1}{4\pi^2N^2}\sum_{j=1}^{N}E\left(|e_k(X_j)|^2\right)+\frac{1}{4\pi^2N^2}\sum_{j,l=1,j\neq l}^{N}E\left(e_k(X_j)\overline{e_k(X_l)}\right)\\
&= \frac{1}{4\pi^2N}E\left(|e_k(X)|^2\right)+\frac{1}{4\pi^2N^2}\sum_{j,l=1,j\neq l}^{N}E\left(e_k(X_j)\right)\overline{E\left(e_k(X_l)\right)}\\
&= \frac{1}{4\pi^2N}E\left(|e_k(X)|^2\right)+\frac{1}{4\pi^2N^2}|\sum_{j=1}^{N}E\left(e_k(X_j)\right)|^2-\frac{1}{4\pi^2N^2}\sum_{j=1}^{N}|E\left(e_k(X_j)|^2\right)\\
&= \frac{1}{4\pi^2N}E\left(|e_k(X)|^2\right)+\frac{1}{4\pi^2}|E\left(e_k(X)\right)|^2-\frac{1}{4\pi^2N}|E\left(e_k(X)\right)|^2.
\end{aligned}
$$

Using the fact $E(e_k(X))=2\pi\bar{c}_k$ for $k=-n,\dots,n$, it follows

$$
E\left(|\hat{c}_k|^2\right)-|c_k|^2=\frac{1}{4\pi^2N}E\left(|e_k(X)|^2\right)-\frac{1}{N}|c_k|^2. \tag{3.15}
$$

Combing (3.14) and (3.15), we have

$$
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right)=\sum_{k=-n}^{n}\left(\frac{1}{4\pi^2N}E\left(|e_k(X)|^2\right)-\frac{1}{N}|c_k|^2\right). \tag{3.16}
$$

Noting that $E\left(|e_k(X)|^2\right)=\int_{-\pi}^{\pi}\frac{1-|a_k|^2}{|1-\bar{a}_ke^{ix}|^2}q(x)dx$, which has lower bound $\frac{1-|a_k|}{1+|a_k|}$ and upper bound $\frac{1+|a_k|}{1-|a_k|}$, it gives

$$
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right)\le\frac{1}{N}\frac{1}{4\pi^2}\sum_{k=-n}^{n}\frac{1+|a_k|}{1-|a_k|}-\frac{1}{N}\sum_{k=-n}^{n}|c_k|^2,
$$

and hence,

$$
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right)\le\frac{1}{2\pi^2N}\sum_{k=-n}^{n}(1-|a_k|)^{-1}.
$$

Finally, we conclude that

$$
E\left(\|\hat{q}_{n,N}-q_n\|_2^2\right)=O\left(\frac{n}{N}\right)=O\left(\frac{1}{\sqrt{N}}\right).
$$

□

Proof of Theorem 3.3

Proof

This estimation can be obtained from (2.7) and Theorem 3.1 by setting $n = \left[N^{\frac{1}{2}}\right]$ and noting the boundedness of $\frac{\ln^3 n}{n^r}$ for any positive integer r. □

Proof of Theorem 3.4

Proof

By Chebyshev's inequality for measure-theoretic statement, we know that

$$E\left(\mu\left\{x \in [-\pi,\pi] : |\hat{q}_{n,N}(x) - q(x)| > \epsilon\right\}\right)$$
$$\leq E\left(\frac{1}{\epsilon^2}\int_{-\pi}^{\pi} |\hat{q}_{n,N}(x) - q(x)|^2 dx\right) = \frac{2\pi}{\epsilon^2} E\left(\|\hat{q}_{n,N} - q\|_2^2\right).$$

By Theorem 3.2, the estimation (3.12) follows. □

3.2. Some adjustments

Because the TM estimator $\hat{q}_{n,N}$ defined in (3.9) is designed for densities supported in the specific interval $[-\pi,\pi]$, some adjustment measurements are adopted for more general density functions supported in more general subset $\mathbb{E} \subseteq \mathbb{R}$.

Suppose that $f \in L^2(\mathbb{R})$ is a function supported in $\mathbb{E} \subseteq \mathbb{R}$ and assume that there exists a bijection

$$\mathcal{M} : t \to g(t),\ t \in \mathbb{E}$$

such that $\mathcal{M}$ maps $\mathbb{E}$ to $[-\pi,\pi]$, where g is a differentiable univariate real-valued function. The estimator for f can be designed by

$$\hat{f}_{n,N}(t) = |g'(t)| \sum_{k=-n}^{n} \hat{f}_k e_k(g(t)),\ t \in \mathbb{E} \tag{3.17}$$

with

$$\hat{f}_k = \frac{1}{2\pi N}\sum_{j=1}^{N} e_k(g(X_j)). \tag{3.18}$$

Here, $X_1, X_2, \cdots, X_N$ are the i.i.d. samples from the random variable X, which has the density f. We remark that $\hat{f}_{n,N}(t)$ is not a pure composition of $\hat{q}_{n,N}$, not only the weighted $|g'|$ but also the difference between $\hat{c}_k$ in (3.10) and $\hat{f}_k$ in (3.18). In the two important cases with $\mathbb{E} = \mathbb{R}$ and $\mathbb{E} = [a,b]$, we take

$$g(t) = 2\arctan t,\ t \in \mathbb{R}$$

and

$$g(t) = \pi\frac{2t - a - b}{b - a},\ t \in [a,b],$$

respectively.

3.3. Algorithm for adaptive choice of parameters

We have known that the continuous TM random estimator $\hat{q}_{n,N}$ can approximate the density function q with approximation order $\frac{1}{\sqrt{N}}$ when $n = O(\sqrt{N})$. But, the presupposition is that the parameter vector $\vec{a} = (a_1, \ldots, a_n)$ should be given beforehand. This issue is particularly important in numerical simulation. We hope to choose the unknown parameter vectors $\vec{a}$ from the samples $\{(x_j, y_j), j = 1, \ldots, N\}$. This can be performed under the principle of maximal coefficient existing in the algorithm theories including matching pursuit (MP), greedy algorithm (GA), and compressed sensing.

The essence of MP (GA) is to obtain a locally optimal choice at each stage with the hope of finding a global optimum. Both MP and GA are based on continuous atoms. Let us explain briefly the common idea in MP and GA. In order to represent a sparse signal $f \in L^2(\mathbb{R})$ in terms of a fixed dictionary $\Phi_I = \{\phi_m : m \in I\}$ with I some index set, where Φ_I is dense in $L^2(\mathbb{R})$, we hope to choose a subset $\Lambda \subset I$ such that the cardinality Λ of is as small as possible and $\sum_{j\in\Lambda}\langle f, \phi_j\rangle\phi_j$ can approximate f with higher precision. In the case of orthonormal basis, the maximal coefficient principle is easy to understand, which naturally requires the maximization of coefficients step by step, that means, the first coefficient $\langle f, \phi_1\rangle$ from $\sup_{\phi_j \in \Phi_I} |\langle f, \phi_j\rangle|$, the second coefficient $\langle f, \phi_2\rangle$ from $\sup_{\phi_j \in \Phi_I\setminus\{\phi_1\}} |\langle f, \phi_j\rangle|$, and so on. The compressed sensing is a discrete version of MP and GA.

Applying the strategy of maximal coefficient principle to the construction of $\vec{a}$ appearing in the system $\{e_k\}$ in terms of the samples $\{(x_j, y_j) : j = 1, \ldots, N\}$, there is a modification compared with the conventional case. Let us taste the difference. In the conventional case, the greedy task mainly emphasizes on the choice of atoms because the atom system (including parameters if available) is given in advance. In our case, only the type of atoms is known, but the parameters are unknown. Our emphasis lies in the construction of parameters in terms of samples.

Set $\triangle = \cup_{j=1}^{N} U(e^{ix_j}, \delta)$, the union of the N open neighborhoods at the points $e^{ix_j} : j = 1, \ldots, N$ on the unit circle. Our algorithm is in the succeeding text.

Algorithm for finding $\vec{a}$

Step 1. Set

$$y_j^{(1)} = y_j e^{-ix_j}, \ j = 1, \dots, N.$$

Find a_1 by

$$a_1 = \arg\sup_{a \in \mathbb{D} \setminus \triangle} \left| \sum_{j=1}^{N} y_j^{(1)} \frac{\sqrt{1-|a|^2}}{1-ae^{-ix_j}} \right|^2$$

Step 2. Set

$$y_j^{(2)} = y_j^{(1)} \frac{e^{-ix_j} - \bar{a}_1}{1 - a_1 e^{-ix_j}}, \ j = 1, \dots, N.$$

Find a_2 by

$$a_2 = \arg\sup_{a \in \mathbb{D} \setminus \triangle} \left| \sum_{j=1}^{N} y_j^{(2)} \frac{\sqrt{1-|a|^2}}{1-ae^{-ix_j}} \right|^2$$

Step k ($k = 1, 2, \dots, n$). Set

$$y_j^{(k)} = y_j^{(k-1)} \frac{e^{-ix_j} - \bar{a}_{k-1}}{1 - a_{k-1} e^{-ix_j}}, \ j = 1, \dots, N.$$

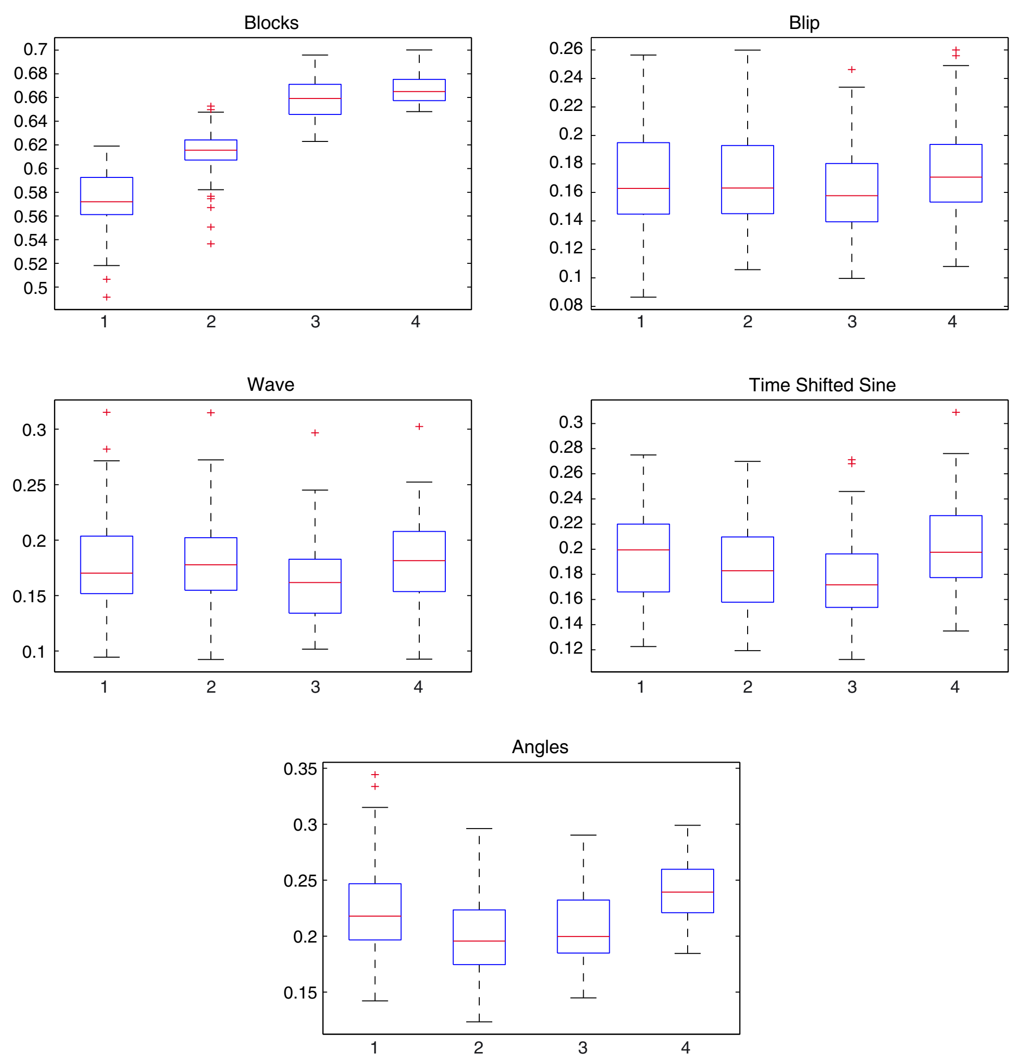

Figure 1. Boxplots of root mean square error (from left to right: Takenaka–Malmquist, wavelet, local linear, and spline).

Find a_k by

$$a_k = \arg\sup_{a\in\mathbb{D}\setminus\triangle} |\sum_{j=1}^{N} y_j^{(k)} \frac{\sqrt{1-|a|^2}}{1-ae^{-ix_j}}|^2.$$

For $k = 1, 2, \dots, n$, denote $y^{(k)} = \left(y_1^{(k)}, \dots, y_N^{(k)}\right)$ and call it the kth samples vector of Y.

About the previous algorithm, two points need to be addressed. Firstly, the sequence $y^{(1)}, \dots, y^{(n)}$ can be obtained iteratively by

$$y^{(k)} = y^{(k-1)}\text{diag}\left(\overline{\tau_{a_{k-1}}(e^{ix_1})}, \dots, \overline{\tau_{a_{k-1}}(e^{ix_N})}\right) \tag{3.19}$$

where the τ_a is the Möbius transform $\tau_a(z) = \frac{z-a}{1-\bar{a}z}$ and $a_0 = 0$. Secondly, a basic pursuit algorithm can be formulated in the succeeding text: for a given complex-valued vector $(t_1, t_2, \dots, t_N) \in \mathbb{C}^N$, find $\hat{a} \in \mathbb{D}$ by

$$\hat{a} = \arg\sup_{a\in\mathbb{D}\setminus\triangle} |\sum_{j=1}^{N} t_j \frac{\sqrt{1-|a|^2}}{1-ae^{-ix_j}}|^2. \tag{3.20}$$

Note that all steps in our algorithm depend on the problem (3.20). To solve it, define the function f by

$$f(a) = |\sum_{j=1}^{N} t_j \frac{\sqrt{1-|a|^2}}{1-ae^{-ix_j}}|^2. \tag{3.21}$$

About the existence of the maximum, we have the following theorem.

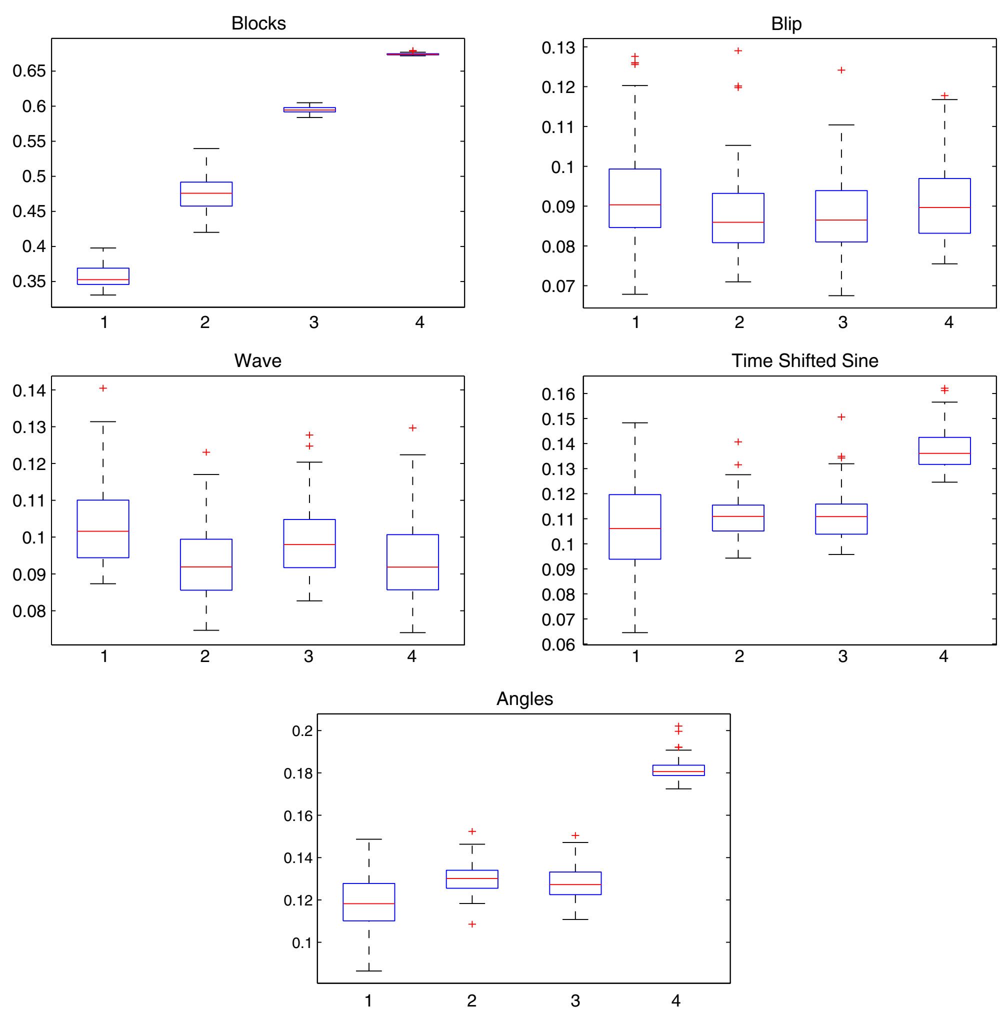

Figure 2. Boxplots of root mean square error (from left to right: Takenaka–Malmquist, wavelet, local linear, and spline).

Theorem 3.5
The function f defined in (3.21) arrives at its maximum in $\mathbb{D} \setminus \triangle$.

Proof
On one hand, as a continuous function in the closed set $\bar{\mathbb{D}} \setminus \triangle$, f has a maximum. On the other hand, $\lim_{a \to z} f(a) = 0$ for all $z \in \partial\mathbb{D} \setminus \{e^{ix_j} : j = 1, \dots, N\}$. We therefore conclude the desired result. □

4. Simulation

We consider the following regression model:

$$Y_t = f(X_t) + \mathcal{E}_t$$

with f being an unknown function. For given samples (X_k, Y_k), $k = 1, \dots, n$, we are to estimate $f(X_k)$, $k = 1, \dots, n$. The precision is measured by root mean square error (RMSE) defined by

$$RMSE = n^{-\frac{1}{2}} \sqrt{\sum_{k=1}^{n} [\hat{f}(X_k) - f(X_k)]^2}.$$

Specifically, we consider some examples that have been studied by Anestis Antoniadis and Jeremie Bigot [22].

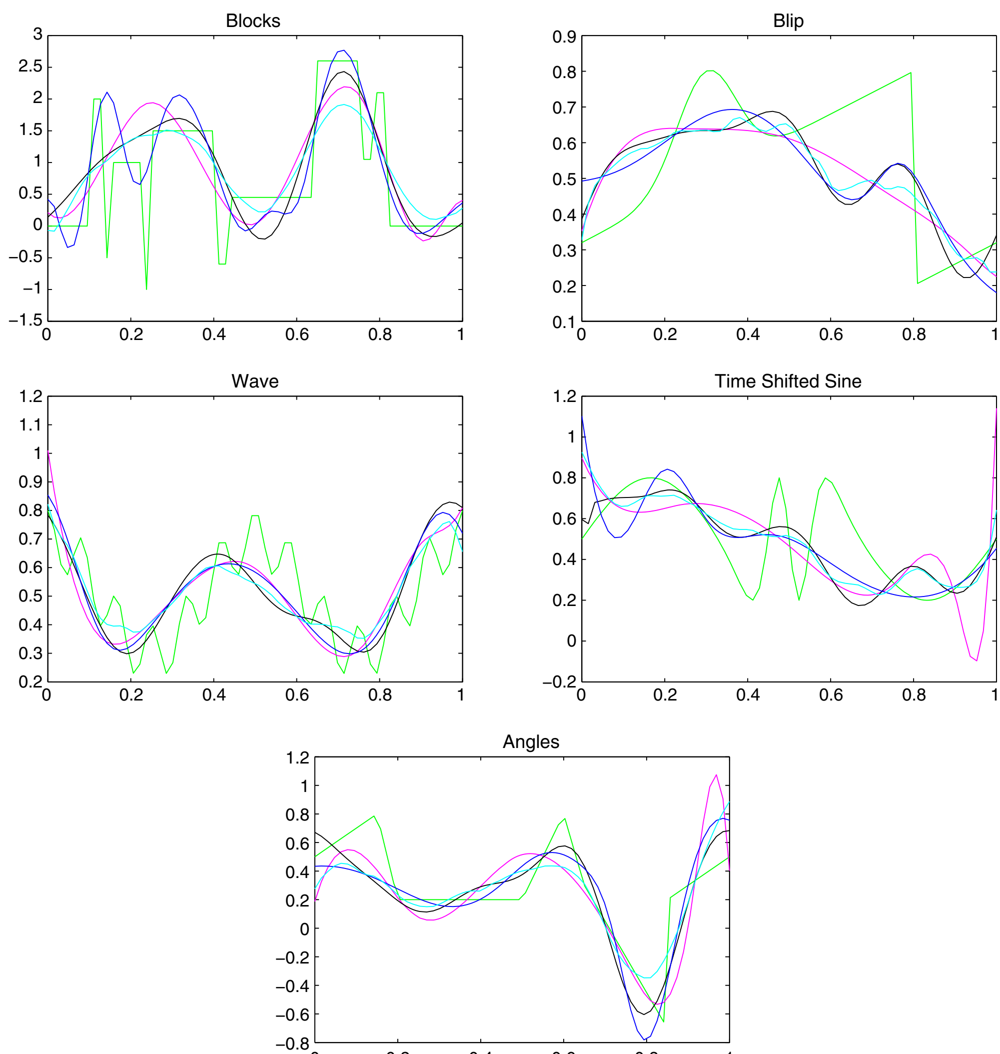

Figure 3. Original (green), Takenaka–Malmquist (blue), wavelet (black), local linear (cyan), and spline (magenta).

Blocks:

$$f_1(x) = \sum_{j=1}^{11} \frac{1}{2} h_j \left(1 + \mathrm{sgn}(x - t_j)\right)$$

with $(h_j : j = 1, \ldots, 11) = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81)$ and $(t_j : j = 1, \ldots, 11) = (4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2)$.

Blip:

$$f_2(x) = \left(0.32 + 0.6x + 0.3e^{-100(x-0.3)^2}\right) \mathbf{I}_{[0,0.8]}(x) + \left(-0.28 + 0.6x + 0.3e^{-100(x-1.3)^2}\right) \mathbf{I}_{(0.8,1]}.$$

Here, $\mathbf{I}_{(a,b)}$ stands for the indicator function of the interval (a, b).

Wave:

$$f_3(x) = 0.5 + 0.2\cos(4\pi x) + 0.1\cos(24\pi x).$$

Time-shifted sine:

$$f_4(x) = 0.3\sin\{3\pi[g(g(g(g(x)))) + x]\} + 0.5,$$

where $g(x) = (1 - \cos(\pi x))/2$.

Angles:

$$\begin{aligned} f_5(x) = {} & (2x + 0.5)\mathbf{I}_{[0,0.15]}(x) + (-12(x - 0.15) + 0.8)\mathbf{I}_{(0.15,0.2]}(x) + \\ & 0.2\mathbf{I}_{(0.2,0.5]}(x) + (6(x - 0.5) + 0.2)\mathbf{I}_{(0.5,0.6]}(x) + \\ & (-10(x - 0.6) + 0.8)\mathbf{I}_{(0.6,0.65]}(x) + (-5(x - 0.65) + 0.3)\mathbf{I}_{(0.65,0.85]}(x) + \\ & (2(x - 0.85) + 0.2)\mathbf{I}_{(0.85,1]}(x). \end{aligned}$$

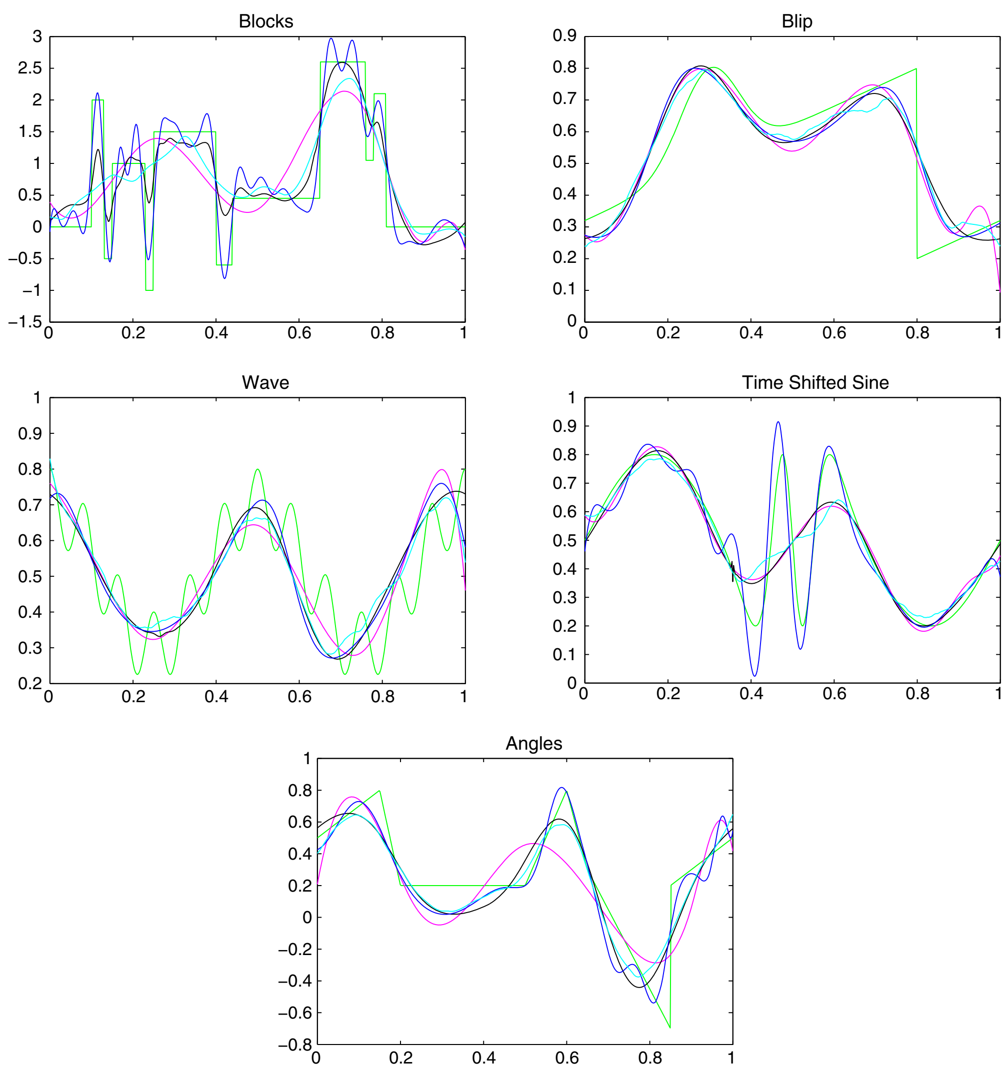

Figure 4. Original (green), Takenaka–Malmquist (blue), wavelet (black), local linear (cyan), and spline (magenta).

Table I. Here, A > B means method A performs better than B; A ≈ B means methods A and B have similar performance (the difference of root mean square errors of A and B is at 10^{-3} level).

Block	TM > wavelet > local linear > spline
Blip	TM ≈ wavelet ≈ local linear > spline
Wave	TM ≈ wavelet ≈ local linear > spline
Time-shifted sine	Local linear > wavelet > TM ≈ spline
Angles	Wavelet ≈ local linear > TM > spline

TM, Takenaka–Malmquist.

Table II. Here, A > B means method A performs better than B; A ≈ B means methods A and B have similar performance (the difference of root mean square errors of A and B is at 10^{-3} level).

Block	TM > wavelet > local linear > spline
Blip	TM ≈ wavelet ≈ local linear ≈ spline
Wave	TM ≈ wavelet ≈ local linear ≈ spline
Time-shifted sine	TM ≈ Wavelet ≈ local linear > spline
Angels	TM > wavelet ≈ local linear > spline

TM, Takenaka–Malmquist.

For simulation, we suppose that X is an equally spaced on $[0, 1]$, $\mathcal{E}_t \sim \mathcal{N}(0, 0.5)$. For each of the previous examples, we adopt the TM method, the wavelet method, the local linear method, and the spline method with sample size $N = 64, 512$, respectively. Set the repeat time to 100. We denote the TM method by 'TM', the wavelet method by 'W', the local linear method by 'LL', and the spline method by 'S'. To compare the performance, we need to calculate the RMSE for each estimation.

For the TM method with $N = 64$, we choose n to be $8, 3, 3, 4, 4$ with respect to the order of the previous examples of density function. Similarly, we choose n to be $20, 3, 4, 10, 10$ when $N = 512$. We first show boxplots of RMSE (from left to right: TM, W, LL, and S) of each subfigure in Figures 1 and 2. Figures 3 and 4 show the original function of each example (green curve) and results with median performance from TM (blue curve), W (black curve), LL (cyan curve), and S (magenta curve).

In the succeeding text is our simulation results with sample size $N = 64$.

From previous illustrations, we see that each method has its pluses and minuses. We conclude it in Table I.

In the case of 'Blocks', the performance of the TM method is much better than those of the other methods. In the cases of 'Blip' and 'Wave', the performances of the TM method, the wavelet method, and the local linear method are similar, and the performance of the spline method is the worst. For the cases of 'Time Shifted Sine' and 'Angles', the performances of the TM method and the spline method are similar and both of them are a little bit worse than those of the wavelet and the local linear methods.

In the succeeding text is our simulation results with sample size $N = 512$.

Similarly, we conclude it in Table II.

In the case of 'Blocks', the performance of the TM method is still the best. For the cases of 'Blip' and 'Wave', all the tested methods behave well. Unlike the results with $N = 64$, the performances of the TM method are better than those of the other methods for the cases of 'Time Shifted Sine' and 'Angles'.

It is noticeable that for all the tested methods, the experimental results with $N = 512$ are obviously better than those for $N = 64$. Comparatively, however, the improvement for the TM method is extraordinary. We note that the TM estimator performance in the experiments may be improved because the way of selecting the involved parameters can be further developed.

Appendix A: The upper half plane case

There is a parallel theory of estimators based on the TM systems on the upper half plane. For the purpose of completeness of our theory, this section investigates this topic. In the aspect of numerical implementation, the TM estimator in the case of the upper half plane has no superior compared with the estimator in the case of the unit disk. Two reasons cause a bad approximation accuracy when the truncation strategy is adopted. One is the infinity of the support of the estimator on the upper half plane and the other is the wide scope of the parameters' distribution.

For the upper half plane case, we say $f \in H^p(\mathbb{C}^+)$, $0 < p < \infty$, if f is analytic on $\mathbb{C}^+$ and

$$\sup_{y>0} \int_{-\infty}^{\infty} |f(x+iy)|^p dx = \|f\|^p_{H^p(\mathbb{C}^+)} < \infty.$$

When $p = \infty$, we write $f \in H^\infty(\mathbb{C}^+)$ for the bounded analytic functions on $\mathbb{C}^+$, and we give $H^\infty(\mathbb{C}^+)$ the norm $\|f\|_{H^\infty(\mathbb{C}^+)} = \sup_{w\in\mathbb{C}^+} |f(w)|$. The relation between $f \in H^p(\mathbb{C}^+)$ and their non-tangential boundary limits on $\mathbb{R}$ is the same as the unit disk case. $H^2(\mathbb{C}^+)$ is equipped with the inner product

$$\langle f, g\rangle_{\mathbb{C}^+} = \int_{-\infty}^{\infty} f(t)\overline{g(t)}dt, \ \ f, g \in H^2(\mathbb{C}^+).$$

For a given parameter sequence $\{\lambda_n\}_{n=0}^{\infty} \subset \mathbb{C}^+$, the corresponding TM system $\{\beta_n\}_{n=1}^{\infty}$ in the upper half plane case is

$$\beta_n(w) = \frac{\sqrt{\frac{1}{\pi} Im\{_{-n}\}}}{w - \overline{\lambda_n}} B_{n,\mathbb{C}^+}(\omega), \ \ w \in \mathbb{C}^+, \ \ n \in \mathbb{Z}_+, \tag{A1}$$

where the Blaschke product on the upper-half plane is defined by

$$B_{n,\mathbb{C}^+}(\omega) = \prod_{j=0}^{n-1} \frac{w - \lambda_j}{w - \overline{\lambda_j}}.$$

Under the condition $\sum_{k=0}^{\infty} \frac{\sqrt{Im(_{-k})}}{1 + |\lambda_k|^2} = \infty$, $\{\beta_n\}$ is an orthonormal base of the Hilbert space $H^2(C^+)$.

A.1. Approximation order of partial sum in the case $\mathbb{C}^+$

For a density function $f \in L^2(\mathbb{R})$, we can expand it in terms of the orthonormal basis $\{\beta_k\}$ as

$$f(x) = \sum_{k\in\mathbb{Z}} d_k\beta_k(x), \ \ x \in \mathbb{R}$$

with

$$d_k = \langle f, \beta_k\rangle = \int_{-\infty}^{\infty} f(x)\beta_k(x)dx.$$

The partial sum operator is defined by

$$S_{n,\mathbb{C}^+} f(x) = \sum_{k=-n}^{n} d_k\beta_k(x), \ x \in \mathbb{R}. \tag{A2}$$

It is necessary to investigate the order of $S_{n,\mathbb{C}^+} f$ converging to f in different measures. As we know, no related literatures are available about this topic. Our approach is to use the Caley transform.

We remark that the Caley transform

$$w = \kappa(z) := i\frac{1-z}{1+z}$$

is a conformal mapping from $\mathbb{D}$ to $\mathbb{C}^+$ and its inverse mapping is

$$z = \frac{i - w}{w + i}: \ \ \mathbb{C}^+ \to \mathbb{D}.$$

An important fact is that Caley transform is an isomorphism between $H^\infty(\mathbb{D})$ and $H^\infty(\mathbb{C}^+)$. Moreover, κ is an isometry from $H^\infty(\mathbb{D})$ to $H^\infty(\mathbb{C}^+)$ that maps one of the two type of TM systems to the other, that is,

$$\beta_n(\kappa(z)) = \frac{1}{2\sqrt{\pi}} e^{i\alpha}(1 + z)e_n(z), \tag{A3}$$

where $e^{i\alpha} = \frac{1+\overline{a_n}}{|1+a_n|}(-i)\prod_{j=0}^{n-1}\frac{-(1+\overline{a_j})}{1+a_j}$.

Similarly to the unit disk case, we need to investigate the one-sided partial sum of the half system (A1)

$$L_{n,\mathbb{C}^+} f(w) = \sum_{k=0}^{n} \langle f, \beta_k\rangle\beta_k(w). \tag{A4}$$

By combing the transform $t = i\frac{1-e^{i\theta}}{1+e^{i\theta}}$ with the identity (A3), a direct calculation gives

$$\begin{aligned}\langle f, \beta_k\rangle &= \int_{-\infty}^{\infty} f(t)\overline{\beta_k(t)}dt \\ &= \int_{-\pi}^{\pi} f\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right)\overline{\beta_k\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right)}d\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right) \\ &= \int_{-\pi}^{\pi} f\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right)\overline{\frac{1}{2\sqrt{\pi}}e^{i\alpha}(1+e^{i\theta})e_k(e^{i\theta})}\frac{2e^{i\theta}}{(1+e^{i\theta})^2}d\theta \\ &= \frac{1}{\sqrt{\pi}}e^{-i\alpha}\int_{-\pi}^{\pi}\frac{f\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right)}{1+e^{i\theta}}\overline{e_k(e^{i\theta})}d\theta \\ &= 2\sqrt{\pi}e^{-i\alpha}\langle g, e_k\rangle\end{aligned}$$

where $g(z) = \frac{f(\kappa(z))}{1+z}$. Then

$$\begin{aligned}L_{n,\mathbb{C}^+}f(w) &= \sum_{k=0}^{n} 2\sqrt{\pi}e^{-i\alpha}\langle g, e_k\rangle\frac{1}{2\sqrt{\pi}}e^{i\alpha}(1+z)e_n(z) \\ &= (1+z)L_{n,\mathbb{D}}g(z).\end{aligned}$$

Hence,

$$L_{n,\mathbb{C}^+}f(w) - f(w) = (1+z)\left(L_{n,\mathbb{D}}g(z) - g(z)\right). \tag{A5}$$

We therefore establish the version in the upper-half plane of Pycak's result.

Proposition A.1
Suppose that $f \in H^{\infty}(\mathbb{C}^+)$, $\frac{f(k(z))}{1+z} \in \mathbb{B}_r\mathbb{H}_1$ for some nonnegative integer r, and $\omega\left(\frac{d}{dz}\frac{f(k(z))}{1+z}, t\right) = O(t^{\mu})$ for some $\mu \in (0,1)$. Then there exists some suitable parameter vector $\lambda = (\lambda_0, \dots, \lambda_{n-1})$ such that

$$\max_{w\in\overline{\mathbb{C}}}|L_{n,\mathbb{C}^+}f(w) - f(w)| \le C\frac{\ln^3 n}{n^{r+1}}, \tag{A6}$$

where C is only dependent on μ.

Proof
This is a consequence of (2.5) and (A5). □

A parallel result of Lemma 2.2 can be established as follows.

Lemma A.2
Suppose that $\lambda_0 = i$. Then for any $f \in L^2(\mathbb{R})$, it holds

$$S_{n,\mathbb{C}^+}f(x) - f(x) = 2\mathrm{Re}\left[L_{n,\mathbb{C}^+}f^+(x) - f^+(x)\right]. \tag{A7}$$

Proof
The proof is the same as that of 2.2 by changing e_k to β_k. □

A pointwise approximation order for $S_{n,\mathbb{C}^+}$ is therefore concluded below.

Theorem A.3
Suppose that $\lambda_0 = i, f^+ \in H^{\infty}(\mathbb{C}^+)$, $\frac{f^+(k(z))}{1+z} \in \mathbb{B}_r\mathbb{H}_1$ for some nonnegative integer r and $\omega\left(\frac{d}{dz}\frac{f^+(k(z))}{1+z}, t\right) = O(t^{\mu})$ for some $\mu \in (0,1)$. Then there exists some suitable parameter vector $\lambda = (\lambda_0, \dots, \lambda_{n-1}) \in \mathbb{C}^n$ such that

$$|S_{n,\mathbb{C}^+}f(x) - f(x)| \le C\frac{\ln^3 n}{n^{r+1}} \tag{A8}$$

for any $x \in \mathbb{R}$, where the constant C is only dependent on μ.

To discuss the approximation order in L^2-norm sense, we need to estimate the decaying rate of the Fourier coefficients $\langle f, \beta_k\rangle$ for functions in $L^2(\mathbb{R})$.

Lemma A.4
Assume that $\lambda_0 = i$ and $f \in L^2(\mathbb{R})$ is Kth continuously differentiable function with $K \ge 3$, then the decaying order of the generalized Fourier coefficients $\langle f, \beta_n\rangle$ is $O\left(\frac{1}{n^{K-1}}\right)$.

Proof
Noting that the operator $\mathcal{T}: H^2(\mathbb{C}^+) \to H^2(\mathbb{D})$ defined by

$$\mathcal{T}f(z) = \frac{2i\pi^{\frac{1}{2}} f\left(i\frac{1-z}{1+z}\right)}{(1+z)}, \quad z \in \mathbb{D} \tag{A9}$$

is an isomorphism and an isometry between $H^2(\mathbb{C}^+)$ and $H^2(\mathbb{D})$, we confirm that

$$\frac{f\left(i\frac{1-e^{i\cdot}}{1+e^{i\cdot}}\right)}{1+e^{i\cdot}} \in L^2(-\pi,\pi).$$

Recalling that (see the calculation between (A4) and (A5))

$$\langle f, \beta_k\rangle = \int_{-\infty}^{\infty} f(t)\overline{\beta_k(t)}dt = \int_{-\pi}^{\pi} \frac{f\left(i\frac{1-e^{i\theta}}{1+e^{i\theta}}\right)}{1+e^{i\theta}}\overline{e_k(e^{i\theta})}d\theta,$$

noting that $\frac{f\left(i\frac{1-e^{i\cdot}}{1+e^{i\cdot}}\right)}{1+e^{i\cdot}}$ is 2π periodic and Kth continuously differential and recalling Proposition 2.4, we obtain that $|\langle f, \beta_n\rangle| = O\left(\frac{1}{n^{K-1}}\right)$. □

The following theorem indicates that the L^2-approximation order of $S_{n,\mathbb{C}^+}$ is $O(n^{-1})$.

Theorem A.5
Assume that $\lambda_0 = i$. Suppose that $f \in L^2(\mathbb{R})$ is third continuously differentiable. Then

$$\|S_{n,\mathbb{C}^+}f - f\|_2^2 = O(n^{-1}). \tag{A10}$$

Proof
This is a direct consequence from Plancherel theorem and Lemma A.4. □

A.2. Estimator in the case $\mathbb{C}^+$

For i.i.d. samples $X_1, X_2, \dots, X_N$ from the random variable X with density p supported in $\mathbb{R}$, define the TM system estimator by

$$\hat{p}_{n,N}(x) := \sum_{k=-n}^{n} \hat{d}_k\beta_k(x),\ x \in \mathbb{R} \tag{A11}$$

where the Fourier coefficients $\hat{d}_k$ are random variable

$$\hat{d}_k := \frac{1}{N}\sum_{j=1}^{N}\beta_k(X_j). \tag{A12}$$

The following theorem shows that $\hat{p}_{n,N}$ is the asymptotically unbiased estimator of p.

Theorem A.6
Assume that $\lambda_0 = i$ and $\sum_{k=-n}^{n}(\mathrm{Im}(\lambda_k))^{-2} = O(n)$. Suppose that the density function p satisfies the conditions in Theorem A.3 with $r > 0$. Then for TM estimator $\hat{p}_{n,N}$, as $n, N \to \infty$ and $\frac{n}{N} \to 0$, we have

(i) $\lim_{n\to\infty} \hat{p}_{n,N}(x) = p(x), \ \ \forall x \in \mathbb{R}$; and
(ii) $Var(\hat{p}_{n,N}(x)) = O\left(\frac{n}{N}\right) = o(1)$.

Proof
(i) By (A2) and (A11), we obtain that $E(\hat{d}_k) = d_k$ and correspondingly $E(\hat{p}_{n,N}) = S_{n,\mathbb{C}^+}p$. By Theorem A.3, we have that, for $x \in \mathbb{R}$,

$$\lim_{n\to\infty} E\left(\hat{p}_{n,N}(x)\right) = \lim_{n\to\infty} S_{n,\mathbb{C}^+}p(x) = p(x).$$

(ii) To estimate $Var(\hat{p}_{n,N}(x))$, we first need to investigate $E(\hat{d}_k - d_k)^2$. A direct calculation gives rise to

$$\begin{aligned}
E(\hat{d}_k - d_k)^2 &= E\left(\hat{d}_k^2\right) - d_k^2 \\
&= E\left(\frac{1}{N}\sum_{j=1}^{N}\beta_k(X_j)\right)^2 - d_k^2 \\
&= \frac{1}{N^2}\sum_{j=1}^{N}E\left[\beta_k^2(X_j)\right] + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{\ell\neq j}E\left[\beta_k(X_j)\beta_k(X_\ell)\right] - d_k^2 \\
&= \frac{1}{N^2}\sum_{j=1}^{N}E\left[\beta_k^2(X_j)\right] + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{\ell\neq j}E[\beta_k(X_j)]E[\beta_k(X_\ell)] - d_k^2 \\
&= \frac{1}{N^2}\sum_{j=1}^{N}E\left[\beta_k^2(X_j)\right] + \frac{1}{N^2}\sum_{j=1}^{N}\sum_{\ell\neq j}d_k^2 - d_k^2 \\
&= \frac{1}{N^2}\sum_{j=1}^{N}E\left[\beta_k^2(X_j)\right] + \frac{1}{N^2}(N^2-N)d_k^2 - d_k^2 \\
&\leq \frac{1}{N^2}\sum_{j=1}^{N}E\left[\beta_k^2(X_j)\right].
\end{aligned}$$

Combining this with the inequality

$$|\beta_k(x)| \leq \sqrt{\frac{1}{\pi}}\,(\text{Im}(\lambda_k))^{-\frac{1}{2}}, \tag{A13}$$

it follows

$$E(\hat{d}_k - d_k)^2 \leq \frac{1}{N^2}\sum_{j=1}^{N}\int_{-\infty}^{\infty}p(x_j)\beta_k^2(x_j)dx_j \leq \frac{1}{\pi}\frac{1}{N}\frac{1}{\text{Im}(\lambda_k)}. \tag{A14}$$

For simplicity, we denote $p_n = S_{n,\mathbb{C}^+}p$. Noting that the variance of the estimator $\hat{p}_{n,N}$ can be written as

$$\begin{aligned}
Var(\hat{p}_{n,N}(x)) &= E[\hat{p}_{n,N}(x) - E(\hat{p}_{n,N}(x))]^2 \\
&= E[\hat{p}_{n,N}(x) - p_n(x)]^2 \\
&= E[\sum_{k=-n}^{n}(\hat{d}_k - d_k)\beta_k(x)]^2 \\
&= \sum_{k=-n}^{n}\beta_k^2(x)E(\hat{d}_k - d_k)^2 + \sum_{k=-n}^{n}\sum_{\ell\neq k}^{n}\beta_k\beta_\ell(x)E[(\hat{d}_k - d_k)(\hat{d}_\ell - d_\ell)] \\
&= \sum_{k=-n}^{n}\beta_k^2(x)E(\hat{d}_k - d_k)^2,
\end{aligned}$$

taking into account (A13) and (A14), we obtain the estimation for the variance of $\hat{q}_{n,N}$

$$V(\hat{p}_{n,N}(x)) \leq \frac{1}{N}\frac{1}{\pi^2}\sum_{k=-n}^{n}(\text{Im}(\lambda_k))^{-2}. \tag{A15}$$

Finally, (ii) can be concluded from the increasing requirement of the parameter sequence λ. □

Remark
The condition $\sum_{k=-n}^{n}(\text{Im}(\lambda_k))^{-2} = O(n)$ is automatically met in one parameter case.

About the mean square error of the estimator $\hat{p}_{n,N}$, we have the following corollary.

Corollary A.7
The mean square error of $\hat{p}_{n,N}$ satisfies

$$E\left(\hat{p}_{n,N}(x) - p(x)\right)^2 = O\left(\frac{1}{\sqrt{N}}\right), \quad n = \sqrt{N}.$$

Proof

Recall that the mean square error of $\hat{p}_{n,N}$ is dominated by the variance of the estimator $\hat{p}_{n,N}$ and the approximation rate of the partial sum $S_{n,\mathbb{C}}p$, that is,

$$E[\hat{p}_{n,N}(x) - p(x)]^2 = V(\hat{p}_{n,N}(x)) + |S_{n,\mathbb{C}}p(x) - p(x)|^2.$$

Combining this with (A8) and (A15), we conclude that

$$\begin{aligned} E\left(\hat{p}_{n,N}(x) - p(x)\right)^2 &\le C\frac{\ln n^3}{n^{r+1}} + \frac{1}{N}\frac{1}{\pi^2}\sum_{k=-n}^{n}(\text{Im}(\lambda_k))^{-2} \\ &\le C\left(\frac{1}{n} + \frac{n}{N}\right) = O\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

□

Regarding the approximation order of the estimator $\hat{p}_{n,\mathbb{C}^+}$ in the sense of L^2-norm, we have the following theorem for the quantity $E\left(\|\hat{p}_{n,N} - p\|_2^2\right)$.

Theorem A.8

Suppose that $\lambda_0 = i$ and $\sum_{k=-n}^{n}(\text{Im}(-_n))^{-1} = \text{O(n)}$. If the density function $p \in L^2(\mathbb{R})$ is third continuously differentiable, then

$$E\left(\|\hat{p}_{n,N} - p\|_2^2\right) = O\left(\frac{1}{\sqrt{N}}\right) \quad \text{with} \quad n = \sqrt{N}. \tag{A16}$$

Proof

By Theorem A.5 and the triangle inequality, we are left to show that

$$E\left(\|\hat{p}_{n,N} - p_n\|_2^2\right) = O\left(\frac{1}{\sqrt{N}}\right). \tag{A17}$$

The proof of (A17) is the same as the periodic case, just by changing e_k to β_k and some modification of constant multiplication. For convenience of readers, we offer the outline in the succeeding text. Firstly, using the same technique as (3.14), we obtain

$$E\left(\|\hat{p}_{n,N} - p_n\|\right) = \sum_{k=-n}^{n}\left(E(|\hat{d}_k|^2) - |d_k|^2\right). \tag{A18}$$

Secondly, a similar approach as (3.15) leads to

$$E\left(|\hat{d}_k|^2\right) - |d_k|^2 = \frac{1}{N}E\left(|\beta_k(X)|^2\right) - \frac{1}{N}|d_k|^2. \tag{A19}$$

Thirdly, combing (A18) and (A19), we have

$$E\left(\|\hat{p}_{n,N} - p_n\|_2^2\right) = \sum_{k=-n}^{n}\left(\frac{1}{N}E\left(|\beta_k(X)|^2\right) - \frac{1}{N}|d_k|^2\right). \tag{A20}$$

Fourthly, direct calculating gives the estimation of $E\left(|\beta_k(X)|^2\right)$

$$\begin{aligned} E\left(|\beta_k(X)|^2\right) &= \int_{\mathbb{R}} p(x)|\beta_k(x)|^2 dx = \int_{\mathbb{R}} p(x)\frac{1}{\pi}\frac{\text{Im}(-_n)}{|x - \bar{\lambda}_n|^2}dx \\ &\le \frac{1}{\pi}(\text{Im}(-_n))^{-1}. \end{aligned}$$

Finally, combing this with (A20) and the assumption on λ_n, it gives

$$\begin{aligned} E\left(\|\hat{p}_{n,N} - p_n\|_2^2\right) &\le \frac{1}{N}\sum_{k=-n}^{n}\frac{1}{\pi}(\text{Im}(\lambda_n))^{-1} - \frac{1}{N}\sum_{k=-n}^{n}|d_k|^2 \\ &= \frac{O(n)}{N} = O\left(\frac{1}{\sqrt{N}}\right). \end{aligned}$$

The proof of this theorem is completed. □

Regrading the quantity $E\left(\mu\left\{x \in \mathbb{R} : |\hat{p}_{n,N}(x) - p(x)| > \epsilon\right\}\right)$, we have the following corollary.

Corollary A.9

Suppose that $\lambda_0 = i$ and $\sum_{k=-n}^{n} (\text{Im}(\lambda_n))^{-1} = O(n)$. If the density function $p \in L^2(\mathbb{R})$ is third continuously differentiable, then

$$E\left(\mu\left\{x \in \mathbb{R} : |\hat{p}_{n,N}(x) - p(x)| > \epsilon\right\}\right) = O\left(\frac{1}{\sqrt{N}}\right) \quad \text{with} \quad n = \sqrt{N}. \tag{A21}$$

Proof

By Chebyshev's inequality for measure-theoretic statement, we know that

$$\begin{aligned} E\left(\mu\left\{x \in \mathbb{R} : |\hat{p}_{n,N}(x) - p(x)| > \epsilon\right\}\right) &\le E\left(\frac{1}{\epsilon^2}\int_{\{x\in\mathbb{R}:|\hat{p}_{n,N}(x)-p(x)|>\epsilon\}} |\hat{p}_{n,N}(x) - p(x)|^2 dx\right) \\ &\le \frac{1}{\epsilon^2} E\left(\|\hat{p}_{n,N} - p\|_2^2\right). \end{aligned}$$

By Theorem A.8, we can conclude (A21). □

Acknowledgements

This work was supported in part by NSFC grant 11571083, Natural Science Foundation of Guangdong Province (Grant no. S2011010004986), Macao Science and Technology Fund (Grant no. FDCT/056/2010/A3), Research Grant of University of Macau (Grant no. UL017/08-Y3/MAT/QT01/FST), Multi-Year Research Grant (MYRG) MYRG115(Y1-L4)-FST13-QT, Multi-Year Research Grant (MYRG) MYRG116(Y1-L3)-FST13-QT, and Macao Government FDCT 098/2012/A3.

References

1. Rosenblatt M. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 1956; **27**(3):832–837.
2. Parzen E. On estimation of a probability density and mode. *Annals of Mathematical Statistics* 1962; **35**:1065–1076.
3. Susarla V, Kumar S. Estimation of partial derivatives of the average of densities belonging to a family of densities. *Journal of the Australian Mathematical Society* 1975; **20**:230–241.
4. Shwartz SC. Estimation of a probability density by an orthogonal series. *Annals of Mathematical Statistics* 1967; **38**:1261–1265.
5. Walter G. Properties of Hermite series estimation of probabililty density. *The Annals of Statistics* 1977; **5**(6):1258–1264.
6. Blum J, Sursala V. Estimating the parameters of a mixing distribution function. *Annals of Probability* 1977; **5**:200–209.
7. Walter GG, Blum J. Probability density estimation using delta sequences. *Annals of Statistics* 1979; **7**:328–340.
8. Hjort NL, Jones MC. Locally parametric nonparametric density estimation. *Annals of Statistics* 1996; **24**(4):1619–1647.
9. Epstein B. *Orthogonal Families of Analytic Functions*. Macmillan: New York, 1965.
10. Takenaka S. On the orthogonal functions and a new formula of interpolation. *Japanese Journal of Mathematics* 1925; **II**:129–145.
11. Walsh JL. *Interpolation and Approximation by Rational Functions in the Complex Domain* (4th ed). American Mathematical Society: Providence, R.I., 1965.
12. Qian T, Wang YB. Adaptive decomposition into basic signals of non-negative instantaneous frequencies – a variation and realization of greedy algorithm. *Advances in Computational Mathematics* 2011; **34**(3):279–293.
13. Akcay H, Ninness B. Orthonormal basis functions for modelling continuous-time systems. *Signal Processing* 1999; **77**:261–274.
14. Mendel J. A unified approach to the synthesis of orthonormal exponential functions useful in systems analysis. *IEEE Transactions on Systems Science and Cybernetics* 1966; **2**:54–62.
15. Wahlberg B, Mäkilä P. On approximation of stable linear dynamical systems using Laguerre and Kautz functions. *Automatica* 1996; **32**:693–708.
16. Edwards RE. *Fourier Series, A Modern Introduction*, Vol. 2, GTM 64. Springer-Verlag: New York, Inc., 1979.
17. Qian T, Chen Q, Tan L. The rational orthogonal systems are Schauder bases. *Complex Variables and Elliptic Equations* 2013; **59**:841–846. DOI: 10.1080/17476933.2013.787532.
18. Chui CK, Shen XC. Degree of rational approximation in digital filter. *Lecture Notes in Mathematics* 1105. Springer-Verlag: Berlin Heidelberg N. Y. Tokyo, 1984; 189–209.
19. Shen XC. *Approximation of Complex Functions*. Science Press of China: Beijing, 1992.
20. Stein E, Shakarchi R. *Fourier Analysis: An Introduction*. Princeton University Press: New Jersey, 2002.
21. Bultheel A, Carrette P. Algebraic and spectral properties of general Toeplitz matrices. *SIAM Journal on Control and Optimization* 2003; **41**:1413–1439.
22. Antoniadis A, Bigot J. Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software* 2001; **6**(6):1–83.