Electronic Journal of Statistics
Vol. 5 (2011) 53–82
ISSN: 1935-7524
DOI: 10.1214/11-EJS597

Multichannel boxcar deconvolution with growing number of channels*

Marianna Pensky†

Department of Mathematics
University of Central Florida
Orlando FL 32816-1353, USA
e-mail: mpensky@pegasus.cc.ucf.edu

and

Theofanis Sapatinas

Department of Mathematics and Statistics
University of Cyprus
P.O. Box 20537
Nicosia CY 1678, Cyprus
e-mail: fanis@ucy.ac.cy

Abstract: We consider the problem of estimating the unknown response function in the multichannel deconvolution model with a boxcar-like kernel which is of particular interest in signal processing. It is known that, when the number of channels is finite, the precision of reconstruction of the response function increases as the number of channels M grow (even when the total number of observations n for all channels M remains constant) and this requires that the parameter of the channels form a Badly Approximable M-tuple.

Recent advances in data collection and recording techniques made it of urgent interest to study the case when the number of channels $M = M_n$ grow with the total number of observations n. However, in real-life situations, the number of channels $M = M_n$ usually refers to the number of physical devices and, consequently, may grow to infinity only at a slow rate as $n \to \infty$. Unfortunately, existing theoretical results cannot be blindly applied to accommodate the case when $M = M_n \to \infty$ as $n \to \infty$. This is due to the fact that, to the best of our knowledge, so far no one have studied the construction of a Badly Approximable M-tuple of a growing length on a specified interval, of a non-asymptotic length, of the real line, as M is growing. Therefore, this generalization requires non-trivial results in number theory.

When $M = M_n$ grows slowly as n increases, we develop a procedure for the construction of a Badly Approximable M-tuple on a specified interval, of a non-asymptotic length, together with a lower bound associated with this M-tuple, which explicitly shows its dependence on M as M is growing. This result is further used for the evaluation of the L^2-risk of the suggested adaptive wavelet thresholding estimator of the unknown response function and, furthermore, for the choice of the optimal number of channels M which minimizes the L^2-risk.

*An earlier version of this paper was circulated as "Diophantine approximation and the problem of estimation of the initial speed of a wave on a finite interval in a stochastic setting, Technical Report TR-07-2009, Department of Mathematics and Statistics, University of Cyprus, Cyprus".

†Supported in part by National Science Foundation (NSF), grant DMS-0652524.

AMS 2000 subject classifications: Primary 62G05, 11K60; secondary 62G08, 35L05.
Keywords and phrases: Adaptivity, badly approximable tuples, Besov spaces, Diophantine approximation, functional deconvolution, Fourier analysis, Meyer wavelets, nonparametric estimation, wavelet analysis.

Received August 2010.

Contents

1. Introduction

We consider the estimation problem of the unknown response function $f(\cdot) \in L^2(T)$ from observations $y(u_l, t_i)$, $l = 1, 2, \ldots, M$, $i = 1, 2, \ldots, N$, where

$$y(u_l, t_i) = \int_T g(u_l, t_i - x) f(x)\, dx + \varepsilon_{li}, \quad u_l \in U, \;\; t_i = (i-1)/N, \tag{1.1}$$

where $U = [a, b]$, $0 < a < b < \infty$, $T = [0, 1]$ and ε_{li} are standard Gaussian random variables, independent for different l and i. We shall be interested in the case when the blurring (or kernel) function $g(\cdot, \cdot)$ is the, so called, boxcar-like kernel, i.e.,

$$g(u, t) = \frac{\gamma(u)}{2}\, \mathbb{I}(|t| < u),$$

where $\gamma(\cdot)$ is some positive function such that

$$\gamma_1 \le \gamma(u) \le \gamma_2, \quad u \in U, \tag{1.2}$$

for some $0 < \gamma_1 \le \gamma_2 < \infty$. (Obviously, this is true if $\gamma(\cdot)$ is a continuous function.) Hence, (1.1) is of the form

$$y(u_l, t_i) = \frac{\gamma(u_l)}{2} \int_0^1 \mathbb{I}(|t_i - x| < u_l) f(x)\, dx + \varepsilon_{li}, \quad u_l \in U, \;\; t_i = (i-1)/N, \tag{1.3}$$

for $l = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$.

In signal processing, this model is referred to as a multichannel deconvolution model, where M is the number of channels and N is the number of observations per channel, so that $n = MN$ is the total number of observations. We assume that the measurements t_i in each channel are equispaced but the observer can choose the number of channels M and the points u_l, $l = 1, \dots, M$, in (1.1) prior to the experiment as a part of experimental design. In order to be able to access convergence rates depending on the number of channels M and the choice of points u_l, $l = 1, 2, \dots, M$, we shall further assume that the total number of observations n is fixed and very large (i.e., $n \to \infty$). The objective is to choose M and u_l, $l = 1, 2, \dots, M$, which ensure the construction of an estimator of the response function f with the highest possible convergence rates in terms of n.

Note that standard deconvolution (i.e., when $a = b$) with the boxcar kernel (i.e., when $\gamma(u) = 1/u$, for some *fixed* $u > 0$) is a common model in many areas of signal and image processing which include, for instance, LIDAR remote sensing and reconstruction of blurred images. LIDAR is a lazer device which emits pulses, reflections of which are gathered by a telescope aligned with the lazer, see, e.g., [24, 13]. The return signal is used to determine distance and the position of the reflecting material. However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. This loss of precision can be corrected by deconvolution. In practice, measured LIDAR signals are corrupted by additional noise which renders direct deconvolution impossible. If $M \geq 2$, then we talk about a multichannel deconvolution model with blurring functions $g_l(t) = g(u_l, t)$.

Although standard deconvolution models are traditionally solved using the Fourier transform or the Fourier series, if the corresponding blurring function $g(\cdot)$ is a boxcar-like kernel, implementation of the standard Fourier series based technique is impossible. This happens when the Fourier transform of $g(\cdot)$ has real zeros, e.g., when $g(\cdot)$ is the boxcar kernel $g(x) = (2u)^{-1}\mathbb{I}(|x| \leq u)$, for some *fixed* $u > 0$. When $M = 1$, [15] and [16] managed to circumvent this obstacle by considering a boxcar kernel $g(\cdot)$ with irrational scale. Their method is based on the fact that the Fourier coefficients of the boxcar kernel do not vanish at frequencies $(\pi k u)$ when u is a *Badly Approximable* (BA) number. An irrational number u is BA if the terms $a_n = a_n(u)$ of its continued fraction expansion $[a_0; a_1, a_2 \dots]$, where a_0 is an integer and $a_1, a_2, \dots$ is an infinite sequence of positive integers, are bounded, i.e., $\sup_n a_n(u) < \infty$. This notion is related to the fact that a BA number cannot be approximated well by a rational number which leads to the fact that f can be recovered reasonably well. Since standard deconvolution is a particular example of linear statistical ill-posed inverse problems in the sense of Hadamard, i.e., the inversion does not depend continuously on the observed data, [16] used number theory to prove that the degree of ill-posedness in boxcar deconvolution is $\nu = 3/2$. Roughly speaking, the degree of ill-posedness specifies how much the error in the right-hand side of the equation is amplified in the solution. For example, if f belongs to a space with a smoothness index $s > 0$ and the degree of ill-posedness is $\nu > 0$, then, the quadratic risk of the best possible estimator of the response function f is of the order $O(n^{-\frac{2s}{2s+2\nu+1}})$.

Following mathematical ideas of [2] and [7], [8] extended the results of [15] and [16] and showed that if M is finite, $M \geq 2$, one of the u_l's is a BA number, and $u_1, u_2, \ldots, u_M$ is a BA M-tuple, then the degree of ill-posedness is $\nu = 1 + 1/(2M)$. The notion of a BA M-tuple refers to a collection of M irrational numbers which are difficult to approximate simultaneosly by fractions with the same denominator. It will be discussed in depth in Section 2. Therefore, in the case of M channels, the estimation problem requires a construction of a BA M-tuple which has been accomplished by the number theory community (it is described in, e.g., [27, 28]).

Recent advances in data collection and recording techniques made it of urgent interest to study the case when the number of channels $M = M_n$ grow with the total number of observations n. It turns out that when the number of channels $M = M_n$ grows fast as the total number of observations n increases, one does not need to make a special choice of the points u_l, $l = 1, 2 \ldots, M$, and it is sufficient to take them to be equidistant. Indeed, [26] considered the discrete multichannel deconvolution model (1.1) as observations on the continuous functional deconvolution model

$$y(u,t) = f * g(u,t) + \frac{1}{\sqrt{n}}\, z(u,t), \quad u \in U, \;\; t \in T, \tag{1.4}$$

where $z(u,t)$ is assumed to be a two-dimensional Gaussian white noise, i.e., a generalized two-dimensional Gaussian field with covariance function

$$\mathbb{E}[z(u_1,t_1)z(u_2,t_2)] = \delta(u_1 - u_2)\delta(t_1 - t_2),$$

where $\delta(\cdot)$ denotes the Dirac δ-function, and

$$f * g(u,t) = \int_T g(u, t-x) f(x)\, dx$$

with the blurring (or kernel) function $g(\cdot,\cdot)$ assumed to be known. If $a = b$, the functional deconvolution model (1.4) reduces to the standard deconvolution model which attracted attention of a number of researchers, e.g., [9, 1, 17, 15, 10, 16, 23, 18, 4, 5], among others.

Formulation of the functional deconvolution model (1.4) allowed [26] to study the interplay between discrete and continuous deconvolution models. The *ideal* continuous deconvolution model (1.4) assumes that one can measure $y(u,t)$ at *any* $u \in U$ and $t \in T$ and that $n^{-1/2}$ marks the precision of these observations. Nevertheless, this does not happen in real-life situations where one observes $y(u,t)$ *only* at the points $u_l \in U$, $\;\; t_i = (i-1)/N$ for $l = 1, 2, \ldots, M$ and $i = 1, 2, \ldots, N$. Furthermore, [26] showed that the degree of ill-posedness in the continuous deconvolution model (1.4) is $\nu = 1$ and that it can be attained in the discrete deconvolution model (1.1) if $M = M_n \geq c_0 n^{1/3}$ for some constant $c_0 > 0$, independent of n. Indeed, in this case, one does not need to employ BA numbers or BA M-tuples: it is sufficient to observe the discrete deconvolution model (1.1) at equidistant points $u_l = a + l(b-a)/M$, $l = 1, 2, \dots, M$. This set up provides the "best possible" minimax convergence rates (under the L^2-risk and over a wide range of Besov balls) in the model.

However, in real-life situations, the number of channels M usually refers to the number of physical devices and, consequently, cannot be very big. Therefore, $M = M_n \geq c_0 n^{1/3}$ may be impossible although it is natural to assume that $M = M_n$ may grow to infinity at a slower rate as $n \to \infty$. Unfortunately, the theoretical results obtained by [8] cannot be blindly applied to accommodate the case when $M = M_n \to \infty$ as $n \to \infty$. This is due to the fact that, to the best of our knowledge, so far no one have studied the construction of a BA M-tuple of a growing length on a specified interval, of a non-asymptotic length, of the real line. Therefore, this generalization requires non-trivial results in number theory.

Our aim is to investigate the situation when $M = M_n$ grows slowly with n and to derive necessary new results in number theory in order to devise a technique which allows to approach minimax convergence rates (under the L^2-risk and over a wide range of Besov balls) in the continuous model (1.4) with a factor which grows slower than any power of n. This situation seems to be of a particular interest nowadays since data recording equipment is getting cheaper and cheaper while overall volumes of data is growing very fast.

When $M = M_n$ grows slowly as n increases, we develop a procedure for the construction of a BA M-tuple on a specified interval, of a non-asymptotic length, together with a lower bound associated with this M-tuple, which explicitly shows its dependence on M as M is growing. This result is further used for evaluation of the L^2-risk of the suggested adaptive wavelet thresholding estimator of the unknown response function and, furthermore, for the choice of the optimal number of channels M which minimizes the L^2-risk.

The theoretical results that we have obtained provide a cross-area between number theory, statistics and signal processing. We hope to alert the number theory community to a new problem of constructing a BA M-tuple on a specified interval, of a non-asymptotic length, of the real line, as M is growing. On the other hand, we believe that our findings will also be of interest to researchers in statistics and signal processing.

The rest of the paper is organized as follows. Section 2 provides some number theory background which is required for understanding the material presented in subsequent sections. Section 3 briefly reviews the adaptive wavelet thresholding estimator introduced in [25]. Section 4 explains the relationship between the L^2-risk of the estimator obtained in Section 3 and the theory of Diophantine approximation, thus, motivating the derivation of the new results in number theory obtained in Section 5. In particular, the objective of Section 5 is the construction of a BA M-tuple on a specified interval when $M = M_n \to \infty$ as $n \to 0$ and the development of related asymptotic bounds which are necessary in order to choose an optimal value of $M = M_n$ in this case. Section 6 provides the asymptotic upper bounds for the L^2-risk of the adaptive wavelet thresholding estimator constructed in Section 3 when $M = M_n$ is a slowly growing function of n. We conclude in Section 7 with a brief discussion while Section 8 contains the proofs of the theoretical results obtained in earlier sections.

2. Background results in number theory

The theory of Diophantine approximation is an important branch of number theory (see, e.g., [11, 19, 21, 28, 29]). One important topic of the above theory is the simultaneous approximation of linear forms, which was pursued as early as mid-19th century by Dirichlet and later studied by a number of profound researchers in the field. In particular, it is known that for any real numbers $\beta_1, \beta_2, \ldots, \beta_M$ there exist integer numbers q and $p_1, p_2, \ldots, p_M$ such that

$$\max_{i=1,2,\ldots,M} |\beta_i q - p_i| < \frac{M}{(M+1)} |q|^{-1/M}. \tag{2.1}$$

The above result was proved by Minkowski and has been expanded in the recent years to cover systems of linear forms (see, e.g., [28], p. 36, pp. 40-41). We note that in the case where $M = 1$, the constant $C(M) = M/(M+1)$ in (2.1) reduces to $1/2$ whereas, by Hurwitz's theorem, the best possible value is $1/\sqrt{5}$ (see, e.g., [28], Theorem 2F, p. 6). For $M = 2$, $C(M)$ takes the value $2/3$; the best possible value is unknown although if $C_0(2)$ denotes the infimum of admissible values of $C(M)$ for $M = 2$, then it is known that $\sqrt{2/7} \le C_0(2) \le 0.615$ (see, e.g., [28], p. 41). Furthermore, the corresponding best constant in the case of systems of linear forms is positive, meaning that it cannot be replaced by arbitrary small constants (see, e.g., [28], Section 4, pp. 41-47).

We, however, are interested in the opposite result. Namely, the real numbers $\beta_1, \beta_2, \ldots, \beta_M$ form a *BA M-tuple* if for any integer numbers $q > 0$ and $p_1, p_2, \ldots, p_M$ one has

$$\max_{i=1,2,\ldots,M} |\beta_i q - p_i| \ge B(M) q^{-1/M}, \tag{2.2}$$

for some constant $B(M) > 0$, dependent on M (and $\beta_1, \beta_2, \ldots, \beta_M$) but independent of q and $p_1, p_2, \ldots, p_M$ (see, e.g., [28], p. 42). It is well-known that the set of all BA M-tuples has Lebesgue measure zero, but nevertheless this set is quite large, namely there are uncountably many BA M-tuples (see [3, 6]) and the Hausdorff dimension of the set of all BA M-tuples is equal to M (see [27]). In the case where $M = 1$, the number $\beta = \beta_1$ which satisfies (2.2) is referred to in the Diophantine approximation literature as a BA number (see, e.g., [28], p. 22); in view of Hurwitz's theorem, the constant $B(1)$ in this case must satisfy $0 < B(1) < 1/\sqrt{5}$ (see, e.g., [28], pp. 41-42). Furthermore, a characterization result exists, namely a real number, that is not an integer, is BA if and only if its continued fraction coefficients are bounded. The latter is often used as a definition of a BA number, however, there is no analogous characterization for $M > 1$ (see, e.g., [28], Theorem 5F, p. 22). The above definitions of BA numbers and BA M-tuples have been also extended to cover BA systems of linear forms (see, e.g., [28], pp. 41) and their existence was proved by Perron, providing also an algorithm for constructing BA linear forms (see, e.g., [28], Theorem 4B, p. 43). Furthermore, it has been established the existence of uncountably many BA systems of linear forms (see [27]).

In what follows, we are interested in the case of BA M-tuples. Although, as indicated above, an algorithm is available for constructing BA M-tuples on the real line, these do not necessarily belong to any specified interval of the real line. Furthermore, if M is strictly fixed (independent of q), one can treat B in (2.2) as a positive constant; this, however, becomes impossible if the value of M is growing. Using the technique described in [28], Section 4, pp. 43-45, we show that one can construct a BA M-tuple $\beta_1, \beta_2, \dots, \beta_M$ of real numbers so that it lies in any specified interval (a, b), $a < b$, of non-asymptotic length, of the real line, and derive a lower bound for $B(M)$ in (2.2) as $M \to \infty$. This result is proved in Section 5.

3. An adaptive wavelet thresholding estimator

Let $\varphi^*(\cdot)$ and $\psi^*(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively, in the real line (see, e.g., [22] or [20]). As usual,

$$\varphi^*_{jk}(x) = 2^{j/2}\varphi^*(2^j x - k), \quad \psi^*_{jk}(x) = 2^{j/2}\psi^*(2^j x - k), \quad j, k \in \mathbb{Z}, \quad x \in \mathbb{R},$$

are, respectively, the dilated and translated Meyer scaling and wavelet (orthonormal) basis functions at resolution level j and scale position $k/2^j$. Similarly to Section 2.3 in [15], we obtain a periodized version of the Meyer wavelet basis, by periodizing the basis functions $\{\varphi^*(\cdot), \psi^*(\cdot)\}$, i.e., for $j \geq 0$ and $k = 0, 1, \dots, 2^j - 1$,

$$\varphi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2}\varphi^*(2^j(x+i)-k), \quad \psi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2}\psi^*(2^j(x+i)-k), \quad x \in T.$$

Let $\langle \cdot, \cdot \rangle$ denote the inner product in the Hilbert space $L^2(T)$ (the space of squared-integrable functions defined on T), i.e., $\langle f, g \rangle = \int_T f(t)\overline{g(t)}dt$ for $f, g \in L^2(T)$. Let $e_m(t) = e^{i2\pi mt}$, $m \in \mathbb{Z}$, and let $f_m = \langle e_m, f \rangle$, $g_m(u) = \langle e_m, g(u, \cdot) \rangle$, $u \in U$. For any $j_0 \geq 0$ and any $j \geq j_0$, let $\varphi_{mj_0k} = \langle e_m, \varphi_{j_0k} \rangle$ and $\psi_{mjk} = \langle e_m, \psi_{jk} \rangle$, where $\{\phi_{j_0,k}(\cdot), \psi_{j,k}(\cdot)\}$ is the periodic Meyer wavelet basis introduced above.

Using the periodized Meyer wavelet basis described above, and for any $j_0 \geq 0$, any (periodic) $f(\cdot) \in L^2(T)$ can be expanded as

$$f(t) = \sum_{k=0}^{2^{j_0}-1} a_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(t), \quad t \in T. \tag{3.1}$$

Furthermore, by Plancherel's formula, the scaling coefficients, $a_{j_0k} = \langle f, \varphi_{j_0k} \rangle$, and the wavelet coefficients, $b_{jk} = \langle f, \psi_{jk} \rangle$, of $f(\cdot)$ can be represented as

$$a_{j_0k} = \sum_{m \in C_{j_0}} f_m \overline{\varphi_{mj_0k}}, \quad b_{jk} = \sum_{m \in C_j} f_m \overline{\psi_{mjk}}, \tag{3.2}$$

where $C_{j_0} = \{m : \varphi_{mj_0k} \neq 0\}$ and, for any $j \geq j_0$, $C_j = \{m : \psi_{mjk} \neq 0\}$. Note that both C_{j_0} and C_j, $j \geq j_0$, are subsets of $(2\pi/3)[-2^{j+2}, -2^j] \cup [2^j, 2^{j+2}]$, i.e.,

$$|m| \in (2\pi/3)\ [2^j, 2^{j+2}] \tag{3.3}$$

due to the fact that Meyer wavelets are band limited (see, e.g., [15], Section 3.1).

Reconstruct the unknown response function $f(\cdot) \in L^2(T)$ in (1.3) as

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{a}_{j_0k}\varphi_{j_0k}(t) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \widehat{b}_{jk}\mathbb{I}(|\widehat{b}_{jk}| \geq \lambda_j)\psi_{jk}(t), \quad t \in T, \tag{3.4}$$

where $\widehat{a}_{j_0k}$ and $\widehat{b}_{jk}$ are the natural estimates of a_{j_0k} and b_{jk}, respectively (see (3.1) and (3.2)), given by

$$\widehat{a}_{j_0k} = \sum_{m \in C_{j_0}} \widehat{f}_m \overline{\varphi_{mj_0k}}, \quad \widehat{b}_{jk} = \sum_{m \in C_j} \widehat{f}_m \overline{\psi_{mjk}}. \tag{3.5}$$

with $\widehat{f}_m$ obtained by

$$\widehat{f}_m = \left(\sum_{l=1}^{M} \overline{g_m(u_l)} y_m(u_l) \right) \Big/ \left(\sum_{l=1}^{M} |g_m(u_l)|^2 \right), \quad u_l \in U, \quad l = 1, 2, \ldots, M;$$

here $g_m(u_l)$ and $y_m(u_l)$, $l = 1, 2, \ldots, M$, are the discrete Fourier coefficients of $y(u, \cdot)$ and $g(u, \cdot)$, respectively, obtained by applying the discrete Fourier transform to the equation (1.3). Note that, in this case,

$$g_0(u_l) = 1 \quad \text{and} \quad g_m(u_l) = \gamma(u_l)\frac{\sin(2\pi m u_l)}{2\pi m}, \quad m \in \mathbb{Z} \setminus \{0\} \quad l = 1, 2, \ldots, M. \tag{3.6}$$

The choices of the resolution levels j_0 and J and the thresholds λ_j will be described in Section 6 when we examine an expression for the L^2-risk of the estimator (3.4) over a collection of Besov balls, leading to an adaptive estimator (i.e., its construction is independent of the Besov ball parameters that are usually unknown in practice).

Among the various characterizations of Besov spaces for periodic functions defined on $L^p(T)$ in terms of wavelet bases, we recall that for an r-regular $(0 < r \leq \infty)$ multiresolution analysis with $0 < s < r$ and for a Besov ball $B^s_{p,q}(A)$ of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has that, with $s' = s + 1/2 - 1/p$,

$$B^s_{p,q}(A) = \left\{ f(\cdot) \in L^p(T) : \left(\sum_{k=0}^{2^{j_0}-1} |a_{j_0k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \leq A \right\}, \tag{3.7}$$

with respective sum(s) replaced by maximum if $p = \infty$ or $q = \infty$ (see, e.g., [15], Section 2.4). (Note that, for the Meyer wavelet basis, considered considered above, $r = \infty$.)

The parameter s measures the number of derivatives, where the existence of derivatives is required in an L^p-sense, while the parameter q provides a further finer gradation. The Besov spaces include, in particular, the well-known Sobolev and Hölder spaces of smooth functions but in addition less traditional spaces, like the space of functions of bounded variation. The latter functions are of statistical interest because they allow for better models of spatial inhomogeneity (see, e.g., [22]).

The precision of the estimator (3.4) is measured by the (maximal) L^2-risk given by

$$R_n(\hat{f}_n) = \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_n - f\|_2^2. \tag{3.8}$$

We are interested in the asymptotic rate of convergence of the estimator $\hat{f}_n$, i.e., we are interested in the following asymptotical upper bounds

$$R_n(\hat{f}_n) \le C\gamma_n \quad \text{as} \quad n \to \infty,$$

where $\{\gamma_n\}_{n=1}^{\infty}$ is a positive sequence converging to 0 as $n \to \infty$ and $C > 0$ is a generic constant, independent of n, which may take different values at different places.

Hereafter, $\|\cdot\|_2$ denotes the L^2-norm, $\hat{f}_n(\cdot)$ is an estimator (i.e., a measurable function) of $f(\cdot) \in L^2(T)$, based on observations from model (1.3), and the expectation in (3.8) is taken under the true $f(\cdot)$.

4. Relation to the theory of Diophantine approximation

By direct evaluations (see also Lemma 5 and its proof in the Appendix), one can show that

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = N^{-1} \sum_{m \in C_j} |\psi_{mjk}|^2 \left[\sum_{l=1}^{M} |g_m(u_l)|^2\right]^{-1}.$$

Since in the case of Meyer wavelets, $|\psi_{mjk}| \le 2^{-j/2}$ and $|C_j| \asymp 2^j$ (see, e.g., [15], p. 565), we derive

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = O\Big(n^{-1}\Delta_1(j)\Big), \tag{4.1}$$

where

$$\Delta_1(j) = \frac{1}{|C_j|} \sum_{m \in C_j} \left[M^{-1} \sum_{l=1}^{M} |g_m(u_l)|^2\right]^{-1} \equiv \frac{1}{|C_j|} \sum_{m \in C_j} [\tau_1(m)]^{-1}$$

with $\tau_1(m) \equiv \tau_1(m; \underline{u}, M) = M^{-1} \sum_{l=1}^{M} |g_m(u_l)|^2$, $\underline{u} = (u_1, u_2, \ldots, u_M)$. By (1.2) and (3.6), one has

$$\tau_1(m) \asymp \frac{1}{m^2 \, M} \sum_{l=1}^{M} \sin^2(2\pi m u_l). \tag{4.2}$$

Here, $u(m) \asymp v(m)$ mean that there exist constants $C_1 > 0$ and $C_2 > 0$, independent of m, such that $0 < C_1 v(m) \le u(m) \le C_2 v(m) < \infty$ for every m.

Therefore, the risk of the estimator $\hat{f}_n(t)$ defined in (3.4) is determined by the rate of growth of $\Delta_1(j)$ as $j \to \infty$ which, in turn, depends on the rate at which $\tau_1(m)$ goes to zero as $m \to \infty$.

It is easy to see that for some choices of M and $\underline{u}$ (e.g., $M = 1$, $u_1 = u = 1$), one has $\min_m \tau_1(m; \underline{u}, M) = 0$ for every m which leads to an infinite variances of the estimated coefficients $\widehat{b}_{jk}$ and, consequently, to an infinite L^2-risk. Hence, the choice of M and the selection of points $\underline{u}$ is of an uttermost importance. In particular, we want to choose points $(u_1, u_2, \ldots, u_M)$ such that $\sum_{l=1}^{M} \sin^2(2\pi m u_l)$ is as large as possible for $m \in C_j$ and large j.

Moreover, for any choice of M and any selection of points $\underline{u}$, one has

$$\tau_1(m; \underline{u}, M) \leq K_1 m^{-2}$$

for some constant $K_1 > 0$ independent of m, the choice of M and the selection points $\underline{u}$, so that, for any j and selection of M and $\underline{u}$,

$$\Delta_1(j) \geq K_2 2^{2j}, \tag{4.3}$$

for some constant $K_2 > 0$, independent of j. It turns out that if $M = M_n$ increases at least as fast as $n^{1/3}$, then, by sampling u_l, $l = 1, 2, \ldots, M$, uniformly on U, i.e., by selecting $u_l = a + (b-a)l/M$, $l = 1, 2, \ldots, M$, one can attain

$$\Delta_1(j) \leq K_3 2^{2j}$$

for some constant $K_3 > 0$, independent of j, so that the upper and the lower bounds in this case coincide up to a constant independent of n (see [26]).

Unfortunately, the above results do not hold for finite values of M or when $M = M_n$ is a slowly growing function of n. Indeed, in the case of small values of M, both $\tau_1(m; \underline{u}, M)$ and $\Delta_1(j)$ have completely different dynamics from large M. Indeed, if $M = 1$, [16] and [15] showed that in the case of $\gamma(u) = 1/u$, $u_1 = u^* = a = b$, one has $\Delta_1(j) \geq K_4 2^{3j}$ for any choice of u^* and some constant $K_4 > 0$, independent of j. Furthermore, [15] also demonstrated that if u^* is selected to be a BA number, then the lower bound for $\Delta_1(j)$ is attainable, i.e., $\Delta_1(j) \leq K_5 2^{3j}$ for some constant $K_5 > 0$, independent of j. Hence, in this case, $\Delta_1(j) \asymp 2^{3j}$.

These results were extended by [8] who studied the multichannel deconvolution model with a boxcar kernel and showed that the convergence rates obtained by [15] for $M = 1$ can be improved by sampling at several different points. In particular, they demonstrated that if M is *finite*, $M \geq 2$, one of the $u_1, u_2, \ldots, u_M$ is a BA number, and $\underline{u}$ is a BA M-tuple defined in (2.2), then

$$\Delta_1(j) \leq C(M)\, j 2^{j(2+1/M)} \tag{4.4}$$

for some positive $C(M)$. In particular, when M is growing with n, the value of $C(M)$ depends on n and, hence, affects the convergence rates of the estimator $\hat{f}_n(\cdot)$ as $n \to \infty$.

The relation between the convergence rates of the estimator $\hat{f}_n(\cdot)$, given by (3.4), of $f(\cdot)$ in the model (1.3) and the theory of Diophantine approximation becomes obvious when one notes that in (4.2), for any $m \in \mathbb{Z} \setminus \{0\}$ and any u_l, $l = 1, 2, \ldots, M$, one has, combining the periodic behavior of the sine function together with a first order (linear) approximation,

$$4\|2mu_l\|^2 \leq \sin^2(2\pi mu_l) \leq \pi^2 \, \|2mu_l\|^2, \quad l = 1, 2, \ldots, M, \tag{4.5}$$

where $\|a\| = \inf\{|a-k|, \, k \in \mathbb{Z}\}$ denotes the distance from a real number a to the nearest integer number. Hence, (4.2) becomes

$$\tau_1(m) \equiv \tau_1(m; \underline{u}, M) \asymp \frac{1}{m^2 \, M} \sum_{l=1}^{M} \|2mu_l\|^2, \tag{4.6}$$

so that the convergence rates of the estimator $\hat{f}_n(\cdot)$ depend on the lower bound, in terms of m, of the expression (4.6).

The value of $C(M)$ in (4.4) is related to the value of $B(M)$ in (2.2). To the best of our knowledge, there has not been developed a procedure for construction of a BA M-tuple on a specified interval, of non-asymptotic length, of the real line, and there are no asymptotic lower bounds, in terms of M, on $B(M)$ in (2.2) when the value of M is growing. For this reason, in order to find upper bounds of estimator (3.4) and choose an optimal relation between the sample size n and the number of channels M when $M = M_n$ is a slowly growing function of n, we need to obtain new original results in Diophantine approximations. In particular, the objective of the next section is to construct a BA M-tuple on the non-asymptotic interval U, of the real line, and to obtain a lower bound on $B(M)$ in terms of M for this BA M-tuple when M grows slowly with n.

5. Construction of a BA M-tuple on a specified interval

Below, we construct a BA M-tuple $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$ of real numbers on a specified interval (a, b), of a non-asymptotic length, of the real line, and derive the lower bound on $B(M)$ in formula (2.2). For this construction, we use the technique described in [28], Section 4, pp. 43-45. In particular, we shall provide an algorithm for construction of an M-tuple $\beta_1, \beta_2, \ldots, \beta_M$ of real numbers such that, as $M \to \infty$,

1. it lies in any specified interval (a, b), $a < b$, of nonasymptotic length, of the real line, and
2. it satisfies

$$\max_{i=1,2,\ldots,M} |\beta_i q - p_i| \geq B_0 \exp(-6M \ln M) q^{-1/M}, \tag{5.1}$$

   for any integer numbers $q > 0$ and $p_1, p_2, \ldots, p_M$, and for some constant $B_0 > 0$, independent of M, q and $p_1, p_2, \ldots, p_M$, so that $B(M) = B_0 \exp(-6M \ln M)$ in (2.2).

Assume that M is large enough, fix a positive integer Q and consider

$$P(x) = (x - Q)(x - 2Q) \cdots (x - MQ) - 1, \tag{5.2}$$

a monic polynomial (i.e., a polynomial with a unit leading coefficient) of the degree M. Let $\xi_1, \xi_2, \ldots, \xi_M$ be the roots of a polynomial (5.2). Recall that ξ is called an *algebraic integer* number if it is a root of some monic polynomial with coefficients being integer numbers. Algebraic integers are called *conjugate* if they are roots of the same monic polynomial with integer coefficients.

Then, the following statement is valid.

Lemma 1. *If $Q \geq 5M$, then $\xi_1, \xi_2, \ldots, \xi_M$ are real conjugate algebraic integer numbers such that*

$$(i - 1/2)Q < \xi_i < (i + 1/2)Q, \quad i = 1, 2, \ldots, M. \tag{5.3}$$

Now, to construct a BA M-tuple, choose $Q \geq 5(M + 1)$ and construct real conjugate algebraic integers $\xi_1, \xi_2, \ldots, \xi_M, \xi_{M+1}$ using the process described in Lemma 1. Let $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ be a solution to the following system of equations:

$$\sum_{i=1}^{M} \xi_k^{i-1} \alpha_i = -\xi_k^M, \quad k = 1, 2, \ldots, M. \tag{5.4}$$

Observe that the determinant of the system of equations (5.4) is a Vandermonde determinant; hence, it is nonzero since $\xi_i \neq \xi_j$ for $i \neq j$. Therefore, the system of linear equations (5.4) has a unique solution $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$, which turns out to be a BA M-tuple.

Lemma 2. *The solution $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ of the system of equations (5.4) is a BA M-tuple such that*

$$|\alpha_k| \leq 30 \exp(3M \ln M), \quad k = 1, 2, \ldots, M, \tag{5.5}$$

and for any integer numbers $q > 0$ and $p_1, p_2, \ldots, p_M$, as $M \to \infty$, one has

$$\max_{i=1,2,\ldots,M} |\alpha_i q - p_i| \geq C_0 \exp(-3M \ln M)\, q^{-1/M} \tag{5.6}$$

with some constant $C_0 > 0$, independent of M, q and $p_1, p_2, \ldots, p_M$.

Lemma 2 provides a BA M-tuple which, however, does not necessarily belong to the specified interval (a, b), of a non-asymptotic length, of the real line. Assume, without loss of generality, that both a and b are rational numbers, otherwise, replace (a, b) by $(a^*, b^*) \in (a, b)$, where a^* and b^* are rational numbers. Let $a = p_a/q_0$ and $b = p_b/q_0$ for some integer numbers p_a, p_b and q_0, and let z be an integer number such that $z - 1 < 30 \exp(3M \ln M) \leq z$. Define

$$\beta_l = a + \alpha_l (b - a)/z, \quad l = 1, 2, \ldots, M, \tag{5.7}$$

where $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ is the BA M-tuple constructed in Lemma 2.

The following theorem confirms that $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$, as constructed above, forms a BA M-tuple on the specified (a, b), of a non-asymptotic length, of the real line.

Theorem 1. *The real numbers $\beta_1, \beta_2, \dots, \beta_M$ defined in (5.7) lie on the interval (a, b), of a non-asymptotic length, and form a BA M-tuple, so that, as $M \to \infty$, one has*

$$\max_{i=1,2,\dots,M} |\beta_i q - p_i| \geq B_0 \exp(-6M \ln M)\, q^{-1/M}, \tag{5.8}$$

for any integer numbers $q > 0$ and $p_1, p_2, \dots, p_M$, and for some constant $B_0 > 0$, independent of M, q and $p_1, p_2, \dots, p_M$, so that $B(M) = B_0 \exp(-6M \ln M)$.

6. Asymptotical upper bounds for the L^2-risk of the adaptive wavelet thresholding estimator

In Section 5, we constructed a BA M-tuple and derived a lower bound on $B(M)$ in (2.2), as $M \to \infty$. We can now choose the resolution levels j_0 and J, the thresholds λ_j in (3.4) and the optimal relation between the total number of observations n and the number of channels $M = M_n$ and derive asymptotical upper bounds for the L^2-risk of the estimator $\hat{f}_n(\cdot)$ given by (3.4) over a collection of Besov balls.

In order to formulate and prove Theorem 2 we first need to obtain some preliminary results. Recall that $\|a\|$ denotes the distance from a real number a to the nearest integer number. For this purpose, we recall the *equidistribution* lemma (see Lemma 3), proved in [16], which we state here for completeness, and formulate a new lemma (i.e., Lemma 4) which is based on application of Lemma 3 to the BA M-tuple.

Lemma 3. *(Lemma 1 in [16]) Let p/q and p'/q' be successive convergents in the continued fraction expansion of a real number a. Let N be a positive integer number with $N + q < q'$. Let h be a non-increasing function. Then*

$$\sum_{i=4}^{q} h(i/q) \leq \sum_{k=N+1}^{N+q} h(\|ka\|) \leq 2 \sum_{i=1}^{q-3} h(i/q) + 6h(1/(2q')).$$

Lemma 4. *Let $\beta_1, \beta_2, \dots, \beta_M$ be a BA M-tuple constructed in Theorem 1 and let β_1 be a BA number. Let r_0 be an arbitrary fixed positive real number. Denote*

$$\aleph_k(j, M) = \sum_{l \in \Omega_j} \left[\|l\beta_1\|^2 + \cdots + \|l\beta_M\|^2 \right]^{-k}, \tag{6.1}$$

where Ω_j is defined as

$$\Omega_j = \left\{ l : 2^j \leq |l| \leq 2^{j+r_0} \right\}. \tag{6.2}$$

If M is large enough, then, as $j \to \infty$,

$$\aleph_k(j, M) = O\left(j\ 2^{j(1+(2k-1)/M)}\ e^{6(2k-1)M \ln M} \right), \quad k = 1, 2, 3, 4. \tag{6.3}$$

We also need the following two lemmas which evaluate the precision of estimation of $a_{j_0 k}$ and b_{jk}.

Lemma 5. *Let $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_M)$ be a BA M-tuple constructed on the interval $(2a, 2b)$ according to Theorem 1 and let one of $\beta_1, \beta_2, \dots, \beta_M$ be a BA number. Let the equation (1.3) be evaluated at the the point $\underline{u}$ with components $u_l = \beta_l/2$, $l = 1, 2, \dots, M$. Then, for all $j \ge j_0$, as $n \to \infty$,*

$$\mathbb{E}|\widehat{a}_{j_0k} - a_{j_0k}|^2 = O\left(n^{-1} j_0 \ M \ 2^{j_0(2+1/M)} \ e^{6M \ln M}\right),$$

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2 = O\left(n^{-1} j \ M \ 2^{j(2+1/M)} \ e^{6M \ln M}\right),$$

$$\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 = O\left(n^{-2} \ j^2 \ M^2 \ 2^{j(4+2/M)} \ e^{42M \ln M}\right).$$

Lemma 6. *Let $u_1, u_2, \dots, u_M$ be as in Lemma 5. If $\eta > 0$ is a constant large enough, then, for all $j \ge j_0$, as $n \to \infty$,*

$$\mathbb{P}(|\widehat{b}_{jk} - b_{jk}|^2 \ge \eta^2 (n_M)^{-1} j 2^{j(2+1/M)} \ln n) = o\left(n^{-\theta}\right),$$

where $n_M = (n/M) \exp(-6M \ln M)$ and $\theta = \eta^2/(2C_\psi)$ with $C_\psi = 2^{-j}|C_j|$.

We are now ready to formulate Theorem 2. Let the resolution levels j_0 and J and the thresholds λ_j be such that

$$2^{j_0} = \ln n, \quad 2^J = (n_M)^{\frac{1}{3+1/M}}, \quad \lambda_j = \eta \, (n_M)^{-1/2} \sqrt{j 2^{j(2+1/M)} \ln n}, \tag{6.4}$$

for some constant $\eta > 0$, where

$$n_M = \frac{n}{M} \exp(-6M \ln M). \tag{6.5}$$

Note that since the construction of j_0, J and λ_j is independent of the Besov ball parameters, s, p, q and A, the suggested wavelet thresholding estimator $\hat{f}_n(\cdot)$ given by (3.4) is adaptive with respect to these parameters.

The following statement provides the asymptotical upper bounds for the L^2-risk, over a collection of Besov balls.

Theorem 2. *Let $s > 1/\min(p, 2)$, $1 \le p \le \infty$, $1 \le q \le \infty$ and $A > 0$. Let $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_M)$ be a BA M-tuple constructed on the interval $(2a, 2b)$ according to Theorem 1 and let one of $\beta_1, \beta_2, \dots, \beta_M$, say β_1, be a BA number. Let the equation (1.3) be evaluated at the the point $\underline{u}$ with components $u_l = \beta_l/2$, $l = 1, 2, \dots, M$. Choose*

$$M = M_n = \nu \sqrt{\ln n/(\ln \ln n)} \tag{6.6}$$

for some $\nu \le 1/\sqrt{6}$, independent of n. Let $\hat{f}_n(\cdot)$ be the adaptive wavelet thresholding estimator defined by (3.4) with j_0, J and λ_j given by (6.4), and n_M given by (6.5). Then, as $n \to \infty$,

$$R_n(\hat{f}_n) \le \begin{cases} C\, n^{-\frac{2s}{2s+3}} \, \alpha_n, & \text{if} \quad s > 3(1/p - 1/2), \\ C \left(\frac{\ln n}{n}\right)^{\frac{s'}{s'+1}} \alpha_n, & \text{if} \quad s \le 3(1/p - 1/2), \end{cases} \tag{6.7}$$

where α_n *is given by*

$$\alpha_n = \exp\left\{\sqrt{\ln n}\sqrt{\ln\ln n}\left[\frac{A_1}{A_2}\left(3\nu + \frac{1}{A_2\nu}\right) + r_n\right]\right\}, \tag{6.8}$$

with

$$\begin{aligned} r_n \;=\; & \frac{3A_1\nu\ \ln\ln\ln n}{A_2\ \ln\ln n}\left(\frac{2\ \ln\nu}{\ln\ln\ln n} - 1\right) + \frac{\sqrt{\ln\ln n}}{\sqrt{\ln n}}\left(\frac{A_3}{A_2} + \frac{A_1}{2A_2} - \frac{3A_1}{A_2^2} - \frac{A_1}{A_2^3\nu^2}\right) \\ & + \frac{\ln\ln\ln n}{\sqrt{\ln n}\sqrt{\ln\ln n}}\left(\frac{3A_1}{A_2^2} - \frac{A_1}{2A_2}\right) = o(1)\ (n\to\infty), \end{aligned}$$

where

$$\begin{array}{lll} A_1 = 2s,\quad A_2 = 2s+3,\quad A_3 = 2s, & \textit{if} & 2 \le p \le \infty, \\ A_1 = 2s,\quad A_2 = 2s+3,\quad A_3 = 4s, & \textit{if} & 6/(2s+3) < p < 2, \\ A_1 = 2s^*,\ A_2 = 2s^*+3,\ A_3 = 4s^*, & \textit{if} & 1 \le p \le 6/(2s+3) \end{array} \tag{6.9}$$

with $s^* = \min(s', s)$, $s' = s + 1/2 - 1/p$.

7. Discussion

We considered the estimation problem of the unknown response function in the multichannel boxcar deconvolution model with a boxcar-like kernel when the number of channels grows as the total number of observations increases. This situation seems to be of a particular interest nowadays since data recording equipment is getting cheaper and cheaper while overall volumes of data is growing very fast. Our aim was to investigate the situation when the number of channels $M = M_n$ grows slowly with the number of observations n.

For this purpose, we obtained new original results in the field of Diophantine approximation in order to devise a technique which allows the reconstruction of the unknown response function with a precision that differs from the best possible convergence rates (which can be attained in the corresponding continuous functional deconvolution model (1.4)) by a factor which grows slower than any power of n.

Specifically, in Section 6, we derived asymptotical upper bounds for the L^2-risk of the adaptive wavelet thresholding estimator (3.4) of $f(\cdot) \in L^2(T)$ in the model (1.3). In comparison, it follows from [26] that the choice of a uniform sampling strategy (i.e., $u_l = a + (b-a)l/M$, $l = 1, 2, \ldots, M$, for $M = M_n \ge (32\pi/3)(b-a)n^{1/3}$, leads to an adaptive wavelet block thresholding estimator $\hat{f}_n^{\mathrm{B}}(\cdot)$ of $f(\cdot)$ with the following convergence rates

$$R_n(\hat{f}_n^{\mathrm{B}}) \le \begin{cases} Cn^{-\frac{2s}{2s+3}}\,(\ln n)^{\varrho}, & \text{if} \quad s > 3(1/p - 1/2), \\ C\left(\frac{\ln n}{n}\right)^{\frac{s'}{s'+1}}\,(\ln n)^{\varrho}, & \text{if} \quad s \le 3(1/p - 1/2), \end{cases} \tag{7.1}$$

for $s > 1/\min(p,2)$, where $s' = s + 1/2 - 1/p$ and

$$\varrho = \begin{cases} \frac{3\max(0,2/p-1)}{2s+3}, & \text{if} \quad s > 3(1/p - 1/2), \\ \max(0, 1-p/q), & \text{if} \quad s = 3(1/p-1/2), \\ 0, & \text{if} \quad s < 3(1/p-1/2). \end{cases}$$

Moreover, its has been shown that the above convergence rates with $\rho = 0$ are the fastest possible ones (see [26]); hence, up to the logarithmic factor $(\ln n)^{\varrho}$, $\hat{f}_n^{\mathrm{B}}(\cdot)$ attains the best possible convergence rates. By comparing the convergence rates (6.7) in Theorem 2 to the fastest possible convergence rates (without the extra logarithmic factor $(\ln n)^{\rho}$ appearing in (7.1)), one concludes that they differ by the extra factor α_n defined in (6.8).

How fast does $\alpha_n \to \infty$ as $n \to \infty$? It can be easily seen that α_n grows slower than any power of n but faster than any power of $\ln n$, i.e., for any $a_1, a_2 > 0$, one has

$$\lim_{n\to\infty} \frac{\alpha_n}{n^{a_1}} = 0, \quad \lim_{n\to\infty} \frac{\alpha_n}{(\ln n)^{a_2}} = \infty.$$

Hence, although choosing $M = M_n \to \infty$ at a rate given by (6.6) improves the convergence rates in comparison with the finite values of M (see [8], Theorem 2), these rates are quite a bit worse than in the case when $M = M_n$ grows at a faster rate as $n \to \infty$. Since, as we have explained in Section 4, this fast growth of $M = M_n$ with n cannot be achieved in a number of practical situations, one has to resign to M_n growing slowly with n, in particular, $M = M_n = o((\ln n)^{\alpha_3})$ for some $\alpha_3 \geq 1/2$.

The interesting question, however, is whether the convergence rates (6.7) can be improved. To uncover an answer to this question, one needs to either come up with another procedure for constructing a BA M-tuple which belongs to a specified interval, of a non-asymptotic length, of the real line and delivers a higher value of $B(M)$ in (2.2), as $M \to \infty$, or to show that no matter what the value of $\underline{u} = (u_1, u_2, \ldots, u_M)$ is, there exist integer numbers q and $p_1, p_2, \ldots, p_M$ such that, as $M \to \infty$,

$$\max_{i=1,2,\ldots,M} |u_i q - p_i| \leq B_1 \exp(-6M \ln M) q^{-1/M},$$

for some positive constant B_1 independent of M, q and $p_1, p_2, \ldots, p_M$. At the moment we are unable to provide answers to either of the above questions; we challenge, however, the number theory community to work on the issue. Derivation of these results will not only enrich the theory of Diophantine approximation but will also be valuable for the theory of statistical signal processing.

8. Appendix: Proofs

Proof of Lemma 1. Observe that for $P(x)$ given by (5.2), one has

$$P((M+1/2)Q) > 0, \quad (-1)P((M-1/2)Q) > 0, \quad \ldots, \quad (-1)^M P(Q/2) > 0,$$

so that $P(x)$ has M real roots $\xi_1, \xi_2, \ldots, \xi_M$ such that (5.3) is valid. By definition, $\xi_1, \xi_2, \ldots, \xi_M$ are algebraic integer numbers. Let us show that no proper subset of $\xi_1, \xi_2, \ldots, \xi_M$ is itself a set of conjugate algebraic integer numbers. For this purpose, note that

$$Q(|j-i| - 1/2) \le |\xi_i - jQ| \le Q(|j-i| + 1/2), \quad i \ne j, \ i, j = 1, 2, \ldots, M. \tag{8.1}$$

Therefore, by (5.2), for any $i = 1, 2, \ldots, M$,

$$0 \le |\xi_i - iQ| = \left[\prod_{\substack{j=1 \\ j \ne i}}^{M} |\xi_i - jQ| \right]^{-1} \le Q^{-(M-1)} \prod_{\substack{j=1 \\ j \ne i}}^{M} (|j-i| - 1/2)^{-1}. \tag{8.2}$$

Now, assume that $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_m}$, $i_1 < \ldots < i_m$ and $m < M$, form a set of conjugate real integer numbers. Then, $P^*_m(i_1 Q) = (\xi_{i_1} - i_1 Q) \ldots (\xi_{i_m} - i_1 Q)$ is an integer number and is not equal to zero, hence, $|P^*_m(i_1 Q)| \ge 1$. On the other hand, by (8.1) and (8.2),

$$1 \le |P^*_m(i_1 Q)| \le Q^{-(M-1)} \prod_{\substack{j=1 \\ j \ne i_1}}^{M} (|j - i_1| - 1/2)^{-1} \ \prod_{k=2}^{m} [Q(|i_k - i_1| + 1/2)].$$

The product in the right-hand side above takes the largest value if $m = M-1$, $i_1 = 1$ and $i_k = k+1$, $k = 2, 3, \ldots, M-1$. In this case, for $M > 2$, combination of the last two inequalities yields

$$1 \le |P^*_m(i_1 Q)| \le 4(M - 1/2) Q^{-1} < 5M/Q,$$

which leads to a contradiction when $Q > 5M$. □

Proof of Lemma 2. Choose $Q = 5(M+1)$ and construct real conjugate algebraic integer numbers $\xi_1, \xi_2, \ldots, \xi_M, \xi_{M+1}$ using the process described in Lemma 1. Then, by (8.2), $\xi_i \approx Qi$. Let $q > 0$ and $\underline{p} = (p_1, p_2, \ldots, p_M)$ be integer numbers and denote

$$H_k(q, \underline{p}) = \sum_{i=1}^{M} \xi_k^{i-1} p_i + \xi_k^M q, \quad k = 1, 2, \ldots, M+1.$$

Note that if $\underline{p}$ is not zero and the components of the vector $\underline{p}/q$ are not integer numbers, then $H_k(q, \underline{p}) \ne 0$, $k = 1, 2, \ldots, M+1$. Furthermore, $H_k(q, \underline{p})$, $k = 1, 2, \ldots, M+1$, are themselves real conjugate algebraic integer numbers and, thus,

$$\prod_{k=1}^{M+1} |H_k(q, \underline{p})| \ge 1.$$

Now, note that (5.4) implies that $\alpha_1, \alpha_2, \ldots, \alpha_M$ are coefficients of the monic polynomial with the roots $\xi_1, \xi_2, \ldots, \xi_M$. Also, it is easy to check that for the solution $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ of the system of equations (5.4), one can write

$$H_k(q, \underline{p}) = \sum_{i=1}^{M} \xi_k^{i-1}(p_i - \alpha_i q), \quad k = 1, 2, \ldots, M. \tag{8.3}$$

Moreover, if one denotes

$$\omega_M = \sum_{i=1}^{M} \alpha_i \xi_{M+1}^{i-1} + \xi_{M+1}^{M}, \tag{8.4}$$

then $H_{M+1}(q, \underline{p})$ can be written as

$$H_{M+1}(q, \underline{p}) = \sum_{i=1}^{M} \xi_{M+1}^{i-1}(p_i - \alpha_i q) + \omega_M. \tag{8.5}$$

Note that (8.4) implies that ω_M is the value of the polynomial

$$\mathcal{P}(x) = \sum_{i=1}^{M} \alpha_i x^{i-1} + x^M = (x - \xi_1) \cdots (x - \xi_M)$$

at the point ξ_{M+1}. Therefore, by (8.1) and (8.2), $|\omega_M| \leq KM!Q^M$ for some constant $K > 0$.

Recall that $\|a\|$ denotes the distance from a real number a to the nearest integer number. Denote $L = \max_{i=1,2,\ldots,M} |\xi_i q - p_i|$. Note that we can assume that p_i/q, $i = 1, 2, \ldots, M$, are not integer numbers. Otherwise, if, for instance, $p_1/q = z$ is an integer number, then $L \geq q|\xi_1 - z| \geq q\|\xi_1\|$ and (5.6) is valid. If $L \geq 1$, then (5.6) is valid. Hence, consider the case of $L < 1$. Then $L < |q|$ and, by (8.3), we have

$$|H_k(q, \underline{p})| \leq L \sum_{i=1}^{M} \xi_k^{i-1} < L\xi_k^M/(\xi_k - 1), \quad k = 1, 2, \ldots, M.$$

Then, using (8.5) and an upper bound for ω_M, we obtain

$$|H_{M+1}(q, \underline{p})| \leq |q|(\xi_{M+1}^M/(\xi_{M+1} - 1) + KM!Q^M).$$

Since $H_1(q, \underline{p}), H_2(q, \underline{p}), \ldots, H_M(q, \underline{p}), H_{M+1}(q, \underline{p})$ are real conjugate algebraic integer numbers, one has

$$1 \leq \prod_{k=1}^{M+1} |H_k(q, \underline{p})| \leq \prod_{k=1}^{M} \left[\frac{L\xi_k^M}{\xi_k - 1}\right] |q| \left[\frac{\xi_{M+1}^M}{\xi_{M+1} - 1} + KM!Q^M\right].$$

Note that, by (8.2) and $Q \geq 5M$, one has $|\xi_{M+1} - (M+1)Q| < Q^{-M}$ and, hence, $|\xi_{M+1}|^M \leq 2Q^M(M+1)^M$. Therefore

$$L \geq K|q|^{-1/M} \prod_{k=1}^{M} \left[\frac{Qk - 1}{k^M Q^M}\right]^{1/M} \left[\frac{(M+1)^M Q^M}{Q(M+1) - 1} + Q^M M!\right]^{-1/M}.$$

Plugging in $Q = 5(M+1)$ into the expression above, we obtain

$$L \geq B(M)|q|^{-1/M}$$

with

$$\begin{aligned} B(M) &= K\,[5(M+1)]^{-(M+1)}(M!)^{-1}\prod_{k=1}^{M}[5k(M+1)-1]^{1/M} \\ &\quad \times \left[\frac{(M+1)^M}{5(M+1)^2-1} + M!\right]^{-1/M}. \end{aligned}$$

Using Stirling formula,

$$M! = \sqrt{2\pi}(M+1)^{M+1/2}\exp(-(M+1))(1+o(1)), \quad \text{as} \quad M \to \infty,$$

(see, e.g., formula 8.327 of [12]) and the fact that $\ln(M+1) < \ln(M) + 1/M$, after some simple algebra, we obtain that, as $M \to \infty$,

$$B(M) \geq C_0 \exp(-3M \ln M),$$

for some constant $C_0 > 0$, independent of M, q and $\underline{p}$, which proves (5.6).

Now, it remains to prove the upper bound (5.5) for α_k, $k = 1, 2, \ldots, M$. For this purpose, recall that $\alpha_1, \alpha_2, \ldots, \alpha_M$ are coefficients of the monic polynomial with roots $\xi_1, \xi_2, \ldots, \xi_M$. Therefore, using (8.2), obtain

$$|\alpha_k| \leq \binom{M}{k} MQ(M-1)Q\cdots(M-k+1)Q = k!\binom{M}{k}^2 Q^k, \quad k = 1, 2, \ldots, M.$$

Since for any $k = 1, 2, \ldots, M$, $5^k/k! \leq 625/24 < 30$, $Q = 5(M+1)$ and $(M+1)(M-j) \leq M^2$ for $j \geq 1$, one has (reading $\prod_{j=0}^{-1} = 1$)

$$\begin{aligned} |\alpha_k| &\leq \frac{5^k}{k!}\,(M+1)^k \prod_{j=0}^{k-1}(M-j)^2 \\ &\leq 30\, M^2[(M+1)(M-k+1)]^2 \prod_{j=1}^{k-2}[(M+1)(M-j)^2] \\ &\leq 30M^{3k} \leq 30e^{3M\ln M}, \quad k = 1, 2, \ldots, M, \end{aligned}$$

which proves (5.5). □

Proof of Theorem 1. It is easy to check that $\beta_1, \beta_2, \ldots, \beta_M$, as defined by (5.7), lie on (a, b). Furthermore, by Lemma 2 and the fact that $z < 30\,\exp(3M \ln M)$, as $M \to \infty$, one has

$$\begin{aligned} \max_{i=1,2,\ldots,M} |\beta_i q - p_i| &= (zq_0)^{-1} \max_{i=1,\ldots,M} |\alpha_k(p_b - p_a)q - (zq_0 p_l - zp_a q)| \\ &\geq (zq_0)^{-1}\, C_0 |(p_b - p_a)q|^{-1/M} \exp(-3M \ln M) \\ &\geq B_0 \exp(-6M \ln M)\; |q|^{-1/M}, \end{aligned}$$

for any integer numbers $q > 0$ and $p_1, p_2, \ldots, p_M$, and for some constant $B_0 > 0$, independent of M, q and $\underline{p}$. □

Proof of Lemma 4. Recall first that any real number a, which is not an integer number, may be uniquely determined by its continued fraction expansion

$$a = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where a_0 is an integer number and $a_1, a_2, \ldots$ are strictly positive integer numbers. The convergents $p_k/q_k = p_k(a)/q_k(a)$, $k = 0, 1, \ldots$, of a are those rational numbers, the continued fraction expansions of which terminate at stage k, that is, $p_0/q_0 = a_0$, $p_1/q_1 = a_0 + 1/a_1$, $p_2/q_2 = a_0 + 1/(a_1 + 1/a_2)$, and so on. The denominators in the above expansions grow at least geometrically

$$\begin{array}{rcll} q_{n+i} & \geq & 2^{(i-1)/2} q_n, & \text{if } \ i \text{ odd}, \\ q_{n+i} & \geq & 2^{i/2} q_n, & \text{if } \ i \text{ even}, \end{array} \tag{8.6}$$

and $a_n < q_n/q_{n-1} \leq a_n + 1$, $n \geq 1$. A real number a is BA if $\sup_n a_n < \infty$, i.e., there exists $\bar{q} > 0$ such that

$$q_n/q_{n-1} \leq \bar{q}, \quad n \geq 1 \tag{8.7}$$

(see, e.g., [28], Sections 3–5, pp. 7–23).

Let p/q and p'/q' be successive principal convergents in the continued fraction expansion of β_1. Let N be a positive integer number with $N + q < q'$. Then, application of Lemma 3 with $h(x) = x^{-1}$ yields

$$\sum_{l=N+1}^{N+q} \|l\beta_1\|^{-1} = O(q \ln q), \tag{8.8}$$

since $q' \leq \bar{q} q$ by (8.7). Now, note that by (5.8)

$$\sum_{l=N+1}^{N+q} \left(\|l\beta_1\|^2 + \cdots + \|l\beta_M\|^2\right)^{-k} \leq \sum_{l=N+1}^{N+q} \|l\beta_1\|^{-1} [\max(\|l\beta_1\|, \ldots, \|l\beta_M\|)]^{-(2k-1)} \tag{8.9}$$

Combination of (5.8), (8.8) and (8.9) implies that, for $k = 1, 2, 3, 4$,

$$\sum_{l=N+1}^{N+q} \left(\|l\beta_1\|^2 + \cdots + \|l\beta_M\|^2\right)^{-k} = O\left(e^{6(2k-1)\, M \ln M} \ q^{(1+(2k-1)/M)} \ln q\right). \tag{8.10}$$

Now, observe that the set of indices l in Ω_j is symmetric about zero, and so are the components of the sum. Hence, we can consider only the positive part of Ω_j which, with some abuse of notation, we keep calling it Ω_j. Let q_i be the denominators of the convergents of β_1, and let l be the smallest number such that $q_l \geq 2^j$. The geometric grows of denominators (8.6) implies that $2^{j+r_0} < 2^{r_0} q_l \leq q_{l+2r_0}$ so that $\Omega_j \subseteq [q_{l-1}, q_{l+2r_0})$. If we denote $D_s = N \cap [q_{l+s-1}, q_{l+s})$, $s = 0, 1, \ldots, 2r_0$, then

$$\Omega_j \subseteq \bigcup_{s=0}^{2r_0} D_s.$$

Since, by (8.6), $q_{i+1} \le \bar{q} q_i$, there are at most $\bar{q}$ disjoint blocks of length q_{l+s-1} that cover D_s. Applying (8.10) to each of those blocks, we derive, for $k = 1, 2, 3, 4$,

$$\sum_{l \in D_s} \left(\sum_{i=1}^{M} \|l \beta_i\|^2 \right)^{-k} = O\left(e^{6(2k-1)\, M \ln M} \, (q_{l+s-1})^{1+(2k-1)/M} \;\; \ln q_{l+s-1} \right).$$

Note that $q_{l-1} \le 2^j$, so that $q_{l+s-1} \le \bar{q}^s q_{l-1} \le \bar{q}^s 2^j$. Therefore,

$$\begin{aligned} \aleph_k(j, M) &= O\left(\sum_{s=0}^{2r_0} \sum_{l \in D_s} \left(\sum_{i=1}^{M} \|l \beta_i\|^2 \right)^{-k} \right) \\ &= O\left(e^{6(2k-1)\, M \ln M} \sum_{s=0}^{2r_0} (\bar{q}^s 2^j)^{(1+(2k-1)/M)} \;\; \ln(\bar{q}^s \, 2^j) \right) \\ &= O\left(e^{6(2k-1)\, M \ln M} \;\; j \; 2^{j(1+(2k-1)/M)} \right), \quad k = 1, 2, 3, 4, \end{aligned}$$

proving, thus, (6.3). □

Proof of Lemma 5. In what follows, we shall only construct the proof for the term involving b_{jk} since the proof for the term involving $a_{j_0 k}$ is very similar. Denote

$$\Delta_\kappa(j) = \frac{1}{|C_j|} \sum_{m \in C_j} \left[\frac{1}{M} \sum_{l=1}^{M} |g_m(u_l)|^2 \right]^{-2\kappa} \left[\frac{1}{M} \sum_{l=1}^{M} |g_m(u_l)|^{2\kappa} \right], \quad \kappa = 1, 2,$$

where $\tau_1(m)$ is given by (4.2) and (4.6). Note that, by (3.2) and (3.5), we have

$$\widehat{f}_m - f_m = N^{-1/2} M^{-1} \left[\sum_{l=1}^{M} |g_m(u_l)|^2 \right]^{-1} \left(\sum_{l=1}^{M} \overline{g_m(u_l)}\, z_{ml} \right),$$

where z_{ml} are standard Gaussian random variables, independent for different m and l. Therefore, since in the case of Meyer wavelets, $|\psi_{mjk}| \le 2^{-j/2}$ and $|C_j| \asymp 2^j$ (see, e.g., [15], p. 565), we derive that $\mathbb{E}|\widehat{b}_{jk} - b_{jk}|^2$ is given by expression (4.1). If $\kappa = 2$, then

$$\begin{aligned} \mathbb{E}|\widehat{b}_{jk} - b_{jk}|^4 &= O\left(\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^4 \right) + O\left(\left[\sum_{m \in C_j} \mathbb{E}|\widehat{f}_m - f_m|^2 \right]^2 \right) \\ &= O\left(N^{-2} 2^{-2j} M^{-4} [\tau_1(m)]^{-4} \sum_{m \in C_j} \sum_{l=1}^{M} |g_m(u_l)|^4 \right) \\ &\quad + O\left(N^{-2} M^{-2} 2^{-2j} [\tau_1(m)]^{-2} \right) \end{aligned}$$

$$
\begin{aligned}
&= O\left(2^{-j} N^{-2} M^{-3} \Delta_2(j) + N^{-2} M^{-2} \Delta_1^2(j)\right) \\
&= O\left(n^{-2}[M^{-1} 2^{-j} \Delta_2(j) + \Delta_1^2(j)]\right). \qquad (8.11)
\end{aligned}
$$

Now, recall that $|g_m(u_l)| \asymp |m|^{-1} \|m\beta_l\|$ by (4.5). Note that, by formula (6.1), $\aleph_k(j, M)$ is increasing in r_0 and recall that, by the definition of the Meyer wavelet basis, one has $|m| \in [(2\pi/3)2^j, (8\pi/3)2^j] \subset \Omega_j$ with $r_0 = 3 + \log_2(\pi/3)$ (see (3.3) and (6.2)). Then, direct calculations yield

$$
\Delta_1(j) = O(2^j M \aleph_1(j, M)) \quad \text{and} \quad \Delta_2(j) = O(2^{3j} M^4 \aleph_4(j, M)). \qquad (8.12)
$$

To complete the proof, combine (6.3), (4.1), (8.11) and (8.12) and note that $M j^{-1} 2^{-j(1-5/M)} = o(1)$ as $n \to \infty$, since $2^j \geq 2^{j_0} = \ln n$ and $M = M_n \to \infty$ as $n \to \infty$. □

Proof of Lemma 6. It is easy to see that $\widehat{b}_{jk} - b_{jk}$ follows a Gaussian distribution with mean zero and variance bounded by $\frac{C_\psi}{2}(n_M)^{-1} j 2^{j(2+1/M)}$. Hence,

$$
\mathbb{P}\left(|\widehat{b}_{jk} - b_{jk}|^2 \geq \frac{\eta^2}{n_M} j 2^{j(2+1/M)} \ln n\right) \leq 2\Phi\left(\frac{\eta}{\sqrt{C_\psi}} \sqrt{\ln n}\right) = O\left(\frac{n^{-\eta^2/(2C_\psi)}}{\sqrt{\ln n}}\right),
$$

where $\Phi(\cdot)$ is the cumulative distribution function of a Gaussian random variable with mean zero and variance one. □

Proof of Theorem 2. Due to the orthogonality of the Meyer wavelet basis, we obtain

$$
\mathbb{E}\|\hat{f}_n - f\|_2^2 = R_0 + R_1 + R_2 + R_3 + R_4,
$$

where

$$
\begin{aligned}
R_0 &= \sum_{k=0}^{2^{j_0}-1} \mathbb{E}(\widehat{a}_{j_0 k} - a_{j_0 k})^2, \quad R_1 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}^2, \\
R_2 &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}(|\widehat{b}_{jk}| \geq \lambda_j)]\, \mathbb{I}(|b_{jk}| < \lambda_j/2), \\
R_3 &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \mathbb{P}(|\widehat{b}_{jk}| < \lambda_j)\, \mathbb{I}(|b_{jk}| \geq 2\lambda_j), \\
R_4 &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}(|\widehat{b}_{jk}| \geq \lambda_j)]\, \mathbb{I}(|b_{jk}| \geq \lambda_j/2), \\
R_5 &= \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \mathbb{P}(|\widehat{b}_{jk}| < \lambda_j)\, \mathbb{I}(|b_{jk}| < 2\lambda_j).
\end{aligned}
$$

Denote

$$
\zeta(s, M) = \frac{2(3 + 1/M)}{2s + 3 + 1/M}
$$

and observe that $\zeta(s,M) < 2$ for $s > 1/\min(p,2)$. First, consider the terms R_0 and R_1. Using Lemma 5, it is easily seen that

$$\begin{aligned} R_0 = O\left(n^{-1} 2^{j_0} \aleph_1(j_0,M)\right) &= o\left(n^{-1}\, j_0\, 2^{j_0(2+1/M)}\ e^{6M\ln M}\right) \\ &= o((M\, n_M)^{-1} \ln^3 n) = o\left((n_M)^{-\frac{2s}{2s+3+1/M}}\right). \end{aligned}$$

Furthermore, it is well-known (see, e.g., [14], Lemma 19.1) that if $f \in B^s_{p,q}(A)$, then for some positive constant c^*, dependent on p, q, s and A only, we have $\sum_{k=0}^{2^j-1} b_{jk}^2 \le c^* 2^{-2js^*}$ and, thus,

$$R_1 = O\left(2^{-2Js^*}\right) = O\left((n_M)^{-2s^*/(3+1/M)}\right).$$

By direct calculations, one can check that if $2 \le p \le \infty$, then $s^* = s$ and hence

$$R_1 = o\left((n_M)^{-2s/(2s+3+1/M)}\right).$$

On the other hand, if $1 \le p < 2$ then $s^* = s + 1/2 - 1/p$. If $\zeta(s,M) \le p < 2$ then $2s^*/(3+1/M) \ge 2s/(2s+3+1/M)$ and, hence,

$$R_1 = O\left((n_M)^{-2s/(2s+3+1/M)}\right).$$

Similarly, if $1 \le p < \zeta(s,M)$, then $2s^*/(3+1/M) \ge 2s^*/(2s^*+2+1/M)$ and therefore

$$R_1 = O\left((n_M)^{-2s^*/(2s^*+2+1/M)}\right).$$

Now, consider the term R_2. Using Lemma 5 and Lemma 6 with $\theta \ge 2$, formula (6.4), and the fact that $e^{M\ln M} = o(n^a)$ for any $a > 0$ as $n \to \infty$, after some simple algebra, one derives

$$\begin{aligned} R_2 &\le \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\widehat{b}_{jk} - b_{jk})^2 \mathbb{I}(|\widehat{b}_{jk} - b_{jk}| \ge \lambda_j/2)\right] \\ &\le \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \sqrt{\mathbb{E}[(\widehat{b}_{jk} - b_{jk})^4}\sqrt{\mathbb{P}(|\widehat{b}_{jk} - b_{jk}|^2 \ge \lambda_j^2/4)} \\ &= O\left(\sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \frac{M e^{21\, M\ln M}\ j 2^{j(2+1/M)}}{n^{1+\theta}}\right) \\ &= O\left(\frac{2^{J(3+1/M)}\ln n\ e^{15\, M\ln M}}{n_M\, n^\theta}\right) = O\left((n_M)^{-1}\right). \end{aligned}$$

For the term R_3, again applying Lemma 6 with $\theta \ge 2$, obtain

$$R_3 \quad \le \quad \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \mathbb{P}(|\widehat{b}_{jk} - b_{jk}| \ge \lambda_j/2) = o\left(\sum_{j=j_0}^{J-1} 2^{-2js^*} n^{-\theta}\right) = o\left(n^{-1}\right).$$

Now, consider the term R_4. Let j_1 be such that

$$2^{j_1} = O\left((n_M)^{1/(2s+3+1/M)} (\ln n)^{\xi_0} \right),$$

for some real number ξ_0. First, consider the case when $p > \zeta(s, M)$. Then,

$$R_4 \quad \leq \quad \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \, \mathbb{I}(|b_{jk}| \geq \lambda_j/2) = R_{41} + R_{42},$$

where

$$\begin{aligned} R_{41} &= \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \, \mathbb{I}(|b_{jk}| \geq \lambda_j/2), \\ R_{42} &= \sum_{j=j_1+1}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}[(\widehat{b}_{jk} - b_{jk})^2 \, \mathbb{I}(|b_{jk}| \geq \lambda_j/2). \end{aligned}$$

Then, Lemma 5 yields

$$\begin{aligned} R_{41} &= O\left(\sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (n_M)^{-1} \; j \; 2^{j(2+1/M)} \right) \\ &= O\left((n_M)^{-2s/(2s+3+1/M)} (\ln n)^{1+\xi_0(3+1/M)} \right). \end{aligned}$$

For term R_{42}, one derives

$$\begin{aligned} R_{42} &= O\left(\sum_{j=j_1+1}^{J-1} \sum_{k=0}^{2^j-1} (n_M)^{-1} \; j \; 2^{j(2+1/M)} \; \frac{|b_{jk}|^p}{|\lambda_j|^p} \right) \\ &= O\left((n_M)^{-(1-p/2)} (\ln n)^{1-p} \sum_{j=j_1+1}^{J-1} 2^{j[(2+1/M)(1-p/2)-s^* p]} \right) \\ &= O\left((n_M)^{-\rho_1} (\ln n)^{\rho_2} \right), \end{aligned}$$

where $\rho_1 = -2s/(2s+3+1/M)$ and $\rho_2 = 1-p-\xi_0 \, [ps^* - (2+1/M)(1-p/2)]$. Now, choosing w $\xi_0 = -2/(2s+3+1/M)$, and combining the above terms, one easily arrives at

$$R_4 = o\left((n_M)^{-\frac{2s}{2s+3+1/M}} (\ln n)^{\frac{2s}{2s+3+1/M}} \right).$$

Now, consider the case when $1 \leq p < \zeta(s, M)$. Note that, same as above,

$$R_{41} = O\left((n_M)^{-2s/(2s+3+1/M)} \; (\ln n)^{1+\xi_0(3+1/M)} \right);$$

but ξ_0 does not need to be the value chosen above. Observe that, since for R_4 one has $|b_{jk}| \leq c^* 2^{-js^*}$ and $|b_{jk}| > \lambda_j/2$, then, combination of these inequalities requires $j \le j_2$ where j_2 satisfies $j_2\ 2^{j_2} = O\left(n_M/\ln n\right)^{\frac{1}{2s^*+2+1/M}}$. Then, $|b_{jk}| \le \lambda_j/2$ if $j \ge j_2+1$ and

$$\begin{aligned} R_{42} &= O\left((n_M)^{-(1-p/2)}\ (\ln n)^{1-p}\ 2^{j_2\,[(2+1/M)(1-p/2)-s^*p]}\right) \\ &= O\left((n_M)^{\rho_3}(\ln n)^{\rho_4}\right), \end{aligned}$$

where $\rho_3 = -2s^*/(2s^*+2+1/M)$ and $\rho_4 = 2s^*/(2s^*+2+1/M) - p/2 - [(2+1/M)(1-p/2) - ps^*]$. Noting that, in this case, $s/(2s+3+1/M) - s^*/(2s^*+2+1/M) > 0$ and one arrives at $R_{41} = o(R_{42})$ as $n \to \infty$. Therefore,

$$\begin{aligned} R_4 &= O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}}\ (\ln n)^{-\frac{p}{2}-[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right) \\ &= O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}}\right), \end{aligned}$$

since the power of $\ln n$ in the expression above is negative.

Finally, consider the term R_5. First, consider the case when $\zeta(s,M) \le p < 2$. Let j_3 be such that

$$2^{j_3} = O\left((n_M)^{\frac{s}{s^*(2s+3+1/M)}}(\ln n)^{\xi_1}\right),$$

for some real number ξ_1. Then,

$$R_5 \le \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} b_{jk}^2\, \mathbb{I}(|b_{jk}| < 2\lambda_j) \le R_{51} + R_{52},$$

where

$$\begin{aligned} R_{51} &= \sum_{j=j_3+1}^{J_1}\sum_{k=0}^{2^j-1} b_{jk}^2 = O\left((n_M)^{-\frac{2s}{2s+3+1/M}}(\ln n)^{-2s^*\xi_1}\right), \\ R_{52} &= \sum_{j=j_0}^{j_3}\sum_{k=0}^{2^j-1} b_{jk}^2\, \mathbb{I}(|b_{jk}| < 2\lambda_j). \end{aligned}$$

Let

$$\Xi(j) = \sum_{k=0}^{2^j-1} b_{jk}^2\ \mathbb{I}(|b_{jk}| < 2\lambda_j).$$

Note that

$$\Xi(j) = O\left(2^j\lambda_j^2\right) = O\left(j\ 2^{j(3+1/M)}\ln n\ (n_M)^{-1}\right)$$

and also

$$\Xi(j) = O\left(\sum_{k=0}^{2^j-1} |b_{jk}|^p\ |b_{jk}|^{2-p}\ \mathbb{I}(|b_{jk}| < 2\lambda_j)\right) = O\left(\lambda_j^{2-p}\ 2^{-jps^*}\right)$$

$$= \quad O\left((n_M)^{p/2-1}\ (\ln n)^{1-p/2}\ j^{1-p/2}\ 2^{j[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right).$$

Let j_4 be such that

$$2^{j_4} = O\left((n_M)^{\frac{1}{2s+3+1/M}}(\ln n)^{\xi_2}\right),$$

for some real number ξ_2. Then

$$\begin{aligned}
R_{52} &= \sum_{j=j_0}^{j_4} \Xi(j) + \sum_{j=j_4+1}^{j_3} \Xi(j) \\
&= O\left(\sum_{j=j_0}^{j_4} j\ 2^{j(3+1/M)}\ \ln n\ (n_M)^{-1}\right) \\
&\quad + O\left(\sum_{j=j_4+1}^{j_3} (n_M)^{p/2-1}(\ln n)^{1-p/2}\ j^{1-p/2}\ 2^{j[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right) \\
&= O\left((n_M)^{-\frac{2s}{2s+3+1/M}}(\ln n)^{2+\xi_2(3+1/M)}\right) \\
&\quad + O\left((n_M)^{-\frac{2s}{2s+3+1/M}}\ (\ln n)^{2-p+\xi_2[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right).
\end{aligned}$$

Since the bound for R_{52} is valid for any value of ξ_2, we choose ξ_2 which minimizes

$$\max(2+\xi_2(3+1/M),\ 2-p+\xi_2[(2+1/M)(1-p/2)-ps^*]),$$

i.e., $\xi_2 = [s^*+1+1/p+1/(2M)]^{-1}\ (2s^*+2/p-1)$. Hence

$$R_{52} = O\left((n_M)^{-\frac{2s}{2s+3+1/M}}(\ln n)^{\frac{2s^*+2/p-1}{s^*+1+1/p+1/(2M)}}\right).$$

Choose now $\xi_1 = -2/(2s+3+1/M)$. Then, combining the R_{51} and R_{52} terms, obtain

$$R_5 = O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s}{2s+3+1/M}}\ (\ln n)^{\frac{2s}{2s+3+1/M}}\right).$$

Now, consider the case when $1 \le p < \zeta(s,M)$. Let j_5 be such that

$$2^{j_5} = O\left((\ln n/n_M)^{\frac{1}{2s^*+2+1/M}}\ (\ln n)^{\xi_3}\right),$$

for some real number ξ_3. Then

$$R_5 \le R_{51} + R_{52} + R_{53},$$

where

$$R_{51} = \sum_{j=j_5+1}^{J-1}\sum_{k=0}^{2^j-1} b_{jk}^2, \quad R_{52} = \sum_{j=j_0}^{j_4} \Xi(j), \quad R_{53} = \sum_{j=j_4+1}^{j_5} \Xi(j).$$

It is immediate that

$$R_{51} = O\left(\sum_{j=j_5+1}^{J-1} 2^{-2js^*}\right) = O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}} (\ln n)^{-2s^*\xi_3}\right).$$

and that

$$R_{52} = O\left(\sum_{j=j_0}^{j_4} \frac{j\ 2^{j(3+1/M)}\ln n}{n_M}\right) = o\left((n_M)^{-\frac{2s^*}{2s^*+2+1/M}}\right).$$

After some simple algebra, one obtains

$$\begin{aligned} R_{53} &= O\left(\sum_{j=j_4+1}^{j_5} (n_M)^{p/2-1}\ (\ln n)^{1-p/2}\ j^{1-p/2}\ 2^{j[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right) \\ &= O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}}\ (\ln n)^{1-p/2+\xi_3[(2+\frac{1}{M})(1-\frac{p}{2})-ps^*]}\right). \end{aligned}$$

Choosing $\xi_3 = -1/(2s^*+2+1/M)$, and combining the above terms, we arrive at

$$R_5 = O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}}\ (\ln n)^{\frac{2s^*}{2s^*+2+1/M}}\right).$$

Finally, consider the case, $2 \le p \le \infty$. In this case, $j_3 = j_4$, and we easily see that

$$R_5 = O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s}{2s+3+1/M}}\right).$$

Combining all the above expressions, we obtain that, as $n \to \infty$,

$$R_n(\hat{f}_n) = \begin{cases} O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s}{2s+3+1/M}}\right), & \text{if} \quad 2 \le p \le \infty, \\ O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s}{2s+3+1/M}}\right)(\ln n)^{\frac{2s}{2s+3+1/M}}, & \text{if} \quad \zeta(s,M) \le p < 2, \\ O\left(\left(\frac{\ln n}{n_M}\right)^{\frac{2s^*}{2s^*+2+1/M}}\right)(\ln n)^{\frac{2s^*}{2s^*+2+1/M}}, & \text{if} \quad 1 \le p < \zeta(s,M). \end{cases} \tag{8.13}$$

Now, note that $6/(2s+3) < \zeta(s,M)$ for any $M > 0$. Hence, if $p \le 6/(2s+3)$, then $p < \zeta(s,M)$. On the other hand, if $p > 6/(2s+3)$, then it is easy to show that for M large enough one has $p > \zeta(s,M)$. Observe also that $p > 6/(2s+3)$ if and only if $s > 3(1/p - 1/2)$.

The upper bound in (8.13) depends on the choice of $M = M_n$. Choose M_n of the form (6.6). Then, from the definition of n_M and formulae (6.9) and (8.13), it follows that

$$R_n(\hat{f}_n) = O\big(\exp\big\{-(A_2+1/M)^{-1}[A_1\ln(n - 6M\ln M - \ln M) - A_3\ln\ln n]\big\}\big). \tag{8.14}$$

Using Taylor expansion, we write $(A_2+1/M)^{-1} = A_2^{-1} - M^{-1}A_2^{-2} + M^{-2}A_2^{-3} + O(M^{-3})$ as $M \to \infty$. Recalling that $\ln M = \ln\nu + 0.5\,\ln\ln n - 0.5\,\ln\ln\ln n$ and plugging expressions for M, $\ln M$ and $(A_2+1/M)^{-1}$ into the argument of exponent in (8.14), by direct calculations, one derives that

$$R_n(\hat{f}_n) = O\big(\exp\left(-(A_2)^{-1}\,A_1\ \ln n + \Delta_n\right)\big),$$

where, as $n \to \infty$,

$$\Delta_n = \sqrt{\ln n}\,\sqrt{\ln\ln n}\left(\frac{A_1}{A_2}\left[3\nu + \frac{1}{A_2\nu}\right]\right) - \frac{3A_1\nu\,\sqrt{\ln n}\ \ln\ln\ln n}{A_2\,\sqrt{\ln\ln n}}\left(\frac{2\,\ln\nu}{\ln\ln\ln n} - 1\right)$$
$$- \ln\ln n\left(\frac{A_3}{A_2} + \frac{A_1}{2A_2} - \frac{3A_1}{A_2^2} - \frac{A_1}{A_2^3\,\nu^2}\right) + \ln\ln\ln n\left(\frac{A_1}{2A_2} - \frac{3A_1}{A_2^2}\right) + O(1).$$

Now, to complete the proof, note that the main term in Δ_n is minimized by $\nu = \nu_{opt} = (3A_2)^{-1/2}$, and that $A_2 \geq 2$ for any $s > 1/\min(p,2)$. $\square$

Acknowledgments

We would like to thank the Editor and the anonymous referee for useful comments and suggestions on improvements to the presentation of the work.

References

[1] Abramovich, F. and Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. MR1627226

[2] Casey, S.D. and Walnut, D.F. (1994). Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms. *SIAM Review* **36** 537–577. MR1306923

[3] Cassels, J.W.S. (1955). Simultaneous Diophantine approximation II. *Proceedings of the London Mathematical Society* **5** 435–448. MR0075240

[4] Cavalier, L. and Raimondo, M. (2007). Wavelet deconvolution with noisy eigenvalues. *IEEE Transactions on Signal Processing* **55** 2414–2424. MR1500172

[5] Chesneau, C. (2008). Wavelet estimation via block thresholding: a minimax study under L^p-risk. *Statistica Sinica* **18** 1007–1024. MR2440401

[6] Davenport, H. (1962). A note on Diophantine approximation. In *Studies in Mathematical Analysis and Related Topics*, (Eds. G. Szegö *et al.*), pp. 77–81, Stanford University Press, Stanford. MR0146145

[7] De Canditiis, D. and Pensky, M. (2004). Discussion on the meeting on “Statistical Approaches to Inverse Problems”. *Journal of the Royal Statistical Society*, Series B **66** 638–640.

[8] De Canditiis, D. and Pensky, M. (2006). Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics* **33** 293–306. MR2279644

[9] Donoho, D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis* **2** 101–126. MR1325535

[10] Donoho, D.L. and Raimondo, M. (2004). Translation invariant deconvolution in a periodic setting. *International Journal of Wavelets, Multiresolution and Information Processing* **14** 415–432. MR2104873

[11] Edixhoven, B. and Evertse, J.-H. (1993). *Diophantine Approximation and Abelian Varieties.* Lecture Notes in Mathematics, Vol. **1566**, Springer-Verlag, Berlin. MR1288998

[12] Gradshtein, I.S. and Ryzhik, I.M. (1980). *Tables of Integrals, Series, and Products.* Academic Press, New York.

[13] Harsdorf, S. and Reuter, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*, Dresden/FRG, June 16–17.

[14] Johnstone, I.M. (2002). *Function Estimation in Gaussian Noise: Sequence Models.* Unpublished Monograph. (`http://www-stat.stanford.edu/~imj/`)

[15] Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004) Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, Series B **66** 547–573 (with discussion, 627–657). MR2088290

[16] Johnstone, I.M. and Raimondo, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Annals of Statistics* **32** 1781–1804. MR2102493

[17] Kalifa, J. and Mallat, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics* **31** 58–109. MR1962500

[18] Kerkyacharian, G., Picard, D. and Raimondo, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statistica Sinica* **7** 317–340. MR2352512

[19] Lang, S. (1966) *Introduction to Diophantine Approximations.* Springer-Verlag, New York. MR1348400

[20] Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing*, 2nd Edition, Academic Press, San Diego. MR2479996

[21] Masser, D., Nesterenko, Yu.V., Schlickewei, H.P., Schmidt, W.M. and Waldschmidt, M. (2003). *Diophantine Approximation.* Lecture Notes in Mathematics, Vol. **1819**, Springer-Verlag, Berlin.

[22] Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge. MR1228209

[23] Neelamani, R., Choi, H. and Baraniuk, R. (2004). Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing* **52** 418–433. MR2044455

[24] Park, Y.J., Dho, S.W. and Kong, H.J. (1997). Deconvolution of long-pulse lidar signals with matrix formulation. *Applied Optics* **36** 5158–5161.

[25] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic case: uniform case. *Annals of Statistics* **37** 73–104. MR2488345

[26] Pensky, M. and Sapatinas, T. (2010). On convergence rates equivalency and sampling strategies in functional deconvolution models. *Annals of Statistics* **38** 1793–1844. MR2662360

[27] Schmidt, W. (1969). Badly approximable systems of linear forms. *Journal of Number Theory* **1** 139–154. MR0248090

[28] Schmidt, W. (1980). *Diophantine Approximation.* Lecture Notes in Mathematics, Vol. **785**, Springer-Verlag, Berlin. MR0568710

[29] Schmidt, W. (1991). *Diophantine Approximations and Diophantine Equations.* Lecture Notes in Mathematics, Vol. **1467**, Springer-Verlag, Berlin. MR1176315