

# Wavelet Estimation: Minimax Theory and Application

Sam Efromovich\* and Ekaterina Smirnova

Department of Mathematical Sciences,  
The University of Texas at Dallas, Richardson, Texas, USA

\*Corresponding Author: efrom@utdallas.edu

Received: 22, May 2013 / Revised: 27, January 2014 / Accepted: 12, February 2014

## ABSTRACT

*For now wavelet estimation is an important part of the nonparametric curve estimation. The minimax theory is well developed and there are many interesting practical applications where wavelets perfectly fit purposes of the data analysis. This paper presents a new theoretical result and a new application. It is well known in the statistical wavelet theory that for a Besov space of order  $\alpha$  the minimax rate of the mean integrated squared error (MISE) convergence is  $n^{-2\alpha/(2\alpha+1)}$ . The developed in the paper local minimax approach shows that this rate is too slow and that any Besov's function of order  $\alpha$  can be estimated with a faster rate. Furthermore, an adaptive estimator is proposed that is minimax under the classical and new minimax approaches. Then an interesting and challenging application of the adaptive wavelet estimator in the fMRI study of neuroplasticity is discussed.*

**Keywords:** Adaptation, Local minimax, Oracle, Regression, Neuroplasticity.

## 1. Introduction

Consider a classical nonparametric regression problem where based on  $n$  observations

$$Y_l = f(l/n) + \sigma\epsilon_l, \quad l = 1, 2, \dots, n \quad (1.1)$$

the problem is to estimate a square-integrable on  $[0, 1]$  function  $f(t)$ . Here  $\epsilon_l$  are independent standard normal random variables,  $\sigma$  is an unknown positive constant which may be referred as the scale ( $\sigma^2$  is also called variance or volatility depending on a particular application), and the quality of an estimate  $\tilde{f}$  of  $f$  is measured by the mean integrated squared error (MISE) defined as  $E\{\int_0^1 (\tilde{f}(t) - f(t))^2 dt\}$ .

We make a traditional in the wavelet literature assumption that the estimated function (signal)  $f$  belongs to a Besov space  $B_{pqQ}^\alpha$  of order  $\alpha$  which will be defined shortly. Note that the familiar Hölder space of order  $\alpha$  corresponds to  $B_{\infty\infty Q}^\alpha$ . The book by Meyer (1992) is a good reference on Besov spaces.

Suppose that an (inhomogeneous) wavelet basis on  $[0, 1]$  is given,  $\{\phi_{j_0 k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k), k = 0, 1, \dots, 2^{j_0} - 1\}$  and  $\{\psi_{jk}(x) = 2^{j/2}\psi(2^jx - k), k = 0, 1, \dots, 2^j - 1, j \geq j_0\}$ . Here  $\phi$  and  $\psi$  are the scaling (father) function and mother wavelet, respectively. Then, for any positive integer  $j_0$  the estimated function has a formal orthogonal expansion (called the wavelet multiresolution expansion)

$$f(t) = \sum_{k=0}^{2^{j_0}-1} \kappa_{j_0 k} \phi_{j_0 k}(t) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \theta_{jk} \psi_{jk}(t) \quad (1.2)$$

where wavelet coefficients are defined by  $\kappa_{j_0 k} = \int_0^1 f(t)\phi_{j_0 k}(t)dt$  and  $\theta_{jk} = \int_0^1 f(t)\psi_{jk}(t)dt$ .

Customary estimates of the wavelet coefficients, recommended in the literature, are empirical wavelet coefficients  $\hat{\kappa}_{jk} = n^{-1} \sum_{l=1}^n Y_l \phi_{j_0 k}(l/n)$  and  $\hat{\theta}_{jk} = n^{-1} \sum_{l=1}^n Y_l \psi_{jk}(l/n)$ , which are in practice are calculated using a cascade algorithm supported by all wavelet software packages (see a discussion in Efromovich 1999, Vidakovic 1999 and Nason 2008).

Now we are in a position to remember the important characterization of a Besov function space  $B_{pqQ}^\alpha$  of order  $\alpha$  via wavelet coefficients. Namely, if the following classical restrictions, that the wavelet regularity is larger than  $\alpha$ ,  $p \geq 1$ ,  $q \geq 1$ , and  $\alpha + 1/2 - 1/p > 0$ , hold, then

$$B_{pqQ}^\alpha = \left\{ f : f(t) = \sum_{k=0}^{2^{j_0}-1} \kappa_{j_0 k} \phi_{j_0 k}(t) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \theta_{jk} \psi_{jk}(t), \quad (1.3) \right.$$

$$\left. \left( \sum_{k=0}^{2^{j_0}-1} |\kappa_{j_0 k}|^p \right)^{1/p} + \left( \sum_{j \geq j_0} (2^{j(\sigma+1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |\theta_{jk}|^p \right)^{1/p} \right)^q \right)^{1/q} \leq Q < \infty \right\}. \quad (1.4)$$

It is known that the global minimax (over a Besov class of order  $\alpha$ ) rate of the MISE convergence is  $n^{-2\alpha/(2\alpha+1)}$ , that is

$$\inf_{\tilde{f}} \sup_{f \in B_{pqQ}^\alpha} E \left\{ \int_0^1 (\tilde{f}(t, Y^n, \alpha, p, q, Q) - f(t))^2 dx \right\} \geq C^* n^{-2\alpha/(2\alpha+1)}, \quad (1.5)$$

where the infimum is taken over all possible estimators  $\tilde{f}$  based on  $n$  observations  $Y^n := (Y_1, \dots, Y_n)$  defined in (1.1) and parameters of the Besov space, and  $C^*$  is a positive constant. Furthermore, there exists a data-driven estimator, based solely on observations  $Y^n$ , whose MISE attains the lower bound (1.5) up to a constant factor. Lower bound (1.5) is the foundation of the modern wavelet literature and it is the main aim of all wavelet estimators to attain this rate of the MISE convergence; see a discussion in Efromovich (1999) and Nason (2008). So far no methodology challenging this minimax approach has been proposed. In this paper we explore the following natural question. In a practical application there is always a single underlying regression function. It is well known that there is no feasible lower bound for the MISE when a single function is estimated because a “lucky” estimate, which is equal to the underlying function, has zero MISE. The latter is the reason why the global minimax over a Besov space is considered. Nonetheless, can we consider a subfamily of Besov functions of order  $\alpha$  that are close in  $L_\infty$ -norm to a single pivotal function and show that the local minimax over this subfamily may increase the rate of the MISE convergence? In other words, can a local minimax methodology challenge the classical rate  $n^{-2\alpha/(2\alpha+1)}$ ? The answer is “yes” and this issue constitutes the main theoretical part of the paper presented in Section 2. Section 3 is devoted to an interesting new application of the developed wavelet estimator in the study of neuroplasticity based on fMRI data. Proofs are deferred to Section 4.

## 2. Local Minimax Theory

We begin with definition of the proposed local Besov space of order  $\alpha$  with the pivot  $f_0$ ,

$$\begin{aligned}
 B_{pq}^\alpha(f_0) := & \left\{ f : f(t) = \sum_{k=0}^{2^{j_0}-1} v_{j_0 k} \phi_{j_0 k}(t) \right. \\
 & + \sum_{j_0 \leq j \leq (2\alpha+1)^{-1} \log_2(\lambda_n n)} \sum_{k=0}^{2^j-1} u_{jk} \psi_{jk}(t) \\
 & \left. + \sum_{j > (2\alpha+1)^{-1} \log_2(\lambda_n n)} \sum_{k=0}^{2^j-1} \theta_{jk} \psi_{jk}(t); \right. \quad (2.1)
 \end{aligned}$$

$$v_{j_0 k} := \int_0^1 f_0(t) \phi_{j_0 k}(t) dt, \quad u_{jk} := \int_0^1 f_0(t) \psi_{jk}(t) dt, \quad (2.2)$$

$$\theta_{jk} := \int_0^1 f(t)\psi_{jk}(t)dt; \tag{2.3}$$

$$\left( \sum_{k=0}^{2^{j_0}-1} |v_{j_0k}|^p \right)^{1/p} + \left( \sum_{j \geq j_0} (2^{j(\alpha+1/2-1/p)} \left( \sum_{k=0}^{2^j-1} |u_{jk}|^p \right)^{1/p} \right)^q \right)^{1/q} =: Q_0 < \infty; \tag{2.4}$$

$$\sum_{j>s} \sum_{k=0}^{2^j-1} u_{jk}^2 > 0, \quad s = j_0, j_0 + 1, \dots; \tag{2.5}$$

$$f \in B_{pqQ_0}^\alpha; \tag{2.6}$$

$$\lambda_n := \max \left\{ \lambda : \lambda \leq \left[ \sum_{j > (2\alpha+1)^{-1} \log_2(\lambda n)} (2^{j(\alpha+1/2-1/p)}) \times \left( \sum_{k=0}^{2^j-1} |u_{jk}|^p \right)^{1/p} \right]^{2/q} \right\}. \tag{2.7}$$

Let us comment on the proposed local Besov class. Lines (2.1)-(2.3) indicate that we consider functions that on the coarsest scales (up to the scale  $(2\alpha + 1)^{-1} \log_2(\lambda_n n)$ ) are equal to the pivot  $f_0$ . Line (2.4) implies a natural restriction on the pivot which should belong to the Besov space  $B_{pqQ_0}^\alpha$ . Please note that  $Q_0$  is the exact parameter of this Besov space corresponding to the pivot  $f_0$ . Line (2.5) is another restriction on the pivot, and it states that the pivot cannot be a parametric function which is defined by a finite number of wavelet coefficients. Line (2.6) is the main restriction on the considered Besov's functions. It states that all functions from the local class must belong to the same Besov space as the pivot does. This is a natural restriction. Finally, line (2.7) defines the boundary between the coarsest scales, where considered Besov's functions are identical to the pivot, and the finest scales where we have a class of possible functions whose wavelet coefficients must satisfy the restriction (2.6).

Now let us explain why the class  $B_{pq}^\alpha(f_0)$  can be called local. It follows from (2.5) and (2.7) that

$$\lambda_n = o_n(1), \quad [\lambda_n n]^{-1} = o_n(1). \tag{2.8}$$

This result, together with (2.1) and (2.6), implies that  $f$  converges to  $f_0$  in  $L_p$ -norm as  $n \rightarrow \infty$  uniformly over all  $f \in B_{pq}^\alpha(f_0)$ . Furthermore, if  $\alpha > 1/2$  then

we have even stronger convergence of  $f$  to a pivot in  $L_\infty$ -norm. This explains why the class (2.7) is called local.

Let us formulate a lower bound for the local Besov space which is analog of the global minimax lower bound (1.5).

**Theorem 2.1** (Lower Bound for Local Besov Space). *Consider regression model (1.1) and a local Besov space (2.7). Suppose that  $p \geq 1$ ,  $q \geq 1$  and  $\alpha + 1/2 - 1/p > 0$ . Then the following lower bound for the local minimax MISE holds,*

$$\inf_{\check{f}} \sup_{f \in B_{pq}^\alpha(f_0)} E \left\{ \int_0^1 (\check{f}(t, Y^n, f_0, \alpha, p, q) - f(t))^2 dt \right\} \geq C_* [\lambda_n^{-1/(2\alpha)} n]^{-2\alpha/(2\alpha+1)}, \quad C_* > 0, \quad (2.9)$$

where the infimum is taken over all estimates  $\check{f}(t, Y^n, f_0, \alpha, p, q)$  based on data  $Y^n$  and parameters  $f_0, \alpha, p, q$  of the local Besov space, and  $\lambda_n$  is defined in (2.7).

What is the practical impact of the above-formulated lower bound? Because for each pivot  $f_0$  the sequence  $\lambda_n$  vanishes as  $n \rightarrow \infty$ , we have a lower bound which indicates the possibility for a faster MISE decrease than the classical minimax lower bound (1.5) indicates. As a result, if we are able to propose an estimator that attains the lower bound (2.9), then this new lower bound dominates the classical (1.5).

Let us consider a wavelet estimator which belongs to a class of blockwise-shrinkage estimators discussed in Efromovich (1999, 2000a). Set

$$\hat{f}(t) := \sum_{k=0}^{2^{j_0}-1} \hat{\kappa}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{J^*} \sum_{s=1}^{2^j/L_{jn}} \frac{\hat{\Theta}_{js}}{\hat{\Theta}_{js} + \hat{\sigma}^2 n^{-1}} I(\hat{\Theta}_{js} > \nu_{jn} n^{-1}) \sum_{k \in T_{jsn}} \hat{\theta}_{jk} \psi_{jk}(t). \quad (2.10)$$

Here  $\hat{\kappa}_{j_0 k}$  and  $\hat{\theta}_{jk}$  are empirical wavelet coefficients defined below line (1.2) in the Introduction,  $\hat{\Theta}_{js} = L_{jn}^{-1} \sum_{j \in T_{jsn}} (\hat{\theta}_{jk}^2 - \hat{\sigma}^2 n^{-1})$ ,  $I(\cdot)$  is the indicator,  $T_{jsn} = \{k : L_{jn}(s-1) \leq k < L_{jn}s\}$ ,  $L_{jn}$  is the length of blocks on  $j$ th resolution scale,  $J^* = \lfloor \log_2 n \rfloor$  and  $\lfloor x \rfloor$  is the rounded down  $x$ , and

$$\hat{\sigma}^2 := n[(1.48)\text{median}(\{\hat{\theta}_{J^* k}, k = 0, \dots, 2^{J^*}\})]^2. \quad (2.11)$$

The estimate of  $\sigma^2$  is the robust estimate based on the sample median for wavelet coefficients on the finest scale; see its discussion in Efromovich (1999). There is a number of appropriate lengths  $L_{jk}$  of blocks  $T_{j_s n}$  and threshold levels  $\nu_{jk}$  that may be used to attain the lower bound (2.9). For instance, we can use classical increasing blocks of the length  $L_{jn} = b_n 2^{\lfloor (j-j_0)/3 \rfloor}$  and  $\nu_{jn} = 1$ , here  $b_n$  is the largest dyadic (i.e.,  $2^k, k = 1, 2, \dots$ ) number which is at most  $\log_2(n)$ . Another possibility is to use “constant” blocks and thresholds, for instance  $L_{jn} = b_n$  and  $\nu_{jk} = 4$ . The following result holds for these two particular examples of the blockwise-shrinkage estimator.

**Theorem 2.2** (Upper Bound). *Consider regression model (1.1), a local Besov space (2.1) and a Besov space (1.4) with parameters  $p \geq 2, q \geq 1$  and  $\alpha > 1/2$ . Then the MISE of blockwise-shrinkage estimator (2.10) satisfies the following two upper bounds: For a local Besov space*

$$\begin{aligned} & \sup_{f \in B_{pq}^\alpha(f_0)} E\left\{ \int_0^1 (\hat{f}(t) - f(t))^2 dt \right\} \\ & \leq C^* [\lambda_n^{-1/2\alpha} n]^{-2\alpha/(2\alpha+1)}, \quad C^* < \infty, \end{aligned} \quad (2.12)$$

where  $\lambda_n$  is defined in (2.7); For a Besov space

$$\sup_{f \in B_{pqQ}^\alpha} E\left\{ \int_0^1 (\hat{f}(t) - f(t))^2 dt \right\} \leq C^{**} n^{-2\alpha/(2\alpha+1)}, \quad C^{**} < \infty. \quad (2.13)$$

We conclude that the lower bound (2.9) is bona fide and it is attainable by a data-driven estimator. This result supports our above-made conjecture that the classical minimax rate  $n^{-2\alpha/(2\alpha+1)}$  of the MISE convergence is too slow due to the global nature of the minimax, and we can replace this rate by a faster one if the local minimax is considered. Theorem 2.2 also indicates that the same wavelet estimator can be both locally and globally minimax.

The obtained result resembles the phenomenon of superefficiency discussed in Efromovich (1999). Superefficiency is the property of an estimator to exhibit faster rates of convergence than the global minimax indicates. As a result, superefficiency is a curvewise property which is not completely surprising because the global minimax cares only about the most difficult regression functions for estimation, and those functions change with  $n$ . The proposed local minimax methodology is innovative and different from the superefficiency

which just states the fact that the rate may be faster. Indeed, the local minimax explains that a Besov's function of order  $\alpha$  is estimated with the rate  $\lambda_n^{1/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)}$  where  $\lambda_n = o_n(1)$  and this sequence is explicitly defined in (2.7).

### **3. Application: Neuroplasticity**

Neuroplasticity is the lifelong ability of the brain to reorganize neural pathways based on new experiences and learning. Functional MRI is widely used to investigate brain networks and develop possible treatment for patients with stroke, Alzheimer disease, traumatic brain injury, multiple sclerosis, and schizophrenia, see a discussion in Dimyan and Cohen (2011), Greicius et al. (2004) and Tung et al. (2013).

Researchers at the Southwestern Medical School explored the possibility of using alterations in resting brain activity for finding potential markers for brain plasticity; details of the study can be found Tung et al. (2013) and here we just briefly describe them. Using 3T MR scanner, blood-oxygenation-level-dependent (BOLD) signal fMRI data were obtained for healthy right-handed adults. The motor cortex model studies were based on a five minute pre-training session, followed by a 23 minute training session, and finished with a 5 minute post-training run. During the training session, a subject was shown a white cross-hair and was instructed to press a button three times with the thumb when a change in color was detected. The color change occurred randomly every 27-32 seconds, and there were a total of 40 trials during the motor task period. The data from the experiment raw data consists of 300 time points for each of pre and post training sessions, and four sessions of 340 time points during the training part; the data was measured every second. All data was transformed into Talairach coordinates using AFNI program and standardized mask ROI applied to all subjects to select the motor cortex voxels' time series. This preprocessed data was used in our analysis.

The aim of the experiments is to determine the changes in resting state motor cortex network following the button clicking training. The increased correlation between the right and left motor cortices in post-training resting state compared to pre-training resting state is a popular indicator of training efficiency. A popular approach in fMRI data statistical analysis is to average the voxels' time series data over all the ROI for each hemisphere and work with a single averaged time course signal. Following this method, for each subject the data

would consist of the right and left motor cortex hemisphere time series during pre-training and post-training resting state. The correlation is computed using Pearson cross correlation for both pre-training and post-training averaged signals. To test for statistical significance, two-tail paired Student's t-test is used for comparisons of pre-training and post-training correlation indices. Results, presented in Tung et al. (2013), indicate that there is a statistically significant increase in the correlation during the post-training resting state.

Clearly, the averaging over all voxels in ROI smooths the time series signal and avoids a problem of multiple comparisons adjustment. On the other hand, in each ROI slice there are 80-130 voxels in the left and right motor cortex areas and there are 11 slices where ROI is clearly visible on fMRI scan. Can this information be used to shed a new light on the neural plasticity?

The proposed solution, based on the developed wavelet estimator, is as follows. The correlation between two functions  $f_1(t)$  and  $f_2(t)$ , observed in corresponding regressions (1.1), can be calculated as the correlation between the proposed wavelet estimates  $\hat{f}_1(t)$  and  $\hat{f}_2(t)$ . This step takes care about extremely large noise in fMRI data. As a result, because the large noise is no longer an issue, the averaging is not needed and we may treat data on a voxel level. Furthermore, using an appropriate thresholding for calculated correlations allows us to adjust for multiple comparisons because, as it has been explained earlier, we are dealing with the case when a number of comparisons (pairs of voxels) is much larger than the number of observations.

Let us present a particular outcome of the study. For 11 fMRI scan slices, where motor cortex area was visible, the correlation matrix for all voxel pairs in right and left motor cortices is calculated. Then for each slice the number of activated pairs (defined as pairs with correlation exceeding a threshold level) is calculated. Results for a particular subject are presented in Table 3.1, and note that a similar outcome has been observed for other subjects. The results clearly indicate that in each slice there are more activated pairs of voxels in the post-training session rather than in the pre-training session.

**Table 3.1:** Numbers of activated pairs of voxels

Slice	1	2	3	4	5	6	7	8	9	10	11
Pre-Traning	47	19	15	12	3	18	22	19	1	2	0
Post-Traning	183	144	196	174	97	75	63	196	103	21	4



We may conclude that the simple clicking exercise does change the brain network. Furthermore, in addition to the conclusion of Tung et al. (2013), that training of the brain increases the correlation between averaged activities in the right and left hemispheres measured by their BOLD signals, we can also conclude that the training dramatically increases the number of voxel pairs with significant correlation. In other words, the training increases not only correlation in activities of the brain network connecting the right and left hemispheres, but the training also increases the number of neurons participating in the brain network. The latter may be of a special interest for finding a cure for the above-mentioned brain diseases because our analysis indicates that new neurons may be trained to substitute disease-affected areas of the brain.

#### 4. Proofs

*Proof of Theorem 2.1.* Denote by  $K := K(n, f_0, \alpha, p, q)$  the smallest dyadic integer larger than  $[\lambda_n n]^{1/(2\alpha+1)}$ , and also set  $J := \log_2(K)$ . The Bessel inequality allows us to write for any estimate  $\check{f}(t)$  considered in Theorem 2.1,

$$E\left\{\int_0^1 (\check{f}(t) - f(t))^2 dt\right\} \geq E\left\{\sum_{k=1}^{K-1} (\check{\theta}_{Jk} - \theta_{Jk})^2\right\}, \quad (4.1)$$

where  $\check{\theta}_{Jk}$  and  $\theta_{Jk}$  are wavelet coefficients of  $\check{f}$  and  $f$ , respectively.

Inequality (4.1) explains the idea of establishing the lower bound (2.9). Namely, we are replacing a local Besov space  $B_{pq}^\alpha(f_0)$  by its  $K$ -dimensional parametric subspace

$$B_K := \left\{ f : f(t) = \sum_{k=0}^{2^{j_0}-1} v_{j_0k} \phi_{j_0k}(t) + \sum_{j_0 \leq j < J} \sum_{k=0}^{2^j-1} u_{jk} \psi_{jk}(t) + \sum_{k=0}^{K-1} \theta_{Jk} \psi_{Jk}(t), \quad K^{\alpha+1/2-1/p} \left( \sum_{k=0}^{K-1} |\theta_{Jk}|^p \right)^{1/p} \leq \lambda_n^{1/2} \right\}. \quad (4.2)$$

In (4.2) notation (2.2)-(2.3) is used.

Our next step is a classical one when the minimax MISE over  $B_K$  is bounded from below by a Bayes MISE with a corresponding distribution of parameters  $\{\theta_{Jk}, k = 0, 1, \dots, K-1\}$ . To do this, we introduce  $K$  independent standard normal random variables  $\zeta_0, \dots, \zeta_{K-1}$ . We begin with proving a pivotal result

which states that for some positive constant  $a$  the probability

$$P\left(K^{\alpha+1/2-1/p}\left(\sum_{k=0}^{K-1} |an^{-1/2}\zeta_k|^p\right)^{1/p} > \lambda_n^{1/2}\right) = o_n(1). \quad (4.3)$$

Note that (4.3) is equivalent to

$$P\left(a^p n^{-p/2} K^{(\alpha+1/2)p-1} \sum_{k=0}^{K-1} |\zeta_k|^p > \lambda_n^{p/2}\right) = o_n(1). \quad (4.4)$$

Using definition of  $K$  and (2.8) we get

$$\begin{aligned} & n^{-p/2} K^{(\alpha+1/2)-1} K \lambda_n^{-p/2} \\ & < n^{-p/2} ([\lambda_n n]^{1/(2\alpha+1)} + 1)^{(\alpha+1/2)p} \lambda_n^{-p/2} = 1 + o_n(1). \end{aligned} \quad (4.5)$$

Set  $a := [(1 - 2\gamma)/E\{|\zeta_1|^p\}]^{1/p}$  where  $\gamma \in (0, 1/4)$ . Using (4.5) and Chebyshev's inequality we get for all sufficiently large  $n$ ,

$$\begin{aligned} & P\left(a^p n^{-p/2} K^{(\alpha+1/2)p-1} \sum_{k=0}^{K-1} |\zeta_k|^p > \lambda_n^{p/2}\right) \leq P\left(a^p K^{-1} \sum_{k=0}^{K-1} |\zeta_k|^p > 1\right) \\ & = P\left(a^p K^{-1} \sum_{k=0}^{K-1} (|\zeta_k|^p - E\{|\zeta_k|^p\}) > 1 + o_n(1) - a^p E\{|\zeta_1|^p\}\right) \\ & = P\left(a^p K^{-1} \sum_{k=0}^{K-1} (|\zeta_k|^p - E\{|\zeta_k|^p\}) > \gamma\right) \leq \gamma^{-2} a^{2p} K^{-2} K \\ & \quad \times E\{(|\zeta_1|^p - E\{|\zeta_1|^p\})^2\} \leq \gamma^{-2} C_p K^{-1}, \quad C_p < \infty. \end{aligned} \quad (4.6)$$

Inequality (4.6) verifies (4.4) which in its turn verifies (4.3). Now we can define new random variables  $Z_k := [(1 - 2\gamma)/E\{|\zeta_1|^p\}]^{1/p} n^{-1/2} \zeta_k$ . Note that, according to (4.2) and (4.3), a stochastic function (process)  $f^*(t)$ , defined as  $f(t)$  in (4.2) only with  $\theta_{Jk}$  being replaced by  $Z_k$ , belongs to  $B_K$  with the probability tending to 1 as  $n$  increases.

We made all preliminary steps that allow us to use results of Efromovich (2000b) and conclude that

$$\inf_f \sup_{f \in B_K} E\left\{\int_0^1 (\check{f}(t) - f(t))^2 dt\right\} \geq C \sum_{k=0}^{K-1} \frac{[an^{-1/2}]^2 n^{-1}}{[an^{-1/2}]^2 + n^{-1}} \quad (4.7)$$

$$\geq Cn^{-1}K \geq Cn^{-1}[\lambda_n n]^{1/(2\alpha+1)} = C[\lambda_n^{-1/(2\alpha)}n]^{-2\alpha/(2\alpha+1)}.$$

Here  $C$ s are generic positive constants. Inequality (4.7), together with the relation  $B_K \subset B_{pq}^\alpha(f_0)$ , prove Theorem 2.1.

*Proof of Theorem 2.2.* In what follows  $C$ s denote generic positive constants. We begin with a projection pseudo-estimate

$$\tilde{f}_P(t) = \sum_{k=0}^{2^{j_0}-1} \hat{v}_{j_0k} \phi_{j_0k}(t) + \sum_{j=j_0}^J \sum_{k=1}^{2^j-1} \hat{\theta}_{jk} \psi_{jk}(t) \quad (4.8)$$

where  $J$  is the smallest integer such that  $2^J > [\lambda_n n]^{1/(2\alpha+1)}$ . Note that the projection estimate uses the same empirical wavelet coefficients as the blockwise-shrinkage estimate.

Further, according to Efromovich (1999,2000a), we have the following inequality for the mean squared errors of the estimates,

$$E\{(\hat{v}_{j_0k} - v_{j_0k})^2\} \leq Cn^{-1}, \quad E\{(\hat{\theta}_{jk} - \theta_{jk})^2\} \leq Cn^{-1}. \quad (4.9)$$

Using the Parseval identity we can get a classical decomposition of the MISE into variance and integrated squared bias terms, which together with (4.9), used to evaluate the variance part, yields

$$E\left\{\int_0^1 (\tilde{f}_P(t) - f(t))^2 dt\right\} \leq C2^{J+1}n^{-1} + \sum_{j>J} \sum_{k=0}^{2^j-1} \theta_{jk}^2. \quad (4.10)$$

Note that

$$2^J n^{-1} \leq 2[\lambda_n n]^{1/(2\alpha+1)} n^{-1} = 2[\lambda_n^{-1/(2\alpha)}n]^{-2\alpha/(2\alpha+1)}. \quad (4.11)$$

This inequality yields that the variance part of the MISE of the projection estimate has the order indicated by the lower bound (2.9). Let us explore the integrated squared bias term in (4.10). The following inequality holds,

$$\sum_{k=1}^{2^j-1} \theta_{jk}^2 \leq 2^{j(1-2/p)} \left[ \sum_{k=0}^{2^j-1} |\theta_{jk}|^p \right]^{2/p}, \quad p \geq 2. \quad (4.12)$$

To check (4.12) we note that it is trivial for  $p = 2$ , and for  $p > 2$  we can write using the Hölder inequality,

$$\sum_{k=1}^{2^j-1} \theta_{jk}^2 = \sum_{k=1}^{2^j-1} (1)\theta_{jk}^2 \leq \left[ \sum_{k=1}^{2^j-1} 1^{p/(p-2)} \right]^{(p-2)/p} \left[ \sum_{k=1}^{2^j-1} \theta_{jk}^{2(p/2)} \right]^{2/p},$$

which yields (4.12).

Also using (2.7), for  $j > J$  and  $f \in B_{pq}^\alpha(f_0)$  we get the following inequality

$$\left(\sum_{k=1}^{2^j-1} |\theta_{jk}|^p\right)^{1/p} \leq 2^{-j(\alpha+1/2-1/p)} \lambda_n^{1/2}. \quad (4.13)$$

Then (4.12) and (4.13) yield

$$\begin{aligned} \sum_{j>J} \sum_{k=1}^{2^j-1} \theta_{jk}^2 &\leq \sum_{j>J} 2^{j(1-2/p)} 2^{-j(2\alpha+1-2/p)} \lambda_n \\ &\leq C[\lambda_n n]^{-2\alpha/(2\alpha+1)} \lambda_n = C[\lambda_n^{-1/(2\alpha)} n]^{-2\alpha/(2\alpha+1)}. \end{aligned}$$

Using this inequality together with (4.11) in (4.10) establishes the important result about local minimax of the projection estimate, namely

$$\sup_{f \in B_{pq}^\alpha(f_0)} E\left\{\int_0^1 (\tilde{f}_P(t) - f(t))^2 dt\right\} \leq C[\lambda_n^{-1/(2\alpha)} n]^{-2\alpha/(2\alpha+1)}. \quad (4.14)$$

Now we are in a position to present a sequence of steps following Efromovich (1986,1999). The first step is to introduce an oracle-estimate

$$\tilde{f}^*(t) := \sum_{k=0}^{2^{j_0}-1} \hat{v}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{J^*} \sum_{k=1}^{2^j-1} \mu_{jk}^* \hat{\theta}_{jk} \psi_{jk}(t) \quad (4.15)$$

where  $J^*$  is defined above line (2.11) and  $\mu_{jk}^*$  are the individual shrinkage oracle-coefficients  $\mu_{jk}^* = \theta_{jk}^2 / (\theta_{jk}^2 + \sigma^2 n^{-1})$ . Note that these are oracle-coefficients because they depend on wavelet coefficients of an estimated regression function and the variance in the regression model (1.1). It is shown in Efromovich (1999,2000a) that this oracle-estimator dominates the projection estimator and then

$$\sup_{f \in B_{pq}^\alpha(f_0)} E\left\{\int_0^1 (\tilde{f}^*(t) - f(t))^2 dt\right\} \leq C[\lambda_n^{-1/(2\alpha)} n]^{-2\alpha/(2\alpha+1)}. \quad (4.16)$$

The oracle-estimate (4.15) is the most accurate one known in the literature, but a data-driven estimate cannot match its performance. This is why in Efromovich (1986) for the case of a trigonometric basis, and then in Efromovich (2000a) for a wavelet basis, a blockwise-shrinkage oracle estimate has been proposed,

$$\hat{f}^*(t) := \sum_{k=0}^{2^{j_0}-1} \hat{\kappa}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{J^*} \sum_{s=1}^{2^j/L_{j_n}} \frac{\Theta_{js}}{\Theta_{js} + \sigma^2 n^{-1}} \sum_{k \in T_{jsn}} \hat{\theta}_{jk} \psi_{jk}(t). \quad (4.17)$$

It is easy to see that the proposed data-driven estimate (2.10) is motivated (it is actually the plug-in estimate) by the blockwise-shrinkage oracle (4.17). While this blockwise-shrinkage oracle is dominated by the oracle (4.15), it is well known and this is verified by a direct calculation that this oracle is also rate minimax, namely

$$\sup_{f \in B_{pq}^\alpha(f_0)} E\left\{ \int_0^1 (\hat{f}^*(t) - f(t))^2 dt \right\} \leq C[\lambda_n^{-1/(2\alpha)} n]^{-2\alpha/(2\alpha+1)}. \quad (4.18)$$

The final step is to invoke the oracle inequality of Efromovich (2000a) which states that the proposed estimator (2.10) mimics performance of the blockwise-shrinkage oracle for any  $f \in B_{pqQ}^\alpha$ , namely

$$E\left\{ \int_0^1 (\hat{f}(t) - f(t))^2 dt \right\} \leq CE\left\{ \int_0^1 (\hat{f}^*(t) - f(t))^2 dt \right\}. \quad (4.19)$$

Because any considered local Besov space is a subspace of a corresponding Besov space, (4.19) also holds for the considered local Besov class  $B_{pq}^\alpha(f_0)$ .

Using (4.18) in (4.19) proves (2.12). The proof of (2.13) is identical only in (4.16) and (4.18) the supremum over  $f \in B_{pqQ}^\alpha$  yields  $Cn^{-2\alpha/(2\alpha+1)}$ . Theorem 2.2 is proved.

## 5. Conclusion

This is a tradition in the wavelet literature to measure quality of a wavelet estimate via comparison of the MISE convergence with the rate  $n^{-2\alpha/(2\alpha+1)}$ . This rate is the lower bound in a global minimax where the supremum is considered over all functions from a Besov space of order  $\alpha$ . On the other hand, in all practical situation we are dealing with a particular underlying function. It is desirable to measure quality of estimation for a particular function at hand, but we do know that such a program cannot be fulfilled because there is no feasible lower bound for a risk calculated at a particular function. To bridge these two approaches and to get a sharper lower bound, it is proposed to consider a local minimax approach where all considered functions belong to the same Besov space as the pivot does and all these function converge to the pivot (in some norm) as the sample size increases. An appropriate local Besov space is proposed, and its feasibility is supported by: A lower bound that indicates the faster rate of the MISE convergence than the traditional  $n^{-2\alpha/(2\alpha+1)}$ ; A data-driven estimator whose MISE attains the faster rate of the MISE convergence

indicated by the lower bound. Furthermore, the proposed estimator adapts to all parameters of the underlying local Besov space. Oracle-inequalities are also established. The developed estimator is tested on fMRI data exploring the phenomenon of neuroplasticity. Future research will be devoted to considering more general regression models, including heteroscedasticity, dependent errors, and missing observations.

### Acknowledgements

A discussion with Nitis Mukhopadhyay is greatly appreciated. The research is supported by NSF Grant DMS-0906790 and NSA Grant H982301310212

### References

1. Lai, T. L. (1995). Sequential Changepoint Detection in Quality Control and Dynamical Systems, *Journal of Royal Statistical Society, Series B* 57: 613–658.
2. Dimyan, M. A., and Cohen L. G. (2011). Neuroplasticity in the Context of Motor Rehabilitation after Stroke, *National Review Neurology* 7: 76-85. DOI: 10.1038/nrneuro.2010.200
3. Efromovich, S. (1986). Adaptive Algorithm of Nonparametric Regression, *Procedures of Second IFAC Symposium on Stochastic Control, Science* 1: 112-114.
4. Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications*, New York: Springer.
5. Efromovich, S. (2000a). Sharp Linear and Block Shrinkage Wavelet Estimation, *Statistics and Probability Letters* 49: 323-329. DOI: 10.1016/S0167-7152(00)00064-X
6. Efromovich, S. (2000b). On Sharp Adaptive Estimation of Multivariate Curves, *Mathematical Methods of Statistics* 9: 117-139.
7. Greicius, M., Srivastava, G., Reiss, A., Menon, V., and Raichele, M. (2004). Default-Mode Network Activity Distinguishes Alzheimer’s Disease from Healthy Aging: Evidence from Functional MRI, *Proceedings of National Academy of Sciences of United States of America* 101: 4637-4642. DOI: 10.1073/pnas.0308627101

8. Meyer, Y. (1992). *Wavelets and Operators*, Cambridge: Cambridge University Press.
9. Nason, G. (2008). *Wavelet Methods in Statistic with R*, London: Wiley. DOI: 10.1007/978-0-387-75961-6
10. Pinsker, M.S. (1980). Optimal Filtering a Square Integrable Signals in Gaussian White Noise, *Problems Information Transmission*, 16: 45-61.
11. Tung, K., Uh, J., Mao D., Xu, F., Xiao G., and Lu, H. (2013). Alternations in Resting Functional Connectivity due to Recent Motor Task, *NeuroImage* 78: 316-324. DOI: 10.1016/j.neuroimage.2013.04.006
12. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, New York: Wiley. DOI: 10.1002/9780470317020

