Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®

ACADEMIC PRESS

Journal of Multivariate Analysis 85 (2003) 267–291

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

Wavelet regression estimation in nonparametric mixed effect models

Claudia Angelini,[a] Daniela De Canditiis,[a,*] and Frédérique Leblanc[b]

[a] D.M.A., Università degli Studi di Napoli, "Federico II", Italy
[b] L.M.C., Université J. Fourier, B.P. 53, 38041 Grenoble, Cedex 9, France

Received 29 March 2000

Abstract

We show that a nonparametric estimator of a regression function, obtained as solution of a specific regularization problem is the best linear unbiased predictor in some nonparametric mixed effect model. Since this estimator is intractable from a numerical point of view, we propose a tight approximation of it easy and fast to implement. This second estimator achieves the usual optimal rate of convergence of the mean integrated squared error over a Sobolev class both for equispaced and nonequispaced design. Numerical experiments are presented both on simulated and ERP real data.

AMS 1991 subject classifications: 62G08; 62G20; 62J07

Keywords: Wavelets; Besov spaces; Regularization; BLUP estimators

1. Introduction and problem statement

In this work we consider the classical nonparametric regression problem with additive noise:

$$(t_i, Y_i), \quad i = 1, \dots, n, \quad Y_i = f(t_i) + \sigma\varepsilon_i, \quad \text{where } \mathbb{E}\varepsilon_i = 0, \ \mathbb{E}\varepsilon_i^2 = 1, \tag{1}$$

*Corresponding author. Instituto per le Applicazioni del Calcolo "Mauro Picone"-Sezione di Napoli, Consiglio Nazionale delle Ricerche, Via Pietro Castellino, 111 Naples 80131, Italy.

E-mail addresses: angelini@iam.na.cnr.it (C. Angelini), danielad@iam.na.cnr.it (D. De Canditiis), f.leblanc@imag.fr (Fré.. Leblanc).

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(02)00055-6

the ε_i's are uncorrelated random variables and the (t_i) is a deterministic (nonnecessarily regular) design. The value of σ may be known or unknown. We wish to estimate the unknown function f in a nonparametric framework. Hence f will be supposed to belong to some smoothness class $\mathscr{F}$. Problem (1) is studied under two different sets of assumptions on the unknown function f. Either f is considered as a deterministic function and the class $\mathscr{F}$ is a ball in a Sobolev space of regularity s or f has the form

$$f(t) = \mu(t) + b^{1/2}z(t), \tag{2}$$

where μ is a deterministic function and z is a stochastic process which will be specified later. In the first case, the data $\underline{Y} = (Y_1, \ldots, Y_n)^T$ are independent observations. In the second case, $\underline{Y}$ are correlated variables since they are noisy observations of discretisation points of the process f. Moreover, according to the specific structure of covariance considered for the stochastic process z, f lies in a larger space than the Sobolev space considered under the deterministic hypothesis for f.

It is a classical and well-known result that when estimating a deterministic f, the minimax rate obtained over some class $\mathscr{F}$ of Sobolev regularity s is of order $n^{-2s/(2s+1)}$ if the mean integrated squared error (MISE) is considered (see for example [19] or [30]).

Many different linear estimation procedures already exist to reach this rate. Some kernel methods, orthogonal projection, local polynomial, wavelet or spline estimators can be found for example in [2,5,6,14,15,17,34]. All these nonparametric estimators depend on an unknown smoothing parameter. Hence, objective methods were developed by a number of authors for selecting an optimal value of the smoothing parameter (see for example [11,20,22,24]). Among these criteria we will use the generalized cross-validation method.

The aim of this paper is twofold. Assuming that f is deterministic we propose a linear estimator of f as a solution of a minimization problem defined in the wavelet domain. We prove that this estimator is the best linear unbiased predictor for a specific mixed effect regression function f given by (2). The computational complexity of the algorithm to determine this estimator of f is of the order $\mathcal{O}(n^2)$ and does not really take advantage of the fast wavelet discrete transform. Hence, we propose a second estimator which is easy and fast to implement. This new estimator is a tight approximation of the first one. Indeed, we prove that it achieves the same rate of convergence for the MISE as the first one up to a constant.

When $\mathscr{F}$ is the classical Sobolev space $H_2^s[0,1]$, where s is a strictly positive integer, it can be defined as a reproducing kernel Hilbert space $\mathscr{H}$ (r.k.h.s.) with some specific reproducing kernel (r.k.). In this framework precisely described in [34], the so called *s-th order smoothing spline* is the minimizer of the functional

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - f(t_i))^2 + \lambda\int_0^1 [f^{(s)}(x)]^2\,dx \tag{3}$$

over the set $\mathscr{H} = H_2^s[0,1]$. It is known that the s-th order smoothing spline achieves the minimax rate of convergence of the MISE when the design is deterministic (see [22]).

Herein we mimic the usual spline approach to generalize the estimation problem over a Sobolev space with noninteger index. When s is a real number larger than $1/2$ (hence including values of s less than one) we state that $H_2^s[0,1]$ is still a r.k.h.s $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$ with a reproducing kernel constructed with a wavelet basis. Next we estimate f with the solution $\hat{f}_\lambda$ of the following penalized minimization problem:

$$\min_{f \in \mathscr{H}} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(t_i))^2 + \lambda ||\mathscr{P}_1 f||^2_{\mathscr{H}}, \tag{4}$$

where $\mathscr{P}_1$ denotes the orthogonal projector (with respect to the scalar product over $\mathscr{H}$) over the subspace $\mathscr{H}_1$.

For such an optimization problem, when taking $\lambda = 0$, the solution will interpolate the points (t_i, Y_i) by a function of H^s with a huge norm in the Sobolev space. Conversely, taking $\lambda = \infty$ leads to a solution with a small norm in H^s but approximating badly the unknown function f. Hence the term $\lambda||\mathscr{P}_1 f||^2_{\mathscr{H}}$ which penalizes the details in the wavelet expansion of f (equivalent to $\lambda \int_0^1 [f^{(s)}(x)]^2 dx$ which penalizes wild oscillations of f in the spline approach) permits to make a compromise between good approximation and smoothness of the investigated solution. It remains to balance the bias and the variance of the resulting nonparametric estimator.

As for the case of the spline regularization problem (3), we show that $\hat{f}_\lambda$ is also solution of a connected problem. We make the assumption that f is the trajectory of the process given by (2) where $z(t)$ is a Gaussian process whose covariance function is defined by the reproducing kernel K^1 associated with $\mathscr{H}_1$ and $b = \lambda n/\sigma^2$. We then prove that for this particular choice of the parameter b the estimator $\hat{f}_\lambda$ is the minimum variance unbiased predictor for the unknown trajectory f. Under assumption (2), f is the sum of a deterministic part and a random one which describes, respectively, the fixed and the random effects with the specific covariance structure described above, we show that it lies almost surely in a Besov space larger than H^s. This phenomenon is intimately linked with the paradox already noticed by many authors (see for example [34,16]) when interpreting a spline estimator in a Bayesian setting. Indeed, in our case $\hat{f}_\lambda$ has also a Bayesian interpretation under the same prior assumption than the one that are considered in mixed models (see [4]).

It is interesting and relevant to examine Bayesian analogs of corresponding wavelet smoothing frequentist models and procedures. The connection between smoothing methods based on penalized likelihood and mixed effect models and their Bayesian interpretation has been often made implicitly in the literature, but to our knowledge a notable exception on a formal use of such connections is Wahba's 1983 paper (see [33]) exploiting the Bayesian interpretation to construct confidence intervals about splines estimates. It is in the same spirit that we have developed our method since the frequentist properties of confidence intervals for wavelet estimates are problematical. To keep the length of this paper reasonable we focus here on the connection between wavelet smoother and BLUP. Our results about Bayesian interpretation and construction of confidence intervals will be the subject of another paper.

Since $\hat{f}_\lambda(t)$ is the solution of the regularization problem (4), due to a general result of Li [21] it is the minimax linear estimator over a suitable Sobolev ball.

To conclude, this first part of the work generalizes the situation described in [34] for Sobolev space H^s, to the case of s real number larger than 1/2. While achieving the revised version of this paper we noticed that independently from us a similar problem was studied in [18].

Practical formulas for the explicit solution $\hat{f}_\lambda$ are computer intensive and would require a specific numerical treatment as it has been done for the smoothing spline, where some special decompositions have been introduced. In fact for the splines case, the Reinsch algorithm (see [16] and references therein) endowed with the GCV criterion (see [11]) reduces the computational cost to the order $\mathcal{O}(n)$. The same idea cannot be directly applied for evaluating $\hat{f}_\lambda$ since the Reinsch algorithm is directly related to the structure of the splines basis. We therefore propose an alternative estimator, $\tilde{f}_\lambda(t)$, which is easier and faster to compute than $\hat{f}_\lambda$. Indeed, $\tilde{f}_\lambda$ requires $\mathcal{O}(n)$ operations to be computed, using fast wavelet transform; moreover, compared with smoothing spline algorithm, it does not suffer of instability and it can be extended to noninteger s.

The approximated estimator $\tilde{f}_\lambda$ was already studied in [2,6] for an equispaced design. We show that $\tilde{f}_\lambda$ is good enough to assure that it reaches the optimal rate $\mathcal{O}(n^{-2s/(2s+1)})$ also in the nonequispaced case.

To finish we present some numerical simulations and real data applications. We furnish comparisons with other nonparametric estimators.

The paper is organized as follows: In Section 2, we recall the main definitions and tools about wavelets and Besov spaces. Moreover, we prove that the Sobolev space of regularity s with nonnecessarily integer index is a reproducing kernel Hilbert space. Next, in Section 3 we define $\hat{f}_\lambda$ as the solution of a regularization problem and we give its formal expression. In Section 4 we state the Gauss–Markov property of $\hat{f}_\lambda$, in the mixed effect model. Finally, Section 5 is devoted to the approximated solution of $\hat{f}_\lambda$, denoted $\tilde{f}_\lambda$, for both equispaced and nonequispaced designs. In Section 6, numerical results on simulated and real data are discussed.

2. Wavelets and Besov spaces

2.1. *Orthogonal wavelets on* $[0,1]$

We start this section by briefly reviewing some useful facts from basic wavelet theory, that will be used to derive our estimators. A general introduction to the theory of wavelets can be found in [9,12,25,32,35]. The construction of orthonormal wavelet bases for $L^2(\mathbb{R})$ is now well understood. There are many families of wavelets. Throughout this paper we will consider compactly supported wavelets such as Daubechies' orthogonal wavelets. For the construction of orthonormal bases of compactly supported wavelets for $L^2(\mathbb{R})$, one starts with a couple of special, compactly supported functions known as the scaling function φ and the wavelet ψ. The collection of functions $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $j, k \in \mathbb{Z}$, then constitutes an orthonormal basis for $L^2(\mathbb{R})$. For fixed $j \in \mathbb{Z}$, the $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$, $k \in \mathbb{Z}$ are an orthonormal basis for a subspace $V_j \subset L^2(\mathbb{R})$. The spaces V_j constitute a multiresolution analysis.

We denote $P_j f = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k}$ the orthogonal projection of f on the approximation space V_j. The multiresolution analysis is said to be r-regular if φ is C^r, and if both φ and its derivatives, up to the order r, have a fast decay. One can prove that if a multiresolution analysis is r-regular, the wavelet ψ is also C^r and has vanishing moments up to the order r (see [12, Corollary 5.2]). Moreover, we suppose that the moments of φ are equal to zero up to the order r except the zeroth one which is equal to one. Such wavelets were constructed by Daubechies in 1990 and are called *coiflets* (see [12]).

The smoother wavelets provide not only orthonormal bases for $L^2(\mathbb{R})$, but also unconditional bases for several function spaces including Besov spaces (see [31]).

Let us consider now orthogonal wavelets on the interval $[0, 1]$. Adapting wavelets to a finite interval requires some modifications as described in [10]. To summarize, for J such that $2^J \geqslant 2r$, the construction in [10] furnishes a finite set of 2^J scaling functions $\varphi_{J,k}$, and for each $j \geqslant J$, 2^j functions $\psi_{j,k}$, such that the collection of these functions forms a complete orthonormal system of $L_2[0, 1]$. With this notation, the $L_2[0, 1]$ reconstruction formula is

$$f(t) = \sum_{k=0}^{2^J-1} \alpha_{J,k}\varphi_{J,k}(t) + \sum_{j \geqslant J} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(t), \tag{5}$$

where $\alpha_{J,k} = \int_{[0,1]} f(t)\varphi_{J,k}(t)\,dt$, $\beta_{j,k} = \int_{[0,1]} f(t)\psi_{j,k}(t)\,dt$ and we denote $||f||^2 = \int_{[0,1]} f^2(t)\,dt$.

2.2. Besov spaces

In the following, we will use Besov spaces on $[0, 1]$, $B^s_{p,q}$ which are rather general and very well described in terms of sequences of wavelet coefficients. In particular for a suitable choice of the three parameters (s, p, q) we can get Sobolev spaces or Hölder spaces. For the definition of Besov spaces, properties and functional inclusions we refer to [31]. Here we just give the following characterization of the Besov space $B^s_{p,q}$ in terms of wavelet coefficients of its elements.

Lemma 2.1. *Let $0 < p, q \leqslant \infty$ and $s > \max\{(1/p - 1), 0\}$. If the scaling function φ and the wavelet function ψ correspond to a multiresolution analysis of $L_2[0, 1]$ that is*

$([s]+1)$-regular (here $[\cdot]$ stands for the integer part), then a function f in $L_p[0,1]$ belongs to the Besov space $B^s_{p,q}$ if and only if it admits the decomposition (5) *such that*

$$||f||_{B^s_{p,q}} \equiv ||(\alpha_{J,k})_k||_{l_p} + \left(\sum_{j\geqslant J} 2^{jq(s+1/2-1/p)}||(\beta_{j,k})_k||^q_{l_p}\right)^{1/q} < +\infty$$

for $J\in\mathbb{N}$. The $||f||_{B^s_{p,q}}$ is equivalent to the Besov space norm.

For a proof see [13].

Let H^s denote the Sobolev space of functions of $L_2[0,1]$ with noninteger s. By classical embedding relations (see for example [31]) we have $H^s = B^s_{2,2}$. Using Lemma 2.1, we can consider the following inner product over H^s for any $f\in H^s$ and $g\in H^s$:

$$\begin{aligned}
f &= \sum_{k=0}^{2^J-1} \alpha^f_{J,k}\varphi_{J,k} + \sum_{j\geqslant J}\sum_{k=0}^{2^j-1} \beta^f_{j,k}\psi_{j,k},\\
g &= \sum_{k=0}^{2^J-1} \alpha^g_{J,k}\varphi_{J,k} + \sum_{j\geqslant J}\sum_{k=0}^{2^j-1} \beta^g_{j,k}\psi_{j,k},\\
\langle f,g\rangle_{H^s} &= \langle (\alpha^f_{J,k})_k,(\alpha^g_{J,k})_k\rangle_{l_2} + \sum_{j\geqslant J} 2^{2js}\langle (\beta^f_{j,k})_k,(\beta^g_{j,k})_k\rangle_{l_2}.
\end{aligned} \tag{6}$$

The set of functions $\{\varphi_{J,k},\ k=0,\dots,2^J-1;\psi_{j,k},\ j\geqslant J, k=0,\dots,2^j-1\}$ is an orthogonal (but not orthonormal) basis of H^s. Indeed, we have

Lemma 2.2. *Let J be an integer. For any $j\geqslant J$, $j'\geqslant J$, $k,k'\in\mathbb{Z}$,*

(i) $\langle \varphi_{J,k},\varphi_{J,k'}\rangle_{H^s} = \delta_{k,k'}$,
(ii) $\langle \varphi_{J,k},\psi_{j,k'}\rangle_{H^s} = 0$,
(iii) $\langle \psi_{j,k},\psi_{j',k'}\rangle_{H^s} = \delta_{j,j'}\delta_{k,k'}2^{2js}$.

The proof of this result is straightforward so it is omitted.

2.3. Wavelet reproducing kernel Hilbert space

For a complete review on r.k.h.s. we refer to [7]. Put J a fixed integer and introduce the two following symmetric real valued functions defined on $[0,1]^2$:

$$K^0(s,t) = \sum_{k=0}^{2^J-1} \varphi_{J,k}(s)\varphi_{J,k}(t) \quad \text{and} \quad K^1(s,t) = \sum_{j\geqslant J}\sum_{k=0}^{2^j-1} \lambda_j\psi_{j,k}(s)\psi_{j,k}(t),$$

where $\forall j\geqslant J$, $\lambda_j\in\mathbb{R}^+$ such that $\sum_{j\geqslant J} 2^j\lambda_j < +\infty$.

Note that, since ψ is bounded and compactly supported, for any fixed $t\in[0,1]$, $\sum_k |\psi^v_{j,k}(t)|$ is bounded by $\mathcal{O}(2^{vj/2})$ for any $v>0$. Thus $\sum_j \lambda_j 2^j < +\infty$ guarantees that

$K^1(s,t)$ is bounded for any $(s,t)\in[0,1]^2$. Moreover, K^0 and K^1 are obviously positive-definite functions. Hence we can define two unique reproducing kernel Hilbert spaces $\mathscr{H}_0$ and $\mathscr{H}_1$ with K^0 and K^1 as their reproducing kernels. We observe that $\mathscr{H}_0 = V_J = span\{\varphi_{J,k}, k = 0, \dots, 2^J - 1\}$. Since

$$\int_{[0,1]^2} [K^0(s,t)]^2\, ds\, dt = 2^J < \infty \quad \text{and} \quad \int_{[0,1]^2} [K^1(s,t)]^2\, ds\, dt < \infty,$$

from the general theory of r.k.h.s. (see [34, Lemma 1.1.1.]), we have that $f\in\mathscr{H}_0$ and $f\in\mathscr{H}_1$ if and only if

$$||f||^2_{\mathscr{H}_0} \stackrel{\text{def.}}{=} \sum_{k=0}^{2^J-1} \alpha^2_{J,k} < \infty \quad \text{and} \quad ||f||^2_{\mathscr{H}_1} \stackrel{\text{def.}}{=} \sum_{j\geqslant J} \sum_{k=0}^{2^j-1} \frac{\beta^2_{j,k}}{\lambda_j} < \infty.$$

We define the direct sum, $\mathscr{H}$, of the two spaces $\mathscr{H}_0$ and $\mathscr{H}_1$ as: $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$ where $\mathscr{H}$ is endowed with the norm $||f||^2_{\mathscr{H}} = ||f||^2_{\mathscr{H}_0} + ||f||^2_{\mathscr{H}_1}$. Due to [7] $\mathscr{H}$ is a r.k.h.s. Since $\mathscr{H}_0$ and $\mathscr{H}_1$ are orthogonal in L_2 by construction and the $\mathscr{H}$-norm is defined through the $\mathscr{H}_0$-norm and $\mathscr{H}_1$-norm, we keep orthogonality (in $\mathscr{H}$) between $\mathscr{H}_0$ and $\mathscr{H}_1$. The associated scalar product will be denoted $\langle .,. \rangle_{\mathscr{H}}$ and its reproducing kernel is given by

$$K(s,t) = \sum_{k=0}^{2^J-1} \varphi_{J,k}(s)\varphi_{J,k}(t) + \sum_{j\geqslant J} \sum_{k=0}^{2^j-1} \lambda_j \psi_{j,k}(s)\psi_{j,k}(t),$$

so for any $f\in\mathscr{H}$:

$$||f||^2_{\mathscr{H}} = ||\mathscr{P}_0 f||^2_{\mathscr{H}_0} + ||\mathscr{P}_1 f||^2_{\mathscr{H}_1} = \sum_{k=0}^{2^J-1} \alpha^2_{J,k} + \sum_{j\geqslant J} \sum_{k=0}^{2^j-1} \frac{\beta^2_{j,k}}{\lambda_j}, \tag{7}$$

where $\mathscr{P}_i$ denotes the projection over $\mathscr{H}_i$, $i = 0, 1$. Moreover, setting $K_t(\cdot) = K(t,\cdot) = K(\cdot,t)$ (resp. $K^l_t(\cdot) = K^l(t,\cdot) = K^l(\cdot,t)$, for $l = 0, 1$), K_t (resp. K^l_t) is the representer of the evaluation functional at t in $\mathscr{H}$ (resp. in $\mathscr{H}_l$), i.e.,

$$\forall f\in\mathscr{H}, \quad \langle f, K_t \rangle_{\mathscr{H}} = f(t) \quad \text{and} \quad \forall f\in\mathscr{H}_l, \quad \langle f, K^l_t \rangle_{\mathscr{H}_l} = f(t).$$

Next, for some particular choices of λ_j we state that the r.k.h.s. $\mathscr{H}$ coincides with the Sobolev space H^s, for any real $s > 1/2$.

Proposition 2.1. *Let $s > 1/2$; put $\lambda_j = 2^{-2js}$ and suppose that (φ, ψ) defines a multiresolution analysis of $L_2[0,1]$ that is $([s]+1)$-regular, then*

$$H^s = \mathscr{H}.$$

The proof of this result is deferred to Section A.1.

3. Regularization approach

In this section we solve a regularization problem (following the *spline regularization approach* developed in [34]) over a reproducing kernel Hilbert space (r.k.h.s.) which will coincide with a Sobolev space endowed with a wavelet orthogonal basis. The solution of (4) is given by the following theorem.

Theorem 3.1. *Let Φ be the $n \times 2^J$ matrix defined by $\Phi_{i,k} = \varphi_{J,k}(t_i)$ for any $i = 1, \dots, n$ and $k = 0, \dots, 2^J - 1$ and Φ_t the 1×2^J row matrix defined by $\Phi_{1,k} = \varphi_{J,k}(t)$ for any $k = 0, \dots, 2^J - 1$ and for any $t \in [0, 1]$. Moreover, let Σ be the $n \times n$ matrix defined by $\Sigma_{i,j} = K^1(t_i, t_j)$ for any $i = 1, \dots, n$ and $j = 1, \dots, n$ and Σ_t the $1 \times n$ matrix defined by $\Sigma_{t,j} = K^1(t, t_j)$ for any $j = 1, \dots, n$ and for any $t \in [0, 1]$.*

The minimizer of problem (4) *is given by*

$$\begin{aligned} \hat{f}_\lambda(t) &= \sum_{k=0}^{2^J-1} \hat{\alpha}_{J,k}\varphi_{J,k}(t) + \sum_{i=1}^{n} \hat{d}_i K^1_{t_i}(t) \\ &= \Phi_t \underline{\hat{\alpha}} + \Sigma_t \underline{\hat{d}}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} &\underline{\hat{\alpha}} = (\hat{\alpha}_{J,0}, \dots, \hat{\alpha}_{J,2^J-1})^T = (\Phi^T \tilde{\Sigma}^{-1} \Phi)^{-1} \Phi^T \tilde{\Sigma}^{-1} \underline{Y}, \\ &\underline{\hat{d}} = (\hat{d}_1, \dots, \hat{d}_n)^T = \tilde{\Sigma}^{-1}(I_n - \Phi(\Phi^T \tilde{\Sigma}^{-1} \Phi)^{-1} \Phi^T \tilde{\Sigma}^{-1})\, \underline{Y}, \\ &\tilde{\Sigma} = \Sigma + n\lambda I_n \quad \textit{and} \quad \underline{Y} = (Y_1, \dots, Y_n)^T. \end{aligned} \tag{9}$$

The proof of this result is deferred to Section A.1.

Note that $\hat{f}_\lambda(t)$ can be written in terms of a wavelet expansion as follows:

$$\hat{f}_\lambda(t) = \sum_{k=0}^{2^J-1} \hat{\alpha}_{J,k}\varphi_{J,k}(t) + \sum_{j \geq J} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(t) \quad \text{where } \beta_{j,k} = \sum_{i=1}^{n} \lambda_j \hat{d}_i \psi_{j,k}(t_i).$$

4. Mixed model approach

Linear predictors of unknown mixed effects, relying on noisy observations $\underline{Y} = (Y_1, \dots, Y_n)^T$ of f at the design points $t_1, \dots, t_n$, are often considered in a large number of applications for their simplicity and their power. In [28] a rather complete survey is presented for parametric estimation in mixed models. Moreover, some examples are studied under mixed model hypothesis and classical regression model. The comparison of these approaches shows clearly why "*mixed model is a good thing*" in certain situations.

We will study (1) in a mixed model framework. Recall that the observed data, $\underline{Y}$, are discretized observations (at fixed points $t_1, \dots, t_n$) of the trajectory of a stochastic process $Y(t)$ given by

$$Y(t) = f(t) + \sigma\varepsilon(t), \quad t \in [0, 1],$$

where f has the form (2) and $\{\varepsilon(t), t \in [0, 1]\}$ is a zero mean Gaussian process with $Cov(\varepsilon(s), \varepsilon(t)) = \delta_{st}$. Moreover, we assume that $\mu(t) = \sum_{k=0}^{2^J-1} \alpha_{J,k}\varphi_{J,k}(t)$ and $\{\sqrt{b}z(t), t \in [0, 1]\}$ is a centered Gaussian process with covariance function $\mathbb{E}(z(s)z(t)) = K^1(s, t)$. Since $\int\int_{[0,1]^2} K^1(t, s)\, ds\, dt < +\infty$, K^1 admits the Karhunen–Loëve expansion and hence the following representation in quadratic mean holds:

$$\sqrt{b}z(t) = \sum_{j \geqslant J} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(t), \tag{10}$$

where β_{jk} are independent and $\beta_{jk} \sim N(0, \lambda_j)$, (see [26] for the proof). Under the assumptions described by (2) and (10), the trajectories of the processes $z(t)$ and $f(t)$ belong to a space of regular functions. We state that the regularity of this space depends on the choice of the sequence λ_j. We have:

Theorem 4.1. *Let $s > 1/2$ and suppose that the wavelet system $\{\psi_{jk}\}_{j,k}$ is fixed and is $[s] + 1$-regular. Consider the stochastic series,*

$$S(t) = \sum_{j \geqslant J} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(t),$$

where β_{jk} are centered independent normal random variables such that $Var(\beta_{jk}) = \lambda_j$; then the following properties are equivalent:

(i) *each sample of the stochastic series $S(t)$ belongs to $B^{s-1/2}_{2,\infty}[0, 1]$, almost surely (a.s.).*

(ii) $\lambda_j = \mathcal{O}(2^{-2js})$.

The proof of this result is deferred to Section A.2. A more general result has been proved in [1].

As for the regularization approach we assume that $\lambda_j = 2^{-2js}$, hence f lies in $B^{s-1/2}_{2,\infty}$ by Theorem 4.1.

Remark 4.1. We observe that in the regularization approach we assumed that, for the choice of $\lambda_j = 2^{-2js}$, the unknown function f belongs to the Sobolev space H^s. In the mixed model approach the unknown f is assumed to be a sample path that belongs a.s. to a larger space, $B^{s-1/2}_{2,\infty}$. This is the well-known paradox described in [16] for the spline case. However, we have proved in [4] that the Bayesian predictor is exactly $\hat{f}_\lambda(t)$, the solution of the regularization approach and lies in the smaller space H^s (indeed, using classical injections, we have $H^s = B^s_{2,2} \subseteq B^{s-1/2}_{2,2} \subseteq B^{s-1/2}_{2,\infty}$).

Herein we present the BLUP (best linear unbiased predictor), $\hat{f}(t)$, for $f(t)$, using only the observed data $\underline{Y}$. The following definition of BLUP for a function f is a natural extension of the parametric case (see [28]).

Definition 4.1. A predictor $\hat{f}(t)$, relying on the noised observation (data) $\underline{Y}$ given in (1), is the BLUP for $f(t)$ in model (2) if and only if the following properties hold:

(i) $\forall t\ \exists L_t = (l_1(t), \dots, l_n(t))$ such that $\hat{f}(t) = L_t \underline{Y}$.
(ii) $\forall t\ \mathbb{E}\hat{f}(t) = \mathbb{E}f(t) = \mu(t)$.
(iii) $\forall t$ and $\forall \tilde{g}$ such that $\tilde{g}(t) = \tilde{L}_t \underline{Y}$ and $\mathbb{E}\tilde{g}(t) = \mathbb{E}f(t)$, $\mathbb{E}[\hat{f}(t) - f(t)]^2 \leqslant \mathbb{E}[\tilde{g}(t) - f(t)]^2$.

The following theorem gives us the explicit form of the BLUP for predicting $f(t)$. Moreover, we state that the solution of the exact regularization problem is the BLUP for a suitable choice of λ.

Theorem 4.2. *The BLUP for the prediction of $f(t)$ in model* (2), *relying on the data $\underline{Y}$, is given by*

$$\hat{f}(t) = L_t^* \underline{Y}, \tag{11}$$

where the vector $1 \times n$, L_t^*, takes the form

$$L_t^* = \Phi_t(\Phi^T M^{-1} \Phi)^{-1} \Phi^T M^{-1} + \Sigma_t M^{-1}(I_n - \Phi(\Phi^T M^{-1}\Phi)^{-1}\Phi^T M^{-1}),$$

Φ, Φ_t, Σ *and* Σ_t *are defined in Theorem* 3.1 *and* $M = (\Sigma + (\sigma^2/b)I_n)$. *Moreover, with the position* $n\lambda = \sigma^2/b$ *the following identity holds*:

$$\forall t \in [0,1] \quad \hat{f}(t) = \hat{f}_\lambda(t) \text{ where } \hat{f}_\lambda(t) \text{ is given by Theorem 3.1.}$$

The proof of this result is deferred to Section A.2.

The predictor in (11) can be expressed in the equivalent form $\hat{f}(t) = \Phi_t \hat{\underline{\alpha}} + b^{1/2}\hat{z}(t)$, where $\hat{\underline{\alpha}} = (\Phi^T M^{-1} \Phi)^{-1} \Phi^T M^{-1} \underline{Y}$ is the least square weighted predictor for the model $Y(t) = \Phi_t \underline{\alpha} + \varepsilon'(t)$ with $\varepsilon'(t) = b^{1/2} z(t) + \sigma\varepsilon(t)$, and $b^{1/2}\hat{z}(t) = \Sigma_t M^{-1}(I - \Phi(\Phi^T M^{-1}\Phi)^{-1}\Phi^T M^{-1})\underline{Y}$ is the predictor of the centered Gaussian effect.

Remark 4.2. Since $\hat{f}_\lambda(t)$ is solution of the variational problem (4) which is connected with that discussed in [21], by Theorems 2.1 and 2.2 of [21] we conclude that $\hat{f}_\lambda(t)$ is the linear minimax estimator of $f(t)$ in the class $\{f \in H^s\colon \|\mathscr{P}_1 f\|^2_{\mathscr{H}_1} \leqslant \frac{\sigma^2}{\lambda}\}$.

5. Wavelet approximated solution

To develop the approximated estimator $\tilde{f_\lambda}$, we assume that f belongs to H^s and without loss of generality $n = 2^N$. Moreover, we assume that there exists a function $h(t) \in H^{s-1}$ and two positive constants h_1 and h_2 such that $0 < h_1 \leqslant h(t) \leqslant h_2 < \infty$ and $\int_{t_i}^{t_{i+1}} h(t)\,dt = 1/n$ for any i. We introduce the function

$$H(t) \stackrel{\text{def.}}{=} \int_0^t h(u)\,du.$$

Note that $H(0) = 0$, $H(1) = 1$, $H(t_i) = i/n$ and in the equispaced case $H(t) = t$. Moreover, since h is strictly positive, then H is invertible. With this assumption, when $f \in H^s$ we have that $f \circ H^{-1}$ belongs to H^s (where $\circ$ denotes the composition of two functions). Note that, since $h(t)$ is bounded from below and from above, we have the following equivalence:

$$||f||_{L_2} \approx ||f \circ H||_{L_2}. \tag{12}$$

For any $f \in L_2[0,1]$ the orthogonal projector in V_N denoted $P_N f$ is

$$P_N f = \sum_{k=0}^{2^N-1} \alpha_{N,k}\varphi_{N,k} = \sum_{k=0}^{2^J-1} \alpha_{J,k}\varphi_{J,k} + \sum_{j=J}^{N-1}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}.$$

The empirical orthogonal projector in V_N, denoted by $\bar{\Pi}_N f$, is defined for a function f known on a general design, $0 \leqslant t_1 < \cdots < t_n \leqslant 1$, as

$$\begin{aligned}\bar{\Pi}_N f &= \sum_{i=0}^{2^N-1}\left(\sum_{k=0}^{2^N-1} \frac{f(t_{k+1})}{\sqrt{n}} \langle \varphi_{N,k}\circ H, \varphi_{N,i}\rangle\right)\varphi_{N,i}, \\ &\stackrel{\text{def.}}{=} \sum_{i=0}^{2^N-1} \alpha^f_{N,i}\varphi_{N,i} = \sum_{k=0}^{2^J-1} \alpha^f_{J,k}\varphi_{J,k} + \sum_{j=J}^{N-1}\sum_{k=0}^{2^j-1} \beta^f_{j,k}\psi_{j,k}.\end{aligned} \tag{13}$$

Expression of $\bar{\Pi}_N$ simplifies when it is applied to a function f known on an equispaced design. In this particular case we denote it Π_N and we have

$$\Pi_N f = \sum_{i=0}^{2^N-1} \frac{f(\frac{i+1}{n})}{\sqrt{n}}\varphi_{N,i}.$$

With this notations we have $\bar{\Pi}_N f = P_N(\Pi_N(f \circ H^{-1})\circ H)$.

Under the assumptions on the multiresolution analysis, the following approximation results hold, for any $f \in H^s$:

$$||f - P_N f||^2_{L_2} \leqslant C||f||_{H^s} 2^{-2sN}, \tag{14}$$

$$||\Pi_N f - P_N f||^2_{L_2} \leqslant C 2^{-2sN}, \tag{15}$$

$$||\bar{\Pi}_N f - P_N f||^2_{L_2} \leqslant C 2^{-2sN}, \tag{16}$$

where C is any constant independent of f, see [6] for the proof of results (14) and (15); while, (16) comes from the following inequalities:

$$||P_N f - \bar{\Pi}_N f||^2_{L_2} \leqslant ||P_N||^2_{L_2} ||f - \Pi_N(f \circ H^{-1}) \circ H||^2_{L_2}$$
$$\approx C||f \circ H^{-1} - \Pi_N(f \circ H^{-1})||^2_{L_2} \leqslant C2^{-2Ns},$$

where the constant C does not depend on N. According to definition (13) we use the projector $\bar{\Pi}_N$ over the process, Y, known on the design $t_1, \dots, t_n$, in the following sense:

$$\bar{\Pi}_N Y = \sum_{i=0}^{2^N-1} \bar{c}_{N,i}\varphi_{N,i} = \sum_{k=0}^{2^J-1} \bar{c}_{J,k}\varphi_{J,k} + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} \bar{d}_{j,k}\psi_{j,k} \tag{17}$$

and similarly over the noise ε as

$$\bar{\Pi}_N \varepsilon = \sum_{i=0}^{2^N-1} \alpha^{\varepsilon}_{N,i}\varphi_{N,i} = \sum_{k=0}^{2^J-1} \alpha^{\varepsilon}_{J,k}\varphi_{J,k} + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} \beta^{\varepsilon}_{j,k}\psi_{j,k}. \tag{18}$$

Due to the linearity of the operator $\bar{\Pi}_N$, we have

$$\bar{c}_{N,k} = \alpha^f_{N,k} + \alpha^{\varepsilon}_{N,k}, \quad \text{hence } \bar{c}_{J,k} = \alpha^f_{J,k} + \alpha^{\varepsilon}_{J,k} \quad \text{and} \quad \bar{d}_{j,k} = \beta^f_{j,k} + \beta^{\varepsilon}_{j,k}.$$

Since $\bar{\Pi}_N Y$ (resp. $P_N f$) belongs to V_N, the coefficients $\{\bar{c}_{J,k},\ k = 0, \dots, 2^J - 1;\ \bar{d}_{j,k},\ j = J, \dots, N-1,\ k = 0, \dots, 2^j - 1\}$ (resp. $\alpha_{J,k}$ and $\beta_{j,k}$) are obtained using the FWT (Mallat's algorithm) from the coefficients $\bar{c}_{N,k}$ (resp. $\alpha_{N,k}$).

In a general design setting, we suppose the function H known or easily estimated and we define the $n \times n$ matrix

$$P_H = (\langle \varphi_{N,k} \circ H, \varphi_{N,i} \rangle)_{i=0,\dots,2^N-1;\ k=0,\dots,2^N-1},$$

then the empirical coefficients in (17) are evaluated by $\underline{\bar{c}}_{N,\cdot} = P_H \underline{Y}/\sqrt{n}$. Note that for the equispaced design we have $P_H = I_n$, moreover in the general case P_H is a sparse matrix and it can be easily computed using the cascade algorithm.

Due to the orthonormality of the wavelet system, using the expression of $\mathscr{H}$-norm in terms of wavelet coefficients given in (7) and approximation results given in (14)–(16), the exact minimization problem (4) can be approximated up to a term of order $\mathcal{O}(2^{-2sN})$ by the following one:

$$||\bar{\Pi}_N Y - P_N f||^2_{L_2[0,1]} + \lambda \sum_{j,k} \frac{\beta^2_{j,k}}{\lambda_j}. \tag{19}$$

Next, using the expansion of $P_N f$ and $\bar{\Pi}_N Y$ in terms of wavelet coefficients, minimizing expression (19) with respect to $f \in \mathscr{H}$ is equivalent to minimize the following expression with respect to the coefficients $(\alpha_{J,k})_k$ and $(\beta_{j,k})_{j,k}$:

$$\sum_{k=0}^{2^J-1} (\alpha_{J,k} - \bar{c}_{J,k})^2 + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} \left[(\beta_{j,k} - \bar{d}_{j,k})^2 + \lambda \frac{\beta^2_{j,k}}{\lambda_j} \right] + \lambda \sum_{j=N}^{\infty} \sum_{k=0}^{2^j-1} \frac{\beta^2_{j,k}}{\lambda_j}.$$

Such an expression is minimum for the coefficients $(\tilde{\alpha}_{J,k})_k$ and $(\tilde{\beta}_{j,k})$ defined by

$$\begin{cases} \tilde{\alpha}_{J,k} = \bar{c}_{J,k}, & k = 0, \dots, 2^J - 1, \\ \tilde{\beta}_{j,k} = \dfrac{\lambda_j}{\lambda_j + \lambda} \bar{d}_{j,k}, & J \leqslant j \leqslant N-1;\ k = 0, \dots, 2^j - 1, \\ \tilde{\beta}_{j,k} = 0, & j \geqslant N;\ k = 0, \dots, 2^j - 1 \end{cases}$$

and the approximated solution denoted $\tilde{f}_\lambda$ is given by

$$\tilde{f}_\lambda = \sum_{k=0}^{2^J-1} \tilde{\alpha}_{J,k} \varphi_{J,k} + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} \tilde{\beta}_{j,k} \psi_{j,k}.$$

Remark 5.3. Note that the expression of $\tilde{f}_\lambda$ coincides with the one obtained for the equispaced design in [2,6].

In [2,6] it is already proved that the approximated estimator $\tilde{f}_\lambda$, in the equispaced design case and for a particular choice of the parameter λ, converges to the unknown function f with the optimal rate $\mathcal{O}(n^{-\frac{2s}{2s+1}})$ when mean integrated squared error is concerned. The rate of $\tilde{f}_\lambda$ is given up to a constant whereas the rate of $\hat{f}_\lambda$ is minimax. The following theorem states that the same property for estimator $\tilde{f}_\lambda$ holds in the nonequispaced design case.

Theorem 5.1. *Under the regularity assumption on the wavelet bases, for $f \in H^s$ with $s > 1/2$, we have*

$$MISE(\tilde{f}_\lambda) = \mathbb{E}||\tilde{f}_\lambda - f||^2 \leqslant \mathcal{O}(2^{-2Ns} + \lambda + 2^{J-N} + 2^{-N}\lambda^{-\frac{1}{2s}}),$$

moreover, when taking $\lambda = \mathcal{O}(n^{-\frac{2s}{2s+1}})$ we have

$$MISE(\tilde{f}_\lambda) = \mathcal{O}(n^{-\frac{2s}{2s+1}}).$$

The proof of this result is deferred to Section A.3.

6. Numerical results

In this section we show some results obtained by applying the "approximated" wavelet estimator $\tilde{f}_\lambda$ to simulated and real data. It is known that the choice of the smoothing parameter λ can deeply affect the optimal rate of convergence of the estimator. In order to preserve the optimal rate property we need to select it in a 'good' way. In the following experiments, the choice of λ has been done according to the GCV criterion (see [34]).

In [2,3] is proved that $\tilde{f}_\lambda$ endowed with GCV reaches the optimal rate of convergence.

6.1. Simulated data

To show the behavior of the estimator $\tilde{f}_\lambda$, both for the equispaced and nonequispaced design, some numerical experiments have been worked out. Data have been generated according to model (1) with signal-to-noise ratio (SNR) equal to 3. In order to measure the performance of $\tilde{f}_\lambda$ we consider index R_n, defined as

$$R_n = \frac{1}{n}\sum_{k=1}^{n}(\tilde{f}_\lambda(t_k) - f(t_k))^2,$$

and we evaluate it, for each test function, over different realizations of noise. Moreover, we compare $\tilde{f}_\lambda$ with the local and global spline estimators recently studied in [8] and with the local polynomial estimator proposed in [29]. Indeed, the local spline estimator for univariate models have a roughness penalty that adapts to spatial heterogeneity of the regression function. The estimates are p-th degree piecewise polynomials with $p-1$ continuous derivatives. A large and fixed number of knots are used and smoothing is achieved by adding a quadratic penalty on the jumps of the pth derivative in the knots. To be spatially adaptive, the penalty is itself a linear spline but with relatively few knots and values in the knots chosen to minimize GCV. For our comparisons we consider the p-spline with $p=2$ which was recognized to be the most efficient in [8]. The global spline estimator is almost equivalent to the classical smoothing spline estimator. The local polynomial estimator is a classical and well suited method to treat the boundary effects which are well recognized drawbacks when using periodic wavelets. Spatial adaptivity of the local polynomial estimator described in [29] is achieved by using a data-dependent bandwidth selection. We have explored many functions (not showed here for the sake of brevity) and as it has been pointed out in [8], the local spline estimator has a really good performance over heterogeneous functions and sometimes outperforms the other estimators. In such cases $\tilde{f}_\lambda$, the global spline and the local polynomial estimators are comparable. However the computational cost of the local spline estimator is often much larger than the one of $\tilde{f}_\lambda$. From our intensive simulations we have found that in many cases the behavior of the four estimators is similar and $\tilde{f}_\lambda$ can also outperform global, local spline and local polynomial estimators, especially for functions which exhibit some singularities. In any case we stress that the advantage of using the "approximated" wavelet estimator is mainly due to its lower computational cost.

We consider three different kinds of test functions defined for $t \in [0,1]$:

- An infinitely times differentiable function:

$$f_1(t) = 0.5 + 0.2\cos(4\pi t) + 0.1\cos(24\pi x).$$

- A function that has a discontinuity in the first derivative:

$$f_2(t) = \exp(-|t-1/2|).$$

- A well-known discontinuous function, Heavysine, that does not satisfy the assumptions of Theorem 5.1, but represents a typical 1-D signal.

In Figs. 1 and 2 we show, for a particular realization of noise and for the test functions $f_1(t)$ and Heavysine, respectively, the noisy observations and the true function, the estimates obtained using the wavelet estimator, the local spline estimator, the global spline estimator and the local polynomial estimator, for a sample size $n = 2^9$ in the case of equispaced data. Moreover, in the lower right part of Figs. 1 and 2 we show the boxplots of R_n using the four methods over 500 simulated samples for the same test functions and grid design. They reveal a better performance of the wavelet estimator when applied to test functions f_1. Indeed, the wavelet estimator $\tilde{f}_\lambda$ is able to keep the second cosine component often regarded as noise (in several samples) when using the other methods. When Heavysine is considered the wavelet estimator reconstructs better the main singularity.

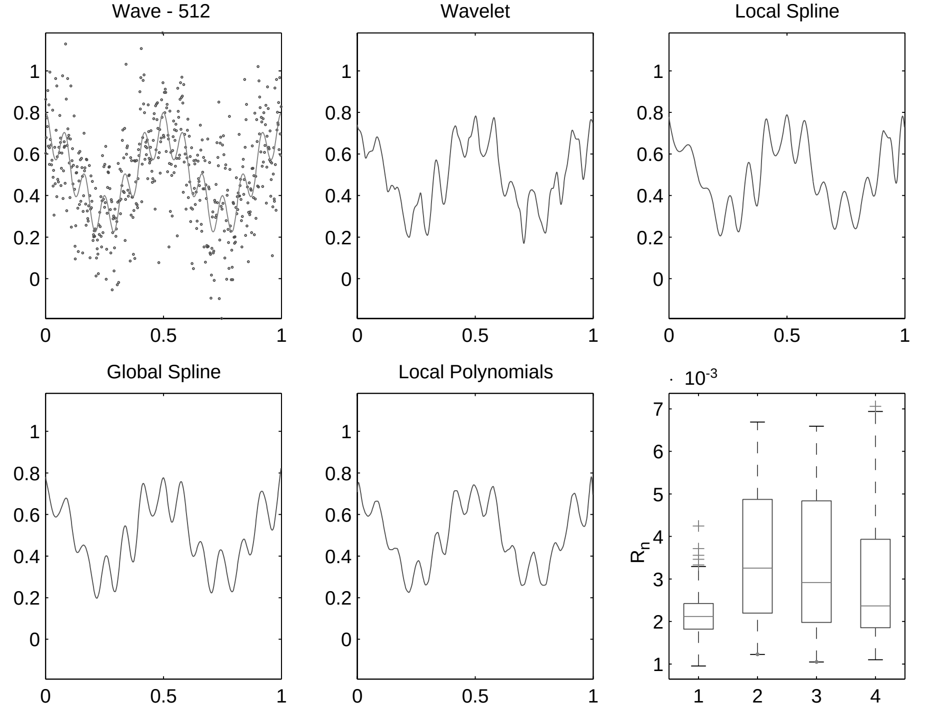

Fig. 1. Plot true and noisy functions (upper left) and the regularized solution obtained using the wavelet estimator endowed with GCV (upper middle), the local spline estimator (upper right), the global spline estimator (lower left), and the local polynomial estimator (lower middle). In the lower right panel the boxplots of R_n computed over 500 simulated samples are showed for the four methods: (1) Wavelet estimator, (2) Local spline estimator, (3) Global spline estimator and (4) Local polynomial estimator. The test function is f_1, $n = 2^9$, SNR = 3 and the design is regular.

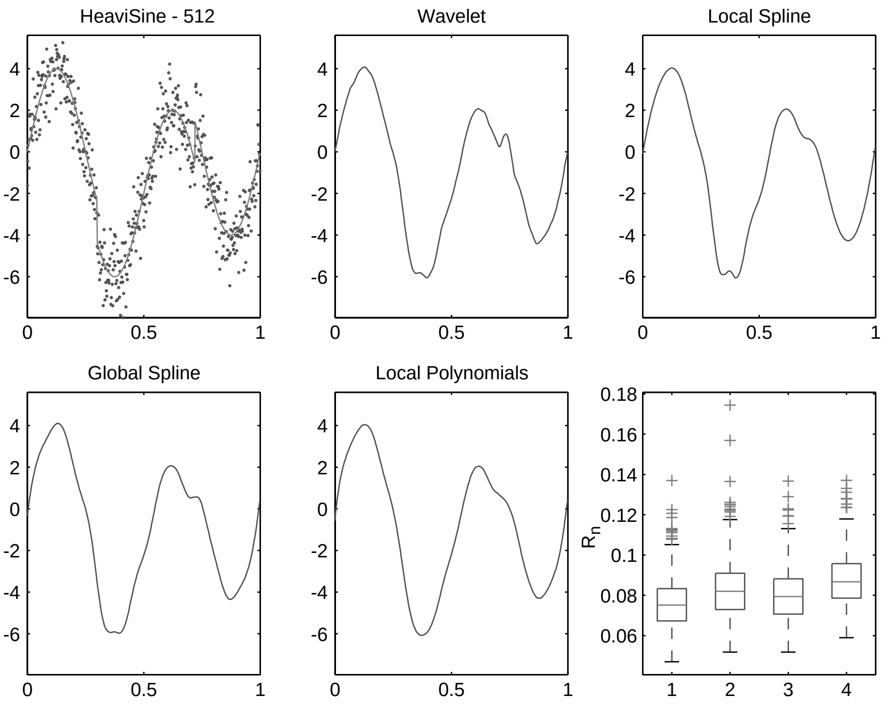

Fig. 2. Plot true and noisy functions (upper left) and the regularized solution obtained using the wavelet estimator endowed with GCV (upper middle), the local spline estimator (upper right), the global spline estimator (lower left), and the local polynomial estimator (lower middle). In the lower right panel the boxplots of R_n computed over 500 simulated samples are showed for the four methods: (1) Wavelet estimator, (2) Local spline estimator, (3) Global spline estimator and (4) Local polynomial estimator. The test function is HeavySine, $n = 2^9$, SNR $= 3$ and the design is regular.

Next, to present the previous estimators in the nonequispaced case, we use the test function $f_2(t)$ and the following grid design:

$$t_i = \frac{e^{i/n} - 1}{e - 1}, \quad i = 1, \ldots, n. \tag{20}$$

The computational results, in this situation are presented in Fig. 3. For the considered sample size $n = 2^9$ the mean behaviors of the four estimators are similar, but when n increases the performance of the wavelet estimator improves.

6.2. A real data application

The estimator $\tilde{f}_\lambda$ has been applied to a real data example concerning with human event-related potentials (ERPs) records described in [23]. Ten subjects were presented five squares, one of which was randomly green colored; then a circle randomly appears inside one of the boxes. If the circle fills the attended green box, then the subject is required to press a thumb button as soon as possible. ERP records

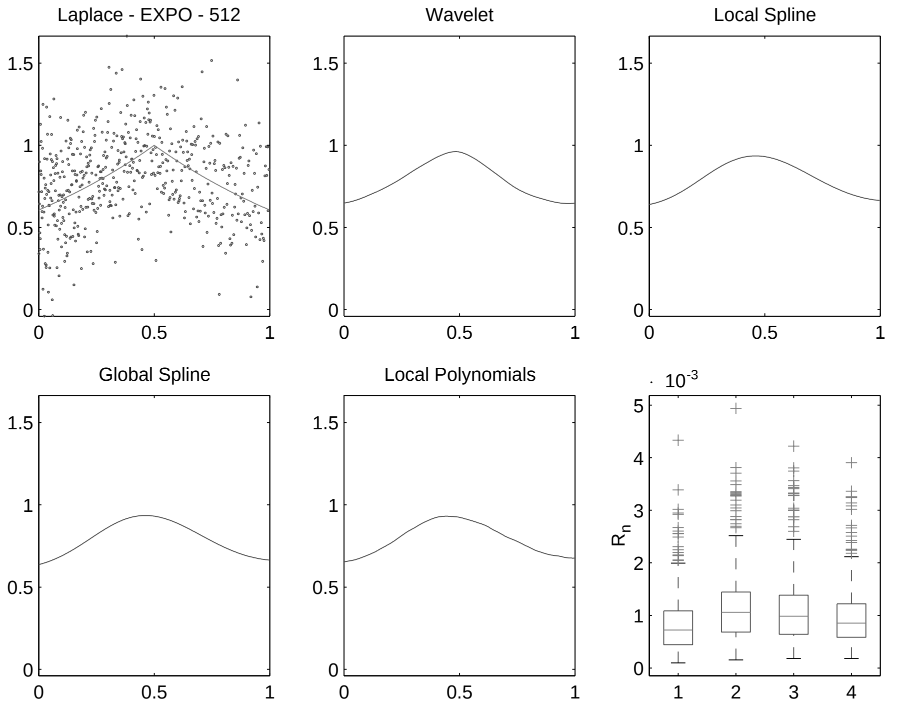

Fig. 3. Plot true and noisy functions (upper left) and the regularized solution obtained using the wavelet estimator endowed with GCV (upper middle), the local spline estimator (upper right), the global spline estimator (lower left), and the local polynomial estimator (lower middle). In the lower right panel the boxplots of R_n computed over 500 simulated samples are showed for the four methods: (1) Wavelet estimator, (2) Local spline estimator, (3) Global spline estimator and (4) Local polynomial estimator. The test function is f_2, $n = 2^9$, SNR $= 3$ and the design is given in (20).

are sampled at 512 Hz for 1000 ms (i.e. each record consists of 512 equispaced noisy observations). The original experiment arises in a different context, here we would like to describe the mean behavior of the observed measurements. In fact, smoothed data are used for classification or other purposes. In Figs. 4 and 5 we show the noisy observations and the estimates obtained applying $\tilde{f}_\lambda$, the local and global spline and the local polynomial estimates, for two different records. Analysis of figures shows a comparable mean behavior of the four estimators. Since a periodic wavelet basis is used in the computation, some boundary effects are visible for the wavelet estimate. This drawback could be reduced by using boundary corrected wavelets. For many applications of interest for this particular data set, the boundary effects are not crucial.

7. Conclusions

In this paper the BLUP is obtained using an orthogonal wavelet basis. A wavelet estimator, $\tilde{f}_\lambda$, fast and easy to compute is proposed. We prove that $\tilde{f}_\lambda$ achieves the

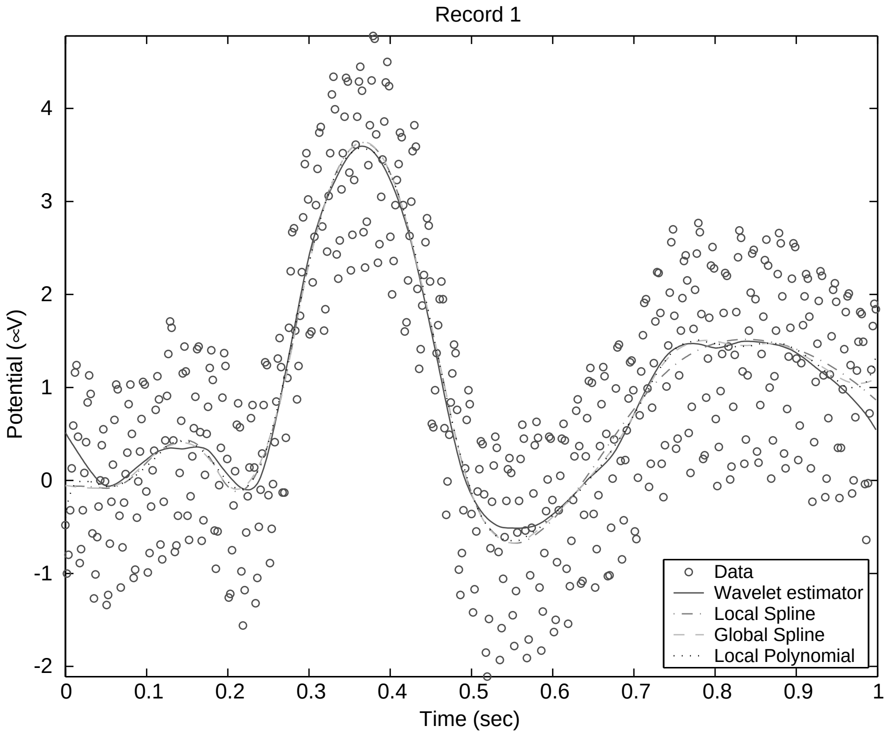

Fig. 4. Plot evoked potential as a function of time: noisy observations (∘) and wavelet estimator endowed with GCV (continuous line), local and spline estimator (dash-dotted and dashed lines respectively) and the local polynomial estimator (dotted line) for record 1.

usual optimal rate of convergence in the MISE sense over a Sobolev class. Numerical experiments on simulated and real data are presented and comparisons with local and global spline and local polynomial estimators are carried out. As a general conclusion, we suggest to use wavelet estimator when the size of the sample is large. The wavelet estimator should be also used when the regression function presents some singularities, global spline estimator when the function is smooth and local spline or local polynomial estimators when heterogeneous smooth functions are considered.

Acknowledgments

Work supported by CNR and CNRS in the framework of the project "*Metodi basati sulle wavelets per la stima non parametrica e la regolarizzazione di problemi malposti*". Authors wish to deeply thank Professor Anestis Antoniadis and Dr. Umberto Amato for stimulating discussions and continuous encouragement to the work. The authors are also grateful to the associated editor and to the referees for their useful comments which significantly improved the presentation of the paper.

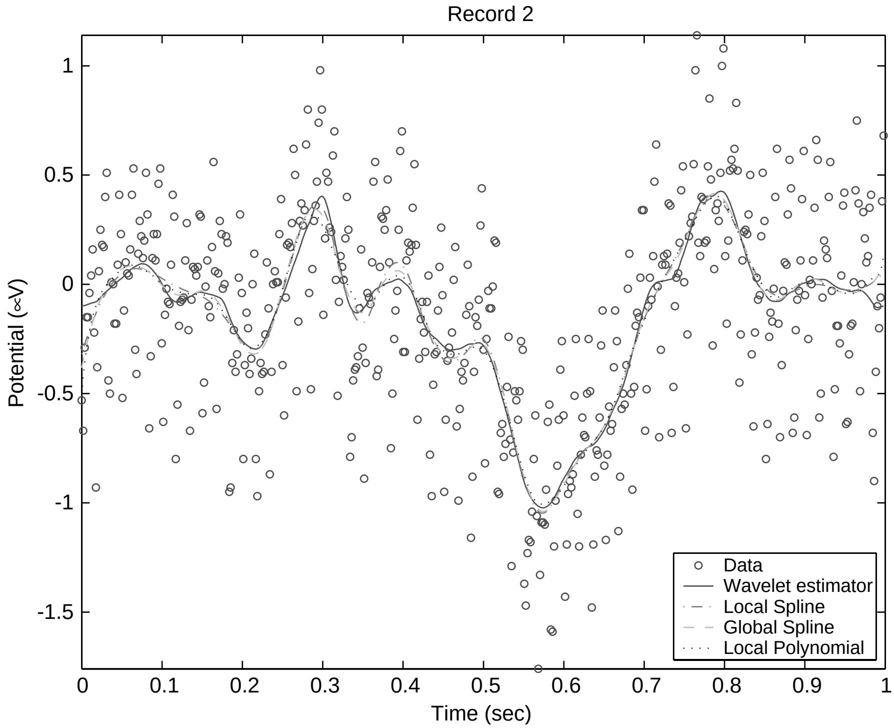

Fig. 5. Plot evoked potential as function of time: noisy observations (◦) and wavelet estimator endowed with GCV (continuous line), local and spline estimator (dash-dotted and dashed lines respectively) and the local polynomial estimator (dotted line) for record 2.

Appendix A

In this section we give the proofs of all main results we have obtained. In the proofs we use the notations $\underline{\mu} = (\mu(t_1), \ldots, \mu(t_n))^T$, $\underline{z} = (z(t_1), \ldots, z(t_n))^T$ and $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

A.1. Proofs of the results in Section 3

Proof of Proposition 2.1. To prove the result we have to check the following properties:

(i) $\forall t \in [0, 1]$, $K_t \in H^s$.

(ii) $\forall f \in H^s$, $\langle K_t, f \rangle_{H^s} = f(t)$.

We have the following wavelet expansion for K_t:

$$K_t(\cdot) = \sum_{k=0}^{2^J-1} \varphi_{J,k}(t)\varphi_{J,k}(\cdot) + \sum_{j \geqslant J} \sum_{k=0}^{2^j-1} \lambda_j \psi_{j,k}(t)\psi_{j,k}(\cdot).$$

Next, using the inner product in H^s defined in (6) we have

$$||K_t||^2_{H^s} = \sum_{k=0}^{2^J-1} \varphi^2_{J,k}(t) + \sum_{j \geqslant J} 2^{2js} \sum_{k=0}^{2^j-1} \lambda_j^2 \psi^2_{j,k}(t).$$

The first term in the right-hand side of the previous expression is obviously finite, because J is finite and $\varphi_{J,k}$ is bounded. Moreover, due to the compact support property of ψ, $\sum_{k=0}^{2^j-1} \psi^2_{j,k}(t)$ is bounded by $\mathcal{O}(2^j)$. Then the choice $\lambda_j = 2^{-2js}$ implies that $\sum_{j \geqslant J} 2^{(2s+1)j} \lambda_j^2 < \infty$ as soon as $s > 1/2$. Hence the second term being also bounded (i) holds.

When using the wavelet expansion of K_t, bilinearity of the inner product and properties given in Lemma 2.2, by straightforward calculations, the choice $\lambda_j = 2^{-2js}$ implies (ii). □

Proof of Theorem 3.1. In this proof we omit the subscript $\mathscr{H}$ in the inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}}$ (or in the norm $|| \cdot ||_{\mathscr{H}}$). By a property of Hilbert space, the minimizer $\hat{f_\lambda}$ of (4) admits the representation

$$\hat{f_\lambda} = \underbrace{\sum_{k=0}^{2^J-1} \hat{\alpha}_{J,k} \varphi_{J,k}}_{\mathscr{P}_0 f} + \underbrace{\sum_{i=1}^{n} \hat{d}_i K^1_{t_i}}_{\mathscr{P}_1 f} + \rho, \tag{A.1}$$

where ρ is some element in $\mathscr{H}_1$ (hence perpendicular to $\mathscr{H}_0$) perpendicular to $K^1_{t_1}, \ldots, K^1_{t_n}$. Before going further on in the proof we remark that since $\rho \in \mathscr{H}_1$ is perpendicular to $K^1_{t_i}$ and $\mathscr{P}_1$ is a self-adjoint operator, for any $i = 1, \ldots, n$, the following equalities hold:

$$\langle \rho, K_{t_i} \rangle = \langle \mathscr{P}_1 \rho, K_{t_i} \rangle = \langle \rho, \mathscr{P}_1 K_{t_i} \rangle = \langle \rho, K^1_{t_i} \rangle = 0,$$
$$\langle K^1_{t_i}, K^1_{t_j} \rangle = \langle K_{t_i}, K^1_{t_j} \rangle = K^1(t_i, t_j).$$

Now we compute the two terms which constitute (4) using expression (A.1) and the previous remark. Since these calculations mimic the proof given in page 12 of Wahba's book (see [34]) we outline the end of the proof.

The first term in (4) can be written as

$$\frac{1}{n} || \underline{Y} - (\Phi \underline{\hat{\alpha}} + \Sigma \underline{\hat{d}}) ||^2_{l_2} = \frac{1}{n} (\underline{Y} - (\Phi \underline{\hat{\alpha}} + \Sigma \underline{\hat{d}}))^T (\underline{Y} - (\Phi \underline{\hat{\alpha}} + \Sigma \underline{\hat{d}})) \tag{A.2}$$

and the second term in (4) is given by $\lambda ||\mathscr{P}_1 f||^2_{\mathscr{H}} = \lambda \underline{\hat{d}}^T \Sigma \underline{\hat{d}} + \lambda ||\rho||^2$.

Solving problem (4) with respect to f is equivalent to minimize the sum of (A.2) and of $\lambda ||\mathscr{P}_1 f||^2_{\mathscr{H}}$ multiplied by n with respect to the function ρ and the vectors $\underline{\hat{d}}$ and $\underline{\hat{\alpha}}$. The solution is given by $\rho = 0$ and $\underline{\hat{d}}$ and $\underline{\hat{\alpha}}$ are as given in the theorem. □

A.2. Proofs of the results in Section 4

Proof of Theorem 4.1. Put $A_j = 2^{2j(s-1/2)} \sum_{k=0}^{2^j-1} (\beta_{j,k}^2 - \lambda_j)$. Since $\beta_{j,k}$, $k = 0, \ldots, 2^j - 1$, are centered Gaussian independent variables with $Var(\beta_{j,k}) = \lambda_j$, we have

$$\mathbb{E}A_j = 0 \quad \text{and} \quad Var(A_j) = 2\ 2^{4j(s-1/4)}\lambda_j^2.$$

Applying Markov inequality for any $\epsilon > 0$ to the r.v.s A_j for all j and summing up the inequalities we obtain

$$\sum_{j \geqslant J} P(|A_j| > \epsilon) \leqslant \frac{2}{\epsilon^2} \sum_{j \geqslant J} 2^{4j(s-1/4)}\lambda_j^2.$$

Suppose now that (ii) is true. Then the right-hand side of the previous inequality is finite, hence the series of the general term $P(|A_j| > \epsilon)$ converges, which implies a.s., convergence of A_j to zero, by Borel Cantelli Lemma. Due to definition of A_j this convergence implies that $\sum_{k=0}^{2^j-1} 2^{2j(s-1/2)}\beta_{j,k}^2$ admits a finite limit, thus $\sup_{j \geqslant J} 2^{2j(s-1/2)} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 < +\infty$ (a.s.), which is equivalent to $S(t) \in B_{2,\infty}^{s-1/2}[0,1]$ by Lemma 2.1.

Conversely, suppose (i) holds; hence $2^{2j(s-1/2)} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \leqslant c$ (a.s.) for some constant c independent of j. Next, taking the expectation of both sides of this inequality we obtain $2^{2j(s-1/2)} \sum_{k=0}^{2^j-1} \lambda_j \leqslant c$ which is equivalent to (ii). □

Lemma A.1. *Let A be a positive definite matrix of order m, B a $m \times k$ matrix and $\underline{u}$ a k-dimensional vector. Denote by S^- any generalized inverse of $B^T A^{-1} B$. Then*

$$\inf_{\{\underline{x} \in R^m \ \text{s.t.}\ B^T\underline{x} = \underline{u}\}} \underline{x}^T A \underline{x} = \underline{u}^T S^- \underline{u}$$

and the infimum is a minimum obtained at $\underline{x}^* = A^{-1} B S^- \underline{u}$.

Proof. See [27]. □

Proof of Theorem 4.2. From properties (i) and (iii) of Definition 4.1, finding the BLUP, $\hat{f}(t)$, is equivalent to solve the following constrained minimization problem:

$$\inf_{\{L_t \ \text{s.t.}\ L_t \in V_t\}} \mathbb{E}[L_t \underline{Y} - f(t)]^2, \tag{A.3}$$

where the constraint on $L_t = (l_1 \ldots l_n)$ is given in order to satisfy property (ii) of BLUP's definition. Such a condition can be explicitly written as

$$L_t \in V_t \iff L_t \underline{\mu} = \mu(t). \tag{A.4}$$

Using the expression of $\underline{Y}$ and $f(t)$ given in (1) and (2), taking in account constraint (A.4) and the independence between $\underline{z}$ and $\underline{\varepsilon}$, we obtain

$$\begin{aligned}\mathbb{E}[L_t\underline{Y} - f(t)]^2 &= \mathbb{E}[L_t(\underline{\mu} + b^{1/2}\underline{z} + \sigma\underline{\varepsilon}) - \mu(t) - b^{1/2}z(t)]^2 \\ &= L_t\, Var(b^{1/2}\underline{z} + \sigma\underline{\varepsilon})L_t^T + b\, Var(z(t)) - 2bL_t\mathbb{E}(\underline{z}z(t)).\end{aligned}$$

It can be easily verified that $Var(b^{1/2}\underline{z} + \sigma\underline{\varepsilon}) = (b\Sigma + \sigma^2 I)$, $Var(z(t)) = K^1(t,t)$ and $\mathbb{E}(\underline{z}z(t)) = \Sigma_t^T$. Substituting these values, we have

$$\mathbb{E}[L_t\underline{Y} - f(t)]^2 = b\{L_tML_t^T + K^1(t,t) - L_t\Sigma_t^T - \Sigma_tL_t^T\},$$

where $M = (\Sigma + (\sigma^2/b)I_n)$. By standard calculation, the latter expression can be rearranged as

$$\mathbb{E}[L_t\underline{Y} - f(t)]^2 = b\{(L_t - \Sigma_tM^{-1})M(L_t - \Sigma_tM^{-1})^T + R\},$$

where $R = K^1(t,t) - \Sigma_tM^{-1}\Sigma_t^T$. Since R is independent of L_t in the optimization problem (A.3), it can be neglected. Hence, solving (A.3) is equivalent to solve the constrained minimization problem

$$\inf_{\{L_t \text{ s.t. } L_t\in V_t\}}[(L_t - \Sigma_tM^{-1})M(L_t - \Sigma_tM^{-1})^T].$$

By Lemma A.1, for $\underline{x} = (L_t - \Sigma_tM^{-1})^T$, $A = M$, $B = \Phi$ and $\underline{u} = \Phi_t^T - \Phi^TM^{-1}\Sigma_t^T$, we immediately obtain the minimizer in (11). The last part of the theorem is trivial. $\square$

A.3. Proofs of the results in Section 5

Lemma A.2. *Under the regularity assumption on the wavelet system and on the function H and setting $J\in\mathbb{N}$, for any $j\geqslant J$ and $k = 0,\dots,j$ we have*:

$$\mathbb{E}(\alpha_{J,k}^{\varepsilon})^2 \leqslant C\frac{\sigma^2}{2^N} \quad \textit{and} \quad \mathbb{E}(\beta_{j,k}^{\varepsilon})^2 \leqslant C\frac{\sigma^2}{2^N}, \quad J\leqslant j\leqslant N-1,$$

where C is a constant independent on N, $\alpha_{J,k}^{\varepsilon}$ and $\beta_{j,k}^{\varepsilon}$ are defined in (18).

Proof. By definition we have the following expression for $\alpha_{J,k}^{\varepsilon}$:

$$\left\langle \sum_{i=0}^{2^N-1}\sum_{l=0}^{2^N-1}\frac{\varepsilon_{l+1}}{\sqrt{n}}\langle\varphi_{N,l}\circ H, \varphi_{N,i}\rangle\varphi_{N,i}, \varphi_{J,k}\right\rangle = \frac{1}{\sqrt{n}}\sum_{l=0}^{2^N-1}\varepsilon_{l+1}s_l^{J,k},$$

where we define $s_l^{J,k} = \sum_i a_{l,i}\langle\varphi_{N,i},\varphi_{J,k}\rangle$ with $a_{l,i} = \langle\varphi_{N,l}\circ H, \varphi_{N,i}\rangle$.

From the independence of the components of $\underline{\varepsilon}$, we obtain

$$\mathbb{E}(\alpha_{J,k}^{\varepsilon})^2 = \mathbb{E}\left(\frac{1}{n}\sum_{l=0}^{2^N-1}\sum_{l'=0}^{2^N-1}\varepsilon_{l+1}\varepsilon_{l'+1}s_l^{J,k}s_{l'}^{J,k}\right) = \frac{\sigma^2}{n}\sum_{l=0}^{2^N-1}(s_l^{J,k})^2.$$

To end the proof of the first part of the lemma we will show that $\sum_l (s_l^{J,k})^2$ is uniformly bounded with respect to N. Note that for fixed l, $\varphi_{N,l} \circ H$ and $\varphi_{N,i}$ have compact supports with a nonempty intersection for a finite number of indices i only. Furthermore, for any l and N this number is bounded by a constant K. Put m_l and M_l the minimum and the maximum index i such that $a_{l,i}$ has a nonzero value. Moreover, there is a constant a such that for any i and l, $a_{l,i} \leqslant \|\varphi_{N,l}(H)\|_\infty \|\varphi_{N,i}(x)\|_\infty |Supp\, \varphi_{N,i}(x)| \leqslant a$. Thus we have

$$s_l^{J,k} = \sum_{i=m_l}^{M_l} a_{l,i} \langle \varphi_{N,i}, \varphi_{J,k} \rangle \leqslant a \sum_{i=m_l}^{M_l} \langle \varphi_{N,i}, \varphi_{J,k} \rangle$$

and

$$(s_l^{J,k})^2 \leqslant a^2 2^{M_l - m_l - 1} \sum_{i=m_l}^{M_l} \langle \varphi_{N,i}, \varphi_{J,k} \rangle^2 \leqslant C \sum_{i=m_l}^{M_l} \langle \varphi_{N,i}, \varphi_{J,k} \rangle^2.$$

Finally, by direct substitution

$$\begin{aligned} \sum_{l=0}^{2^N-1} (s_l^{J,k})^2 &\leqslant C \sum_{l=0}^{2^N-1} \sum_{i=m_l}^{M_l} \langle \varphi_{N,i}, \varphi_{J,k} \rangle^2 = C \sum_{i=0}^{2^N-1} \sum_{l=m_i^*}^{M_i^*} \langle \varphi_{N,i}, \varphi_{J,k} \rangle^2 \\ &\leqslant C \max_i |M_i^* - m_i^*| \sum_{i=0}^{2^N-1} \langle \varphi_{N,i}, \varphi_{J,k} \rangle^2 \leqslant C_1 \|\varphi_{J,k}\|_{L_2}^2 \leqslant C_1. \end{aligned}$$

In the previous step we have used the fact that if both m_l and M_l are strictly increasing sequences such that $\max_l |M_l - m_l| \leqslant K$ then we have $\max_i |M_i^* - m_i^*| \leqslant K$. For the second part of the lemma we put

$$\beta_{j,k}^{\varepsilon} = \frac{1}{\sqrt{n}} \sum_{l=0}^{2^N-1} \varepsilon_{l+1} \underbrace{\sum_{i=0}^{2^N-1} \langle \varphi_{N,l} \circ H, \varphi_{N,i} \rangle \langle \varphi_{N,i}, \psi_{j,k} \rangle}_{r_l^{j,k}}.$$

As in the previous case, we state $\sum_l (r_l^{j,k})^2 \leqslant C_2$. $\square$

Proof of Theorem 5.1.

$$\begin{aligned} \mathbb{E}\|\tilde{f}_\lambda - f\|^2 &\leqslant 2^2 \|f - \Pi_N(f \circ H^{-1}) \circ H\|^2 + \|\Pi_N(f \circ H^{-1}) \circ H - \bar{\Pi}_N f\|^2 \\ &\quad + \mathbb{E}\|\bar{\Pi}_N f - \tilde{f}_\lambda\|^2 \stackrel{\text{def.}}{=} 2^2(B_1 + B_2 + B_3). \end{aligned}$$

Let us consider the term B_1. When using equivalence in (12) and results in (14) and (15), we obtain $B_1 \leqslant C 2^{-2Ns}$. Next we consider the term B_2. Since the function

$\Pi_N(f \circ H^{-1}) \circ H \in H^s$, due to (14) $B_2 \leqslant C2^{-2Ns}$. Let us consider now the term B_3. By definition we have

$$\mathbb{E}||\bar{\Pi}_N f - \tilde{f}_\lambda||^2 = \sum_{k=0}^{2^J-1} \mathbb{E}(\alpha_{J,k}^\epsilon)^2 + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} \left(\frac{\lambda_j}{\lambda_j + \lambda}\right)^2 \mathbb{E}(\beta_{j,k}^\epsilon)^2 + \sum_{j=J}^{N-1} \sum_{k=0}^{2^j} \left(\frac{\lambda}{\lambda + \lambda_j} \beta_{j,k}^f\right)^2.$$

Then applying Lemma A.2

$$\begin{aligned}
&\mathbb{E}||\bar{\Pi}_N f - \tilde{f}_\lambda||^2 \\
&\leqslant C\left(2^J \frac{\sigma^2}{2^N} + \frac{\sigma^2}{2^N} \sum_{j=J}^{N-1} 2^j \left(\frac{\lambda_j}{\lambda_j + \lambda}\right)^2 + \frac{\lambda}{2} \sum_{j=J}^{N-1} \sum_{k=0}^{2^j-1} 2^{2js} (\beta_{j,k}^f)^2\right) \\
&\leqslant C'\left(2^J \frac{\sigma^2}{2^N} + \frac{\sigma^2}{2^N} \int_1^\infty \frac{1}{(1 + \lambda y^{2s})^2} dy + \lambda ||\bar{\Pi}_N f||_{H^s}^2\right).
\end{aligned}$$

The theorem is proved observing that the integral which appears in the previous expression is $\mathcal{O}(\lambda^{\frac{-1}{2s}})$ and that $||\bar{\Pi}_N f||_{H^s}^2$ is uniformly bounded independently of N. Indeed, for any function $g \in H^s$, using (15), the following inequality holds:

$$||\Pi_N g||_{H^s} \leqslant ||\Pi_N g - P_N g||_{H^s} + ||P_N g||_{H^s} \leqslant C2^{Ns} ||\Pi_N g - P_N g|| + ||g||_{H^s}.$$

Due to regularity of H, $||\Pi_N(f \circ H^{-1}) \circ H||_{H^s} \leqslant C$ and applying the previous inequality to $g = \Pi_N(f \circ H^{-1}) \circ H$ we conclude. $\square$

References

[1] F. Abramovich, T. Sapatinas, B.W. Silverman, Wavelet thresholding via Bayesian approach, J. Roy. Soc. Stat. B (1998) 725–749.

[2] U. Amato, D.T. Vuza, Wavelet approximation of a function from samples affected by noise, Rev. Roumaine Math. Pures Appl. 42 (1997) 481–493.

[3] U. Amato, D.T. Vuza, Smoothing data with correlated noise by wavelet regularization, Rev. Roumanie Math. Pure Appl. 43 (1998) 207–222.

[4] C. Angelini, D. De Canditiis, F. Leblanc, Bayesian interpretation for a regularized wavelet estimator of the regression function, Rapp. Tecnico 226/01, Istituto per Applicazioni della Matematica, 2001. Italy, Current version available in the anonymous ftp site: ftp://ftp.iam.na.cnr.it/pub/RT193.pdf

[5] A. Antoniadis, Smoothing data with coiflets, Statist. Sinica 4 (1994) 651–678.

[6] A. Antoniadis, Smoothing noisy data with tapered coiflets series, Scand. J. Statist. 23 (1996) 313–330.

[7] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950) 337–404.

[8] R.J. Caroll, D. Ruppert, Spatially-adaptive penalties for Spline Fitting, Austral. New Zealand, J. Statist. 42 (2000) 205–223.

[9] C.K. Chui, An Introduction to Wavelets, Academic Press, New York, 1992.

[10] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, Appl. Comput. Harm. Anal. 1 (1993) 54–81.

[11] P. Craven, G. Wahba, Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of GCV, Numer. Math. 31 (1979) 377–403.
[12] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61, SIAM, Philadelphia, 1992.
[13] B. Delyon, A. Juditsky, Estimating wavelet coefficients, in: A. Antoniadis and G. Oppenheim (Eds.), Wavelets and Statistics, Springer, Berlin, 1995.
[14] R.L. Eubank, Spline Smoothing and Nonparametric Regression, Dekker, New York, 1988.
[15] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[16] P.J. Green, B.W. Silverman, Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman & Hall, London, 1994.
[17] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, Cambridge, 1990.
[18] S.Y. Huang, H.H.S. Lu, Bayesian wavelet shrinkage for nonparametric mixed-effects models, Statist. Sinica 10 (2000) 1021–1040.
[19] I.A. Ibragimov, R.K. Khasminskii, Bounds for the risks of non-parametric regression estimates, Theory Probab. Appl. 27 (1982) 84–99.
[20] O. Lepski, Asymptotically minimax adaptive estimation I: upper bounds. Optimally adaptive estimates, Theory Probab. Appl. 36 (1991) 682–697.
[21] K.C. Li, Minimaxity of the method of regularization on stochastic processes, Ann. Statist. 10 (1982) 937–942.
[22] K.C. Li, Asymptotic optimality of C_p and GCV in ridge regression with application to spline smoothing, Ann. Statist. 14 (1986) 1101–1112.
[23] S. Makeig, M. Westerfield, J. Covington, J. Townsend, T.J. Sejnowski, E. Courchesne, Functionally independent components of the late positive event-related potential during visual spatial attention, J. Neurosci. 19 (1999) 2665–2680.
[24] C.L. Mallows, Some comments on C_p, Technometrics 15 (1973) 661–675.
[25] Y. Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.
[26] J. Neveu, Processus Aléatoires Gaussiens Séminaire de mathématiques supérieures, 1968.
[27] G. Rao, Linear Statistical Inference and its Applications, Vol. XVIII, Wiley, New York, 1965.
[28] G.K. Robinson, That BLUP is a good thing: the estimation of random effects, Statist. Sci. 6 (1991) 15–51.
[29] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, J. Amer. Statist. Assoc. 90 (1995) 1257–1270.
[30] J.C. Stone, Optimal global rates of convergence for nonparametric estimators, Ann. Statist. 10 (1982) 1040–1053.
[31] H. Triebel, Theory of Function Spaces II, Birkhauser, Basel, 1983.
[32] B. Vidakovic, Statistical Modeling by Wavelets, Wiley, New York, 1999.
[33] G. Wahba, Bayesian "Confidence intervals" for the cross-validated smoothing splines, J. R. Statist. Soc. B 45 (1983) 133–150.
[34] G. Wahba, Spline models for observational data, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, PA, 1990.
[35] G.G. Walter, Wavelets and Other Orthogonal Systems with Applications, CRC Press, Boca Raton, FL, 1994.