Published on *Computational Statistics and Data Analysis* (2012), 56, 1482-1498.
Publisher's version available at http://dx.doi.org/10.1016/j.csda.2011.12.016

Wavelets in Functional Data Analysis: estimation of multidimensional curves and their derivatives

Davide Pigoli, Laura M. Sangalli[1]
MOX - Dipartimento di Matematica, Politecnico di Milano

Abstract

A wavelet-based method is proposed to obtain accurate estimates of curves in more than one dimension and of their derivatives. By means of simulation studies, this novel method is compared to another locally-adaptive estimation technique for multidimensional functional data, based on free-knot regression splines. This comparison shows that the proposed method is particularly attractive when the curves to be estimated present strongly localized features. The multidimensional wavelet estimation method is thus applied to multi-lead electrocardiogram records, where strongly localized features are indeed expected.

Key words: Multidimensional curve fitting, Derivative estimation, Electrocardiograms.

1. Introduction

Functional Data Analysis (FDA) is the branch of statistics which focuses on data that can be seen as the observed value of a functional random variable (see, e.g., Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). However, from a practical point of view, every data is observed on a discrete grid and a measurement error is also present. A crucial step of the analysis thus consists

[1]Corresponding author.
Postal address: Piazza Leonardo da Vinci 32, 20133, Milano, Italy
Email address: laura.sangalli@polimi.it
Phone: +39 02 2399 4554
Fax: +39 02 2399 4606

in the estimation of the continuous functional data starting from its discrete observation. Here, in particular, we are especially interested in the estimation of multidimensional curves and their derivatives. Sangalli et al. (2009) proposed a smoothing technique based on free-knot splines, that was shown to provide very accurate estimates of multidimensional curves and their derivatives, even when the curve are characterized by spatial inhomogeneities. In this work we instead describe a technique based on wavelet bases expansion, that is also capable of accurate estimation of multidimensional curves presenting strongly localized features, such us peaks and oscillations.

Wavelet bases have already proved to be very useful in functional data analysis, dealing with one-dimensional curves, thanks largely to their natural local-adaptivity, that allows them to accommodate a wide variety of functional forms. Besides the problem of curve estimation (see, e.g., Antoniadis et al., 1994), other settings where wavelets have been used include for instance the estimation parameters of stochastic processes (see, e.g., Frías and Ruiz-Medina, 2011, and reference therein), the framework of functional regression (see, e.g., Aguilera et al., 2008), as well as functional anova (see, e.g., Yang and Nie, 2008) and functional classification (see, e.g., Wang et al., 2007; Berlinet et al., 2008; Antoniadis et al., 2010; Timmermans et al., 2011, and references therein). Due to the strong increase in the recording of multidimensional functional data, it seems of interest to extend the field of application of the above mentioned methods to the multidimensional case, developing a wavelet-based estimation technique that can accurately handle multidimensional curves.

Likewise in Sangalli et al. (2009), being our data noisy and discrete observations of some p-dimensional curve, we look for an estimate that is itself a proper p-dimensional curve. This means that we discard the simplistic idea of obtaining an estimate by juxtaposition of p separate smoothing of the p functional coordinates of the curve. In fact, if the curve has a significant feature at some point of the physical space, we expect that this will be to some extent reflected on all p coordinates concurrently; for instance, if the curve has at some point a discontinuity in some of the derivatives, this will be present on all p coordinates. For this reason, we develop of a novel estimation procedure which takes into account simultaneously all the space coordinates of the multidimensional curve. Moreover, we also consider the case where the components of error in the p dimensions are correlated, and show how to efficiently deal with this issue. The proposed estimation technique also provides consistent estimates of the curve derivatives. It should be noticed that

wavelet bases have been so far mainly applied in problems where there was no interest in derivatives, because of the absence of close analytical forms for smooth wavelet bases, issue that has until now restricted their application to a confined part of the FDA field. To overcome this limitation, we resort here to a numerical method (see Strang and Nguyen, 1996) that allows to obtain derivatives of wavelet estimated data. An additional contribution of the paper is a scheme for aligning an orthogonal basis so that the sampled curve values are a better approximation to the scaling coefficients needed to initialize the discrete wavelet transform, thus obtaining a more accurate estimation of curve derivatives. An extended abstract of this work appeared in Pigoli and Sangalli (2011).

1.1. Motivating applied problem: analysis of multi-lead ECG data

Our research has been stimulated by the analysis of Electro Cardio Gram (ECG) records, collected by the 118 Dispatch Center (the medical operating emergency unit) in Milano, Italy, as part of the PROMETEO project "PROgetto sull'area Milanese Elettrocardiogrammi Teletrasferiti dall'Extra Ospedaliero". See Ieva et al. (2011) for details. The aim of this Project is to anticipate diagnostic time in heart ischemia, in order to improve the prognosis of reperfusive treatments and reduce infarction complications. In particular, we consider a sample of multi-lead tele-transmitted ECG records, both physiological and pathological. The estimates here derived via the proposed multidimensional smoothing technique are thus used in Ieva et al. (2011), where a semi-automatic diagnostic procedure is proposed, based on the ECG morphology, that is able to classify physiological and pathological traces.

ECG data have a multidimensional nature, because these records provide potential differences, named leads, between multiple electrodes; in fact, as it will be described in Section 6, ECG traces can be seen as eight-dimensional functional data, whose eight coordinates, corresponding to eight leads, measure different projections of the same physical dynamics in different directions. Smoothing of these data hence calls for a technique that takes into account simultaneously the eight coordinates of this functional data; besides helping in detecting significant features which reflect on more than one lead, thus enhancing patter recognition, such procedure provides coherent estimates, where the different projections of the heart dynamics are among them consistent. Moreover, as it will be clarified later, the components of error on

the eight leads are correlated, issue that can be appropriately taken into account within our technique working jointly on the p coordinates. It should also be noticed that wavelet basis are particularly well suited to capture ECG shapes, that are characterized by localized strong peaks and oscillations.

As mentioned in the previous section, we devote particular attention to the computation of estimates that are accompanied by good estimates of their derivatives. This is payed off in Ieva et al. (2011), where it is shown that, to better study ECG morphology and efficiently distinguish between physiological and pathological ECG traces, it is necessary to take into account both the ECG traces and also their first derivatives.

1.2. Paper outline

The paper is organized as follows. In Section 2 we briefly recall wavelet bases, we review a numerical method that allows to compute pointwise values of a wavelet and its derivatives, and we summarize wavelet smoothing for one dimensional functional data; in this section we moreover derive an optimal translation of the orthogonal basis so that the sampled curve values are a better approximation to the scaling coefficients at the finer scale. Section 3 accurately extends wavelet-based estimation techniques to the case of curves in more than one dimension. Section 4 illustrates the good performances of the proposed technique, especially in the case of multidimensional functional data characterized by strongly localized features. In Section 5 we consider the case where the components of error in the p dimensions are correlated. Section 6 is devoted to the application to the multi-lead ECG data, that have been stimulus to this research. Finally, some conclusive considerations are drawn in Section 7.

2. Wavelets for smoothing 1D data

2.1. An overview on wavelets

We briefly recall wavelet bases for $L^2(\mathbb{R})$. For a systematic introduction to wavelets, see, e.g., Mallat (1999) or Nason (2008). Wavelets are defined starting from an orthogonal multiresolution:

Definition 2.1 *Let $\{V_j\}_{j\in\mathbb{Z}}$ be a sequence of closed subspaces $V_j \subseteq L^2(\mathbb{R})$ and let $\varphi \in V_0$. An orthogonal multiresolution for $L^2(\mathbb{R})$ is a couple $(\{V_j\}_j, \varphi)$ such that:*

1. $V_j \subset V_{j+1}$

2. $\overline{\bigcup_j V_j} = L^2(\mathbb{R})$ *and* $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$

3. $\{l \mapsto f(l)\} \in V_j \;\; \Leftrightarrow \;\; \{l \mapsto f(2l)\} \in V_{j+1}$

4. $\{\varphi(l-k)\}_{k\in\mathbb{Z}}$ *is an orthonormal basis for* V_0 *and* $\int_{\mathbb{R}} \varphi \neq 0$.

The projections of $f \in L^2(\mathbb{R})$ on the sequence $\{V_j\}_j$ give a progressively better approximation of f as j increases. The function φ is called *scaling function* or *father wavelet*. Thanks to property 3 above, $\{2^{j/2}\varphi(2^j l - k)\}_k$ is an orthonormal basis for V_j. However, it is often more useful exploring the detail information needed to go from the space V_j to the space V_{j+1}, starting from a coarse space V_0. This is the reason to introduce the sequence of the complement spaces $W_j = V_{j+1} \backslash V_j$. A *mother wavelet* is a function $\psi \in W_0$ so that $\{\psi(l-k)\}_k$ is a basis for W_0. As a consequence, $L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j$ and $\{\psi_{j,k}(l)\}_k = \{2^{\frac{j}{2}}\psi(2^j l - k)\}_k$ is an orthonormal basis for $L^2(\mathbb{R})$. Therefore, for each $f \in L^2(\mathbb{R})$, we have

$$f = \sum_j \sum_k \langle f, \psi_{j,k}\rangle \psi_{j,k} = \sum_k \langle f, \varphi_{j_0,k}\rangle \varphi_{j_0,k} + \sum_{j=j_0}^{+\infty} \sum_k \langle f, \psi_{j,k}\rangle \psi_{j,k} =$$

$$= \sum_k s_{j_0,k}\varphi_{j_0,k} + \sum_{j=j_0}^{+\infty} \sum_k d_{j,k}\psi_{j,k},$$

where $\langle \cdot, \cdot \rangle$ is the scalar product in $L^2(\mathbb{R})$, $s_{j_0,k} := \langle f, \varphi_{j_0,k}\rangle$ and $d_{j,k} := \langle f, \psi_{j,k}\rangle$. The coefficients $\{s_{j_0,k}\}_{k\in\mathbb{Z}}$, $\{d_{j,k}\}_{j\in\mathbb{Z}\cap\{j>=j_0\},k\in\mathbb{Z}}$ are called *discrete wavelet transform* of f. It can be shown that φ and ψ satisfy the dilation/refinement equations

$$\varphi(l) = \sum_k \sqrt{2} h_k \varphi(2l-k) \qquad \text{and} \qquad \psi(l) = \sum_k \sqrt{2} g_k \varphi(2l-k)$$

for some sequences $\{h_k\}_k$ and $\{g_k\}_k$, named respectively *scaling filter* and *wavelet filter*. These equations are essential for the development of the so-called *fast wavelet transform*, which computes the discrete wavelet transform in $O(n)$ operations. It is important to note that smooth and compactly supported wavelet bases have no analytical form, and they are instead defined via their scaling and wavelet filters.

2.2. Computation of pointwise value of wavelets and their derivatives

As mentioned in the Introduction, wavelet basis has so far been mainly confined to problems in which derivative estimates were not required, this limitation being due to the absence of a close analytical form for wavelet basis functions smooth enough for this purpose.

Wavelets have been used in Leung et al. (1998) to compute approximate function derivatives as difference between the scaling coefficients coming from different scaling/wavelet bases. This procedure exploits the fact that different wavelet bases cause different shifts of the function projected on the subspace V_{J-1}, so that a pair of bases can be found whose difference approximates the function variation at that scale. Anyway, this method offers a derivative estimation on a space that is coarser than the original one. Since higher order derivatives are estimated through an iterative procedure, evaluation points become fewer and fewer, resulting in not accurate estimates of the derivatives. Moreover, a high observation noise strongly worsen this issue, being necessary to use coarser level scaling coefficients.

Here we instead resort to the numerical method illustrated by Strang and Nguyen (1996) that allows to compute pointwise values of scaling functions and their derivatives, even in the absence of a close analytical form for the wavelet basis. This method is based on a common approach for solving dilation equations. Its starting point is constituted by the scaling filter $\{h_k\}_k$ and the wavelet filter $\{g_k\}_k$, which are available for all the wavelet bases of interest. Let $0, \ldots, N$ be the integers in the support of φ, where N is in fact the largest value in its support. The vector of evaluations of φ on $0, \ldots, N$ can be computed as eigenvector of a matrix H, whose elements are obtained from the coefficients of the scaling filter; in particular, the $(i, j)-$th entry of the matrix H, for $i, j = 1, \ldots, N + 1$, is given by $H_{i,j} = \sqrt{2}h_{2i-j-1}$ if $0 \leq 2i - j - 1 < N + 1$ and $H_{i,j} = 0$ otherwise (see Strang and Nguyen, 1996, for details). Using the dilation equation, φ can then be evaluated at the middle points $\frac{1}{2}, \ldots, N - \frac{1}{2}$:

$$\varphi\Big(\frac{l}{2}\Big) = \sum_k \sqrt{2}h_k\varphi(l - k).$$

Iterating this procedure, is thus possible to compute the desired refinement of φ.

Similarly, starting from the k-th derivative of the dilation equation, the vector of evaluation of the k-th derivative $\varphi^{(k)}$ on $0, \ldots, N$ can also be com-

puted as eigenvector of the matrix H corresponding to the eigenvalue $1/2^k$. Likewise for φ, also $\varphi^{(k)}$ can thus be evaluated on progressively finer grids thanks to the k-th derivative of the dilation equation:

$$\varphi^{(k)}\left(\frac{l}{2}\right) = 2^k \sum_k \sqrt{2} h_k \varphi^{(k)}(l-k).$$

The computational cost of these procedure is not high (see Strang and Nguyen, 1996, for details), since the refinement is computed simply by the scalar product between the previously computed values of φ (or its derivatives) and a suitable upsampling of the filter h_k, and matrix H is only $(N+1)\times(N+1)$. For example, N is equal to 19 for the Daubechies wavelet basis with 10 vanishing moment, which will be used in the simulation studies of Section 4.

2.3. Wavelet smoothing of functional data

As we recalled in Section 2.1, every function $f \in L^2(\mathbb{R})$ can be represented by a scaling/wavelet basis. This can be used to obtain an estimator of a functional data, starting from its discrete observation. Let the statistical model be

$$w_i = f(l_i) + \varepsilon_i, \quad i = 1, \ldots, n, \;\; n = 2^J, \;\; J \in \mathbb{N}, \tag{1}$$

where f is the true curve, to be estimated, $l_i = i/n$ are evenly spaced points and ε_i are independent and identically distributed (i.i.d.) errors with $N(0, \sigma^2)$ distribution. The wavelet smoothing procedure consists in changing over to the wavelet domain, where the model becomes

$$d_{j,k} = d^0_{j,k} + \rho_{j,k}$$

$d_{j,k}$ being the empirical coefficients corresponding to the data, $d^0_{j,k}$ the true wavelet coefficients of f, and $\rho_{j,k} \sim N(0, \sigma_d^2)$ the wavelet transforms of the error. Section 2.4 describes how the empirical coefficients $d_{j,k}$, and the coefficients $s_{j_0,k}$ are computed from the data $w_1, \ldots, w_n$. Here we instead recall how the estimates $\hat{d}_{j,k}$ of the true wavelet coefficients $d^0_{j,k}$ can be obtained starting from the empirical coefficients $d_{j,k}$. A first idea consists in the so-called *hard-thresholding* estimator, which fixes a threshold t and considers all coefficients below this threshold as coming only from noise, thus setting $\hat{d}_{j,k} = d_{j,k}\mathbb{I}_{\{|d_{j,k}|>t\}}$. A more refined idea consists in also shrinking the coefficients above this threshold, with the aim of removing their component due

to noise. In particular, the wavelet estimator becomes

$$\hat{d}_{j,k} = \text{sign}(d_{j,k})(|d_{j,k}| - t)_+$$

which is called *soft-thresholding* estimator. The corresponding estimate of the true function f is then given by

$$\widehat{f}(l) = \sum_{k=1}^{2^{j_0}} s_{j_0,k}\varphi_{j_0,k}(l) + \sum_{j=j_0}^{j=J-1} \sum_{k=1}^{2^j} \hat{d}_{j,k}\psi_{j,k}(l). \tag{2}$$

It should be noticed that since wavelet bases are by construction localized in both space and frequency, (2) very naturally provides a locally adaptive estimate of the function f.

The level j_0 in (2) is the lowest for which thresholding is applied; wavelets coefficients of levels lower than j_0 do not undergo any thresholding. The choice of this smoothing coefficient depends on the signal to noise ratio of the data and on the problem under analysis; in general lower signal to noise ratios lead to choosing lower values of j_0 (see, e.g., Nason, 2008). Many strategies have been instead proposed for the choice of the threshold t (see, e.g., Cai and Zhou, 2009; Donoho and Johnstone, 1995; Donoho et al., 1995), among which a popular one is the *universal threshold*:

$$t = \hat{\sigma}_d \sqrt{2 \log n}\,.$$

The estimation of σ_d is based on the fact that, in the wavelet transform, the wavelet coefficients of the finer level $J-1$ are essentially pure noise. Donoho et al. (1995) proposed to use a robust estimator, that is given by the median of the absolute deviation from the median (MAD) of these coefficients, i.e.,

$$\hat{\sigma}_d = \frac{\text{median}(|d_{J-1} - \text{median}(d_{J-1})|)}{0.6745}. \tag{3}$$

2.4. Computation of the empirical coefficients and translation of scaling and wavelet basis functions

To initialize the estimating algorithm, it is necessary to compute the empirical coefficients $d_{j,k}$ from the data $w_1, \ldots, w_n$. The fast wavelet transform algorithm, described in Beylkin et al. (1991), allows to compute all the scaling and wavelet coefficients, starting from the coefficients $s_{J,1}, \ldots, s_{J,n}$. A common procedure consists in approximating the latter coefficients with

the data values $w_1, \ldots, w_n$. This is justified by the fact that the support of $\varphi_{J,i}$ is localized around the point l_i, if J is large enough (i.e., if n is large enough). For a general discussion on the adequacy of this choice see, e.g., Nason (2008). Anyway, since every wavelet basis is defined up to a translation, it is convenient to look for an appropriate translation τ of the basis functions such that the error of this approximation is as small as possible. Here we use the translation $\tau = \int_{\mathbb{R}} y\varphi(y)\mathrm{d}y$. We motivate this choice by the following argument. Approximating f by its first order Taylor expansion, we obtain

$$s_{J,i} = \int_{\mathbb{R}} f(l)\varphi_{J,i}(l+\tau)\,\mathrm{d}l \approx \int_{\mathbb{R}} (f(l_i) + lf'(l_i))\varphi_{J,i}(l+\tau)\,\mathrm{d}l$$

that, exploiting the fact that $\int_{\mathbb{R}} \varphi_{J,i}(l+\tau)\,\mathrm{d}l = 1$, leads to

$$s_{J,i} \approx f(l_i) + f'(l_i) \int_{\mathbb{R}} l\varphi_{J,i}(l+\tau)\,\mathrm{d}l.$$

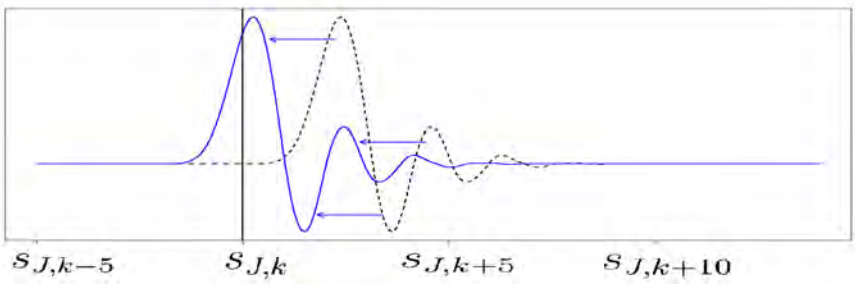


Figure 1: $\varphi_{J,k}$ for the original (dashed black line) and τ-translated (solid blue line) Daubechies wavelet basis with 10 vanishing moments.

This means that, if the translation τ is such that

$$\int_{\mathbb{R}} l\varphi(l+\tau)\,\mathrm{d}l = 0, \tag{4}$$

then $f(l_i)$ is a good approximation of $s_{J,i}$, so that, in this case, using the data value w_i to approximate $s_{J,i}$ seems well justified. If we thus impose the supplementary condition (4), we obtain (by substitution in the integral of

$y = l + \tau$)

$$\tau = \int_{\mathbb{R}} y\varphi(y)\mathrm{d}y. \tag{5}$$

We therefore use the basis associated to the translation τ found above. Figure 1, for instance, shows the effect of this translation on a basis function of level J, for the Daubechies wavelet basis with 10 vanishing moments. For this choice of the wavelet basis, the optimal translation suggested by (5) is $\tau = 2.130968$.

The estimate of the true function f hence becomes

$$\widehat{f}(l) = \sum_{k=1}^{2^{j_0}} s_{j_0,k}\varphi_{j_0,k}(l+\tau) + \sum_{j=j_0}^{j=J-1}\sum_{k=1}^{2^j} \hat{d}_{j,k}\psi_{j,k}(l+\tau)$$

and the corresponding estimate of the first derivative is

$$\widehat{f}\,'(l) = \sum_{k=1}^{2^{j_0}} s_{j_0,k}\varphi'_{j_0,k}(l+\tau) + \sum_{j=j_0}^{j=J-1}\sum_{k=1}^{2^j} \hat{d}_{j,k}\psi'_{j,k}(l+\tau)$$

where pointwise values of $\varphi_{j_0,k}$, $\psi_{j,k}$, $\varphi'_{j_0,k}$ and $\psi'_{j,k}$ are computed as detailed in Section 2.2. Subsequent derivative estimates are obtained analogously.

3. Wavelet estimation for curves in more than one dimension

We now extend wavelet-based estimation techniques to the case of curves in more than one dimension. The function $\mathbf{f}$ we want to estimate has the form

$$\mathbf{f} : \mathbb{R} \ni l \mapsto \big(f_1(l), \ldots, f_p(l)\big) \in \mathbb{R}^p$$

which describes parametric curves in p dimensions. The observed values are generated by the model

$$\mathbf{w}_i = \mathbf{f}(l_i) + \boldsymbol{\varepsilon}_i \quad i = 1, \ldots, n = 2^J \tag{6}$$

where $\boldsymbol{\varepsilon}_i$ are i.i.d. multinormal errors with mean the null vector $\mathbf{0} \in \mathbb{R}^p$ and variance-covariance matrix $\sigma^2\mathbb{I}_p$. Our aim is to estimate the function $\mathbf{f}$ and its derivatives. As anticipated in the Introduction, we discard the simplistic idea of estimating each coordinate function $f_1, \ldots, f_p$ independently, applying separately on each coordinate the procedure described in the previous

section. In fact, if the curve has a significant feature at some point of the physical space, we expect that this will be reflected on all p coordinates concurrently. For this reason, we develop an estimation technique that takes into account the vectorial structure of the function to be estimated. In particular, the proposed estimation technique is such that the same wavelet basis functions are used for the estimation of all coordinate functions $f_1, \ldots, f_p$ of $\mathbf{f}$; a specific wavelet basis function, with a specific frequency and location, is either used for each of the coordinate functions, in order to capture a feature of the p-dimensional function $\mathbf{f}$, or is not used for any of the coordinate functions, if unnecessary to capture relevant features of $\mathbf{f}$. Thus, the proposed soft-thresholding works on p-dimensional wavelet coefficients, so that these coefficients are set to the null vector $\mathbf{0}$, or undergo an appropriate shrinkage, that takes accurately into account all p coordinates.

3.1. Generalization of Universal Threshold in p dimensions

Starting from model (6) and using the orthogonality of the wavelet transform, we have that

$$\mathbf{d}_{j,k} = \mathbf{d}^0_{j,k} + \boldsymbol{\rho}_{j,k},$$

with $\mathbf{d}_{j,k}, \mathbf{d}^0_{j,k}, \boldsymbol{\rho}_{j,k} \in \mathbb{R}^p$, where $\mathbf{d}_{j,k}$ are the vectors of the empirical wavelet coefficients corresponding to the data, $\mathbf{d}^0_{j,k}$ are the vectors of the true wavelet coefficients of the p-dimensional function $\mathbf{f}$, and $\boldsymbol{\rho}_{j,k}$ are the wavelet transforms of the noise and have multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_d^2 \mathbb{I}_p$. In order to decide if $\mathbf{d}^0_{j,k}$ can be estimated as the null vector, we focus on the euclidian norm $||\mathbf{d}_{j,k}||$ of the empirical coefficients and try to fix a threshold for this quantity. In particular, we want to find an estimation procedure that generalizes the 1D universal threshold, which is based on the following result.

Proposition 3.1 *(Donoho et al., 1995) Let $\{X_n\}_n$ be a sequence of i.i.d. $N(0,1)$ random variables and $A_n = \{\max_{i=1,\ldots,n} |X_i| \leq \sqrt{2 \log n}\}$. Then*

$$\mathbb{P}(A_n) \to 1 \text{ for } n \to +\infty.$$

Thanks to Proposition 3.1, we have in fact that if the number n of observations in model (1) is large enough, then the universal threshold $t = \hat{\sigma}_d \sqrt{2 \log n}$ contains with high probability all the coefficients coming from noise; recall in fact that the wavelet transforms $\rho_{j,k}$ of the error have distribution $N(0, \sigma_d^2)$.

In the p-dimensional case, we know that $||\boldsymbol{\rho}_{j,k}/\sigma_d||^2 \sim \chi^2(p)$. We shall thus look for a threshold which contains, with high probability, n observations from a random variable having $\chi^2(p)$ distribution. To find such threshold we exploit the following well-known theorem on random processes (see, e.g., Leadbetter et al. (1983), Theorem 1.5.1).

Theorem 3.1 *Let $\{Y_n\}_n$ be a sequence of i.i.d. random variables with cumulative distribution function F. Let $\{u_n\}_n$ be a real sequence such that*

$$n(1 - F(u_n)) \to \tau \quad \text{for } n \to +\infty,$$

for some $0 \leq \tau < +\infty$. Then

$$P[\max_{1 \leq i \leq n} Y_i \leq u_n] \to e^{-\tau}.$$

Proof.

$$P[\max_{1 \leq i \leq n} Y_i \leq u_n] = \{F(u_n)\}^n =$$

$$= \{1 - (1 - F(u_n))\}^n \sim (1 - \frac{\tau}{n} + o(\frac{1}{n}))^n \to e^{-\tau} \quad \text{for } n \to +\infty.$$

□

Applying Theorem 3.1 it is possible to prove Proposition 3.1 and also the following more general result.

Proposition 3.2 *Let $\{Y_n\}_n$ be a sequence of i.i.d. $\chi^2(p)$ random variables and $A_n = \{\max_{i=1,\ldots,n} Y_i \leq c_p \log n\}$, where*

$$c_p = \begin{cases} 2 & \text{if } p = 1 \\ 3 & \text{if } p \geq 2. \end{cases}$$

Then

$$\mathbb{P}(A_n) \to 1 \text{ for } n \to +\infty.$$

Proof. Denoting by F and f respectively the cumulative distribution function and the density function of $\chi^2(p)$, we have

$$\lim_{n \to +\infty} n(1 - F(c_p \log n)) = \lim_{n \to +\infty} \frac{1 - F(c_p \log n)}{\frac{1}{n}} =$$

$$= \lim_{n\to+\infty} \frac{-f(c_p \log n)\frac{c_p}{n}}{-\frac{1}{n^2}} = \lim_{n\to+\infty} \frac{2^{-p/2}}{\Gamma(p/2)} c_p^{p/2} n (\log n)^{p/2-1} e^{(-\frac{c_p \log n}{2})}$$

$$= \lim_{n\to+\infty} c_p^{p/2} \frac{2^{-p/2}}{\Gamma(p/2)} n (\log n)^{p/2-1} n^{-c_p/2} =$$

$$= \lim_{n\to+\infty} \begin{cases} 2^{1/2} \frac{2^{-1/2}}{\Gamma(1/2)} n (\log n)^{1/2-1} n^{-2/2} = \frac{1}{\Gamma(1/2)} \frac{1}{\sqrt{\log n}} = 0 & \text{if } p = 1 \\ 3^{p/2} \frac{2^{-p/2}}{\Gamma(p/2)} n (\log n)^{p/2-1} n^{-3/2} = 3^{p/2} \frac{2^{-p/2}}{\Gamma(p/2)} \frac{(\log n)^{p/2-1}}{\sqrt{n}} = 0 & \forall p \geq 2. \end{cases}$$

Then, using Theorem 3.1, we obtain

$$P[\max_{1\leq i\leq n} Y_i \leq c_p \log n] \to e^{-0} = 1.$$

□

For $p = 1$ (i.e., when Y_i is the square of a Gaussian random variable), Proposition 3.2 leads to the same threshold supported by Proposition 3.1, i.e., the universal threshold $t = \hat{\sigma}_d \sqrt{2 \log n}$ on $|d_{j,k}|$. In the multidimensional case $p \geq 2$, Proposition 3.2 supports instead the threshold $\hat{\sigma}_d^2 (3 \log n)$ on $||\mathbf{d}_{j,k}||^2$ or, equivalently, the threshold

$$t_p = \hat{\sigma}_d \sqrt{3 \log n}$$

on $||\mathbf{d}_{j,k}||$, where $\hat{\sigma}_d$ is estimated from the coefficients of the finer level via the MAD estimator defined in (3), pooling together the coefficients corresponding to the p directions. A simple estimator for multidimensional case is therefore the following hard-thresholding scheme:

$$\hat{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{0} & \text{if } ||\mathbf{d}_{j,k}|| \leq t_p \\ \mathbf{d}_{j,k} & \text{if } ||\mathbf{d}_{j,k}|| > t_p. \end{cases} \tag{7}$$

3.2. Generalization of soft-thresholding in p dimensions

Likewise hard-thresholding in the 1D case, estimator (7) does not take into account that also the coefficients larger than the threshold contains a component due to noise. To obtain a more refined result, we shall use a soft-thresholding estimator which applies a shrinkage to the coefficients larger than the threshold t_p. However, the generalization to the p-dimensional setting of the simple shrinkage operation usually performed in 1D case is not straightforward.

Recall that the empirical wavelet coefficient $\mathbf{d}_{j,k}$ is a vector of $\mathbb{R}^p$ and

t_p is a threshold on its euclidian norm; in particular, t_p identifies an upper bound for the error contribute to $||\mathbf{d}_{j,k}||$. Starting from the hypothesis that the variance of the error on the coefficients is the same in all p directions, we can consider the direction of the vector $\mathbf{d}_{j,k}$ to be mainly determined by that of the true coefficient $\mathbf{d}^0_{j,k}$. Thus, we choose to estimate $\mathbf{d}^0_{j,k}$ keeping unchanged the direction of the empirical coefficient $\mathbf{d}_{j,k}$ and diminishing its norm by t_p i.e.,

$$\text{if} \quad ||\mathbf{d}_{j,k}|| > t_p \quad \text{then} \quad ||\hat{\mathbf{d}}_{j,k}|| = ||\mathbf{d}_{j,k}|| - t_p$$

with the aim of removing the component of the empirical wavelet coefficients due to noise. Setting $||\hat{\mathbf{d}}_{j,k}|| = ||c\mathbf{d}_{j,k}||$, we get

$$c = 1 - \frac{t_p}{||\mathbf{d}_{j,k}||}$$

so that the soft-thresholding estimator will be

$$\hat{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{0} & \text{if } ||\mathbf{d}_{j,k}|| \leq t_p \\ (1 - \frac{t_p}{||\mathbf{d}_{j,k}||})\mathbf{d}_{j,k} & \text{if } ||\mathbf{d}_{j,k}|| > t_p \end{cases}$$

i.e.,

$$\hat{\mathbf{d}}_{j,k} = \left(1 - \frac{t_p}{||\mathbf{d}_{j,k}||}\right)_+ \mathbf{d}_{j,k}\,.$$

Geometrically, this soft-thresholding procedure works as follows. Consider a p-dimensional sphere with radius t_p and centered in the origin; if the p-dimensional vector $\mathbf{d}_{j,k}$ lies completely inside the sphere, then the estimated wavelet coefficient $\hat{\mathbf{d}}_{j,k}$ is set to $\mathbf{0}$; otherwise, $\hat{\mathbf{d}}_{j,k}$ is obtained from $\mathbf{d}_{j,k}$ by removing the part of $\mathbf{d}_{j,k}$ that lies inside the sphere. Figure 2 gives a visual representation of this procedure for $p = 3$.

4. Simulation studies

In this section we illustrate, via a two-case simulation study, the good performances of the proposed wavelet fitting technique for multi-dimensional functional data, particularly when the true curves to be estimated are characterized by strongly localized features. In the implementation of the technique, we use here the Daubechies wavelet basis with 10 vanishing moments,

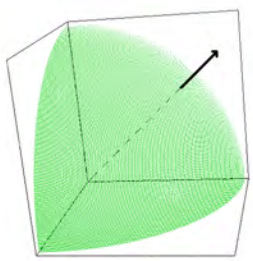


Figure 2: Visual representation in three dimensions of the soft-thresholding procedure: only the part of the vector $\mathbf{d}_{j,k}$ that lies outside the sphere with radius t_p is retained as significant.

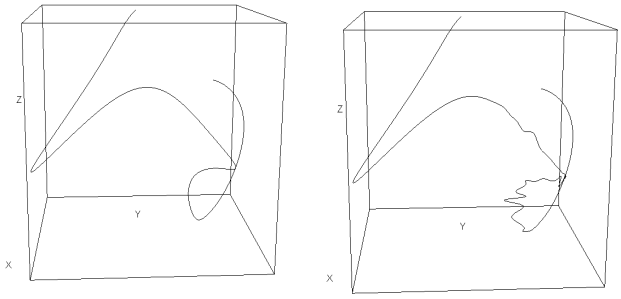


Figure 3: Left panel: curve $\mathbf{c}_1$, first dataset. Right panel: curve $\mathbf{c}_{51}$, second dataset; obtained from curve $\mathbf{c}_1$, left panel, by adding strongly localized features.

because this basis is compactly supported and smooth enough to allow the estimation of second derivatives (see Daubechies, 1988, for details).

As a comparison, we use another locally adaptive regression technique based on free-knot splines, that has been shown to give functionally very accurate estimates of multidimensional curves and their derivatives, even when the curve are characterized by spatially inhomogeneities, having parts where the curve varies more rapidly and others where it varies more slowly; see Sangalli et al. (2009). The comparison is carried out on two different datasets, each consisting of 50 simulated curves. Both the curves in the first and in the second dataset are spatially inhomogeneous and with varying roughness, but the curves in the second dataset also present strongly localized features, which are instead absent in the curves of the first dataset.

First dataset. The curves of the first dataset, $\mathbf{c}_1, \ldots, \mathbf{c}_{50}$, are generated in

the following way. Independently for $i = 1, \ldots, 50$, we generate three order-5 splines, $\mathrm{x}_i(l), \mathrm{y}_i(l), \mathrm{z}_i(l)$, for $l \in [-1, 2]$, with $\mathrm{x}_i, \mathrm{y}_i, \mathrm{z}_i$ having a common knot vector $\mathbf{k}_i = (k_{i\,1}, \ldots, k_{i\,20})$; the locations of the 20 knots in $\mathbf{k}_i$ are obtained via i.i.d. sampling from a uniform distribution on $[-1, 2]$, and the coefficients of the corresponding spline-basis expansions that yield $\mathrm{x}_i, \mathrm{y}_i$ and z_i are obtained via i.i.d. sampling from a Gaussian distribution with mean 0 and variance 0.5. We then apply the following non-linear transformation,

$$x_i(l) = \log_{10}(\mathrm{x}_i(l) + 5) \quad y_i(l) = \log_{10}(\mathrm{y}_i(l) + 5) \quad z_i(l) = \log_{10}(l + \mathrm{z}_i(l) + 5)$$

where the constant 5 is added to make the logarithm well defined; we thus consider the curves $\mathbf{c}_i(l) = \{x_i(l), y_i(l), z_i(l)\}$ for $l \in [0, 1]$. Note that, thanks to this non-linear transformation, the 50 curves $\mathbf{c}_i$, for $i = 1, \ldots, 50$, are no longer splines. Figure 3, left panel, gives a 3D visualization of the first generated curve, $\mathbf{c}_1$. We hence simulate from each curve $\mathbf{c}_i$ on an equispaced grid of $n = 2^8$ points along l, $l \in [0, 1]$, adding independent normally distributed errors $\boldsymbol{\varepsilon} = \{\varepsilon^{[x]}, \varepsilon^{[y]}, \varepsilon^{[z]}\}$ with mean $\mathbf{0} = (0, 0, 0)$ and variance-covariance matrix $\sigma^2 \mathbb{I}_3$, where $\sigma = 2 \cdot 10^{-4}$, thus obtaining a noisy and discrete observation of the curve: $\{(x_{iu}, y_{iu}, z_{iu}) : u = 1, \ldots, n = 2^8\}$. Figure 4 shows the noisy and discrete observation of curve $\mathbf{c}_1$: the top panel displays the three space coordinates $\{x_1(l), y_1(l), z_1(l)\}$ (black), superimposed to sampled data (grey); the center panel displays the first derivatives $\{x_1'(l), y_1'(l), z_1'(l)\}$, superimposed to first central differences of sampled data, and the bottom panel displays the second derivatives $\{x_1''(l), y_1''(l), z_1''(l)\}$, superimposed to second central differences of sampled data. First and second central differences are here and in the following displayed, since they give a rough indication, computable from raw data, of the first and second derivatives behavior.

Second dataset. Let us denote the curves in the second dataset by $\mathbf{c}_{51}, \ldots, \mathbf{c}_{100}$. For $i = 51, \ldots, 100$, the curve $\mathbf{c}_i$ is obtained from the corresponding curve $\mathbf{c}_{i-50}$, in the first dataset, by adding to $\mathbf{c}_{i-50}$ a three-dimensional curve characterized by strongly localized features, generated as follows. Consider three-dimensional wavelets having coefficients $\mathbf{d}_{j,k} = (d_{j,k}^{[\mathrm{x}]}, d_{j,k}^{[\mathrm{y}]}, d_{j,k}^{[\mathrm{z}]})$, where $d_{j,k}$ are coefficients associated to the Daubechies wavelet functions with 10 vanishing moments. Independently for $i = 51, \ldots, 100$, we randomly select 6 wavelet coefficients $\mathbf{d}_{j,k}$ among levels $j = 4$ and $j = 5$, and sample the values of these coefficients from a Gaussian distribution with mean $\mathbf{0}$ and variance-covariance matrix $0.0003^2 \mathbb{I}_3$; the remaining coefficients are set to $\mathbf{0}$. We then apply the inverse fast wavelet transform to the three

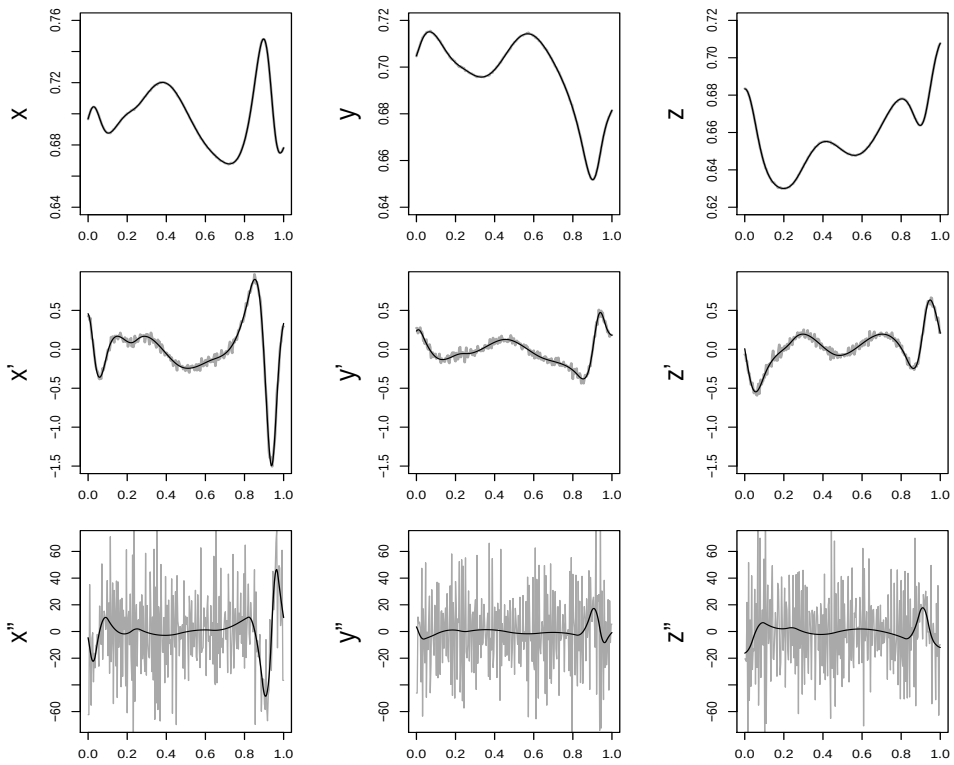


Figure 4: Space coordinates and derivatives of the curve $\mathbf{c}_1$, first dataset. Top: three space coordinates $\{x_1(l), y_1(l), z_1(l)\}$ (black), superimposed to raw data (grey). Center: first derivatives $\{x_1'(l), y_1'(l), z_1'(l)\}$ (black), superimposed to first central differences (grey). Bottom: second derivatives $\{x_1''(l), y_1''(l), z_1''(l)\}$ (black), superimposed to second central differences (grey).

coordinates of this wavelet representation to obtain a 3D wavelet $\mathbf{w}_i$ on an equispaced grid of 2^8 points over $[0, 1]$. The curve $\mathbf{c}_i$ is hence given by

$$\mathbf{c}_i(l) = \mathbf{c}_{i-50}(l) + \log(6\mathbf{w}_i(l) + \mathbf{2}) \qquad t \in [0, 1].$$

Note that, also in this case a non-linear transformation (log) is applied to the added components (so that the obtained curves are neither splines, nor wavelets, nor a combination of the two). Figure 3, right panel, gives a 3D visualization of the first generated curve of this second dataset, $\mathbf{c}_{51}$. Likewise for the first dataset, we hence simulate from each curve $\mathbf{c}_i$, for $i = 51, \ldots, 100$, on the equispaced grid of $n = 2^8$ points over $[0, 1]$, adding independent nor-

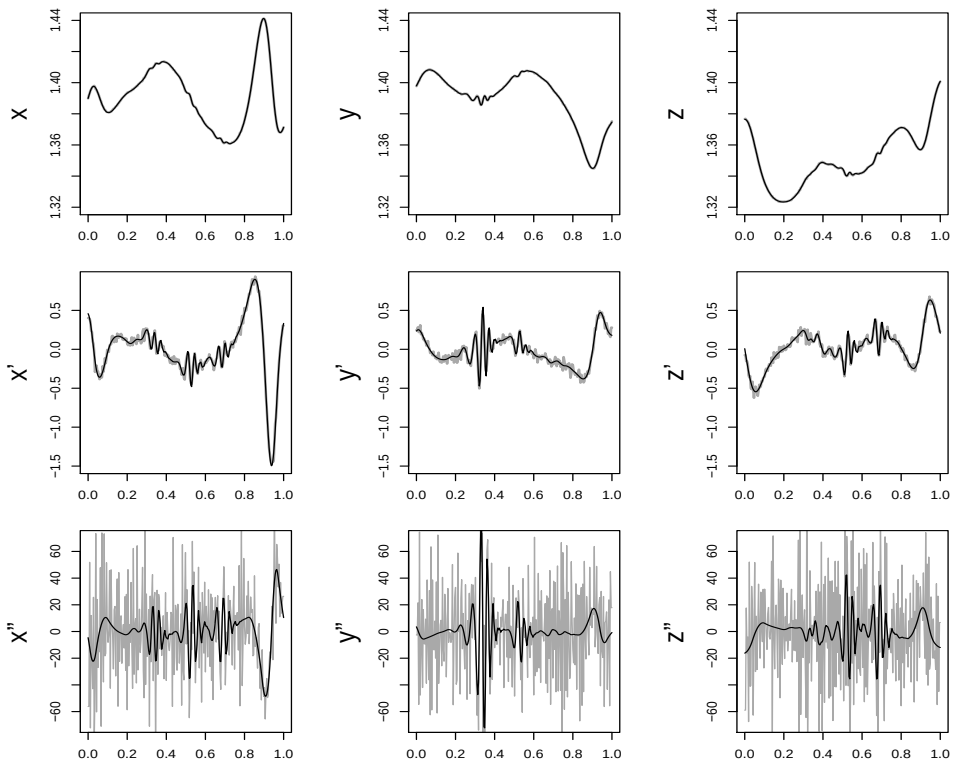


Figure 5: Space coordinates and derivatives of the curve $\mathbf{c}_{51}$, second dataset. Top: three space coordinates $\{x_{51}(l), y_{51}(l), z_{51}(l)\}$ (black), superimposed to raw data (grey). Center: first derivatives $\{x'_{51}(l), y'_{51}(l), z'_{51}(l)\}$ (black), superimposed to first central differences (grey). Bottom: second derivatives $\{x''_{51}(l), y''_{51}(l), z''_{51}(l)\}$ (black), superimposed to second central differences (grey).

mally distributed errors with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2\mathbb{I}_3$, where $\sigma = 2 \cdot 10^{-4}$, thus obtaining a noisy and discrete observation of the curve: $\{(x_{iu}, y_{iu}, z_{iu}) : u = 1, \ldots, n = 2^8\}$. Figure 5 shows the noisy and discrete observation of curve $\mathbf{c}_{51}$, together with its true space coordinates and derivatives; a 3D visualization of $\mathbf{c}_{51}$ is given in Figure 3, right panel.

To compare the different smoothing methods on equal terms, we shall compare the estimates they provide for the same level of data-adaptation. We measure the data-adaptation of an estimate $\hat{\mathbf{c}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)$ of the curve $\mathbf{c}_i = (x_i, y_i, z_i)$ by the Root Average Squared Error of the estimate with

respect to the *data*:

$$\mathrm{RASE}_{DATA}(\hat{\mathbf{c}}_i) =$$
$$= \sqrt{\frac{1}{n-2m}\sum_{u=1+m}^{n-m}\left[\left(x_{iu}-\hat{x}_i(l_u)\right)^2+\left(y_{iu}-\hat{y}_i(l_u)\right)^2+\left(z_{iu}-\hat{z}_i(l_u)\right)^2\right]}$$

with m=15 boundary grid points not considered in the computation of RASE_{DATA}. The performances of the different methods will thus be measured by the errors with respect to the *true* curve. In particular, we shall consider the following goodness of fit measures.
Root Mean Squared Error of the curve estimate with respect to true curve:

$$\mathrm{RMSE}_{TRUE}(\hat{\mathbf{c}}_i) =$$
$$= \sqrt{\frac{1}{n-2m}\sum_{u=1+m}^{n-m}\left[\left(x_i(l_u)-\hat{x}_i(l_u)\right)^2+\left(y_i(l_u)-\hat{y}_i(l_u)\right)^2+\left(z_i(l_u)-\hat{z}_i(l_u)\right)^2\right]};$$

Root Mean Squared Error of the estimate of first derivative with respect to true first derivative:

$$\mathrm{RMSEder1}_{TRUE}(\hat{\mathbf{c}}_i) =$$
$$= \sqrt{\frac{1}{n-2m}\sum_{u=1+m}^{n-m}\left[\left(x_i'(l_u)-\hat{x}_i'(l_u)\right)^2+\left(y_i'(l_u)-\hat{y}_i'(l_u)\right)^2+\left(z_i'(l_u)-\hat{z}_i'(l_u)\right)^2\right]};$$

Root Mean Squared Error of the estimate of second derivative with respect to true second derivative:

$$\mathrm{RMSEder2}_{TRUE}(\hat{\mathbf{c}}_i) =$$
$$= \sqrt{\frac{1}{n-2m}\sum_{u=1+m}^{n-m}\left[\left(x_i''(l_u)-\hat{x}_i''(l_u)\right)^2+\left(y_i''(l_u)-\hat{y}_i''(l_u)\right)^2+\left(z_i''(l_u)-\hat{z}_i''(l_u)\right)^2\right]}.$$

Note that we are here using the term "Root Average Squared Error" to denote errors with respect to observed *data*, and the term "Root Mean Squared Error" to denote errors with respect to *true* curve values. Notice also that in the computation of this goodness of fit measures, we disregard m boundary grid points, because both spline and wavelet estimates may display some

irregularity in the behavior at the boundary. For instance, it is known that increasing the smoothness and regularity of the wavelet basis, thus enlarging its support, may affect negatively the estimate near to the boundary (see, e.g., Antoniadis et al., 1994). When there is an interest in an accurate estimation near the boundary, some corrections can be introduced (see, e.g., Oh et al., 2001). We do not pursue this here because in the applied problem which motivated our research the range of the grid of observation points is much larger than the range of interest, thus making this issue not relevant for our problem; see Section 6.

Analogously to the selection of the smoothing parameter in Sangalli et al. (2009), we choose the level j_0 in our wavelet fitting technique according to qualitative considerations, along a Goldilocks approach where this smoothness parameter is selected neither to undersmooth nor to oversmooth the data, but to lead to a just right degree of smoothness for the data under analysis. In particular, we choose a level j_0, common to all curves, by visually comparing the first derivative estimates, corresponding to various j_0, to the first central differences; we thus select the level j_0 that allows, for most curves, to fully capture the strong peaks and troughs in the central differences, without fitting also the very-high frequency variation (due to noise). This qualitative choice was ex-post comforted by noticing that the chosen j_0 yielded estimates whose RASE_{DATA} has the same order of magnitude of the error standard deviation σ used for data generation. After selection of the level j_0 for the wavelet estimates, we then choose the smoothing parameter in free-knot splines, common to all curves in each dataset, in order to obtain comparable values of RASE_{DATA}.

Figure 6 shows the boxplots of RASE_{DATA}, RMSE_{TRUE}, RMSEder1_{TRUE} and RMSEder2_{TRUE} for the estimates of the curves in the first dataset, obtained by free-knot regression splines and wavelet smoothing. In this case, the two methods gives comparable results: if allowed the same level of data-adaptation, the estimates provided by the two methods have comparable errors with respect to the true curves and their derivatives. Figure 7 shows instead the results found for the second dataset. In this case, wavelet fitting provides more accurate estimates. In fact, wavelet estimates, even if allowed comparable or even worse levels of data-adaptation, still provide better estimates of the curves and their derivatives. The comparative advantage of the wavelet estimation method over free-knot splines is here due to the fact that wavelets can better capture the strongly localized features of the curves.

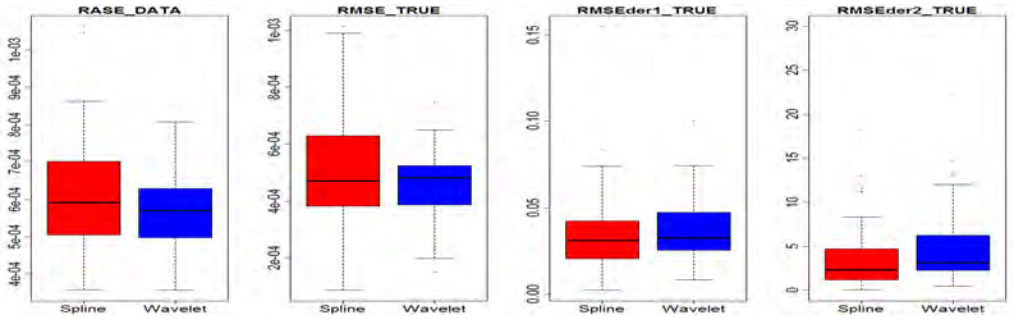


Figure 6: First Dataset. Boxplots of RASE_{DATA}, RMSE_{TRUE}, RMSEder1_{TRUE} and RMSEder2_{TRUE}.

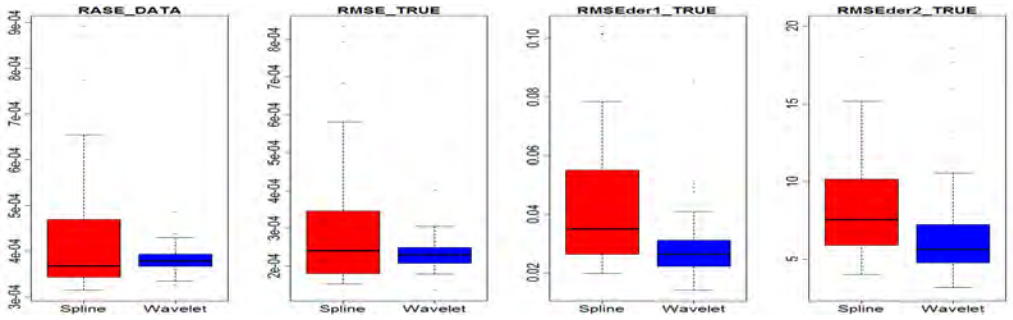


Figure 7: Second Dataset. Boxplots of RASE_{DATA}, RMSE_{TRUE}, RMSEder1_{TRUE} and RMSEder2_{TRUE}.

Some of these features may instead be missed even by a locally adaptive estimation technique such free-knot regression splines, that has been shown to provide very accurate estimates of spatially inhomogeneous curves. Figures 8 and 9 illustrate this issue. Figure 8 displays the estimates of the three space coordinates and of the first and second derivatives of $\mathbf{c}_{51}$, the first curve in the second dataset; wavelet estimates (blue) and spline estimates (red) are superimposed to the true curve (black). The figure highlights that some strongly localized features of the curve, that are evidenced by large oscillations of the derivatives, are well captured by wavelet estimates, whilst are missed by spline estimates, that smooth them away. This can be better appreciated in the left panel of Figure 9 that, as an example, zooms in the estimates of the second derivative in the z direction, in correspondence of one of these features (blue, wavelet estimate; red: spline estimate; black: true curve). The right panel of the same figure displays the residuals $z''_{51} - \widehat{z}''_{51}$ for the

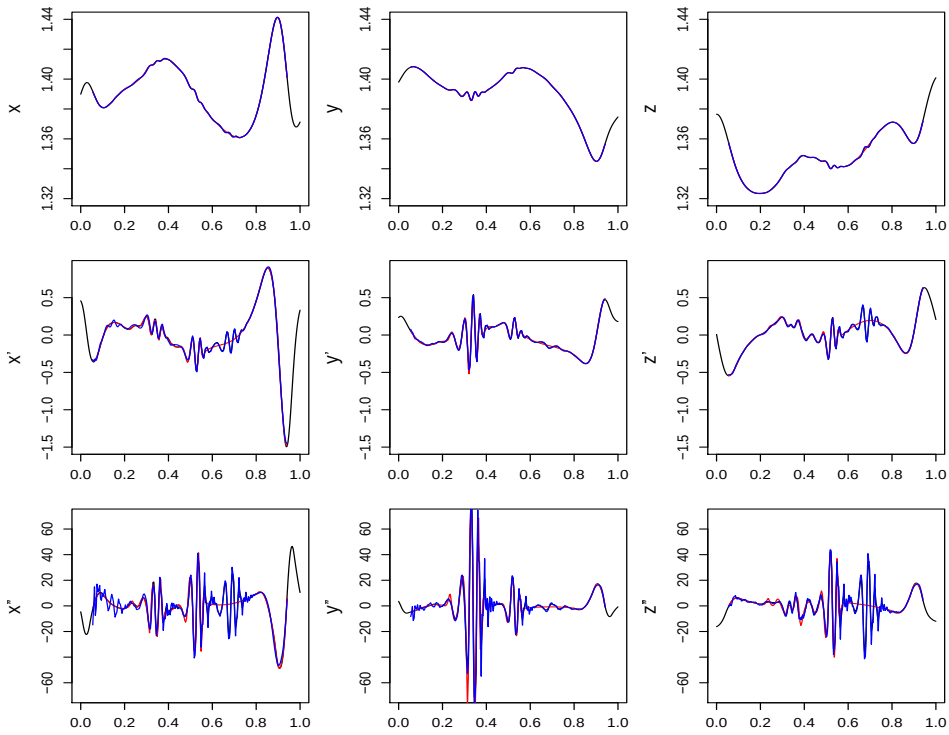


Figure 8: Estimates of the three space coordinates (top) and first and second derivatives (center and bottom) of $\mathbf{c}_{51}$, obtained by wavelet smoothing (blue line) and free-knot regression splines (red), superimposed to true curve (black)

wavelet estimate (blue) and the spline estimate (red); also this figure highlights the smaller errors committed by wavelet estimates in correspondence of the strongly localized features. It should be mentioned that considering higher levels of data-adaptation does not improve spline estimates, because the estimates start interpolating also the noise.

This two-case simulation study shows that the proposed wavelet-based estimation procedure for multidimensional curves is particularly attractive when the data are characterized by strongly localized features. In the absence of these characteristics, the proposed method provides estimates that have a level of accuracy comparable to that of free-knot regression splines, the latter technique having though the advantage of not being bound to evenly spaced grids of 2^J points.

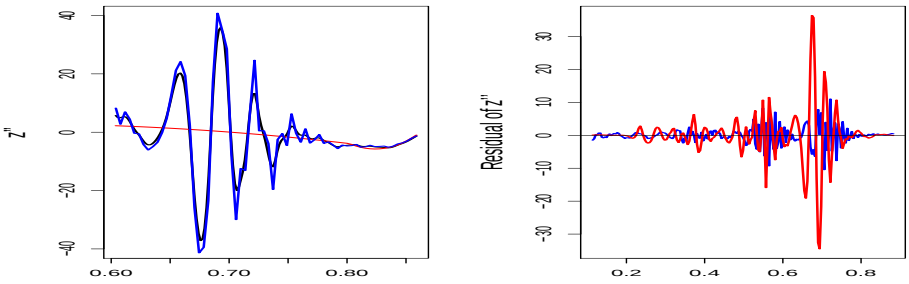


Figure 9: Left: Zoom of the estimates of the second derivative in the z direction of $\mathbf{c}_{51}$, in correspondence of one of the strongly localized feature; wavelet estimate (blue line) and spline estimate (red), superimposed to true derivative (black). Right: Residuals $z''_{51} - \widehat{z}''_{51}$ for the wavelet estimate (blue) and the spline estimate (red).

5. Errors correlated in the p dimensions

The method proposed in Section 3 for the estimation of multidimensional wavelet coefficients assumes that the p components of the error in the p dimensions are incorrelated, i.e., $Var(\boldsymbol{\varepsilon}_i) = \sigma^2 \mathbb{I}_p$. However, in many applications it might be useful to allow for correlation of the components of error in the various directions, since these may capture the same source of noise. This is the case, for instance, of the ECG data, whose analysis has motivated our research. In fact, as it will be clearer from next section, the nature of these data is such that its $p = 8$ functional coordinates represent different projections of the same physical dynamic; for this reason, we expect that also the errors involve different projections of the noise on the original physical signal, as well as some independent components due to the measurement device and other sources or noise.

In this Section we thus generalize the estimation method proposed in Section 3, to account for correlation of the components of the error in the p-directions. The model we assumes is the same as in (6), always with i.i.d. multinormal errors $\boldsymbol{\varepsilon}_i$ with mean the null vector $\mathbf{0} \in \mathbb{R}^p$, but now having some general variance-covariance matrix S. Since we use an orthogonal wavelet transform, each vector of wavelets coefficients $\mathbf{d}_{j,k}$ also has a multivariate normal distribution with mean $\mathbf{0} \in \mathbb{R}^p$ and variance-covariance matrix S. Therefore, we have that $\mathbf{d}'_{j,k} S^{-1} \mathbf{d}_{j,k} \sim \chi^2(p)$, so that the threshold $3 \log n$ is appropriate for this quantity, according to Proposition 3.2. The main difference, with respect to the case considered in Section 3, is that we now

have to estimate the full variance-covariance matrix S of the p-dimensional coefficients, and not simply the value σ. However, likewise for incorrelated error components, we can estimate the matrix S from level $J-1$ coefficients; in particular, in this work we use the simple estimate provided by the sample covariance matrix of coefficients $\mathbf{d}_{J-1,k}$ (a robust estimator could also be used; see, e.g, Brown et al., 2005).

It should be mentioned that the wavelet fitting technique here described bears some similarities with the method introduced in Downie and Silverman (1998), the context though being different. In Downie and Silverman (1998), the authors aim at estimating a monodimensional function via a decomposition in multiple wavelet bases, to obtain better estimates with respect to the usual wavelet decomposition; we instead pursue the opposite goal of estimating a multidimensional functional data and to this goal we use the same wavelet decomposition in each coordinate direction, for the reasons highlighted earlier in the paper.

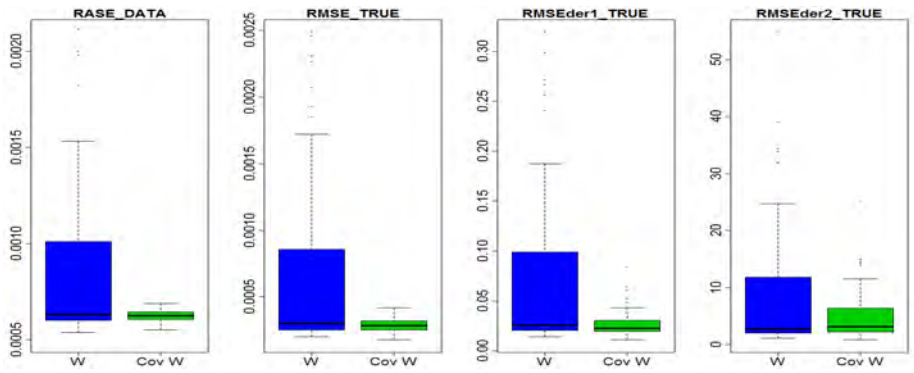


Figure 10: Boxplots of RASE_{DATA}, RMSE_{TRUE}, RMSEder1_{TRUE} and RMSEder2_{TRUE} for wavelet (left) and wavelet with estimation of the covariance structure (right).

To illustrate the good performances of the proposed technique when the p components of the error are correlated, we consider the same curves used to obtain Dataset 1, but with errors $\boldsymbol{\varepsilon}_i$ i.i.d. multinormal with $\mathbf{0}$ mean and variance-covariance matrix $S = (2, 0.25, 1; 0.25, 0.35, 3; 1, 3, 1.5) \times 10^{-7}$. Figure 10 shows the boxplots of RASE_{DATA}, RMSE_{TRUE}, RMSEder1_{TRUE} and RMSEder2_{TRUE} on this dataset, for the multidimensional wavelet technique proposed in Section 3 and for its modification described in the current section. This simple simulation study shows that, when the components of error in the p dimensions are indeed correlated, appropriately taking this correla-

tion into account leads to far more robust estimates than if this correlation was neglected.

6. Application to ECG data

In this section we apply the proposed multidimensional wavelet fitting technique for the estimation of ECG records collected by the 118 Dispatch Center in Milano within the PROMETEO project; see Ieva et al. (2011) and Ieva (2011). These ECG traces have been tele-transmitted from ambulances during emergency rescue operations (in Italy most emergency rescue operations are connected to ischemic heart diseases, that alone cause more than 40% of the overall deaths in the country). One of the main goals of the PROMETEO project is the development of statistical tools capable of classifying ECG traces based on their morphology, without the need for clinical criteria; this would help the early detection of heart failures and thus facilitate a positive outcome. Indeed, well-timing is fundamental to obtain good prognosis for this pathologies, and early diagnosis via ECG tele-transmission is the most important factor to access an effective reperfusive treatment, as shown for instance in Ieva and Paganoni (2010) and Grieco et al. (2011), who analyzed data recorded by a network of more than 20 Cardiology Units of the Milanese urban area together with the 118 Dispatch Center.

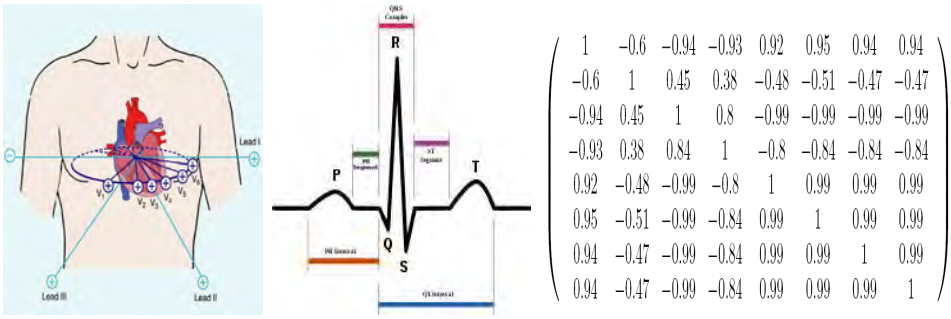
The processing of ECG records as functional data is becoming increasingly important with the advent of statistical techniques that exploit curves shapes in the analysis of these records (see, e.g., Boudaoud et al., 2007; Trigano et al., 2010). As mentioned in the Introduction, these data have a multidimensional nature, because the ECG records provide potential differences, named leads, between multiple electrodes. In particular, ten electrodes are used for a standard "twelve-leads" ECG. Among the twelve leads provided by the experimental device, eight leads are jointly needed to capture all the information concerning the complex heart dynamics:

- Leads I and II jointly describe heart activity on the sagittal plane; they are called limb leads because the electrodes for these signals are located on the limbs.

- Leads V1, V2, V3, V4, V5 and V6 jointly record heart electric activity on the horizontal plane; these leads are called precordial and the electrodes that measure them are placed on the chest.

These eight leads thus constitute projections in different directions of the same physical dynamics. The left panel of Figure 11 shows the positions of electrodes and leads, whilst the central panel of the same figure gives the scheme of the typical structure of a physiological Lead I, indicating the so-called P wave, QRS complex and T wave.

The considered dataset, coming from PROMETEO datawarehouse, includes the ECG records of $n = 198$ subjects, among which 101 are physiological, whilst 97 are affected by Bundle Branch Block, an abnormality in cardiac conduction; in particular, 49 subjects are affected by Right Bundle Branch Block (RBBB) and 48 by Left Bundle Branch Block (LBBB). Figure 12 displays the raw data of the eight significant ECG leads for a patient affected by RBBB; superimposed, in blue color, are the estimates of this eight-dimensional functional data, yielded by the technique detailed in Sections 3 and 5. The estimates are obtained using a Daubechies wavelet basis with 10 vanishing moments, and are limited to the interval corresponding to the central $2^{10} = 1024$ observation points, that anyway covers all important features of the ECG record.



$$\begin{pmatrix}
1 & -0.6 & -0.94 & -0.93 & 0.92 & 0.95 & 0.94 & 0.94 \\
-0.6 & 1 & 0.45 & 0.38 & -0.48 & -0.51 & -0.47 & -0.47 \\
-0.94 & 0.45 & 1 & 0.8 & -0.99 & -0.99 & -0.99 & -0.99 \\
-0.93 & 0.38 & 0.84 & 1 & -0.8 & -0.84 & -0.84 & -0.84 \\
0.92 & -0.48 & -0.99 & -0.8 & 1 & 0.99 & 0.99 & 0.99 \\
0.95 & -0.51 & -0.99 & -0.84 & 0.99 & 1 & 0.99 & 0.99 \\
0.94 & -0.47 & -0.99 & -0.84 & 0.99 & 0.99 & 1 & 0.99 \\
0.94 & -0.47 & -0.99 & -0.84 & 0.99 & 0.99 & 0.99 & 1
\end{pmatrix}$$

Figure 11: Left: Scheme of the directions along which the potential difference is measured for every lead. Center: Template of a physiological ECG record on Lead I. Right: Estimated correlation matrix for the eight components of the error, for the patient whose ECG traces are shown in Figure 12

.

Figure 11, right panel, reports the estimated correlation of the eight components of error for this patient (i.e., the sample correlation matrix of the empirical wavelets coefficients $\mathbf{d}_{J-1,k}$), whose corresponding estimated covariance matrix has been used to obtain this patient ECG estimate; the matrix highlights the strong correlation structure of the components of error on the different leads; as noted in the previous Section, this was expected,

since, given the nature of these data, the errors involve different projections of the noise on the original physical signal. Figure 13 shows for the same patient the corresponding estimated first derivatives, superimposed to first central differences. The obtained estimates of the eight-lead traces and of their derivatives, for the 198 records in the PROMETEO database, are the starting point of the analysis described in Ieva et al. (2011). In particular, with the aim of developing a ECG classification procedure based on ECG morphology, Ieva et al. (2011) carry out an unsupervised clustering of the the 198 ECG estimates, proceeding as follows. First, the estimated ECG traces of the various subjects (and, correspondingly, their estimated first derivatives) are aligned via landmark registration, where the employed landmarks identify respectively the P wave, the QRS complex and the T wave. The analysis thus focus on the registered estimated ECG traces and their first derivatives, for the 198 subjects, restricted to the time interval T going from the offset of the P wave to the offset of the T wave,

$$\mathbf{F}_j(t) = \{F_j^{[p]}(t)\}_{p=1}^8 = \big(I_j(t), II_j(t), V1_j(t), V2_j(t), V3_j(t), V4_j(t), V5_j(t), V6_j(t)\big)$$
$$\mathbf{F'}_j(t) = \{F_j^{\prime[p]}(t)\}_{p=1}^8$$

where $t \in T$ and $j = 1, \ldots, 198$. Next, the aligned $\mathbf{F}_j(t)$ are clustered via a multidimensional functional k-mean algorithm using the following distance between eight-dimensional curves:

$$d\big(\mathbf{F}_j, \mathbf{F}_l\big) = \sqrt{\sum_{p=1}^{8} \Big\{ \int_T \big(F_j^{[p]}(t) - F_l^{[p]}(t)\big)^2 dt + \int_T \big(F_j^{\prime[p]}(t) - F_l^{\prime[p]}(t)\big)^2 dt \Big\}}$$

that is the natural distance in the Hilbert space $H^1(T; \mathbb{R}^8)$. As shown in Ieva et al. (2011), this clustering procedure suggests the presence of three clusters of curves; it turns out that these clusters identify the three groups of physiological, RBBB and LBBB traces, with a confusion matrix where only 11 out of 198 are missclassified. The authors also show via a cross-validation analysis that the classification based on the distance above leads to the smallest missclassification cost, compared to analogous procedures based on distances involving only the registered estimated ECG traces or only their registered first derivatives. See Ieva et al. (2011) for details. The very interesting results achieved within this study, that first uses estimates provided by our wavelet-based multidimensional curve fitting method, constitute another illustration

of its efficiency and good performances.

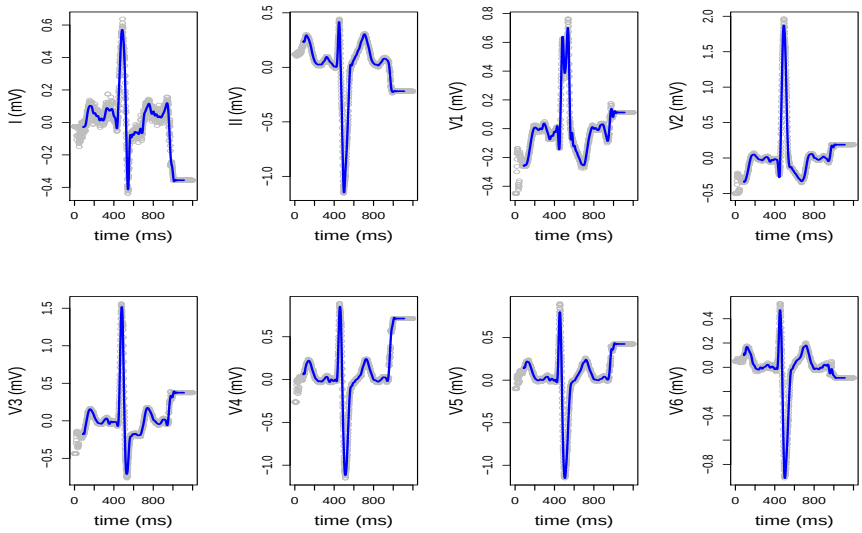


Figure 12: Eight significant leads in a twelve-leads ECG for a patient affected by Right Bundle Branch Block; raw data (grey) and multidimensional wavelet estimate (blue).

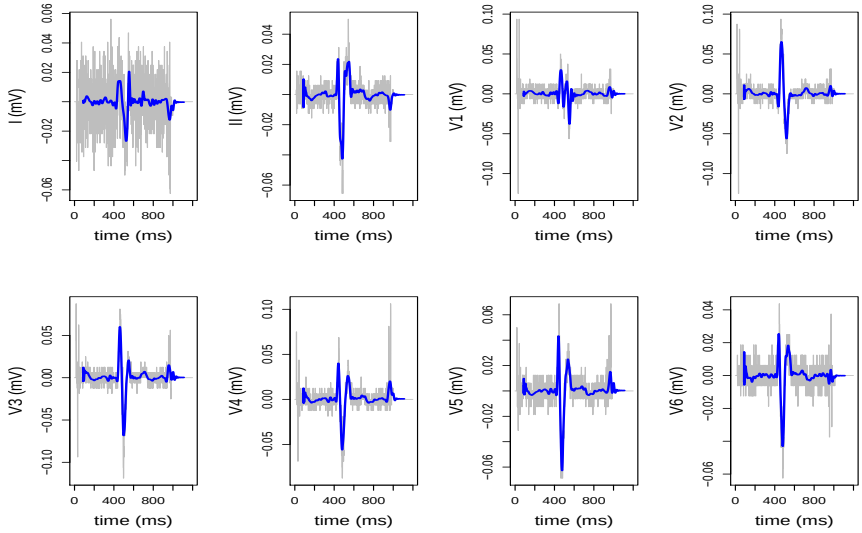


Figure 13: Estimate of first derivatives for the eight leads (blue), superimposed to first central differences of raw data (grey).

7. Discussion

We have described a wavelet-based method for the accurate estimation of multidimensional curves and their derivatives; the method also allows for correlation of the components of error in the p dimensions. As illustrated by means of simulation studies, the proposed estimation technique is particularly attractive when the multidimensional functional data are characterized by strongly localized features. In particular, the motivating application for this research concerned the fitting of multi-lead ECG records, an important example of real life data having this nature. The estimated ECG traces and their derivatives are the starting point of the analysis carried out in Ieva et al. (2011), where they undergo an unsupervised clustering, based on a measure between curves that involves both the curve and its first derivatives, with the aim distinguishing physiological and pathological traces.

Thanks to the estimation of the curve derivatives, the proposed procedure can in fact also be used as data smoothing step prior to discrimination procedures based on semi-metrics in Hilbert spaces, such as those described in Ferraty and Vieu (2003). Always in a function classification setting, the described method could be used to extend to the multidimensional case the scope of the classification procedures referred to in the Introduction. Moreover, this wavelet based technique can for instance also be implemented in the framework of functional regression (see Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Ferraty et al., 2007, and reference therein); e.g., it can be used to fruitfully generalize to the multidimensional case the wavelet regression model illustrated in Aguilera et al. (2008). In general, the proposed method can be of interest, as smoothing step prior to the analysis or combined directly within the functional data analysis procedures, for all those application where, because of the nature of the multidimensional functional data, a wavelet-based description of the data is attractive.

The R codes (R Development Core Team, 2009) for the implementation of the proposed method, as well as one of the simulated examples, are available at `http://mox.polimi.it/~sangalli/`.

Acknowledgments
We are very grateful to Piercesare Secchi and James O. Ramsay for a careful reading of the present manuscript and many helpful suggestions. We would also like to thank Marco Verani for constructive discussions. The data analyzed in Section 6 have been provided by 118 Milan Dispatch Center and Mortara Rangoni Europe s.r.l..; we wish to thank Anna Maria Paganoni, leader of the statistical

group within the PROMETEO Project, for support on the analysis of these data. This work has been funded by the Research Program Dote Ricercatore Politecnico di Milano - Regione Lombardia, research project "Functional data analysis for life sciences", and by MIUR Ministero dell'Istruzione dell'Università e della Ricerca, FIRB research project "Advanced statistical and numerical methods for the analysis of high dimensional functional data in life sciences and engineering".

References

Aguilera, A., Ocaña, F. and Valderrama, M. (2008), "Estimation of Functional Regression Models for Functional Responses by Wavelet Approximation", in Functional and Operatorial Statistics, Contributions to Statistics, Springer Physica-Vergal, 15–21.

Antoniadis, A., Brossat, X., Cugliari, J. and Poggi, J.M. (2010), "Clustering Functional Data Using Wavelets", Electronic Proceedings of COMPSTAT'2010, SpringerLink, 697–704.

Antoniadis, A., Gregoire, G. and McKeague, I.W.(1994), " Wavelet Methods for Curve Estimation", Journal of the American Statistical Association, **89**, 1340–1353.

Berlinet, A., Biau, G. and Rouvière, L. (2008), "Functional Supervised Classification with Wavelets", Annales de l'I.S.U.P., **52**, 61–80.

Beylkin, G., Coifman, R. and Rokhlin, V. (1991), "Fast wavelet transforms and numerical algorithms I", Communications on Pure and Applied Mathematics **44**, 141-183.

Boudaoud, S., Rix, H., Meste, O., Heneghan, C. and O'Brien, C. (2007), "Corrected Integral Shape Averaging Applied to Obstructive Sleep Apnea Detection from the Electrocardiogram", EURASIP Journal on Advances in Signal Processing Volume 2007.

Brown, C.L., Brcich, R.F. and Debes, C. (2005), "Adaptive M-estimators for use in structured and unstructured robust covariance estimation", 2005 IEEE/SP 13th Workshop on Statistical Signal Processing, 573 – 578.

Cai, T.T. and Zhou, H.H. (2009), "A data-driven block thresholding approach to wavelet estimation", Annals of Statistics, **37**, 569–595.

Daubechies, I. (1988), "Orthonormal basis of compactly supported wavelets", Communictions on Pure and Applied Mathematics, **41**, 909–996.

Donoho, D.L. and Johnstone, I.M. (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage", Journal of the American Statistical Association, **90**, 1200–1224.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995), "Wavelet Shrinkage: Asymptopia", Journal of the Royal Statistical Society, Ser. B **57**, 301–369.

Downie, T.R. and Silverman, B.W. (1998), "The discrete multiple wavelet transform and thresholding methods", IEEE Transactions in Signal Processing, **46**, 2558–2561.

Ferraty, F., Mas, A. and Vieu, P. (2007), "Nonparametric regression on functional data: inference and practical aspects", Australian & New Zealand Journal of Statistics, **49**, 267–286.

Ferraty, F. and Vieu, P. (2003), "Curves discrimination: a nonparametric functional approach", Computational Statistics & Data Analysis, **44**, 2003, 161–173.

Ferraty, F. and Vieu, P. (2006), "Nonparametric functional data analysis: theory and practice", Springer Series in Statistics, Springer.

Frías, M.P. and Ruiz-Medina, M.D. (2011), "Computing functional estimators of spatiotemporal long-range dependence parameters in the spectral-wavelet domain", Journal of Statistical Planning and Inference,**141**, 2417–2427.

Grieco, N., Ieva, F. and Paganoni, A.M. (2011), "Performance assessment using mixed effects models: a case study on coronary patient care", IMA Journal of Management Mathematics, in press. Available at: http://imaman.oxfordjournals.org/content/early/2011/05/27/imaman.dpr007.full.pdf

Ieva, F. and Paganoni, A.M. (2011), "Depth Measures For Multivariate Functional Data", Tech. Rep. MOX 37/2011, Dipartimento di Matematica, Politecnico di Milano. Available at: http://mox.polimi.it/it/progetti/pubblicazioni/view.php?id=314&en=en

Ieva, F., Paganoni, A.M. (2010),"Multilevel models for clinical registers concerning STEMI patients in a complex urban reality: a statistical analysis of MOMI^2 survey", Communications in Applied and Industrial Mathematics, **1**, 128 - 147.

Ieva, F., Paganoni, A.M., Pigoli, D., and Vitelli, V. (2011), "Multivariate Functional Clustering for the Morphological Analysis of ECG Curves". Quaderni di Dipartimento, Dipartimento di Matematica, Politecnico di Milano. http://www1.mate.polimi.it/biblioteca/qddview.php?id=1434&L=i

Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983), Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York.

Leung, A.K., Chau F. and Gao, J. (1998), "Wavelet Transform: a Method for Derivative Calculation in Analytical Chemistry", Analitical Chemistry, **70**, 5222–5229.

Mallat, S. (1999), A Wavelet Tour of Signal Processing, Accademic Press.

Nason, G.P. (2008), Wavelet Methods in Statistics with R, Springer, New-York.

Oh, H-S., Naveau, P. and Lee, G. (2001) "Polynomial boundary treatment for wavelet regression", Biometrika, **88**, 291-298.

Pigoli, D. and Sangalli, L.M. (2011), "Wavelets smoothing for multidimensional curves", in Recent Advances in Functional Data Analysis and Related Topics, Contributions to Statistics, Springer Physica-Vergal, 255–262.

R Development Core Team (2009), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Ramsay, J.O. and Silverman, B.W. (2005), Functional data analysis (2nd ed.), Springer, New York.

Sangalli, L.M., Secchi, P., Vantini S. and Veneziani, A.(2009), "Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines", Journal of the Royal Statistical Society, Ser. C, **58**, 285–306.

Strang, G. and Nguyen, T. (1996), Wavelets and Filter Banks (2nd ed.), Wellesley Cambridge Press.

Timmermans, C., Delsol, L. and von Sachs, R. (2011), "Bases Giving Distances. A New Semimetric and its Use for Nonparametric Functional Data Analysis", in Recent Advances in Functional Data Analysis and Related Topics, Contributions to Statistics, Springer Physica-Vergal, 307–313.

Trigano, T., Isserles, U., Montagu, T. and Ritov, Y. (2010), "Semiparametric curve alignment and shift density estimation: ECG data processing revisited", Signal Processing, Sebastian Miron (Ed.), INTECH.

Wang, X., Ray, S. and Mallick, B.K. (2007), "Bayesian Curve Classification Using Wavelets", Journal of American Statistical Association, **102**, 962–973.

Yang, X. and Nie, K. (2008), "Hypothesis testing in functional linear regression models with Neyman's truncation and wavelet thresholding for longitudinal data", Statistics in Medicine, **27**, 845-863.