MOX–Report No. 09/2011

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

PIGOLI, D.; SANGALLI, L.

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

Davide Pigoli, Laura M. Sangalli

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
davide.pigoli@mail.polimi.it
laura.sangalli@polimi.it

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, we compare this novel method 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.

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. 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 in the estimation of the continuous functional data starting from its discrete observation. In this process the functional basis plays an essential role. Usual choices are Fourier bases and spline bases (see, e.g., Ramsay and Silverman, 2005). 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. This issue has restricted their application to a

1

small part of the FDA field. To overcome this limitation, we resort here to a numerical method that allows to obtain derivatives of wavelet estimated data. We moreover extend traditional wavelet estimators to curves in general dimensions; this requires the development of a new estimation procedure which takes into account simultaneously all the space coordinates of the multidimensional curve.

The paper is organized as follows. In Section 2 we briefly recall wavelet bases. In Section 3 we describe a numerical method that allows to compute point-wise values of a wavelet and its derivatives, even in the absence of a close analytical form for the wavelet basis. Section 4 reviews wavelet smoothing for one dimensional functional data; in this section, we also describe the choice of the translation of the wavelet basis functions, that minimizes the approximation error in the wavelet representation of a function. Section 5 accurately extends wavelet-based estimation techniques to the case of curves in more than one dimension. Section 6 illustrates the good performances of the proposed technique, especially in the case of multidimensional functional data characterized by strongly localized features. An important example of real life data having this nature is given by multi-lead Electro Cardio Grams (ECG); Section 7 shows an application to a dataset of multi-lead ECGs collected by the 118 Dispatch Center (the medical emergency unit) in Milano, Italy. Finally, some conclusive considerations are drawn in Section 8.

All simulations and analysis of real data are performed in R 2.10.1 (R Development Core Team, 2009), with extensive use of the package `wavethresh` (Nason, 2010).

2 An overview on wavelets

In this section 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. *$\{l \mapsto f(l)\} \in V_j \quad \Leftrightarrow \quad \{l \mapsto f(2l)\} \in V_{j+1}$*

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

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

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

where $< \cdot, \cdot >$ is the scalar product in $L^2(\mathbb{R})$, $s_{j_0,k} :=< f, \varphi_{j_0,k} >$ and $d_{j,k} :=< f, \psi_{j,k} >$. 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 del dilation/refinement equations:

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

$$\psi(l) = \sum_k g_k \sqrt{2} \varphi(2l - k) \tag{1}$$

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.

3 Wavelet derivatives

Wavelet bases have already proved to be very useful in functional data analysis, thanks largely to their natural local-adaptivity, that allows them to accommodate a wide variety of functional forms. See for instance Antoniadis et al. (2010) and Wang et al. (2007), and references therein, for contributions in the frequentist and bayesian literature, respectively. As mentioned in the Introduction, though, their application 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 bases smooth enough for this purpose.

Wavelets have been used by 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, an high observation noise strongly worsen this issue, being necessary to use coarser level scaling coefficients.

Here we instead resort to a numerical method that allows to compute pointwise values of wavelets and their derivatives, even in the absence of a close analytical form for the wavelet basis. The method is based on a common approach for solving dilation equations (see, e.g., Strang, 1989). 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. First of all, recall that if the filter $\{h_k\}_k$ is of finite length N, then the support of the scaling function φ is $[0, N-1]$. This happens because of the dilation equation:

$$\varphi(l) = \sqrt{2} \sum_{k=0}^{N-1} h_k \varphi(2l - k). \tag{2}$$

In fact, supposing that the support of φ is $[a, b]$, we have

- $\varphi(2l)$ has support $[\frac{a}{2}, \frac{b}{2}]$

- $\varphi(2l - 1)$ has support $[\frac{a+1}{2}, \frac{b+1}{2}]$
  $\vdots$

- $\varphi(2l - (N-1))$ has support $[\frac{a+N-1}{2}, \frac{b+N-1}{2}]$.

The right hand side of equation (2) has therefore support $[\frac{a}{2}, \frac{b+N-1}{2}]$. However, this must coincide with the support $[a, b]$ of the left hand side. Thus,

$$\begin{cases} a = \frac{a}{2} \\ b = \frac{b+N-1}{2} \end{cases} \Rightarrow \begin{cases} a = 0 \\ b = N - 1. \end{cases}$$

We describe directly how to obtain point-wise values of wavelet derivatives. Deriving the dilation equation and evaluating it at the integers in the support of φ, we obtain

$$\begin{aligned} \varphi'(0) &= 2\sqrt{2}(h_0 \varphi'(0)) \\ \varphi'(1) &= 2\sqrt{2}(h_0 \varphi'(2) + h_1 \varphi'(1) + h_2 \varphi'(0)) \\ &\vdots \\ \varphi'(N-1) &= 2\sqrt{2}(h_{N-1} \varphi'(N-1)). \end{aligned}$$

This linear system can be rewritten as

$$
\begin{pmatrix} \varphi'(0) \\ \vdots \\ \varphi'(N-1) \end{pmatrix} = 2H \begin{pmatrix} \varphi'(0) \\ \vdots \\ \varphi'(N-1) \end{pmatrix}
$$

where H is the matrix formed by the filter coefficients. For example, for N = 6, the matrix H turns out to be

$$
H = \sqrt{2} \begin{pmatrix}
h_0 & 0 & 0 & 0 & 0 & 0 \\
h_2 & h_1 & h_0 & 0 & 0 & \\
h_4 & h_3 & h_2 & h_1 & h_0 & 0 \\
0 & h_5 & h_4 & h_3 & h_2 & h_1 \\
0 & 0 & 0 & h_5 & h_4 & h_3 \\
0 & 0 & 0 & 0 & 0 & h_5
\end{pmatrix}
$$

It becomes then clear that the vector $(\varphi'(0), \ldots, \varphi'(N-1))$ is the eigenvector associated to the eigenvalue 2 of the matrix H. It is thus possible to obtain the values of φ' on the integers $0, \ldots, N-1$, thanks to known methods for computing eigenvectors (see, e.g., Quarteroni et al., 2000). Since eigenvectors are defined up to a scaling coefficient, it is therefore necessary to impose a normalization. Being $[l\varphi(l)]_{-\infty}^{+\infty} = 0$ and $\int \varphi(l) = 1$, we have that $\int l\varphi'(l) = [l\varphi(l)]_{-\infty}^{+\infty} - \int \varphi(l) = -1$; hence, a suitable normalization for the eigenvector is $\sum_{k=0}^{N-1} k\varphi'(k) = -1$. Using the dilation equation a second time, we get

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

so that φ' can be evaluated at the middle points $\frac{1}{2}, \ldots, N - \frac{3}{2}$. Iterating this procedure, it is possible to compute the desired refinement of φ'. Once φ' is obtained, ψ' is computed deriving the wavelet dilation equation (1):

$$
\psi'(l) = 2\sqrt{2} \sum_k g_k \varphi'(2l-k).
$$

By the same procedure, taking the second and subsequent derivatives of the dilation equation, it is possible to obtain the second and subsequent derivatives of φ and ψ.

4 Wavelet smoothing of functional data

As we recalled in Section 2, 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

$$
w_i = f(l_i) + \varepsilon_i, \quad i = 1, \ldots, n, \ n = 2^J, \ J \in \mathbb{N}, \tag{3}
$$

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 wavelet domain, where the model becomes

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

$d_{j,k}$ being the empirical coefficients corresponding to the data, $d_{j,k}^0$ the true wavelet coefficients of f, and $\rho_{j,k} \sim N(0, \sigma_d^2)$ the wavelet transforms of the error. Section 4.1 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 $d_{j,k}^0$ 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

$$\widehat{f}(l) = \sum_{k=1}^{2^{j_0}} s_{j_0,k}\varphi_{j_0,k}(l) + \sum_{j=j_0}^{j=J-1} \sum_{k=1}^{2^j} \hat{d}_{j,k}\psi_{j,k}(l). \tag{4}$$

It should be noticed that since wavelet bases are by construction localized in both space and frequency, (4) very naturally provides a locally adaptive estimate of the function f.

The level j_0 in (4) 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*:

$$t = \hat{\sigma}_d\sqrt{2\log n}\,. \tag{5}$$

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 of these coefficients, i.e.,

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

Since wavelet bases are localized for definition both in space and frequency, estimation of the wavelet coefficients provides naturally a locally adaptive estimation of the function f.

4.1 Computation of the empirical coefficients and optimal translation for 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 these 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

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

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

$$\Downarrow$$

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

since $\int_{\mathbb{R}} \varphi_{J,i}(l + \tau) \, \mathrm{d}l = 1$. Hence, if the translation τ is such that

$$\int_{\mathbb{R}} l \varphi(l + \tau) \, \mathrm{d}l = 0, \tag{6}$$

then $f(l_i)$ would be a good approximation of $s_{J,i}$, so that using the data value w_i to approximate $s_{J,i}$ seems well justified in this case. If we thus impose the supplementary condition (6), we get

$$\int_{\mathbb{R}} l \varphi(l + \tau) \, \mathrm{d}l = 0 \Rightarrow \int_{\mathbb{R}} (y - \tau) \varphi(y) \mathrm{d}y = 0 \Rightarrow$$
$$\Rightarrow \int_{\mathbb{R}} y \varphi(y) \mathrm{d}y - \tau \int_{\mathbb{R}} \varphi(y) \mathrm{d}y = 0 \Rightarrow \tau = \int_{\mathbb{R}} y \varphi(y) \mathrm{d}y. \tag{7}$$

We therefore use the basis associated to the translation τ found above. Empirical results (not reported here for sake of brevity) confirm that this is an optimal

translation. 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 point-wise values of $\varphi'_{j_0,k}$ and $\psi'_{j,k}$ are computed as detailed in Section 3. Subsequent derivative estimates are obtained analogously.

5   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

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

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

$$\mathbf{w}_k = \mathbf{f}(l_k) + \boldsymbol{\varepsilon}_k \quad k = 1, \dots, n = 2^J \tag{8}$$

where $\boldsymbol{\varepsilon}_k$ is a multinormal error 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. A first idea could be to estimate each coordinate function $f_1, \dots, f_p$ independently, applying separately on each coordinate the procedure described in the previous sections. However, 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 $f_1, \dots, f_p$ 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.

5.1 Generalization of Universal Threshold in p dimensions

Starting from model (8) and using the orthogonality of the wavelet transform, we have that

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

with $\mathbf{d}_{j,k}, \mathbf{d}_{j,k}^0, \boldsymbol{\rho}_{j,k} \in \mathbb{R}^p$, where $\mathbf{d}_{j,k}$ are the vectors of the empirical wavelet coefficients corresponding to the data, $\mathbf{d}_{j,k}^0$ 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 $\mathbf{d}_{j,k}^0$ can be estimated as the null vector, we focus on $||\mathbf{d}_{j,k}||_2^2$ 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 5.1 *(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 5.1, we have in fact that if the number n of observations in model (3) is large enough, then the universal threshold (5) contains, with high probability, all the coefficients coming from noise.

In the p-dimensional case, we know that $||\frac{\boldsymbol{\rho}_{j,k}}{\sigma_d}||_2^2 \sim \chi^2(p)$. 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 5.1 *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

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

Proof.
$$P[\max_{1 \leq k \leq n} Y_k \leq u_n] = \{F(u_n)\}^n =$$

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

$\square$

Indeed, applying Theorem 5.1, we can prove the following result.

Proposition 5.2 *Let $\{Y_n\}_n$ be a sequence of i.i.d. $\chi^2(p)$ random variables and $A_n = \{\max_{i=1,\dots,n} Y_i \le c_p \log n\}$, where*

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

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

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

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

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

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

Then, using Theorem 5.1, we obtain

$$P[\max_{1 \le k \le n} Y_k \le c_p \log n] \to e^{-0} = 1.$$

$\square$

Applying Theorem 5.1 it is also possible to prove Proposition 5.1. It should be noticed that Proposition 5.2 when $p = 1$ (i.e., when Y_i is the square of a Gaussian random variable) gives the same threshold supported by Proposition 5.1. In the multidimensional case $p \ge 2$, Proposition 5.2 supports instead the following threshold on $||\mathbf{d}_{j,k}||_2^2$:

$$t_p = \hat{\sigma}_d^2 (3 \log n).$$

A simple estimator for multidimensional case is therefore the following hard-thresholding scheme:

$$\hat{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{0} & \text{if } ||\mathbf{d}_{j,k}||_2^2 \le t_p \\ \mathbf{d}_{j,k} & \text{if } ||\mathbf{d}_{j,k}||_2^2 > t_p. \end{cases}$$

Likewise hard thresholding in the 1D case, this estimator 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. 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 $\mathbf{d}_{j,k}^0$. Thus, we choose to estimate $\mathbf{d}_{j,k}^0$ keeping unchanged the direction of the empirical coefficient $\mathbf{d}_{j,k}$ and diminishing its norm by $\sqrt{t_p}$, i.e.,

$$\text{if} \quad ||\mathbf{d}_{j,k}||_2^2 > t_p \quad \text{then} \quad ||\hat{\mathbf{d}}_{j,k}||_2 = ||\mathbf{d}_{j,k}||_2 - \sqrt{t_p} \,.$$

Setting $||\hat{\mathbf{d}}_{j,k}||_2 = ||c\mathbf{d}_{j,k}||_2$, we get

$$c = 1 - \frac{\sqrt{t_p}}{||\mathbf{d}_{j,k}||_2} \,,$$

so that the soft-thresholding estimator will be

$$\hat{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{0} & \text{if } ||\mathbf{d}_{j,k}||_2^2 \leq t_p \\ (1 - \frac{\sqrt{t_p}}{||\mathbf{d}_{j,k}||_2})\mathbf{d}_{j,k} & \text{if } ||\mathbf{d}_{j,k}||_2^2 > t_p \end{cases}$$

i.e.,

$$\hat{\mathbf{d}}_{j,k} = \left(1 - \frac{\sqrt{t_p}}{||\mathbf{d}_{j,k}||_2}\right)_+ \mathbf{d}_{j,k} \,. \tag{9}$$

Geometrically, this soft-thresholding procedure works as follows. Consider a p-dimensional sphere with radius $\sqrt{t_p}$ 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 1 gives a visual representation of this procedure for $p = 3$.

6   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, 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 (see, e.g., Luo and Wahba, 1997; Miyata and Shen, 2003); this technique has been shown to give functionally very accurate estimates of
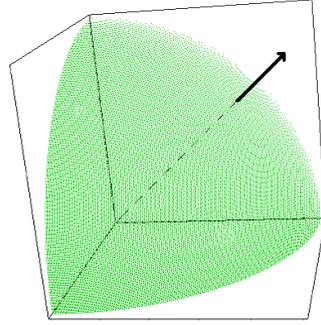
Figure 1: Visual representation in three dimensions of the soft-thresholding procedure: only the part of the vector $\mathbf{d}_{j,k}$ that lies outside the sphere with radius $\sqrt{t_p}$ is retained as significant.
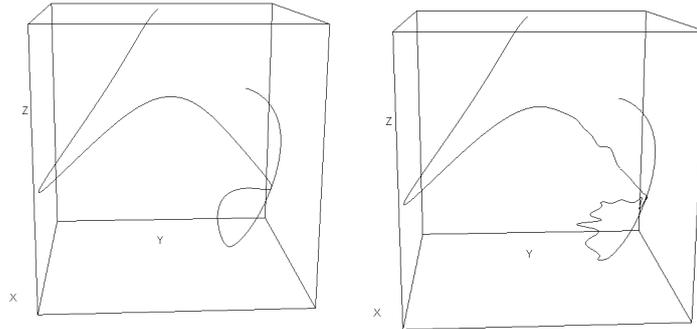
Figure 2: Left panel: curve $\mathbf{c}_1$, first dataset. Right panel: curve $\mathbf{c}_{51}$, second dataset; obtained from curve $\mathbf{c}_1$, left panel, by adding strongly localized features.

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, $x_i(l), y_i(l), z_i(l)$, for $l \in [-1, 2]$, with x_i, y_i, 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 x_i, 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,

$$x_i(l) = \log_{10}(x_i(l) + q) \quad y_i(l) = \log_{10}(y_i(l) + q) \quad z_i(l) = \log_{10}(l + z_i(l) + q),$$

where q is a fixed positive constant that makes 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 2, 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 $\sigma_e^2 \mathbb{I}_3$, where $\sigma_e = 2 \cdot 10^{-4}$, thus obtaining a noisy and discrete observation of the curve: $\{(x_{iu}, y_{iu}, z_{iu}) : u = 1, \ldots, n = 2^8\}$.
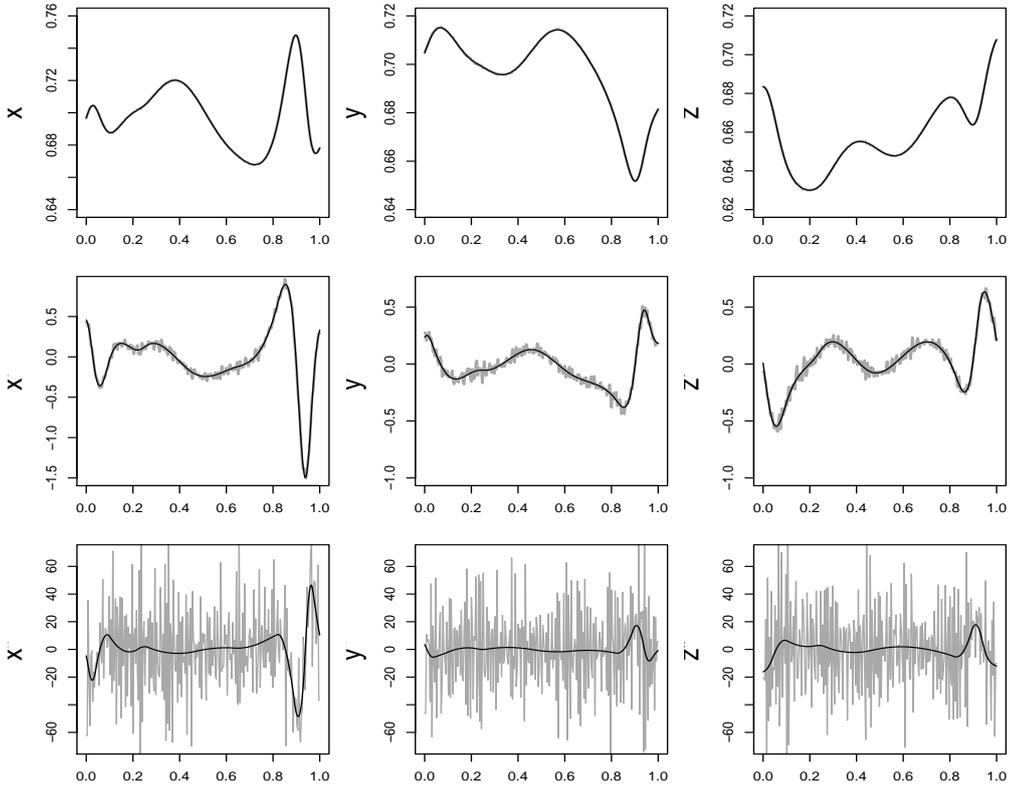
Figure 3: Space coordinates and derivatives of the curve $\mathbf{c}_1$, first dataset. Top: three space coordinates $\{x_1(l), y_1(l), z_1(l)\}$ (black), superimposed to raw data (grey). Center: first derivatives $\{x_1'(l), y_1'(l), z_1'(l)\}$ (black), superimposed to first central differences. Bottom: second derivatives $\{x_1''(l), y_1''(l), z_1''(l)\}$ (black), superimposed to second central differences.

13

Figure 3 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 $\{x_1'(l), y_1'(l), z_1'(l)\}$ (black), superimposed to first central differences of sampled data, and the bottom panel displays the second derivatives $\{x_1''(l), y_1''(l), z_1''(l)\}$ (black), superimposed to second central differences of sampled data.

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 $\mathbf{d}_{j,k} = (d_{j,k}^{[x]}, d_{j,k}^{[y]}, d_{j,k}^{[z]})$, 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 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

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

Figure 2, 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 normally distributed errors with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_e^2 \mathbb{I}_3$, where $\sigma_e = 2 \cdot 10^{-4}$, 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 $\mathbf{c}_{51}$.

$$\mathrm{RASE}_{DATA}(\hat{\mathbf{c}}_i) =$$

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

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:

$$\mathrm{RMSE}_{TRUE}(\hat{\mathbf{c}}_i) =$$

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

14
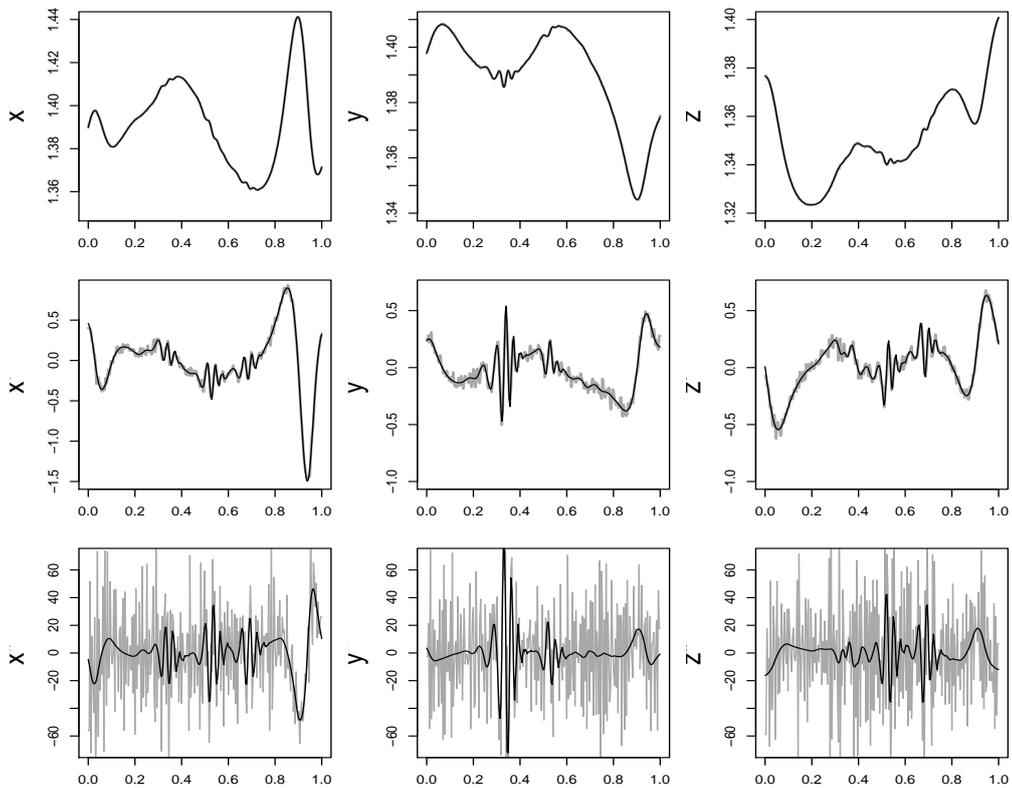
Figure 4: Space coordinates and derivatives of the curve $\mathbf{c}_{51}$, second dataset. Top: three space coordinates $\{x_{51}(l), y_{51}(l), z_{51}(l)\}$ (black), superimposed to raw data (grey). Center: first derivatives $\{x'_{51}(l), y'_{51}(l), z'_{51}(l)\}$ (black), superimposed to first central differences. Bottom: second derivatives $\{x''_{51}(l), y''_{51}(l), z''_{51}(l)\}$ (black), superimposed to second central differences.

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

$$\text{RMSEder1}_{TRUE}(\hat{\mathbf{c}}_i) =$$

$$= \sqrt{\frac{1}{n - 2m} \sum_{u=1+m}^{n-m} \left[\left(x'_i(l_u) - \hat{x}'_i(l_u)\right)^2 + \left(y'_i(l_u) - \hat{y}'_i(l_u)\right)^2 + \left(z'_i(l_u) - \hat{z}'_i(l_u)\right)^2 \right]};$$

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.

Figure 5 shows the boxplots of RASE_{DATA}, RMSE_{TRUE}, $\mathrm{RMSEder1}_{TRUE}$ and $\mathrm{RMSEder2}_{TRUE}$ 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.
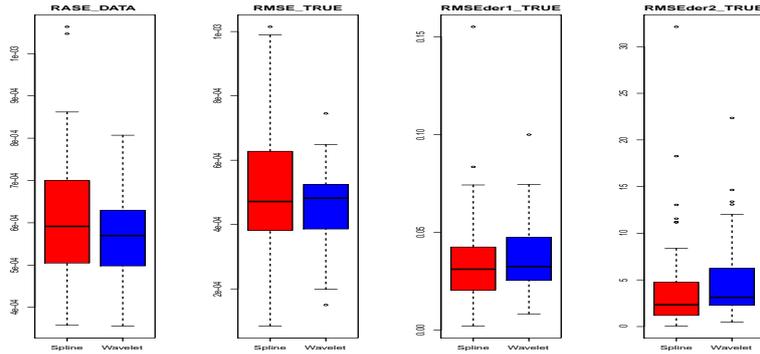
Figure 5: First Dataset. Boxplots of RASE_{DATA}, RMSE_{TRUE}, $\mathrm{RMSEder1}_{TRUE}$ and $\mathrm{RMSEder2}_{TRUE}$.
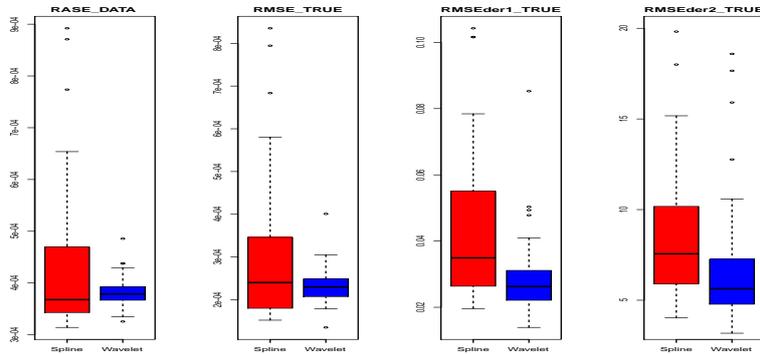
Figure 6: Second Dataset. Boxplots of RASE_{DATA}, RMSE_{TRUE}, $\mathrm{RMSEder1}_{TRUE}$ and $\mathrm{RMSEder2}_{TRUE}$.

Figure 6 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; 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.
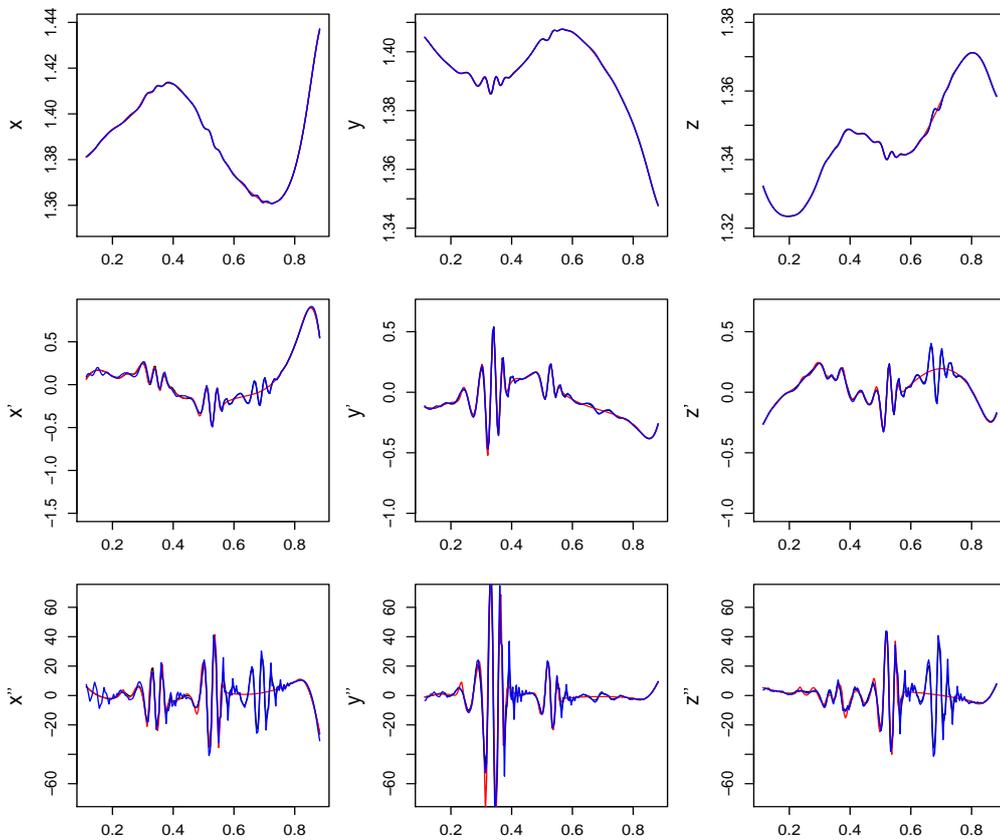
Figure 7: Estimates of the three space coordinates (top) and first and second derivatives (center and bottom) of $\mathbf{c}_{51}$, obtained by wavelet smoothing (blue) and free-knot regression splines (red), superimposed to true curve (black).

Figures 7 and 8 illustrate this issue. Figure 7 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

17

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 8 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 $z''_{51} - \hat{z}''_{51}$ for the 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 has shown 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.
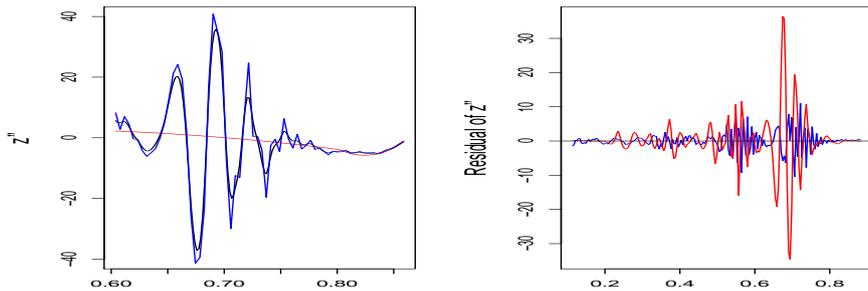
Figure 8: Left: Zoom of the estimates of the second derivative in the z direction of $\mathbf{c}_{51}$, in correspondence of one of the strongly localized feature; wavelet estimate (blue) and spline estimate (red), superimposed to true derivative (black). Right: Residuals $z''_{51} - \hat{z}''_{51}$ for the wavelet estimate (blue) and the spline estimate (red).

7 Application to ECG data

In this section we apply the proposed multidimensional wavelet fitting technique for the estimation of electrocardiogram (ECG) records. The data come from the 118 Dispatch Center, the medical operating emergency unit, operating in Milano, Italy. These records are collected as part of the PROMETEO (PROgetto Milano Ecg Teletrasmessi ExtraOspedaliero) Project (see Barbieri et al., 2010; Ieva et al., 2010), whose aim is to anticipate diagnostic time in heart attacks, in order to improve the prognosis of reperfusive treatments and reduce infarction complications.

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). 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 "12-leads" ECG. Among the 12 leads provided by the experimental device, eight leads are needed to capture all the information:

- 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.

Figure 9, left panel, shows the positions of electrodes and leads. Since the eight significant leads jointly describe the complex heart dynamics, when smoothing these data it is appropriate to use a technique which takes into account all the eight significant leads simultaneously. Moreover, this helps in detecting significant features, which reflect on more than one lead. Furthermore, wavelet basis are particularly suited to capture ECG shapes, because these are characterized by localized strong oscillations. Figure 9, right panel, gives a scheme of the typical structure of Lead I. Figure 10 shows one of the ECG records stored in the PROMETEO database. In particular, the figure display the raw data of the eight significant ECG leads for a patient affected by ST Elevation Myocardial Infarction; superimposed are the estimates of these eight-dimensional functional data, obtained by the proposed technique using Daubechies wavelet basis with 10 vanishing moments and the generalized soft-thresholding estimator (9).
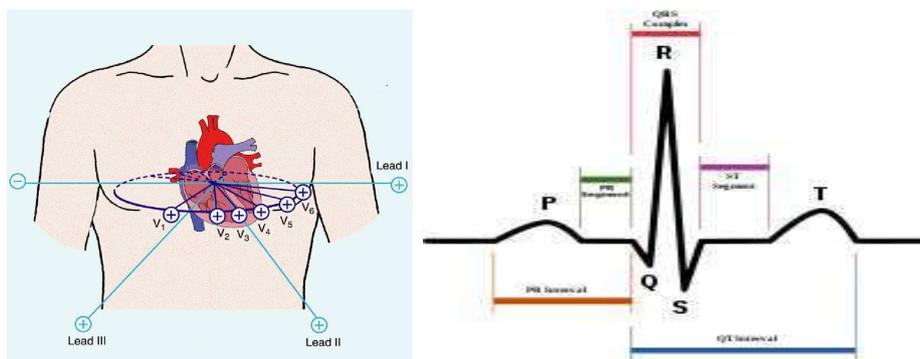
Figure 9: Left: Scheme of the directions along which the potential difference is measured for every lead. Right: Template of a physiological ECG record on Lead I.

Figure 11 shows the estimated first and second derivatives of Lead I for this patient. The obtained estimates of the eight-lead traces and of their derivatives, for the records in the PROMETEO database, are the starting point of extensive analyses that aim at identifying existing pathologies via ECG shapes, as well as exploring epidemiologic correlations among different cardiovascular diseases. A first promising result in this respect is for instance the identification of patients affected by Bundle Branch Blocks (Ieva et al., 2010). These analyses are not described here as they will be the object of a dedicated publication.
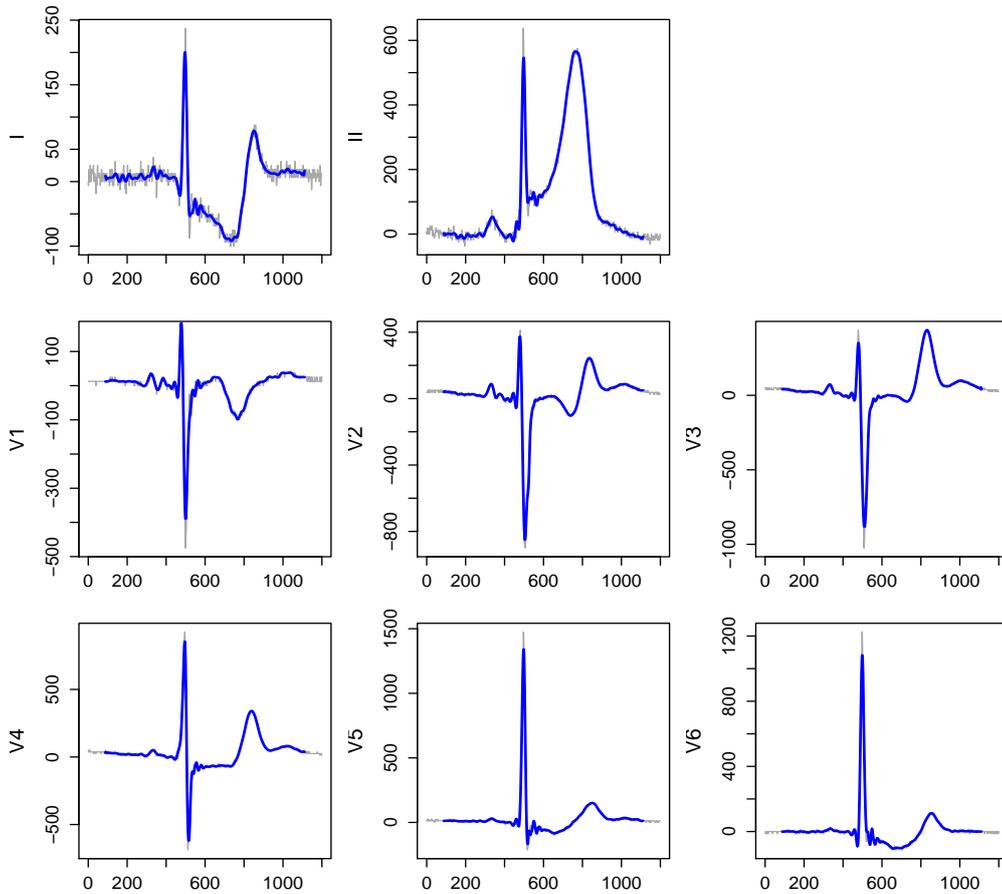
Figure 10: Eight significant leads in a 12-leads ECG for a patient affected by ST Elevation Myocardial Infarction; raw data (grey) and multidimensional wavelet estimate (blue).
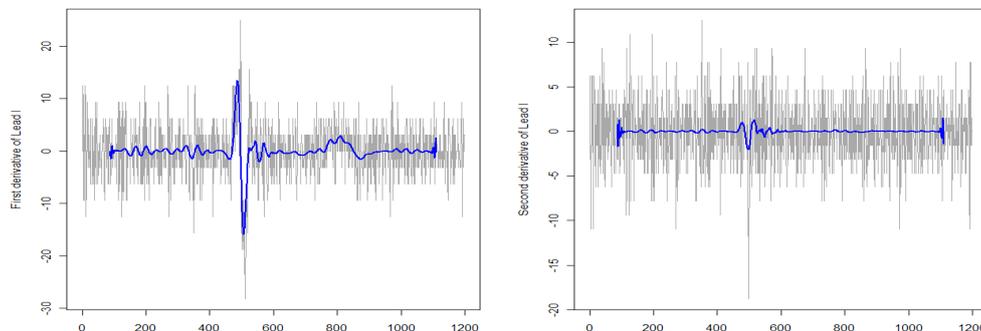
Figure 11: Left: Estimate of first derivative of Lead I (blue), superimposed to first central differences of raw data (grey). Right: Estimate of second derivative of Lead I (blue), superimposed to second central differences of raw data (grey).

8 Discussion

We have described a wavelet-based method for the accurate estimation of multi-dimensional curves and their derivatives. 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, a stimulating application for this research concerned the fitting of multi-lead ECG records.

Other contributions, such us Storlie et al. (2010), have recently appeared, on locally adaptive smoothing techniques able to accurately deal with functional data having spatial inhomogeneities and varying roughness, highlighting the growing interest for this topic.

References

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

Backman, G., Beckenstein, E. and Narici, L. (2000), Fourier and Wavelet Analysis. Springer-Verlag, New York.

Barbieri, P., Grieco, N., Ieva, F., Paganoni, A.M. and Secchi, P. (2010), "Exploitation, integration and statistical analysis of Public Health Database and STEMI archive in Lombardia Region", in Complex Data Modeling and Computationally Intensive Statistical Methods, Springer-Verlag, pp. 41–55.

Beylkin, G., Coifman, R., 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.

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.

Ieva, F., Paganoni, A.M., Pigoli, D., and Vitelli, V. (2010), "Statistics on ECGs: wavelet smoothing, registration and classification". Tech. Rep. MOX, Dipartimento di Matematica, Politecnico di Milano.

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.

Luo, Z. and Wahba, G. (1997), "Hybrid adaptive splines". Journal of the American Statistical Association, **92**, 107–116.

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

Miyata, S. and Shen, X. (2003), "Adaptive Free-Knot Splines", Journal of Computational and Graphical Statistics, **12**, 197-213.

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

Nason, G.P. (2010), wavethresh: Wavelets statistics and transforms. R package version 4.5. http://CRAN.R-project.org/package=wavethresh.

Quarteroni, A., Sacco, R. and Saleri, F. (2000), Numerical Mathematics, Springer, Berlin.

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.

Storlie, C.B., Bondell, H.D. and Reich, B.J. (2010), "A locally adaptive penalty for estimation of functions with varying roughness". Journal of Computational and Graphical Statistics, **19** 569–589.

Strang, G. (1989), "Wavelets and Dilation Equations: A Brief Introduction", SIAM Review, **31**, 614–627.

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.

MOX Technical Reports, last issues

09/2011 PIGOLI, D.; SANGALLI, L.
Wavelets in Functional Data Analysis: estimation of multidimensional curves and their derivatives

08/2011 IEVA, F.; PAGANONI, A.M.; SECCHI, P.
Mining Administrative Health Databases for epidemiological purposes: a case study on Acute Myocardial Infarctions diagnoses

07/2011 ARIOLI, G.; GAMBA, M.
Automatic computation of Chebyshev polynomials for the study of parameter dependence for hyperbolic systems

06/2011 SECCHI, PIERCESARE; STAMM, AYMERIC.; VANTINI, SIMONE
Large p Small n Data: Inference for the Mean

05/2011 ARIOLI, GIANNI; GAMBA, MONICA
An algorithm for the study of parameter dependence for hyperbolic systems

04/2011 IEVA, FRANCESCA ; PAGANONI, ANNA MARIA; PIGOLI, DAVIDE; VITELLI, VALERIA
Multivariate functional clustering for the analysis of ECG curves morphology

03/2011 GAREGNANI, GIULIA; ROSATTI, GIORGIO; BONAVENTURA, LUCA
Mathematical and Numerical Modelling of Fully Coupled Mobile Bed Free Surface Flows

02/2011 LASSILA, TONI; QUARTERONI, ALFIO; ROZZA, GIANLUIGI
A reduced basis model with parametric coupling for fluid-structure interaction problems

01/2011 DALLA ROSA, M.; SANGALLI, LAURA M.; VANTINI, SIMONE
Dimensional Reduction of Functional Data by means of Principal Differential Analysis

43/2010 PENNATI, GIANCARLO; DUBINI, GABRIELE; MIGLIAVACCA, FRANCESCO; CORSINI, CHIARA; FORMAGGIA, LUCA;, QUARTERONI,ALFIO; VENEZIANI, ALESSANDRO
Multiscale Modelling with Application to Paediatric Cardiac Surgery