European Commission

JRC TECHNICAL REPORTS

Study of Lipschitz regularity

Feature extraction on regular and irregular grids

Christophe Damerval

2012

Joint Research Centre

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*): 00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/.

Printed in Italy

Study of Lipschitz regularity
Feature extraction on regular and irregular grids

C. Damerval

Abstract

In this paper we study the pointwise Lipschitz regularity, covering several aspects: theoretical and practical, methods for its estimation on regular and irregular grids. The relevance of this value of regularity lies in its invariance properties to several transformations, and its fast computation thanks to wavelets. We study the influence of scale on wavelets transforms and show invariance properties this value of regularity. We also put forward an original technique for its estimation on regular grids. We also address the issue of irregular grids, based on the behavior of smoothing kernels with respect to scale. The obtained results emphasize the usefulness of such features for the applications, and motivate further work on this topic.

Keywords: Lipschitz regularity, wavelets, smoothing kernels, robust feature extraction, regular and irregular grids.

1 Introduction

The pointwise Lipschitz regularity is a feature which quantifies the regularity a function, associated for instance to a time series. This value $\alpha \in \mathbb{R}$ allows to measure the sharpness of edges and the smoothness of variations. Various works in signal processing studied its properties and applications [1, 2, 3, 4]. Here we are interested in the pointwise Lipschitz regularity since it makes a robust feature potentially useful in various applications, especially in EU policies to which the Joint Research Centre of the European Commission contributes (for instance monitoring of trade, antifraud and antimoney-laundering, financial time series, economic trends). Here the notion of robustness means that the quantities computed are slightly affected by transformations (linear or non-linear). The notion of feature is related to the extraction of quantities that are characteristic of an interpretable entity.

This paper is organized as follows. We first recall the notion of regularity and show some invariance properties. Since its classical computation is based on wavelets, we recall the definition of the Continuous Wavelet Transform and establish properties on its behavior with respect to scale. Then we describe the classical computation of regularity based on wavelets, and we present a new technique that distinguishes regular areas and irregular ones, applicable on regular grids. Besides we address the context of irregular grids, for which wavelets are no longer adapted. For this purpose we introduce an approach based on smoothing kernels, which use their behavior with respect to scale to compute a feature – using an algorithm similar to the computation of Lipschitz regularity with wavelets. The values of this new feature are then studied and compared with the theoretical one. Results show a certain relevance of this approach that extends the estimation of regularity features to more general settings (regular and irregular grids).
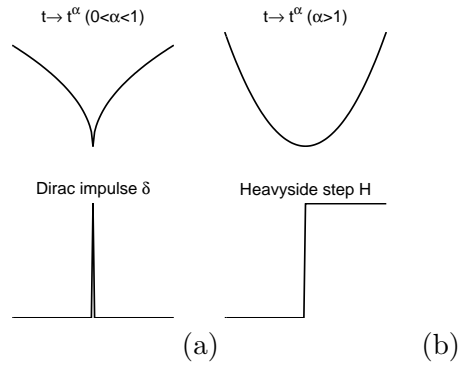
Figure 1: (a) Classical patterns in 1D, associated to different values of Lipschitz regularity ($\alpha = -1$ for the Dirac, $\alpha = 0$ for the step); (b) Classical wavelets functions: first derivative of Gaussian $\psi(t) = -te^{-t^2/2}$, Sombrero wavelet $\psi(t) = -te^{-t^2/2}$, Meyer wavelet.

2   Notion of regularity

2.1   Definition

Definition (Lipschitz regularity)
A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is α-Lipschitz at $t_0 \in \mathbb{R}$

- for $\alpha \in]0, 1[$, if there exists a neighborhood V of t_0 and $A > 0$ such that

$$\forall t \in V, |f(t) - f(t_0)| \leq A|t - t_0|^{\alpha} \tag{1}$$

- for $\alpha > 1$ (α non integer), denoting $n = \lfloor \alpha \rfloor$, if there exists a neighborhood V of t_0, $A > 0$ and a polynomial $P_n(t)$ of order n ($n \leq \alpha < n+1$, P_n depending on t_0) such that

$$\forall t \in V, |f(t) - P_n(t)| \leq A|t - t_0|^{\alpha} \tag{2}$$

- for $\alpha = n \in \mathbb{N}^*$, if f is C^n at t_0.

More generally, the Lipschitz regularity can be defined for $\alpha \leq 0$ using the theory of distributions, see [5] for details.

Definition (Regularity α)
Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $t_0 \in \mathbb{R}$. The regularity α of f at t_0 is defined as

$$\alpha = \inf\{\alpha_0 \in \mathbb{R},\ f\ \alpha_0\text{-Lipschitz at } t_0\} \tag{3}$$

The point of studying the pointwise Lipschitz regularity is the detection and identification of singularities [6], such as the patterns represented on Figure 1(a). As an example, the real function $t \mapsto \sqrt{|t|}$ is C^∞ at any $t \neq 0$ ($t \in \mathbb{R}$), and presents a singularity at $t = 0$, with an associated regularity $\alpha = 1/2$.

2.2 Invariance properties

Here we present invariance properties of the regularity α under different transformations. Let us first present properties bearing on α-Lipschitz functions, and then infer invariance properties of the regularity α (proofs are detailed in appendix).

Proposition 1. *Let $t_0 \in \mathbb{R}$ and $f, g : \mathbb{R} \to \mathbb{R}$ related by $g(t) = c(t)f(t) + d(t)$, where $c, d : \mathbb{R} \to \mathbb{R}$. Assuming c, d are C^∞ and c is locally bounded,*

$$\forall \alpha > 0, \ f \ \alpha\text{-Lipschitz at } t_0 \Rightarrow g \ \alpha\text{-Lipschitz at } t_0 \tag{4}$$

Proposition 2. *Let $t_0 \in \mathbb{R}$ and $f, g : \mathbb{R} \to \mathbb{R}$ related by $g(t) = f(u(t))$ where $u : \mathbb{R} \to \mathbb{R}$. Assuming u is 1-Lipschitz,*

$$\forall \alpha > 0, f \ \alpha\text{-Lipschitz at } u(t_0) \Rightarrow g \ \alpha\text{-Lipschitz at } t_0 \tag{5}$$

Note that such transformations do not correspond to a smoothing or a sharpening, which would alter the Lipschitz regularity.

Proposition 3. *(Invariance to multiplication and addition of C^∞ functions)*
Let $f, g : \mathbb{R} \to \mathbb{R}$, related by: $\forall t \in \mathbb{R}$

$$g(t) = c(t) \cdot f(t) + d(t), \ \text{with } c, d : \mathbb{R} \to \mathbb{R} \ \text{are } C^\infty \ \text{and } c \neq 0 \tag{6}$$

Then, at any $t_0 \in \mathbb{R}$, f and g have the same regularity α (for $\alpha > 0$).

Proposition 4. *(Invariance to dilatation and translation)*
Let $f, g : \mathbb{R}^2 \to \mathbb{R}$, related by: $\forall x \in \mathbb{R}^2$

$$g(ct + d) = f(t), \ \text{with } c, d \in \mathbb{R}, c \neq 0 \tag{7}$$

Then, at any $t_0 \in \mathbb{R}$, f and g have the same regularity α (for $\alpha > 0$). respectively at t_0 and $z_0 = ct_0 + d$

$$\alpha(f, t_0) = \alpha(g, ct_0 + d) \tag{8}$$

Conclusions.

We established invariance properties of the regularity α: to constant translation and dilatation, and also to the multiplication by a C^∞ function.

3 Properties of wavelet transforms

To estimate the regularity α, the wavelet framework shown its efficiency [6, 7]: these multiscale methods allow to detect singularities and to estimate α robustly. We first present results on the behavior of the wavelet transform with respect to scale, when the function has a simple analytical expression. Then we recall the known link between Lipschitz regularity and wavelets [4, 8], which provides the classical method to compute the regularity α. We represent on Figure 1(b) examples of wavelet functions, and we now recall definitions coming from wavelet theory [8].

Definition (Continuous Wavelet Transform)
Given a wavelet function $\psi : \mathbb{R} \to \mathbb{R}$, i.e., verifying the admissibility condition

$$\int_{\mathbb{R}} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} dw < +\infty \tag{9}$$

the continuous wavelet transform (CWT) with normalization L^p (integer $p > 0$) associated to a function $f : \mathbb{R} \to \mathbb{R}$ is defined as $Wf : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}$, $\forall u \in \mathbb{R}, \forall s > 0$

$$Wf(u,s) = \frac{1}{s^{1/p}} \int_{\mathbb{R}} f(t)\psi\left(\frac{t-u}{s}\right) dt \tag{10}$$

Remark: denoting $\psi_s(u) = \frac{1}{s^{1/p}}\psi\left(\frac{u}{s}\right)$ and $\overline{\psi_s}(u) = \frac{1}{s^{1/p}}\psi\left(\frac{-u}{s}\right)$, the CWT can be expressed as the convolution product

$$Wf(u,s) = (f * \overline{\psi}_s)(u) \tag{11}$$

We study the asymptotic behavior of the wavelet coefficients when the scale s tends to infinity. In particular we obtain equivalents of Wf (when $s \to +\infty$) when we know an analytical expression of the f.

3.1 Behavior of the CWT with respect to scale

Here we study the CWT, focusing on its behavior with respect to scale. We show novel properties whose proofs are detailed in annex.

Proposition 5. *(Boundedness of the CWT) Let $q > 0$ defined by $\frac{1}{p} + \frac{1}{q} = 1$. Assuming $f \in L^q$ and $\psi \in L^p$, there exist $C > 0$ so that*

$$\forall u \in \mathbb{R}, \forall s > 0, \quad Wf(u, s) \leq C \tag{12}$$

Proposition 6. *(Behavior of the CWT when the scale tends to infinity) Assuming $\psi \in L^1 \cap L^2 \cap L^\infty$,*

$$||Wf(., s)||_\infty \longrightarrow 0 \text{ when } s \to +\infty \tag{13}$$

in all cases when $f \in L^1$, for $p < 2$ if $f \in L^2$ and for $p < 1$ if $f \in L^\infty$.

Proposition 7. *(Behavior of the CWT when the scale tends to zero) Assuming $\psi \in L^1$, the property*

$$|Wf(u, s)| \to 0 \text{ when } s \to 0 \tag{14}$$

holds provided one of the following conditions:
***General case:** when $f \in L^\infty$, provided $p > 1$*
***Particular case:** when f is α-Lipschitz (with $\alpha > 0$), provided $p > \frac{1}{\alpha+1}$*

Proposition 8. *(Asymptotic behavior of the CWT) We suppose that $\psi(0) \neq 0$ (which is the case for many wavelets, including derivatives of Gaussian). At any fixed $u \in \mathbb{R}$, when $s \to +\infty$*
(a) for the Dirac δ_0

$$|W\delta_0(u, s)| \sim \frac{\psi(0)}{s^{1/p}} \quad (s \to +\infty) \tag{15}$$

(b) for the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) = \begin{cases} 1 \text{ for } t \in [0, 1] \\ 0 \text{ elsewhere} \end{cases} \tag{16}$$

$$|Wf(u, s)| \sim \frac{\psi(0)}{s^{1/p}} \quad (s \to +\infty) \tag{17}$$

(c) for the function $f_\alpha : \mathbb{R} \to \mathbb{R}$, with $\alpha > 0$, defined by

$$f_\alpha(t) = \begin{cases} t^\alpha(1 - t^\alpha) \text{ for } t \in [0, 1] \\ 0 \qquad\qquad \text{ elsewhere} \end{cases} \tag{18}$$

$$Wf_\alpha(u, s) \sim \frac{\alpha}{(\alpha + 1)(2\alpha + 1)} \frac{\psi(0)}{s^{1/p}} \quad (s \to +\infty) \tag{19}$$

Remark: the function f_α (with $\alpha > 0$) is uniformly α-Lipschitz with compact support.

3.2 Link between Lipschitz regularity and wavelets

Here we present known results on wavelets, which makes up a basis for the the classical method of regularity estimation – see [1, 9] for details.

Proposition 9. *(Jaffard) Let $f : \mathbb{R} \to \mathbb{R}$, $f \in L^2$. We consider its continuous wavelet transform Wf, defined as: $\forall s > 0, \forall u \in \mathbb{R}$*

$$Wf(u,s) = \frac{1}{s} \int_{\mathbb{R}} f(t)\psi\left(\frac{t-u}{s}\right) dt \tag{20}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a wavelet function with n vanishing moments. If f is α-Lipschitz (with $\alpha \leq n$) at t_0 then there exists $A > 0$ such that: $\forall u \in \mathbb{R}, \forall s > 0$,

$$|Wf(u,s)| \leq As^\alpha \left(1 + \left|\frac{u-t_0}{s}\right|^\alpha\right) \tag{21}$$

Conversely, if $\alpha < n$ is not an integer and there exist $A > 0$ and $\alpha' < \alpha$ such that: $\forall u \in \mathbb{R}, \forall s > 0$,

$$|Wf(u,s)| \leq As^\alpha \left(1 + \left|\frac{u-t_0}{s}\right|^{\alpha'}\right) \tag{22}$$

then f is α-Lipschitz at x_0.

3.3 Classical estimation of the regularity α

Let us recall a classical method providing a fast computation of the regularity α. Starting from eq.(21) we get $|Wf(u,s)| \leq As^\alpha$ and then $\log|Wf(u,s)| \leq \log A + \alpha \log s$. Now, this latter inequality becomes a quasi-equality at fine scales [6]. So the regularity α can be estimated by performing a regression based on the formula

$$\log|Wf(u,s)| = \alpha \log s + C \tag{23}$$

More precisely α and C are computed by a regression of $\log|Wf(u,s)|$ over $\log s$ at fine scales (generally three scales are sufficient for an accurate estimation). When dealing with numerical data, a basic approach consists on applying the formula (23), for instance on scales $s = 1, 2, 3$. This gives one value of regularity at each data location. A more precise approach concerning regularity consists on computing values of regularity at pointwise singularities. These can be defined as modulus maxima (MM), i.e., locations where the modulus of the continuous wavelet transform is locally maximum.
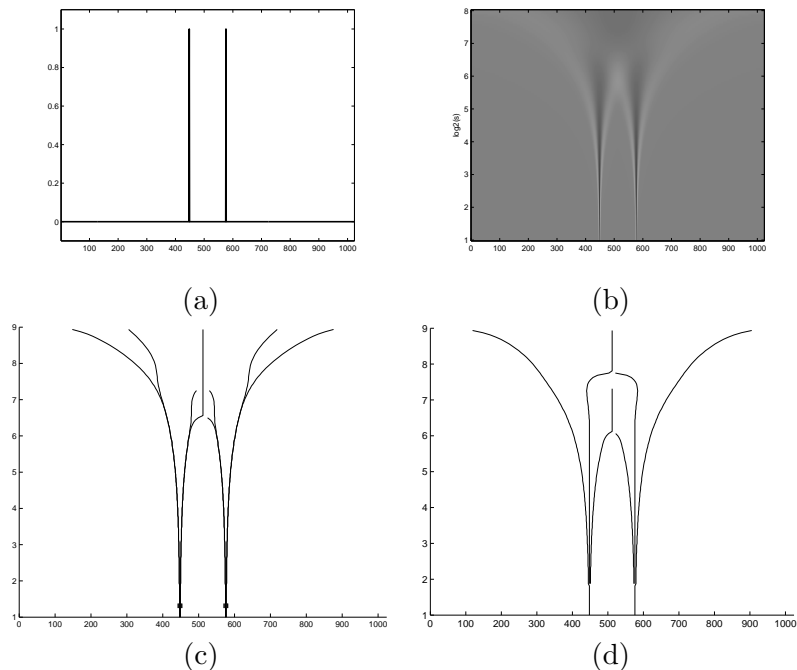
7

Figure 2: (a) Data: sum of two Dirac impulses; (b) Continuous Wavelet Transform of these data; (c) Structure of wavelet maxima lines, using the first derivative of Gaussian as wavelet; (d) *idem*, using the second derivative of Gaussian as wavelet.

Using an appropriate wavelet such as a n-th derivative of Gaussian, these MM are connected in the scale-space domain and make up wavelet maxima lines. The computation of the regularity α is done by using the formula (23), along these wavelet maxima lines. From an algorithmic point of view, denoting N the size of the data, this can be performed in $O(N)$ operations thanks to fast wavelet algorithms.

The relevance the wavelet maxima lines comes from the fact they allow to track singularities through scales [10, 11]. Let us see illustrations of the structure of these wavelet maxima lines, which depend on the signal and also on the wavelet. For illustration purposes, we represent on Figure 2 their structure in the case of a simple pattern: the sum of two Dirac impulses, see Fig.2(a). After the computation of the Continuous Wavelet Transform (CWT) (see Fig.2(b)), it is possible to construct numerically wavelet maxima lines. We represent them using two different wavelets: using the first

derivative of Gaussian, on Fig.2(c); using the second derivative of Gaussian, on Fig.2(d). Let us emphasize an essential principle: if a feature bears theoretical invariance properties, then it should present a certain robustness in practice. We present here a methodology that allows to quantify the robustness of a feature to various transformations.

4 New estimation of the regularity α

We present here a novel approach for the computation of the regularity α. We recall that previous works proposed to focus on singularities, the regularity α being estimated on these singularities. and also to estimate the regularity α at each location (treating data as a whole set). Taking advantage of these two techniques, we propose here an original approach which agrees with the idea of regularity areas between singularities. Assuming ψ is compactly supported on $[-C, C]$ ($C > 0$), let us split the scale-space $(u, s) \in \mathbb{R} \times \mathbb{R}_+^*$ into two sets. On the one hand, the cones of influence of the detected singularities

$$\mathcal{C} = \left\{ (u, s) \in \mathbb{R} \times \mathbb{R}_+^*, |u - u_0| \leq Cs \right\} \tag{24}$$

where u_0 is the location of a singularity, corresponding to MM at the finest scale. On the other hand, the remaining area of the scale-space, outside the cones of influence

$$\mathcal{O} = \mathbb{R} \times \mathbb{R}_+^* \setminus \mathcal{C} = \left\{ (u, s) \in \mathbb{R} \times \mathbb{R}_+^*, u_0 + Cs \leq u \leq u_1 + Cs \right\} \tag{25}$$

where u_0 and u_1 are locations of two successive singularities. Now, let us describe how to estimate the regularity α within these two sets. As mentioned in the preceding section, using three fine scales, we perform a linear regression based on the formula

$$\log W f(u, s) = \alpha \log s + K \tag{26}$$

Estimation in $\mathcal{C}$: so as to estimate the regularity α at one pointwise singularity located at u_0, we propose a technique based on wavelet coefficients. We define the set

$$\forall s > 0, \quad C(s) = \{u \in \mathbb{R}, u_0 - Cs < u < u_0 + Cs\} \tag{27}$$

and the quantity

$$\forall s > 0, \quad H(s) = \sup_{u \in C(s)} |Wf(u, s)| \tag{28}$$

so that at each scale s, we take the highest wavelet coefficients (in absolute value) within the cone of influence of u_0. Then, we perform a regression on fine scales, using the formula

$$\log H(s) = \alpha \log s + K \tag{29}$$

This gives an estimation of the pointwise regularity α at the location u_0.

Estimation in $\mathcal{O}$: so as to estimate the regularity α between two successive singularities, we propose a technique based on wavelet coefficients located outside the cones of influence of these singularities. Given two successive singularities located at u_0 and u_1, we define the set

$$\forall s > 0, \quad O(s) = \{u \in \mathbb{R}, u_0 + Cs < u < u_1 - Cs\} \tag{30}$$

and define the quantity

$$\forall s > 0, \quad L(s) = \inf_{u \in O(s)} |Wf(u, s)| \tag{31}$$

so that we consider the lowest wavelet coefficients within $D(s)$ (at one fixed scale s). Then, we perform a regression on fine scales using on the formula

$$\log L(s) = \alpha \log s + K \tag{32}$$

This gives an estimation of the regularity α over the interval $]u_0, u_1[$.

Note: the regularity α is computed through two different manners: pointwise estimations at singularities (estimation in $\mathcal{C}$) and local estimations between successive singularities (estimation in $\mathcal{O}$). This approach assumes that the data is made of discontinuities separated by smooth variations, as opposed to multifractal structures presenting singularities at all scales. As a consequence this approach can turn as relevant for data known to be quite stable (economic indicator, agricultural), less for very chaotic data (seismic activity, financial time series).
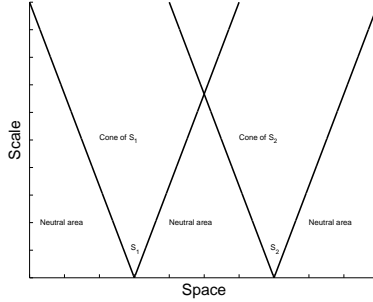
Figure 3: Given two singularities located at S_1 and S_2, the scale-space can be divided into two domains: areas corresponding to cones of influence of these singularities, and the remaining neutral areas.

Algorithm. We consider data made of real values $(x_i)_{i=1..N}$ given on integer locations $t_i = 1..N$. The wavelet coefficients can be efficiently computed through a spectral method, whose computational cost is $O(N \log N)$ – this assumes the Fourier transform of the wavelet is known analytically. Compared to the well-known algorithm Fast Wavelet Transform put forward by S. Mallat and implemented by filterbanks, this allows a high flexibility since any scale $s > 0$ can be considered (thus scales are not necessarily dyadic). For the determination of the singularities, we consider the modulus maxima at the finest scale, using the condition

$$|Wf(u, s)| > |Wf(u + h, s)| \quad \forall h \in \{-H, , ..., H\} \text{ with } H \in \mathbb{N}^* \tag{33}$$

so that for each location u, we compare the modulus $|Wf(u, s)|$ at u to its value at the nearest neighbors in space (using a window of size $2H + 1$). This allows to partition the data into two sets: $\mathcal{C}$ associated to pointwise singularities and $\mathcal{O}$ associated to a certain local regularity.

Concerning the practical estimation of α in $\mathcal{C}$, we select at each scale the highest coefficient within the cone of influence. Concerning the practical estimation in $\mathcal{O}$, we consider at each scale the mean of coefficient in the neutral area between two cones of influence. This allows to estimate the regularity α by the aforementioned regression at fine scales.

11

5 Generalization to non-regular grids

Here we generalize the preceding tools known on regular grids. We present here an original approach for regularity estimation on non-regular grids. In the following we consider data observations $(t_i, x_i)_{1 \leq i \leq N}$ with $t_i \in [-1, 1]$ and $x_i \in \mathbb{R}$.

5.1 Decompositions using smoothing kernels

Here we define a decomposition based on smoothing kernels, which can be integrated into the *lifting scheme* framework (also known as second generation wavelets). This multiscale decomposition presents some flexibility, since the kernel and the scale parameters can be chosen. Let us first recall the notion of smoothing kernels, along with their classical application in non-parametric statistics [12] . A smoothing kernel is a function $K : \mathbb{R} \to \mathbb{R}_+$, with bounded support or fast decaying, $0 < \int_{\mathbb{R}} K < +\infty$. The kernel estimator associated to the data $(t_i, x_i)_{1 \leq i \leq N}$ is defined as

$$\widehat{f_h}(t) = \frac{\sum_{i=1}^{N} K\left(\frac{t - t_i}{h}\right) x_i}{\sum_{i=1}^{N} K\left(\frac{t - t_i}{h}\right)} \tag{34}$$

where $h > 0$ is a bandwidth parameter. Typically, a cross-validation method allows to compute an appropriate value of h. Assuming that data are sampled from a probability density f, this gives a robust estimation of f by a smooth function $\widehat{f_h}$, see Fig.4(a). Here we use smoothing kernels differently, using a multiscale version, see Fig.4(b). Given different scales $s_j > 0$ ($j = 1..J$), we define at each level j ($j = 1..J$) the function $P_j : \mathbb{R} \to \mathbb{R}$ as

$$P_j(t) = \frac{\sum_{i=1}^{N} K\left(\frac{t - t_i}{s_j}\right) x_i}{\sum_{i=1}^{N} K\left(\frac{t - t_i}{s_j}\right)} \tag{35}$$
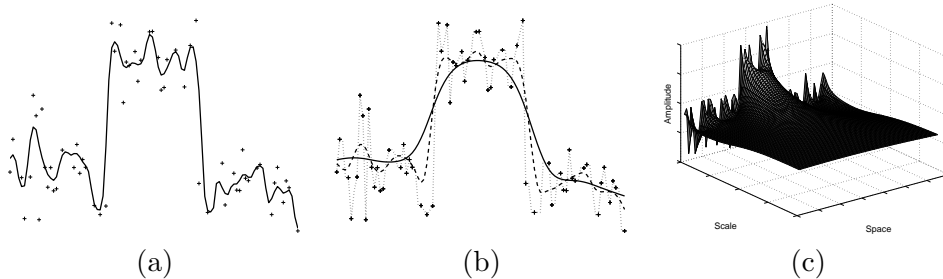
12

Figure 4: Smoothing data using a Gaussian kernel. (a) Estimator $\widehat{f}_{h^*}$ with h^* computed by cross-validation; (b) Multiscale version $P_j(t)$ at different scales s_j (c) 3D view of $(t_i, s_j, P_j(t))$.

The scales $(s_j)_{j=1..J}$ are similar to the bandwidth h seen in eq.(34). Large bandwidths imply coarse scale while small bandwidths imply fine scales; however, in contrast with the kernel estimation method, where cross-validation leads to an intermediate value of h, we use here small scales so as to analyze the local regularity. Now, starting from eq.(35), we define approximation and detail coefficients for each level $j = 1..J$ and each $k = 1..N$

$$a_{jk} = P_j(t_k) \qquad (36)$$
$$d_{jk} = P_{j+1}(t_k) - P_j(t_k) \qquad (37)$$

Classically these details d_j. correspond to differences between two successive approximations a_{j+1}. and a_j.. Note there is no upsampling here, since coefficients are known at each data location (these being indexed by k). This contrasts with multiresolution analysis and presents some similarity with the continuous wavelet transform: the coefficients d_{jk} play the role of the wavelet coefficients $Wf(u, s)$. Eventually the main point of this transformation is to define detail coefficients on any type of grid (regular or irregular) with a view to use them for regularity estimation.
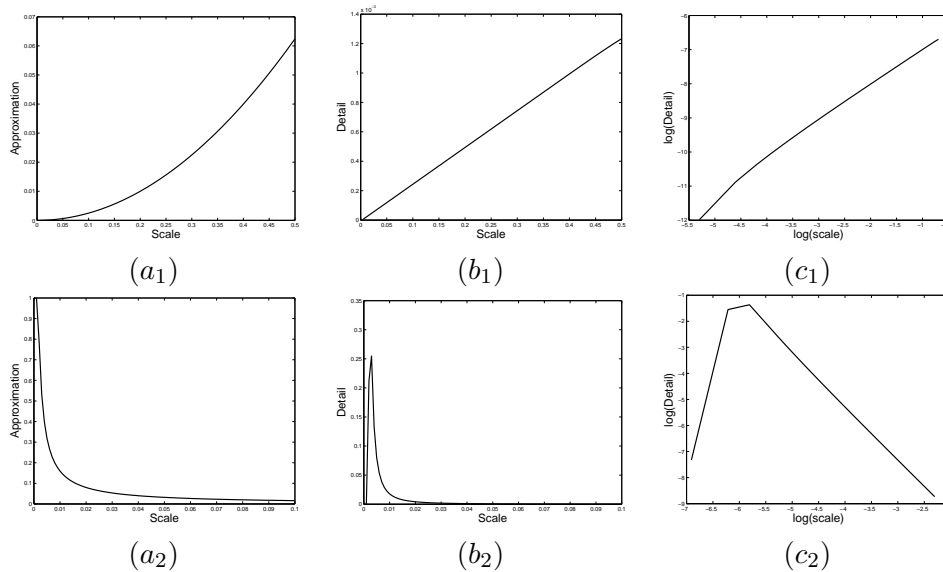
13

Figure 5: Evolution with respect to scale of approximation and detail coefficients: (a_1, a_2) Approximation coef. (b_1, b_2) Detail coef. (c_1, c_2) Detail coef. in logarithm representation; for the function $t \mapsto t^2$ (a_1, b_1, c_1) and for a Dirac impulse in (a_2, b_2, c_2).

5.2 Evolution of coefficients with respect to scale

Let us study the evolution with respect to scale of the decomposition coefficients, in the context of regular grids. With a view to show the difference between regular and irregular patterns, we consider the smooth function $f : t \mapsto t^2$ and the Dirac impulse δ. Starting from sampled data (100 samples in our experiments), we compute approximation and detail coefficients by using the proposed decomposition (using a fine range of scales to analyze their behavior). Now, we represent the evolution of approximation and detail coefficients with respect to scale on Figure 5: on top for the smooth function $t \mapsto t^2$ (at $t = 0$), at bottom for the Dirac impulse. Concerning approximation coefficients, these are increasing steadily for the smooth function f (see Fig.5(a_1)) whereas they are decreasing sharply for the Dirac impulse (see Figure 5(a_2)). Concerning detail coefficients, these are increasing steadily for the smooth function f (see Fig.5(b_1)) while they behave in a more complex manner for the Dirac impulse (see Fig.5(b_2)); more precisely, as the scale increases, the detail coefficients increase sharply until a scale

14

$s_0 > 0$, and decrease sharply after s_0. This appears clearly in log representation, see Fig.5(c_2). Let us give an interpretation of this behavior: since the scales can be freely chosen in the decomposition, the use of very fine scales leads to frequencies under the sampling rate, so that only one sample (ti, x_i) will be taken into account in the formula (35). This is valid until the above-mentioned scale s_0 is reached, which will be very small for a high sampling rate. Then, for fine and intermediate scales, more samples will be taken into account in the formula (35), which will result in a different behavior: a decay related to the irregularity of the Dirac impulse. This is very similar to the behavior of wavelet coefficients seen in section 3. For properties of the kernel estimator with respect to scale, see appendix B.

5.3 Comparison with wavelets.

Considering the two former patterns $f : t \mapsto t^2$ and δ, let us write their wavelet transforms (using L^1 normalization)

$$Wf(u,s) = \frac{1}{s} \int_{\mathbb{R}} t^2 \psi\left(\frac{t-u}{s}\right) dt = \int_{\mathbb{R}} (sv+u)^2 \psi(v) dv \qquad (38)$$

$$Wf(u,s) = s^2 \int_{\mathbb{R}} v^2 \psi(v) dv + 2s \int_{\mathbb{R}} v\psi(v) dv \qquad (39)$$

$$W\delta(u,s) = \frac{1}{s}\psi(-u/s) \qquad (40)$$

In particular, at $t = 0$ we have

$$Wf(0,s) = s^2 \int_R v^2 \psi(v) dv + 2s \int_{\mathbb{R}} v\psi(v) dv \qquad (41)$$

$$W\delta(0,s) = s^{-1}\psi(0) \qquad (42)$$

These behaviors are similar to the ones of detail coefficients of the original decomposition based on smoothing kernels: an increase of the response with scale for a regular pattern (regularity $\alpha > 0$), and decrease of the response with scale for an irregular pattern (regularity $\alpha < 0$). As we saw earlier, using a wavelet transform appropriately normalized, the evolution of the wavelet coefficients with scales is related to the regularity. This decomposition being similar to wavelets, we observe similar properties. However, it is important to note that the more general context of irregular grids leads to a more complex behavior with respect to scale.

5.4 Estimation of regularity features

We explain here how it is possible to compute features in the general context of data given on a non-necessarily regular grid. As seen in section 3, the wavelet framework provides an estimation of pointwise regularity, thanks to the link between the regularity and the decay of wavelet coefficients at fine scales. Here we use the estimation technique known in the context of wavelets, so as to obtain features at every location. This original approach can be applied to any distribution of points. We perform the following algorithm:

1. Input: data observations $(t_i, x_i)_{i=1..N}$

2. Perform the multiscale decomposition previously defined, using fine scales $(s_j)_{j=1..J}$. We recommend linear scales: $s_j = j \cdot P/N$ (N: size of the data, $P \in \mathbb{N}^*$: a parameter ensuring that on average, we take into account P points among the $(t_i)_{i=1..N}$ at the finest scale).

3. Given the detail coefficients $\{d_{jk}, \; 1 \leq j \leq J, \; 1 \leq k \leq N\}$ apply a linear regression based on the formula

$$\log |d_{jk}| = \beta \log s_j + C \tag{43}$$

   at fine scales $(s_j)_{j=1..J}$, for each index k.

4. Output: regularity features $(\beta_i)_{i=1..N}$, $\beta_i \in \mathbb{R}$ being related to the local regularity around x_i.

Note: it is appropriate to choose fine scales, and necessary to avoid too small values (see the discussion on the behavior at very fine scales in section 5.2).

5.5 Analysis of computed values.

Now, we investigate how the computed values β_i are related to the regularity α. For this purpose we consider simple patterns where the regularity is known *a priori*, see Fig.6(a): irregular patterns such as a Dirac impulse δ (regularity $\alpha = -1$) and a Heaviside step H (regularity $\alpha = 0$); regular patterns such as the real function $t \mapsto |t|^\alpha$ with $\alpha > 0$ (focusing at $t = 0$), for $\alpha = 1/2$, $\alpha = 1$ and $\alpha = 2$. Considering these five patterns, we apply the preceding algorithm to compute one value β (corresponding to the location $t = 0$). Then we represent on Figure 6(b-f) the evolution of the computed value β with the sample size N. Generally the computed value decreases as the sample size increases. Concerning the irregular patterns (Dirac impulse and Heaviside step), we note a significant influence of the sampling, even for high values. Concerning the more regular patterns ($t \mapsto |t|^\alpha$ with $\alpha > 0$), we note a convergence to a limit value from $n = 200$. Having noted that a certain density of points is desirable for the meaningfulness of the estimated values, let us address the link between the computed value β with the theoretical value α, using fixed values of sampling. In this regard let us first choose $n = 1000$ and consider the patterns $t \mapsto t^\alpha$ for numerous values of $\alpha > 0$. We then represent on Figure 5.5(a) the evolution of the computed value β with respect to the theoretical regularity α. This smooth curve shows there are strong link between these two quantities. Considering now several values of n, we can compute values β_i^n associated with values $\alpha_i > 0$ and several values of $n \in \mathbb{N}$. With a view to explain statistically these values β_i^n with α_i and independent of n, let us consider the regression model

$$\beta = a(\sqrt{\alpha} - 1) + b \text{ with } a, b \in \mathbb{R} \tag{44}$$

Results are reported in Fig. 5.5(b). The high value of the determination coefficient in all cases emphasizes the link between β and α. Heuristically, this can be summarized as

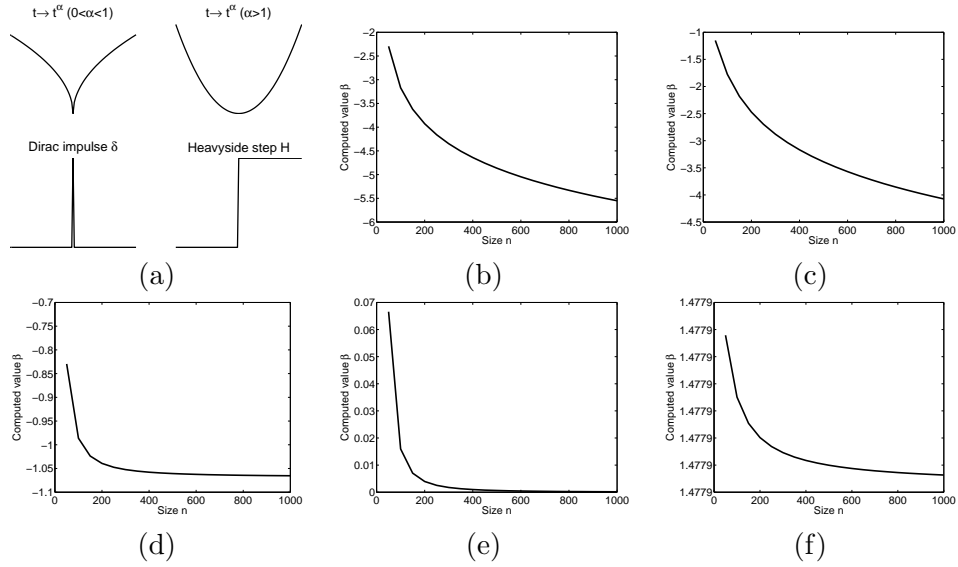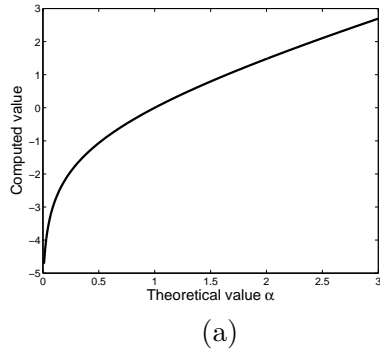$$\beta \approx 3.8(\sqrt{\alpha} - 1) \tag{45}$$

Figure 6: (a) Patterns associated with a specific regularity α; (b-f) Evolution of the computed value β with respect to the sample size, for the patterns. (b): Dirac (spike), (c): Heaviside (level-shift), (d): $t \mapsto \sqrt{t}$, (e): $t \mapsto |t|$, (f): $t \mapsto t^2$.

Regression model $\beta = a\sqrt{\alpha} + b$			
n	a	b	R^2
2000	3.85	-3.94	0.991
1000	3.83	-3.91	0.993
500	3.80	-3.86	0.995
200	3.70	-3.73	0.999
100	3.56	-3.53	0.999
50	3.32	-3.15	0.992

(a)
(b)

Figure 7: (a) Evolution of computed values (β_i) with respect to theoretical values (α_i) for $n = 1000$; (b) Results of the regression of computed values (β_i) against theoretical values (α_i), using the model $\beta = a\sqrt{\alpha} + b$, for different values of n (size of the data).

6   Conclusions, perspectives and applications

In this paper we studied properties of the Lipschitz regularity from theoretical and practical point of view. On the theoretical side we established several properties concerning the behavior of the continuous wavelet transform and the invariance of the Lipschitz regularity. On the practical side, we addressed the issue of the regularity estimation on regular and irregular grids. Concerning regular grids we put forward an original technique that distinguishes regular areas and irregular ones – the robustness of this technique should be further explored. Concerning irregular grids we defined a novel feature using a multiscale method based on smoothing kernels. Although this method is different from the wavelet estimation, it shares an important element: it is based on the behavior of coefficients with respect to scale. We assessed in particular how this computed feature is influenced by the theoretical regularity. Results show strong links between this original feature and the Lipschitz regularity. This encourages new developments on the estimation of regularity on irregular grids. In terms of applications, let us mention the extraction of robust features from time series, either regularly sampled or irregularly. This would be all the more relevant that in the applications real datasets present many missing observations. Using appropriate methodologies, these features would then allow to define indicators of how regular (or irregular) are datasets. This would be absolutely relevant for issues relevant for EU policy such as antifraud, antimoney-laundering and detection of abnormal behavior in trade data, areas to which the Joint Research Centre of the European Commission contributes.

References

[1] A. Arneodo, E. Bacry, S. Jaffard, J. F. Muzy, Singularity spectrum of multifractal functions involving oscillating singularities, Journal of Fourier Analysis and Applications 4 (2) (1998) 159–174.

[2] A. Benassi, S. Cohen, J. Istas, S. Jaffard, Identification of filtered white noises, Stochastic Processes and their Applications 75 (1) (1998) 31–49.

[3] S. Deguy, C. Debain, A. Benassi, Classification of texture images using multi-scale statistical estimators of fractal parameters, British Machine Vision Conference.

[4] S. Jaffard, Y. Meyer, Wavelet methods for pointwise regularity and local oscillations of functions, American Mathematical Society.

[5] W. Rudin, Functional Analysis, McGraw-Hill, 1991.

[6] S. Mallat, W. L. Hwang, Singularity detection and processing with wavelets, IEEE Trans. on Information Theory 38 (2) (1992) 617–643.

[7] S. Mallat, S. Zhong, Characterization of signals from multiscale edges, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (7) (1992) 710–732.

[8] S. Mallat, A wavelet tour of signal processing, Academic Press, 1998.

[9] A. Arneodo, E. Bacry, S. Jaffard, J. F. Muzy, Oscillating singularities and fractal functions, Spline functions and the theory of wavelets, CRM Proc. Lecture Notes 18 (1999) 315–329.

[10] S. Mallat, W. Hwang, Singularity detection and processing with wavelets, IEEE Transactions on Information Theory 38 (1992) 617–643.

[11] C. Damerval, S. Meignen, Study of a robust feature: the pointwise lipschitz regularity, International Journal of Computer Vision 88 (3) (2010) 363–381.

[12] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

[13] G. Hardy, J. Littlewood, G. Plya, Inequalities, Cambridge University Press, 2nd edition, 1988.

A Study of the CWT with respect to scale

Proof. (concerning prop.5)
According to Young's inequalities on convolution products [13], we can write
$\forall u \in \mathbb{R}, \forall s > 0,$

$$
\begin{aligned}
|Wf(u,s)| &= |f * \overline{\psi}_s| \leq ||f||_q ||\overline{\psi}_s||_p & (46) \\
&\leq ||f||_q \left(\int_{\mathbb{R}} \left| \frac{1}{s^{1/p}} \psi(-x/s) \right|^p dx \right)^{1/p} & (47) \\
&\leq ||f||_q \left(\int_{\mathbb{R}} |\psi(y)|^p \, dy \right)^{1/p} & (48) \\
|Wf(u,s)| &\leq ||f||_q ||\psi||_p & (49)
\end{aligned}
$$

ensuring that the CWT is bounded on $\mathbb{R} \times \mathbb{R}_+^*$ $\qquad\square$

Proof. (concerning prop.6)
First, we can write

$$
\begin{aligned}
||\overline{\psi}_s||_\infty &= \frac{1}{s^{1/p}} ||\psi(-\tfrac{\cdot}{s})||_\infty = \frac{1}{s^{1/p}} ||\psi||_\infty & (50) \\
||\overline{\psi}_s||_2 &= \frac{1}{s^{1/p}} ||\psi(-\tfrac{\cdot}{s})||_2 = \frac{1}{s^{1/p}} \sqrt{\int_{\mathbb{R}} \psi^2(-x/s)\, dx} = \frac{1}{s^{1/p}} \sqrt{\int_{\mathbb{R}} \psi^2(y)\, s\, dy} & (51) \\
||\overline{\psi}_s||_2 &= \frac{1}{s^{(1/p)-1/2}} ||\psi||_2 & (52) \\
||\overline{\psi}_s||_1 &= \frac{1}{s^{1/p}} ||\psi(-\tfrac{\cdot}{s})||_1 = \frac{1}{s^{1/p}} \int_{\mathbb{R}} |\psi(-x/s)|\, dx = \frac{1}{s^{1/p}} \int_{\mathbb{R}} |\psi(y)|s\, dy & (53) \\
||\overline{\psi}_s||_1 &= \frac{1}{s^{(1/p)-1}} ||\psi||_1 & (54)
\end{aligned}
$$

leading to the following properties when $s \longrightarrow +\infty$:

$$
\begin{cases}
||\overline{\psi}_s||_\infty &\longrightarrow & 0 \\
||\overline{\psi}_s||_2 &\longrightarrow & 0 \text{ provided } p < 2 \\
||\overline{\psi}_s||_1 &\longrightarrow & 0 \text{ provided } p < 1
\end{cases}
$$

Now, assuming $f \in L^1$, or $f \in L^2$ or alternatively $f \in L^\infty$, since $\psi \in L^1 \cap L^2 \cap L^\infty$, the inequalities of Young concerning convolution products [13] hold:

$$||f * \overline{\psi}_s||_\infty \quad \leq \quad ||f||_1 ||\overline{\psi}_s||_\infty \tag{55}$$

$$||f * \overline{\psi}_s||_\infty \quad \leq \quad ||f||_2 ||\overline{\psi}_s||_2 \tag{56}$$

$$||f * \overline{\psi}_s||_\infty \quad \leq \quad ||f||_\infty ||\overline{\psi}_s||_1 \tag{57}$$

So that, for $f \in L^1$ and $p < 1$, or $f \in L^2$ and $p < 2$, or alternatively $f \in L^\infty$ and any $p > 0$:

$$||Wf||_\infty \longrightarrow 0 \quad \text{when } s \to +\infty \tag{58}$$

$\square$

Proof. (concerning prop.7)
Let us distinguish the general case and the particular one.
General case. Since

$$||f * \overline{\psi}_s||_\infty \quad \leq \quad ||f||_\infty ||\overline{\psi}_s||_1 \leq \frac{1}{s^{(1/p)-1}} ||f||_\infty ||\psi||_1 \tag{59}$$

$$|Wf(u,s)| \quad \leq \quad s^{1-1/p} ||f||_\infty ||\psi||_1 \tag{60}$$

assuming $f \in L^\infty$ and $\psi \in L^1$, and provided $p > 1$

$$|Wf(u,s)| \to 0 \text{ when } s \to 0 \tag{61}$$

Particular case. Assuming f is α-Lipschitz, then there exists $A > 0$ such that: $\forall u \in \mathbb{R}, \forall s > 0$,

$$|Wf(u,s)| \leq As^{\alpha+(1-1/p)} \tag{62}$$

So that, provided $p > \frac{1}{\alpha+1}$

$$|Wf(u,s)| \to 0 \text{ when } s \to 0 \tag{63}$$

$\square$

Proof. (concerning prop.8)
Let us distinguish the different cases (a), (b) and (c).
(a) The CWT associated to the Dirac δ_0 is

$$Wf(u,s) \quad = \quad \delta * \overline{\psi}_s = \overline{\psi}_s(0) = \frac{1}{s^{1/p}} \psi \left(\frac{-u}{s} \right)$$

Since $\psi(0) \neq 0$, for any fixed $u \in \mathbb{R}$,

$$|Wf(u,s)| \sim \frac{\psi(0)}{s^{1/p}} \text{ when } s \to +\infty \tag{64}$$

22

(b) The CWT associated to the function f is

$$
\begin{aligned}
Wf(u,s) &= f * \overline{\psi}_s = \int_0^1 \frac{1}{s^{1/p}} \psi\left(\frac{t-u}{s}\right) dt \\
&= \frac{1}{s^{1/p}} \left[\frac{1}{s} \int_{-u/s}^{(1-u)/s} \psi(z) dz\right]
\end{aligned}
$$

We recall that the distribution T_s defined by

$$
< T_s, \psi > = \frac{1}{s} \int_{-u/s}^{(1-u)/s} \psi(z) dz \tag{65}
$$

converges in the sense of distribution to the Dirac δ_0 when s tends to infinity. Since $\psi(0) \neq 0$, for any fixed $u \in \mathbb{R}$,

$$
|Wf(u,s)| \sim \frac{\psi(0)}{s^{1/p}} \text{ when } s \to +\infty \tag{66}
$$

(c) The CWT associated to f_α is

$$
\begin{aligned}
Wf(u,s) &= f * \overline{\psi}_s \\
&= \frac{1}{s^{1/p}} \int_0^1 t^\alpha (1 - t^\alpha) \psi\left(\frac{t-u}{s}\right) dt
\end{aligned}
$$

Moreover we define the quantity K_α as

$$
\begin{aligned}
K_\alpha &= \int_0^1 t^\alpha (1 - t^\alpha) dt = \int_0^1 t^\alpha - t^{2\alpha} dt \\
&= \left[\frac{t^{\alpha+1}}{\alpha+1} - \frac{t^{2\alpha+1}}{2\alpha+1}\right]_{t=0}^{t=1} = \frac{1}{\alpha+1} - \frac{1}{2\alpha+1} \\
K_\alpha &= \frac{\alpha}{(\alpha+1)(2\alpha+1)}
\end{aligned}
$$

Since $K_\alpha > 0$ ($\alpha > 0$) and $\psi(0) \neq 0$, general theorems on limits of integrals ensure that

$$
Wf(u,s) \sim K_\alpha \frac{\psi(0)}{s^{1/p}} \text{ when } s \to +\infty \tag{67}
$$

$\square$

23

B Properties of the kernel estimator

Let us consider data $(t_i, x_i)_{i=1..N}$ and fixed $t \in [0, 1]$. We define the function $Q :]0, +\infty[\to \mathbb{R}$ as

$$Q(s) = \frac{\sum_i p_i(s)x_i}{\sum_i p_i(s)} \text{ with } p_i(s) = K\left(\frac{t - t_i}{s}\right) \tag{68}$$

The link between this function and the coefficients is straightforward

$$a_{jk} = P_j(t_k) = Q(s_j)_{|t=t_k} \tag{69}$$

$$d_{jk} = P_{j+1}(t) - P_j(t) = Q(s_{j+1})_{|t=t_k} - Q(s_j)_{|t=t_k} \tag{70}$$

Let us study the function $s \mapsto Q(s)$, assuming $(t_i)_{i=1..N}$ are distinct values and considering a Gaussian kernel $K(t) = \exp(t^2/2)$ ($\forall t \in \mathbb{R}$).

Proposition 10. *(Behavior of Q)*
*(i) **Behavior at infinity**:*

$$\lim_{s \to +\infty} Q(s) = \frac{1}{N} \sum_i t_i \tag{71}$$

*(ii) **Behavior at zero**: depending on the value of t, there exists either one or two indexes associated to a minimum of $(|t - t_i|)_{i=1..N}$. In particular, if $t = t_k$ for one $k \in \{1..N\}$, this index k corresponds to the only minimum of $(|t - t_i|)_{i=1..N}$.*
In the case of one index k

$$\lim_{s \to 0} Q(s) = x_k \tag{72}$$

In the case of two indexes k_1 and k_2

$$\lim_{s \to 0} Q(s) = \frac{x_{k_1} + x_{k_2}}{2} \tag{73}$$

*(iii) **Majoration over** $]0, +\infty[$: there exists $C_1 > 0$ so that*

$$|Q(s)| \leq C_1 \quad \forall s > 0 \tag{74}$$

Proof. (i) **Behavior at infinity**

For all $i = 1..N, p_i(s) \longrightarrow p_i(0) = K(0) > 0$ when $s \to +\infty$. Therefore

$$\lim_{s \to +\infty} Q(s) = \frac{\sum_i K(0) t_i}{\sum_i K(0)} = \frac{1}{N} \sum_i t_i \tag{75}$$

(ii) **Behavior at zero**: let us distinguish between two cases.

First case: $t \in \{t_i, i = 1..N\}$. There exists $k \in \{1..N\}$ so that $t = t_k$. Thus $p_k(s) = K(0)$, and for $i \neq k$, $p_i(s) \longrightarrow 0$ when $s \to 0$. This allows to write $\lim_{s \to 0} \sum_i p_i(s) x_i = K(0) x_k$ and $\lim_{s \to 0} \sum_i p_i(s) = K(0)$. Since $Q(s) = \frac{\sum_i p_i(s) x_i}{\sum_i p_i(s)}$

$$\lim_{s \to 0} Q(s) = x_k \tag{76}$$

Second case: $t \notin \{t_i, i = 1..N\}$. We can write $Q(s) = \sum_i A_i x_i$ with

$$A_i = \frac{K(\frac{t - t_i}{s})}{\sum_i K(\frac{t - t_i}{s})} = \frac{\exp(-\frac{(t - t_i)^2}{2s^2})}{\sum_k \exp(-\frac{(t - t_k)^2}{2s^2})} \tag{77}$$

Denoting $\lambda_i = (t - t_i)^2/2$ and $u = 1/s^2$,

$$A_i = \frac{\exp(-\lambda_i u)}{\sum_k \exp(-\lambda_k u)} = \frac{1}{\sum_k \exp(-(\lambda_k - \lambda_i) u)} \tag{78}$$

Let us point out that when $s \to 0$ $(s > 0)$, $u \to +\infty$ and so

$$\text{if } \lambda_k - \lambda_i > 0, \quad \exp(-(\lambda_k - \lambda_i) u) \longrightarrow 0 \tag{79}$$

$$\text{if } \lambda_k - \lambda_i = 0, \quad \exp(-(\lambda_k - \lambda_i) u) = 1 \tag{80}$$

$$\text{if } \lambda_k - \lambda_i < 0, \quad \exp(-(\lambda_k - \lambda_i) u) \longrightarrow +\infty \tag{81}$$

Considering the sum $\sum_k \exp(-(\lambda_k - \lambda_i) u)$ at the denominator of A_i, let us note that this sum never tends to zero since $\lambda_k - \lambda_i = 0$ for $k = i$. This sum tends to infinity provided there exists one k so that $\lambda_k < \lambda_i$. This is always the case except when the index i corresponds to the minimum of $\left((t - t_l)^2/2\right)_{l=1..N}$, also corresponds to the minimum of $(|t - t_l|)_{l=1..N}$.

Now let us define the set

$$I^* = \{i^*, \ |t - t_{i^*}| \leq |t - t_l| \quad \forall l = 1..N\} \tag{82}$$

This set is never empty. Since the $(t_i)_{i=1..N}$ are distinct, this set contains at most two values. It contains exactly two values when $t = (t_{i_1^*} + t_{i_2^*})/2$ for some indexes i_1^* and i_2^*. In this case, when $s \to 0$,

$$A_i \longrightarrow \begin{cases} \frac{1}{2} \text{ if } i \in \{i_1^*, i_2^*\} \\ 0 \text{ if } i \notin \{i_1^*, i_2^*\} \end{cases} \tag{83}$$

This set contains exactly one value in all other cases: $I^* = \{i^*\}$. In this case, when $s \to 0$,

$$A_i \longrightarrow \begin{cases} 1 \text{ if } i = i^* \\ 0 \text{ if } i \neq i^* \end{cases} \tag{84}$$

Since $Q(s) = \sum_i A_i x_i$,

$$\lim_{s \to 0} Q(s) = \begin{cases} x_{i^*} & \text{if } I^* = i^* \\ \frac{x_{i_1^*} + x_{i_2^*}}{2} & \text{if } I^* = \{i_1^*, i_2^*\} \end{cases} \tag{85}$$

(iii) **Majoration over** $]0, +\infty[$: as seen earlier, denoting $\lambda_i = (t - t_i)^2/2$ and $u = 1/s^2$,

$$0 \leq A_i = \frac{\exp(-\lambda_i u)}{\sum_k \exp(-\lambda_k u)} \leq \frac{\exp(-\lambda_i u)}{\exp(-\lambda_i u)} = 1 \tag{86}$$

Since $Q(s) = \sum_i A_i(s) x_i$,

$$|Q(s)| \leq \sum_i |A_i(s)| \cdot |x_i| \leq \sum_i |x_i| \leq C_1 \tag{87}$$

$\square$

Abstract

In this paper we study the pointwise Lipschitz regularity, covering several aspects: theoretical and practical, methods for its estimation on regular and irregular grids. The relevance of this value of regularity lies in its invariance properties to several transformations, and its fast computation thanks to wavelets. We study the influence of scale on wavelets transforms and show invariance properties this value of regularity. We also put forward an original technique for its estimation on regular grids. We also address the issue of irregular grids, based on the behaviour of smoothing kernels with respect to scale. The obtained results emphasize the usefulness of such features for the applications, and motivate further work on this topic.

LB-NA-25492-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office