JRC TECHNICAL REPORTS

Study of pointwise regularity and application to trade data

Christophe Damerval

2012

Report EUR 25489 EN

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Christophe Damerval
Address: Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy
E-mail: christophe.damerval@jrc.ec.europa.eu or spyros.arsenis@jrc.ec.europa.eu
Fax: +39 0332 78 5154

http://theseus.jrc.ec.europa.eu
http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC72824

EUR 25489 EN

ISBN 978-92-79-26261-6

ISSN 1831-9424

doi:10.2788/4550

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

Study of pointwise regularity and application to trade data

C. Damerval

Institute for the Protection and Security of the Citizen
Joint Research Center of the European Commission
Via Enrico Fermi 2749, 21027 Ispra (VA) ITALY
Email: christophe.damerval@jrc.ec.europa.eu

Abstract

In this paper we focus on the pointwise Lipschitz regularity in 1D and 2D. We put the emphasis on its invariance properties to a wide range of transformations. Wavelets algorithms provide fast computations, which is desirable in the applications. In addition to theoretical properties, a practical evaluation of its robustness is possible in practice. This leads to the conclusion that the regularity stands out as a robust pointwise features in 1D as well as in 2D. As an application, we use it to extract features that are indicators of potential fraud, through the processing of trade data.

Keywords: Lipschitz regularity, wavelets, feature extraction

1 Introduction

The pointwise Lipschitz regularity is a feature that quantifies the regularity a function, associated for instance to a 1D signal or time series, a 2D surface or an image. This value $\alpha \in \mathbb{R}$ allows to measure the sharpness of edges and the smoothness of variations. Various works in signal and image processing studied its properties and applications [1, 3, 6, 8].

Here we are interested in the pointwise Lipschitz regularity since it makes a robust feature useful in the applications. Here the notion of robustness means that the quantities computed are only slightly affected by transformations either geometric or photometric. The notion of feature is related to the extraction of quantities that are characteristic of an interpretable entity: for instance the size

of an object, or the regularity of a pattern[5, 14]. This feature extraction appears as relevant for fraud detection, an important field to which the Joint Research Centre of the European Commission contributes.

This paper is organized as follows: we first recall the notion of regularity and its classical computation with wavelets; then we present new theoretical properties of the regularity and practical results showing it makes up robust feature for certain geometrical transformations. Finally we present an application to the processing of trade data, defining and matching features based on regularity.

2 Notion of regularity

2.1 Monodimensional case.

Definition 1. *(1D 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.*

Remark: for $\alpha \leq 0$, it can be defined thanks to the theory of distributions [16].

Definition 2. *(1D regularity α) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $t_0 \in \mathbb{R}$. The regularity α of f at t_0 is*

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

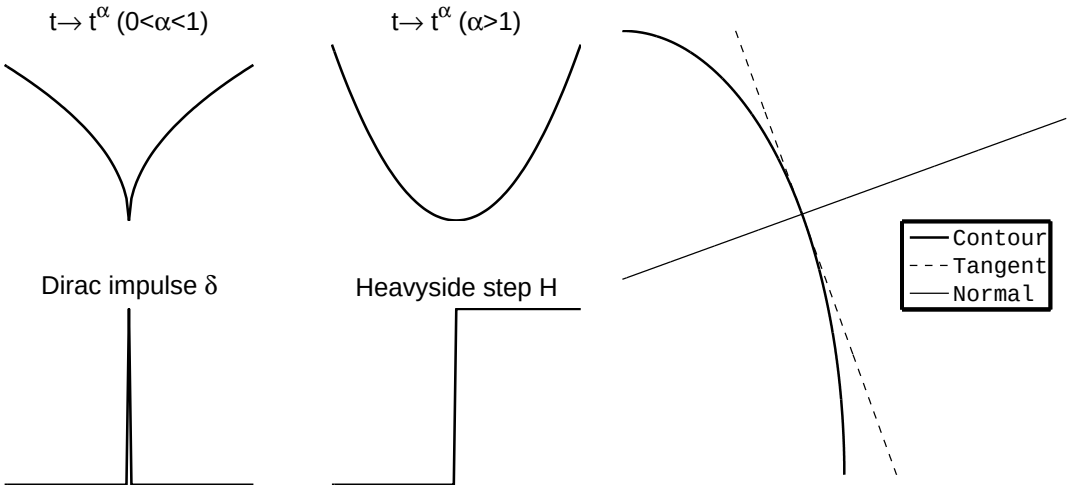


Figure 1: (a) Classical patterns in 1D, associated to different values of Lipschitz regularity ($\alpha = -1$ for the Dirac, $\alpha = 0$ for the step). (b) 2D contour line, associated to a regular pattern along the contour and an irregular pattern in the perpendicular direction.

The point of studying the pointwise Lipschitz regularity is the detection and identification of singularities [12], 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 Bidimensional case.

The notion of regularity in the 2D case can be defined in different ways: either isotropic (regularity in a neighborhood), or directional (regularity in a certain direction). In the case of an image representing a contour (a line for example), the intensity f is regular along the tangent of the contour and irregular along the normal direction, see Fig.1(b). So as to grasp this direction of maximum irregularity, it is more relevant to consider a directional definition. In this context, the 2D definition uses the one in 1D.

Definition 3. *(2D Lipschitz regularity) Let* $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *and* $x_0 \in \mathbb{R}^2$*. For any* $\theta \in [0, \pi[$*, denoting* $u_\theta = (\cos\theta, \sin\theta) \in \mathbb{R}^2$*, we define* $f_\theta : \mathbb{R} \longrightarrow \mathbb{R}$ *as* $f_\theta(h) = f(x_0 + hu_\theta)$ $(\forall h \in \mathbb{R})$*. Now, for* $\alpha \in \mathbb{R}$*,* f *is* α*-Lipschitz at* $x_0 \in \mathbb{R}^2$ *if*

$$\exists \theta \in [0, \pi[, f_\theta \ \alpha\text{-Lipschitz at } 0 \tag{4}$$

Definition 4. *(2D regularity* α*)*

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *and* $x_0 \in \mathbb{R}^2$*. The regularity* α *of* f *at* x_0 *is defined as*

$$\alpha = \inf\{\alpha_0 \in \mathbb{R}, \ f \ \alpha_0\text{-Lipschitz at } x_0\} \tag{5}$$

From here, we consider an image given by its intensity function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$. We focus on the pointwise Lipschitz regularity of f (denoted regularity α). First we present the notion of regularity α in 2D, as a generalization of the 1D case. Then we investigate its invariance properties, especially in the case of geometric transformations. Finally note that when $\alpha \in]0, 1[$, assuming f is α-Lipschitz at x_0, we can write (with $h > 0$ in a neighborhood of 0)

$$|f(x_0 + (h\cos\theta, h\sin\theta)) - f(x_0)| = |f_\theta(h) - f_\theta(0)| \leq Ah^\alpha \tag{6}$$

so that definition 4 agrees with the usual definition of the Lipschitz regularity.

2.3 Numerical estimation

To estimate the regularity α, the wavelet framework shown its efficiency [12, 13]: these multiscale methods enable to detect singularities and to estimate α robustly. We recall the known link between Lipschitz regularity and wavelets [8, 11] and the application to the computation of the regularity α. This result can be applied in 1D, and also in 2D since the bidimensional definition of the regularity α uses the monodimensional definition.

Proposition 1. *(Jaffard) Let $f : \mathbb{R} \to \mathbb{R}$, $f \in L^2$. We consider its continuous wavelet transform Wf, defined as: $\forall s > 0, \forall u \in \mathbb{R}$*

$$Wf(u,s) = \frac{1}{s} \int_{\mathbb{R}} f(t) \psi \left(\frac{t-u}{s} \right) dt \tag{7}$$

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: $\forall u \in \mathbb{R}, \forall s > 0$,

$$|Wf(u,s)| \leq A s^{\alpha} \left(1 + \left| \frac{u - t_0}{s} \right|^{\alpha} \right) \tag{8}$$

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

$$|Wf(u,s)| \leq A s^{\alpha} \left(1 + \left| \frac{u - t_0}{s} \right|^{\alpha'} \right) \tag{9}$$

then f is α-Lipschitz at x_0.

2.3.1 Practical computation of α.

(Mallat) Starting from eq.(8) we get $|Wf(u,s)| \leq A s^{\alpha}$ and then $\log |Wf(u,s)| \leq \log A + \alpha \log s$. Now, this latter inequality becomes a quasi-equality at fine scales [12]. So the regularity α can be estimated by performing a regression based on the formula

$$\log |Wf(u,s)| = \alpha \log s + C \tag{10}$$

More precisely α and C are computed by a regression of $\log |Wf(u,s)|$ over $\log s$ at fine scales. At each scale we use the coefficient which is maximum in a neigborhood. Generally three scales are sufficient for an accurate estimation. This corresponds to an estimation of the regularity α along maxima lines.

2.3.2 Monodimensional case.

A classical approach in 1D consists on computing values of regularity α 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. The computation of the regularity α is done by using the formula (10), where Wf is a wavelet transform with an appropriate wavelet. From an algorithmic point of view, thanks to wavelet algorithms, this can be performed in $O(N)$ operations (N denoting the size of the data).

2.3.3 Bidimensional case.

A known approach in 2D consists on computing values of regularity α at contour points, which are singularities in a particular direction. These contour points can be defined as the maxima in the sense of Canny (MC), i.e., the locations $(x, y) \in \mathbb{R}^2$ where the magnitude of the gradient attains a local maximum in the direction of the gradient [4]. Fast and efficient computations are provided by a wavelet formulation [13], that defines a multiscale gradient magnitude $Mf(., ., s)$. The computation of the regularity α is done by using the formula (10), where $|Wf|$ is replaced by the gradient magnitude Mf. From an algorithmic point of view, denoting N the size of the data (for an image $n \times n$, $N = n^2$), this can be performed in $O(N)$ operations.

3 Regularity as a robust feature

We present here original properties of the regularity α. First we show theoretical properties on the invariance of the regularity; secondly, we present a methodology which quantifies the robustness of the regularity in practice. This allows to say that the regularity α is a feature robust to some geometrical transformations.

3.1 Theoretical invariance properties

Here we present invariance properties of the regularity α under different transformations, geometric and photometric. 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 2. *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,*

$$\forall \alpha > 0,\ f\ \alpha\text{-Lipschitz at } t_0 \Rightarrow g\ \alpha\text{-Lipschitz at } t_0 \tag{11}$$

Proposition 3. *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,*

$$\forall \alpha > 0, f\ \alpha\text{-Lipschitz at } u(t_0) \Rightarrow g\ \alpha\text{-Lipschitz at } t_0 \tag{12}$$

More generally the Lipschitz regularity of a function will not be altered by a transformation, provided it does not correspond to a smoothing or a sharpening. We will see it is the case for geometric transformations, such as affine deformations in 2D.

3.1.1 Monodimensional case.

Here we consider the regularity α in 1D (see def.2) and we investigate its invariance properties.

Proposition 4. *(Invariance to multiplication and addition of C^∞ functions) Let $f, g : \mathbb{R} \to \mathbb{R}$, related by: $\forall t \in \mathbb{R}$*

$$g(t) = c(t) \cdot f(t) + d(t) \tag{13}$$

with $c, d : \mathbb{R} \to \mathbb{R}$ are C^∞ and $c \neq 0$. Then, at any $t_0 \in \mathbb{R}$, f and g have the same regularity α (for $\alpha > 0$).

Proposition 5. *(Invariance to dilatation and translation) Let $f, g : \mathbb{R}^2 \to \mathbb{R}$, related by: $\forall x \in \mathbb{R}^2$*

$$g(ct + d) = f(t),\ \text{with } c, d \in \mathbb{R}, c \neq 0 \tag{14}$$

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

$$\alpha(f, t_0) = \alpha(g, ct_0 + d) \tag{15}$$

3.1.2 Bidimensional case.

Here we consider the regularity α in 2D (see def.4) that focuses on directions of maximum irregularity. We investigate its invariance properties to different transformations, in particular affine deformations.

Proposition 6. *(Invariance to contrast and illumination change) Let $f, g : \mathbb{R}^2 \to \mathbb{R}$, related by: $\forall x \in \mathbb{R}^2$*

$$g(x) = cf(x) + d, \text{ with } c \neq 0 \text{ and } d \in \mathbb{R} \tag{16}$$

Then, at any $x_0 \in \mathbb{R}^2$, f and g have the same regularity α.

Proposition 7. *(Invariance to affine deformations) Let $f, g : \mathbb{R}^2 \to \mathbb{R}$, related by: $\forall x \in \mathbb{R}^2$*

$$g(x) = f(Bx), \text{ with } B \text{ a } 2 \times 2 \text{ invertible matrix} \tag{17}$$

Then, for any $x_0 \in \mathbb{R}^2$, f and g have the same regularity α, respectively at x_0 and $y_0 = B^{-1}x_0$

$$\alpha(f, x_0) = \alpha(g, B^{-1}x_0) \tag{18}$$

3.1.3 Conclusions.

We established invariance properties of the regularity α: in 1D, to constant translation and dilatation, and also to the multiplication by a C^∞ function; in 2D, to photometric transformation (contrast and illumination change) and to geometric ones (affine deformation). This motivates a practical study.

3.2 Practical robustness properties

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. This is inspired by repeatability tests, a tool in the field of computer vision allowing to evaluate the robustness of interest regions [10, 14]. Here we consider the bidimensional case, and we are interested in evaluating the robustness of the regularity α (pointwise feature) to changes in imaging conditions (geometric and photometric transformations). We assume that the detector gives values α_i (regularity α) at locations $(x_i, y_i) \in \mathbb{R}^2$ (singularities), each location being associated with one value of regularity. It is the case of the algorithm mentioned in section 2.3, in particular on natural images the associated detector gives many values of the regularity α, which is useful for applications in computer vision.

We study here the robustness of the estimation of α in 2D, more precisely natural images. We define different scores that reflect the robustness of the regularity α for one given transformation (the higher score, the more robust). Given an image of grey-level intensity f, we can obtain the set

$$\left\{(x_i, y_i, \alpha_i) \in \mathbb{R}^3, 1 \leq i \leq n_f\right\} \tag{19}$$

where n_f is the number of detected edge points (x_i, y_i), each being associated to a value α_i (this can be performed by the afore-mentioned detector). To evaluate the robustness of the regularity, we consider image sequences representing the same scene viewed under different imaging conditions. The main idea is the following: provided transformations from one image to another is known *a priori*, it is possible to perform point-to-point correspondences and then to compare the values of regularity. Ideally these values should remain stable, even for significant degrees of transformation.

3.2.1 Methodology.

Here we consider one sequence of 6 images $(X_k)_{0\leq k\leq 5}$, representing the same scene. Each image X_k $(1 \leq k \leq 5)$ corresponds to a transformation applied to the reference image X_0, such as a known change of viewpoint. We carry out the following steps:

1. For each image $(X_k)_{0\leq k\leq 5}$, detect edge points and compute associated values of regularity α: $p_i^k = (x_i^k, y_i^k, \alpha_i^k), 1 \leq i \leq n_k$.

2. For fixed k $(1 \leq k \leq 5)$, determine a set of $\mathcal{C}^k$ of point-to-point correspondences between edge points of X_0 and X_k (thanks to the known homography between these images)

$$\mathcal{C}^k = \left\{ \begin{array}{c} (p_i^0, p_j^k) \text{ matched according} \\ \text{to a geometric criterion} \end{array} \right\} \tag{20}$$

   This leads to a certain number of correspondences (NC) $\#\mathcal{C}^k$.

3. Select the subset $\mathcal{C}_\epsilon^k$ of correspondences for which regularities are sufficiently close (according to a parameter $\epsilon > 0$)

$$\mathcal{C}_\epsilon^k = \Big\{ (p_i^0, p_j^k) \in \mathcal{C}^k, d_\alpha = |\alpha_i^0 - \alpha_j^k| < \epsilon \Big\} \tag{21}$$

   and compute the matching score, representing the proportion of correct matches:

$$S_k = \frac{\#\mathcal{C}_\epsilon^k}{\#\mathcal{C}^k} \tag{22}$$

This score reflects the robustness of the regularity α. It depends on the degree of the transformation, indexed by k ($k = 1$ corresponds to a small change of perspective, while $k = 5$ is associated to a significant change of perspective). It also depends on the parameter ϵ, a tolerance parameter playing the same role as the confidence level of a statistical test.

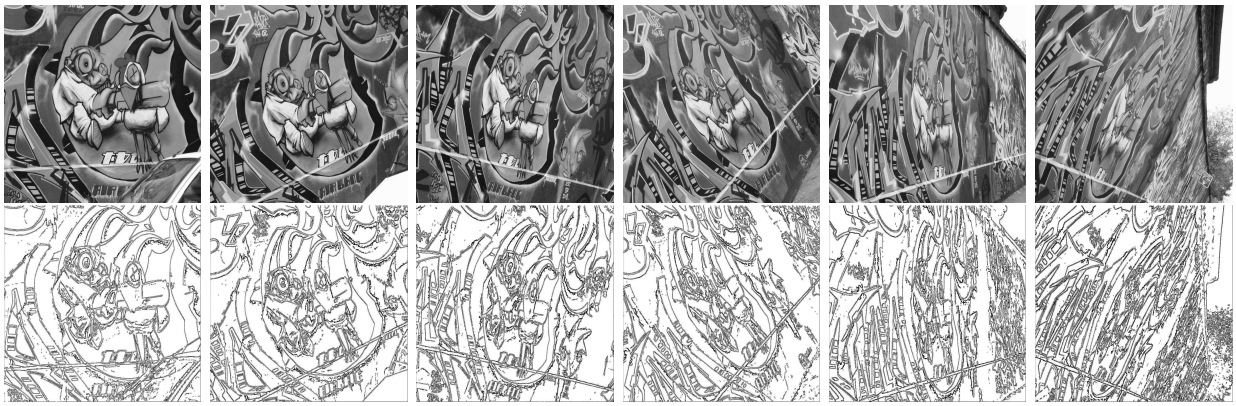
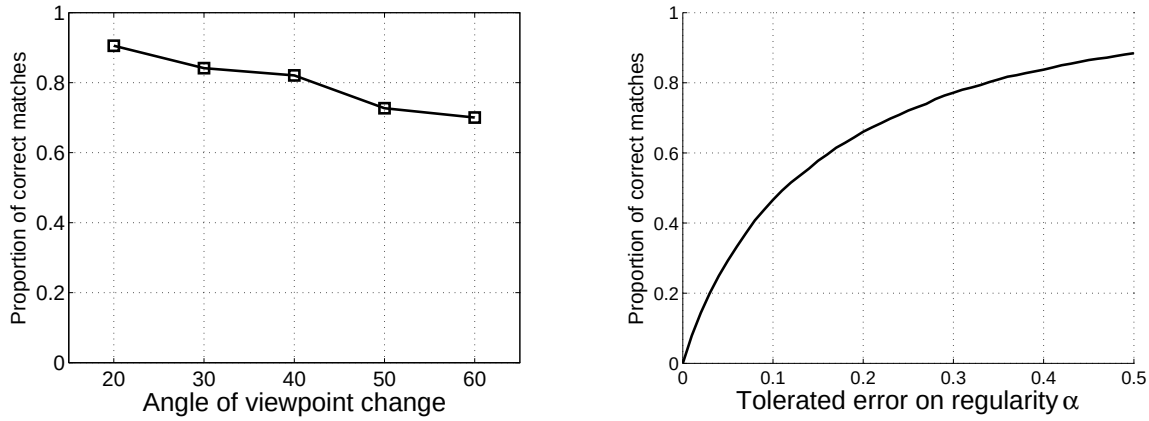


Figure 2: Top: image test sequence X_0,...,X_5 associated to a viewpoint change; Bottom: detected edges, one value of the regularity α being computed at each edge point. In our methodology, each image $(X_k)_{1\leq k\leq 5}$ is compared with the reference image X_0.



(a) (b)

Figure 3: (a) Evolution of the matching score with respect to the index k representing the degree of the transformation (for $\epsilon = 0.25$) (b) Evolution of the matching score with respect to the tolerance parameter ϵ (for $k = 6$).

3.2.2 Results.

We represent on Fig.2 the image sequence $(X_k)_{0\leq k\leq 5}$ and the detected edges. This sequence corresponds to the same scene viewed under different angles of perspectives. This transformation correspond to actual camera operations, taking pictures of a natural scene from different angles

of perspective. For additional results on various imaging conditions, see [5]. Using the proposed methodology, we compute the matching scores S_k corresponding to the comparison between X_0 and X_k ($1 \leq k \leq 5$). We represent on Figure 3(b) the evolution of the matching score with respect to the index k representing the degree of the transformation, for fixed ϵ (here $\epsilon = 0.25$). This curve remains at a high level, which means that the computed values (α_i) are slightly affected by this change of perspective. Moreover we represent on Figure 3(a) the evolution of the matching score with respect to the tolerance parameter ϵ for one fixed k (here $k = 6$). This curve increases sharply, which means high matching scores for a moderate values of the tolerance parameter: for instance 70% of the correspondences lead to values for which $d_\alpha \leq 0.25$. These good results show the regularity α remains stable under a change of perspective, which is a very general transformation.

3.2.3 Conclusions.

From the theoretical point of view, the regularity α is invariant under a change of contrast and illumination (prop.6). It is also invariant under an affine deformation (prop.7). This includes in particular rotation and scale change, widely studied transformations in Scale-Space theory [10]. From the practical point of view, the regularity α turns out as a robust feature, especially to geometric transformations such perspective changes (and also to other geometric and photometric transformations [5]). This underlines the relevance of the regularity α for applications in image processing and computer vision, for instance the extraction of relevant descriptors.

4 Application: feature extraction from trade data

The identification of fraud or money laundering is a difficult task, to which the Joint Research Center of the European Commission contributes. In particular, a relevant source of information for this task is trade data: large amount of data collected by EU customs, reporting price and volume of a variety of products exchanged between EU and third countries. Given this data, an important problem is the definition of relevant indicators of potential fraud or money laundering. While approaches based on statistical outliers showed their efficiency [15], we propose here original features defined thanks to the regularity α.

4.1 Methodology for extracting features

We consider import trade flows extracted from the COMEXT database, covering the period 1999–2008. This database is designed by Eurostat and contains detailed information on trading in goods of all EU member states with third countries. This dataset gives the price and the volume of a product exchanged between two countries, reported at different dates, for a certain number of triplets (Product,Origin,Destination). Each triplet corresponds to data $\{x_i, y_i \in \mathbb{R}, 1 \leq i \leq N\}$, x_i: price per unit of volume, y_i: volume in tons (here we consider 1346 triplets with $N = 120$). For each triplet, and for each of these two time series, we apply the following steps:

1. Given one chosen scale of interest (typically a fine scale, here $s = 1$), compute the wavelet transform Wf associated to the data (x_i) or (y_i)

2. Identify the set of singularities $\mathcal{S}$, defined as locations associated to modulus maxima of the wavelet transform and also to a response higher than a certain threshold. To surpass the noise level, we set it as $thresh = \text{MAD}/0.6745$, where MAD is the median absolute deviation of the wavelet coefficients [7].

$$\begin{aligned} \mathcal{S} \quad = \quad & \{u \in \mathbb{R}, |Wf(., s)| \text{ locally maximum} \\ & \text{and } |W(u, s)| > thresh\} \end{aligned} \tag{23}$$

3. Compute values of regularity (α_i) at each of these locations. We use the algorithm seen in section 2.3, in particular the formula (10) using $s = 1, 2, 3$.

4. Split into three groups: spikes, if $\alpha_i < -1/2$; level shifts, if $\alpha_i \in [-1/2, 1/2]$; regular patterns, if $\alpha_i > 1/2$ (we recall that $\alpha = -1$ corresponds to a Dirac impulse, while $\alpha = 0$ corresponds to a Heaviside step).

5. This allows to define the following features, associated to one time series:

$$\begin{cases} \text{Type} & : & \text{"Spike", "Level shift"} \\ & & \text{or "Regular pattern"} \\ \text{Regularity} & : & \alpha_i \in \mathbb{R} \\ \text{Location} & : & u_i \in \mathbb{R} \\ \text{Strength} & : & \frac{N \cdot |Wf(u_i,s)|}{\sum_i |Wf(u_i,s)|} > 0 \end{cases} \tag{24}$$

4.2 Results

When applied to large datasets, this methodology allows to automatically extract a set of features containing relevant information, through the detection of such or such pattern. For illustration purposes, we represent one of these time series and the detected singularities on Figure 4(a). Besides, we represent on Figure 4(b) a scatterplot of the response $|Wf|$ and the regularity α computed at every location, highlighting modulus maxima and responses above the threshold. The type of pattern is determined according to the regularity. The strength defined in eq.(24) gives additional information on the relative weight of the wavelet coefficient compared to the others (a high strength corresponding to a significant singularity). Now, focusing only at the regularity α computed at singularities (as defined in eq.(23)), let us now apply this methodology on the whole dataset, with a view to compare time series associated to value (TSV) and those associated to a quantity (TSQ). We represent on Figure 5(a) the distribution of the regularity α associated to the singularities detected in the TSV (the distribution is similar for the TSQ); this corresponds to typical values of α obtained in various applications. In Table 1 we precise statistics on obtained values, distinguishing spikes, level shifts and regular patterns. Now, let us compare value and quantity. For each triplet (Product,Origin,Destination), we can obtain two sets of features (Type,Regularity, Location, Strength) and establish correspondences by selecting those having the same location:

$$(\alpha_V, u_V) \longleftrightarrow (\alpha_Q, u_Q) \quad \text{with } u_V = u_Q \tag{25}$$

Then we compare the obtained values and consider it is a correct match if

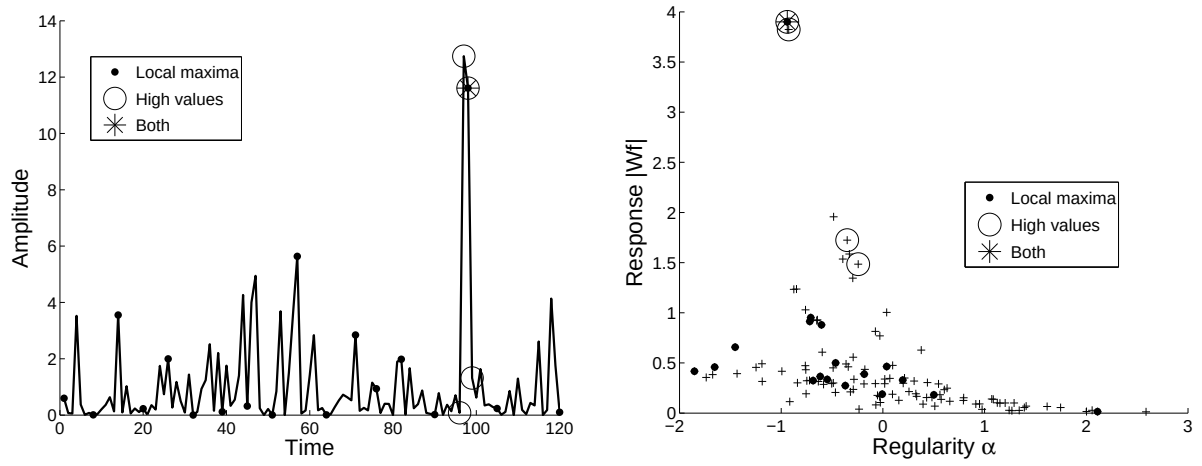
$$d_\alpha = |\alpha_V - \alpha_Q| < \epsilon \text{ with a tolerance parameter } \epsilon > 0 \tag{26}$$

This comparison is very similar to the one described in section 3.2 (first establish correspondences, then compare values of regularity). We represent on Figure 5(b) the proportion of the correct matches (over all matches), tracking its evolution with the parameter ϵ. In addition, we represent the same curve on Figure 5(c) distinguishing the different types of patterns. In all of these we observe a sharp increase with respect to ϵ of all these curves: a high proportion of correct matches is attained for low values of ϵ. We note in particular that $d_\alpha < 0.3$ for 60% of the correspondences. From these results we conclude that these time series reporting price and quantity lead to a significant number of features, many of them sharing the same location, and for which the computed values of regularity α are close.

Conclusions. Features based on the regularity α appear as relevant indicators adapted to the processing of trade data. The presented features lays the emphasis on regularity α, gives additional information on the underlying pattern (like a spike or a level shift) and on its relative importance (using the modulus of wavelet coefficients). In addition, comparing time series reporting both price and volume, we found a significant number of features in common. This motivates further analysis of this data, exploiting simultaneously price and volume data so as define relevant indicators.

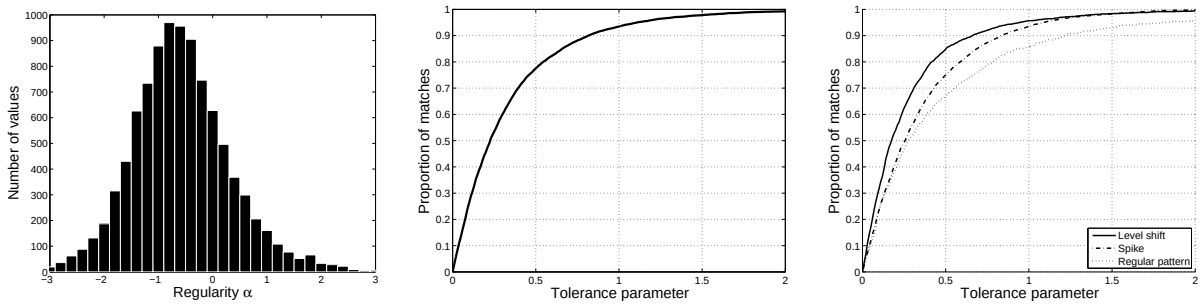
	Spike		Level shift		Regular pattern	
	V	Q	V	Q	V	Q
Number of singularities	5454	5071	3143	3382	1085	1107
mean(α)	-1.18	-1.17	-0.089	-0.086	1.12	1.18
std(α)	0.52	0.52	0.28	0.25	0.63	0.70
Average strength	2.44	2.53	1.66	1.74	0.53	0.52

Table 1: Summary of the values α computed at singularities for time series reporting value V and quantity Q (1346 time series covering a 12 year period with monthly data).



(a) (b)

Figure 4: Illustration of feature extraction from time series: (a) Data reporting a price over time, represented in (Amplitude, Time) and detected singularities; (b) Representation of this data in (Response $|Wf|$, Regularity α).



(a) (b) (c)

Figure 5: Analysis of features extracted from time series reporting value and quantity, the regularity α being computed at singularities: (a) Histogram of computed (α_i), concerning value; (b) Proportion of the correct matches with respect to the tolerance parameter; (c) *Idem*, representing separately spikes, level shifts, and regular patterns.

5 Conclusions and perspectives

In this paper we studied the Lipschitz regularity in 1D and 2D, focusing on the influence of different transformations. Both theoretical properties and empirical results show that this feature is robust to various transformations, especially geometrical ones. This underlines its usefulness for many applications; in particular we proposed an algorithm allowing to extract features from trade data, which can be further interpreted as indicators of potential fraud. In this regard, we successfully matched features extracted from time series reporting respectively value and quantity. This encourages further processing of trade data for applications in the fields of fraud detection and anti-money laundering. From the theoretical point of view, it appears as interesting to establish invariance properties of the Lipschitz regularity for negative value of α. From the practical point of view, many applications in signal and image processing can benefit from the Lipschitz regularity, pointwise feature which contrasts with methods focusing on local features.

References

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

[2] O. Barriere, and J. Levy Vehel, Local Holder regularity-based modeling of RR intervals 21th IEEE International Symposium on Computer-Based Medical Systems, June 17-19, 2008, Jyvaskyla, Finland, 2008.

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

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

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

[6] S. Deguy, C. Debain, and Albert Benassi. Classification of texture images using multi-scale statistical estimators of fractal parameters. *British Machine Vision Conference*, 2000.

[7] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[8] S. Jaffard and Y. Meyer. Wavelet methods for pointwise regularity and local oscillations of functions. *American Mathematical Society*, 1996.

[9] H.S. Lee, Y. Cho, H. Byun, J. Yoo, An Image Enhancement Technique Based on Wavelets Proceedings of the First IEEE International Workshop on Biologically Motivated Computer Vision (BMVC '00), Seoul, Korea, May 15-17, 2000.

[10] T. Lindeberg. *Scale Space Theory in Computer Vision*. Kluwer, 1994.

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

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

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

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 62(1):43–72, 2005.

[15] M. Riani, A. Cerioli, A. Atkinson, D. Perrotta, and F. Torti. Fitting mixtures of regression lines with the forward search. In *Mining Massive Data Sets for Security*, pages 271–286. IOS Press, Amsterdam (Netherlands), 2008.

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

[17] H. Shekarforoush, Image Registration Using Directional Local Holder Regularity, Proceedings of the 34th Conference on Information Sciences and Systems, Princeton, March 2000.

Appendix. Properties of the regularity α

Here we explain the proofs of the propositions mentioned in section 3.

Proof. (of prop.2) Let us assume that f is α-Lipschitz at t_0 with $\alpha \in]0,1[$.

$$\begin{aligned} |g(t)-g(t_0)| &\leq |c(t)f(t)-c(t_0)f(t_0)| + |d(t)-d(t_0)| \\ &\leq |c(t)| \cdot |f(t)-f(t_0)| + |f(t_0)| \cdot |c(t)-c(t_0)| + |d(t)-d(t_0)| \end{aligned} \tag{27}$$

Since c, d are C^∞ and therefore α-Lipschitz

$$|d(t)-d(t_0)| \leq A_0|t-t_0|^\alpha \tag{28}$$

$$|c(t)-c(t_0)| \leq A_1|t-t_0|^\alpha \tag{29}$$

Since c is bounded and f α-Lipschitz

$$|c(t)| \cdot |f(t)-f(t_0)| \leq A_2|t-t_0|^\alpha \tag{30}$$

Denoting with $A = A_0 + A_1 + A_2 + |f(t_0)|$, combining the preceding inequalities leads to

$$|g(t)-g(t_0)| \leq A|t-t_0|^\alpha \tag{31}$$

Using the polynomial formulation seen in def.1, this can be generalized to $\alpha > 1$. □

Proof. (of prop.3) Let us assume that f is α-Lipschitz at $u(t_0)$ with $\alpha \in]0,1[$. Since u is 1-Lipschitz, we can write (with $A > 0$)

$$|g(t)-g(t_0)| = |f(u(t))-f(u(t_0))| \tag{32}$$

$$\leq A|u(t)-u(t_0)|^\alpha \tag{33}$$

$$\leq A|(|t-t_0|^1)|^\alpha = A|t-t_0|^\alpha \tag{34}$$

This shows that for $\alpha \in]0,1[$

$$f \text{ } \alpha\text{-Lipschitz at } u(t_0) \Rightarrow g \text{ } \alpha\text{-Lipschitz at } t_0 \tag{35}$$

Using the polynomial formulation seen in def.1, this can be generalized to $\alpha > 1$. □

Proof. (of prop.4) First, since let us note that

$$g(t) = c(t) \cdot f(t) + d(t) \tag{36}$$
$$f(t) = \frac{1}{c(t)} g(t) + \frac{-d(t)}{c(t)} \tag{37}$$

Then, applying twice prop.2, we get

$$\forall \alpha > 0, f \text{ } \alpha\text{-Lipschitz at } t_0 \Leftrightarrow g \text{ } \alpha\text{-Lipschitz at } t_0 \tag{38}$$

Since the regularity α is defined as the infimum of the α-Lipschitz regularity at one point t_0, we can write

$$\alpha(f, t_0) = \alpha(g, t_0) \tag{39}$$

□

Proof. (of prop.5) Let $u, v : \mathbb{R} \to \mathbb{R}$ defined as $u(t) = ct + d$ and $v(t) = (t - d)/c$ (for all $t \in \mathbb{R}$). These two functions are 1-Lipschitz and invertible, and $u^{-1} = v$. Applying twice prop.3, we can obtain

$$f \text{ } \alpha\text{-Lipschitz at } u(t_0) = ct_0 + d \Rightarrow g \text{ } \alpha\text{-Lipschitz at } t_0 \tag{40}$$
$$g \text{ } \alpha\text{-Lipschitz at } t_0 \Rightarrow f \text{ } \alpha\text{-Lipschitz at } v^{-1}(t_0) = ct_0 + d \tag{41}$$
$$f \text{ } \alpha\text{-Lipschitz at } ct_0 + d \Leftrightarrow g \text{ } \alpha\text{-Lipschitz at } t_0 \tag{42}$$

From this last equation, taking the infimum on α, we conclude

$$\alpha(f, ct_0 + d) = \alpha(g, t_0) \tag{43}$$

□

Proof. (of prop.6) Denoting $\alpha \in \mathbb{R}$ (resp. α') the pointwise regularity of f (resp. of g) at x_0, we can write

$$\alpha' = \inf\{\alpha_0 \in \mathbb{R}, \text{ } g \text{ } \alpha_0\text{-Lipschitz at } x_0\}$$
$$\alpha' = \inf\{\alpha_0 \in \mathbb{R}, \exists \theta \in [0, \pi[, g_\theta \text{ } \alpha_0\text{-Lipschitz at } 0\}$$

with $g_\theta = g(x_0 + hu_\theta) = cf(x_0 + hu_\theta) + d = cf_\theta + d$. Since $g_\theta = cf_\theta + d$ and $f_\theta = \frac{1}{c}g_\theta - \frac{d}{c}$, the monodimensional functions g_θ and f_θ have the same Lipschitz regularity: g_θ α-Lipschitz $\Leftrightarrow$ f_θ α-Lipschitz (see prop. 2). Therefore f and g have the same regularity: $\alpha = \alpha'$. □

Proof. (of prop.7) Denoting $\alpha(f, x_0) = \alpha \in \mathbb{R}$ the regularity of f at x_0, let us study the regularity of g at $y_0 = B^{-1}x_0$. According to def. 4, there exists $A > 0$ so that

$$\exists \theta \in [0, \pi[, \quad |f(x_0) - f(x_0 + hu_\theta)| \leq Ah^\alpha \tag{44}$$

For this $\theta \in [0, \pi[$ we have

$$|g(y_0) - g(y_0 + hu_\theta)| \tag{45}$$
$$= |g(B^{-1}x_0) - g(B^{-1}x_0 + hu_\theta)| \tag{46}$$
$$= |f(x_0) - f(x_0 + hBu_\theta)| \tag{47}$$

Since $Bu_\theta = \lambda u_{\theta'}$ with $\lambda \in \mathbb{R}$ ($\lambda \neq 0$) and $\theta' \in [0, \pi[$:

$$\begin{aligned} |g(y_0) - g(y_0 + hu_\theta)| &= |f(x_0) - f(x_0 + h\lambda u_{\theta'})| \\ &\leq (A|\lambda|^\alpha)h^\alpha \end{aligned} \tag{48}$$

So there exists $A' > 0$ so that

$$|g(y_0) - g(y_0 + hu_\theta)| \leq A'h^\alpha \tag{49}$$

and g is α-Lipschitz at $y_0 = B^{-1}x_0$. Since the regularity α is defined as an infimum (see def.5), we can infer

$$\alpha(g, y_0) \leq \alpha(f, x_0) \tag{50}$$

Now, given the invertibility of the matrix B, we can repeat the preceding reasoning for g, so as to infer also $\alpha(g, y_0) \geq \alpha(f, x_0)$. This allows to conclude

$$\alpha(g, y_0) = \alpha(f, x_0) \tag{51}$$

□

European Commission
EUR 25489 EN Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **Study of pointwise regularity and application to trade data**

Authors: Christophe Damerval

Luxembourg: Publications Office of the European Union

2012 – 25 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-26261-6

doi:10.2788/4550

Abstract

In this paper we focus on the pointwise Lipschitz regularity in 1D and 2D. We put the emphasis on its invariance properties to a wide range of transformations. Wavelets algorithms provide fast computations, which is desirable in the applications. In addition to theoretical properties, a practical evaluation of its robustness is possible in practice. This leads to the conclusion that the regularity stands out as a robust pointwise features in 1D as well as in 2D. As an application, we use it to extract features that are indicators of potential fraud, through the processing of trade data.

LB-NA-25489-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.

ISBN 978-92-79-26261-6

9 789279 262616

Publications Office