Appl. Comput. Harmon. Anal. 31 (2011) 392–409

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Curve denoising by multiscale singularity detection and geometric shrinkage

Matt Feiszli *, Peter W. Jones

Yale University, Department of Mathematics, 10 Hillhouse Ave, New Haven, CT, United States

ARTICLE INFO

Article history:
Received 9 July 2010
Revised 22 January 2011
Accepted 5 February 2011
Available online 18 February 2011
Communicated by Charles K. Chui

Keywords:
Curve denoising
Shrinkage
Wavelet
Geometric statistics
Multiscale

ABSTRACT

We propose a method for denoising piecewise smooth curves, given a number of noisy sample points. Using geometric variants of wavelet shrinkage methods, our algorithm preserves corners while enforcing that the smoothed arcs lie in an L^2 Sobolev space H^α of order α chosen by the operator. The reconstruction is scale-invariant when using the Sobolev space $H^{3/2}$, adapts to the local noise level, and is essentially free of tuning parameters. In particular, our noise-adaptivity ensures that there is no arbitrarily-chosen "diffusion time" parameter in the denoising. Further, in cases where the distinction between signal and noise is unclear, we show how statistics gathered from the curve can be used to identify a finite number of "good" choices for the denoising.

1. Introduction

Given a number of noisy sample points taken along a curve, we seek to reconstruct the original curve in the best way possible. This problem, in various forms, has been studied extensively in the literature; we describe some previous work below. In this paper we pose the problem as a statistical estimation problem: we formulate a model for curves where each point on the curve is either a corner or it lies on a smooth arc, and we construct a near-optimal estimator for the original locations of the sample points given our model.

In general, there are many, many approaches to the denoising problem. Curvature-based diffusion methods have been used extensively in the literature for deblurring and denoising images; see, for example, [14,16,17]. These methods are closely related to the smoothing functionals we describe later; like these methods, our functionals also seek to minimize high-frequency oscillations, while still preserving some kind of fidelity to the original data. (In fact, the geometric heat equation [7] can be derived as a limiting case of a curve evolution which seeks to drive our multiscale Sobolev-type norms on curves to zero.) A disadvantage, for certain applications, of these methods is that naive application of curvature-based methods will fail to preserve singular points like corners and cusps. In this work we add a shock detector which works along the curve in order to identify and preserve these features. There are of course existing methods for identifying shocks. Shock-preserving methods like geometric essentially non-oscillatory (GENO) polynomials [18] will preserve corners, but it is difficult to make strong statements about the local regularity of the piecewise polynomial approximations.

Our methods also bear similarity to the curvelet [12] and beamlet [5] approaches of Donoho, Starck, et al. for image processing. These both provide multiscale representations suited for analysis of plane curves; the beamlet approach in

* Corresponding author.
E-mail address: mattfeiszli@gmail.com (M. Feiszli).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2011.02.001

particular is a multiscale method for detecting features in images. Our work here differs in that we work in one dimension, that is, on the curve itself, using a wavelet-like representation of local oscillation adapted for use on locally Lipschitz curves, and we assume that the curve has already been extracted from some image or dataset. In this way our work can be viewed as a parsing stage occurring after feature extraction has been performed.

1.1. Overview

The remainder of the paper is organized as follows. In Section 2 we collect necessary background material and describe the tools we use to perform multiscale analysis of plane curves. In Section 3 we describe our proposed methods for detecting singularities and denoising piecewise smooth curves. In Section 4 we discuss numerical solution of the problem and provide experimental results.

2. Background

Here we provide the relevant theorems and background material.

2.1. Characterizations of differentiability

The constructions here are motivated by the following theorem (see Stein [19]):

Theorem 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be in $L^2 \cap C^1$. Then*

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|f(x+t)+f(x-t)-2f(x)|^2}{|t|^2}\frac{dt}{|t|^n}\,dx = c_n \int_{\mathbb{R}^n} \sum_{k=1}^{n} \left|\frac{\partial f}{\partial x_k}\right|^2 dx$$

where c_n depends only on the dimension.

We remark that changing the power of $|t|$ reweights the various scales and will increase or decrease the number of derivatives being taken. For $0 < \alpha < 2$ we have that the quantity

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{|f(x+t)+f(x-t)-2f(x)|^2}{|t|^{2\alpha}}\frac{dt}{|t|^n}\,dx$$

is a multiple of the squared L^2 norm of the α-th derivatives of f.

These integrals also yield a local statement of regularity:

Theorem 2. *For almost every $x \in \mathbb{R}^n$, the function $f : \mathbb{R}^n \to \mathbb{R}$ has a weak derivative in the L^2 sense at x iff*

$$\int_{\mathbb{R}^n} \frac{|f(x+t)+f(x-t)-2f(x)|^2}{|t|^2}\frac{dt}{|t|^n} < \infty$$

In other words, this integral in t controls the regularity of f about x. Again, if we introduce the parameter α we get corresponding statements for the α-th derivatives.

2.2. β angles and Sobolev norms

Here we introduce versions of the previous results more suited to analysis of curves. Let $\gamma(s)$ be an arc-length[1] parametrization of a smooth curve $\Gamma \subset \mathbb{R}^2$ and let $\{z_n\}_{i=1}^N$ be a collection of sample points taken at equal arc-length along Γ (we are not yet assuming any noise in the samples). Given the parametrization $\gamma(s) = x(s) + iy(s)$, we define the angle $\beta(s,t)$ with respect to γ by

$$\beta(s,t) \equiv \arg \frac{\gamma(s+t)-\gamma(s)}{\gamma(s)-\gamma(s-t)}$$

[1] Arc-length sampling is not at all necessary; we assume it mostly for a concise presentation. The results can easily be modified if, for example, the sampling intervals are bounded above and below.

We introduce the α "norm"[2] for curves, defined for closed curves as

$$\|\gamma\|_\alpha^2 = \int_0^L \int_0^{L/2} \beta(s,t)^2 \frac{dt}{|t|^{2\alpha-1}}\, ds \tag{1}$$

where L is the length of Γ, and for open curves as

$$\|\gamma\|_\alpha^2 = \int_0^L \int_0^{\min(s,L-s)} \beta(s,t)^2 \frac{dt}{|t|^{2\alpha-1}}\, ds \tag{2}$$

In making computations on discretized closed curves we will use the equivalent dyadic variant

$$\|\gamma\|_\alpha^2 = \sum_{n=1}^{N} \sum_{k=1}^{K} \beta(n,k)^2 2^{-2k(1-\alpha)} \tag{3}$$

(although we truncate the inner sum near the endpoints for open curves), where

$$\beta(n,k) = \arg \frac{\gamma_{n+2^{k-1}} - \gamma_n}{\gamma_n - \gamma_{n-2^{k-1}}}$$

These constructions give geometric variants of Theorem 1; the β's are to curves what the second differences are to functions. The next section gives these results in the dyadic case.

Remark. When passing from an integral like (2) to a dyadic sum like (3), it is helpful to keep the following intuition in mind. Consider an integral of the form

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} f(s,t)^2 \frac{dt\, ds}{t^2}$$

In the discrete setting, using point spacing of Δs in the s coordinate and Δt in the t coordinate, we might approximate this integral as

$$\sum_{n\in\mathbb{Z}} \sum_{k=1}^{\infty} f_{n,k}^2 \frac{\Delta t \Delta s}{(k\Delta t)^2}$$

where $f_{n,k}$ is, for example, the value

$$f_{n,k} = f(n\Delta s, k\Delta t)$$

An integral with respect to the measure dt/t puts equal weight on each octave. If we wish to use only dyadic scales 2^k, we have only one sample per octave in scale, and we must increase the weight per octave. Hence, when using dyadic scales, the equivalent sum is obtained by dropping one factor of Δt in the denominator, to achieve equal weight on each scale; i.e. we wish to use the sum

$$\sum_{n\in\mathbb{Z}} \sum_{k=1}^{\infty} f_{n,k}^2 \frac{\Delta t \Delta s}{2^k \Delta t}$$

In our work here the factors Δt and Δs are unimportant constants which we omit. One may verify that this produces the sum (3) from (2).

2.2.1. The mapping from β's to wavelet coefficients

Consider $\Gamma \subset \mathbb{C}$ which is the graph of a Lipschitz function; i.e. $\Gamma = \{x + iA(x)\} \subset \mathbb{C}$ where $A : \mathbb{R} \to \mathbb{R}$ satisfies $\|A'\|_\infty < M$. If $\gamma : \mathbb{R} \to \mathbb{C}$ is an arc-length parametrization of Γ we define $\phi : \mathbb{R} \to [-\arctan M, \arctan M]$ a.e. by writing

$$\gamma'(s) = e^{i\phi(s)}$$

[2] It is not correct to refer to this as a norm on curves; we have not imposed a linear structure on the space of curves. However, these integrals of β angles play a role analogous to the Sobolev norms on function spaces and the β's themselves share many features with wavelet coefficients from the point of view of analysis. We hope the reader will forgive this abuse of terminology.

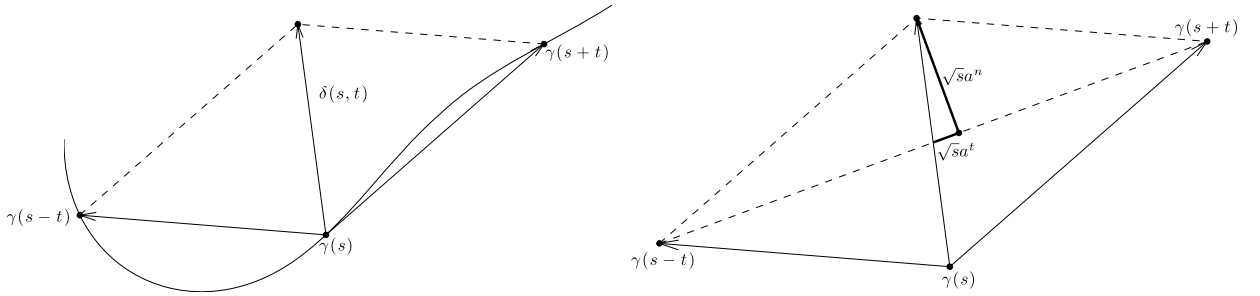

Fig. 1. Wavelet coefficients of complex-valued γ' in the continuous setting. Left: The wavelet coefficients $a(s,t)$ are, up to the L^2 rescaling, given by the vector $\delta(s,t)$ joining $\gamma(s)$ to the opposite vertex of the parallelgram. Right: Again up to rescaling, the "normal" and "tangential" parts a^n, a^t are given by the projections of δ onto the line ℓ joining $\gamma(s-t)$ to $\gamma(s+t)$.

Let $\mathcal{D}$ be the set of dyadic intervals of $\mathbb{R}$. Given $I \in \mathcal{D}$, we write x_I for the midpoint of I and x_I^+, x_I^- for the right and left endpoints, respectively. Denote by h_I the Haar wavelet associated with I:

$$h_I(x) = \begin{cases} |I|^{-1/2}, & x \in [x_I, x_I^+] \\ -|I|^{-1/2}, & x \in [x_I^-, x_I) \end{cases}$$

We will generally write x_I for intervals in the domain of a function and s_I for intervals in the domain of a curve. Given our curve $\gamma(s)$, we may consider the dyadic β angles

$$\beta_I \gamma = \arg \frac{\gamma(s_I^+) - \gamma(s_I)}{\gamma(s_I) - \gamma(s_I^-)}$$

as well as the Haar coefficients of γ'

$$a_I = \langle \gamma', h_I \rangle$$

We see that in the sense of distributions, the Haar coefficients of γ' in fact agree with the second differences of γ, up to a rescaling:

$$\begin{aligned} a_I &= \langle \gamma', h_I \rangle \\ &= |I|^{-1/2} \big(\gamma(s_I^+) + \gamma(s_I^-) - 2\gamma(s_I) \big) \end{aligned}$$

Geometrically, a_I is a vector joining $\gamma(s_I)$ to the opposing vertex of the parallelogram defined by the three points $(\gamma(s_I^+), \gamma(s_I), \gamma(s_I^-))$. The next theorem shows that on Lipschitz graphs, the β's are almost the wavelet coefficients a_I. While this is pointwise false, it becomes true after integrating. We first need a lemma. Decompose a_I into "tangential" and "normal" components

$$a_I = a_I^t + a_I^n$$

where

$$\begin{aligned} a_I^t &= \operatorname{proj}_{\ell_I} a_I \\ a_I^n &= \operatorname{proj}_{\ell_I^\perp} a_I \end{aligned}$$

and ℓ is the line joining $\gamma(s_I^-)$ to $\gamma(s_I^+)$ (see Fig. 1). We think of ℓ as a rough tangent line, thus the terminology. We will use the fact that $|a_I|^2 = |a_I^t|^2 + |a_I^n|^2$. Our next lemma says that the β's and the normal components a_I^n agree up to a rescaling.

Lemma 1 *(β is a_I^n). With β, γ, ϕ, a_I defined as above,*

$$\left| a_I^n \right| \sim |\beta_I| |I|^{1/2}$$

with constant that depends only on M.

Proof. Let $x_I^\pm = \operatorname{Re} \gamma(s_I^\pm)$. Since $\{(x, A(x))\}$ is a Lipschitz graph, we have $ds \sim dx$ and hence there exist $k, m > 0$, depending only on M, such that $\gamma(s_I)$ is located in a rectangle $\{(x,y)\colon x_I^- + m|I| < x < x_I^+ - m|I|, -k|I| < y < k|I|\}$. Now compose with a rotation R so $R(\gamma(s_I^\pm))$ lie on the real axis. Observe that $R(\gamma(s_I))$ lies in a tilted rectangle whose boundary makes an

angle with the real axis which is bounded below by a constant depending on M. Elementary geometry shows that $|I|\beta_I$ is comparable to the distance of $R(\gamma(s_I))$ from the real axis. However, this distance is the normal part of the second difference of γ on I and hence equal to $|I|^{1/2}|a_I^n|$, and the conclusion follows. □

In fact, under certain conditions, we can use β in place of the entire coefficient a_I. Pointwise, this is false, but on Lipschitz graphs, a_I^n controls a_I^t after integrating, when α is large enough. The next theorem makes this precise. Let $\mathcal{I}$ be a dyadic partition of the domain of A.

Theorem 3 *(β's are almost wavelet coefficients). For $0 < \alpha < 1$,*

$$\sum_{I\in\mathcal{I}} |a_I|^2 |I|^{-2\alpha} \sim \sum_{I\in\mathcal{I}} \beta_I^2 |I|^{1-2\alpha}$$

with constant that depends only on the Lipschitz constant M.

Proof. The lower bound is trivial. By Lemma 1, we have

$$\begin{aligned} |a_I|^2 &= \left|a_I^t\right|^2 + \left|a_I^n\right|^2 \\ &\geqslant \left|a_I^n\right|^2 \\ &\geqslant C_M \beta_I^2 |I| \end{aligned}$$

Similarly, the β's give an upper bound on the sum of the a_I^n. Hence, we must show that there is C_M so that

$$\sum_{I\in\mathcal{I}} \left|a_I^t\right|^2 |I|^{-2\alpha} \leqslant C_M \sum_{I\in\mathcal{I}} \beta_I^2 |I|^{1-2\alpha}$$

Let p be the projection of $\gamma(s_I)$ onto the line passing through $\gamma(s_I^{\pm})$ and set

$$\begin{aligned} r_{\pm} &= \left|\gamma\left(s_I^{\pm}\right) - p\right| \\ q_{\pm} &= \left|\gamma\left(s_I^{\pm}\right) - \gamma(s_I)\right| \end{aligned}$$

We claim that $r_- - r_+$ and $q_- - q_+$ are comparable. Indeed, by the Pythagorean theorem applied to the triangles $(\gamma(s_I^-), p, \gamma(s_I))$ and $(\gamma(s_I^+), p, \gamma(s_I))$, one finds

$$r_-^2 - q_-^2 = r_+^2 - q_+^2$$

which gives us

$$\frac{r_- - r_+}{q_- - q_+} = \frac{q_- + q_+}{r_- + r_+}$$

The right-hand side is bounded below by 1 and bounded above by some constant depending on M, so we conclude that

$$\frac{1}{c} < \frac{r_- - r_+}{q_- - q_+} < c$$

as claimed. We now proceed to estimate the tangential parts of our wavelet coefficients.

$$\begin{aligned} \left|a_I^t\right| &= |I|^{-1/2} |r_- - r_+| \\ &\leqslant c|I|^{-1/2} |q_- - q_+| \\ &= c|I|^{-1/2} \left|q_- - |I|/2 + |I|/2 - q_-\right| \\ &\leqslant c|I|^{-1/2} \left(\left(|I|/2 - q_-\right) + \left(|I|/2 - q_+\right)\right) \end{aligned} \tag{4}$$

The next step is essentially the Lipschitz graph case of the second author's traveling salesman theorem; we use the β's to control the length of arcs of Γ. For any $J \in \mathcal{I}$, let Γ_J be the corresponding arc of Γ and let ℓ_J be the line segment joining $\gamma(s_J^-)$ to $\gamma(s_J^+)$. We will consider a succession of piecewise linear approximations to Γ_I. Let $\mathcal{F}_n$ be the collection of generation-n dyadic children of I. Define the n-th generation approximation

$$\Gamma_n = \bigcup_{J\in\mathcal{F}_n} \ell_J$$

with Γ_0 being the line segment ℓ_I of length $L(\Gamma_0) = r_+ + r_-$. Then by the Pythagorean theorem,

$$L(\Gamma_1) = \sqrt{r_+^2 + h^2} + \sqrt{r_-^2 + h^2}$$

where $h = |\gamma_{s_I} - p|$. Continuing,

$$\begin{aligned} L(\Gamma_1) &= r_+\sqrt{1 + \left(\frac{h}{r_+}\right)^2} + r_-\sqrt{1 + \left(\frac{h}{r_-}\right)^2} \\ &\leqslant r_+\left(1 + \left(\frac{h}{r_+}\right)^2\right) + r_-\left(1 + \left(\frac{h}{r_-}\right)^2\right) \\ &\leqslant r_+ + r_- + c\left(r_+\beta_I^2 + r_-\beta_I^2\right) \\ &\leqslant L(\Gamma_0) + c|I|\beta^2 \end{aligned}$$

where we used the facts that on Lipschitz graphs, $r_\pm \sim |I|$ and also that $h/r_\pm$ is bounded and comparable to β. Applying this recursively we see

$$L(\Gamma_{n+1}) - L(\Gamma_n) \leqslant c \sum_{J \in \mathcal{F}_n} |J|\beta_J^2$$

and hence by summation,

$$L(\Gamma_I) \leqslant L(\Gamma_0) + c \sum_{J \subset I} |J|\beta_J^2 \tag{5}$$

Applying (5) to the left and right dyadic children $I_\pm$ of I gives

$$|I|/2 \leqslant q_\pm + c \sum_{I' \subset I_\pm} |I'|\beta_{I'}^2$$

and substitution into (4) yields

$$\left|a_I^t\right| \leqslant c|I|^{-1/2} \sum_{J \subset I} |J|\beta_J^2 \tag{6}$$

A similar argument shows that there is c so that

$$\sum_{I' \subset I} |I'|\beta_{I'}^2 \leqslant c|I| \tag{7}$$

again using the fact that Γ is a Lipschitz graph. The next and final step essentially says that points on Γ cannot be brought too near each other unless Γ is wrinkly. Applying our estimates above, we see

$$\begin{aligned} \sum_{I \subset \mathcal{I}} \left|a_I^t\right|^2 |I|^{-2\alpha} &\leqslant c \sum_{I \subset \mathcal{I}} |I|^{-1} \left(\sum_{J \subset I} |J|\beta_J^2\right)^2 |I|^{-2\alpha} \quad \text{by (6)} \\ &\leqslant c \sum_{I \subset \mathcal{I}} \sum_{J \subset I} |J|\beta_J^2 |I|^{-2\alpha} \quad \text{by (7)} \\ &= c \sum_{J \subset \mathcal{I}} |J|\beta_J^2 \sum_{I \supset J} |I|^{-2\alpha} \\ &\leqslant c \sum_{J \subset \mathcal{I}} |J|\beta_J^2 |J|^{-2\alpha} \sum_{k=1}^{\infty} 2^{-2k\alpha} \\ &\leqslant c \sum_{J \subset \mathcal{I}} |J|^{1-2\alpha} \beta_J^2 \end{aligned}$$

where the geometric series in the second-to-last line converges for $\alpha > 0$. This completes the proof. □

Remark. The condition $\alpha > 0$ in Theorem 3 is necessary. Consider the curve of length 2 obtained by joining the line segment $[-1, 0] \subset \mathbb{R}$ to 2^k triangle functions with sides of slope $\pm M$, as shown in Fig. 2. Let γ be an arc-length parametrization of this curve. Note that the rightmost endpoint of the curve lies at $(\frac{1}{\sqrt{1+M^2}}, 0)$, independent of k, so the tangential part of the largest a_I is non-zero. If $\mathcal{I}$ is a dyadic partition of $[-1, 1]$ we see there are 2^k non-zero β_I's which share some common value β. We have

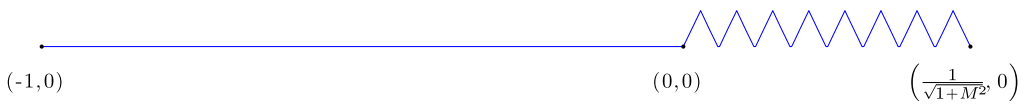

Fig. 2. The condition $\alpha > 0$ is necessary in Theorem 3. For $\alpha < 0$, the cost of the wiggly part vanishes as the number of wiggles increases.

$$\sum_{I\in\mathcal{I}} \beta_I^2 |I|^{1-2\alpha} = \beta^2 2^k \left(2^{-k}\right)^{1-2\alpha}$$
$$= \beta^2 2^{2\alpha k}$$

If $\alpha < 0$ we can make this arbitrarily small by increasing k, so no variant of Theorem 3 can hold.

In light of Theorem 3 it follows that statements about local regularity of $\gamma(s)$ which can be obtained via decay of wavelet coefficients can be translated immediately into an equivalent statement about the β's. For example, Theorem 3 has an immediate corollary:

Corollary 4 *(Characterization of Sobolev spaces via β). For $0 < s < 1$, $\gamma \in H^{1+s}$ iff*

$$\sum_{I\in\mathcal{I}} \beta_I^2 |I|^{1-2\alpha} \leqslant \infty$$

Proof. By Theorem 3, the fractional derivative seminorms induced by the a_I and the β's are comparable, and γ is always in L^2 by construction, hence this is simply a restatement of the definition of H^{1+s}. □

These results are more-or-less implicit in the results of [10,11]. We plan to expand on and extend these techniques in an upcoming paper.

In fact, more is true. The mapping from one wavelet basis to another, and hence the mapping from the β's to essentially any "reasonable" wavelet basis, is "almost diagonal" in the following sense. Choose two mother wavelets $\psi, \widetilde{\psi} \in L^1 \cap L^\infty$. Then given some $f \in L^2$, we may expand f in either basis

$$f = \sum c_I \psi_I$$
$$= \sum \widetilde{c}_I \widetilde{\psi}_I$$

and one finds immediately that the coefficients are related by

$$\widetilde{c}_I = \sum_{J\in\mathcal{D}} c_J \langle \psi_J, \widetilde{\psi}_I \rangle$$

If we fix a scale $|J| = |I|$, the decay as we translate J away from I can be made as rapid as one wishes by choosing mother wavelets with rapid decay. The next result demonstrates that the inner product always decays exponentially in scale, given only our modest assumptions on the mother wavelets. It is in this sense (rapid decay of the off-diagonal terms in the change-of-basis matrix) that we mean the change of basis is an almost diagonal transformation.

Theorem 5 *(Change of wavelet basis is almost diagonal). For any $\psi, \widetilde{\psi} \in L^1 \cap L^\infty$, we have the following*:

(1) *Let $I, J \in \mathcal{D}$ with $|I| = 2^{-k}$, $|J| = 2^{-\tilde{k}}$. Then*

$$\langle \psi_I, \widetilde{\psi}_J \rangle \leqslant C_1 2^{-|k-\tilde{k}|/2}$$

(2) *Assume further that $|\psi(x)|, |\widetilde{\psi}(x)| < M(x)$ for some symmetric $M : \mathbb{R} \to \mathbb{R}_+$, decreasing for $x \geqslant 0$. If $|I| = |J|$, then*

$$\langle \psi_I, \widetilde{\psi}_J \rangle \leqslant C_2 \int_{|x_I - x_J|/(2|J|)}^{\infty} M(x)\, dx$$

where C_1, C_2 each depend on both ψ and $\widetilde{\psi}$. In other words, the off-diagonal terms always decay exponentially in scale, and the translates may be made to decay as rapidly as one wishes.

Proof. We prove part (1) first. Consider $|J| \leqslant |I|$:

$$\begin{aligned}\langle \psi_I, \widetilde{\psi}_J \rangle &= \frac{1}{|J|^{1/2}|I|^{1/2}} \int \psi\left(\frac{x-x_I}{|I|}\right)\widetilde{\psi}\left(\frac{x-x_J}{|J|}\right)dx \\ &\leqslant \frac{\|\psi\|_{L^\infty}}{|J|^{1/2}|I|^{1/2}} \int \widetilde{\psi}\left(\frac{x-x_J}{|J|}\right)dx \\ &\leqslant \left(\frac{|J|}{|I|}\right)^{1/2} \|\psi\|_{L^\infty}\|\widetilde{\psi}\|_{L^1}\end{aligned}$$

Exchanging the roles of ψ_I and $\widetilde{\psi}_J$ in the above, one sees

$$\langle \psi_I, \widetilde{\psi}_J \rangle \leqslant \left(\frac{|I|}{|J|}\right)^{1/2} \|\widetilde{\psi}\|_{L^\infty}\|\psi\|_{L^1}$$

Taking $C_1 = \min(\|\psi\|_{L^\infty}\|\widetilde{\psi}\|_{L^1}, \|\widetilde{\psi}\|_{L^\infty}\|\psi\|_{L^1})$ proves (1). We now prove (2).

$$\begin{aligned}\langle \psi_I, \widetilde{\psi}_J \rangle &= \frac{1}{|J|}\int \psi\left(\frac{x-x_I}{|J|}\right)\widetilde{\psi}\left(\frac{x-x_J}{|J|}\right)dx \\ &= \int \psi\left(y-\frac{x_I}{|J|}\right)\widetilde{\psi}\left(y-\frac{x_J}{|J|}\right)dy \\ &\leqslant \int M\left(y-\frac{x_I}{|J|}\right)M\left(y-\frac{x_J}{|J|}\right)dy \\ &\leqslant 2\|M\|_\infty \int\limits_{|x_I-x_J|/2|J|}^{\infty} M(y)\,dy\end{aligned}$$

This completes the proof. □

In summary, studying the β angles and their decay at small scales is equivalent to studying the wavelet coefficients of the derivative γ'. Any statements about regularity obtained using one construction translates immediately into a corresponding statement about the other.

3. Denoising: Feature detection, local scales and near-optimal estimation

In this section we present the method for feature detection and denoising. We begin by discussing our noise model. Then we discuss feature detection and how to break a piecewise smooth curve into its smooth subarcs. Finally derive a near-optimal method for denoising the smooth arcs, and then describe the resulting algorithm. Our strategy is the following:

(1) Define a set of singular features we wish to preserve.
(2) Determine what these features look like in the β-angle representation.
(3) Apply statistical testing to the β angles to detect the features.
(4) Reconstruct a curve which is near the original in a least-squares sense, preserves singularities, and is as smooth as possible away from singular points.

3.1. Noise on piecewise smooth curves

In this section we describe our model for curves and describe methods for singularity detection and denoising piecewise smooth curves. We first discuss our model for noise, and justify some approximations we will use in the remaining sections. We relate our techniques to wavelet shrinkage methods for function estimation and then derive our proposed algorithms.

Consider a curve $\Gamma \subset \mathbb{R}^2$ which is a finite union of smooth (say, at least C^2) arcs. Let $\gamma(s)$ be an arc-length parametrization. Assume we are given a number of sample points $\{z_n\}_{n=1}^N$

$$z(n) = \gamma(s_n) + \eta(n)$$

where

$$s_n = \frac{n-1}{N}\,\mathrm{Length}(\Gamma)$$

and $\eta(n) \in \mathbb{R}^2$ is random noise. We assume that $\eta(n)$ is given by

$$\eta(n) = v\big(\gamma(s_n)\big)$$

where $v : \mathbb{R}^2 \to \mathbb{R}^2$ is a Gaussian random vector field whose covariance is such that the expected value of the increments locally satisfies

$$E\big\|v(z) - v(w)\big\|^2 = O\big(\|z - w\|^p\big) \tag{8}$$

for some exponent $p \in \mathbb{R}$. For example, $p = 0$ amounts to assuming our measurements have been corrupted by Gaussian white noise, while $p = 1$ produces a field which is continuous almost everywhere, but not differentiable. In the case $p = 2$ the field admits one derivative in the mean-square sense almost everywhere.

We wish to understand how this noise looks after we perform our multiresolution analysis.

3.1.1. β angles

In this section we work in the continuous setting and study the noise as it affects the angles $\beta(s,t)$. We show that while the noise is certainly not Gaussian, it tends to Gaussian when the noise is small relative to the scale under observation. Given our noise model, it is straightforward to compute the difference $\Delta\beta(s,t)$ between the observed and true β's. About any point $\gamma(s)$, consider the vector field

$$w_s(t) = v\big(\gamma(s+t)\big) - v\big(\gamma(s)\big)$$

Then we have the following:

Proposition 1. *At any fixed s and for $t > 0$, we have*

$$\Delta\beta(s,t) = \frac{1}{|\gamma(s+t) - \gamma(s)|}\,\mathrm{proj}_{\hat{\tau}}\, w_s(t) + o\big(\big\|w_s(t)\big\|\big) \tag{9}$$

where $\mathrm{proj}_{\hat{\tau}}\, w_s(t)$ is the scalar projection onto $\hat{\tau}$, a unit vector orthogonal to the line joining $\gamma(s)$ to $\gamma(s+t)$.

Proof. Elementary geometry shows that $\Delta\beta$ is related to the noise vector field by

$$\Delta\beta(s,t) = \arctan\left(\frac{\mathrm{proj}_{\hat{\tau}}\, w_s(t)}{|\gamma(s+t) - \gamma(s)| + \mathrm{proj}_{\hat{n}}\, w_s(t)}\right)$$

where $\mathrm{proj}_{\hat{n}}\, w$ and $\mathrm{proj}_{\hat{\tau}}\, w$ are the projections of the vector $w_s(t)$ onto the unit vector parallel to $z(s+t) - z(s)$ and onto the orthogonal vector, respectively. In other words, $\tan\Delta\beta$ obeys a Gaussian ratio distribution. We regard the right-hand side as a function $g(a,b,c) = \arctan(\frac{a}{b+c})$ which has gradient

$$\nabla g(a,b,c) = \left(\frac{1}{b+c}, -\frac{a}{(b+c)^2}, -\frac{a}{(b+c)^2}\right)$$

The proposition statement is thus the first-order Taylor series for g about the point $(0, |\gamma(s+t) - \gamma(s)|, 0)$. □

The scalar projection $\mathrm{proj}_{\hat{\tau}}\, w$, as a convex combination of Gaussians, is again Gaussian. Hence, the errors $\Delta\beta$ will follow a Gaussian distribution to first order. Whether this is a reasonable approximation in practice, of course, depends on the situation. We remark that the true distribution of $\tan\Delta\beta$ is a Gaussian ratio distribution, which tends to Cauchy in the large-noise case, in which case $\Delta\beta$ tends to a uniform distribution on $[-\pi, \pi]$. From the point of view of denoising, this is most likely either an unreasonable assumption or a hopeless situation; hence we feel the small-noise approximation is likely to be more relevant in applications. That said, we recognize that there may be some benefit in assuming heavier tails, particularly at the smallest scales where the noise may appear larger. We do not pursue this in the work below.

We will use one further approximation to the noise in the β's: we assume that the variance is not $O(\|z(s) - z(s+t)\|^p)$, but rather that it is $O(|t|^p)$. This is simply another first-order approximation; point separation is locally comparable to arc-length on any smooth arc.

3.2. Singularity detection and local scales using β angles

As above, we consider a curve $\Gamma \subset \mathbb{R}^2$ which is a finite union of smooth (at least C^2) arcs which meet at points of singularity. Here, the only singularity we attempt to detect is a corner, i.e. a pair of lines meeting at an angle. However, it should be clear how to extend the algorithm described here to handle additional types of singularities, e.g. cusps, or corners formed by a pair of circular arcs.

As before, we are given samples $z(n)$ from a curve Γ, corrupted by Gaussian noise. We will write $b(n,m)$ for the full collection of observed angles (not just those at dyadic scales):

$$b(n,m) = \arg\frac{z(n+m) - z(n)}{z(n) - z(n-m)}$$

Using the full collection of angles increases power in our statistical testing. The angles $b(n,m)$ contain noisy information about the local geometry of Γ about each point. We wish to use the b's to classify each point as either a corner or a point on a smooth arc. We define corners to be the points where the true angles β locally look constant:

Definition 1. We say the point $\gamma(s)$ is a *corner point* if

$$\beta(s,t) = \theta + o(t)$$

for $\theta \in (-\pi, \pi)$.

On the other hand, if $\gamma(s)$ lies along a smooth arc, then we will have the following:

Proposition 2 *(β increases linearly on circular arcs). If $\gamma(s)$ is locally the graph of a C^2 function on some neighborhood of s, then*

$$\beta(s,t) = t/\kappa + o(t)$$

where κ is the Euclidean curvature of γ at s.

Proof. We remark that this follows directly from elementary geometry in the case that γ is in fact locally a circular arc. In the general case, let $f(x)$ be a function which describes γ locally, with $x = x(t)$ so that $x = 0$ corresponds to $t = 0$. Then this is a direct calculation using

$$\beta(s,t) = \arctan \frac{f(x(t))}{x(t)}$$

combined with

$$\frac{dx}{dt} = \frac{1}{\sqrt{1 + (f'(x(t)))^2}}$$

and the usual formulas for curvatures of plane curves. □

Thus our task is to decide, at various scales and after accounting for noise, whether β at each point looks more like a constant or more like a straight line decaying to zero. We expect that at different scales the answer will be different; for example, a straight line corrupted by noise may locally appear to have corners but at large scales will still be well-approximated by a straight line. Conversely, if you round off the corners of a polygon, they will only look smooth at small scales, whereas at large scales the β's will still look constant.

3.2.1. Detecting corners

Our ability to discriminate corners from arcs depends on the local level of noise near $\gamma(s)$. Using our small-noise model (9) we now derive a likelihood-ratio test for identifying corners against the default hypothesis that the curve is locally a smooth arc. If we assume that the local variance of the noise is known at each scale, then this is the most powerful test for identifying corners at a given significance level.

First, choose minimum and maximum lengths of interest:

$$0 < M_{\min} < M_{\max} < \lfloor N/2 \rfloor$$

About each point $z(n)$, for each $M \in [M_{\min}, M_{\min}+1, M_{\min}+2, \ldots, M_{\max}]$ we solve two weighted least-squares problems to approximate $b(n,m)$ as a function of m on the set $[0,1,2,\ldots,M]$. Given a choice of M, we look for the best linear approximation

$$L_M(n,m) = c(n,M)m$$

for $c(n,M) \in \mathbb{R}$ and also the best constant approximation $\mu_M(n)$. That is, we seek $L_M(n,m)$ and $\mu_M(n)$ which minimize the squared residuals

$$R_L^2(n,M) = \sum_0^M \left| L_M(n,m) - b(n,m) \right|^2 m^{2-p}$$

$$R_\mu^2(n,M) = \sum_0^M \left| \mu_M(n) - b(n,m) \right|^2 m^{2-p}$$

We see that the least-squares weights m^{2-p} are inversely proportional to the variance of $\Delta\beta$. Now given a threshold value δ, apply the decision rule that declares "$z(n)$ is a corner at scale M" when

$$R_L^2 - R_\mu^2 > \delta \tag{10}$$

This test has the following justification given our noise model:

Proposition 3 *(Likelihood ratio test). The decision rule which says $z(n)$ is a corner when*

$$R_L(n,m)^2 - R_\mu(n,m)^2 > \delta \tag{11}$$

is a likelihood ratio test for corners against the default hypothesis, when m is less than the inter-corner distance along z.

Proof. Let $p(\beta)$ and $q(\beta)$ be the likelihoods of the observed β under the assumptions that $\gamma(s_n)$ is a corner point or a point of smoothness, respectively. According to our model the errors $\Delta\beta$ are Gaussian. In this case, we have

$$p(\beta) = k\exp\{-R_\mu^2\}$$
$$q(\beta) = k\exp\{-R_L^2\}$$

where the constant $k \in \mathbb{R}$ is the same in both cases since we assume the noise is $O(|t|^p)$ and it is additive with identical variance in either circumstance. This in turn means that

$$\log\frac{p(\beta)}{q(\beta)} = R_L^2 - R_\mu^2$$

and the conclusion follows. □

In practice, the choice of threshold

$$R_L^2 - R_\mu^2 > \hat{\sigma}(n,m)^2$$

works very well in virtually every case we have tried, where $\hat{\sigma}(n,m)^2$ is an estimate for the local variance of the noise at scale m. We estimate $\sigma(n,m)$ using the heuristic that most points lie on smooth arcs and thus averaging the residuals R_L^2 at points in a neighborhood of $z(n)$ should give a good estimate for the level of local noise. That is, we compute

$$\hat{\sigma}^2(n,m) = \frac{1}{2m}\left(\sum_{j=n+m}^{n+2m} R_L^2(j,m) + \sum_{j=n-2m}^{n-m} R_L^2(j,m)\right)$$

We avoid averaging over $z(n)$ in order to reduce errors at corner points.

3.2.2. Local scales and local noise

The "local scale" associated with a point on an object is, loosely, the "size" at which different features appear on the object. That is, if a curve locally looks like a smooth arc of a circle for a certain length L, we might say the curve is locally a circle at scale L, and call L the local scale. Associated with the local scale L is some measure of local noise, which is deviation from circularity on the arc of length L. The difference between this estimate for local noise and the method used in the preceding section is that we now have more information about the point; we can now use the corner residuals or the smooth residuals as appropriate, whereas above we did not yet know which was a better choice.

Finding and smoothing local scales on images has been discussed in [9]; a thorough development of methods for detecting and smoothing local scales on curves is a substantial topic on its own and is a subject for a separate paper. Here we make the above loose definitions precise by using goodness-of-fit measures to identify scales at which a feature, i.e. a circular arc or a corner, is most likely present. We suggest the following simple method based on the R-squared statistic:

(1) For each $n \in 1, 2, 3, \ldots, N$, examine the appropriate residuals $R_L^2(n,m)$ or $R_\mu^2(n,m)$ to find the scales at which the R-squared statistic achieves a local maximum. Do not consider large scales M where the arcs $[z(n-M), z(n)]$ or $[z(n), z(n+M)]$ cross corners. Record a list of features detected along with the length scales at which they were detected.
(2) For each point $z(n)$, do the following:
   (a) If $z(n)$ lies on a smooth arc, set the local scale $K_{loc}(n)$ to be the scale of the longest smooth arc detected in step 1 which contains $z(n)$.
   (b) If $z(n)$ is a corner point, set the local scale $K_{loc}(n)$ to be the scale of the largest corner centered at $z(n)$ detected in step 1.

Note that the largest possible scale that can be associated with a smooth point is the length of the smooth arc of Γ on which it lies, *not* the distance to the nearest corner. This allows for the removal of large-scale noise near corners, which is impossible if the local scale is limited by distance to the nearest corner.

Once the local scales $K_{loc}(n)$ have been chosen, we need to estimate the local variance of the noise. We use the fact that for small noise, displacement of points is $O(t\Delta\beta(s,t))$ (see the proof of Proposition 3). Hence we compute

$$\hat{\sigma}_z\big(n, K_{loc}(n)\big) = \sqrt{\frac{1}{K_{loc}(n)} \sum_{m=0}^{K_{loc}(n)} \big(m\Delta\beta(n,m)\big)^2}$$

3.3. Shrinkage methods and near-optimal estimation

We now apply methods of Donoho, Johnstone and Silverman (see [3,8]) to construct a shrinkage-type denoising algorithm for curves. The goal is to find an estimator $\hat{\gamma}(n)$ in order to recover the true locations $\gamma(s_n)$ of our noisy sample points $z(n)$. Note that we explicitly do not wish to use ℓ^2 error as a measure of risk; we wish to use a Sobolev-type measure of error in order to correctly capture the oscillations of the curve.

Given sampled curves η, γ with angles β_η, β_γ, define the distance

$$d_\alpha^2(\beta_\eta, \beta_\gamma) = \sum_{n=1}^{N}\sum_{k=1}^{K}\big(\beta_\eta(n,k) - \beta_\gamma(n,k)\big)^2 \frac{1}{2^{2k(\alpha-1)}}$$

We also need some ℓ^2 type measure of distance. Set

$$d_2^2(\eta, \gamma) = \sum_{n=1}^{N} \sigma^{-2}(n)\big\|\eta(n) - \gamma(n)\big\|^2$$

where the weights $\sigma^2(n)$ are operator chosen. We take them to be the variance of the noise near $\gamma(n)$. Finally, we define the risk for our estimator as

$$R(\hat{\gamma}, \gamma) = E\big(d_2^2(\hat{\gamma}, \gamma) + Cd_\alpha^2(\beta_{\hat{\gamma}}, \beta_\gamma)\big)$$

We view this as a Sobolev-type norm on curves, since d_2 is an ℓ^2-type measure of error in the point locations, and d_α is a measure of difference in the order-α regularity of the curves, using the β angles defined earlier. C is a constant chosen by the operator.

3.3.1. Regularity and β angles

As per our noise model described above, we assume that we are given observations $\beta_z(n,k)$ of the true β angles

$$\beta_z(n,k) = \beta_\gamma(n,k) + \epsilon_k\xi(n,k)$$

where the noises $\xi(n,k)$ are now (correlated) normal $\mathcal{N}(0,1)$ which are stationary in n (this occurs, for example, in the case $p=0$ of small white noise), and $\epsilon_k > 0$ is the noise level at each scale. We assume for the moment that the noise level is known. Consider the problem of constructing an estimator $\hat{\beta}$ for just the angles β_γ, regardless of whether $\hat{\beta}$ produces a collection of angles which could have arisen from a curve (in fact, the estimates described below will almost surely not arise from any curve). Begin by defining the hard and soft threshold operators

$$\eta_H(\beta, \lambda) = \beta\mathbb{1}_{[|\beta|>\lambda]}$$
$$\eta_S(\beta, \lambda) = \operatorname{sgn}(\beta)\big(|\beta| - \lambda\big)_+$$

It is shown in [8] that for fixed k, the estimator

$$\hat{\beta}(\cdot, k) = \eta_S\big(b(\cdot,k), \epsilon_k(2\log N)^{1/2}\big)$$

satisfies

$$E\big\|\hat{\beta}(\cdot,k) - \beta_\gamma(\cdot,k)\big\|_2^2 \leqslant (2\log N + 1)\left\{\epsilon_k^2 + \sum_{j=1}^{N}\min\big(\beta(n,k)^2, \epsilon_k^2\big)\right\}$$

where $\|\cdot\|_2$ is the ℓ^2 norm on N points (the same holds for the hard threshold η_H). Since ϵ_k^2 is the expected squared error for estimating a single parameter, this result says that the thresholding estimator is optimal in an ℓ^2 sense, up to essentially $2\log N$. Even better, this estimator is also optimal in the Sobolev sense we take as our risk. For example, it follows from [4] that the estimated $\hat{\beta}$'s look as "smooth" as the β_γ's of the original curve; i.e. with probability approaching 1 as $N \to \infty$, we have a constant C_1 so that

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\hat{\beta}(n,k)^2\frac{1}{2^{2k(\alpha-1)}} \leqslant C_1\sum_{n=1}^{N}\sum_{k=1}^{K}\beta_\gamma(n,k)^2\frac{1}{2^{2k(\alpha-1)}} \tag{12}$$

Note that this estimate holds without any a-priori assumptions on the smoothness of the original curve. There are many results along these lines for wavelets, including local versions, which all pass over directly to our geometric constructions, see [6] for details.

In other words, $\hat{\beta}$ is nearly optimal for estimating the true β's under our Sobolev risk.

3.3.2. The ℓ^2 term

We want our estimator $\hat{\gamma}$ to produce a curve which is near the original both in terms of regularity, for which we use d_α, and in terms of absolute location, for which we use an ℓ^2 type distance d_2. Let $\hat{p}(n)$ denote our estimates for the true point locations $\gamma(n)$. In the current work, we simply use the observed point locations $z(n)$; i.e. we set

$$\hat{p}(n) = z(n)$$

This is a minimum-risk estimator if we consider each point separately, corrupted by mean-zero white noise.

Remark. There are certainly other methods for handling ℓ^2 risk, which we hope to explore in future work. First, one might simply require that the algorithm preserve the large-scale β's, as is typically done with wavelet shrinkage. However, while the analogy between β angles and wavelet coefficients is ideally-suited for the study of local regularity properties of curves, it is less natural for ℓ^2-type estimates. It is certainly true that β angles at large scales (indeed, at any scale) determine the curve up to rotation, translation, and scale, so we could simply require that our denoising method attempt to preserve the β's at large scales. However, it seems awkward, for example, to adapt the β's to handle local noise, as the β angles at large scales about a single point do not reflect distance between points.

We feel it will be both more practical and illuminating to consider another method, which provides a way of generalizing "shrinkage toward zero" to geometric situations. Shrinking small β's or wavelet coefficients to zero works precisely because small coefficients are typically due to noise. However, shrinking point locations to zero is obviously pointless. Instead, a natural analog of zero is provided by the best-fit least-squares arcs identified earlier, when doing geometric feature detection. Recall that at each point, we identified the best-fit circular arc at various scales and then chose the best scale based on the R^2 statistic. One might choose $\hat{p}$ to be, say, the nearest point to $z(n)$ on the local arc of best fit (or some average over several local best-fit arcs) and then apply shrinkage: move $z(n)$ towards $\hat{p}(n)$ if $\|z(n) - \hat{p}(n)\|$ is below the local noise threshold. We hope to explore this in future work.

3.3.3. The estimator $\hat{\gamma}$ is near-optimal

Ultimately we want to construct a curve $\hat{\gamma}$, not a collection of angles and approximate point locations. We do so by projecting the point locations and $\hat{\beta}$'s onto the space of curves; i.e. we minimize the total distance

$$d^2\big(\eta,(\hat{p},\hat{\beta})\big) \equiv d_2^2(\eta,\hat{p}) + Cd_\alpha^2(\beta_\eta,\hat{\beta})$$

over curves η to define the projection $\Pi(\hat{p},\hat{\beta})$:

$$\Pi(\hat{p},\hat{\beta}) \equiv \arg\min_{\eta} d^2\big(\eta,(\hat{p},\hat{\beta})\big)$$

for some operator-chosen constant $C \in \mathbb{R}$. This, then, is our estimator:

$$\hat{\gamma} \equiv \Pi(\hat{p},\hat{\beta})$$

From this definition we see that the angles of $\hat{\gamma}$ will look like the $\hat{\beta}$'s, plus an error term which lives in the Sobolev space H^α. Assuming the parameter α is chosen equal to or greater than the exponent of the Sobolev space in which γ lies, we lose essentially no regularity in taking this projection. The following easy estimate shows that $\hat{\gamma}$ is as regular as $\hat{\beta}$, and we know $\hat{\beta}$ is itself as regular as β_γ with high probability by (12).

Lemma 2.

$$d\big(\hat{\gamma},(\gamma,\beta_\gamma)\big) \leqslant 2d\big((\hat{p},\hat{\beta}),(\gamma,\beta)\big)$$

and

$$R\big(\hat{\gamma},(\gamma,\beta_\gamma)\big) \leqslant 4R\big((\hat{p},\hat{\beta}),(\gamma,\beta_\gamma)\big)$$

Proof. The first statement is just the triangle inequality:

$$\begin{aligned} d\big(\hat{\gamma}, (\gamma, \beta_\gamma)\big) &\leqslant d\big(\hat{\gamma}, (\hat{p}, \hat{\beta})\big) + d\big((\hat{p}, \hat{\beta}), (\gamma, \beta_\gamma)\big) \\ &\leqslant 2d\big((\hat{p}, \hat{\beta}), (\gamma, \beta_\gamma)\big) \end{aligned}$$

where $d(\hat{\gamma}, (\hat{p}, \hat{\beta})) \leqslant d((\hat{p}, \hat{\beta}), (\gamma, \beta_\gamma))$ by definition of $\hat{\gamma}$. The second statement is then just the definition of risk:

$$R\big(\hat{\gamma}, (\gamma, \beta_\gamma)\big) = Ed^2\big(\hat{\gamma}, (\gamma, \beta)\big) \qquad \square$$

Thus, this result guarantees that the approximate curve is expected to lie near the original curve in the Sobolev-type sense given by our risk. Further, like wavelet shrinkage for function estimation, this estimator is also nearly optimal, since our noise model is such that the ideal risk is identical to the case of function estimation.

Remark. Donoho and Johnstone obtained essentially the best possible statement about function estimation. Why are their methods a "right" thing to do on curves? For example, an important fact in their original development is that white noise on sampled functions values passes through to a white noise on the empirical wavelet coefficients. Indeed, this follows immediately from the fact that the mapping from samples to wavelet coefficients is orthogonal. Unfortunately, the transformation from wavelet coefficients of, say, parametrized curves to β angles is not orthogonal (indeed, it is not even linear). Why should we expect their methods to work for curve estimation?

The justification lies in the results in Section 2: Theorem 3 and Proposition 5 show that the analogy between β angles and wavelet coefficients is not just philosophical. The mapping from β's to wavelet coefficients is in fact almost diagonal, with good estimates for the off-diagonal terms. Hence one may, in principle, view the β's as wavelet coefficients. As a rule, methods which are designed for operating on wavelet coefficients will translate directly into corresponding methods for operating on β's, up to some typically very mild error terms.

3.4. The denoising algorithm

We now collect the results above and summarize the algorithm. It works as follows:

(1) Singularity detection:
   (a) Choose a largest and smallest distance $0 < M_{\min} < M_{\max}$.
   (b) At each point $z(n)$ and each scale $M_{\min} < M < M_{\max}$, compute residuals $R_L^2(n, M)$ and $R_\mu^2(n, M)$ after fitting least-squares lines and constants.
   (c) At each point $z(n)$ and scale $M_{\min} < M < M_{\max}$, estimate $\sigma^2(n, M)$ as the local average of residuals

$$\hat{\sigma}^2(n, M) = \frac{1}{2M}\left(\sum_{j=n+M}^{n+2M} R_L^2(j, M) + \sum_{j=n-2M}^{n-M} R_L^2(j, M) \right)$$

   and declare $z(n)$ to be a "corner at scale M" when

$$R_L^2(n, M) - R_\mu^2(n, M) > \hat{\sigma}^2(n, M)$$

   (d) For each point $z(n)$, there will be a set of "existence intervals" $I_{n,j} = [a_{n,j}, b_{n,j}]$ where $z(n)$ was identified as a corner for $M \in I_{n,j}$. Optionally, reject corners where the largest such interval has $\log_2 \frac{b_{n,j}}{a_{n,j}} < 1$ (or some other operator-chosen value). This heuristic tends to eliminate corners which are due only to noise.
   (e) Optionally, merge corners which are within distance $M_{\min}$ (or other operator-chosen threshold) of adjacent corners. At large scales, many adjacent points near a corner will be flagged as corners; this step may or may not be necessary depending on the method of reconstruction chosen. An effective criterion seems to be to choose the "true" corner to be the one with the largest existence intervals.

(2) Denoising:
   (a) Choose a number $C \in \mathbb{Z}_+$ of additional scales to add to the curve. If we had N sample points on Γ to begin with, the denoised curve will have $2^C N$ points.
   (b) Let S denote the set of upsampled curves:

$$S = \left\{ \{w_j\}_{j=1}^{2^C N} \colon\ w_j \in \mathbb{R}^2 \right\}$$

   Let N_s be the indices of points not flagged as corners. Let $K(n)$ be the scale of the largest dyadic interval about w_n which contains no corner points. Now choose parameters $\alpha \in \mathbb{R}$ and $\lambda_k \in \mathbb{R}$ for $k = 1, 2, 3, \ldots, 2^C N$ and solve the minimization problem

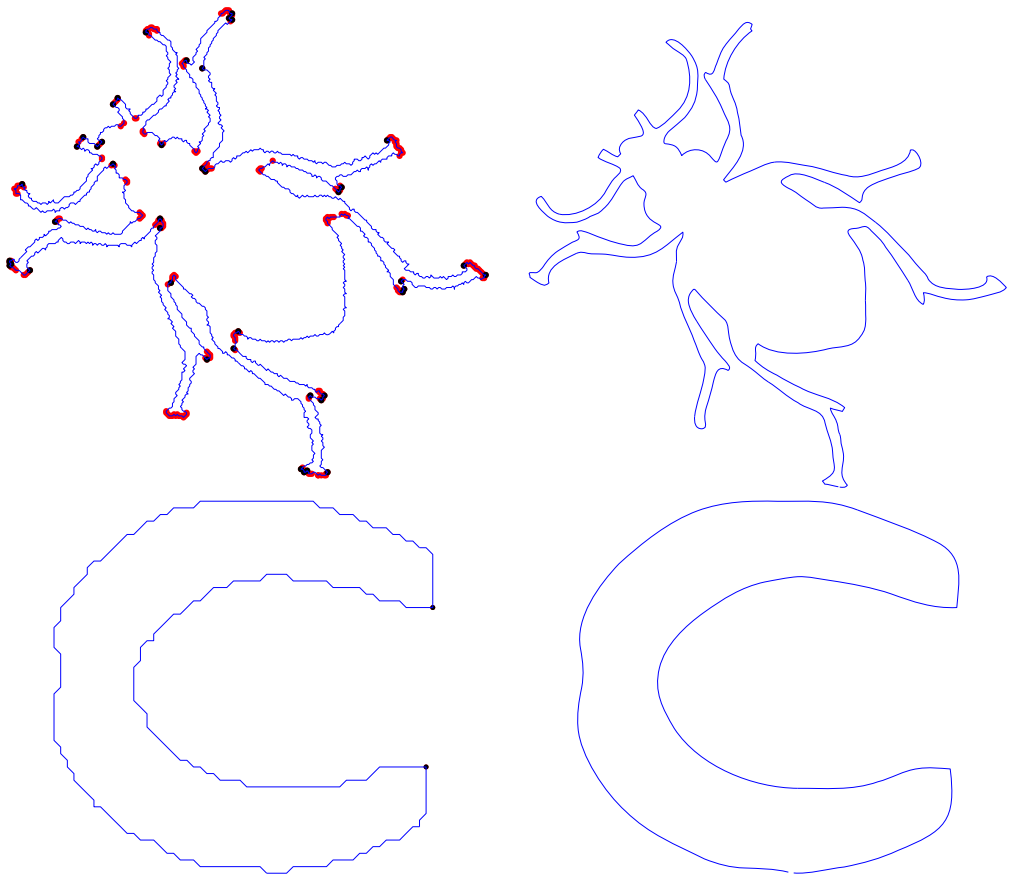

Fig. 3. Above: Curves from the MPEG-7 shape database. Left: Original curve with corners; black dots indicate corners which appear corner-like at the smallest detectable scale. Light dots are corners at larger scale. Right: Denoised curves with $\lambda_n = 1/2\hat{\sigma}_z$.

$$w^* = \arg\min_{\{w\in S\}} \sum_{n\in N_s} \sum_{k=1}^{K(n)} \big(\beta(n,k) - \hat{\beta}(n,k)\big)^2 2^{-2k\alpha} + \sum_{n=1}^{N} \lambda_n^2 \big| w_{1+(n-1)2^c} - z_n \big|^2 \tag{13}$$

to find the denoised curve w^*. The angles $\beta(n,k)$ are deduced from the curve w.
The weights λ_n may be chosen by the operator as tuning parameters based on prior knowledge, or they may be chosen adaptively by the algorithm. A natural adaptive choice, used in our numerical experiments below, is to choose λ_n inversely proportional to the expected magnitude of the local noise, i.e. set λ_n proportional to $1/\hat{\sigma}_{z(n)}$.

In the final step, the second term in the functional serves to keep the reconstructed curve near the original in a weighted least-squares sense; allowing λ_n to vary with n lets us account for the local level of noise (we say more about this below). The first term smooths by driving the β's to the estimated $\hat{\beta}$'s.[3]

The multiresolution analysis of corner-like features above may be analyzed further to make final decisions about the existence of corners on the curve. For example, the procedure above may identify a number of corners which are indeed corners (as opposed to noise), but which only exist at small scales, while there are simultaneously points which are corners at large scales. There seem to be at least two ways to make use of this information. First, there is information contained in the corner existence intervals. See Fig. 5 for an example where a simple analysis of the histogram of right-endpoints of the corner existence intervals contains enough information to identify the two obvious choices of scale for denoising. While one could look at all possible choices of noise threshold, the existence intervals tend to suggest the few "good" choices. Another kind of information comes from measures of goodness-of-fit of our models for corners and arcs; in Section 3.2.2 we describe a method using the R^2 statistic to identify the "best" choices of scale.

Finally, it is worth mentioning that this same algorithm works well for detecting corners at various resolutions whether the original data has been corrupted by noise or not. Indeed, the small-scale structure of a curve may well be considered

[3] It is also effective to replace the first term with the simpler term $\sum_{n\in N_s} \sum_{k=1}^{K(n)} \beta(n,k)^2 2^{-2k\alpha}$. This loses some small-scale details, but always gives a smoother reconstruction which may be preferable and provides a faster algorithm.

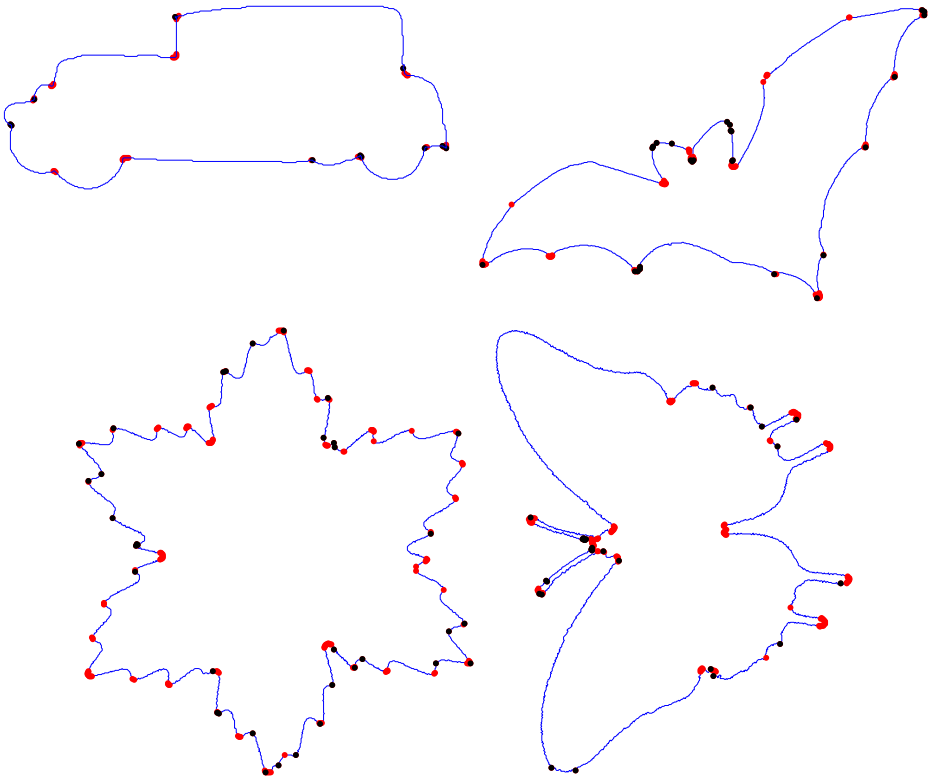

Fig. 4. The corner detector. Black dots indicate corners which appear corner-like at the smallest detectable scale. Light dots are corners at larger scale.

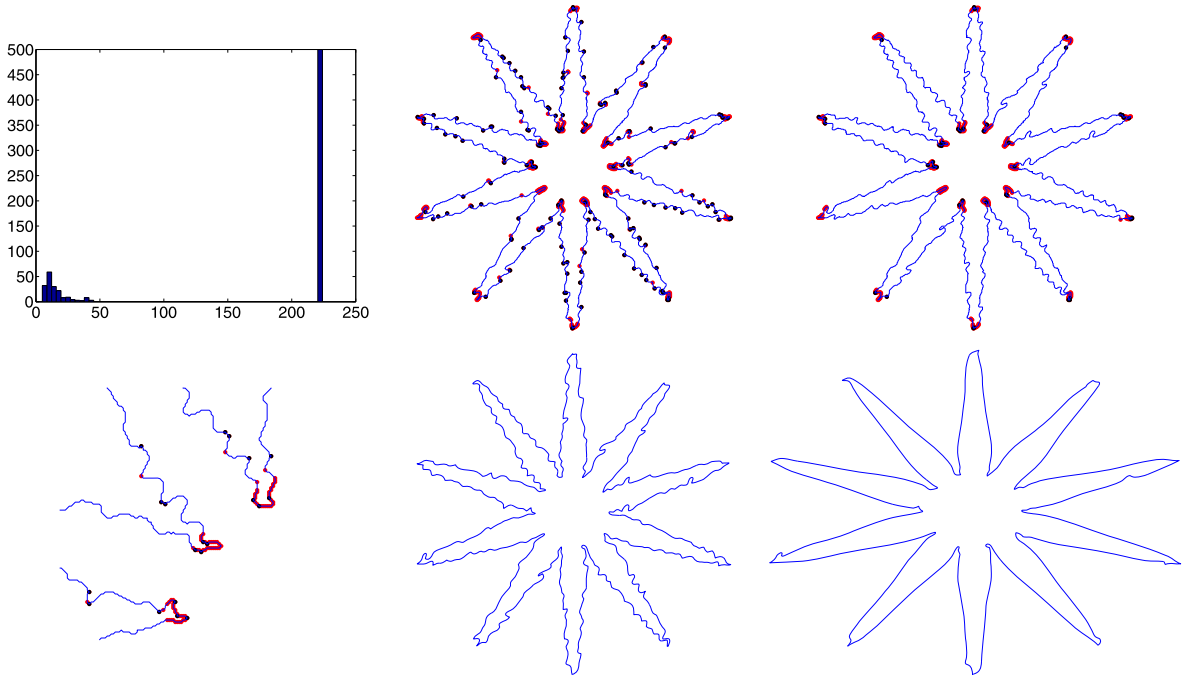

Fig. 5. Are the small bumps on the starfish signal or noise? There is no right or wrong answer. This curve contains 4480 points and looks quite smooth at small scales, as shown in the lower left detail view. However, the histogram of corner scales (upper left) screams "there are exactly two good choices". Denoising with all corners yields the middle pictures; large-scale only yields the right-side pictures. The corner existence intervals can be used to prune out corners at scales one is not interested in. Upper left: the histogram of largest corner-like scales. Upper middle: corners at all scales. Upper right: corners at scales above 200 points. Lower left: original curve detail view with corners. Lower middle: denoised curve when using all corners. Lower right: denoised curve using only corners at scales above 200 points. In both denoisings, λ_n was chosen with the same rule of $\lambda_n = 1/2\hat{\sigma}_z$. More displacement is allowed in the second case since estimates for local noise increase.

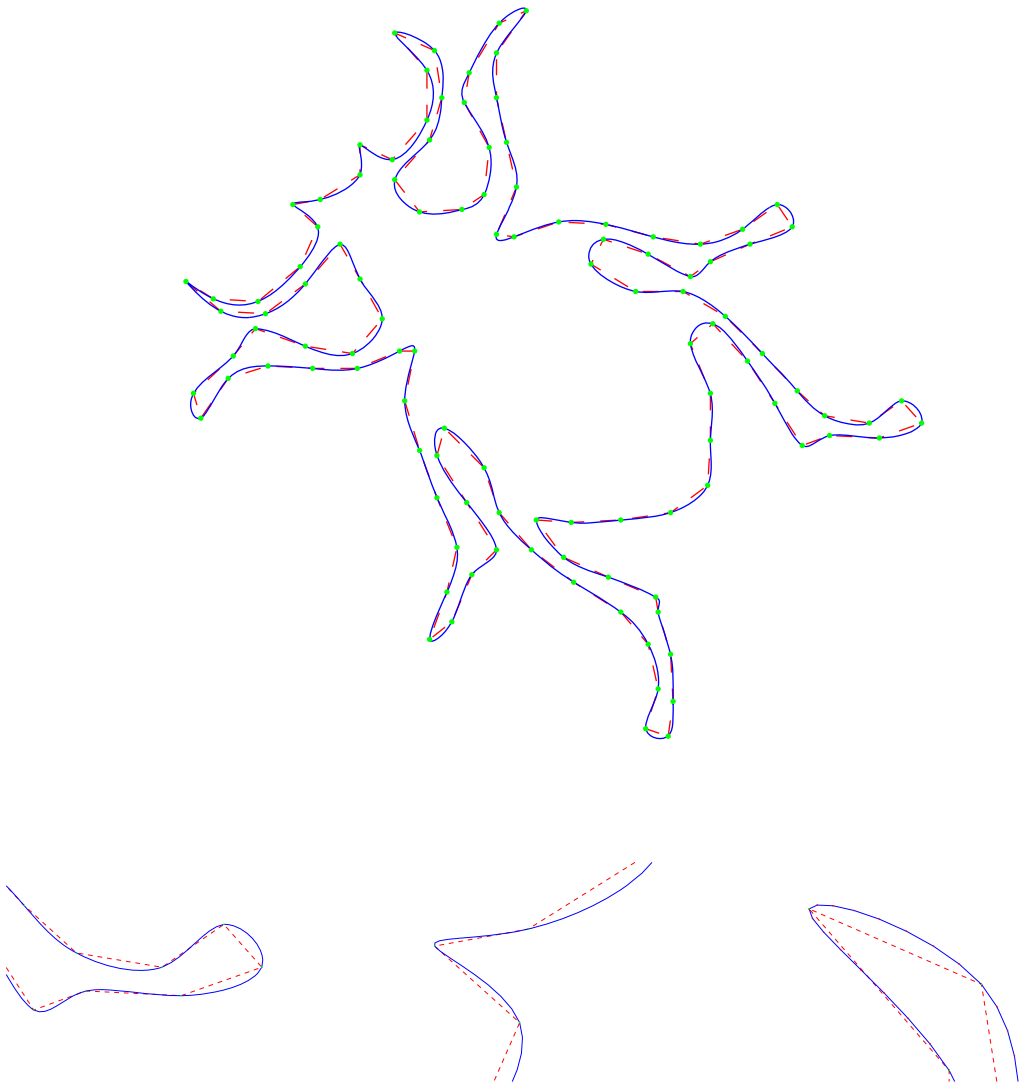

Fig. 6. Above: An upsampled beetle from the MPEG-7 shape database without corner detection. To create this image, all but 100 of the original 2232 sample points were discarded. The green points and dashed line are the low-resolution polygon which was upsampled by three additional scales to create the smoother blue curve. Below: Closeups of a leg, a horn, and an antenna; the smooth overshoot is a result of running the algorithm without corner detection. Corners would be difficult to detect accurately in this situation given the scarcity of sample points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

noise when the curve is viewed at large scales. In this case, the most reasonable choice of parameter p for the noise seems to be $p = 0$, or white noise, with the assumption that the variance is relatively small compared to the scale. The idea is that small-scale structure on an arc which is smooth at large scales often looks like noise whose variance is independent of location on the arc; consider the graph of $\sin(x)$, which looks like a straight line at large scales, for a motivating example.

4. Examples

We present some examples of denoising curves obtained from the MPEG-7 shape 1b dataset in Figs. 3–6. This dataset consists of a number of binary images at various sizes and resolutions; to obtain curves we used the MATLAB built-in function “bwboundaries”. Thus, our contours contain noise typically from two sources: pixelation noise at fine scales, arising from our use of a naive boundary-tracking algorithm, and additional noise at larger scales, intentionally added during the creation of the dataset. We emphasize that all examples shown were constructed with identical choices of parameters: we used the adaptive method for choosing weights λ_n with constant of proportionality $1/2$, and the noise variance parameter was $p = 0$ (white noise). The solutions were obtained with a straightforward gradient descent algorithm based on the functional (13). In these examples, the sample points were taken at arc-length with respect to the noisy curve, not with respect to the original (unknown) curve. In order to accommodate this we allowed the reconstructed points to lie at non-

uniform distance along the reconstructed curve. As we remarked above, the arc-length assumption is made primarily to simplify the exposition; up to constants, the theory carries over to the non-uniform case as long as the original curve was sampled at a rate which was bounded above and below. In all examples we tried, no special care was needed to enforce this additional requirement on the reconstructed curve.

5. Conclusion and comments

We have presented a novel and effective method for denoising plane curves. The results are scale-invariant when working in the Sobolev space $H^{3/2}$, and the denoising method is essentially free of tuning parameters. In future work we plan to extend these methods in several ways:

(1) Extension to surfaces. While the algorithm we designed here is tailored for curves (in particular, we rely on the natural arc-length parametrization on curves), the underlying theory passes over to handle curves and surfaces in higher dimensions (see [1,2], for example). In higher dimensions, the theory tells you that it is possible to denoise by fitting pieces of planes, low-order polynomials, or other simple geometric objects at each point and scale, and studying the L^2 residuals, which behave in some sense like wavelet coefficients. We also observe that the second differences in Theorem 1 could be used directly to construct a version of our current algorithm for curves in higher dimensions.
(2) Improved corner detection. Currently, the corner detector considers β over all scales, but only at each point taken individually. This generally works well, but occasional mistakes are made. It would be better, and not much more difficult, to consider the behavior of β in the *neighborhood* of a potential corner and extend the likelihood ratio test to work on this data. Since the behavior of β in a neighborhood of a corner is easy to compute, this should increase statistical power significantly with only a mild increase in algorithmic complexity.

We also wish to make one remark in the case where one is given simply a set of sample points from a curve or surface, rather than a parametrization of the object. In this case, we could still denoise, but we would need to solve a manifold fitting problem, either prior to or concurrently with denoising. In the case of a plane or space curve, this is very much like the second author's solution to the traveling salesman problem [11,15]. Adapting that proof to our situation would amount to locally approximating the sample points with straight lines at all locations and scales and using the ℓ^2 residuals to infer a curve containing (or simply near, when denoising) the sample points. Other variants are possible, say by fitting circles, and initial experiments seem promising. For a simple curve it appears that this problem will not be especially thorny, but if the curve has many self-intersections, or near-intersections, this quickly becomes difficult. In the case of surfaces, the possible kinds of singularities and self-intersections are even more troublesome. Still, similar techniques seem to show promise in high-dimensional problems [20,13]. We look forward to extending the methods described here to broader classes of problems.

References

[1] G. David, S. Semmes, Harmonic analysis and the geometry of subsets of $\mathbb{R}^n$, Publ. Mat. 35 (1991) 237–249.
[2] G. David, S. Semmes, Analysis of and on Uniformly Rectifiable Sets, American Mathematical Society, 1993.
[3] D. Donoho, I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1993) 425–455.
[4] David L. Donoho, De-noising by soft-thresholding, 1994.
[5] David L. Donoho, Xiaoming Huo, Beamlets multiscale image analysis, in: Multiscale and Multiresolution Methods, Springer, 2001, pp. 149–196.
[6] David L. Donoho, Iain M. Johnstone, Gžrard Kerkyacharian, Dominique Picard, Wavelet shrinkage: Asymptopia? 1995.
[7] M. Grayson, The heat equation shrinks embedded plane curves to round points, J. Differential Geom. 26 (1987) 285–314.
[8] I.M. Johnstone, B.W. Silverman, Wavelet threshold estimators for data with correlated noise, J. Roy. Statist. Soc. Ser. B 59 (2) (1997) 319–351.
[9] P. Jones, T. Le, Local scales and multiscale image decompositions, Appl. Comput. Harmon. Anal. 26 (3) (2009) 371–394.
[10] Peter W. Jones, Square functions, Cauchy integrals, analytic capacity, and harmonic measure, in: Springer Lecture Notes in Math., vol. 1384, 1989, pp. 24–68.
[11] Peter W. Jones, Rectifiable sets and the traveling salesman problem, Invent. Math. 102 (1) (1990) 1–15.
[12] Jean Luc Starck, Emmanuel J. Candes, David L. Donoho, The curvelet transform for image denoising, 2000.
[13] M. Maggioni, Intrinsic dimensionality estimation and multiscale geometry of data sets, in: Applied Mathematics Seminar, Yale University, February 2010.
[14] B. Merriman, J.K. Bence, S. Osher, Diffusion Generated Motion by Mean Curvature, Dept. of Mathematics, University of California, Los Angeles, 1992.
[15] K. Okikiolu, Characterization of subsets of rectifiable curves in $\mathbb{R}^n$, J. Lond. Math. Soc. (2) 46 (1992) 336–348.
[16] S. Osher, L.I. Rudin, Feature-oriented image enhancement using shock filters, SIAM J. Numer. Anal. 27 (4) (1990) 919–940.
[17] L. Rudin, S. Osher, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1–4) (1992) 259–268.
[18] K. Siddiqi, B.B. Kimia, Chi-Wang Shu, Geometric shock-capturing ENO schemes for subpixel interpolation, computation, and curve evolution, in: International Symposium on Computer Vision, 1995, p. 437.
[19] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, 1970.
[20] Teng Zhang, Arthur Szlam, Yi Wang, Gilad Lerman, Randomized hybrid linear modeling by local best-fit flats, CoRR, abs/1005.0858, 2010.