LOCAL REGULARITY-BASED IMAGE DENOISING

P. Legrand, J. Lévy Véhel

Irccyn, 1 Rue de la Noé, BP 92101, 44321 Nantes, France
{Pierrick.Legrand, Jacques.Levy-Vehel}@irccyn.ec-nantes.fr

ABSTRACT

We present an approach for image denoising based on the analysis of the local Hölder regularity. The method takes the point of view that denoising may be performed by increasing the Hölder regularity at each point. Under the assumption that the noise is additive and white, we show that our procedure is asymptotically minimax, provided the original signal belongs to a ball in some Besov space. Such a scheme is well adapted to the case where the image to be recovered is itself very irregular, e.g. nowhere differentiable with rapidly varying local regularity. The method is implemented through a wavelet analysis. We show an application to SAR image denoising where this technique yields good results compared to other algorithms.

1. INTRODUCTION

A large of number of techniques have been proposed for image denoising. The basic frame is as follows. One observes a image Y which is some combination $F(X, B)$ of the signal of interest X and a "noise" B. Making various assumptions on the noise, the structure of X and the function F, one then tries to derive a method to obtain an estimate $\hat{X}$ of the original image which is optimal in some sense. Most commonly, B is assumed to be independent of X, and, in the simplest case, is taken to be white, Gaussian and centered. F usually amounts to convoluting X with a low pass filter and adding noise. Assumptions on X are almost always related to its regularity, e.g. X is supposed to be piecewise C^n for some $n \geq 1$. Techniques proposed in this setting resort to two domains: functional analysis and statistical theory. In particular, wavelet based approaches, developed in the last ten years, may be considered from both points of view [1, 2].

Our approach in this work is different from previous ones in several respects. First, we do not require that X belongs to a given *global* smoothness class but rather concentrate on its *local* regularity. More precisely, we view denoising as equivalent to increasing the Hölder function α_Y (see section 2 for definitions) of the observations. Indeed, it is generally true that the local regularity of the noisy observations is smaller than the one of the original image, so that in any case, $\alpha_{\hat{X}}$ should be greater than α_Y. If the Hölder function of X happens to be known, it may serve as a target for the algorithm. If this is not the case, it can be estimated from Y provided sufficient information on F and B is available (e.g. independent additive noise of known law). More generally, the largest $\alpha_{\hat{X}}$, the more regular the estimate will be, and the smoother it will look. We thus define our estimate $\hat{X}$ to be the image which minimizes the risk under the constraint that it has the desired Hölder function. Note that since the Hölder exponent is a local notion, this procedure is naturally adapted for images which have sudden changes in regularity, like discontinuities. In addition, this scheme is appropriate when one tries to process images for which it is important that the right regularity structure be recovered. An example of this situation is when denoising is to be followed by image segmentation based on textural information: Suppose we wish to differentiate highly textured zones (appearing for instance in MR or radar imaging) in a noisy image. Applying an denoising technique which assumes that the original image is, say, piecewise C^1, will induce a loss of the information which is precisely the one needed for segmentation: The denoised image will not contain much texture, and cannot be used for segmentation.
Our denoising technique is thus well suited to the case where the original signal X displays the following features:

- X is everywhere irregular.
- The regularity of X (as measured by its Hölder function) varies rapidly in space
- The local regularity of X bears essential information for subsequent processing.

The remaining of this paper is organized as follows. Section 2 recalls some basic facts about Hölder regularity analysis, which is the basis of our approach. The denoising method is explained in section 3. Section 4 gives some theoretical results. Numerical experiments are displayed in section 5.

2. HÖLDER REGULARITY ANALYSIS

A popular way of evaluating the regularity of an image is to consider a family of nested functional spaces, and to determine the ones it actually belongs to. A usual choice is to consider Hölder spaces, either in their local or pointwise version. To simplify notations, we deal with 1D signals, and we assume that our signals are nowhere differentiable. Generalization to differentiable signals simply requires to introduce polynomials in the definitions [3].

Definition 1 *Pointwise Hölder exponent*
Let $\alpha \in (0, 1)$, and $x_0 \in K \subset \mathbf{R}$. A function $f : K \to \mathbf{R}$ is in $C^{\alpha}_{x_0}$ if for all x in a neighbourhood of x_0,

$$|f(x) - f(x_0)| \leq c|x - x_0|^{\alpha} \tag{1}$$

where c is a constant.

The pointwise Hölder exponent of f at x_0, denoted $\alpha^p_f(x_0)$, is the supremum of the α for which (1) holds.

Let us now introduce the local Hölder exponent: Let $\alpha \in (0, 1)$, $\Omega \subset \mathbf{R}$. One says that $f \in C^{\alpha}_l(\Omega)$ if:

$$\exists\, C : \forall x, y \in \Omega : \frac{|f(x) - f(y)|}{|x - y|^{\alpha}} \leq C$$

Let: $\alpha_l(f, x_0, \rho) = \sup\{\alpha : f \in C_l^\alpha(B(x_0, \rho))\}$ $\alpha_l(f, x_0, \rho)$ is non increasing as a function of ρ.

We are now in position to give the definition of the local Hölder exponent :

Definition 2 *Let f be a continuous function. The* local Hölder exponent *of f at x_0 is the real number:*

$$\alpha_f^l(x_0) = \lim_{\rho \to 0} \alpha_l(f, x_0, \rho)$$

Since α^p and α^l are defined at each point, we may associate to f two functions $x \to \alpha_f^p(x)$ and $x \to \alpha_f^l(x)$ which are two different ways of measuring the evolution of its regularity.

These regularity characterizations are widely used in fractal analysis because they have direct interpretations both mathematically and in applications. It has been shown for instance that α^p indeed corresponds to the auditive perception of smoothness for voice signals. Similarly, computing the Hölder exponents at each point of an image gives a good idea of its structure, as for instance its edges [4]. More generally, in many applications, it is desirable to model, synthesize or process signals which are highly irregular, and for which the relevant information lies in the singularities more than in the amplitude. In such cases, the study of the Hölder functions is of obvious interest.

In [5], a theoretical approach for signal denoising based on the use of the pointwise Hölder exponent and the associated multifractal spectrum was investigated. The work in [6] proposes an approach similar to the one we explain below, but in a functional frame. We shall develop here a practical denoising technique in a statistical setting. More precisely, we will assume a definite model for the noise and its interaction with the image, and we shall derive the denoising procedure in a *minimax* frame.

3. IMAGE DENOISING

Let X denote the original image and Y the degraded observations. Our model supposes that $Y = X + B$ with B a centered white Gaussian noise independent of X. The precise marginal law of B is however not important, and the computations below could be adapted with minor modifications to non-Gaussian cases. We seek a denoised version $\hat{X}$ of Y that meets the following constraints:

1. the risk $E(\|\hat{X} - X\|^2)$ is "small",
2. the Hölder function of $\hat{X}$ is prescribed.

If α_X is known, we choose $\alpha_{\hat{X}} = \alpha_X$. In some situations, α_X is not known but can be estimated from Y (see [7]). Otherwise, we just set $\alpha_{\hat{X}} = \alpha_Y + \delta$, where δ is a user-defined positive function, so that the regularity of $\hat{X}$ will be everywhere larger than the one of the observations.

Two problems must be solved in order to obtain $\hat{X}$. First, we need a procedure that computes the Hölder function of a signal from discrete observations. Second, we need to be able to manipulate the data so as to impose a specific regularity.

Both these aims may be reached through a wavelet analysis. Indeed, under some assumptions, one may estimate and control the Hölder regularity via wavelet coefficients. More precisely, let $\{\psi_{j,k}\}_{j,k}$ be an orthonormal wavelet basis, where as usual j denotes scale and k position, and assumes that ψ is regular enough and has sufficiently many vanishing moments. For ease of notation, we shall explain the method in one dimension (generalization to images is straightforward). Let $X_n = (x_1^n, \ldots x_{2^n}^n)$ denote a regular sampling over the 2^n points $(t_1^n, \ldots t_{2^n}^n)$ of the original signal. Let $(c_{j,k})$ denote the wavelet coefficients of X, $(d_{j,k})$ denote the coefficients of Y, and $(\hat{c}_{j,k})$ denote the coefficients of $\hat{X}$. From our noise model, and since we use orthonormal wavelets, we have: $d_{j,k} = c_{j,k} + \frac{\sigma}{\sqrt{2^n}} z_{j,k}$, where σ is the standard deviation of the noise B and $z_{j,k}$ are iid Gaussian variables with unit variance.

For $p = 1 \ldots 2^n$, we consider the point $i = t_p^n$ and the wavelet coefficients which are located "above" it, i.e. $d_{j,k(j,i)}$ with $k(j,i) = \lfloor (i-1)/(2^{n+1-j}) \rfloor + 1$. In general, estimating the Hölder exponents at a given i from the wavelet coefficients is not an easy task, because one needs to take into account all the coefficients in a neighbourhood of i. In this work, we shall assume that our signals verify $\alpha_l(i) = \alpha_p(i)$ at all i. The implications of such an assumption are discussed in [**?**]. One then has:

$$\alpha_l(i) = \alpha_p(i) = -\frac{1}{2} + \text{liminf}_{j \to \infty} \frac{\log(|d_{j, \lfloor (i-1)/(2^{n+1-j}) \rfloor + 1}|)}{-j} \quad (2)$$

The equality above means that when the local and pointwise Hölder exponents at i coincide, we may compute their common value by looking only at the wavelet coefficients above i. In the sequel, we shall denote the common value by $\alpha(i)$. To simplify the exposition, we assume in addition that the liminf in (2) is a limit (the more general case may be handled with techniques presented in [8]). When this is the case, the Hölder exponent may be estimated through a linear regression of the logarithm of the wavelet coefficients above i with respect to scale.

We may now formulate our denoising scheme as follows: For a given set of observations $Y = (Y_1, \ldots, Y_{2^n})$ and a target Hölder function α, find $\hat{X}$ such that the risk $R = E(||\hat{X} - X||^2)$ is minimum and the regression of the logarithm of the wavelet coefficients of $\hat{X}$ above any point i w.r.t. scale is $-(\alpha(i) + \frac{1}{2})$. Note that we must adjust the wavelet coefficients in a global way. Indeed, each coefficient at scale j subsumes information about roughly 2^{n-j} points. Thus we cannot consider each point i sequentially and modify the wavelet coefficients above it to obtain the right regularity, because point $i+1$, which shares many coefficients with i, requires different modifications. The right way to control the regularity is to write the regression constraints simultaneously for all points. This yields a system which is linear in the logarithm of the coefficients:

$$ML = A$$

where M is a $(2^n, 2^{n+1} - 1)$ matrix of rank 2^n, and

$$\begin{aligned} L &= (\log|\hat{c}_{1,1}|, \log|\hat{c}_{2,1}|, \log|\hat{c}_{2,2}|, \ldots \log|\hat{c}_{n,2^n}|), \\ A &= -\frac{n(n-1)(n+1)}{12}\left(\alpha(1) + \frac{1}{2}, \ldots, \alpha(2^n) + \frac{1}{2}\right) \end{aligned}$$

Since we use an orthonormal wavelet basis, the requirements on the $(\hat{c}_{j,k})$ may finally be written as:

minimize: $\quad R = E(\sum_{j,k} (\hat{c}_{j,k} - c_{j,k})^2)$

subject to: $\quad \forall\, i = 1 \ldots 2^n$,

$$\sum_{j=1}^{n} s_j \log(|\hat{c}_{j, \lfloor (i-1)/(2^{n+1-j}) \rfloor + 1}|) = -M_n(\alpha(i) + \frac{1}{2}) \quad (3)$$

where the coefficients $s_j = j - \frac{n+1}{2}$, $M_n = \frac{n(n-1)(n+1)}{12}$ and equation (3) are deduced form the requirement that the linear regression of the wavelet coefficients of $\hat{X}$ above position i should equal $-(\alpha(i)+\frac{1}{2})$.

Searching the most general solution to the program above is not an easy task. We consider instead in this paper the following special case: We impose that, for all (j,k),

$$\hat{c}_{j,k} = B_j d_{j,k} \qquad (4)$$

where the multipliers B_j are real numbers belonging to the interval $(0,1]$. The main motivation for the restriction on the form of the $\hat{c}_{j,k}$ is of course that it leads to a simple solution. The choice of the range of the B_j parallels an idea at work in classical denoising by wavelet shrinkage, namely that we seek to reduce the variance of the estimator by decreasing the absolute value of the coefficients. Now, by definition:

$$\sum_{j=1}^{n} s_j \log(|d_{j,\lfloor (i-1)2^{j+1-n}\rfloor}|) = -M_n(\alpha_Y(i)+\frac{1}{2}) \qquad (5)$$

Subtracting (5) to (3) and using (4), we get:

$$\forall i=1,\ldots,2^n, \sum_{j=1}^{n} s_j \log(B_j) = M_n(\alpha_Y(i)-\alpha(i))$$

Thus the ansatz (4) imposes that the desired increase in regularity is uniform along the path, i.e. $\delta(i)=\delta=$ constant. This restriction can be weakened by a classical block technique. Our problem now reads:

minimize: $R = E(\sum_{j,k}(B_j d_{j,k} - c_{j,k})^2)$

subject to: $\sum_{j=1}^{n} s_j \log(B_j) = \beta$ and $0 < B_j \leq 1$

where $\beta = M_n(\alpha_Y(i)-\alpha(i))$. The risk is computed as follows:

$$\begin{aligned} R &= E\left(\sum_{j,k} B_j(c_{j,k} + \frac{\sigma}{\sqrt{2^n}} z_{j,k}) - c_{j,k}\right)^2 \\ &= \sum_{j,k}\left(c_{j,k}^2(1-B_j)^2 + B_j^2\frac{\sigma^2}{2^n}\right) \\ &= \sum_j e_j^2(1-B_j)^2 + \sigma^2\sum_j 2^{j-n}B_j^2 \end{aligned}$$

where $e_j^2 = \sum_k (c_{j,k})^2$ is the energy of X at scale j.

The constrained minimization has a unique solution $B^* = (B_1^*,\ldots,B_n^*)$. The values are classically found using a Lagrange multiplier. One gets:

$B_j^* = \frac{1\pm\sqrt{\Delta}}{2}a_j$ with $\Delta = 1 - \frac{\lambda 2^{\beta+1}(j-\frac{n-1}{2})}{a_j e_j^2}$ and $a_j = \frac{e_j^2}{e_j^2+2^{j-n}\sigma^2}$.

Here, λ is the Lagrange multiplier. No closed form exist for λ. It is obtained through a numerical procedure (see [7] for more details).

It is interesting to compare our scheme with the soft-thresholding policy. Recall that soft-thresholding replaces the noisy coefficient $d_{j,k}$ by $e_{j,k} = \text{sgn}(d_{j,k})(|d_{j,k}|-\lambda)_+$, where λ is a threshold that depends, among other things, on n and the type of noise. Denoting $\beta_j = -\log(B_j)$, we see that in our case:

$$\log(|\hat{c}_{j,k}|) = \log(|d_{j,k}|) - \beta_j, \text{with } \text{sgn}(\hat{c}_{j,k}) = \text{sgn}(d_{j,k})$$

The regularity based enhancement is thus a kind of shrinkage on the logarithm of the wavelet coefficients, and the restriction (4) may be interpreted as a requirement that the threshold must depend only on scale and not on position.

4. THEORETICAL RESULTS

In order to obtain convergence results, we need to assume that X has some regularity. A convenient frame is provided by the Besov spaces $\mathcal{B}^{\alpha}_{p,\infty}$ (see for instance [2] for an account on Besov spaces).

Proposition 1 *Assume that $X \in \mathcal{B}^{\alpha}_{p,\infty}$, with $p \in [2,\infty], \alpha > \frac{1}{p}$. Then the Hölder based denoising is asymptotically minimax, i.e.*

$$R = O(2^{-n\frac{2\alpha}{2\alpha+1}}) \qquad \textit{when } n \to \infty$$

See [7] for a proof.

5. NUMERICAL EXPERIMENTS

Our examples deal with synthetic aperture radar (SAR) images. A huge literature has been devoted to the difficult problem of enhancing these images, where the noise, called speckle, is non Gaussian, correlated and multiplicative. A fine analysis of the physics of the speckle suggests that it follows a K distribution [9]. Classical techniques specifically designed for SAR image denoising include geometric filtering and Kuan filtering. Wavelet shrinkage methods have also been adapted to this case [10].

SAR imaging of natural landscapes is a good test for our technique, since the original signal is itself irregular. Although the noise is not additive, it is interesting to see how the Hölder based denoising performs on such data. We display on figure 1 an original image along with its denoising using a) Kuan filtering, b) classical wavelet hard-thresholding, and c) the algorithm developed in this paper. Figure 2 shows a Hölder based denoising on another SAR image.

6. REFERENCES

[1] R.A. Devore and B. Lucier, "Fast wavelet techniques for near optimal image processing," *IEEE Military Communications Conference*, vol. 2–12, 1992.

[2] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory 41*, pp. 613–627, 1994.

[3] Y. Meyer, "Wavelets, Vibrations and Scalings," *American Mathematical Society, CRM Monograph Series*, vol. 9, 1997.

[4] J. Lévy Véhel, "Fractal Approaches in Signal Processing," *Fractals*, vol. 3(4), pp. 755–775, 1995.

[5] J. Lévy Véhel and B. Guiheneuf, "Multifractal Image Denoising," *SCIA*, 1997.

[6] J. Lévy Véhel, "Signal enhancement based on Hölder regularity analysis," *IMA Volumes in Mathematics and its Applications*, vol. 132, pp. 197–209, 2002.

[7] P. Legrand and J. Lévy Véhel, "Statistical denoising of irregular signals," Tech. Rep., Inria, 2002.

[8] P. Legrand and J. Lévy Véhel, "Signal processing with FracLab," *preprint*.

[9] C.J Oliver, “Information from SAR images,” *J. Phys. D*, pp. 1493–1514, 1991.

[10] L. Gagnon and F. Drissi Smaili, “Speckle noise reduction of airborne SAR images with symetric Daubechies wavelets,” *Signal and Data Processing of Small Targerts, Proc. SPIE 2759*, 1996.

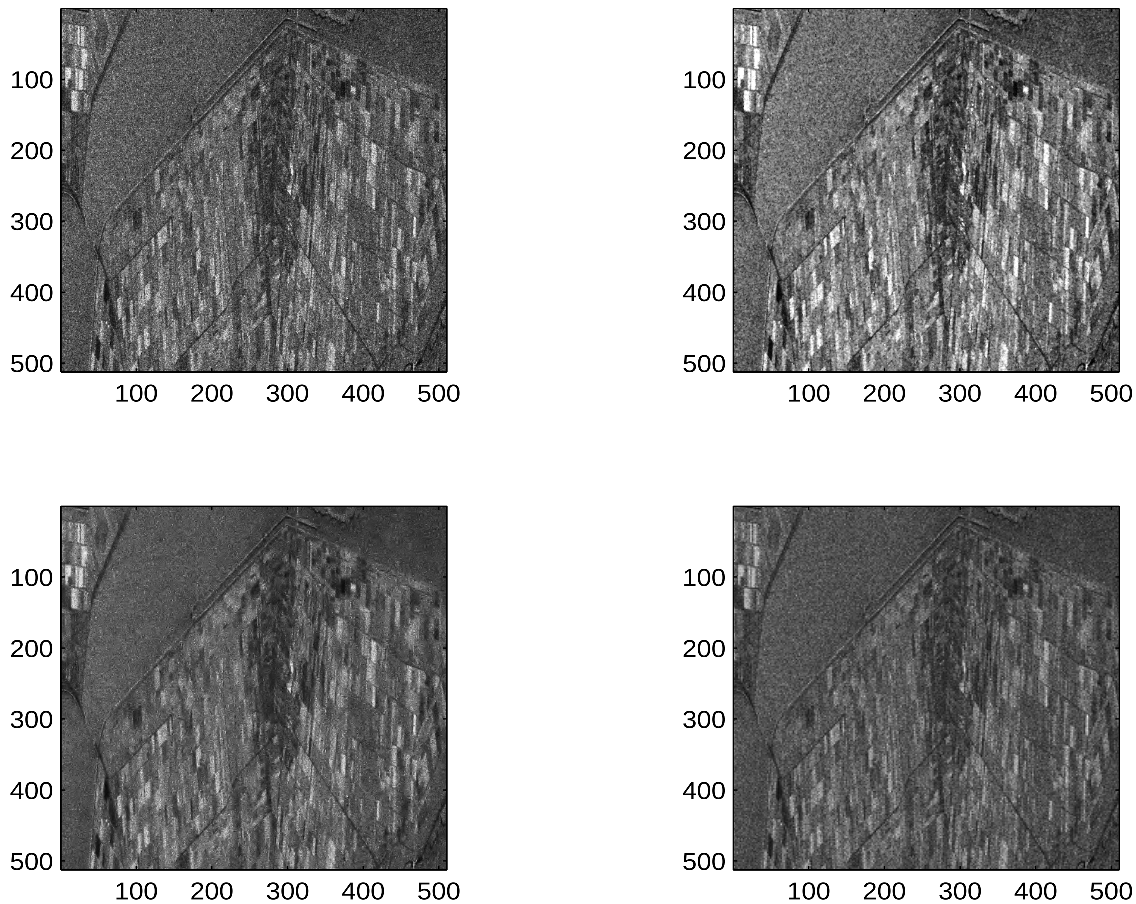

Fig. 1. *Original SAR image (top left), Kuan median filtering (top right), wavelet shrinkage (bottom left), local regularity based method (bottom right).*

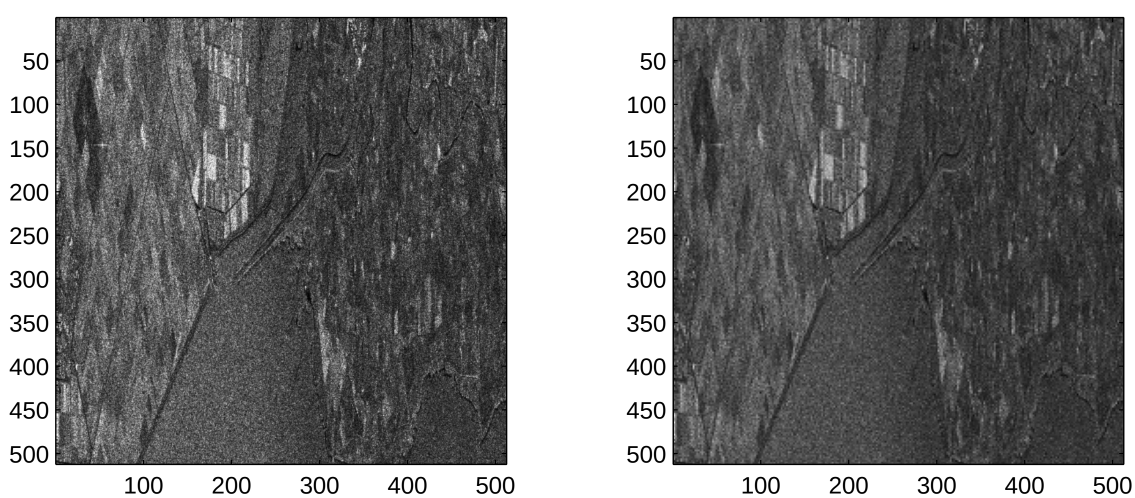

Fig. 2. *Original SAR image (left), local regularity based method (right).*