UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

A Noise Selection Approach of Denoising

Francois Malgouyres

February 2001

CAM Report 01-04

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

A noise selection approach of denoising*

F. Malgouyres†

February 12, 2001

Résumé

Cet article traite d'une méthode de débruitage. Cette dernière n'est qu'une variante des méthodes de seuillages en ondelettes traditionelles. Cependant, elle présente l'avantage de permettre l'usage de plusieurs bases et ce de façon à sélectionner ce qui est percu comme de l'information par une base ou une autre base ou une autre base,... pour autant de bases que l'on peut en souhaiter. Le coup numérique de cette méthode réside essentiellement dans le calcul des coordonées du signal (ou de l'image) dans les bases considérées.

Abstract

This paper deals with a denoising method. This latter is only a small modification from the usual wavelet thresholding ones. However, it has the significant advantage to allow the use of several bases in such a way that we select what is considered as information by a basis or another basis or another basis, and so on for as many bases as we want. The computational cost of the method is mainly the computation of the coordinates of the signal (or image) in the bases.

1 Introduction

This paper is mainly concerned with image or signal denoising. More precisely, we are going to introduce a method which permits to generalize the usual wavelet thresholding procedure to a kind of thresholding of the projection of the signal (image) on elements of a dictionary. Basically, the dictionary will be a set of functions which does not necessarily have to be a basis. Our method permits to preserve information as soon as it yields a large scalar product with one of the element of the dictionary. In this sense it is heuristically close to best basis and matching pursuit algorithms. However, our method can only be used for the issue of denoising (not compression) but it has the advantage of being fast and easy to implement.

Here, we will understand denoising as methods whose aim is to recover a signal (or image) $u \in \mathbb{R}^N$, from a data

$$v_l = u_l + b_l$$

for N an integer, $l \in \{0, ..., N-1\}$ and b a noise (we will assume it Gaussian).

There is an large number of papers dealing with this problem. Among these papers two "families" are often opposed : the wavelet and variational approaches. Among variational approaches, the ones based on the minimization of the total variation appears to be the most efficient. This particular energy was introduced in [11] and has been, since then, studied under various aspects. Wavelet methods was introduced by Donoho and Johnstone and are studied and extended in several papers (see [5, 9, 12]).

Concerning the particular aspect of using several bases for the denoising, we can first mention the average over translations (see [2, 8]). There exists also some papers on methods, which use several bases, whose results are probably close to the method we are proposing here. For instance the best basis and matching pursuit aims at selecting the information with regards to its projection on elements of several bases [1, 3, 9, 10]. These methods cost more computations and are more complex to implement than the one we propose here. However, our "noise selection" approach does not make sense in the framework of signal (or image) compression.

*This work has been financed by NSF grant DMS-9973341.

†UCLA Dept of Mathematics, 6363 Math science building, Box 951555, Los Angeles, CA 90095-1555, USA;
http://www.math.ucla.edu/~malgouy
malgouy@math.ucla.edu

2 Noise selection

Let us now introduce the new method under consideration. This method is based on the remark that in order to separate the noise and the information it is simpler to characterize the noise than the information.

Indeed, usual similar methods (for instance wavelet ones) generally use the fact that a given basis gives a sparse representation of an image. Knowing that and the fact that a Gaussian noise has, in practice, a bounded l^∞ norm in any considered basis. These methods basically only preserve the large coefficients in this appropriate basis. They are generally supported by arguments claiming that a given basis yields a sparse representation of a given class of function.

The method we are proposing in this paper is in fact only a small modification of these latest. It consists simply in taking a "noise point of view". More precisely, another interpretation of the preceding methods consists in saying that we want to remove from the image all what is not considered as information by a given basis. Having said that, it is clear that we do not need to restrict ourselves to a basis and that we can simply remove all what is not considered as information by a given dictionary. As a consequence, our algorithm does not try to recover the information present in an image, it tries to determine the noise (and then subtract it to the initial data). Let us state that mathematically.

Let $\mathcal{D} = \{\psi_l\}_{0\leq l\leq P}$ be a dictionary of functions $\psi_l \in \mathbb{R}^N$ which are not identically zero (typically, $\mathcal{D}$ can be a union of bases). For simplicity, we assume that $\|\psi_l\|_2 = 1$, for all $l = 0, ..., P$. For any $\sigma > 0$, we define, for $v \in \mathbb{R}^N$, the operator B_σ, which goes from $\mathbb{R}^N$ into itself, by the result of the following iterative process:

$$v_0 = v,$$

and for $l \in \{0, \cdots, P-1\}$, we let $v_{l+1} = v_l - \delta_{l+1}\psi_{l+1}$ with

$$\delta_{l+1} = \begin{cases} 0 & , \text{if } |\langle v_l, \psi_{l+1}\rangle| \leq \sigma \\ \langle v_l, \psi_{l+1}\rangle - \sigma & , \text{if } \sigma \leq \langle v_l, \psi_{l+1}\rangle \\ \langle v_l, \psi_{l+1}\rangle + \sigma & , \text{if } \langle v_l, \psi_{l+1}\rangle \leq -\sigma \,. \end{cases} \tag{1}$$

Then, we let $B_\sigma(v) = v_P$.

Remark that, when $\mathcal{D}$ contains only one wavelet basis, the function $v - B_\sigma(v)$ is exactly the usual wavelet soft thresholding. Moreover, we could modify (1) in such a way that $v - B_\sigma(v)$ is the hard thresholding (we simply have to set $\delta_{l+1} = \langle v_l, \psi_{l+1}\rangle$ when $|\langle v_l, \psi_{l+1}\rangle| > \sigma$).

Note also that, applying (1), we are not sure that

$$|\langle B_\sigma(v), \psi_l\rangle| \leq \sigma\,, \tag{2}$$

for all $l \in \{0, ..., P\}$. However, if we recursively define $v^{k+1} = B_\sigma(v^k)$, with $v^0 = v$, we can extract from the sequence $(v^k)_{k\in\mathbb{N}}$, of elements of $\mathbb{R}^N$, a sequence which converges to a function satisfying (2) (indeed, $\|v^{k+1}\|_2 \leq \|v^k\|_2$, for $k \in \mathbb{N}$).

Moreover, it is clear that the result of B_σ strongly depends on the order in which the functions are stored in the dictionary $\mathcal{D}$. For instance, if we take for the first elements of $\mathcal{D}$ the basis made of Dirac delta functions on sampling points, for any bright image, after the noise selection, we get a noise which is uniformly equal to σ. It seems therefore advisable to first use functions which highly decorrelate information and noise (such as the elements of a usual wavelet basis). However, it is clear that we can switch any consecutive functions such that

$$\langle \psi_l, \psi_{l+1}\rangle = 0\,.$$

First, this means that, if $\mathcal{D}$ is a union of orthonormal bases, we can compute the coefficient of our image in the first basis, select the noise on all coefficients, without taking care of their order, and then redo this for the other bases. (Instead of computing one coefficient, modify it, compute the second coefficient, ... as described in (1).) Second, this can be used to modify the order of the functions of $\mathcal{D}$ in order to accelerate the process yielding the result. For instance, if $\mathcal{D}$ contains a wavelet basis and a wavelet packet basis, we can use the wavelet decomposition of the image in order to compute its wavelet packet decomposition without changing the result. This yields a "multilevel" kind of wavelet packet noise selection which is equivalent to the composition of a wavelet noise selection and several wavelet packet noise selection.

Remark that, when adding a new function to our dictionary, the computational cost is basically increased by the computation of one scalar product with this function. It even happens that the computation of this scalar product is an intermediate step for the computation of another scalar product; in which case it does almost cost nothing. For instance, if we take the example of the wavelet packet basis based on an admissible binary tree of full depth J, it almost costs nothing to make the noise selection in all the intermediate bases of levels lower than J (of course, here, we do not make the noise selection on the elements of the bases which correspond to low frequencies).

The simplicity of this method permits to envisage the use of very large dictionaries. We can for instance quote the dictionaries generally used in matching pursuit and best basis pursuit (see [9]) : wavelet packets, Gabor dictionary, translations of wavelet and wavelet packet bases. However, it is clear that this method will much more improve its results when used with a dictionary made of different kind of bases. For instance, it could be interesting to add a wavelet packet dictionary with other bases such as curvelets or ridgelets (see [12]), or even the Fourier basis. It could also be interesting to define some criterion (probably derived from the usual criterion saying that a wavelet basis yields a sparse representation of the information) on dictionaries to characterize whether they sufficiently select decorrelated noise or not. Moreover, we can also use texture dictionaries, dictionaries made of characteristic functions (maybe modified to have a mean equal to 0), dictionaries adapted to particular type of images (for instance if we are only interested in denoising images of faces, it seems accurate to use a dictionary made of eyes, noses, mouths,...), we are mainly restricted by our imagination, computation issues and the fact that the elements of our dictionary must be "autocorrelated".

Note also that it is possible to compose this noise selection with some algorithms of other types. Indeed, given a denoising algorithm yielding a result $Algo(v)$, we can apply the noise selection described above to the function $v - Algo(v)$ and then remove $B_\sigma(v - algo(v))$ to v as a final result. We present in the next section some results when combining a noise selection with the Rudin-Osher-Fatemi[1] denoising. Note that doing so, we simply denoise the main structure with the Rudin-Osher-Fatemi method and then denoise the texture with our noise selection. This is quit similar to the idea of representing an image as a sketch plus some texture (see [6]).

At last, we can adapt the method to the issue of deconvolution. Similarly to the existing wavelet packet deconvolution methods, the difficulty in this case is that we have to restrict ourselves to dictionaries such that $\langle h^{-1}(n), \psi_{l+1}\rangle$ are "almost independent variables" for a noise n and for h^{-1} the "pseudo-inverse" of the convolution (see [7]). Note that it is also possible to apply some ideas similar to the ones introduced in [4] by modifying (1) in such a way that the test is made on $\langle v_l, \tilde{\psi}_{l+1}\rangle$, instead of $\langle v_l, \psi_{l+1}\rangle$, for an convenient function $\tilde{\psi}_{l+1}$. Remark that we can generally do the same comments when the noise is not Gaussian or white.

3 Numerical results

We present here some experiments for the noise selection method. All the parameters (the thresholds and the parameter of the Rudin-Osher-Fatemi method) have been estimated empirically. However, when using a larger dictionary we have observed that the method is much more robust to the increase of the threshold. The considered wavelet basis is obtained with a cubic spline wavelet and is of depth 4.

We display on Figure 1 an extracted part of some experiments made on the image "Barbara". They correspond to:

- Up-Left : The initial image.
- Up-Right : The preceding image plus a Gaussian noise of standard variation 30.
- Middle-Left : An image denoised with a wavelet soft thresholding for a threshold $\sigma = 75$.
- Middle-Right : An image denoised by the Rudin-Osher-Fatemi method with a parameter $\lambda = 0.0005$.

[1]This method is based on the minimization of

$$\|\nabla u\|_1 + \lambda \|u - v\|_2^2 , \tag{3}$$

among $u \in \mathbb{R}^N$, and was introduced in [11].

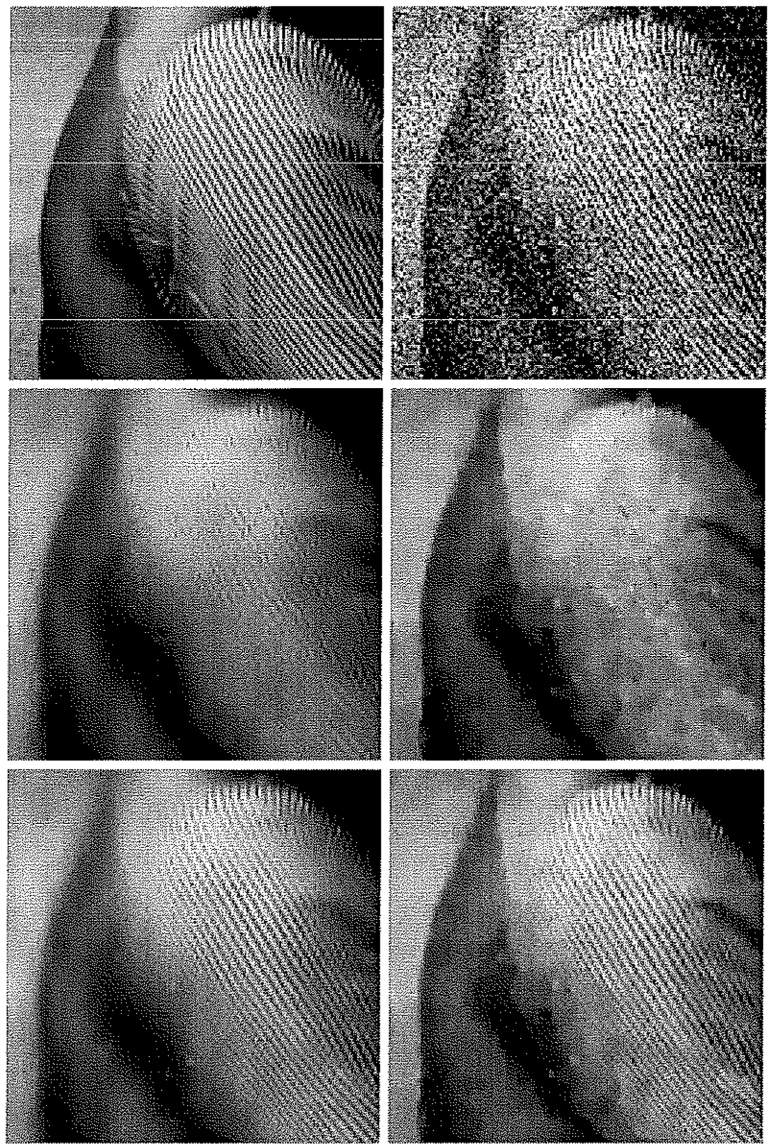

Figure 1: Up-Left : The initial image. Up-Right : The noisy image. Middle-Left : wavelet soft thresholding. Middle-Right : Rudin-Osher-Fatemi method. Down-Left : The noise selection in a wavelet basis, all the wavelet packet bases of full depth 2, 3 and 4 and a Fourier basis. Down-Right : Rudin-Osher-Fatemi method "composed" with the preceding noise selection.

image, method or dictionary	noisy image	wavelet thresholding	wavelet wavelet packets noise selection	wavelet wavelet packets Fourier	Rudin-Osher-Fatemi method	ROF W+WP+F noise selection
SNR	6.36	9.03	10.85	11.47	10.43	13.01
MSE	893	315	214	186	239	135

Table 1: Signal to noise ration and mean square error between the image and the initial image (ROF, W, WP, F refer respectively to Rudin-Osher-Fatemi, wavelet, wavelet packet and Fourier.

- Down-Left : The noise selection with a dictionary[2] containing a wavelet basis, all the wavelet packet bases of full depth 2, 3 and 4 (again with a cubic spline wavelet) and a Fourier basis. We choose a parameter $\sigma = 95$.
- Down-Right : The noise selection similar to the preceding one computed on what has been considered like noise by Rudin-Osher-Fatemi method (with $\lambda = 0.0005$).

It is clear that the use of several bases permits a better preservation of the textures. Indeed, the texture on the pants yields small wavelet coefficients and is considered like noise by such a basis. However, when we test this noise in a wavelet packet basis with a better frequencial localization, this texture is this time considered as information and is therefore preserved. Moreover, when preliminary using the Rudin-Osher-Fatemi method, the noise selection approach permits to properly restore edges. Furthermore, the difference between this result (Down-Right) and the preceding one (Down-Left) is mostly concentrated on large wavelet coefficients. It can be understood as an alternative to soft and hard thresholding.

We also summarize some statistics between the different denoised images and the initial image in the Following table.

It is clear here that when more bases are considered we get better statistics. The best result being the one for the combination of Rudin-Osher-Fatemi method and the noise selection in several bases. Note that we have not displayed the image obtained by a noise selection in wavelet and wavelet packets bases only since it is visually similar to the image displayed on Figure 1 Down-Left.

References

[1] S. Chen and D. Donoho. Atomic decomposition by basis puirsuit. In *SPIE, international Conference on Wavelets*, San Diego, July 1995.

[2] R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. *A. Antoniadis and G. Oppenheim, editors, Wavelets and statistics*, pages 125–150, 1995. NewYork, Springer Verlag.

[3] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE, Transactions on Information Theory*, 38(2):713–718, March 1992.

[4] D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.

[5] D. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. Technical report, Departement of Stat., Stanford University, 1992.

[6] J. Froment. A compact and multiscale image model based on level sets. *Control, Optimisation and Calculus of Variation*, 4:473–495, August 1999.

[7] J. Kalifa. *Restauration minimax et déconvolution dans une base d'ondelettes miroirs.* PhD thesis, Ecole Polytechnique, 1999. Available at http://www.cmap.polytechnique.fr/~kalifa.

[2] Remark that despite the size of the dictionary the computational cost is almost the one of the calculation of the wavelet packet coefficients in the basis of full depth equal to 4 and the Fourier basis.

[8] F. Malgouyres. Convolution approximation by mean of an operator diagonal in a wavelet packet basis and application to image deblurring. CAM report at UCLA, available at http://www.math.ucla.edu/~malgouy, November 2000.

[9] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, Boston, 1998.

[10] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE, Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[11] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[12] J.L. Starck, E.J Candès, and D.L. Donoho. The curvelet transform for image denoising. Technical report, Dept of Statistics of Stanford University, November 2000.