A Greedy Algorithm for a Sparse Scalet Decomposition

Albert Bijaoui

A Greedy Algorithm for a Sparse Scalet Decomposition

Albert Bijaoui
University of Nice Sophia Antipolis
Département Cassiopée UMR CNRS 6202
Observatoire de la Côte d'Azur, B.P. 4229 F-06304 Nice Cedex4
Email: Albert.Bijaoui@oca.eu

***Abstract*—Sparse decompositions were mainly developed to optimize the signal or the image compression. The sparsity was first obtained by a coefficient thresholding. The matching pursuit (MP) algorithms were implemented to extract the optimal patterns from a given dictionary. They carried out a new insight on the sparse representations. In this communication, this way is followed. It takes into account the goal to obtain a sparse multiscale decomposition with the different constraints: i/ to get a sparse representation with patterns looking like to Gaussian functions, ii/ to be able to decompose into patterns with only positive amplitudes, iii/ to get a representation from a translated and dilated pattern, iv/ to constrain the representation by a threshold, v/ to separate the sparse signal from a smooth baseline. Different greedy algorithms were built from the use of redundant wavelet transforms (pyramidal and à trous ones), for $1D$ signals and $2D$ images.**

Experimentations on astronomical images allow one a gain of about two in sparsity compared to a classical DWT thresholding. A fine denoising is obtained. The results do not display any wavy artifacts. This decomposition is an efficient tool for astronomical image analysis.

I. Introduction

Many kinds of sparse decompositions were developed these two last decades mainly in order to optimize the signal or the image compression [1]. Among them the wavelet transform appeared as a very efficient way to carry out sparse representations for a large class of signals [2]. Since different new transforms were proposed for optimizing sparsity for a given signal class [3]. The sparsity was first obtained with an available coefficient thresholding. The Discrete Wavelet Transform (DWT) concentrates the significant information onto the minimum of coefficients without any redundancy between the data and the transform.

Sparsity can be also derived from overcomplete representations, by choosing optimally the projection vectors [4]. Matching pursuit (MP) algorithms were implemented to extract the representation elements from a given dictionary [5]. The non-orthogonality between the alphabet patterns can lead to convergence problems. The orthogonal matching pursuit (OMP) algorithm [6] allows one to keep the orthogonality between the representation and the residuals. Different algorithms were proposed in order to extract in parallel the patterns [7], [8].

The basis pursuit [9] has given a new insight on the sparse representations. The introduction of a ℓ_1 constraint on the representation has reduced significantly the number of coefficients. Some algorithms were proposed in order to solve the associated variational problem. The relation between the different ways to obtain a sparse decomposition has been carefully studied. If the sparsity is sufficient, the basis pursuit algorithms minimizes also the norm ℓ_0 [10] [11]

Here another insight is given, taking into account our goal, obtaining a sparse multiscale decomposition of large images with the following constraints:

- The dictionary has to be built from well-defined patterns. Our goal does not consist in the search of the largest sparsity but in the search of how to get a representation with simple patterns. This is required for a further signal/image analysis for which a decomposition into separate objects is needed. The alphabet is made of monomodal, positive and symmetric patterns, looking like to Gaussian functions.
- The patterns were chosen compact. This allows one to increase the separation between them and thus improve the pattern extraction.
- The pattern amplitudes can all be positive. This constraint is connected to our scientific goal, i.e. to describe a signal/image into positive separate objects. In specific cases these amplitudes could all be negative.
- Our goal is also to get a representation as covariant as possible with translations and dilations. This leads one to generate the pattern set by translations and dilations of a basic pattern.
- The decomposition is constrained by a threshold. In the case of a noisy signal, this threshold is linked to false alarm detections at each scale.
- The signal representation is the sum of the sparse decomposition with a baseline. This baseline is such that no significant pattern can be detected from it at each examined scale, for the given threshold.

This last constraint plays an essential role in the algorithm. This characteristic is introduced in order to take into account our goal, the extraction of objects in astronomical images. In this framework, a sky background is supposed to be superimposed on the objects field. The background is first computed and then removed. Here, the background is considered as a spurious component which is simultaneously determined with the searched components.

II. The greedy algorithms.

Hereafter, the algorithm is presented only in its $1D$ pyramidal version.

A. The basic patterns.

The multiresolution pyramid is defined by [12]:

$$F(i,k) = \frac{1}{2^i} < f(x), \phi(\frac{x}{2^i} - k) > . \quad (1)$$

where the scaling function $\phi(x)$ satisfies the dilation relation [13]:

$$\frac{1}{2}\phi(\frac{x}{2}) = \sum_{n=n1,n2} h(n)\phi(x-n), \quad (2)$$

Here $h(n)$ is symmetric and $n_1 = -n_2$. The following recursive relation is easily derived:

$$F(i+1,k) = \sum_{n=n1,n2} h(n)F(i, 2k+n). \quad (3)$$

The coefficients can be also written as:

$$F(i,k) = \sum_{m=m1,m2} h(i,m)F(2^i k + m). \quad (4)$$

The patterns $h(i,m)$ are obtained by the following recursive relation:

$$h(i+1,m) = \sum_{n=n1,n2} h(i, m-2^i n)h(n). \quad (5)$$

At scale 0 we have $h(0,m) = \delta(m)$, which leads to $h(1,n) = h(n)$. The bounds (m_1, m_2) increase exponentially ($m_1 = n_1 2^i$ and $m_2 = n_2 2^i$). The pyramidal algorithm carries out the set of correlations of the signal with the filters $\{h(i,m)\}$. These patterns are built by a pyramidal rule (Equation 5), so, we call them *pyrels* for (*pyr*amid *el*ements). These pyrels may also be called *scalet pyrels* taking into account their construction from the scaling function. It can be denoted that the patterns $\{h(i,m)\}$ derives from the scaling function, but it does not correspond to its sampled values.

B. The background removal.

Let us consider the wavelet transform which results from the differences between two successive approximations:

$$w(i+1,k) = F(i,k) - \tilde{F}(i+1,k) \quad (6)$$

where $\tilde{F}(i+1,k)$ is the approximation at scale $i+1$ before the decimation. The wavelet coefficients are independent of the constant adding. Taking into account the relation (4) we can write as:

$$w(i+1,k) = \sum_{m=m1,m2} g(i+1,m)F(2^i k + m) \quad (7)$$

where:

$$g(i+1,m) = \sum_{m=m1,m2} h(i,m) - h(i+1,m) \quad (8)$$

Thus, the correlations in each point between the signal and the patterns $g(i,m)$ are obtained with this wavelet transform. The highest amplitude coefficients can be identified and they can be used for the representation in the framework of a matching pursuit algorithm. This is equivalent to threshold the wavelet transform and to restore the signal by inversion. The result displays generally wavy artifacts. An iterative algorithm can be applied on the residuals to reduce them, until we get non significant wavelet coefficients. This process reduces the sparsity and it carries out negative artifacts without physical meaning. In order to avoid this difficulty, the signal will be restored using pyrel patterns, even if the identification is done from the wavelet coefficients. The wavelet transform will only play the role to identify the significant pyrels in presence of a background.

C. The pyrel identification.

The pyrels are selected on a criterion based on the highest SNR. In the hypothesis of a white Gaussian noise, Relation (7) allows one to estimate the $w(i,k)$ standard deviation due to the noise and thus its SNR. The wavelet coefficient $w(i_0,k_0)$ having the highest SNR identifies a pyrel at $(i_0, 2^{i_0-1}k_0)$. Let us consider a pyrel $a_0(i_0,k_0)$. Its wavelet transform is:

$$\bar{w}(i,k) = a_0 v(i, i_0, k, k_0)) \quad (9)$$

$v(i,i_0,k,k_0)$ is the pyramidal wavelet transform of $h(i_0, m + 2^{i_0-1}k_0)$. Due to the decimation from a scale to the following one, this function depends on $k - 2^{i_0-i}k_0$, up to scale i_0. The pyrel amplitude is first estimated by:

$$a(i_0,k_0) = \frac{w(i_0,k_0)}{v(i_0,i_0,0,0)} \quad (10)$$

In the case of isolated pyrels, the amplitude is correctly estimated by the relation (10). If two pyrels at the same scale are too close, it is not possible to separate them directly by scanning the extrema of the wavelet transform. But, if the pyrels are not at the same scale, the largest scale pyrel can be considered as background for the smallest one, and a separation is then possible. The matching pursuit algorithm can proceed at each step on the extrema set. All the extrema having a SNR greater than the chosen threshold are simultaneously considered. The extrema which correspond to the extrema of the wavelet coefficients along the scale are kept, are called suprema. Thus, the pyrel identification consists in two steps:

1) Identification, scale by scale, of the local extrema of the wavelet transform which have a SNR greater than the threshold;
2) Identification of the suprema, corresponding to the extremum along the scale of the wavelet transform.

D. The pyrel coupling.

The pyrels are previously assumed to be separated. A pyrel is identified, its amplitude evaluated and the pyrel subtracted, and so one. But the pyrels are not orthogonal patterns. They are identified from the wavelet coefficients, and the wavelet functions are not also orthogonal. So a coupling exists between the pyrels. The signal is restored by the relation

$$\bar{F}(k) = \sum_l a(i_l,k_l)h(i_l, k - 2^{i_l-1}k_l) \quad (11)$$

where l is the pyrel index, $(l \in (1, L)$. The wavelet transform of the reconstructed signal is:

$$\bar{w}(i,k) = \sum_{l} a(i_l, k_l) v(i, i_l, k, k_l) \qquad (12)$$

The pyrel amplitudes being determined from their wavelet coefficients at the same location, both in scale and position, the following coupling equation results:

$$\bar{w}(i_l, k_l) = \sum_{l'} a(i_{l'}, k_{l'}) v(i_l, i_{l'}, k_l, k_{l'}) \qquad (13)$$

An implicit procedure is done for inverting Equation (13). It takes into account the wavelet transform of the image reconstructed from the pyrels. A Van Cittert correction [14] of the amplitude is done by adding iteratively on each amplitude a term proportional to the difference between the original wavelet coefficient and its restored one.

E. The greedy algorithm from a pyramidal representation.

The pyrel representation from the pyramidal wavelet transform is derived from the previous analysis:

1) Computation of the pyramidal wavelet transform
2) Determination of the extrema of the wavelet transform at each scale
3) Determination of the suprema.
4) Estimation of the maximum SNR, Q_{max}. A threshold equal to αQ_{max} is chosen for the suprema selection. If Q_{max} is little than a given threshold, the loop is stopped and the algorithm goes to step [9].
5) The detected wavelet coefficients $w(i_j, k_j)$, $j \in (1, J)$, are multiplied by a factor which takes into account the ratio between the pyrel amplitude and the wavelet one.
6) The identified pyrels are added to the previous detected ones, which leads to a set $\{a^{(0)}(i_l, k_l)\}$ where l corresponds to the set of identified pyrels for all the iterations.
7) An inverse algorithm is done in order to determine the new set $\{a^{(\infty)}(i_l, k_l)\}$ such that the signal reconstructed with these pyrels carries out after a wavelet transform which has the wavelet coefficients $w(i_l, k_l)$ at the pyrel locations.
8) An image is built from the pyrels. Its pyramidal wavelet transform is computed and it is subtracted to the original one. A test is done if significant wavelet coefficients still exist. If yes, the algorithm goes to step [2].
9) If the restored signal has to be close to the original one, a baseline is added. The baseline is estimated as the difference between the last approximation of the signal and the last approximation of the pyrel representation. As the sampling is progressively reduced, the baseline is reinterpoled by a set of correlations with $2h(n)$, after the insertion at each step a 0 between two values. At the largest scale, the baseline can be reduced to a constant.

F. The algorithm tuning.

The algorithm depends on different parameters.

The low-pass filter. The choice of the low-pass filter is governed by the following constraints:

- Its length must be as short as possible in order to reduce the computations and the boundary artifacts.
- The corresponding scaling function $\phi(x)$ must be always positive and must display only one maximum, in order to associate a pyrel to a maximum.
- The filter must be symmetric, in order to get available pyrel positions.
- $\phi(x)$ must be as regular as possible, taking into account its length. This will avoid to introduce pyrels due to the irregularities of the scaling function.

The binomial filter $\{h(n) = C_{2l}^{n+l}\}$, for $-l \leq n \leq l$, corresponds to the centered B-spline of order $2l - 1$ for the scaling function. It fully satisfies the constraints. In our application, we use the filter for $l = 2$ that corresponds to the cubic centered B-spline.

The α factor. In the algorithm description, an α factor has been introduced in order to select the suprema which have to be processed in parallel. In the program, it was set to 0.5. With this value it is assumed not to select suprema whose the values would be too contaminated by the new identified pyrels. If this factor decreases, the algorithm may consider the bumps of the coupling function. Experimentally, this factor appears also as a good compromise between sparsity (which increases with α) and computing time (which also increases with α).

The break parameter in the inversion. The pyrel amplitudes are determined by inversion which takes into account the coupling matrix. This inversion is done by a Van Cittert iterative algorithm. The iterations are stopped when the highest residual is greater than rQ_{max}. A value of $r = 0.25$ was experimentally a good compromise .

The Van Cittert convergence factor. The Van Cittert iterative algorithm was designed in order to solve $Y = AX$, where A is smoothing operator. The iterations are written as:

$$X^{(n+1)} = X^{(n+1)} + \beta[Y - AX^{(n)}]. \qquad (14)$$

β is not necessary equal to 1. The stability needs $0 < \beta < 2$, generally $\beta = 1$. The algorithm instability in case of a bad matrix conditioning is well known in deconvolution problems. After experiments, we set in the algorithm $\beta = 0.5$.

G. The monosign pyrel decomposition.

In the previous algorithm the pyrel amplitude could be either positive or negative, according to the supremum sign. Our scientific purpose was to build objects from pyrels. The more often astronomical images display positive objects drowned in a background. It is thus convenient, and easy, to modify the algorithm in order to force the detection of only positive (or only negative) peaks.

We note that the wavelet transform of a real negative peak leads to two positive peaks due to the bumps of the wavelet function. Therefore, in case of a positive decomposition, a peak which significantly corresponds to a negative object can be identified as positive structures. Nevertheless, the bump amplitudes are lower than the central peak and some significant negative peaks can remain after the decomposition. Tests on the residual wavelet coefficients can alarm that significant negative (or positive) peaks still remain.

H. Some properties.

The effect of translations. Due to the decimation, the decomposition is not shift-covariant. It is possible to get this covariance if no decimation is done from one scale to the following one. The algorithm complexity is increased.

Dyadic dilation or contraction. The pyrels are generated by dilation of a generic pattern. If the signal is dilated by the same dilation, as a first approximation, the same pyrels would be identified at the above scales. But, in details some small changes may appear. For the contraction, the pyrel decomposition can differ sensibly from a simple shift along the scales.

Non linearity of the decomposition. The algorithm is covariant with the scalar multiplication if the threshold is multiplied by the same scalar. Let us consider now two signals $\{F_1(k)\}$ and $\{F_2(k)\}$ leading to pyrel decompositions $\{a_{1l}(i_l, k_l)\}$ and $\{a_{2l}(i_l, k_l)\}$. The signals are added:

- For pyrels in the signals at the same location, the resulting pyrel amplitude is the sum of their amplitudes. If they have not the same sign, the new amplitude may be less than the threshold and the pyrel will not be detected.
- Added pyrels may merge into a larger one, more significant.
- If pyrels of each signal have compact separate support, the adding is kept. If not, the decomposition is modified by the pyrel coupling.

The pyrel decomposition of the result is thus generally different of the union of the pyrels coming from each decomposition. Due to the thresholding and the coupling between the pyrels, this decomposition is not linear.

Invariance to a baseline addition. This algorithm is built to be invariant to the addition of a constant. More generally, the algorithm may be invariant to the adding of a non constant baseline whereas its wavelet transform does not modify the detection at the chosen threshold, in the scale range.

One pyrel identification. Let us consider a signal composed only of a pyrel $a_0(i_0, k_0)$. If the amplitude is sufficiently high, the algorithm only detects it. The other extrema (due to the wavelet bumps) are removed taking into account the suprema identification rules. The wavelet amplitude is converted into one pyrel. The pyrel is computed and subtracted. The residuals are null, apart the computational errors. A single pyrel is correctly restored.

Multiple pyrels recognition. Let us consider now a signal composed by L pyrels $\{a_l(i_l, k_l)\}$. The question is to know if the algorithm would restore these pyrels. Evidently this could be possible only if the wavelet coefficients related to the pyrels at the same scales and locations would be greater than the threshold.

But even if this condition is satisfied, there are reasons for which the full identification is not possible:

- The merging of pyrels into larger ones;
- The coupling between them which leads to increase the number of identified pyrels.

Nevertheless, it is clear that the recognition of input pyrels works well their supports are disjoined. The decomposition algorithm can also recognize pyrels even if this condition is not satisfied. The main difficulty lies into its capability to detect the pyrels at their correct location (position and scale) during the identification phase.

A basic condition to recognize a set of pyrels is their linear independence. If the pyrels are not independent, their Gram-Schmidt matrix is singular and the inversion fails.

Stability of the decomposition. Let us consider a signal which leads to a set of pyrels $\{a_l^{(0)}(i_l^{(0)}, k_l^{(0)})\}$ after decomposition with a threshold T. The signal is reconstructed from the pyrels and the decomposition algorithm is applied on it, with the same threshold. The decomposition algorithm is stable on this signal if the restored set $\{a_l^{(1)}(i_l^{(1)}, k_l^{(1)})\}$ is equal to the initial one.

An exact stability is not generally strictly reached, taking into account the computational accuracy. The decomposition is also sensitive to the threshold. In the case of pyrels having amplitudes close to the threshold, computational errors may also lead to remove these features. Numerical experiments showed that stability was not assumed on studied signals. The restored signals are quite identical, but the number of pyrels, their positions, their scales and their amplitudes may sensibly differ.

I. Relation to a variational approach.

This instability shows that this greedy algorithm does not bring the global minimum of a given functional. The algorithm carries out a decomposition interesting for a given analysis. An associated variational approach, if it exists, would improve its stability.

The data attachment. This attachment is generally done by the l_2 distance between the observed data and the restored ones. Here, the removal of the background avoids the application of this principle. The algorithm identifies pyrels from the extrema in the wavelet space, up to a given scale. The identification is based on a thresholding. The data attachment is thus connected to the pseudo-norm:

$$W_{\infty,I} = \max_{i=1,I,k=1,K} \left| \frac{w(i,k) - \overline{w}(i,k)}{\sigma_i} \right| \tag{15}$$

The condition $W_{\infty,I} \leq t$ indicates that a variational version would be connected to a minimax approximation [15] in the wavelet transform space (WTS). A function in this space $w(i, k)$ is decomposed with the set $\{v(i, i_p, k, k_p)\}$, where p covers all the scales and positions for a pyrel.

The prior condition. Today the prior on the decomposition is often linked to the ℓ_1 norm, i.e. the sum of the absolute coefficient values. The basis pursuit corresponds to this variational constraint [9]. An algorithm which minimizes ℓ_1 taking into account $W_{\infty,I}$ would converge to a sparse decomposition. Nevertheless, the proposed algorithm is based on a direct identification of the pyrels, starting from the most significant ones. The goal is to reduce the number of elements allowing one to represent the image such that $W_{\infty,I}$ reaches a given value. So, the prior is not connected to the sum of the absolute values of the coefficients, but to their number.

It was shown previously that the results were not stable, but the variations correspond to quite insignificant coefficients resulting of progressive approximations. An algorithm which constrains to minimize the number of pyrels with $W_{\infty,I} \leq t$, t being a given threshold, would bring the stability. Greedy algorithms which proceed by decreasing the identification thresholds are a natural way to solve the problem, but they do not allow to carry out the stable solution which would correspond to the global minimum in case of non convexity.

Minimax approximations are extensively used in order to approximate functions by rational approximations. The exchange algorithm allows one to get the decomposition, if the Haar condition on the decomposition is satisfied [15]. In the present variational problem, this condition is not satisfied, so that another class of algorithm has to be developed.

J. Two dimensional algorithm.

The $2D$ multiresolution pyramid. The multiresolution pyramid is defined by:

$$F(i,k) = \frac{1}{4^i} < f(x), \phi(\frac{x}{2^i} - k, \frac{y}{2^i} - l) > . \qquad (16)$$

The following recursive relation is derived in case of separate variables:

$$F(i+1,k,l) = \sum_n h(n)h(m)F(i, 2k+n, 2l+m). \qquad (17)$$

The derived patterns. Taking into account the separation between the variables, the coefficients can be also written as:

$$\begin{aligned} F(i+1,k) &= \sum_{p=p1,p2} \sum_{q=q1,q2} h(i,p)h(i,q) \\ & \quad F(2^{i+1}k+p, 2^{i+1}k+q), \end{aligned} \qquad (18)$$

the filters $h(i,p)$ being the $1D$ pyrels.

The identification from the wavelet transform. Similarly, the wavelet transform is built on the difference between two successive approximations. The suprema of the resulting transform are identified, taking into account a threshold, derived from the noise level.

In order to avoid identifying first spurious pyrels, the threshold is progressively decreased, with the same rule that the one used for the $1D$.

The algorithm. Taking into account the previous considerations, the $2D$ algorithm is simply copied from the $1D$ one. The image can be decomposed with only positive (or negative) pyrels, taking into account the suprema sign.

III. An application on an astronomical image.

On Figure 1 left, the image of the planetary nebula NGC40, taken in the near infrared, is plotted. This image is characteristic of the astrophysical images. A bright central star illuminates a shell, which splitted in two parts. The image contains star-like objects and extended diffuse ones.

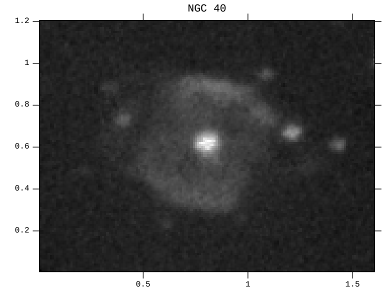

Fig. 1. The infrared image of the planetary nebula NGC40.

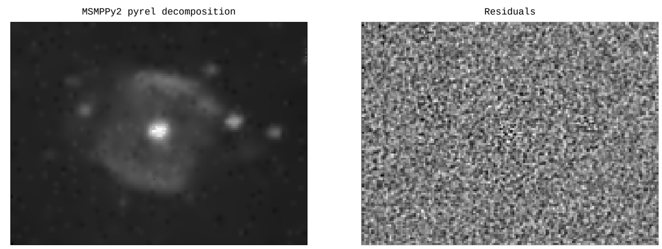

Fig. 2. The reconstructed image of the planetary nebula NGC40 from pyrels identified with the pyramidal algorithm (left). The residual image looks like to a white noisy image (right).

On Figure 2 the image of the planetary nebula NGC40 after the application of this decomposition algorithm, with the unsigned analysis. Here, the number of pyrels corresponds to 3.1% of the number of pixels.

On Figure 3 the image is reconstructed from the decomposition based on only positive pyrels. Now, the number of pyrels corresponds to only 2.6% of the number of pixels.

Using this decomposition, the reduction of the number of significant coefficients is a factor 2 compared to the thresholded orthogonal discrete wavelet transform. Not only the sparsity is seriously increased, but also the reconstruction does not display the wavy artifacts connected to the inverse wavelet transform.

IV. Conclusion

In the present paper, we propose a new way to decompose a signal or an image. Its main specificity is the local background

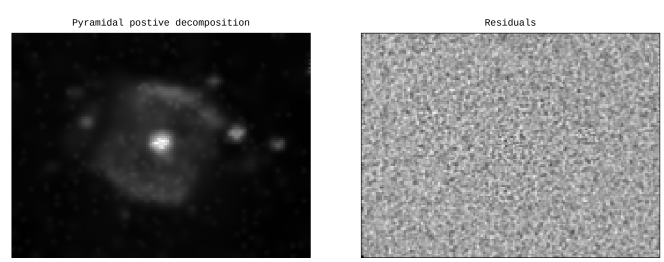

Fig. 3. The reconstructed image of the planetary nebula NGC40 from positive pyrels identified with the pyramidal algorithm (left). The residual image is quite a white noisy image (right). Nevertheless in the central part quite significant structural features can be discerned.

removal. In the framework of a multiresolution analysis that leads to examine the suprema of the wavelet transform in order to reconstruct with scalet patterns, here called pyrels. The matching pursuit is then done simultaneously for a set of coefficients, while their amplitudes are greater than a threshold which progressively decreases up to a given level. Sufficiently separated suprema are kept at each step. An amplitude correction is also made in order to recover the observed wavelet coefficients at the pyrel locations.

The algorithm (MSMPPy1) was presented for a $1D$ signal, with the use of an pyramidal wavelet transform. The application of the undecimated wavelet transform increased the complexity, but it leads to a shift invariant transform (MSMPAT1). The two-dimensional algorithms (MSMPPy2 and MSMPAT2) are their natural extensions to the two-dimensional field.

MSMPAT1 is quite fast and it can be applied on a large size signal. Nevertheless, MSMPPy1 is more appropriate to process long series. For an image, MSMPPy2 is the useful tool for the analysis of large images, but MSMPAT2 brings a better description. A careful analysis of small images is more convenient with this tool. For these four algorithms a signed decomposition (positive or negative) can be done.

The pyramidal algorithms carry out a sparse decomposition. For example, on the planetary nebula image, a gain around 2 was obtained in the number of coefficients, compared to a classical wavelet thresholding. The different experiments on astronomical images shows that this gain was the more often higher. The image compression is thus a direct application of these algorithms.

These algorithms were built for the analysis of multiband astrophysical images. For each band, it is easy to determine the amplitude of each pyrel taking into account the same identification set. The merging of pyrels into objects allows one to give a full description of the images. A complete multiband vision was then derived [16].

REFERENCES

[1] Donoho, D.L.. *Compressed sensing*. IEEE Trans. on Information Theory, 52:4, 12891306, 2006.

[2] Usevitch, B.E., *A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000*. Signal Processing Magazine, IEEE, 18:5, 22-35, 2001

[3] Welland, G.V. (eds). *Beyond wavelets*. Studies in computational Mathematics, 10, Associated Press, 2003.

[4] Donoho, D.L., Elad, M., Temlyakov, V.N.. *Stable recovery of sparse overcomplete representations in the presence of noise*. IEEE Transactions on Information Theory, 52:1, 6-18, 2006.

[5] Mallat, S., Zhang, Z.. *Matching pursuits with time-frequency dictionaries*. IEEE Transactions on Signal Processing, 41, 3397-3415, 1993.

[6] Pati, Y., Rezaiifar, R., Krishnaprasad, P.. *Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition*. Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers, Nov. 1993.

[7] Feichtinger, H.G., Turk, A., Strohmer, T.. *Hierarchical parallel matching pursuit*. Proc. SPIE Vol. 2302, 222-232, Image Reconstruction and Restoration, Timothy J. Schulz, Donald L. Snyder, Eds, 1994.

[8] Donoho, D.L., Tsaig Y., Drori I., Starck, J.L. *Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit*. IEEE Trans. on Info. Theory, subm. 2006.

[9] Chen, S.S., Donoho, D.L., Saunders, M.A.., *Atomic Decomposition by Basis Pursuit*. SIAM Journal on Scientific Computing, 20:1, 33-61, 1999.

[10] Tropp, J.A., *Greed is good: algorithmic results for sparse approximation*. IEEE Trans. Inform. Theory, 50:10, 2231-2242, 2004.

[11] Gribonval, R., Nielsen, M., *Beyond sparsity: Recovering structured representations by l1l1 minimization and greedy algorithms*. Advances in Computational Mathematics, 28:1, 23-41, 2008.

[12] Mallat, S.G.. *A Theory for Multiresolution Signal Decomposition: The Wavelet Representation*. IEEE, PAMI, 11:7, 674-693, 1989.

[13] Strang, G.. *Wavelets and Dilation Equations: A Brief Introduction*. SIAM Review, 31:4, 614-627, 1989.

[14] Van Cittert, P.H., *Zum einfluss der spaltbreite auf die intensitäts verteilung in spektrallinen II*, Z. Phys., 69, 298, 1931.

[15] Powell, M.J.D., Approximation theory and methods, chap. 7, Cambridge University Press, 1981.

[16] Bijaoui, A., *Fusion images and Analysis of multiband astronomical images*, Astronomical Data Analysis IV, ed. C. Surace, http://www.oamp.fr/conf/ada4/pub/18/DSD/ADAIV_Bijaoui.pdf, 2006.