Sparse representations versus the matched filter

F. Martinelli, J.L. Sanz

Sparse representations versus the matched filter

F. Martinelli
Università degli Studi di Perugia,
via Vanvitelli, 1, I-06123 Perugia, Italy
Istituto di Scienza e Tecnologie dell'Informazione,
Area della Ricerca CNR di Pisa
Via G. Moruzzi, 1, I-56124 Pisa,
Email: martinelli@dipmat.unipg.it

J.L. Sanz
Instituto de Fisica de Cantabria (CSIC-UC),
Av. de los Castros s.n., 39005-Santander, Spain
Email: sanz@ifca.unican.es

***Abstract*—We have considered the problem of detection and estimation of compact sources immersed in a background plus instrumental noise. Sparse approximation to signals deals with the problem of finding a representation of a signal as a linear combination of a small number of elements from a set of signals called dictionary. The estimation of the signal leads to a minimization problem for the amplitude associated to the sources. We have developed a methodology that minimizes the l_p-norm with a constraint on the goodness-of-fit and we have compared different norms against the matched filter.**

I. Introduction

The detection and estimation of intensity of compact objects immersed on a background plus instrumental noise is relevant in different contexts, e. g. astrophysics, cosmology, medicine, teledetection, etc. On the one hand, the use of frequentist methods in astrophysics/cosmology has been proven to be useful: a) the standard matched filter and generalizations [1], [2], [3], [4]; b) continous wavelets like the standard mexican hat [5] and other members of the family [6]; c) filters based on the Neyman-Pearson approach using the distribution of maxima [7]. Some of them have been applied, in the astrophysical context, to real data like WMAP and simulated data [8] for the upcoming experiment on board the Planck satelite. On the other hand, Bayesian methods have also recently developped [9].

Sparse approximations to signals (i.e. the problem of finding a representation of a signal as a linear combination of a small number of elements from an over-complete set of vectors or signals often called a dictionary) in the presence of noise is a subject that has been recently considered in the mathematical literature [10], [11], [12], [13], [14], [15], [16]. The minimization of the norm l_p assuming a constraint on the goodness-of-fit (residuals) is the relevant assumption of this methodology.

We will adopt such a methodology including correlations but the applications, in this paper, will be done only for uniform noise. On the other hand, we are interested in a positive signal that is corrupted by noise. In such a case, we shall consider a dictionary that contains elements in such a way that a sparse representation is always possible with positive coefficients. For instance, for the case we are interested of point sources the dictionary will be constituted by Gaussians with the appropriate normalization.

In all the cases considered, the estimation of the signal leads to a minimization problem for the amplitude associated to the sources. The lagrangian to be minimized has two terms, a quadratic one related to the goodness-of-fit and other that is a regularizer functional related to the l_p-norm.

As a test of the new methodology, we shall compare different norms ($p = 0, 1, 2$) and also versus the use of the matched filter. Regarding detection, we will compare: number of true detections and false detections, reliability, completeness and F-score. Regarding estimation, we will compare the bias, error in amplitude and goodness-of-fit.

We remark the importance of our development of compact source detection/estimation in different areas: astrophysics/cosmology (detection/estimation of emission/absorption lines in spectra of galaxies and quasars in different bands, detection/estimation of galaxies and clusters of galaxies in different bands in 2D-images), medicine (detection/estimation of spots in 2D-images related to tumors in Doppler and X-rays, 3D-images obtained with magnetic resonance), speech (detection/estimation of especial sounds in 1D signals), teledetection (detection of features on 2D-images), etc.

In Section II we comment about l_p norm and matched filter approach.

In Section III we describe our method based on a functional minimization. Finally, in Section IV we comment about the numerical simulations.

II. Sparse Methodology

A. l_p-norm approach

Sparse representations deals with the problem of finding a representation of a signal $s \in \mathbb{R}^N$ in terms of a linear combination of a small number of elements (i.e set of signals or atoms) that is usually called a dictionary:

$$s = \Phi a, \tag{1}$$

where $\Phi \in \mathbb{R}^{N \times n}$ and $a \in \mathbb{R}^n$ is a vector given by the coefficients of the representation. We are interested in the case $n << N$, N representing the number of pixels in an image. Let us consider a set of data described by the $N \times 1$ matrix

d. These data contain a signal linearly corrupted by noise, described by $z \in \mathbb{R}^N$,

$$d = s + z. \quad (2)$$

We assume a Gaussian uniform noise, characterized by zero mean and known variance σ^2. Moreover, we are interested in neither a generic signal nor a dictionary but in a positive signal that is corrupted by noise. In such a case, we shall consider a dictionary that contains elements in such a way that the sparse representation is always possible with positive coefficients. For instance, for the case we are interested of point sources the dictionary, will be constituted by Gaussians with the appropriate normalization

$$\Phi^i_\alpha = A_\alpha e^{-\frac{(i-\alpha)^2}{2R^2}}, \quad (3)$$

where i is the pixel position, α is the center of the Gaussian and

$$A_\alpha = \left[\sum_{k=1}^{N} e^{-\frac{(k-\alpha)^2}{R^2}}\right]^{-1/2}.$$

We will assume a l_p-norm approach subject to the constraint that the error in the residual to be less or equal a fixed value

$$\min_{a>0} ||a||_p \;\; s.t. \;\; \epsilon \leq N\delta, \quad (4)$$

$$||a||_p \equiv \left[\sum_{\alpha=1}^{N} |a_\alpha|^p\right]^{1/p}, \epsilon = ||d-s||^2_{\xi^{-1}} \equiv (d-s)^t \xi^{-1}(d-s),$$

where $\xi = \sigma^2 I_{N\times N}$ represent the correlation matrix of the noise. We will take $\delta = 1$, tipically this means that we have a sparse representation for the sources.

This approach is equivalent to a constrain minimization problem with a penalty term

$$\min_{a>0} L(a) = \min_{a>0} \left[\frac{1}{2}(a^t M a - 2D^t a) + \lambda \|a\|^p_p\right], \quad (5)$$

where

$$M \equiv \Phi^t \xi^{-1} \Phi, \;\; D \equiv \Phi^t \xi^{-1} d, \quad (6)$$

and $\lambda > 0$ is an adequately chosen multiplier satisfying the constrain

$$\epsilon = (d - \Phi a)^t \xi^{-1} (d - \Phi a) = N\delta. \quad (7)$$

The lagragian L is given by the sum of two terms: the first one is a quadratic term related to the goodness-of-fit, the other one is a regularizer functional related to the l_p-norm. If there is not overlapping between the point sources, $M = \frac{1}{\sigma^2} I_{N\times N}$. Hereinafter, we will assume, without loss of generality, that the noise has unit variance ($\sigma = 1$).

The minimization of the constrained lagrangian leads to the following equation

$$\sum_{\beta=1}^{n} M_{\alpha\beta} a_\beta + \lambda p a_\alpha^{p-1} = D_\alpha, \;\; \alpha = 1, \ldots, N, \quad (8)$$

and the constraint dealing with the goodness-of-fit gives the value for the multiplier λ

$$\lambda = [p \sum_{\alpha=1}^{n} a^p_\alpha]^{-1} [f - D^t a - N\delta]. \quad (9)$$

Two necessary conditions to find a sparse representation for the signal s are

$$a_1, ..., a_n > 0; \quad \lambda > 0. \quad (10)$$

Moreover, the number n of sources to be detected, is not fixed a priori. Thus, we will estimate (a, λ) such that n is the minimum number of positive coefficients satisfying 8, 9 and such that λ is a positive parameter.

B. Matched Filter

Given the properties of the noise, a maximum likelihood approach allows to estimate the parameters. For instance, assuming Gaussian noise for the N-probability density function

$$p(d|a) \propto e^{-\frac{1}{2}||d-\Phi a||^2_{\xi^{-1}}}, \quad (11)$$

the log-likelihood is

$$l \equiv -ln\, p(a|d) = c + \frac{1}{2}(a^t M a - 2D^t a), \quad (12)$$

where

$$M \equiv \Phi^t \xi^{-1} \Phi, \;\; D \equiv \Phi^t \xi^{-1} d, \quad (13)$$

and $c \in \mathbb{R}$. Therefore, the minimization of l leads to the least squares solution

$$Ma = D. \quad (14)$$

The previous equations can be interpreted in the following form: D represents the data filtered with a filter adapted to the dictionary $\Phi^t \xi^{-1}$. While the dictionary contains only elements related to the profile of the signal, such a filter, F, becomes proportional to the standard matched filter at n positions

$$F \equiv \Phi^t \xi^{-1}, \quad D = Fd. \quad (15)$$

III. Proposed Algorithm

We estimate (a, λ) alternatively minimizing the lagragian L with respect to the coefficients and evaluating the parameter λ according to the constraint. In particular, we choose (a, λ) as a fixed point of the method

$$\begin{aligned} a &= \arg\min_{x>0} \left[\frac{1}{2}(x^t M x - 2D^t x) + \lambda \|x\|^p_p\right], \\ \lambda &= [p \sum_\alpha a^p_\alpha]^{-1} [f - D^t a - N\delta]. \end{aligned}$$

The chosen minimizing algorithm is the Successive Over-Relaxation (SOR) [17]. This is a gradient descent algorithm, which uses local quadratic approximations to determine optimal step size. In particular, for the iterative minimization of the lagrangian L, the k-th iteration is

$$a_i^k = a_i^{k-1} - \omega \frac{1}{T_i} \frac{\partial L(a)}{\partial a_i},\ \forall\, i = 1, \dots, n, \tag{16}$$

where ω is a parameter governing the speed of convergence and T_i is an upper bound of the second derivative of L. We fix $\omega = 0.5$ and

$$T_i = M_{i,i} + \lambda p(p-1) a_i^{(p-2)}$$

We consider the vector m of positive local maxima of the filtered signal D, sorted in a descending order. We fix n as the minimum integer such that $m_1, \dots, m_n$ satisfy

$$\sum_{\beta=1}^{n} m_\beta + \lambda p m_\alpha^{p-1} = D_\alpha,\ \ \alpha = 1, \dots, n,$$

and

$$\lambda = [p \sum_{\alpha=1}^{n} m_\alpha^p]^{-1} [f - D^t a - N\delta],$$

corresponding to the minimization of the constrained lagrangian in the case of no-overlapping sources. Moreover, we choose the vector $(m_1, \dots, m_n)$ as initial point of the method above.

According to the method described in section II-B, the vector m represents the estimated coefficients with the matched filtered data, thus the amplitude of estimated sources with the matched filter correspond to the positive peaks of the filtered signal.

IV. Simulation

We perform 1D numerical simulations for uniform noise to test the cases $p = 10^{-2}, 1, 2$ against the standard matched filter. We choose a signal of length 2^{12} and we introduce 100 point sources in random positions, the amplitudes ranging from $SNR = 0.5$ to 5 following a uniform distribution. In order to compare the results regarding detection, we determine the number of true and false detections and the reliability; regarding estimation, we take into account the bias and the goodness-of-fit. Figure 1 represent the estimated amplitude versus the number of true detection greater then the fixed amplitude. We can see that for amplitude higher than 2.5, the matched filter and the l_p-norm approach have the same behaviour. For lower amplitudes the matched filter can detect more true sources. However, in this case, the number of false ones is bigger then the number of the simulated sources, as we can see in figure 1. Instead, the l_p-norm approach introduce few false sources also for amplitudes lower then 2.5. This behaviour can be inferred also observing the plot in figure 4(a): reliability is defined as the ratio between the number of true detections and the number of all detections. For the l_p-norm approach, reliability is always greater then 0.9, becoming 0.95 at amplitude 2.5. Figure 4(b) represents

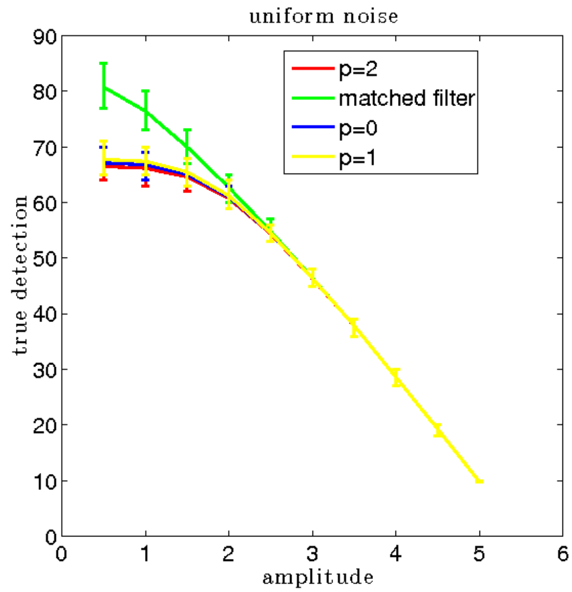

Fig. 1. Number of true detections versus estimated amplitude.

bias over amplitude . Finally in table IV we report the error in the fit for the l_p-norm approach and the matched filter. Finally, taking into account the global performance, we can remark that the sparse representations $p = 10^{-2}, 1, 2$ work similarly regarding reliabilty as compared to the matched filter whereas in this case the bias is lesser, being the worst case $p = 2$.

Summing up, the matched filter and the p-norm give similar results above a $SNR > 2$, below this value the p-norm performs better regarding reliability whereas regarding bias the oposite is found (the $p = 2$ case gives the highest bias and the case $p = 10^{-2}$ is better than $p = 1$). Regarding the goodness-of-fit the matched filter gives a better result but introduces a lot of spurious sources.

	$p = 10^{-2}$	p=1	p=2	MF
Mean value	4116.5	4117.2	4115.2	3727.4
Standard deviation	19.92	24.21	22.03	57.49

TABLE I
Error in the fit: mean value and standard deviation

References

[1] M. Tegmark and A. de Oliveira-Costa, "Removing point sources from cosmic microwave background maps," *Astrophysical Journal Letters*, vol. 500, no. L83, 1998.

[2] J. L. Sanz, D. Herranz, and E. Martínez-Gónzalez, "Optimal detection of sources on a homogeneous and isotropic background," *The Astrophysical Journal*, vol. 552, pp. 484–492, 2001.

[3] D. Herranz, J. L. Sanz, M. P. Hobson, R. B. Barreiro, J. M. Diego, E. Martínez-González, and A. N. Lasenby, "Filtering techniques for the detection of sunyaev-zel'dovich clusters in multifrequency maps," *Monthly Notices of the Royal Astronomical Society*, no. 336, pp. 1057–1068, 2002.

[4] D. Herranz and J. L. Sanz, "Matrix filters for the detection of extragalactic point sources in cosmic microwave background images," *IEEE Journal of Selected Topics in Signal Processing*, no. in press, 2008.

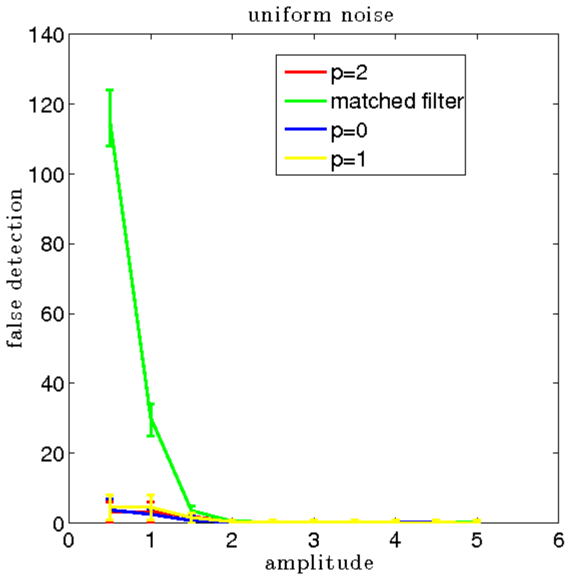

Fig. 2. Number of false detections versus estimated amplitude.

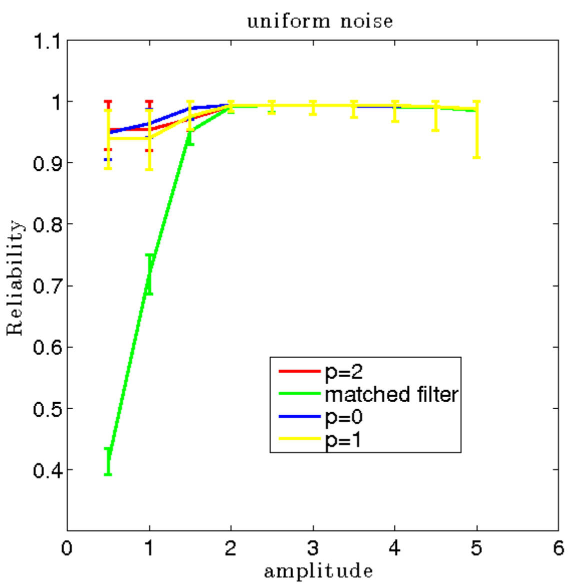

Fig. 3. Reliability versus estimated amplitude.

[5] L. Cayón, J. L. Sanz, R. B. Barreiro, E. Martínez-González, P. Vielva, L. Toffolatti, J. Silk, J. M. Diego, and F. Argeso, "Isotropic wavelets: a powerful tool to extract point sources from cmb maps," *Monthly Notices of the Royal Astronomical Society*, vol. 315, pp. 757–761, 2002.

[6] J. González-Nuevo, F. Argeso, M. López-Caniego, L. Toffolatti, J. L. Sanz, P. Vielva, and D. Herranz, "The mexican hat wavelet family: application to point-source detection in cosmic microwave background maps," *Monthly Notices of the Royal Astronomical Society*, no. 369, pp. 1603–1610, 2006.

[7] M. López-Caniego, D. Herranz, R. B. Barreiro, and J. L. Sanz, "Filter design for the detection of compact sources based on the neyman-pearson detector," *Monthly Notices of the Royal Astronomical Society*, vol. 359, pp. 993–1006, 2005.

[8] M. López-Caniego, D. Herranz, J. González-Nuevo, R. B. B. J. L. Sanz, P. Vielva, F. Argeso, and L. Toffolatti, "Comparison of filters for the detection of point sources in planck simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 370, pp. 2047–2063, 2006.

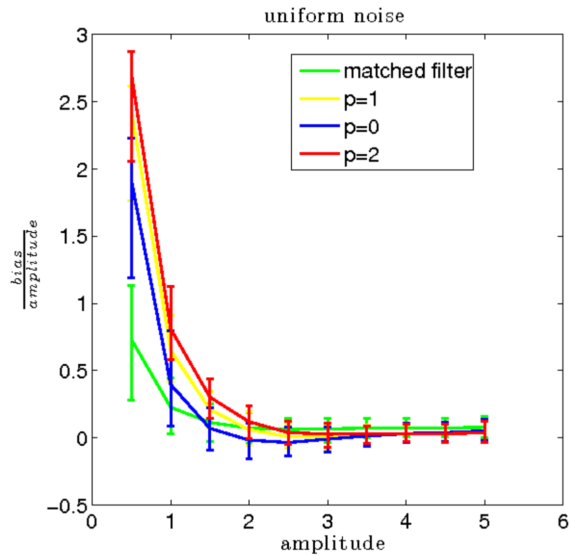

Fig. 4. Bias over amplitude versus estimated amplitude.

[9] P. Carvalho, G. Rocha, Graca, and M. P. Hobson, "A fast bayesian approach to discrete object detection in astronomical datasets - powellsnakes i," *ArXiv e-prints, arXiv:0802.3916v1*, vol. 802, 2008.

[10] J.-J. Fuchs, "Recovery of exact sparse representations in the prsence of noise," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. II, pp. 533–536, 2004.

[11] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, no. 59, pp. 1207–1223, 2006.

[12] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Processing Letters*, vol. 13, pp. 409–412, 2006.

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, no. 52, pp. 6–18, 2006.

[14] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Information Theory*, pp. 1030–1051, 2006.

[15] A. K. Fletchera, S. Rangan, V. K. Goyal, and K. Ramchandran, "Denoising by sparse approximation: error bounds based on rate-distortion theory," *Eurasip Journal of Applied Signal Processing*, pp. 1–19, 2006.

[16] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," in *Proc. Allerton Conference on Communication, Control and Computing*, 2006.

[17] A. Blake and A. Zissermann, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.