

# DETECTION AND ESTIMATION OF SUPERIMPOSED SIGNALS.

Jean-Jacques FUCHS

IRISA/Université de Rennes I  
Campus de Beaulieu - 35042 Rennes Cedex - France  
fuchs@irisa.fr

## ABSTRACT

The problem of fitting a model composed of a number of superimposed signals to noisy observations is addressed. An approach allowing us to evaluate both the number of signals and their characteristics is presented. The idea is to search for a parsimonious representation of the data. The parsimony is insured by adding to the maximum likelihood criterion a regularization term built upon the  $\ell_1$ -norm of the weights. Different equivalent formulations of the criterion are presented. They lead to appealing physical interpretations. Due to limited space, we can only sketch here an analysis of the performance of the algorithm that has been successfully applied to different classes of problems [6],[7].

## 1. INTRODUCTION

In many different domains, one faces the problem of fitting a model composed of a small number of superimposed signals to noisy observations. Among the problems that belong to this set one can cite the detection and estimation of the parameters of closely spaced sinusoids in noise, the estimation of width, position (delay) and amplitude of superimposed pulses of known shape in noise. If the additive noise is gaussian and white (or of known covariance), one can solve the problem using the maximum likelihood (ML) approach. But this approach has at least two drawbacks : it requires the prior knowledge of the number of signals that are present and, in difficult situations, a precise initial point must be known. This second drawback is indeed the major issue since the search for a good initial point, which is needed because the ML function has many local extrema, is extremely time consuming and it is this initialization procedure that is the crucial problem. The local improvement achieved by the ML criterion is of marginal interest.

Most of the iterative algorithms including gradient type searches [1] or expectation-maximize type algorithms will converge to an extremum that will essentially depend upon the initial point that was used. A dynamic programming approach [2] has been recently designed to circumvent this difficulty, it heavily relies on a local (limited) interaction signal model (LISMO) that can be seen as an *ad hoc* assumption in a dynamic programming context. Its robustness with respect to this assumption is evaluated in [3] and appears quite reasonable.

Our approach is completely different, it amounts to rewrite the problem as a quadratic program with essentially two parameters to tune. It similarly relies on a *ad hoc* assumption which somehow measures the difficulty of a

scenario with respect to this approach and seems to be satisfied reasonably often. In a multipath time-delay estimation context [6], it seems to outperform other approaches. It is worth noting that a whole set of bayesian approaches [5] developed in general to handle problems from geophysics actually address similar problems with yet fully different techniques. The aim there is to restore sparse spike trains distorted by a linear system in additive noise.

## 2. THE MODEL

One observes a noisy vector that can be modeled as the sum of an unknown number  $P$  of signals belonging to one or several parametrized families of known vector functions and white noise  $e$ . In the simplest case, one observes a  $m$ -dimensional vector  $\hat{b}$  :

$$\hat{b} = \sum_{i=1}^P \alpha_p a(\tau_p) + e = b + e$$

where  $a(\tau)$  is a known family of vectors parametrized by a scalar  $\tau$  and  $e$  is the white noise vector with covariance matrix  $\sigma_e^2 I$ . The objective is to recover the deterministic part  $b$  of  $\hat{b}$  i.e. to estimate  $P$  the number of components and to identify the parameters  $\theta_p = (\alpha_p, \tau_p)$ . It is assumed that  $P \ll m$  and that  $\tau$  belongs to a known compact set. The different difficulties one can face in this type of problem are the detection of weak isolated components (low signal to noise ratios), the resolution of several closely spaced components (close  $\tau$ 's), the detection of a weak component hidden by a stronger one, and any combination of these difficulties.

Without loss of generality, we consider that the euclidean norm of  $a(\tau)$  is independant of  $\tau$  and equal to one, that the order of magnitudes of the amplitudes  $\alpha$ 's is around one so that one should think of  $\sigma_e$  as being quite small.

Let us further note that, since  $m$  the dimension of the observation vector is constant, evaluate the asymptotic performance of an estimator amounts, in this setting, to evaluate it as the signal to noise ratio (SNR) goes to infinity and thus as  $\sigma_e \rightarrow 0$ . Indeed, one can think of  $\hat{b}$  as being a vector estimated from  $T$  scalar observations,  $e$  is then the estimation error whose variance is of order  $1/T$  that goes to zero as  $T \rightarrow \infty$ .

If the noise  $e$  is assumed to be gaussian, the ML estimates of the parameters  $\theta_k = (\alpha_p, \tau_p)$  are obtained by solving the following non-linear least-squares problem :

$$\min_{\alpha, \tau} \|\hat{b} - \sum_{p=1}^q \alpha_p a(\tau_p)\|_2^2 \quad (\text{ML})$$

In this approach one needs to fix  $q$  the number of components and, for difficult scenarios, an initial point close to the global optimum must be known. If the number  $q$  of fitted components is larger than the true number  $P$ , the additional degrees of freedom are used to somehow model the noise realization. It is easy to check that the amplitudes associated with such false components will in our case be of the order of  $\sigma_e$  and that the decrease each induces on the value of the criterion is of the order of  $\sigma_e^2$ . These values are obtained from standard regression analysis and are in agreement with the widely used Akaike's information criterion that could be used to estimate the true number of components in a ML type approach.

### 3. THE CRITERION

In order to transform the non-linear least-square problem (ML) into an easier one, we propose to uniformly discretize the values of  $\tau$  over its compact domain. We denote  $h$  the discretization step and  $n$  the number of resulting potential values of  $\tau$ . With the  $n$  vectors  $a_j$  (where  $a_j = a(\tau_j)$  with  $\tau_j = jh + \tau_0$ , for instance), we build a matrix  $A$  of dimension  $(m, n)$  with  $n \gg m$  and we propose to solve the following quadratic program, for an adequately tuned positive real parameter  $\lambda$  :

$$\min_X \|\hat{b} - AX\|_2^2 + \lambda \|X\|_1 \quad (\text{QP})$$

where  $\|X\|_1$  denotes the  $\ell_1$ -norm :  $\|X\|_1 = \sum_1^n |x_i|$ . This can indeed be transformed into a quadratic program and its unique global minimum can be obtained using standard algorithms available from any scientific program library. We will assume that any subset of  $m$  columns in  $A$  forms a set of linearly independent vectors. This is not to be considered as a limitation. As a matter of fact, there is a huge gap between the ML criterion (ML) and this one (QP) and the link between the minimum of (QP) and the problem under investigation may not be obvious. Since there are now more unknowns than observations  $AX = \hat{b}$  has an infinite number of solutions. The  $\ell_1$ -norm in the additive term in (QP) is there to select sparse solutions and the parameter  $\lambda$  must further be tuned to possibly select the sparse and true solution. The two parameters to be tuned are  $\lambda$  and the discretization step  $h$ . The value of  $h$  will mainly be fixed by the accuracy we want to achieve.

#### 3.1. About the optimum of QP

One can make the following remarks about the solution  $X^*$  of (QP) as  $\lambda$  goes from 0 to  $+\infty$ .

- For  $\lambda = 0$ , one is left with  $\min_X \|\hat{b} - AX\|_2^2$ , and since there are more unknowns than equations ( $n > m$ ), the value of the minimum is zero and it is attained for all points in a convex set (a linear variety), some of them having at most  $m$  non-zero components.

- For  $\lambda = 0^+$  i.e. for  $\lambda$  positive and arbitrarily close to zero, the solution is attained at the point(s) in the previous set having least  $\ell_1$ -norm (we do not prove this). Again there is a solution having at most  $m$  non-zero components.

- For all  $\lambda \geq 2 \|A^T \hat{b}\|_\infty$ , the solution is  $X^* = 0$ . This is an easy consequence of a result presented below in section 3.2. One can easily check that for such  $\lambda$ 's, by moving a

component  $x_i$  away from zero, the decrease in the  $\ell_2$ -term in (QP) is lower than the increase in the  $\ell_1$ -term.

- For  $\lambda$  between these two values, there might be solutions having any number of non-zero components and namely  $P$  or better  $2P$  non-zero components. We seek this type of solution.

Remember that  $P$  denotes the number of components in the scenario, since these components will generically all fall in between two points of the discretization grid, the two neighboring columns in  $A$ , at the least, will be needed to approximatively reconstruct each component. In fact the interesting property of this criterion, that is due to the presence of the  $\ell_1$ -norm in the additive term, is that it favors among the many different ways to approximatively reconstruct a non-existing component, the way that uses just the two neighbors. The value of  $\lambda$  actually allows us to tune the degree with which an approximate reconstruction is permitted. For  $\lambda = 0$ , the reconstruction must be exact.

The best we can thus expect from our approach when applied to a scenario with  $P$  components is to get a solution  $X^*$  with  $2P$  non-zero (or significant) components appearing by pairs (neighboring columns). To deduce from such an  $X^*$  the estimates of the parameters  $\theta_p$ , we then associate which each pair of non-zero components, an amplitude  $\hat{\alpha}_p$  equal to their sum and an estimate  $\hat{\tau}_p$  obtained by linear interpolation of the "indices" of the two columns. The transformation that converts the optimum  $X^*$  into estimates  $\hat{\theta}$  is denoted  $G(\cdot)$  i.e.  $\hat{\theta} = G(X^*)$  and will be detailed later.

#### 3.2. Optimality conditions for QP

The criterion (QP) is convex but not continuously differentiable. A necessary and sufficient condition for  $X^*$  to be a global optimum is that the vector 0 is a sub-gradient of the criterion at  $X^*$  [8]. A vector  $\gamma$  is a sub-gradient of  $f$  at  $X^*$  if  $f(X) \geq f(X^*) + \gamma^T(X - X^*)$ . If the vector 0 is a sub-gradient at  $X^*$  then  $f(X) \geq f(X^*)$  and  $X^*$  is clearly a minimum.

Since (QP) is non-smooth at zero only, one must distinguish in  $X^*$  the zero components from the non-zero components. For the non-zero components, we denote  $\bar{X}^*$ , the sub-gradient is unique and equal to the gradient. Replacing then  $\|X\|_1$  by  $X^T \text{sgn} X$  (where  $\text{sgn} X$  stands for sign of  $X$ ) and denoting  $\bar{A}^*$  the columns in  $A$  associated with  $\bar{X}^*$ , we get by nulling the gradient of (QP) with respect to  $\bar{X}^*$  :

$$-2\bar{A}^T(\hat{b} - \bar{A}\bar{X}^*) + \lambda \text{sgn}\bar{X}^* = 0$$

which we rewrite :

$$\bar{A}^T(\hat{b} - \bar{A}\bar{X}^*) = \lambda/2 \text{sgn}\bar{X}^*$$

Omitting some details, this leads to the following expression for the non-zero components of  $X^*$  :

$$\bar{X}^* = (\bar{A}^T \bar{A})^{-1} \bar{A}^T \hat{b} - \lambda/2 (\bar{A}^T \bar{A})^{-1} \text{sgn}\bar{X}^*$$

Note that the solution of (QP) can only be obtained through an iterative search. This is by no means a relation giving the solution. It is only an implicit relation since  $\bar{X}^*$  appears on both sides. It nevertheless allows to compute the optimum  $X^*$  if one knows beforehand which components will be non zero and their signs. The second term in this expression is a bias term that is due to the penalty term in (QP). Once

$\bar{X}^*$ , and thus  $\bar{A}^*$ , are known, it is easy to compute this bias and to subtract it from  $X^*$ . We will always do so when applying this scheme.

The vector 0 is a sub-gradient for the zero components in  $X^*$ , if the criterion increases when they are taken non zero. This is the case if the absolute value of the partial derivative of the quadratic term in (QP) is smaller than  $\lambda$ :

$$|a_j^T(\hat{b} - \bar{A}\bar{X}^*)| = |a_j^T(\hat{b} - A X^*)| < \lambda/2 \quad \forall a_j \notin \bar{A}$$

The following condition thus holds for all the components:

$$\|A^T(\hat{b} - AX^*)\|_\infty \leq \lambda/2 \quad (C)$$

where  $\|X\|_\infty$  denotes the  $\ell_\infty$ -norm,  $\|X\|_\infty = \max_{1 \leq i \leq n} |x_i|$ .

Condition (C) holds with equality for the "active" columns of  $A$  (those in  $\bar{A}$ ) and with inequality for the others.

One can establish that the optimum of (QP) also solves:

$$\min_X \|AX\|_2^2 \quad \text{s.t.} \quad \|A^T(AX - b)\|_\infty \leq \frac{\lambda}{2}$$

(We do not prove this.) The criterion (QP) thus selects the solution  $X^*$  which minimizes  $\|AX\|_2^2$  and achieves (C), a condition on the outputs of the matched filter when applied to the residues. It has an interesting (and quite illuminating) physical interpretation. After having subtracted  $AX^* = \bar{A}\bar{X}^*$  from the observations  $\hat{b}$ , the residues are such that their correlation with the columns of  $A$  (the output of a matched filter) never exceeds  $\lambda/2$ . This seems to be a far more sensible way to match the observations than the one used in matching pursuit type algorithms [9], which work in an iterative way and subtract successively the best matching component from the current residues until a criterion (threshold) is satisfied. Here a similar result is achieved in a single shot.

### 3.3. The noise free case

Even in the absence of noise, nothing guaranties that the true scenario can be retrieved from the minimum of our criterion for an adequate value of  $\lambda$ . Let us consider the quite simple case where  $\hat{b} = b = \sum_1^P \alpha_p a_p = A_o X_o = \bar{A}_o \bar{X}_o$  i.e. there is no noise present and all the  $P$  components of the scenario are columns of  $A$ . At best, we then expect that there is a value of  $\lambda$  for which the solution  $X^*$  to (QP) is a biased version of  $X_o$  i.e. the non-zero components of  $X_o$  and  $X^*$  are in the same position and have the same sign.

We have seen above that the optimum of (QP) is completely defined if one knows the indices and the signs of the non-zero components. Introducing this knowledge into the relations giving the non-zero components  $\bar{X}^*$  of the solution  $X^*$  we obtain the following expression for the solution we seek:

$$\bar{X}^* = \bar{X}_o - \lambda/2 (\bar{A}_o^T \bar{A}_o)^{-1} \text{sgn} \bar{X}_o$$

with  $\bar{X}_o = \bar{A}_o^+ b$  where  $\bar{A}_o^+ = (\bar{A}_o^T \bar{A}_o)^{-1} \bar{A}_o^T$  denotes the pseudo-inverse of  $\bar{A}_o$ . It remains to check if this solution completed by zeroes satisfies the  $n$  conditions in (C).

They hold by construction for the  $P$  rows in  $\bar{A}_o^T$ , let us check the  $n - P$  other rows of  $A^T$ . Replacing  $\bar{X}^*$  by its expression given above, the following expression is obtained:

$$|a_j^T \bar{A} (\bar{A}_o^T \bar{A}_o)^{-1} \text{sgn} \bar{X}_o| \leq 1 \quad \text{for} \quad a_j \notin \bar{A}_o$$

With  $d = \bar{A}_o (\bar{A}_o^T \bar{A}_o)^{-1} \text{sgn} \bar{X}_o$ , the last relation becomes:  $|a_j^T d| \leq 1$ . Note that  $\lambda$  is no longer present in these conditions and only appears in the expression of  $\bar{X}^*$ .

In summary, for (QP) to yield the *good* solution in this case the following conditions have to be satisfied:

- the scenario dependent vector  $d = \bar{A}_o (\bar{A}_o^T \bar{A}_o)^{-1} \text{sgn} \bar{X}_o$  must be such that  $|a_j^T d| \leq 1$  for  $a_j \notin \bar{A}_o$ ,

- $\lambda$  has to be chosen (small enough) such that  $\text{sgn} \bar{X}_o = \text{sgn} \bar{X}^*$ . The first of these two conditions is an identifiability conditions that depends on the difficulty of the scenario. It can be given a geometric interpretation in terms of a *separating hyperplane*, the vector  $d$  defines a hyperplane that must separate the true columns in  $\bar{A}_o$  from the other columns: for the true columns it satisfies  $|\bar{A}_o^T d| = 1$  while for the others it should satisfy  $|a_j^T d| < 1$ .

If the scenario dependent vector  $d$  satisfies this property, there is a whole domain of  $\lambda$  that yields the *good* solution, i.e. for which the true columns in  $A$  are selected with the good signs. The fact that this *good* solution  $X^*$  is indeed biased is not hampering since the bias is easily removed.

### 3.4. The general case

Let us now sketch what happens in the general case, for  $\hat{b} = \sum_1^P \alpha_p a(\tau_p) + e$ . In this case, the sought-for solution is completely defined by selecting for each true component  $a(\tau_p)$  the two neighboring columns in  $A$  and by assigning them the sign of  $\alpha_p$ .

*Strictly speaking, this is only valid if  $\sigma_e$  and  $h$  are small enough, since otherwise the noise may be such that the best (ML) estimate of a component no longer lies between the so-defined two columns.*

One proceeds as in the section above. There are two additional terms in  $X^*$ , one induced by the noise, the other one by the reconstruction error. Due to limited space we do not enter into the details. The conclusion is essentially the same as in the section above. If an hyperplane (defined in a similar way) is *separating*, there exists a domain in  $\lambda$  for which the optimum of the criterion has the sought-for property i.e. has its non-zero components with the right sign in the correct positions. The lower bound of the domain is now no longer zero but depends, among other factors, on the noise.

## 4. CHOICE OF THE DISCRETIZATION STEP

We already indicated that the discretization step  $h$  had to be chosen small enough to be able to achieve the desired accuracy. We choose it such that it does not preclude our method to attain the Cramer Rao bounds. We analyze this point in this section.

### 4.1. Linear interpolation and linear approximation

For a component  $a_\tau$ , let  $\bar{A}$  be the matrix built upon the two neighboring columns that are associated with the parameter values  $\tau_o$  and  $\tau_1$ , where  $\tau_o < \tau < \tau_1$  and  $\tau_1 - \tau_o = h$ . The component  $a_\tau$  can then be approximatively reconstructed by linear interpolation:

$$a_\tau \simeq \frac{\tau_1 - \tau}{\tau_1 - \tau_o} a_o + \frac{\tau - \tau_o}{\tau_1 - \tau_o} a_1 = \bar{A} \bar{X}_l$$

For a regular enough vector function  $a(\tau)$ , the reconstruction error  $\delta_\tau = a_\tau - \bar{A} \bar{X}_l$  is easily shown to be of order  $h^2$  and to attain its maximum for  $\tau = (\tau_1 + \tau_o)/2$ .

But one can also consider the best linear approximation which minimizes:  $\|a_\tau - \bar{A}\bar{X}\|^2$ . The optimum is attained for  $\bar{X}_o = \bar{A}^+ a_\tau$ , and the reconstruction error denoted  $\Delta_t$  is at most of order  $h^2$ , since it is smaller than  $\delta_t$ .

#### 4.2. Bias in $\hat{\theta}$

For the ease of exposition, we consider the simplest (single component) case  $\hat{b} = \alpha a_\tau$ . Applying our scheme to  $\hat{b}$ , we seek a solution that selects the two neighboring columns. With  $\bar{A}$  the matrix built upon these two columns, we have the following expression for the two non-zero component of the optimum of (QP):  $\bar{X}^* = \alpha \bar{A}^+ a_\tau - \frac{\lambda}{2} (\bar{A}^T \bar{A})^{-1} [1 \ 1]^T$ . We have then to associate with  $\bar{X}^*$  an estimate  $\hat{\theta} = G(\bar{X}^*)$  of  $\alpha$  and  $\tau$ . Before we do so, we remove the bias (second) term in  $\bar{X}^*$  and define  $\bar{X}_u^* = \alpha \bar{A}^+ a_\tau$  where the subscript  $_u$  stands for unbiased. The expression of the transformation  $G(\cdot)$  is :

$$\hat{\theta} = \begin{bmatrix} \hat{\alpha} \\ \hat{\tau} \end{bmatrix} = \begin{bmatrix} x_o + x_1 \\ \frac{x_o \tau_o + x_1 \tau_1}{x_o + x_1} \end{bmatrix} = G(\bar{X}_u^*)$$

Our aim is to estimate the order of magnitude of the error  $\tau - \hat{\tau}$  as a function of  $h$ . Using Taylor series expansion, it is lengthy but not difficult to show that this error is of order  $h^2$ .

#### 4.3. Variance in $\hat{\theta}$

Here we analyze the effect upon the variance of the estimates of the discretization of  $\tau$ . For a single or isolated component  $\hat{b} = \alpha a_\tau + e$  the ML estimates are obtained through:  $\min_{\theta} \|\hat{b} - \alpha a(\tau)\|^2$ . In our approach, using the same notation as above,  $\bar{X}_u^*$  will at best realize:  $\min_{\bar{X}} \|\hat{b} - \bar{A}\bar{X}\|^2$ . We assume again that  $\lambda$  is such that only the two neighboring columns are selected and describe here the result obtained after removal of the bias in  $\bar{X}^*$ .

The analysis is more intricate than for the bias, we propose to write  $\alpha a_\tau = \bar{A}\bar{X}_l + \delta_t$  where  $\bar{X}_l$  is defined in section 4.1 and is such that  $G(\bar{X}_l)$  yields the exact parameter. The unbiased solution can then be written:  $\bar{X}_u^* = \bar{X}_l + \bar{A}^+ \delta_t + \bar{A}^+ e$ . The classical error term, that yields the standard deviation of the estimates, is the last one, it is of order  $\sigma_e$ . The second term is the one due to the discretization. He can be made negligible with respect to the third by choosing  $h = O(\sigma_e)$  since (section 4.1) it is of order  $h^2$ .

Since this same choice reduces the deterministic bias term analyzed above to a negligible quantity, we decide to take  $h = O(\sigma_e)$  in our scheme.

One can further show that, at least in this simple case and provided -as assumed above- that the true columns are selected by our procedure, the estimate  $\hat{\theta} = G(\bar{X}_u^*)$  attains then the Cramer Rao bounds.

### 5. HOW TO TUNE $\lambda$

We have indicated in section 3.4 that provided the scenario was *separable*, there is a domain of  $\lambda$  for which only the "true" columns have non-zero weights. Remember that "true" means in general for each component present in  $b$  the two neighboring columns. The analysis we have sketched there, indeed yields a lower bound for this domain that

depends mainly upon the noise standard deviation and an upperbound that depends upon the signal to noise ratio.

A straightforward way to arrive at the same qualitative conclusion is as follows. The first step is to note that (QP) is equivalent to the following optimization problem :

$$\min_X \|AX - \hat{b}\|_2^2 \quad \text{s.t.} \quad \|X\|_1 \leq B$$

this holds because both problems are convex and the Lagrangien of this problem is nothing but (QP). There is an unknown implicit relation between  $\lambda$  in (QP) and the bound  $B$  above. For a given  $B$ , the value of  $\lambda$  that renders (QP) equivalent to this problem, is thus the value of the Lagrangien multiplier of this problem at its optimum. But the Lagrangien multiplier of this problem has an interpretation in terms of sensitivity of the value of the minimum with respect to a variation of the second member of the constraint. Using then the order of magnitudes indicated at the end of section 2 for spurious components in the ML approach, we obtain that for  $\lambda = O(\sigma_e)$ , the optimum  $X^*$  of (QP) should just have a few small ( $O(\sigma_e)$ ) components besides all the true ones.

### 6. CONCLUDING REMARKS

Due to lack of space, the analysis, as it is presented here, is incomplete. Many details have been left out and no simulations are presented. The performance of the approach has however already been tested in [4],[6] and the interested reader can check these references for the implementation issues and performance evaluation of similar schemes.

### 7. REFERENCES

- [1] B.M. Bell and T.E. Ewart. Separating multipaths by global optimization of a multidimensional matched filter. *IEEE-T-ASSP*, 37:378-391, mar. 1986.
- [2] S.F. Fau and Y. Bresler. ML parameter estimation of superimposed signals by dynamic programming. *IEEE-T-SP*, 41:804-819, feb. 1993.
- [3] S.F. Fau and Y. Bresler. On the robustness of parameter estimation of superimposed signals by dynamic programming. *IEEE-T-SP*, 44:2825-2836, nov. 1996.
- [4] J.J. Fuchs. Extension of the Pisarenko method to sparse linear arrays In *Proc. ICASSP*, III, pp. 2100-2103, may 1995. An extended version will appear in *IEEE-T-SP* Oct. 1997.
- [5] F. Champagnat, Y. Goussard and J. Idier. Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE-T-SP*, 44:2988-2988, dec. 1996.
- [6] J.J. Fuchs. Multipath time-delay estimation. In *Proc. ICASSP*, I, pp. 527-530, Munich, April 1997. An extended version is submitted to *IEEE-T-SP* Mar 1997.
- [7] N. Moal and J.J. Fuchs. Sinusoids in white noise : a quadratic programming approach. submitted to *Proc. ICASSP*, oct. 1997.
- [8] M. Minoux. *Programmation mathématique : théorie et algorithmes. tome 1* Dunod, Paris, 1983.
- [9] S.G. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE-T-SP*, 41: 3397-3415, Dec. 1993.