

EXTENSION OF THE GLOBAL MATCHED FILTER TO STRUCTURED GROUPS OF ATOMS: APPLICATION TO HARMONIC SIGNALS.

Jean-Jacques FUCHS

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

ABSTRACT

When seeking a sparse representation of a signal on a redundant basis, one might want to convey available a priori information on the observations to the optimization criterion. If one observes a sum of harmonic signals in noise, taking into account the structure of each individual harmonic signal definitely improves the efficiency of an estimator of the fundamental frequencies. More or less efficient or elegant solutions to these problems have been proposed, modifying the penalty term in a sparse representation criterion is one of them. We show how to translate prior information by modifying the penalization term of the usual $\ell_2 - \ell_1$ regularized criterion, we indicate how to tune the corresponding hyper-parameters by forming the dual of these modified criterion and we evaluate the associated performance on the sum of harmonics example.

Index Terms— Sparse representations, structured sparsity, duality, harmonic signals, fundamental frequencies estimation.

1. INTRODUCTION

Sparse representation approaches enjoy considerable popularity in signal and image processing. The approach consists in decomposing a signal into a small number of components chosen from a user-designed over-complete set of vectors. It is either used to obtain a simple and sparse approximate model of a complex signal for denoising, compression or coding purposes in audio or video signal processing, or to solve linear inverse problems for which one knows that the solution is indeed sparse. Identification or parameter estimation problems can often be cast into the second set of applications together with direction of arrivals estimation, source separation and many other problems. In statistics, this technique has been developed independently to select variables in the basic linear regression model and the two areas contribute to the development of new algorithms. Of particular interest is presently the possibility to introduce some structure into the redundant dictionary or equivalently to favor groups in the selection [1, 2]. Once an atom from such a group has been selected, prior information tells us that the whole group should potentially be included into the selection. However, since the same atom may belong to several groups, some flexibility has to be preserved. These issues are considered in this paper. In section II, we introduce the modified criterion we will be using, we then evaluate the dual criterion in section III and indicate how to tune the hyper-parameters in the modified criterion. In section IV, an application is proposed, where one identifies the fundamental frequencies of a sum of harmonic signals in white noise.

2. THE MODIFIED CRITERION

2.1. Preliminary remarks

Given an observation $b \in R^n$ one seeks a sparse representation of b , in terms of the columns a_k of a $n \times m$ matrix A , with $m \gg n$. Provided A is a full row-rank matrix, there are an infinity of representations x such that $b = Ax$ and to select a sparse one, one solves the quadratic program :

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + h \|x\|_1, \quad h > 0. \quad (1)$$

where $\|x\|_p$ denotes the ℓ_p norm of a vector x , $\|x\|_p = [\sum_1^m |x_j|^p]^{1/p}$ for $p \geq 1$. This criterion known as LASSO [3] in statistics, Basis Pursuit Denoising [4] in applied mathematics, was introduced simultaneously to the Signal Processing community as the Global Matched Filter (GMF) [5, 6]. In (1), the ℓ_2 -part is clearly linked to the Maximum Likelihood criterion in the Gaussian noise case and the ℓ_1 -penalty term together with the hyper-parameter h allows to monitor the sparsity of the representation, i.e. the number of non zero components in the optimal x . The Lagrangian dual of (1) [7, 6] has a nice physical interpretation

$$\min_x \|Ax\|_2^2 \quad \text{subject to} \quad \|A^T(Ax - b)\|_\infty \leq h \quad (2)$$

which lead to the name GMF and permits to tune h in a justified way by establishing a link with the Generalized Likelihood Ratio Test (GLRT) [8]. For the atoms, i.e. the columns a_k of A , to play an identical role, it already follows from (2) that they need to have the same ℓ_2 -norm. In the sequel they are assumed to be normed to one.

2.2. Non-overlapping groups

In many applications in signal processing, one might want to introduce links between atoms, to favor groups of atoms or to introduce hierarchies. In the simplest case, one groups the atoms in A into non-overlapping sets. Without loss of generality and to simplify the notations, we consider non-overlapping pairs as an example of the case of disjoint sets. If x is m -dimensional with $m = 2p$, one gets

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + h \sum_{k=1}^p (x_{2k-1}^2 + x_{2k}^2)^{1/2}. \quad (3)$$

Introducing $A_k = [a_{2k-1} \ a_{2k}]$ and $X_k = [x_{2k-1} \ x_{2k}]^T$, (3) becomes

$$\min_x \frac{1}{2} \|b - \sum_{k=1}^p A_k X_k\|_2^2 + h \sum_{k=1}^p \|X_k\|_2. \quad (4)$$

An example where this is definitely of interest is the basic sum of sinusoids plus white noise model, where given a observation vector b one wants to determine the number of sinusoids and their characteristics (amplitude, frequency and initial phase). For this example and using (1), to solve the non-linear initial phase problem, one has to resort to a_k vectors that are sines spread over a two dimensional grid (frequency, initial phase) unless one uses (3) and considers pairs A_k of (sine, cosine) vectors over a one-dimensional frequency grid.

The penalty term in (4) with a (non-squared) ℓ_2 -norm is indeed an efficient way to induce sparsity at the level of the groups (pairs) but of course not inside the weights attributed to the elements of a group. One can deduce from results established below in section 3 that the dual of (4) is:

$$\min_x \|Ax\|_2^2 \quad \text{subject to} \quad \|A_k^T(Ax - b)\|_2 \leq h \quad \forall k. \quad (5)$$

2.3. Overlapping groups

For more generality and to be able to handle a larger set of applications, overlapping groups are of interest. To keep the exposition simple, we consider overlapping pairs of (contiguous) vectors. For x of dimension m , one gets

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + h \sum_{k=1}^{m-1} (x_k^2 + x_{k+1}^2)^{1/2}.$$

In this criterion, all unknowns x_k but the two extreme ones, appear twice in the penalty term. A different criterion is indeed obtained if one duplicates these unknowns as in

$$\min_{x, x'} \frac{1}{2} \|b - a_1 x_1 - \sum_{k=2}^{m-1} a_k (x'_k + x_k) - a_m x'_m\|_2^2 + h \sum_{k=1}^{m-1} (x_k^2 + x_{k+1}^2)^{1/2}. \quad (6)$$

This amounts to get back to the non overlapping case by duplicating the vectors in A and the unknowns in x to get C and y , with

$$C = [a_1 \ a_2 \ a_2 \ a_3 \ \dots \ a_{i-1} \ a_i \ a_i \ a_{i+1} \ \dots \ a_{m-1} \ a_{m-1} \ a_m]$$

$$y = [x_1 \ x'_2 \ x_2 \ x'_3 \ \dots \ x_{i-1} \ x'_i \ x_i \ x'_{i+1} \ \dots \ x'_{m-1} \ x_{m-1} \ x'_m]^T.$$

This is not forbidden since the A matrix is never full column rank anyway. It is easy to prove that if, at the optimum, the weights x_k and x'_k associated with the same column in A are non zero, then they have the same sign. The dual of (6) is

$$\min_z \|Az\|_2^2 \quad \text{subject to} \quad \|B_k^T(Az - b)\|_2 \leq h \quad \forall k. \quad (7)$$

with $B_k = [a_k \ a_{k+1}]$, $k = 1$ to $m - 1$ and z of (the reduced initial) dimension m .

2.4. Other structures

The previous section is actually representative for quite general overlapping groups, let us consider now a mixture of (1) and (3)

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + h_1 \|x\|_1 + h_2 \sum_{k=1}^p (x_{2k-1}^2 + x_{2k}^2)^{1/2}, \quad (8)$$

that favors both individual atoms and pairs of non overlapping atoms. As opposed to the criterion (3) or (6) this criterion really favors sparsity of the solution since one allows again an isolated atom to be selected without appearing necessarily in a pair.

In (8) each unknown appears twice in the penalty terms and another possibility is to optimize the following criterion where the number of unknowns is doubled

$$\min_{x, x'} \frac{1}{2} \|A(x + x') - b\|_2^2 + h_1 \|x'\|_1 + h_2 \sum_{k=1}^p (x_{2k-1}^2 + x_{2k}^2)^{1/2}. \quad (9)$$

While in (1), (4) and (6) the results are quite robust with respect to the tuning of h , in (8) and (9), the relative tuning of h_1 and h_2 is indeed crucial. Let us investigate the tuning of the hyper-parameters issue by evaluating the dual of (9).

3. THE DUAL CRITERION

3.1. Evaluation

We will detail the dual of (9). This then leads to the dual of (1) and (4) and also of (6). As already alluded to, the name Global Matched Filter comes from the dual (2) of (1). The constraint in the dual (2)

$$\|A^T(Ax - b)\|_\infty \leq h$$

says that at the optimum, say x^* the output of the matched filter applied to the reconstruction error $r^* = Ax^* - b$ is everywhere smaller than or equal to h . In the present context, one might wish to solve the following natural extension of (2) to get what could be called the extended GMF

$$\min_x \|Ax\|_2^2 \quad (10)$$

$$\text{under} \quad \|A^T(Ax - b)\|_\infty \leq h_1, \quad \|A_k^T(Ax - b)\|_2 \leq h_2, \quad \forall k.$$

While the first constraint corresponds in some precise way to the Generalized Likelihood Ratio Test (GLRT) [8], the last set of constraints corresponds to the extension of the GLRT to groups of atoms (in the Gaussian case). Somehow surprisingly the dual of (10) is (9) which thus might appear as the natural extension of (1).

Lemma 1. The dual of the convex program (10) is (9) and vice versa, since the problems are convex. \square

Proof: We first rewrite (10) as

$$\min_x \frac{1}{2} \|Ax\|_2^2 - \frac{1}{2} \|b\|_2^2 \quad \text{under} \quad z = A^T(b - Ax), \quad \|z\|_\infty \leq h_1$$

$$\text{and} \quad y_k = A_k^T(b - Ax), \quad \|y_k\|_2 \leq h_2, \quad \forall k.$$

The Lagrangian $\ell(x, z, y, v, s, \mu, \lambda)$ of this problem is then

$$\frac{1}{2} \|Ax\|_2^2 - \frac{1}{2} \|b\|_2^2 - s^T(z - A^T(b - Ax)) + v(\|z\|_\infty - h_1) - \sum_k \lambda_k^T(y_k - A_k^T(b - Ax)) + \sum_k \mu_k(\|y_k\|_2 - h_2)$$

and defining $\phi(v, s, \mu, \lambda) = \min_{x, z, y} \ell(x, z, y, v, s, \mu, \lambda)$, the dual problem is $\max_{v \geq 0, \mu \geq 0, s, \lambda} \phi(s, \lambda)$.

In order to evaluate $\phi(s, \lambda)$, we zero the gradient with respect to x and get

$$A^T Ax = A^T As + A^T A\lambda \Rightarrow Ax = A(s + \lambda)$$

The minimum in z is obtained by zeroing the sub-gradient

$$s = v \partial \|z\|_\infty \Rightarrow v = \|s\|_1 \geq 0 \quad \text{and} \quad s^T z = v \|z\|_\infty,$$

and the minimum in y by zeroing the gradient with respect to y

$$\lambda_k = \mu_k \frac{y_k}{\|y_k\|_2} \Rightarrow \mu_k = \|\lambda_k\|_2 \geq 0 \quad \text{and} \quad \lambda^T y = \sum \mu_k \|y_k\|_2$$

By substitution of these results into $\ell(x, z, y, v, s, \mu, \lambda)$, the dual problem is found to be

$$\max_{s, \lambda} -\frac{1}{2} \|A(s + \lambda) - b\|_2^2 - h_1 \|s\|_1 - h_2 \sum_k \|\lambda_k\|_2$$

and hence the announced result. \square

3.2. Comments

The natural extension of the dual (2) of the GMF (1) to (5), (7) and now to (10) makes us believe that, in case of overlap, one should duplicate the unknowns.

In the sequel, we consider an application in which the groups to be introduced are quite naturally overlapping and we will indeed build successively the different groups without at any time checking the possible overlap with previous groups. To solve the (primal) optimization problem, we transform it into a Second Order Cone Program (SOCP) which allows to handle complex data as well.

3.3. Tuning the hyper-parameters

In presence of groups of different sizes as in (8) the tuning of the associated hyper-parameters h_i is crucial if one wants to keep a certain coherency in the selection of the atoms. In [6] and [8], one explains that the tuning can be done by looking at the dual criterion and has indeed much in common with the tuning of the threshold in the GLRT. For the case of the criterion (8), it follows from (10) assuming that the additive noise present in the observation vector b (denoted e below) is white and Gaussian with unit variance, one has to compare the distributions of

$$\max_k |a_k^T e| \quad \text{and} \quad \max_k \sqrt{(a_k^T e)^2 + (a_{k+1}^T e)^2}.$$

Given a desired probability of false alarm, i.e. the probability to select an atom a_k when the observation vector is just noise $b = e$, these distributions allows to fix h_1 and h_2 respectively. The ratio h_2/h_1 between the two depends upon the correlation between the columns a_k and a_{k+1} . If these two columns were equal, the ratio would be $\sqrt{2}$, in practice the ratio should thus be taken smaller and function of the expected correlations between the vectors within the groups. For a arbitrary size n_g it is thus recommended to take h_g/h_1 slightly smaller than $\sqrt{n_g}$. For a 10% probability of false alarm, taking $h_1 = \sqrt{2 \ln 2n}$, as recommended usually, works well.

4. PARAMETRIC ESTIMATION OF A SUM OF HARMONIC SIGNALS

4.1. Problem description

Within the last decade, the problem of estimating the fundamental frequency of an harmonic signal in noise has attracted considerable attention because of its importance in speech denoising [9]. We consider the case where several such signals are buried in white Gaussian noise. We consider the case where the signals are complex and the additive noise complex white circular Gaussian with variance σ_e^2 :

$$b_j = \sum_{k=1}^K s_k(j) + e_j, \quad s_k(j) = \sum_{\ell=1}^L \alpha_{\ell,k} e^{i\omega_k \ell j} \quad (11)$$

which we write in vector form as

$$b = \sum s_k + e, \quad s_k = \sum \alpha_{\ell,k} a(\omega_k \ell) = A(\omega_k) \alpha_k.$$

with $a(\omega) = [1 e^{i\omega} \dots e^{i(n-1)\omega}]^T$ and $A(\omega)$ an $n \times L$ matrix associated with an harmonic signal with fundamental frequency $\omega/(2\pi)$,

$A(\omega) = [a(\omega) a(2\omega) \dots a(L\omega)]$. The problem is in general to identify the fundamental frequencies $f_k = \omega_k/(2\pi)$ but the other parameters are of interest as well. If the true order is known it is useful to introduce this knowledge into the modeling step.

4.2. Proposed solution

In the context developed in the first part of the paper, b is of course the observation vector and the columns a_k of the A matrix in (1) are the vectors $a(\omega)$ with ω taken on a grid. For harmonic signal of maximal order L , the groups are quite naturally defined and are associated with the block matrices $A(\omega)$. Depending upon the grid used for the fundamental frequencies, these groups will be more or less overlapping. Though complex, we will consider that the frequencies range from 0 to .5 and for $L = 3$, this means that the fundamental frequencies range from 0 to .5/3 i.e., .15 say.

With the notations introduced above, we propose to solve

$$\min_{Z_k} \frac{1}{2} \|b - \sum_m B_m Z_m\|_2^2 + h_L \sum_m \|Z_m\|_2, \quad (12)$$

with $B_m = A_m/\sqrt{n}$, the $n \times L$ matrix A_m -matrix with normalized columns whose first column is say a_m , i.e. the vector $a(\omega)$ evaluated at the m -th point of an M -point grid that covers the potential fundamental angular frequencies range. According to section 3, the dual of (12) is

$$\min_z \|Bz\|_2^2 \quad \text{under} \quad \|B_m^H(b - Bz)\|_2 \leq h_L \quad \forall m, \quad (13)$$

where B and z are obtained by concatenation of all the B_m and Z_m blocks respectively. Or, since these blocks are somehow overlapping by keeping in the union of the columns of these blocks only the distinct ones, the two formulations are equivalent and lead to the same optimal value. There are thus ML complex unknowns in z in the convex optimization problem (12), that is solved by transforming it into, a second order cone program (SOCP). To get the optimum, we use the subroutines SeDuMi [10], it is a freely accessible, second order cone programming library that is released under the GNU/GPL open source license.

As recommended in section 3.3, the parameter h_L in (12) is taken equal to $\sigma_e (.9\sqrt{L})\sqrt{2 \ln 2n}$ where we assume σ_e to be known. Only the order of magnitude of σ_e is actually needed since the results are quite robust with respect to the tuning the h_L . When b is composed of a the single harmonic signal, one group or two neighboring groups will then in general have non-zero weights. In the last case, the fundamental frequency estimate is obtained by linear interpolation of the frequencies associated with the two groups.

4.3. Simulation results

We have conducted a set of Monte Carlo simulations in order to evaluate the statistical properties of the extended GMF compared to the HMUSIC and the Cramer Rao lower bound. In the first of these simulations, we simulate only one harmonic signal, consider $n = 80$ observations samples and $L = 3$ harmonically related cisoids with each unit amplitude and zero initial phase ($\alpha_{\ell,1} = 1$ in (11)) and fundamental frequency $f_1 = .0945$. This example is drawn from [11]. Complex white Gaussian circular noise was added to the signal such that the SNR defined as

$$\text{SNR} = 10 \log_{10} \frac{\sum_{\ell=1}^L |\alpha_{\ell,1}|^2}{\sigma_e^2}$$

has the required value. We ran 400 realizations for different SNR's and the results are presented in Figure 1 in terms of the MSE

of the angular frequency. Note that f_1 is not on the proposed grid but falls in the middle of two neighboring points. This is indeed of little consequence since in any case some interpolation is performed to deduce the estimates from the location of the non zero components in z and that, in any case due to the additive noise, the ML estimates are likely to be on a different grid point or couple of grid points than the expected one. The Cramer Rao bounds are sensitive to all the parameters for harmonic signals and a precise development has been implemented following [12]. The extended GMF remains close to the CRB over the whole range of SNR. Some outliers appear at -10 dB.

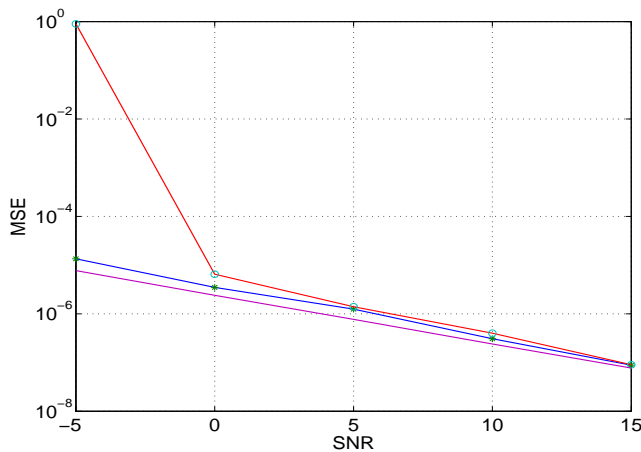


Fig. 1. A single harmonic signal. Estimate of the fundamental angular frequency. Cramer Rao bound (lower curve) and MSE for HMU-SIC (o) and proposed method (*) as a function of the SNR.

In another set of simulations, we test the resolution properties of the proposed algorithm. The sum of two harmonic signals is considered with an SNR of 20 dB each. As above, we take $L = 3$, unit amplitudes and zero initial phases for both harmonic signals. The fundamental frequency of the first, f_1 , is at .1 and the fundamental frequency of the second is $f_2 = f_1 + 1/120$ and the number of samples varies from $n=20$ to $n=120$. The Cramer Rao bound of the angular frequency estimate of $\omega_2 = 2\pi f_2$ and the MSE averaged over 1000 independent noise realizations is presented in Figure 2.

5. CONCLUDING REMARKS

In the sparse representations community, there is currently a growing interest for criterion - and associated algorithms [13]- that allow to introduce prior knowledge about the observations as for instance a structure among the basis vectors that leads to grouping these vectors and seeking sparsity at the group level.

The tuning of the hyper-parameters is then important and it is explained how this should be done by introducing the duals. Different criterion are proposed and it is argued that if an atom belongs to several groups, it should be duplicated and allocated several weights.

A typical example that is considered, is the estimation of the fundamental frequencies of a sum of harmonic signals where, if a fundamental frequency is selected, all its harmonics should be retained in the selection. The performance are close to the CRB over a range that is larger than for previously proposed algorithms.

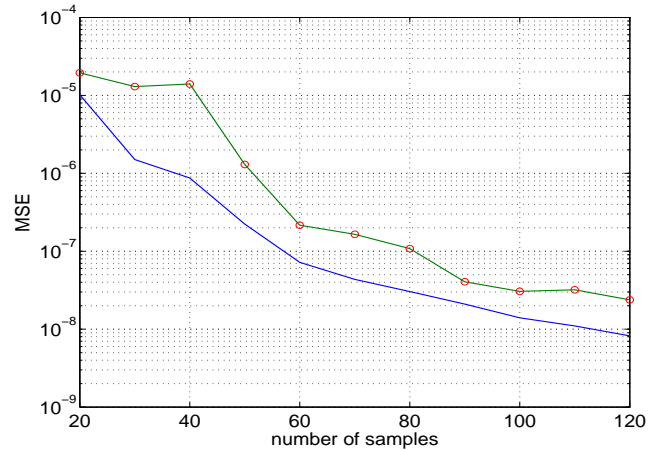


Fig. 2. A two harmonic signals scenario with fundamental frequency separation 1/120. Estimate of the (upper) fundamental angular frequency . Cramer Rao bound (lower curve) and MSE as a function of the number of observed samples.

6. REFERENCES

- [1] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables" *J. Royal. Statist. Soc B.*, Vol. 88, No. 1, pages 49-67, 2007.
- [2] M. Kowalski and B. Torresani, "Structured sparsity : from mixed norms to structured shrinkage" *Spars'09*, St. Malo, 2007.
- [3] R. Tibshirani, "Regression shrinkage and selection via the lasso." *J. Royal. Statist. Soc B.*, Vol. 58, No. 1, pages 267-288, 1996.
- [4] S. Chen, D. Donoho and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. on Scientific Comput.*, 20, 1, 33-61, 1999.
- [5] J.J. Fuchs, "Detection and estimation of superimposed signals." In *Proceedings ICASSP*, vol. III, pp. 1649-1652, Seattle, 1998.
- [6] J.J. Fuchs. "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE-T-SP*, 49, p. 702-709, Apr. 2001.
- [7] M. Osborne, B. Presnell and B. Turlach. "On the LASSO and its dual," *J. Comput. Graph. Statist.*, 9, 3319-337, 2000.
- [8] J.J. Fuchs, "The generalized likelihood ratio test and the sparse representations approach," *ICISP*, . *ICISP 2010*, Trois Rivieres, Quebec, LNCS 6134, pp. 245-253. Springer, Heidelberg (2010)
- [9] M. G. Christensen and A. Jakobsson, "Multi-Pitch Estimation," *Morgan and Claypool Publishers*, 2009.
- [10] SeDuMi, <http://sedumi.mcmaster.ca>
- [11] J.R. Jensen, M. G. Christensen and S.H. Jensen, "Fundamental frequency estimation using polynomial rooting of a subspace based method," *EUSIPCO*, Aalborg, pp. 502-506, August 2010.
- [12] H.Li, P. Stoica and J. Li, "Computationally efficient parameter estimation for harmonic sinusoidal signals," *Signal Processing*, 80, 1937-1944, 2000.
- [13] A.T. Puig, A. Wiesel and A. Hero, "A multidimensional shrinkage thresholding operator," *SSP'09*, Cardiff, pp. 113-116, August 2009.