Blind cluster structured sparse signal recovery: A nonconvex approach

(Article begins on next page)

01 May 2024

Blind Cluster Structured Sparse Signal Recovery: a Nonconvex Approach

Damiana Lazzaro*, Laura B. Montefusco, Serena Papi[1]

Department of Mathematics, University of Bologna, Piazza di Porta S. Donato, 40123 Bologna (Italy)

Abstract

We consider the problem of recovering a sparse signal when its nonzero coefficients tend to cluster into blocks, whose number, dimension and position are unknown. We refer to this problem as *blind cluster structured sparse recovery*. For its solution, differently from the existing methods that consider the problem in a statistical context, we propose a deterministic neighborhood based approach characterized by the use both of a nonconvex, nonseparable sparsity inducing function and of a penalized version of the iterative ℓ_1 reweighted method. Despite the high nonconvexity of the approach, a suitable integration of these building elements led to the development of MB-NFCS (*Model Based Nonlinear Filtering for Compressed Sensing*), an iterative fast, self-adaptive, and efficient algorithm that, without requiring any information on the sparsity pattern, adjusts at each iteration the action of the sparsity inducing function in order to strongly encourage the emerging cluster structure. The effectiveness of the proposed approach is demonstrated by a large set of numerical experiments that show the superior performance of MB-NFCS to the state-of-the-art algorithms.

Keywords: Compressed Sensing, Nonconvex Minimization, Cluster Structured Sparsity, Neighborhood Based Method, Penalization Method, Reweighted Methods.

1. Introduction

This paper deals with the problem of recovering a sparse signal when its nonzero coefficients tend to cluster into blocks, whose number, dimension and position are unknown. We refer to this problem as *blind cluster structured sparse recovery*. This kind of sparse pattern is encountered often in practical applications, including gene expression levels, DNA microarrays, MIMO channel equalization, magnetoencephalography, and has recently drawn considerable attention in the Compressed Sensing research area [2, 5, 19, 41].

The basic paradigm of conventional Compressed Sensing is the K-sparsity of an N-dimensional signal. By exploiting this sparsity it is possible to obtain robust signal recovery from $M = O(K \log(\frac{N}{K}))$ linear measurements, by selecting from the signals that agree with the measurements the one with minimum

---

ℓ_1 norm, [6, 7, 8, 9]. While this result provides a significant improvement over Nyquist-rate sampling, *model based and structured* Compressed Sensing represent new directions of the traditional theory which extend the standard sparsity prior to include structural dependencies between the values and locations of the nonzero signal coefficients [2, 5, 19, 28, 41]. Two typical examples of structure are wavelet trees [25] and block-sparsity [19, 20]. Structured sparsity models reduce the degrees of freedom of a sparse signal by permitting only certain configurations of the support of the nonzero coefficients. This allows for a higher reduction of the number of measurements required for stable signal recovery. Several papers have been devoted to integrating structured sparsity into Compressed Sensing. In [2, 19, 20] new theoretical results and algorithms are presented for the above mentioned structure, while [1, 3, 19, 23, 24, 26, 34, 36, 37, 49, 50] focus in particular on the block-structure. In these papers several combinations of norms are used to induce both overlapping and non-overlapping group sparsity and different strategies, such as variable splitting and alternating direction method of multipliers (ADMM), are proposed to solve the corresponding convex optimization problem. In the attempt to improve the results based in convex optimization, in [12], [52], nonconvex approaches have been proposed for group sparsity with sparse group. All the mentioned methods solve the cluster structured sparse signal recovery problem efficiently, but each of them requires some a priori information about the block partition. The only papers that, to the best of our knowledge, do not require any information on the sparsity structure are [43] and [48] , which present neighborhood based approaches developed in statistical Bayesian frameworks. Their methods, implemented in the Cluss-VB and EBSBL-BO packages, represent the best state-of-the-art algorithms for blind cluster structured sparse signal recovery. The only drawback of these algorithms is that are computationally demanding.

Motivated by the need to significantly reduce the computing time, this paper focuses on the blind recovery problem with the objective of creating a fast and efficient reconstruction algorithm. We propose two new ideas to achieve this goal: the first is to use in the compressed sensing reconstruction problem, in place of the ℓ_1 norm, a new neighborhood based nonconvex and nonseparable sparsity inducing function, that allows us to capture during the reconstruction process the local interdependency structure of the sparse signal support; the second is to integrate the iterative reweighted ℓ_1 scheme (IRℓ_1) within the iterative penalization approach used for the solution of the corresponding constrained nonconvex minimization problem. Even if the use of a suitable sparsity inducing function to encode prior information on the structure of the signal sparsity is not new, [24, 29, 44], the blind approach is very different. The novelty of our proposal relies on the fact that we do not require any *a priori* information on the signal structure, since the iterative integration of the penalization and reweighted approach allows the algorithm to discover at each iteration the hidden local structure and suitably adapt the action of the sparsity inducing function by changing the weights accordingly. The resulting MB-NFCS algorithm is thus iterative, and at each iteration of the penalization method the proposed nonseparable sparsity inducing function adaptively encodes the locally emerging cluster structure and encourages the next reconstruction to possess the discovered sparsity pattern. Extensive experiments

and comparisons with the best state-of-the-art algorithms for blind structured sparse signal recovery show that the proposed iterative, deterministic approach is very effective, since MB-NFCS is very fast, but still succeeds in achieving perfect reconstruction from a very low number of measurements.

The paper is organized as follows: in Section II we briefly recall the basic results of the classical and Model Based Compressed Sensing Theory. The proposed nonconvex approach to the blind cluster structured sparsity problem is presented in Section III, and in Section IV we give the solution method and the scheme of the corresponding algorithm. Section V presents a brief extension of the proposed approach to the two-dimensional case and in Section VI we give numerical results demonstrating the effectiveness of the MB-NFCS algorithm. Section VII closes the paper with a short conclusion.

2. The Compressed Sensing Problem: Summary of theory.

The Compressed Sensing formulation for the problem of recovering a K-sparse N-dimensional signal from M linear measurements $y = \Phi x$, with $M \ll N$ and Φ an $M \times N$ matrix, is the following: find $\hat{u} \in \mathbb{R}^N$ that satisfies:

$$\hat{u} = \arg \min_{u \in \mathbf{R}^N} F(u) \text{ subject to } \Phi u = y, \tag{1}$$

where the sparsity inducing function $F(u)$ allows us to select, among the infinitely many solutions of the underdetermined linear system $\Phi u = y$, the desired one. The most natural choice for $F(u)$ is $F(u) = ||u||_0$, but the corresponding combinatorial search is numerically prohibitive. A common approach is therefore to use $F(u) = ||u||_1$ as a convex relaxation of the ℓ_0 norm and to solve the corresponding constrained convex optimization problem. Theoretical results have shown the equivalence of the two formulations, provided that the matrix Φ possesses the Restricted Isometry Property of order $2K$ ($2K$-RIP), and the number of measurements M satisfies a Φ-dependent bound of the kind $M = O(K log(\frac{N}{K}))$, (see [16] and references therein).

The need to further reduce the number of measurements has recently encouraged the emergence of new research directions for the choice of the sparsity inducing function. Among them, *nonconvex* Compressed Sensing is based on a nonconvex approach to problem (1). Allowing the function $F(u)$ to be nonconvex makes it possible to better approximate the ℓ_0-norm, thereby obtaining fast and stable reconstructions with a lower number of measurements ([11, 30, 31, 33, 35, 40, 42, 45, 46]). Another direction known as *model-based* Compressed Sensing accounts for the fact that the nonzero coefficients of some sparse signals tend to cluster into groups or other structured patterns. In order to exploit the interdependency of the nonzero coefficients, model-based Compressed Sensing makes use of suitable sparsity inducing functions $F(u)$ that encode the known sparsity structure, favoring the arrangement of coefficients into the desired pattern. Several papers have been proposed to add the structured sparsity constraint into problem (1), both

by using linear combinations of norms on subsets of variables or specifically constructed sparsity inducing functions, and by adding suitable constraints on the set of nonzero coefficients (see e.g. [2, 18, 20, 27, 29, 38]). The main practical interest of inserting structured sparsity constraints into the Compressed Sensing recovery problem is that structured sparsity models reduce the degrees of freedom of sparse signals, thus decreasing the minimal number of measurements M necessary for their stable recovery.

In the classical Compressed Sensing approach, an N-dimensional K-sparse signal x lives in $\Sigma_K \subset \mathbb{R}^N$, which is a union of $\binom{N}{K}$ subspaces of dimension K. The only constraint is its sparsity K. By adding further constraints permitting only certain patterns in the support of the signal, we create a structured sparsity model $\mathcal{M}_K$, where certain K-dimensional subspaces are allowed and other disallowed [5], [28]. The structured sparse signal lives in $\mathcal{M}_K$, the union of m_K K-dimensional subspaces, with $m_K \ll \binom{N}{K}$ [2]. Accordingly, also the $2K$-RIP becomes a *model based-* RIP ($\mathcal{M}_K$-RIP) and the corresponding minimal number of measurements M decreases. In [5], Blumensath and Davis show that with probability $1 - e^{-t}$, a subgaussian random measurement matrix Φ satisfies $\mathcal{M}_K$-RIP with constant $\delta_{\mathcal{M}_K}$, provided that

$$M \geq \frac{2}{c\delta_{\mathcal{M}_K}}(\log(2 \cdot m_K) + K \log \frac{12}{\delta_{\mathcal{M}_K}} + t). \tag{2}$$

By supposing that the coefficients of the K-sparse signal are clustered into G groups, the corresponding cluster structured model $\mathcal{M}_{K,G}$ turns out to be the union of $m_{K,G}$ subspaces, where $m_{K,G}$, as evaluated in [10] using a run-length coding approach, is given by

$$m_{K,G} = \begin{pmatrix} N+1-K \\ G \end{pmatrix} \begin{pmatrix} K-1 \\ G-1 \end{pmatrix}. \tag{3}$$

By inserting this estimate in the general model-based formulation (2), we obtain the following theoretical bound on the minimum number of measurements M for stable cluster structured sparse signal recovery:

$$M = O\left(K + G\log\left(\frac{N}{G}\right)\right). \tag{4}$$

It is interesting to note that this sampling bound for $\mathcal{M}_{K,G}$ still depends on N, but the dependency is much weaker than in the classical Compressed Sensing bounds. When $G \approx K$, we again obtain the classical bound

$$M = O\left(K\log\left(\frac{N}{K}\right)\right). \tag{5}$$

3. A nonconvex approach to the blind cluster structured sparsity problem

In this subsection we focus on the blind cluster structured sparse recovery problem, namely on the problem of recovering a cluster structured sparse N-dimensional signal from a low number of linear measurements without any information on the structure details, such as size, position, sparsity and number of clusters.

For this problem, to best of our knowledge, few algorithms have been proposed in the recent literature (see [51] and references therein). All algorithms are neighborhood based and use statistical approaches that during the reconstruction process, allow them to learn the necessary structure details according to the chosen statistical model. These algorithms obtain good results but at the expense of computing time. Motivated by the desire to create a fast algorithm for blind cluster structured sparsity recovery and aware of the efficiency of the nonconvex framework proposed in [33], we propose a neighborhood based method realized by means of a deterministic, iterative self-adaptive procedure that discovers the unknown structure during the reconstruction process. The novelty of our proposal relies on the integration of the structured sparsity concept into a nonconvex local approach. This integration is achieved by using a neighborhood based nonconvex and nonseparable sparsity inducing function that encourages certain configurations of the support of nonzero signal coefficients and discourages others. Specifically, we fix an integer $d \geq 1$ and, for each sample value u_i, we consider the neighborhood set $I_i = \{\ell \in \mathbb{N} | l = i - d, ..., i + d\}$, that contains the indices of u_i and its $2d$ neighbors. We then extend to the nonseparable case the nonconvex $F(u)$ proposed in [33] as follows:

$$F_\mu(u) = \sum_{i=1}^{N} \psi_{\mu_i} \left(|u_\ell|, \ell \in I_i \right), \tag{6}$$

where we denote with $\psi_{\mu_i} \left(|u_\ell|, \ell \in I_i \right) = \psi_\mu \left(|u_\ell|, \ell \in I_i \right)$, and the function

$$\psi_\mu : \mathbb{R}_+^{2 \cdot d + 1} \to \mathbb{R}_+$$

is defined as:

$$\psi_\mu(\mathbf{t}) = \frac{1}{\log(2)} \log \left(\frac{2}{1 + e^{-\frac{||\mathbf{t}||_1}{\mu}}} \right), \quad \mu > 0 \tag{7}$$

where $\mathbf{t} \in \mathbb{R}_+^{2 \cdot d + 1}$.

The behaviour of the nonseparable ψ_μ defined in (7) is shown in Fig. 1(a), in the simple case of increasing data, for $d = 1$ and different values of the parameter μ, and in Fig. 1(b) compared with both the nonconvex separable and the ℓ_1 sparsity inducing functions.

It is easy to show that the function $\psi_\mu(\mathbf{t})$ enjoys the following properties, which are highly desirable for a sparsity-promoting function [21]:

- $\psi_\mu(\mathbf{t})$ is concave and non decreasing for $\mathbf{t} \geq 0$;

- $\psi_\mu(\mathbf{t})$ has continuous bounded partial derivatives for $\mathbf{t} \neq \mathbf{0}$ and $\psi_\mu(\mathbf{t})$ is singular for $\mathbf{t} = \mathbf{0}$;

- its partial derivatives are given by:

$$\frac{\partial \psi_\mu(\mathbf{t})}{\partial |t_h|} = \frac{1}{\mu \log(2)} \frac{1}{1 + e^{\frac{||\mathbf{t}||_1}{\mu}}}, \quad h = 1, ..., 2 \cdot d + 1$$

5

and satisfy

$$\begin{cases} \frac{\partial \psi_\mu(\mathbf{t})}{\partial |t_h|} & \rightarrow & 0 & \text{for } ||\mathbf{t}||_1 \gg \mu \\ \frac{\partial \psi_\mu(\mathbf{t})}{\partial |t_h|} & \rightarrow & \frac{1}{2\mu \log(2)} & \text{for} ||\mathbf{t}||_1 \ll \mu \end{cases}.$$

The aforementioned properties of $\psi_\mu(\mathbf{t})$ induce similar properties in the sparsity inducing function $F_\mu(u)$, which turns out to be both nonconvex and nonseparable. The latter property represents the mathematical tool that allows the algorithms to discover and encourage the unknown group structure. In fact, while the separable nonconvex $F(u)$ proposed in [33] treats each coefficient independently from the others, the nonseparable nonconvex $F_\mu(u)$ defined in (6) acts very differently. Its effect on a coefficient depends on the behavior of its neighbors, and this represents a means to strongly encourage the desired cluster structured sparsity patterns. Specifically, if the ℓ_1 norm of the vector of coefficients whose indices belong to the neighborhood I_i is greater than μ, then the coefficient u_i is strongly encouraged to be nonzero, because it likely belongs to a cluster; otherwise it is encouraged to be zero, because it could be an isolated artifact.

Remark 3.1. *It is important to note that, despite its nonseparability, the action of the proposed penalty function remains local, because it concerns only a small neighborhood of radius d of the considered coefficient. Moreover, differently from the penalties used in the group sparsity literature[17, 36, 37, 38, 41, 50, 52], where the composition of convex or nonconvex functions with the ℓ_2 norm of groups is always used, we consider the ℓ_1 norm of the vector of coefficients whose indices belong to the neighborhood I_i. This allows the algorithm to better distinguish between signal and noise, to more closely approximate the ℓ_0 norm and even to promote sparsity inside the groups.*

In Fig. 2 we consider the reconstruction of a K-sparse N-dimensional signal whose non-zero coefficients are randomly clustered into two groups, (N=256, K=30, G=2, details are given in the Numerical Results Section) and we compare the results yielded by the two nonconvex sparsity inducing functions: the separable $F(u)$, as given in [33], and the nonseparable $F_\mu(u)$ given in (6). As expected, the use of the nonseparable $F_\mu(u)$ allows the algorithm to reach 90% Probability of Perfect Reconstruction with only $M = 60 = 2 \cdot K$ measurements and 100% for $M = 70$. In contrast, the separable $F(u)$ needs at least $M = 90$ acquisitions to reach 90% Probability of Perfect Reconstruction.

The nonseparability of the sparsity inducing function is also advocated in [46, 47], where the authors prove, in the context of classical sparse estimation problems, the uniform superiority of the nonseparable approach over the ℓ_1 solution and show that the use of nonseparable sparsity inducing functions drastically reduces the number of local minima of the original ℓ_0 minimization problem. Moreover, the use of the nonseparable sparsity inducing function allows the algorithm to exploit the eventual intra-block correlation thus improving the recovery performance.

Another key element for the success of the recovery is the dependence of the sparsity inducing function (6) on the parameter μ. It allows us to adapt the action of $F_\mu(u)$ according to the available number

6

of measurements. When M is very low compared with the signal sparsity, a low value for μ allows the function to better reproduce the action of the ℓ_0-norm, while a slightly higher value is enough when more measurements are available.

It is important to mention that in other papers suitable nonseparable, parametric penalty functions have been used to induce group sparsity for the denoising of group sparse signals. In [13] an overlapping translation invariant group shrinkage method (OGS) has been developed by minimizing a convex group sparsity promoting penalty function and in [14], to improve the overlapping group shrinkage and to more strongly promoting sparsity, the authors use a nonconvex parametrized penalty, the parameters of which are constrained to ensure the convexity of the total cost function. The extension of these approaches to the reconstruction problem, however, is actually known only for the sparse signal case [39] .

4. A Self Adaptive Iterative Algorithm

The insertion of the nonconvex, nonseparable sparsity inducing function (6) in problem (1) leads to the challenging constrained minimization problem

$$\min_{u \in \mathbb{R}^N} \sum_{i=1}^{N} \psi_{\mu_i} \left(|u_\ell|, \ell \in I_i \right) \ \text{ subject to } \ \Phi u = y. \tag{8}$$

For its solution we propose to integrate an IRℓ_1 scheme into a penalization framework and to exploit the iterative nature of both these approaches in order to take advantage of the nonseparability of the penalty function for adaptively discovering the unknown cluster structure. Specifically, we consider the sequence of unconstrained nonconvex minimization problems

$$u^k = \arg \min_{u \in R^N} \mathcal{P}(F_\mu, \lambda_k, u) = \arg \min_{u \in R^N} \left\{ F_\mu(u) + \frac{1}{2\lambda_k} ||\Phi u - y||_2^2 \right\}, \tag{9}$$

where $\lambda_1 > \lambda_2 > ... > \lambda_{min}$ is a decreasing sequence of penalization parameter values that ensures the convergence of the penalized solution to that of the original problem, [22]. Thus, iteratively, for each new value λ_{k+1} of the penalization parameter we evaluate the weights w_i^{k+1}, in terms of the partial derivatives of the nonconvex, nonseparable sparsity inducing function F_μ evaluated at the previous iterate u^k.

$$w_i^{k+1} = \left. \frac{\partial F_\mu(u)}{\partial |u_i|} \right|_{u=u^k} = \sum_{\ell \in I_i} \frac{\partial}{\partial |u_i|} \psi_{\mu_\ell} \left(|u_h^k|_{h \in I_\ell} \right) = \sum_{\ell \in I_i} \frac{1}{\mu \log(2)} \left(\frac{1}{1 + e^{\frac{\sum_{h \in I_\ell} |u_h^k|}{\mu}}} \right) \tag{10}$$

and we replace the nonconvex, nonseparable sparsity inducing function $F_\mu(u)$ by its convex local linear majorizer

$$f_\mu(u) = F_\mu(u^k) + \sum_{i=1}^{N} \left. \frac{\partial F_\mu(u)}{\partial |u_i|} \right|_{u=u^k} (|u_i| - |u_i^k|); \tag{11}$$

that satisfies

$$f_\mu(u^k) = F_\mu(u^k) \quad \text{and} \quad f_\mu(u) \geq F_\mu(u)$$

7

for u in a suitable neighborhood of u^k. We thus obtain the following Penalized Reweighted ℓ_1 iterative scheme (PRℓ_1)

> Starting from $u^0 = \phi^T y$, $\lambda_1 > \lambda_{min} > 0$
>
> `for` k=0,1,2,...
> $$w_i^{k+1} = \left.\frac{\partial F_\mu(u)}{\partial |u_i|}\right|_{u=u^k}, i = 1, \cdots, N$$
> $$u^{k+1} = \arg\min_{u \in R^N} \mathcal{P}(f_\mu, \lambda_{k+1}, u) = \arg\min_{u \in R^N} \left\{ \sum_{i=1}^N w_i^{k+1} |u_i| + \frac{1}{2\lambda_{k+1}} ||\Phi u - y||_2^2 \right\}, \tag{12}$$
> choose λ_{k+2} such that $\lambda_{k+2} < \lambda_{k+1}$
>
> `until convergence`

The convergence of the PRℓ_1 procedure to a local minimum of the original problem is proven in [33] by showing that the descent property of the penalization method is maintained when, at each iteration, the nonconvex sparsity inducing function $F_\mu(u)$ is replaced by its convex local majorizer $f_\mu(u)$.

The iterative procedure PRℓ_1 is the main iterative structure of the algorithm Model-Based - NFCS (MB-NFCS) displayed in Table 1. In that context the updating of the penalization parameter λ has been done with the adaptive rule developed in [32], where it is shown that a good strategy to decrease the penalization parameter value is to follow the behavior of the objective functional $P(f_\mu, \lambda, u)$, namely

$$\lambda_{k+2} = \lambda_{k+1} \cdot \frac{P(f_\mu, \lambda_{k+1}, u^{k+1})}{P(f_\mu, \lambda_k, u^k)} = \frac{\lambda_2}{P(f_\mu, \lambda_1, u^1)} \cdot P(f_\mu, \lambda_{k+1}, u^{k+1}) = c \cdot P(f_\mu, \lambda_{k+1}, u^{k+1})$$

where $c = \frac{\lambda_2}{P(f_\mu, \lambda_1, u^1)}$.

The second iterative structure of the algorithm, to be performed at each step of the outer loop, is the iterative solution of the unconstrained convex minimization problem (12). In fact, since the objective function of this minimization problem is the sum of two convex functionals, one of which is non-smooth, the minimization is obtained using the Forward-Backward splitting approach [15] combined with FISTA acceleration strategy [4].

Remark 4.1. *The integration of the reweighted scheme into the penalization approach and the use of a non-separable sparsity inducing function are crucial for the effectiveness of the proposed MB-NFCS algorithm for blind structured signal recovery. They allow the algorithm to iteratively discover the unknown structure driven by the weight values. For this purpose, the latter are adaptively modified at each outer iteration according to the value of the derivative of the nonseparable sparsity inducing function evaluated at the previous approximate reconstruction, and then used to strongly encourage in the current iteration the emerging sparsity pattern. These subsequent updates of the weights values lead the proposed algorithm to become self adaptive, since its action on the coefficients changes at each outer iteration driven by the emerging cluster structure. This adaptive behavior of the weights is clearly shown in Fig. 3, where we consider the blind recovery from $M = 60$ acquisitions of a 30-sparse 256-length signal, whose nonzero coefficients are clustered*

8

into two groups, and we display, for each iteration of the penalization approach, the evolution of the weight values and the corresponding signal reconstruction.

4.1. Choice of parameter values

The algorithm MB-NFCS is characterized by some problem dependent parameters: the initial values u_0 and λ_0, the parameters μ and β, the number of neighbors to be considered, the constant γ of the inner loop exit test. Guided by a large number of numerical experiments, we have used the following strategies: we have first evaluated $u_0 = \Phi^T y$ and then chosen the starting value of the penalization parameter as $\lambda_0 = 0.1||u_0^0||_\infty$ in noiseless experiments, and $\lambda_0 = 0.25||u_0^0||_\infty$ in noisy cases. The constant μ, which models the shape, and consequently the action of the sparsity inducing function $F_\mu(u)$, has been set to $\mu = 0.8 \cdot ||u_0^0||_\infty$. The parameter β depends on the spectral radius of the matrix $\Phi^T\Phi$ [15] and must satisfy the condition $0 < \beta < \frac{2}{\lambda_{max}(\Phi^T\Phi)}$ for the convergence of the Forward-Backward splitting method. In the experiments presented in the Section 4 we have always used random Gaussian matrices (with normalized columns), so the choice $\beta = 1$ satisfies the convergence requirements. The neighborhood radius d, which characterizes the nonseparability of the sparsity inducing function, is crucial for discovering the group structure. It must be different from zero and not too large, to prevent the loss of the local nature of the method. Motivated by an extensive experimentation, that we partially present in Section 6.1, we have decided to use in the following always $d = 1$, but the choice $d = 2$ is equivalently effective.

The most important parameter for the speed of the proposed algorithm is γ. It represents a means of tuning the precision of the iterative proximal splitting approach and heavily influences the efficiency of the algorithm. Since we are working in the context of a penalization method, the precision requested in the inner loop is $\gamma \cdot \lambda_k$, namely, dependent on the value of the penalization parameter. Nevertheless, the value of γ is important to suitably adjust the precision according to the difficulty of the reconstruction problem. When the number of acquisitions is very low, a high precision is necessary in the inner loop, while higher values are allowed when more measurements are available. In our experiments we have used values of γ of the order of 10^{-5} for $M \approx 2K$, and values of γ of the order of 10^{-3} for $M \approx 3K$.

The last free parameter, which is not explicitly mentioned in the proposed algorithm, represents the exit test of the outer loop. In noiseless simulations we used the exit test $NMSE <= 10^{-8}$, which corresponds to "perfect reconstruction", namely, to a relative reconstruction precision of about 10^{-5} (for the definition of NSME see the next section), while the algorithm is stopped after $maxit= 3000$ iterations if the exit test was not satisfied.

In noisy experiments we used the exit test

$$||\Phi \cdot u^k - y||_2 \leq \epsilon, \tag{13}$$

where $\epsilon = \sigma \cdot \sqrt{M}$, and σ is the standard deviation of the Gaussian white noise, while only 400 iterations

was considered if the computed solution did not reach the precision requested by the exit test (13).

5. Extension to the two-dimensional case

In this section, following the suggestions of the reviewers, we briefly outline how the proposed approach to the blind cluster structured sparsity problem can be easily extended to the two-dimensional case. In this case, we consider an unknown signal $u \in \mathbb{R}^{N \times N}$ that is a piecewise smooth image, whose gradient is sparse and with the nonzero values clustered along the image edges. We denote with Δu the image gradient, and we fix an integer $d \geq 1$. For each sample value $u_{i,j}$, we consider the neighborhood set

$$I_{i,j} = \{(\ell, m) \in \mathbb{N} \times \mathbb{N} | l = i - d, ..., i + d, m = j - d, ..., j + d\},$$

that contains the indices of $u_{i,j}$ and its $(2d+1)^2 - 1$ neighbors. The two-dimesional nonconvex nonseparable sparsity inducing function is then defined as:

$$F_\mu(u) = \sum_{i=1}^{N} \sum_{j=1}^{N} \psi_{\mu_{i,j}} \left(|\Delta u_{\ell,m}|, \ell, m \in I_{i,j}\right), \tag{14}$$

where we denote with $\psi_{\mu_{i,j}} \left(|\Delta u_{\ell,m}|, \ell, m \in I_{i,j}\right) = \psi_\mu \left(|\Delta u_{\ell,m}|, \ell, m \in I_{i,j}\right)$, and the function

$$\psi_\mu : \mathbb{R}_+^{(2 \cdot d + 1) \times (2 \cdot d + 1)} \to \mathbb{R}_+$$

is given by

$$\psi_\mu(\mathbf{t}) = \frac{1}{\log(2)} \log \left(\frac{2}{1 + e^{-\frac{\sum_{r,s} |t_{r,s}|}{\mu}}} \right), \quad \mu > 0$$

where $\mathbf{t} \in \mathbb{R}_+^{(2 \cdot d + 1) \times (2 \cdot d + 1)}$.

In order to extend the Penalized Reweighted Iterative ℓ_1 scheme to the two-dimensional case, the weigths $w_{i,j}^{k+1}$ are calculated, as in one-dimensional setting, in terms of the partial derivatives of the sparsity inducing function F_μ evaluated at the previous iterate u^k as

$$w_{i,j}^{k+1} = \sum_{\ell, m \in I_{i,j}} \frac{1}{\mu \log(2)} \left(\frac{1}{1 + e^{\frac{\sum_{h,n \in I_{\ell,m}} |\Delta u_{h,n}^k|}{\mu}}} \right). \tag{15}$$

In this way, the changes to be made in the MB-NFCS algorithm regard only the calculation of the weights in the outer iteration loop and the Backward minimization step in the inner iteration. In the latter, a weighted Total Variation minimization has to be performed. The effectiveness of the extension of the proposed algorithm to the two-dimensional case is proved in Experiment 7, presented in Section 6.8.

6. Numerical Results

This section is devoted to demonstrate the effectiveness of the proposed MB-NFCS algorithm by showing the results of several numerical experiments that clearly highlight its reconstruction capabilities, stability and speed. The experiments that we present are designed to emphasize the benefits of using the proposed nonconvex, nonseparable sparsity inducing function for the reconstruction of blind cluster structured sparse signals, both as regards the required number of measurements for exact signal reconstruction and as regards the computing time. We have compared our results with those of CLUSS-VB [48], EBSBL-BO [43], the better state-of-the-art algorithms for the nonconvex recovery of blind cluster structured sparse signal. The MATLAB package for CLUSS-VB is available at *https://sites.google.com/site/link2yulei/cs*, while the MATLAB software for EBSBL-BO has been kindly provided by the author.

The results presented in the next subsections were obtained by repeating each experiment 100 times. In each trial, unless stated otherwise:

- a synthetic cluster structured K- sparse N-length signal x was considered, whose G nonzero blocks, randomly generated according to a multivariate Gaussian distribution with zero mean and covariance matrix Σ, as in [43], are of different random dimension, with only the constraint that the sum of their dimensions equals the sparsity K. Different intra-block correlations were obtained by suitably manipulating the covariance matrix. Correlation values have been randomly uniformly chosen from 0 to 1;

- a zero mean random Gaussian measuring matrix Φ was generated, with columns normalized to unit ℓ_2 norm;

- the measurements y were obtained as $y = \Phi \cdot x$ and the reconstruction was denoted by $\hat{u}$.

In noiseless experiments we used the Probability of Perfect Reconstruction (PPR) as performance index, defined as the ratio of total number of successful trials over the total number of trials. A trial was considered successful when the Normalized Mean Square Error (NMSE), given by

$$\text{NMSE} = ||\hat{u} - x||_2^2 / ||x||_2^2,$$

satisfies NMSE $\leq 10^{-8}$.

As this high precision seemed to penalize the performance of the other considered algorithms, according to the reviewer suggestion, in Experiment 3 we used, as further performance index, the Probability of Good Reconstruction (PGR) defined using NMSE$\leq 10^{-3}$.

In noisy experiments the accuracy of each reconstruction was evaluated through the SNR between the original sparse signal x and its reconstruction $\hat{u}$, defined as

$$\text{SNR} = 10 \cdot \log_{10}\left(\frac{1}{\text{NMSE}}\right).$$

The performance index of an experiment was then evaluated by the *mean* SNR, defined as the mean of the SNR values of all the reconstructions obtained by running the considered experiment, including even those where the algorithm does not reach $||\Phi \cdot u^k - y||_2 \leq \epsilon$ in 400 iterations.

We remark that, even if not justified by theoretical results, in our experimentation we have used an adaptive rule for the parameter μ. Specifically we have slightly decreased its value at each change of the penalization parameter λ_k, in order to adapt the weights to the trend of the newly computed solution. In [33] the authors have experimentally demonstrated the superior performance of this strategy with respect to the results obtained with fixed μ. In this case, in order to ensure the descent property of the algorithm, an acceptance test must be satisfied before taking the new value of μ.

All the experimental results were obtained by running a Matlab implementation of the proposed algorithm on a Intel Core 2 Duo, CPU P8700 (2.53 GHz) notebook computer with 4 GB of RAM under Windows 7.

6.1. Experiment 0: Influence of the radius d of the considered neighborhood on the algorithm efficiency.

The aim of this experiment is to show how the choice of the dimension of the set I_i influences the efficiency of the proposed algorithm. We have fixed $N = 512$, $K = 60$, $M = 120$ and, in order to work with groups of different size, we have considered G ranging from 2 to 8 with step 2. Each experiment has been performed with $d = 0, 1, 2$. In this setting, as expected, by using $d = 0$ the algorithm does not reach the Perfect Reconstruction. The results obtained with $d = 1$ and $d = 2$, presented in Table 2, show that the radius of the neighborhood minimally affects the performance of the algorithm. In fact, even if the value of $d = 2$ sometimes reduces the number of iterations, each of them requires an higher computational cost. Larger values of d were not considered because they are not suited to the local nature of the proposed method.

6.2. Experiment 1: Comparisons with the better state-of-the-art algorithms.

This set of experiments is devoted to comparing the efficiency and the robustness to noise of the proposed algorithm with those of CLUSS-VB, EBSBL-BO ($h = 4$, $h = 8$). In Experiment 1-1, we have fixed $N = 256$, $K = 40$, and $G = 8$, and we have considered values of M ranging from 80 to 150 with step 10.

Fig. 4(a) shows the reconstruction results obtained by MB-NFCS and by the other considered algorithms. The comparison between them shows that the proposed algorithm has the greatest efficiency. In fact, with only $M = 90$ acquisitions, it exactly recovers the blind cluster structured signals, while none of the others methods reaches the value of PPR=1.

The setting of Experiment 1-2 is the same as in Experiment 1-1, but the acquisitions are corrupted by white noise with standard deviation $\sigma = 0.01$. The results, displayed in Fig.5(a), show that for a very low number of acquisitions MB-NFCS exhibits better accuracy than the others, while for higher values of M all methods achieve similar results in terms of reconstruction quality. Concerning the average reconstruction

time of both experiments, Fig. 4(b) and Fig. 5(b) highlight the high speed of MB-NFCS, whose computing time is about two order of magnitude lower than that needed by the other algorithms. In fact, MB-NFCS requires, for each inner iteration, only two matrix-vector multiplications in the Forward Minimization step, and a soft-thresholding operation to update the solution in the Backward step. Since the outer iterations are usually of order of about ten, the computational cost of the whole algorithm is mainly due to that of the inner loop. Cluss-VB [48] requires, instead, the solution of an ill-posed inverse problem of dimension $M \times M$ at each iteration step, while the computational complexity of EBSBL-BO is due to the Expectation Maximization (EM) method on which is based the updating of the γ_i learning rule in the sparse Bayesian Learning.

6.3. Experiment 2: Adaptivity of MB-NFCS

In this experiment we demonstrate the flexibility of the proposed method to adapt itself to various kinds of signal structure. For this purpose we consider unstructured K-sparse spike signals with $N = 256$, $K = 30$, (a sample signal is displayed in Fig. 6(a)) and values of M ranging from 60 to 140 with step 10. We then display in Fig. 6(b) the reconstruction results yielded by MB-NFCS, WNFCS [33], CLUSS-VB, EBSBL-BO (h = 4, h=8). The best reconstruction results are clearly obtained by WNFCS, which uses a nonconvex separable sparsity inducing function, but also MB-NFCS, in spite of the nonseparability of F_μ, works quite well, as can be seen from its performance very similar to that of WNFCS. This self-adaptive behavior of MB-NFCS is a crucial property of the proposed method, since it justifies its effectiveness in blind cluster structured sparse signal recovery.

6.4. Experiment 3: Effect of the number of clusters on the performance.

The aim of this experiment is to investigate the effect of increasing the number of clusters on the performance of the proposed MB-NFCS algorithm. In this simulation, we fix $N = 512$, $K = 60$ and we consider two values of the number of acquisitions, $M = 120$ and $M = 150$. For each M we let the number of nonzero groups G range from 2 to 16, with step 2. The results obtained by MB-NFCS, CLUSS-VB, EBSBL-BO ($h = 4$, $h = 8$) for $M = 150$ and $M = 120$, are displayed in Fig.7(a) and Fig.7(b), respectively, when we use the PPR as performance index and in Fig. 7(c) and Fig. 7(d) when we use the PGR . They show that, differently from the other experimented algorithms, the performance of MB-NFCS depends very weakly on the required precision and on the increasing of the number of groups, even for low number of acquisitions, while the precision play an important role for the other methods that, in the experiments with lower precision, exhibit better performance than that shown in higher precision experiments.

6.5. Experiment 4: Robustness to different noise levels.

This experiment demonstrates the robustness of the proposed algorithm to different noise levels. We generate a cluster structured sparse signal with $N = 512$, $K = 60$, $G = 10$. We consider $M = 100$

measurements corrupted by white noise with standard deviation σ ranging from $\sigma = 0.01$, to $\sigma = 0.09$, (which corresponds to a SNR approximately ranging from 32 to 12 dB). For each noise level, we repeat the experiment 100 times with same sparse signal and acquisition matrix Φ. The results, depicted in Fig.8(a), show that the SNR of recovery by MB-NFCS is proportional to the noise bound with proportionality constant ≥ 1 and in some cases it exhibits improved accuracy compared to other algorithms. Fig. 8(b) highlights again the higher speed of the proposed algorithm over the others. Interestingly, in this experiment, in which we do not requires Perfect Reconstruction, all methods give comparable results, while in Experiment 3, when Perfect Reconstruction is requested, the performance of the other methods decay greatly.

6.6. Experiment 5: Robustness to intra-block correlation.

The purpose of this experiment is to show that the performance of the proposed algorithm is independent from the intra-block correlation of the signal. To this aim we have considered three possible cases for intra-block correlation: (Experiment 5.1) correlation values equal to zero; (Experiment 5.2) correlation values randomly uniformly chosen from 0 to 1; (Experiment 5.3) correlation values randomly uniformly chosen from 0.95 to 1. For all the three experiments we fix $N = 512$, $K = 80$ and $G = 10$ and M ranging from 140 to 200 with step 10. The results are depicted in Fig. 9 and confirm that the proposed algorithm exhibits the same performance in all the three cases and that its performance is the best of the experiment. This demonstrates the self-adaptivity property and effectiveness of MB-NFCS.

6.7. Experiment 6: Minimum number of acquisitions for Perfect Reconstruction

This experiment verifies that, when using a nonconvex non separable sparsity inducing function, the minimum number of acquisitions $M_{\min}$ for stably blind clustered sparse signal recovery meets the theoretical bound given in eq. (4). More precisely, if we write

$$M_{\min} = a \cdot \left(K + G \log \left(\frac{N}{G} \right) \right) \tag{16}$$

with $a \geq 1$, a small constant, we find that with the value $a = 1.8$ relation (16) represents an upper bound for our empirical estimation of $M_{\min}$.

In this simulation, we fix $K = 100$, $G = 10$ and for each value of $N = 2^n$, n ranging from 9 to 13 with step=1, we found the lowest number of measurements $M_{\min}$ for which we have 100% Probability of Perfect Reconstruction. The resulting values of the minimum number of acquisitions $M_{\min}$ are displayed in Fig. 10 as a function of n. As can be seen from this figure, the values obtained for $M_{\min}$ by our algorithm remain below the theoretical estimate given in (16), with $a = 1.8$.

6.8. Experiment 7: Reconstruction of 2D-data

In this experiment we show the efficiency of the MB-NFCS algorithm in the case of 2D-data. To test the proposed algorithm, we have used the synthetic Phantom image, shown in Fig. 11 together with its gradient.

14

This image, of size 256×256, is a piecewise smooth image that satisfies the assumption that its gradient is sparse and with the nonzero values clustered along the edges. The experiments are performed using subsampled frequency acquisitions. As acquisition mask, we have used a radial mask with 6 rays (2.5 % undersampling). Fig. 12 demonstrates the efficiency of the proposed algorithm when we use very highly undersampled data. In fact, while with the ℓ_1 penalty the reconstruction is poor (SNR=20.85) (Fig. 12(b)), the quality improves using the nonconvex separable penalty function (SNR=69.3) (Fig. 12(c)), but only by making use of the proposed nonseparable $F_\mu(u)$ ($d = 1$) we reach the Perfect Reconstruction (SNR=91.4) (Fig. 12(d)) in only 148 seconds.

7. Conclusions

We have proposed a new algorithm for the blind cluster structured recovery problem. It is based on a deterministic iterative neighborhood based approach, which uses a nonconvex, nonseparable sparsity inducing function that incorporates the structured sparsity constraints by encouraging solutions which possess the desired structured sparsity properties. For the solution of the nonconvex, constrained minimization problem we integrate the iterative reweighted ℓ_1 method into the penalization approach, still maintaining the descent property of the latter. This guarantees the convergence of the method to a local minimum of the original problem. The resulting MB-NFCS algorithm is iterative and adaptively adjusts the action of the sparsity inducing function according to the emerging cluster structure. A large set of numerical experiments demonstrate that MB-NFCS is very fast, robust to model mismatch and to noise, insensitive to the increase of the number of groups and to different intra-block correlations and significantly outperforms the best state-of-the-art algorithms.

References

[1] F. Bach, R. Jenatton, J. Mairal and G. Obozinski, Optimization with Sparsity-Inducing penalties, *Found. Trends Machine Learn*, vol.4, N. 1, pp. 1–106, 2012.

[2] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde: Model-Based Compressive Sensing, *IEEE Trans. on Information Theory* 56(4), pp. 1982-2001, 2010.

[3] I. Bayram, Mixed Norms with Overlapping Groups as Signal Priors, in *Proc. IEEE Int. Conf. Acoust., Speech SIgnal Process. (ICASSP)*, pp. 4036–4039, 2011.

[4] A. Beck, M. Teboulle, Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems, *IEEE Trans. on Image Processing*, Vol.18, N.11, pp.2419–2434, 2009.

[5] T. Blumensath and M. E. Davies, Sampling Theorems for Signals from the Union of Finite-Dimensional Linear Subspaces", *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 18721882, Apr. 2009.

[6] E.J. Candès, Compressive Sampling, *Proceedings of the International Congress of Mathematicians*, Vol.3, pp.1433–1452, Madrid, 2006.

[7] E. J. Candès, J. Romberg, and T. Tao, Near Optimal signal recovery from random projections: Universal encoding strategies?,*IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, 2006.

[8] E.J. Candès, J. Romberg, T. Tao, Stable Signal Recovery from Incomplete and Inaccurate Measurements, *Communications on Pure and Applied Math.*, Vol. 59, Issue 8, pp. 1207–1223, 2006.

[9] E.J. Candès, J. Romberg, T. Tao, Robust Uncertainty Principle: Exact Signal Reconstruction From Highly Incomplete Frequency Information, *IEEE Transactions on Information Theory*, Vol .52, n. 2, pp. 489–509, 2006.

[10] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, Recovery of Clustered Sparse Signals from Compressive Measurements, paper presented to the *Conference Sampling Theory and Applications (SAMPTA)*, Marseille, 2009.

[11] R. Chartrand, Fast Algorithms for Nonconvex Compressive Sensing: MRI Reconstruction from Very Few Data, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp.262-265, 2009.

[12] R. Chartrand and B. Wohlberg, A Nonconvex ADMM Algorithm for Group Sparsity with Sparse Groups, ICASS 2013, pp. 6009-6012.

[13] P. Chen and I. Selesnick, Translation-invariant shrinkage/thresholding of group sparse signals, *Signal Processing*, Vol. 94, pp. 476-489, 2014.

[14] P. Chen and I. Selesnick, Group-sparse signal denoising: Non-convex regularization, convex optimization, *IEEE Trans. on Signal Processing*,Vol. 62, No. 13, pp. 3464–3478,July 2014.

[15] P.L. Combettes and V.R. Wajs , Signal Recovery by Proximal Forward-Backward Splitting, *SIAM Journal on Multiscale Modelling and Simulation*, Vol. 4, No. 4, pp. 1168–1200, 2005.

[16] M. Davenport, Random Observations on Random observations: Sparse Signal Acquisition and Processing, PhD Thesis, Rice University, Aug. 2010.

[17] W. Deng, W. Yin, and Z. Zhang, Group Sparse Optimization by Alternating Direction Method, *Proc. SPIE 8858, Wavelets and Sparsity XV, 88580R (September 26, 2013)*; doi:10.1117/12.2024410.

[18] M. Duarte and Y. C. Eldar, Structured Compressed Sensing: From Theory to Applications, an overview paper in *IEEE Trans. on Signal Processing*,Vol.59, issue 9, pp. 4053-4085, Sept. 2011.

[19] Y. C. Eldar, M. Mishali, Robust Recovery of Signals From a Structured Union of Subspaces, *IEEE Trans. on Information Theory*, Vol.55, N.11, pp.5302–5316, 2009.

[20] Y. C. Eldar, P. Kuppinger, and H. Blcskei, Block-Sparse Signals: Uncertainty Relations and Efficient Recovery, *IEEE Transaction on Signal Processing*, Vol.58, N.6, pp.3042–3054, 2010.

[21] J. Fan, and R. Li, R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *J. Amer. Statist. Assoc.* 96, pp. 1348–1360, 2001.

[22] N. I. M. Gould, On the Convergence of a Sequential Penalty Function Method for Constrained Minimization", *SIAM J. Numer. Anal.*, vol. 26, no. 1, pp. 107-128, 1989

[23] L. Jacob, G. Obozinski, and J. P. Vert, Group Lasso with Overlap and Graph-Lasso, in *Proc. Int. Conf. Machine Learn (ICML)*, pp. 433–440, 2009.

[24] R. Jenatton, J. Audibert, F. Bach, Structured Variable Selection with Sparsity-Inducing Norms, *Journal of Machine Learning Research* Vol. 12, pp. 2777–2824, 2011.

[25] L. He, L. Carin, Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing, *IEEE Transaction on Signal Processing*, Vol.57, N.9, pp.3488–3497, 2009.

[26] J. Huang, T. Zhang, and D. Metaxas, Learning with Structured Sparsity, in *ICML*, pp. 417–424, 2009.

[27] M. Kowalski and B. Torresani, Sparsity and Persistence: Mixed Norms Provide Simple Signal Models with Dependent Coefficients, *Signal, Image and Video Processing.* Vol. 3, N. 3, 2009.

[28] Y. M. Lu and M. N. Do, Sampling Signals from a Union of Subspaces, *IEEE Signal Processing Mag.*, vol. 25, pp. 4147, Mar. 2008.

[29] C.A. Micchelli, J. M. Morales, M. Pontil, A Family of Penalty Functions for Structured Sparsity, NIPS 2010.

[30] H. Mohimani, M. Babaie-Zadeh, I. Gorodnitsky, C. Jutten, Sparse Recovery using Smoothed ℓ^0 (SL0): Convergence Analysis, preprint 2010

[31] H. Mohimani, M. Babaie-Zadeh, C. Jutten, A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed ℓ^0 Norm, *IEEE Trans. on Signal Processing*, vol. 57, N. 1, January 2009

[32] L.B.Montefusco and D.Lazzaro, An Iterative L_1-Based Image Restoration Algorithm With an Adaptive Parameter Estimation, *IEEE Trans. on Image Process.*, vol. 21, no. 4, pp. 1676–1686, April 2012.

[33] L.B.Montefusco, D.Lazzaro and S.Papi, A Fast Algorithm for Nonconvex Approaches to Sparse Recovery Problems,*Signal Processing*, vol. 93, no. 9, pp. 2636-2647, 2013.

[34] S. Mosci, S. Villa, A. Verri, and L. Rosasco. A Primal-Dual Algorithm for Group Sparse Regularization with Overlapping Groups, *Advances in Neural Information Processing Systems* , Vol 23, 2010.

[35] N. Mourad ,J. P. Reilly, Minimizing Nonconvex Functions for Sparse Vector Reconstruction, *IEEE Trans. on Signal Processing*, Vol. 58, n.7, pp.3485–3496, July 2010.

[36] G. Peyre and J. Fadili, Group Sparsity with Overlapping Partition Functions, in *Proc. Eur. Signak Image Process. Conf. (EUSPICO)*, Aug. 29 - Sep 2, 1022.

[37] A. T. Puig, A. Wiesel, G. Fleury, and O. Hero, Multidimensional Shrinkage-Thresholding Operator and Group Lasso Penalties, *IEEE Signal Process. Lett.*, Vol. 18, N.6, pp. 363-366, 2011.

[38] Z. T. Qin, D. Goldfarb, Structured Sparsity via Alternating Direction Methods, *Journal of Machine Learning Research*, Vol. 13, pp. 1435–1468, 2012.

[39] I. Selesnick and I. Bayram, Sparse Signal Estimation by Maximally Sparse Convex Optimization, *IEEE Trans. Signal Processing*, Vol. 62, No.5, pp. 1078–1092, March 2014.

[40] W. Shao, H. Deng, and Z. Wei, Nonconvex Compressed Sampling of Natural Images and Applications to Compressed MR Imaging, *ISRN Computational Mathematics*, Vol. 2012, Article ID 982792, doi:10.5402/2012/982792, 2012.

[41] M. Stojnic, F. Parvaresh, and B. Hassibi, On the Reconstruction of Block-Sparse Signals with an Optimal Number of Measurements, *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3075-3085, Aug. 2009.

[42] J. Trzasko, A. Manduca, Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic ℓ_0-Minimization, *IEEE Transactions on Medical Imaging*, Vol. 28, N. 1, January 2009

[43] Z. Zhang, B. D. Rao, Extension of SBL Algorithms for the Recovery of Block Sparse Signals with Intra-Block Correlation, *IEEE Trans. on Signal Processing*, vol. 61, no. 8, pp. 2009 - 2015, 2013.

[44] P. Zhao, G. Rocha, and B. Yu, The composite absolute penalties family for grouped and hierarchical variable selection, *Annals of Statistics*, Vol. 37 (6A), pp. 3468–3497, 2009.

[45] H. Zou and R. Li, One-step Sparse Estimates in Nonconcave Penalized Likelihood Models, *The Annals of Statistics*, Vol. 36, N. 4, pp. 1509–1533, 2008

[46] D. Wipf and S. Nagarajan, Iterative Reweighted l_1 and l_2 Methods for Finding Sparse Solutions, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 4, n. 2, pp. 317–329 April 2010

[47] D. Wipf, B. D. Rao, and S. Nagarajan, Latent Variable Bayesian models for promoting sparsity, *IEEE Transactions on Information Theory*, Vol. 57, n. 9, pp. 6236–6255, September 2011

[48] L. Yu, J.P. Barbot, G. Zheng, and H. Sun, Compressive Sensing for Cluster Structured Sparse Signals: Variational Bayes Approach, available on http://hal.archives-ouvertes.fr/docs/00/57/39/53/PDF/

[49] M.Yuan and Y. Lin, Model Selection and Estimation in Regression with Grouped Variables, *J. R. Statist. Soc. B*, vol. 68, pp.59–67, 2006.

[50] L. Yuan, Jun Liu, and J. Ye, Efficient Methods for Overlapping Group Lasso, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 9, 2013.

[51] L. Yu, H. Sun, J.P. Barbot, G. Zheng, Bayesian Compressive Sensing for Cluster Structured Sparse Signals", *Signal Processing* Vol. 92, Issue 1, pp. 259–269, January 2012.

[52] S. Xiang, X, Shen, J. Ye, Efficient Sparse Group Feature Selection via Nonconvex Optimization, *Journal of Machine Learning Research*,Vol. 28, N. 1, 284-292, 2013.

Table 1: Proposed MB-NFCS algorithm

Inputs: CS matrix Φ, measurements y, the non separable sparsity inducing function $F_\mu(\lambda)$ and its partial derivatives $\frac{\partial F_\mu(u)}{\partial |u_i|}, i = 1, ..., N$, Fista acceleration algorithm [4] A_{τ_n}, $const > 0$, $\gamma > 0$, $\mu > 0$, $\lambda_0 > 0$, $\beta > 0$

$u_0^0 = \tilde{u}_0^0 = \Phi^T y,$ $\{Initialize\}$

$w_i^0 = \left.\frac{\partial F_\mu(u)}{\partial |u_i|}\right|_{u=u_0^0} \quad i = 1, .., N,$ $\{First\,weights\,estimation\}$

Outer penalization Loop

I. `for` k`=0,1,...` `do`

     $n = 0$

Inner FBS minimization loop

II. $\{Forward\,updating\,step\}$

$$v_{n+1}^k = u_n^k + \beta \Phi^T \left(y - \Phi u_n^k \right)$$

$\{Backward\,minimization\,step\}$

$$\tilde{u}_{n+1}^k = \arg\min_{u \in R^N} \left\{ \lambda_k \sum_{i=1}^N w_i^k |u_i| + \frac{1}{2\beta} ||u - v_{n+1}^k||_2^2 \right\}$$

$\{Acceleration\,Strategy\}$

$$u_{n+1}^k = A_{\tau_n}(\tilde{u}_{n+1}^k, \tilde{u}_n^k)$$

$\{Inner\text{-}Loop\,exit\,test\}$

if $\frac{|\mathcal{P}\left(f_\mu, \lambda_k u_{n+1}^k,\right) - \mathcal{P}\left(f_\mu, \lambda_k, u_n^k,\right)|}{|\mathcal{P}\left(f_\mu, \lambda_k, u_{n+1}^k\right)|} > \gamma \cdot \lambda_k, \quad n = n+1$

     goto II.

     else

$$\lambda_{k+1} = const \cdot \mathcal{P}\left(f_\mu, \lambda_k, u_{n+1}^k \right) \qquad \{Updating\,\lambda_k\}$$

$$w_i^{k+1} = \left.\frac{\partial F(u)}{\partial |u_i|}\right|_{u=u_{n+1}^k} \qquad \{Updating\,the\,weights\}$$

$$u_0^{k+1} = u_{n+1}^k$$

`until convergence`

Output: $\hat{u} = u_0^{k+1}$ as an approximation of x.

Table 2: Influence of the value of d on the performance of MB-NFCS algorithm

G	$d=1$		$d=2$	
	Iterations	Time in sec.	Iterations	Time in sec.
2	283	0.17	138	0.15
4	454	0.21	329	0.21
6	273	0.17	275	0.21
8	211	0.16	181	0.17

(a)   (b)
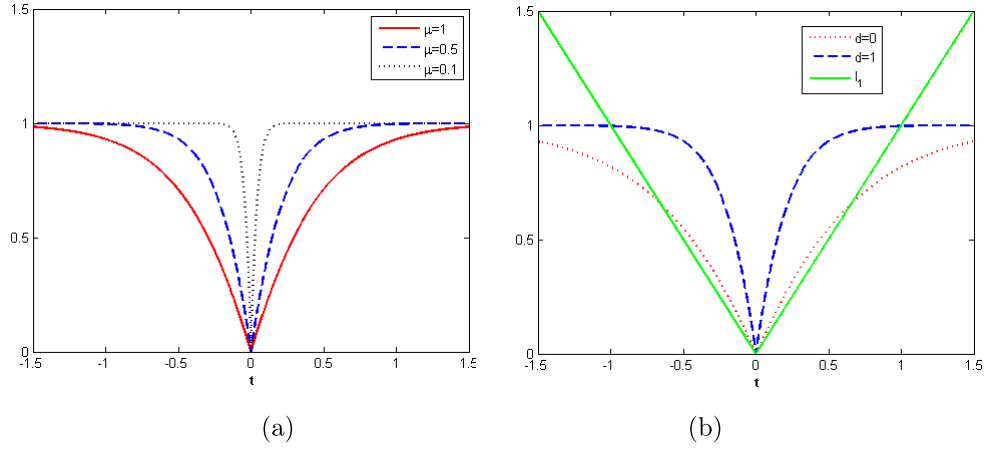
Figure 1: Nonconvex nonseparable sparsity inducing function ψ_μ: (a) evaluated for $d = 1$ and different values of μ, (b) compared, for fixed $\mu = 0.5$, with both the nonconvex separable ($d = 0$) and the l_1 sparsity inducing functions.

1

(a)   (b)
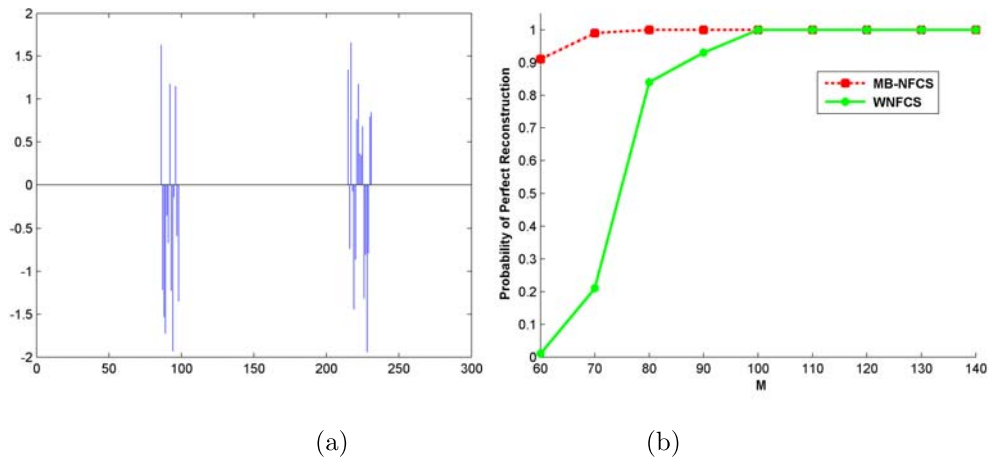
Figure 2: Cluster Structured Signal: $N = 256, K = 30, G = 2, M \in [60, 140]$. Comparison between the results obtained by MB-NFCS and WNFCS. (a) A sample signal. (b). Probability of Perfect Reconstruction as a function of the number of measurements.

$u^0\downarrow$ $u^1\downarrow$ $u^2\downarrow$ $u^3\downarrow$

$w^1\nearrow$ $w^2\nearrow$ $w^3\nearrow$ w^4

x $\hat{u}=u^5$
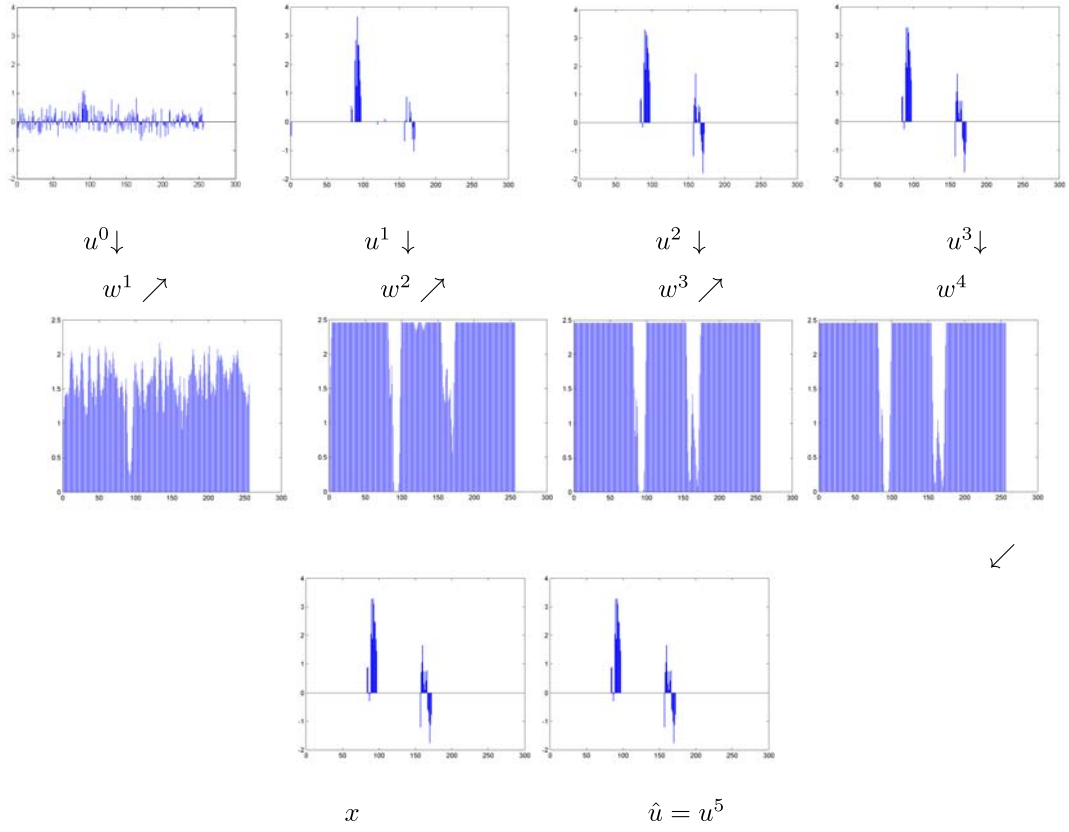
Figure 3: Evolution of the weights $w_i^{k+1}=\frac{\partial F(u)}{\partial|u_i|}\Big|_{u=u^k}$ for each change of the penalization parameter λ_k, ($N=256$, $K=30$, $G=2$, $M=60$); (first row) reconstruction u^k; (second row) weights w_i^{k+1}; (third row) original signal and final perfect reconstruction.
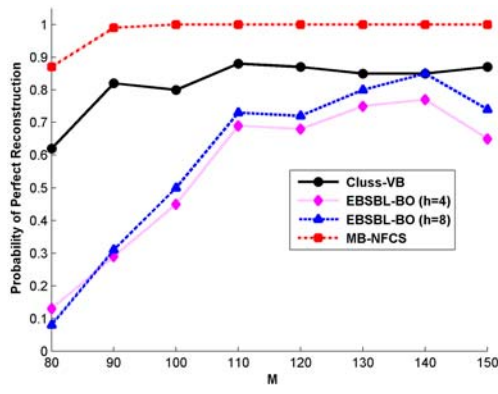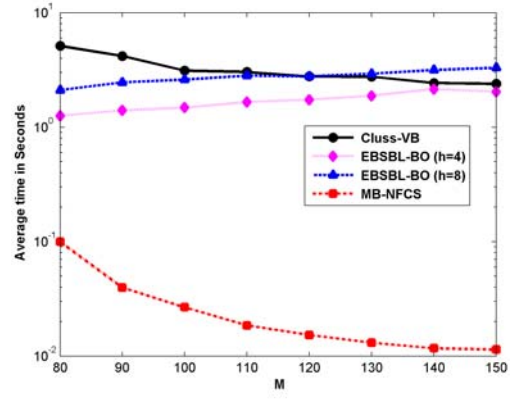
(a)   (b)

Figure 4: Experiment 1-1: $N = 256, K = 40, G = 8, M \in [80, 150]$. Comparison between reconstruction results obtained by MB-NFCS and those obtained by CLUSS-VB, EBSBL-BO($h = 4, h = 8$). (a) Probability of Perfect Reconstruction as a function of the number of acquisitions M. (b) The average perfect reconstruction time in seconds.
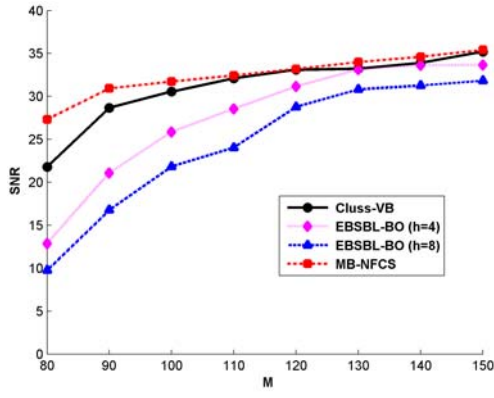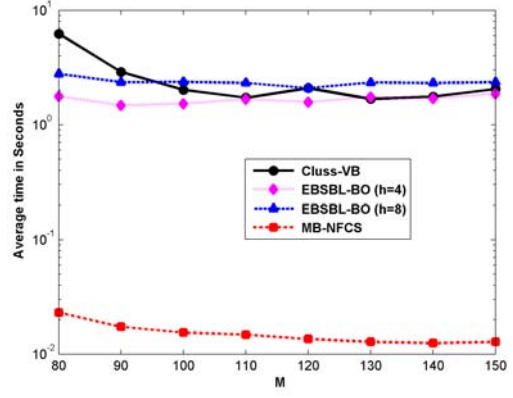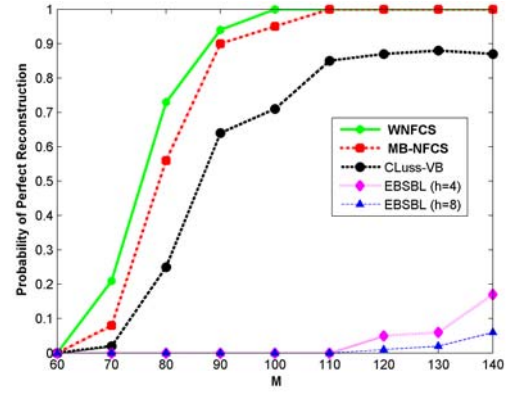
(a)   (b)

Figure 5: Experiment 1-2: $N = 256, K = 40, G = 8, M \in [80, 150]$. Measurements corrupted by a white noise with standard deviation $\sigma = 0.01$. Comparison between reconstruction results obtained by MB-NFCS and those obtained by CLUSS-VB, EBSBL-BO ($h = 4, h = 8$). (a) Mean SNR as a function of the number of acquisitions M. (b) The average time in seconds.
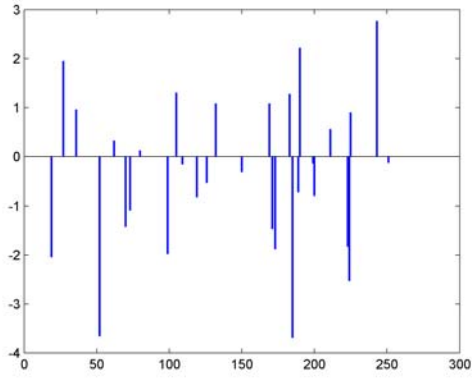
(a)   (b)

Figure 6: Spike Signal: $N = 256, K = 30, M \in [60, 140]$. Comparison between the results obtained by MB-NFCS, WNFCS, Cluss-VB, EBSBL-BO ($h = 4$, $h = 8$). (a) Sample signal. (b). Probability of Perfect Reconstruction as a function of the number of measurements M.
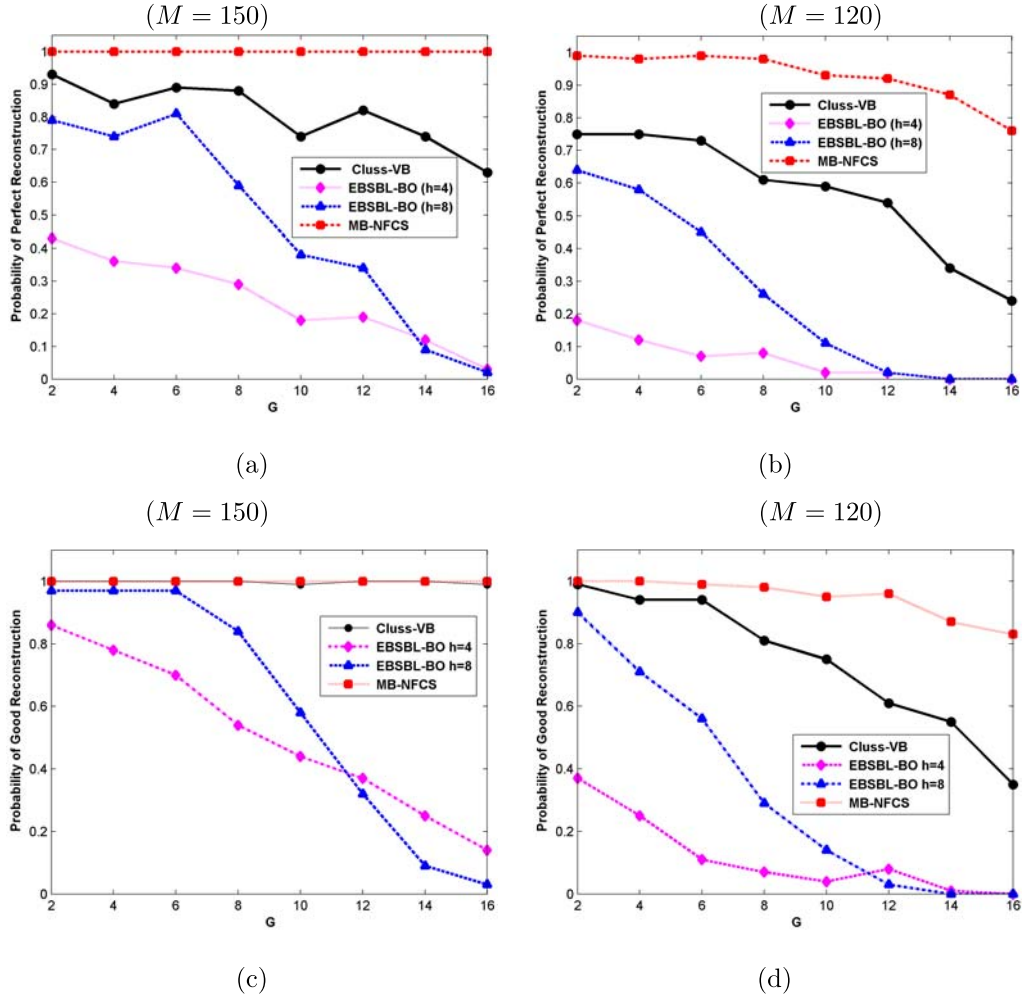
$(M = 150)$ $(M = 120)$

(a)   (b)

$(M = 150)$ $(M = 120)$

(c)   (d)

Figure 7: Experiment 3: $N = 512, K = 60, G \in [2, 4, 6, 8, 10, 12, 16]$. Comparison between reconstruction results obtained by MB-NFCS, CLUSS-VB, EBSBL-BO ($h = 4$, $h = 8$). (a) , (b) Probability of Perfect Reconstruction (NMSE$\leq 10^{-8}$) as a function of the number of groups G, $M = 150$ and $M = 120$ respectively; (c) ,(d) Probability of Good Reconstruction (NMSE$\leq 10^{-3}$) as a function of the number of groups G, $M = 150$ and $M = 120$ respectively.
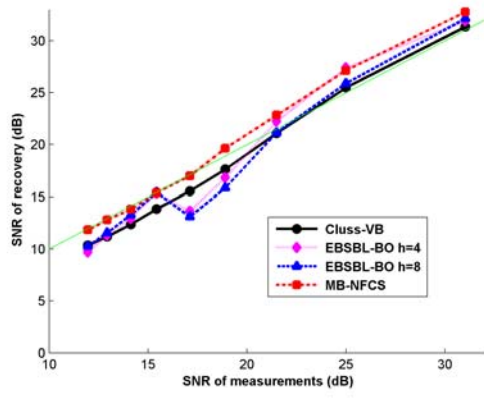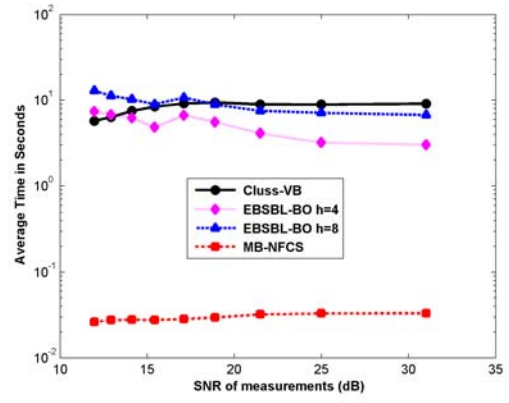
(a)   (b)

Figure 8: Experiment 4: $N = 512$, $K = 60$, $G = 10$, $M = 100$ measurements corrupted by white noise with standard deviation σ ranging from $\sigma = 0.01$, to $\sigma = 0.09$. (a) Mean SNR as a function of the number of measurements. (b) The average time in seconds.
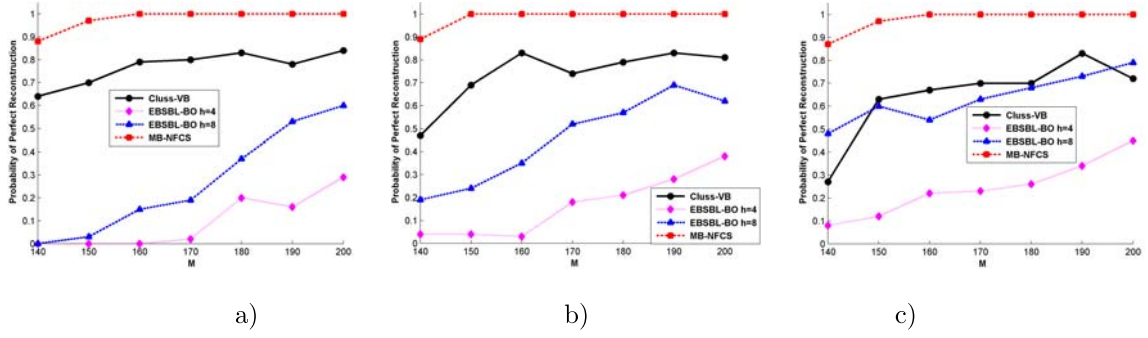
a)   b)   c)

Figure 9: Experiment 5: $N = 512$, $K = 80$ and $G = 10$, $M \in [140, 200]$. Performance comparisons: MB-NFCS , CLUSS-VB, EBSBL-BO ($h = 4$, $h = 8$). Probability of Perfect Reconstruction as a function of the number of measurements. (a) Experiment 5-1: Intra-block correlation equal to zero. (b) Experiment 5-2: Intra-block correlation uniformly varying between 0 and 1. (c) Intra-block correlation uniformly varying between 0.95 and 1.
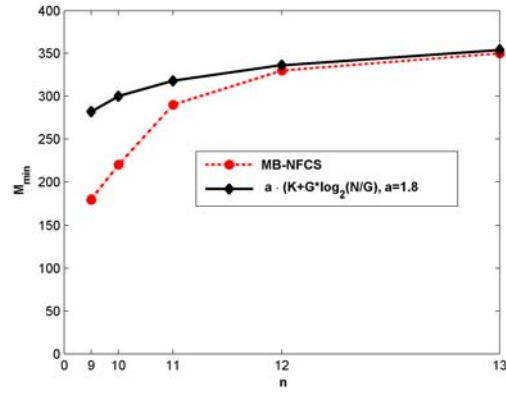
Figure 10: Experiment 6: Minimum number of acquisitions M_{min} required to achieve perfect signal reconstruction by MB-NFCS algorithm and theoretical bound (eq. (16) with $a = 1.8$), as a function of $N = 2^n$. (The displayed results correspond to N = 512, 1024, 2048, 4096, 8192).

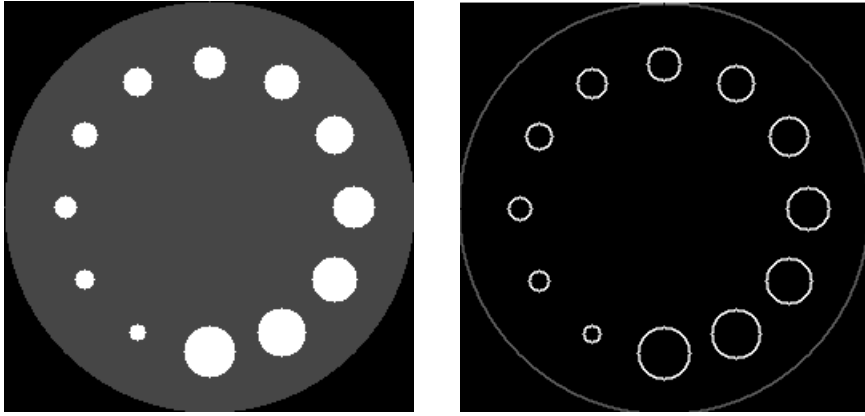Figure 11: 256 ×256 Original Phantom Image and its Gradient.

(a)   (b)   (c)
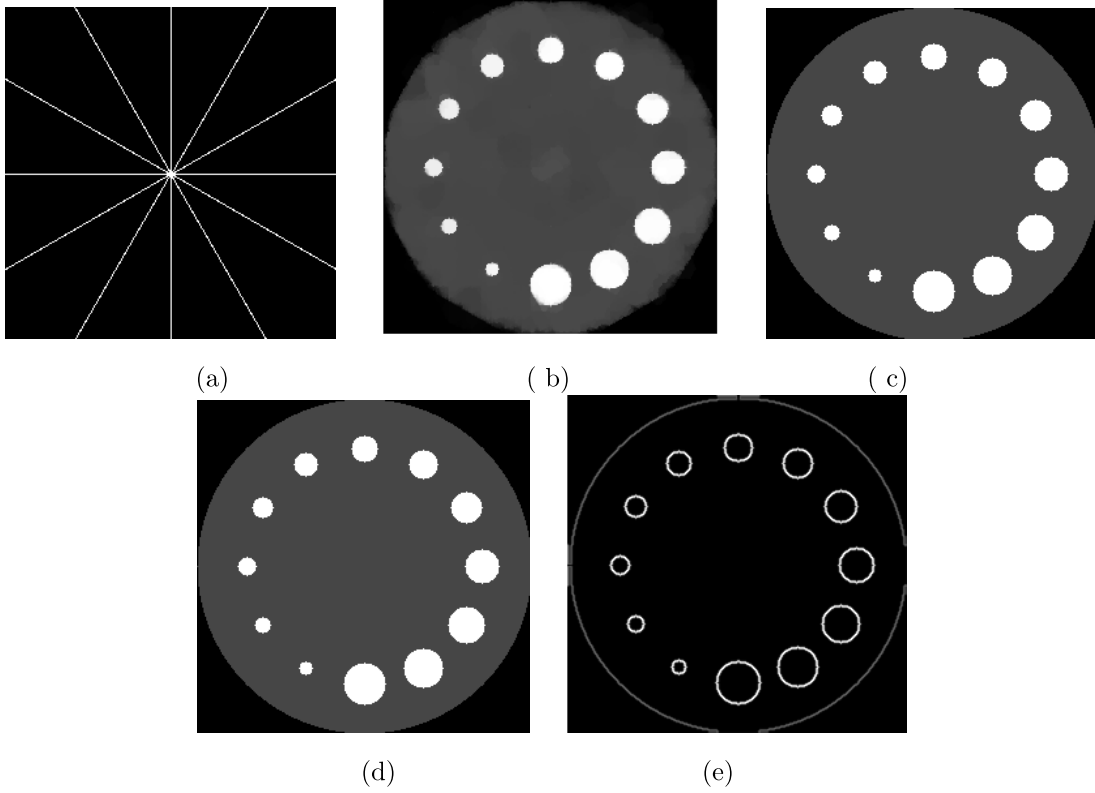

(d)   (e)

Figure 12: Experiment 7: Two-Dimensional Case:(a) Radial Mask 6 rays (2.5 %), (b) l_1 Reconstruction SNR=20.85 Time 70 sec. (c) Nonconvex Separable Reconstruction SNR=69.3, Time=316 sec. (d) Nonconvex Nonseparable Reconstruction (d=1) SNR=91.48 Time 148 sec. (e) Gradient of Nonconvex Nonseparable Reconstruction (d=1).