Greedy methods for simultaneous sparse approximation

Leila Belmerhnia, El-Hadi Djermoune, David Brie

GREEDY METHODS FOR SIMULTANEOUS SPARSE APPROXIMATION

Leila Belmerhnia, El-Hadi Djermoune, David Brie

CRAN, Université de Lorraine, CNRS
Boulevard des Aiguillettes, BP 70239, 54506 Vandoeuvre, France
firstname.lastname@univ-lorraine.fr

ABSTRACT

This paper extends greedy methods to simultaneous sparse approximation. This problem consists in finding good estimation of several input signals at once, using different linear combinations of a few elementary signals, drawn from a fixed collection. The sparse algorithms for which simultaneous versions are proposed are namely CoSaMP, OLS and SBR. These approaches are compared to Tropp's S-OMP algorithm using simulation signals. We show that in the case of signals exhibiting correlated components, the simultaneous versions of SBR and CoSaMP perform better than S-OMP and S-OLS.

Index Terms— Simultaneous sparse approximation, Greedy algorithms, Orthogonal Matching Pursuit

1. INTRODUCTION

Sparse representation has become widely used in recent years, to solve many ill-posed problems in several domains such as compressive sensing, denoising, audio processing [1, 2] and spectral unmixing [3]. This is a linear inverse problem in which we seek to represent an input signal using the minimum number of vectors from an overcomplete dictionary. The problem of finding the sparsest vector $\mathbf{x}$ corresponding to a given observation vector $\mathbf{y}$ is a NP-hard problem so it can not be solved in a polynomial time [4]. Thus, several sub-polynomial approaches have been explored and have shown a good reconstruction performance. Typically, those approaches are based on convex relaxation, non-convex local optimization or greedy pursuit strategy. Indeed, exploiting the sparsity constraint in such problems ensure that when the signal $\mathbf{x}$ is sufficiently sparse, it can be uniquely determined from a set of known measurements $\mathbf{y} = \mathbf{\Phi x}$ where the dictionary $\mathbf{\Phi}$ is a finite collection of elementary signals. The potential of this type of representation motivated many researchers to consider several extensions of sparse representation models. In this context, simultaneous sparse approximation appeared as a natural extension of the standard single measurement vector (SMV) approaches. This sparse representation allows a good approximation of a set of sparse signal vectors sharing common non-zeros support.

The simultaneous sparse representation models, called also multiple measurement vectors (MMV) has been studied from several angles of view [5], and different approaches have been proposed, using greedy strategies [6] such as S-OMP [7], convex relaxation methods [8], randomized algorithms such as REduce MMV and BOost (ReMBo) [9]. The MMV problem has also been approached using non parametric algorithm, namely M-FOCUSS [10], and with bayesian strategy using M-SBL (Multiple Sparse Bayesian Learning) [11], as well as other union of subspace models such as block sparsity and tree structured sparsity [12–14].

In this paper we focus on greedy pursuit approaches. This choice was motivated by the fact that methods based on greedy pursuit are in general more practical in terms of computational demands and it has been proven that this type of approaches leads to correct solution under well-defined conditions. Thus, we present extension of some methods known for their efficiency in solving standard sparse approximation to the simultaneous case. These methods are S-CoSaMP, S-OLS and the forward-backward method S-SBR (Simultaneous Single Best Replacement, see [15] for standard SBR). The remainder of this paper is as follows. Section 2 presents the simultaneous sparse approximation problem. In section 3, we introduce the simultaneous versions of the aforementioned algorithms. Simulation results are presented in section 4 and the conclusions are drawn in section 5.

2. PROBLEM FORMULATION

Let $\mathbf{Y} \in \mathbb{C}^{M\times K}$ be the observation matrix and $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ a sensing matrix (called overcomplete dictionary when $M < N$) such that $\mathbf{Y} = \mathbf{\Phi X}$ for some $\mathbf{X} \in \mathbb{C}^{N\times K}$. The simultaneous sparse approximation problem is formulated as:

$$\min_{\mathbf{X}} ||\mathbf{X}||_0 \text{ s.t. } \mathbf{\Phi X} = \mathbf{Y} \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K]$ and $\mathbf{x}_i$ is the i-th column of $\mathbf{X}$. $||\mathbf{X}||_0 = |\Omega|$, where $\Omega = \text{supp}(\mathbf{X}) = \{1 \leq i \leq N \,|\, \mathbf{x}^i \neq 0\}$ is the set of active atoms, and $\mathbf{x}^i$ stands for the i-th row of $\mathbf{X}$.

Thus, the simultaneous sparse approximation problem consists in recovering the matrix $\mathbf{X}$ knowing the signal matrix $\mathbf{Y}$ and dictionary $\mathbf{\Phi}$ under the assumption that all the columns $\mathbf{x}_i$ of $\mathbf{X}$ have the same sparsity profile. This can

be formulated in the coefficient matrix $\mathbf{X}$ having a minimum number of non-zeros rows. One of the first algorithms developed to solve the simultaneous sparse approximation problem is S-OMP [7]. This is an iterative procedure that selects an atom at each iteration and updates the residual. The selected atom is the one that maximizes the correlation with the current residual. This algorithm exhibits a very low computational complexity but it lacks accuracy for highly correlated signal components. The aim of this paper is to propose other alternatives (although based on the same principle than S-OMP), which are extensions of CoSaMP, OLS and SBR to the simultaneous case. It will be shown that these approaches performs better than S-OMP for signals containing very correlated components.

3. EXTENSION OF SMV GREEDY ALGORITHMS FOR SIMULTANEOUS SPARSE APPROXIMATION

3.1. Simultaneous CoSaMP and OLS

The CoSaMP [16] and OLS [17] algorithms are based on greedy pursuit just like OMP. CoSaMP incorporates some combinatorial techniques that guarantees speed of the solution and provides rigorous error bounds. We propose an extension of this approach to solve the simultaneous sparse approximation problem. The resulting algorithm is called Simultaneous CoSaMP (S-CoSaMP). OLS seems to be very similar to OMP algorithm, in some way that a lot of confusion has been made between the two [18]. Indeed, both algorithms achieve the same steps, the difference is only in the selection procedure. The greedy selection in OMP algorithm is based on the inner products between residual and the column vector $\boldsymbol{\phi}_n$ of $\boldsymbol{\Phi}$ that leads to the minimal residual error. In OLS, the selection is also greedy, but this step is performed after orthogonal projection of the signal onto the selected elements. Thus, the orthogonal projection of the elements $\boldsymbol{\phi}_n$ before the calculation of the inner products characterizes the OLS procedure and guarantees to select the element resulting in the smallest error after projection.The difference between the standard algorithms and the simultaneous ones lies in the selection and estimation steps. Indeed, in the selection step, the simultaneous algorithms compute the correlation between columns of $\boldsymbol{\Phi}$ and the current residual using the Frobenius norm instead of the Euclidean norm. The Frobenius norm is also used in the estimation of the target signal to solve the least-squares problem. A summary of the S-CoSaMP algorithm and S-OLS are given in Fig. 1 and 2.

Fig. 1. Simultaneous CoSaMP algorithm

- **Input:** Data matrix $\mathbf{Y} \in \mathbb{C}^{M\times K}$; Dictionary $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$; sparsity parameter s; setting parameter μ; a stopping criterion.
- **Output:** Set of active atoms Ω; $s-$sparse matrix $\mathbf{X} \in \mathbb{C}^{N\times K}$ and the residual matrix $\mathbf{R}$.

1. **Initialization**: $\mathbf{R}_0 = \mathbf{Y}, \Omega_0 = \emptyset, \mathbf{X}_0 = 0$.
2. **For** $i = 1;\ i = i + 1$ until the stopping criterion is satisfied **do**:
   (a) Find the μs columns of $\boldsymbol{\Phi}$ that are most correlated with residual:
   $$\Gamma \in \arg\max_{|T|\leq \mu s} \sum_{n\in T} ||\mathbf{R}_{i-1}^t \boldsymbol{\phi}_n||_F^2$$
   (b) Let $T = \Omega_{i-1} \cup \Gamma$. Determine the best coefficients for approximating the residual matrix: $\mathbf{U}_i = \arg_{\mathbf{U}} \min ||\mathbf{R}_{i-1} - \boldsymbol{\Phi}_T \mathbf{U}||_F$
   (c) Save the s largest rows of $\mathbf{U}_i$ in the Euclidean norm sense: $\mathbf{X}_i = [\mathbf{U}_i]_s$
   (d) Set $\Omega_i = \text{supp}(\mathbf{X}_i)$, and update the residual $\mathbf{R}_i = \mathbf{Y} - \boldsymbol{\Phi}_{\Omega_i}\mathbf{X}_i$
3. **Return** Ω_i and $\mathbf{X}_i$ and $\mathbf{R}_i$.

Fig. 2. Simultaneous OLS algorithm

- **Input:** Data matrix $\mathbf{Y} \in \mathbb{C}^{M\times K}$; Dictionary $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$; sparsity parameter s.
- **Output:** Set of active atoms Ω; $s-$sparse matrix $\mathbf{X} \in \mathbb{C}^{N\times K}$ and the residual matrix $\mathbf{R}$. .

1. **Initialization** : $\mathbf{R}_0 = \mathbf{Y}, \Omega_0 = \emptyset$
2. **For**: $i = 1;\ i = i + 1$ until the stopping criterion is satisfied **do**:
   (a) $n_{\max} = \arg_n \min_{\Omega_i^n = \Omega_{i-1}\cup n} \|\mathbf{Y} - \boldsymbol{\Phi}_{\Omega_i^n}\boldsymbol{\Phi}_{\Omega_i^n}^{\dagger}\mathbf{Y}\|_F$
   (b) $\Omega_i = \Omega_{i-1} \cup n_{\max}$
   (c) $\mathbf{X}_{\Omega_i} = \boldsymbol{\Phi}_{\Omega_i}^{\dagger}\mathbf{Y}$
   (d) $\mathbf{R}_i = \mathbf{Y} - \boldsymbol{\Phi}_{\Omega_i}\mathbf{X}_{\Omega_i}$
3. **Return** Ω_i and $\mathbf{X}_{\Omega_i}$ and $\mathbf{R}_i$.

3.2. Simultaneous Single Best Replacement

In its single measurement form, single best replacement [15] is an algorithm based on minimization of a cost function $\mathcal{I}(\mathbf{x}, \lambda)$, such that:

$$\arg\min_{\mathbf{x}\in\mathbb{C}^N}\{\mathcal{I}(\mathbf{x}, \lambda) = \mathcal{E}(\mathbf{x}) + \lambda||\mathbf{x}||_0\} \tag{2}$$

where $\mathcal{E}$ is the corresponding least-squares error: $\mathcal{E}(\mathbf{x}) = ||\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}||_2^2$ and λ is an hyperparameter controlling the level of sparsity in the desired solution.

Unlike the previous approaches, the selection is not irreversible as it allows adding and removing a new atom to/from the set of active ones. Indeed, SBR selects the index n that minimizes the cost function in (2). So, at each iteration the algorithm tests the N possible single replacements $\Omega \bullet n$, $n = 1, \ldots, N$ then selects the one that yields to maximal decrease of $\mathcal{I}(\mathbf{x}, \lambda)$, i.e. $\mathcal{I}_{\Omega\bullet n}(\lambda) = \mathcal{E}_{\Omega\bullet n} + \lambda\text{Card}(\Omega \bullet n)$. Thus, $\Omega \bullet n$ may be an insertion or a removal of an index in/from the set of active indices Ω such that:

$$\Omega \bullet n = \begin{cases} \Omega \cup \{n\} & \text{if } n \notin \Omega \\ \Omega \setminus \{n\} & \text{otherwise} \end{cases} \tag{3}$$

The version proposed here, consists on adapting this algorithm to solve simultaneous sparse approximation problem. This is done by replacing the norm 2 in $\mathcal{E}$ by the Frobenius norm taking advantage of the joint support of the columns of $\mathbf{X}$. In addition, the support of the coefficient matrix repre-

Fig. 3. Simultaneous SBR algorithm

- **Input:** Data matrix $\mathbf{Y} \in \mathbb{C}^{M\times K}$; Dictionary $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$; regularization parameter λ.
- **Output:** Set of active atoms Ω; $s-$sparse matrix $\mathbf{X} \in \mathbb{C}^{N\times K}$.

1. **Initialization**: $\mathbf{R}_0 = \mathbf{Y}$, $\Omega_0 = \emptyset$.
2. **For** $i = 1$; $i = i + 1$ until the set Ω_i is no longer updated **do:**
   (a) $n_i \in \arg_n \min\{\mathcal{I}_{\Omega_{i-1}\bullet n}(\lambda) = \mathcal{E}_{\Omega_{i-1}\bullet n} + \lambda \mathrm{Card}(\Omega \bullet n)\}$.
   (b) **if** $\mathcal{I}_{\Omega_{i-1}\bullet n_i}(\lambda) < \mathcal{I}_{\Omega_{i-1}}(\lambda)$ **then**
   $\Omega_i = \Omega_{i-1} \bullet n_i$.
   End if
   (c) $\mathbf{X}_{\Omega_i} = \mathbf{\Phi}^{\dagger}_{\Omega_i}\mathbf{Y}$
3. **Return** Ω_i and $\mathbf{X}_{\Omega_i}$.

sents in this version the number of non-zero rows in $\mathbf{X}$. Thus the simultaneous SBR seeks to minimize the following cost function:

$$\arg\min_{\mathbf{X}\in\mathbb{C}^{N\times K}} \{\mathcal{I}(\mathbf{X},\lambda) = \mathcal{E}(\mathbf{X}) + \lambda||\mathbf{X}||_0\} \tag{4}$$

where $\mathcal{E}(\mathbf{X}) = ||\mathbf{Y} - \mathbf{\Phi}\mathbf{X}||_F^2$ and, again, λ is a hyperparameter controlling the tradeoff between data-fitting and sparsity. S-SBR terminates when no index yields to decreasing the cost function in (4) anymore. This algorithm has been shown to be very efficient for dictionary with high correlated atoms, however it involves the calculation of the inverse of the Gram matrix $\mathbf{\Phi}_\Omega^T\mathbf{\Phi}_\Omega$ so, instabilities may occurs when in case of ill-conditioned dictionary for low λ [15]. The algorithm is summarized in Fig. 3.

4. SIMULATION RESULTS

This section aims at evaluating the performances of the proposed algorithms using simulated signal trying to mimic situations arising is spectroscopic mixture analysis. In order to measure the quality of the reconstruction, we propose to compare the reconstruction of the signal matrix with the best sparse approximation obtained knowing exactly the position and amplitudes of the s-largest entries of $\mathbf{X}$ (active atoms). So, we propose to test the three developed algorithms to solve simultaneous sparse approximation on 20 observation vectors at different signal-to-noise ratios (SNR) levels and evaluate the performances of each in comparison with the well known S-OMP. The signal matrix $\mathbf{Y} \in \mathbb{R}^{20\times 20}$ is built as follows: Each column in $\mathbf{Y}$ is the sum of 3 Gaussian functions with standard deviation 0.5. Their weights are drawn from 3 sinusoids with random frequencies and amplitude 1. The centers of the Gaussians are set to $(8 - d, 8, 8 + 2d)$, where $d \in \{0.66, 1, \ldots, 5\}$. The variable d controls the correlation between the signal components. This scheme is represented in figure 4 for 3 values of d. Then, the matrix $\mathbf{Y}$ is perturbed by an additive white Gaussian noise $\mathbf{E}$ such that the SNR defined by $\mathrm{SNR} = \|\mathbf{\Phi}\mathbf{X}\|_F^2/\|\mathbf{E}\|_F^2$ varies between -20 and 20 dB. The overcomplete dictionary $\mathbf{\Phi} \in \mathbb{R}^{20\times 60}$ is

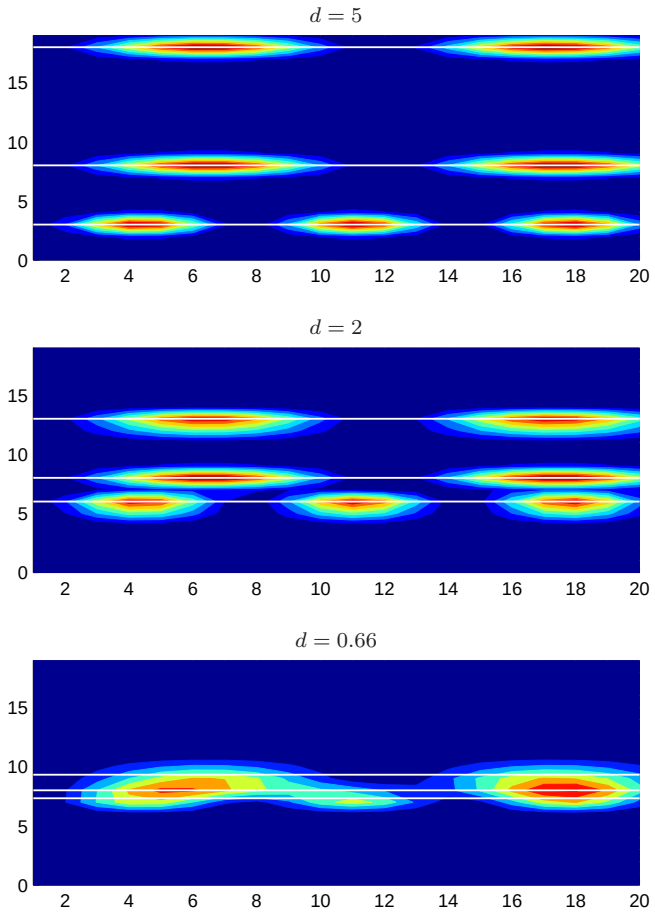

Fig. 4. Representation of the data matrix $\mathbf{Y}$ for $d = 5$, $d = 2$, and $d = 0.66$. The signal is composed of 3 Gaussians with sinusoidal amplitudes along x-axis

composed of 60 normalized Gaussian atoms with a standard deviation of 0.5. Their centers are ranging from 0 to 19.66 with a step of 0.33. As a consequence, the correlation coefficient between two consecutive atoms (mutual coherence [4]) is about 0.93, meaning that the dictionary is far from orthogonality. The setting parameters are: $s = 6$, $\mu = 0.5$ and $\lambda = 10^{-5}$. Here the objective is not to evaluate performances in the worst case (using the RIP) but to localize the exact recovery regions (defined by d and the SNR) for each method. So we use the detection rate (related to the number of signal components correctly identified) as a measure of exact recovery.

The results are shown in figure 5. It is obvious that all the methods achieve exact recovery for large d and high SNR (100% detection rate). However, detection rates differ from one to another when the SNR is low and/or the signal components are close one to the other (low d). S-OMP exhibits the smallest exact recovery region as S-OLS, but the latter shows better performances for moderate SNRs. S-CoSaMP leads to better detection rates than S-OMP for high d and low SNR

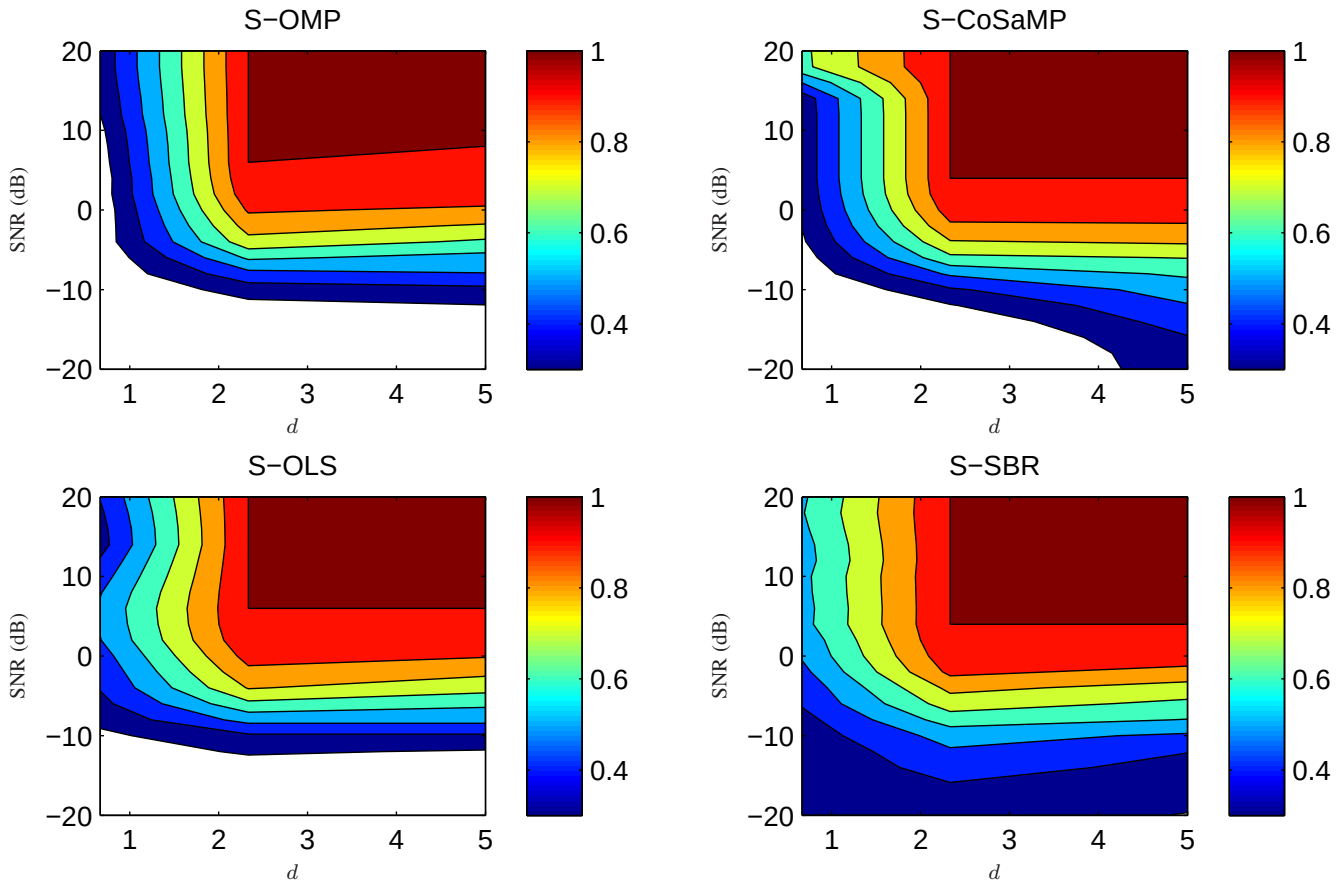

Fig. 5. Detection rate of the simultaneous methods versus SNR and d for 100 Monte Carlo simulations

scenario. Nevertheless, those three methods failed completely to recover one right atoms for signal sorely corrupted as represented by the white region in figure 5. By contrast, S-SBR exhibits the best behavior at low SNR and/or low d for sufficiently small λ. Generally speaking, S-SBR and S-CoSaMP show the best performances, as their SMV versions, but at the cost of higher numerical demands. To show the advantage of the simultaneous approximation, we present in figure 6 the results obtained for the same simulation using the standard SBR on each column of $\mathbf{Y}$ independently. We observe that the performances are dramatically increased by taking advantage of the fact that all the columns $\mathbf{Y}$ have the same support.

The evolution of Frobenius norm of residual returned by each method is represented in Fig. 7, for the following scenario: SNR $= 10$ and $d = 2.33$. It appears that algorithms such as S-OMP, S-CoSaMP and S-OLS failed to achieve the desired accuracy for signal reconstruction even for a large number of iterations, whereas the reconstruction error is decreasing using S-SBR. However, we note that S-CoSaMP algorithm is much faster than others as it achieves the lowest error after only eight iterations while S-OLS and S-SBR need one more iteration and S-OMP two more iterations.

5. CONCLUSION

We presented extensions of greedy algorithms to solve the simultaneous sparse approximation problem. In the case of correlated signal components or low SNR, S-SBR and S-CoSaMP, support recoveries than S-OLS and Tropp's S-OMP in terms of detection rates and thus reconstruction error. Algorithms such as S-OLS and S-CoSaMP have stronger capacities of exploration in comparison with S-OMP, while the objective function of S-SBR is guaranteed to decrease performing both insertion and removal of atoms in/from the support. Furthermore, we found that for highly correlated dictionaries they outperform S-OMP and require less iterations to meet recovery conditions but at the expense of computational demands. As future work, we are planning to use these algorithms in hyperspectral unmixing and multidimensional harmonic retrieval of spectroscopy data [19].

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, 2006.

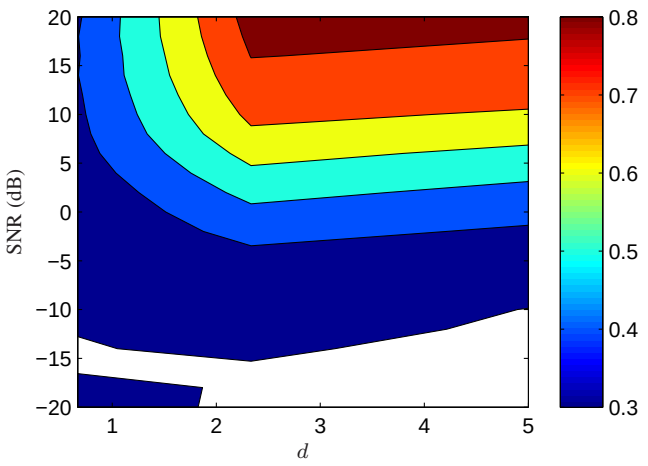

Fig. 6. Detection rate of the standard SBR versus SNR and d for 100 Monte Carlo simulations

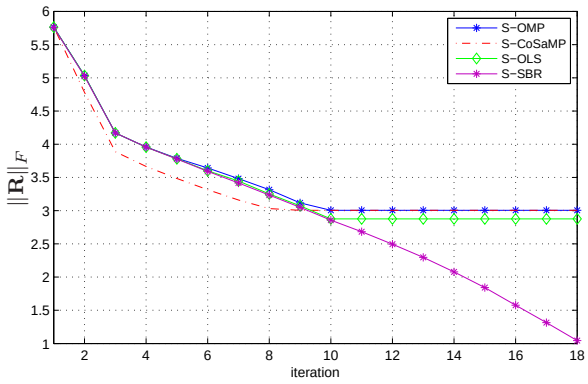

Fig. 7. Evolution of the reconstruction error $\mathbf{R} = \mathbf{Y} - \mathbf{\Phi}_\Omega \mathbf{X}_\Omega$ according to the number of iterations (SNR $=$ 10 dB and $d = 2.33$)

[2] E.J. Candès, J. Romberg, and T. Tao, "Stable signal recovery for incomplete and annucarate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, May 2006.

[3] M. D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, pp. 2014–2039, June 2011.

[4] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, pp. 948–958, June 2010.

[5] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, 2006.

[6] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct 2004.

[7] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, pp. 572–588, 2006.

[8] J. A.Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, pp. 589–602, March 2006.

[9] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, pp. 4692–4702, Oct 2008.

[10] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, pp. 2477–2488, July 2005.

[11] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, pp. 3704–3716, July 2007.

[12] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, pp. 5302–5316, Nov 2009.

[13] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, pp. 3042–3054, Jun 2010.

[14] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.

[15] C. Soussen, J. Idier, D. Brie, and J. Duan, "From bernoulli-gaussian deconvolution to sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 59, pp. 4572–4584, Oct 2011.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Anal.*, vol. 26, pp. 301–321, May 2009.

[17] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least-squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, pp. 1873–1896, 1989.

[18] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," Tech. Rep., University of Edinburgh, March 2007.

[19] S. Sahnoun, E. H. Djermoune, and D. Brie, "Sparse modal estimation of 2-D NMR signals," in *IEEE ICASSP*, Vancouver, Canada, 2013, pp. 8751–8755.