A SPARSE APPROACH FOR DOA ESTIMATION WITH A MULTIPLE SPATIAL INVARIANCE SENSOR ARRAY

M-A. Bisch[*], *S. Miron*[*], *D. Brie*[*], *D. Lejeune*[+]

[*]Centre de Recherche en Automatique de Nancy
Université de Lorraine, Vandœuvre-lès-Nancy, France F-54500
Email: marc-abel.bisch@univ-lorraine.fr
[+]Ministère de la défense, Paris, France

ABSTRACT

In this paper, we introduce a sparse direction-of-arrival (DOA) estimation algorithm for sensor arrays presenting multiple scales of spatial invariance. We exploit the Khatri-Rao structure of the over-complete steering vector dictionary, corresponding to this array geometry, in order to devise a computationally efficient sparse estimation approach. This approach is based on an iterative refinement and pruning strategy of the dictionary. We show, in numerical simulations, that our approach outperforms the state-of-the-art approach based on a Candecomp/Parafac (CP) decomposition, proposed by Miron *et al.* in 2015.

Index Terms— Multidimensional signal processing, direction-of-arrival estimation, Khatri-Rao dictionary, multi-invariance array, sparse estimation.

I. INTRODUCTION

Large-aperture sensor arrays is a hot topic in array processing, as more and more domains (telecommunications, radar, astronomy, etc.) are using them lately. For example, the Murchison Wide-field Array (MWA) is a radio-telescope composed of more than 3 000 antennas, spread over more than 3 km^2 in the Australian desert. MWA is capable of enhancing the angular resolution to about 2 arcmin. The data flow produced by this type of radio telescope approaches the GB/s, representing a major challenge for the processing algorithms. A particular type of large-aperture array, presenting multiple spatial scale invariance, *i.e.,* a *multi-scale array*, was introduced by Miron *et al.* in [1]; an efficient Candecomp/Parafac (CP) based algorithm for direction-of-arrival (DOA) estimation with such an array was also proposed in [1]. In this paper we propose an alternative solution for DOA estimation with a multi-scale array, based on a sparse estimation approach. In an earlier work, Malioutov *et al.* [2] showed that the DOA estimation can be formulated as a sparse estimation problem, by discretizing the DOA domain. Our approach is based on the same idea; the novelty is the exploitation of the Khatri-Rao structure of the over-complete dictionary of steering vectors, specific to the multiple spatial scale invariance array, in order to propose a computationally efficient algorithm. In Section II, we briefly present the structure of a multi-scale sensor array, as introduced in [1], as well as the associated data model. In Section III we introduce the principle of sparse DOA estimation and the properties of the Khatri-Rao dictionary. The proposed KR-SOLS algorithm is described in section IV and his performances are studied on numerical simulations in section V. Finally, some conclusions are drawn in section VI.

II. THE MULTIPLE SPATIAL SCALE INVARIANCE SENSOR ARRAY

Consider a *level-1* sensor array composed of L_1 isotropic, identical sensors, indexed by $l_1 = 1, ..., L_1$. By arbitrary, but known translations, we replicate this array L_2 times in space. The newly obtained sensor array is said of "level-2", and comprises L_2 *level-1* subarrays. This operation can be repeated as many times as necessary, in order to generate N *scale levels.* Fig. 1 illustrates such an array with 3 levels and a regular geometric structure; this sensor array will be used in the simulations of section V.

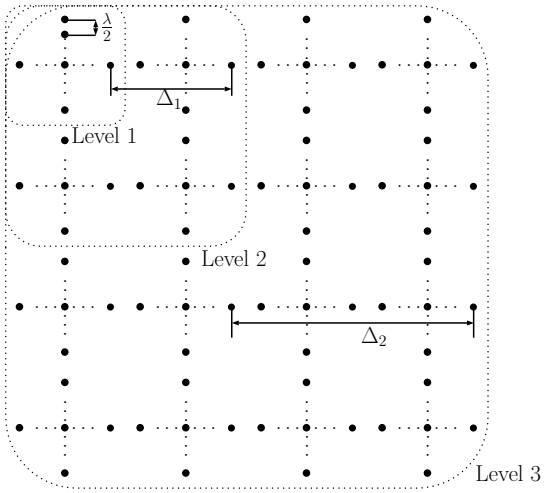

Fig. 1. Multi-scale sensor array with 3 levels.

In the Cartesian system $OXYZ$ associated with this array, the DOA of a source impinging on the array is given by [1] the angles $\theta \in [0, \pi]$ (elevation), measured from the positive Z-axis, and $\varphi \in [0, 2\pi]$ (azimuth), measured from the positive X-axis, or equivalently, by the unit-norm vector $\mathbf{k} = [u \ v \ w]^T$, where u, v, w are the direction-cosines. Denote by $\mathbf{d}_{l_n}^{(n)}$ the spatial displacement of the l_n^{th} subarray of *level-n* with respect to the first one. Under the narrow band assumption, the array manifold $\mathbf{a}(\mathbf{k})$, is a vector of size $(L \times 1)$, with $L = L_1 \cdot \ldots \cdot L_N$, given by:

$$\mathbf{a}(\mathbf{k}) = \mathbf{a}_1(\mathbf{k}) \otimes \ldots \otimes \mathbf{a}_N(\mathbf{k}). \quad (1)$$

In eq. (1), $\mathbf{a}_n(\mathbf{k}) = [\mathrm{e}^{j(2\pi/\lambda)\mathbf{k}^T\mathbf{d}_1^{(n)}}, \ldots, \mathrm{e}^{j(2\pi/\lambda)\mathbf{k}^T\mathbf{d}_{L_n}^{(n)}}]^T$ is the array manifold of size $(L_n \times 1)$ corresponding to the n^{th} scale level, and "$\otimes$" symbolizes the Kronecker product. Consider P sources impinging on the array with distinct DOAs and K time snapshots of the array. Denote $\mathbf{A}_n = [\mathbf{a}_n(\mathbf{k}_1) \ \ldots \ \mathbf{a}_n(\mathbf{k}_P)]$ the $(L_n \times P)$ matrix having on its columns the *level-n* steering vectors for the P sources, and $\mathbf{S}$, of size $(K \times P)$, the matrix containing on its columns the time sequences of the P sources. Then, the data recorded on the array can be expressed as the $(L \times K)$ matrix:

$$\mathbf{Z} = (\mathbf{A}_1 \odot \ldots \odot \mathbf{A}_N)\mathbf{S}^T + \mathbf{N}, \quad (2)$$

where $\mathbf{N}$ is the noise matrix and "$\odot$" symbolizes the Khatri-Rao product of two matrices, *i.e.* the Kronecker product column-wise. Eq. (2) expresses a CP model of the data; a multilinear DOA estimation approach based on this model was proposed in [1].

III. SPARSE DOA ESTIMATION WITH A KHATRI-RAO DICTIONNARY

III-A. Simultaneous sparse approximation for DOA estimation

It was shown [2], [3] that the DOA estimation problem can be reformulated as a simultaneous sparse approximation problem:

$$\min_{\mathbf{X}} \|\mathbf{Z} - \mathbf{\Phi}\mathbf{X}^T\|^2 \text{ s.t. } \|\mathbf{X}\|_{2,0} = P, \quad (3)$$

where $\|\mathbf{X}\|_{2,0}$ represents the number of active columns in $\mathbf{X}$ (which correspond to the columns of $\mathbf{S}$), and $\mathbf{\Phi}$ is the dictionary containing the atoms (steering vectors) obtained by the discretization of the DOA angles θ and φ. Thus, the DOA estimation problem can be recasted as a problem of finding active atoms in an over-complete dictionary. In this work we use the simultaneous greedy-like sparse algorithm called SOLS (Simultaneous Orthogonal Least Squares) [4], to solve the problem in (3). Although the principle of the method is quite simple, the algorithm implementation poses a certain number of problems. Consider N_θ values of θ and N_φ values of φ; then the dictionary $\mathbf{\Phi}$ has $M = N_\theta \cdot N_\varphi$ atoms. $\mathbf{\Phi}$ is constructed in the same manner as the steering vector matrix *i.e.* $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_M]$, with $\boldsymbol{\phi}_m = \boldsymbol{\phi}_{1,m} \odot \cdots \odot \boldsymbol{\phi}_{N,m}$ and $\boldsymbol{\phi}_{n,m} = [\mathrm{e}^{j(2\pi/\lambda)\mathbf{k}_m^T\mathbf{d}_1^{(n)}}, \ldots, \mathrm{e}^{j(2\pi/\lambda)\mathbf{k}_m^T\mathbf{d}_{L_n}^{(n)}}]^T$, where $m = 1, \cdots, M$ and $n = 1, \cdots, N$. To achieve a high DOA resolution, it is necessary to use a small discretization step of the angles θ and φ, *i.e.,* large values of N_φ and N_θ. In this case, the size M of the dictionary grows dramatically and consequently the computational burden.

Besides the computational aspect, the discretization step has also an impact on the accuracy of the sparse estimation algorithm, via the dictionary coherence. For a dictionary $\mathbf{\Phi}$, if its columns $\boldsymbol{\phi}_m (m = 1, \cdots, M)$ have unit norm, the dictionary coherence is defined as $\mu(\mathbf{\Phi}) = \max_{i \neq j} | < \boldsymbol{\phi}_i, \boldsymbol{\phi}_j > |$, with $0 \leq \mu(\mathbf{\Phi}) \leq 1$. It is well-known (see *e.g.,* [5]) that the accuracy of the sparse estimation algorithms is linked to the dictionary coherence; the smaller the coherence, the better the performance. In our case, a small discretization step implies a high dictionary coherence. We propose in this paper an algorithmic strategy for efficiently handling the computational burden and the dictionary coherence, by exploiting the special properties of the Khatri-Rao dictionary presented in the next subsection.

III-B. Khatri-Rao dictionary

The particular structure of a multi-scale sensor array leads to a matrix of steering vectors presenting a Khatri-Rao structure see (eq. (2)). This particular structure is also found in the over-complete dictionary $\mathbf{\Phi}$, which can be written in a similar way as:

$$\mathbf{\Phi} = \mathbf{\Phi}_1 \odot \ldots \odot \mathbf{\Phi}_N, \quad (4)$$

with $\mathbf{\Phi}_n$, $(n = 1, \ldots, N)$, the dictionary corresponding to the n^{th} scale level. Over-complete dictionaries presenting a Kronecker structure, *i.e.* $\mathbf{\Phi} = \mathbf{\Phi}_1 \otimes \ldots \otimes \mathbf{\Phi}_N$, have already been studied in several papers. In [6], Jokar and Mehrmann proved that the global coherence of such dictionaries is given by $\mu(\mathbf{\Phi}) = \max(\mu(\mathbf{\Phi}_1), \cdots, \mu(\mathbf{\Phi}_N))$. Hence, for Kronecker dictionaries, the dictionary with the greatest coherence dominates the global coherence. Going back to our Khatri-Rao dictionary in eq. (4), the following result can be proven:

$$\mu(\mathbf{\Phi}_1) \geq \mu(\mathbf{\Phi}_1 \odot \mathbf{\Phi}_2) \geq \cdots \geq \mu(\mathbf{\Phi}_1 \odot \cdots \odot \mathbf{\Phi}_N). \quad (5)$$

Proof: Denote

$$\mathbf{\Phi}_n = [\boldsymbol{\phi}_{n,1}, \cdots, \boldsymbol{\phi}_{n,M}]. \quad (6)$$

By definition:

$$\mu(\mathbf{\Phi}_1 \odot \mathbf{\Phi}_2) = \max_{i \neq j} | < \boldsymbol{\phi}_{1,i} \otimes \boldsymbol{\phi}_{2,i}, \boldsymbol{\phi}_{1,j} \otimes \boldsymbol{\phi}_{2,j} > |. \quad (7)$$

Yet, (see *e.g* in [6]),

$$< \boldsymbol{\phi}_{1,i} \otimes \boldsymbol{\phi}_{2,i}, \boldsymbol{\phi}_{1,j} \otimes \boldsymbol{\phi}_{2,j} > = < \boldsymbol{\phi}_{1,i}, \boldsymbol{\phi}_{1,j} > \cdot < \boldsymbol{\phi}_{2,i}, \boldsymbol{\phi}_{2,j} > . \quad (8)$$

We can deduce that

$$\begin{aligned} \mu(\mathbf{\Phi}_1 \odot \mathbf{\Phi}_2) &= \max_{i \neq j} | < \boldsymbol{\phi}_{1,i}, \boldsymbol{\phi}_{1,j} > \cdot < \boldsymbol{\phi}_{2,i}, \boldsymbol{\phi}_{2,j} > | \\ &\leq \max_{i \neq j} | < \boldsymbol{\phi}_{1,i}, \boldsymbol{\phi}_{1,j} > | \cdot \max_{i \neq j} | < \boldsymbol{\phi}_{2,i}, \boldsymbol{\phi}_{2,j} > | \end{aligned}$$

$$= \mu(\mathbf{\Phi}_1) \cdot \mu(\mathbf{\Phi}_2) \tag{9}$$

In general

$$\mu(\mathbf{\Phi}) \leq \prod_{n=1}^{N} \mu(\mathbf{\Phi}_n). \tag{10}$$

As $\mu(\mathbf{\Phi}_n) \leq 1$ $(n = 1, \cdots, N)$ we have:

$$\mu(\mathbf{\Phi}_1) \geq \mu(\mathbf{\Phi}_1) \cdot \mu(\mathbf{\Phi}_2) \geq \cdots \geq \prod_{n=1}^{N} \mu(\mathbf{\Phi}_n), \tag{11}$$

and using eq. (9) we obtain:

$$\mu(\mathbf{\Phi}_1) \geq \mu(\mathbf{\Phi}_1 \odot \mathbf{\Phi}_2) \geq \cdots \geq \mu(\mathbf{\Phi}_1 \odot \cdots \odot \mathbf{\Phi}_N) = \mu(\mathbf{\Phi}). \quad \square \tag{12}$$

Thus, for a given discretization step of the grid, the minimal coherence achievable corresponds to the global dictionary $\mathbf{\Phi}$, whereas the maximal coherence is obtained for the dictionary $\mathbf{\Phi}_1$.

IV. THE KR-SOLS ALGORITHM

In the light of the results of the previous section, one can see that it is computationally inefficient to design a sparse approach for DOA estimation with a multi-scale sensor array, using the same discretization step for all the N levels. Therefore, we propose hereafter, an algorithm that incorporates every scale level iteratively; at iteration step n, the data corresponding to the n^{th} scale level is added to the dataset and the discretization step is decreased, *i.e.,* the grid is further refined. From the acquired data $\mathbf{Z}$, we build N datasets, that can be expressed (using the Matlab notation) as:

$$\begin{aligned} \mathbf{Z}_1 &= \mathbf{Z}(1 : L_1, :) \\ \mathbf{Z}_2 &= \mathbf{Z}(1 : L_1 \cdot L_2, :) \\ &\vdots \\ \mathbf{Z}_N &= \mathbf{Z}(1 : L_1 \cdot L_2 \cdot \cdots \cdot L_N, :) = \mathbf{Z}. \end{aligned} \tag{13}$$

Each data matrix $\mathbf{Z}_n$ $(n = 1, \cdots, N)$ contains the snapshots of the subarrays from levels 1 to n. Accordingly, the sparse estimation problem can be expressed for each dataset $\mathbf{Z}_n$ as:

$$\mathbf{Z}_n = (\mathbf{\Phi}_1^{(n)} \odot \cdots \odot \mathbf{\Phi}_n^{(n)})\mathbf{X}^T, \tag{14}$$

where $\mathbf{\Phi}_j^{(i)}$ represents the dictionary for the j^{th} scale level with a discretization step corresponding to iteration i. In order to efficiently solve the sparse problem in eq. (3), we propose a dictionary adaptation strategy. This strategy is composed, at each iteration, of two steps: a dictionary refinement procedure and a pruning procedure. For the dictionary refinement procedure we start by fixing the minimum grid step δ_{min}, corresponding to the final dictionary $\mathbf{\Phi} = \mathbf{\Phi}_N^{(N)} = \mathbf{\Phi}_1^{(N)} \odot \ldots \odot \mathbf{\Phi}_N^{(N)}$ and the maximum grid step δ_{max}, corresponding to $\mathbf{\Phi}_1^{(1)}$. The value of δ_{min} fixes the maximum resolution achievable by our method. Given δ_{max}, δ_{min} and N, a factor q, that gives the ratio between the grid steps of two adjacent levels can be computed using $\delta_{max} = \delta_{min}\, q^{N-1}$. The dictionary refinement procedure can thus be summarized as follows:

- Fix the maximal resolution, *i.e.,* the minimum grid step δ_{min} for the last dictionary $\mathbf{\Phi}_N^{(N)}$.
- Build initial dictionary $\mathbf{\Phi}_1^{(1)}$ as a sub-dictionary of $\mathbf{\Phi}_1^{(N)}$ with a decimation factor q^{N-1} and solve the problem in (3) with $\mathbf{Z} = \mathbf{Z}_1$ and $\mathbf{\Phi} = \mathbf{\Phi}_1^N$.
- Proceed similarly with the other datasets $(\mathbf{Z}_2, \cdots, \mathbf{Z}_N)$ and the corresponding dictionaries (with the adequate decimation step).

The main drawback of this approach is that the size of the dictionary grows exponentially with the upper index (n). To avoid this phenomenon, we propose the use of a pruning procedure.

Since our approach solves iteratively a sparse estimation problem on a multi-grid, once we have a solution on the coarse grid, we resample the grid to obtain a finer one. However, the second grid is bigger in size than the first one, resulting in a higher computational cost. In order to decrease this computational burden the pruning procedure conserves only a few points on the grid from one iteration to another. The pruning procedure (see *e.g.* [7]), allows to activate at iteration $(n + 1)$ only the atoms in the neighborhood of the atoms estimated at iteration (n); for example: in this paper we consider at iteration $(n + 1)$ the $2q$ atoms in the neighborhood of each one of the P atoms estimated at iteration (n). Fig. 2 illustrates the dictionary refinement and pruning procedure over one dimension, for three scale levels and a factor $q = 3$. To have a better idea of the computational gain in our case, without the pruning procedure, the over-complete dictionary at the last iteration, contains $N_\theta \cdot N_\varphi$ atoms, whereas using the pruning procedure, the number of the atoms drops at $(2q + 1)^2 \times P$. For example, in the case $N_\theta = N_\varphi = 3^6$, $q = 3$, $N = 3$, $P = 2$, the final dictionary contains 98 atoms using the pruning procedure, against 531441 atoms for the full-size dictionary.

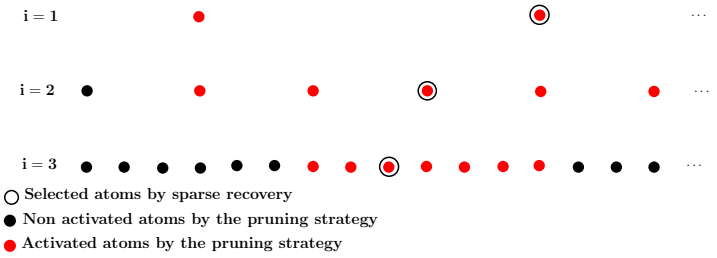

Fig. 2. Pruning procedure with grid refinement mechanism

We summarize hereafter the main steps of the proposed algorithm:

Algorithm 1 Algorithm KR-SOLS

INPUT: $\mathbf{Z}$: Data matrix of size $(L \times K)$; P: Number of sources; N: Number of levels; q: Decimation factor; $\mathbf{\Phi}_1^{(N)}, \ldots, \mathbf{\Phi}_N^{(N)}$: Maximal resolution dictionaries

1: Initialization: Set of active atoms $\Omega = (1 : q^{N-1} : N_\theta) \times (1 : q^{N-1} : N_\varphi)$
2: **for** n = 1: N **do**
3: $\quad \mathbf{\Phi}_j^{(n)} = \mathbf{\Phi}_j(\Omega) \qquad j = 1, \cdots, n$
4: $\quad \mathbf{\Phi} = \mathbf{\Phi}_1^{(n)} \odot \cdots \odot \mathbf{\Phi}_n^{(n)}$
5: $\quad \mathbf{X} = SOLS(\mathbf{\Phi}, \mathbf{Z}_n, P)$
6: $\quad \Omega = ADAPT(\mathbf{X})$;
7: **end for**

OUTPUT: Ω

The *ADAPT* function in **Algorithm 1** implements the refinement and the pruning procedures described previously.

V. SIMULATIONS

In this section, we compare the proposed KR-SOLS algorithm to the CP-based method of [1] and to the Cramér-Rao bound. The simulated sensor array is illustrated in Fig. 1 and consists of 3 hierarchical scale levels: a level-1 array of 21 cross-shaped sensors, half-wavelength spaced, reproduced over a 2×2 square grid at extended spacing $\Delta_1 = 10\lambda$. This level-2 array is reproduced again over a 2×2 square grid, at an extended spacing $\Delta_2 = 10\Delta_1$. Consider two sources impinging on the array from $(u_1 = 0.01, v_1 = 0.707)$ and $(u_2 = 0.5, v_2 = 0.5)$. Two hundred Monte-Carlo runs are used for each point plotted on the figures. Fig. 3 plots the Root Mean Square Error (R.M.S.E.) vs. the number of snapshots; our algorithm performs better than CP, especially for a small number of snapshots. Fig. 4 shows the evolution of the R.M.S.E. with respect to the Signal to Noise Ratio (SNR). KR-SOLS yields better results than CP for low SNR, which makes it well-suited for the detection of low-power signals. Fig. 5 plots the estimation error for the two methods vs. the SNR, for two different final resolutions. One can see that the accuracy of our method is bounded by the resolution of the grid, while the error for the CP method decreases monotonically with the SNR. Future works will focus on how to efficiently adapt the grid resolution to the SNR, to avoid this "saturation" behavior.

VI. CONCLUSIONS

An iterative sparse algorithm for DOA estimation, with a multiple spatial invariance sensor array was introduced. We exploited the particular Khatri-Rao structure of the dictionary of steering vectors to design a computationally efficient DOA estimation method, based on an iterative grid refinement and a pruning strategy. A coherence result for the Khatri-Rao dictionary was also provided. We showed in numerical simulations, that our method yields better results than the method of [1], especially for low SNR and few snapshots.

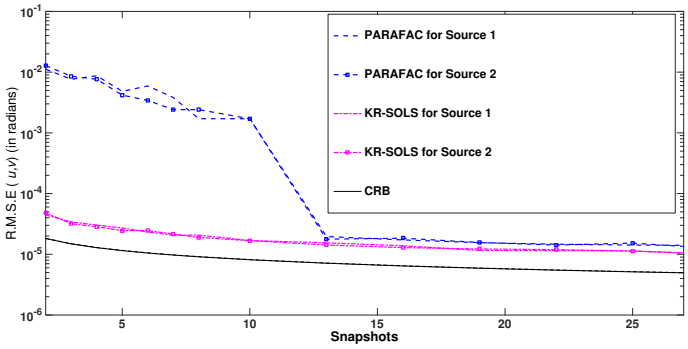

Fig. 3. R.M.S.E. for two non-coherent sources vs. number of snapshot, SNR = 20 dB

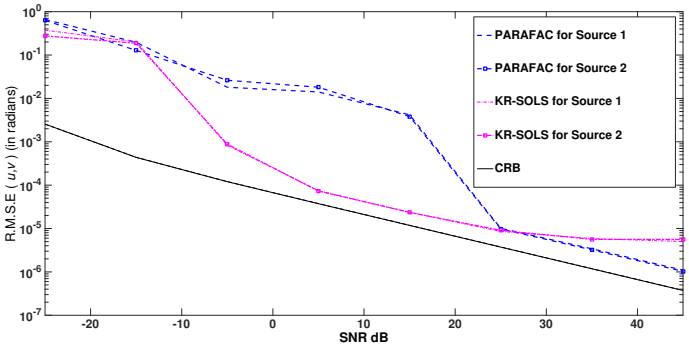

Fig. 4. R.M.S.E. for two non-coherent sources vs. SNR, 50 snapshots

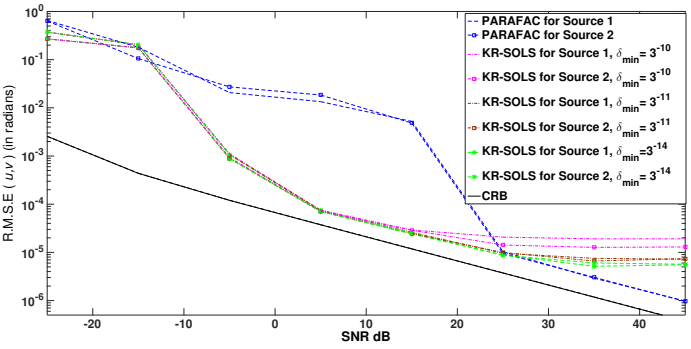

Fig. 5. R.M.S.E. for two non-coherent sources vs. SNR, 50 snapshots, for different grid resolutions

VII. REFERENCES

[1] S. Miron, S. Yang, D. Brie, and K. T. Wong, "A multilinear approach of direction finding using a sensor-array with multiple scales of spatial invariance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 7, pp. 3475–3481, 2015.

[2] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[3] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed L2,0 norm approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4646–4655, 2010.

[4] L. Belmerhnia, E.-H. Djermoune, and D. Brie, "Greedy methods for simultaneous sparse approximation," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, 2014, pp. 1851–1855.

[5] J. D. Blanchard, M. Cermak, and Y. Hanle, D.and Jing, "Greedy algorithms for joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1694–1704, 2014.

[6] S. Jokar and V. Mehrmann, "Sparse solutions to underdetermined Kronecker product systems," *Linear Algebra and its Applications*, vol. 431, no. 12, pp. 2437–2447, 2009.

[7] S. Sahnoun, E.-H. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Adv. Signal Process.*, pp. 1–10, 2012.