Appl. Comput. Harmon. Anal. 41 (2016) 362–383

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Sparse frame DOA estimations via a rank-one correlation model for low SNR and limited snapshots

Cao Zeng [a], Shengqi Zhu [a], Shidong Li [b,a,*], Quisheng Liao [a], Lanmei Wang [a]

[a] National Laboratory of Radar Signal Processing, XiDian University, Xi'An 710071, China
[b] Department of Mathematics, San Francisco State University, San Francisco, CA 94132, United States

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 12 January 2016
Accepted 23 February 2016
Available online 9 March 2016
Communicated by Spec. Issue Guest Editor

Keywords:
DOA
Frames
Dual frames
Optimal dual frames
Sparse dual frames
Sparse representation
Compressed sensing with frames
Redundant dictionary
ℓ_1-analysis
ℓ_1-synthesis

ABSTRACT

As a typical problem of sparse frame representations or compressed sensing with frames, direction of arrival (DOA) estimations, via sparse recovery methodologies such as nonlinear optimizations or various greedy algorithms, suffer seriously under severe noisy measurement without proper treatment. In this article, a crucial and effective correlation operation is outlined to mitigate the severe noise effect from the sparse representation point of view. A fast and super resolution method for DOA estimations is therefore established under very low signal-to-noise ratio, and by a sparse recovery technique of null space tuning in the context of compressed sensing and sparse representations. Improvement to the technique using the thresholding difference and alternating grid refinements are also outlined to estimate the source/target numbers dynamically and to improve the precision. Another characteristic is that only a very small number of snapshots are needed in comparison with other techniques. Simulation results demonstrate apparent advantages of the proposed technique over known approaches. The proposed method has the effectiveness at simultaneous high resolution, robustness to noise, and source number estimations. The algorithm is also computationally efficient, which is critical for large array systems and/or mission critical real-time applications.

1. Introduction

In array signal processing, direction of arrival (DOA) estimations have been one of the main topics. It is widely encountered in application platforms such as radar, communication, navigation and sonar etc. [1]. Subspace-based methods such as multiple signal classifications (MUSIC) [2], effective estimations of signal parameters via rotation invariance techniques (ESPRIT) [3] and iterative adaptive algorithm (IAA) [4] can achieve relatively high resolution with enough snapshots under moderately high signal-to-noise ratio

* Corresponding author. Fax: +415 338 1461.
E-mail addresses: czeng@mail.xidian.edu.cn (C. Zeng), sqzhu@xidian.edu.cn (S. Zhu), shidong@sfsu.edu (S. Li), liaogs@xidian.edu.cn (Q. Liao), lmwang@mail.xidian.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.acha.2016.02.002
1063-5203/

(SNR). These methods also require that the source number is precisely known and the sources are not strongly correlated.

Among new techniques, sparse-representation-based (SR) methods perform relatively well when observations carry high signal-to-noise ratio (SNR). Typical techniques in these new methods include several classes. The foremost class is ℓ_1-norm based convex optimization (CVX) such as [5]. Among others, "greedy algorithms" are another class of popular approaches of finding sparse solutions. Two typical representative approaches are Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP), e.g., [6–8]. In addition, a number of variants of the greedy pursuit algorithms have also been proposed by various authors, e.g., stagewise orthogonal matching pursuit (StOMP) [9], compressive sampling matching pursuit (CoSaMP) [10] and subspace pursuit (SP) [11], etc.

Another class of algorithms for sparse solutions to underdetermined linear inverse problems are iterative thresholding/shrinkage algorithms, which are known for their simplicity. Most iterative thresholding/shrinkage algorithms are motivated by minimizing a cost function, which combines a quadratic error term with a sparsity-promoting regularization term, for instance,

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1.$$

Various iterative hard/soft thresholding algorithms [12–19], gradient-descent methods [20–22], and Bregman iterations [23,24] are representatives. Among this class of works, an algorithm [25] that replaces the ℓ_1-regularization term by $\lambda\|x\|_0$ is proposed by Blumensath and Davies. An iterative hard thresholding (IHT) algorithm within the majorization minorization (MM) framework is analyzed [26]. It was also shown that IHT converges to a local minimum of the ℓ_0-regularized cost function under some conditions.

Recent developments of iterative algorithms include also a hard thresholding pursuit (HTP) algorithm [27], and the latest null space tuning algorithm with feedbacks (NST + HT + FB) [28]. In essence, HTP can be regarded as a hybrid of IHT and CoSaMP. Both HTP and NST + HT + FB converge in finite many steps for strictly sparse vectors.

Applied directly to DOA applications, alternating direction method (ADM) [29] and various variations are yet another class.

However, when directly applied to the linear measurement of an DOA equation, performances of these methods degrade substantially under typically very noisy observations in DOA application platforms.

Dealing with noises in other applications include techniques, for instance, proximity algorithms [30–32]. In the DOA arena, the method of ℓ_1-based singular value decomposition (ℓ_1-SVD) [33] can improve the performance of SR-based methods in low SNR circumstances. The cost, however, is its rather time-consuming computational complexities, and the rather large number of snapshots/sampling times needed.

Thanks to a reviewer, the article [34] came to our close attention during the revision. It is about the analysis of the (best) mathematical resolution of problems (applicable to DOA and others) by ℓ_1 minimization procedures in general.

In this article, a fast and super-resolution DOA estimation approach based on a crucial correlation operation and sparse frame representations is proposed. Typical DOA formulations by sparse frame representations or compressed sensing with frames suffer seriously under severe noisy measurement which is common in the platform of DOA applications. To mitigate the low SNR issue in a typical DOA application, a crucial and effective rank-one correlation operation is proposed from the sparse frame representation point of view. The proposed DOA recovery technique is based on the new iterative null space tuning method with feedbacks (NST + HT + FB) [28], a method of the ℓ_0 spirit [35]. An improvement to the adaptive scheme of the NST + FB is also proposed to estimate the source number dynamically, by a technique of thresholding differences. Another improvement to enhance the estimation accuracy by a procedure of alternating grid optimization is also part of the integrated work.

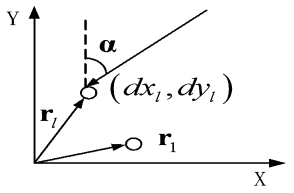


Fig. 1. Planar array.

Simulation results demonstrate apparent advantages of the proposed technique over known approaches. The proposed method has the characteristics of simultaneous high resolution, robustness to noise, and effectiveness at estimating the number of sources. Another notable contribution of this approach is that the algorithm requires very few number of snapshots. This, in addition to computational efficiency of the algorithm, makes this approach feasible in real-time and mission critical applications.

2. Narrow band array signal model

In typical planar wave treatment and approximation, a wave function $\mathbf{s}(t)$ propagates at the space position vector $\mathbf{r}$ along the direction of $\boldsymbol{\alpha}$ in the form of $\mathbf{s}(\mathbf{r}, t) = \mathbf{s}(t - \boldsymbol{\alpha}^T\mathbf{r})$.

Narrow band signals $\mathbf{s}(t)$ have generally the form of $\mathbf{s}(t) = g(t)e^{-j\omega_0 t}$. Its propagation, following the planar wave treatment, is consequently modeled by

$$\mathbf{s}(\mathbf{r}, t) \approx g(t)e^{j\omega_0 t}e^{-j\omega_0\boldsymbol{\alpha}^T\mathbf{r}} = \mathbf{s}(\mathbf{0}, t)e^{-j\omega_0\boldsymbol{\alpha}^T\mathbf{r}} = \mathbf{s}(t)e^{-j\omega_0\boldsymbol{\alpha}^T\mathbf{r}}, \tag{2.1}$$

where $\mathbf{s}(t) \equiv \mathbf{s}(\mathbf{0}, t)$ is the wave function at the space position $\mathbf{r} = \mathbf{0}$ (the origin).

Array formulation Consider now a general (curved) array of N sensors in the xy-plane. Let the position vector of the lth sensor be denoted by, as seen in *Fig. 1*,

$$\mathbf{r}_l = (dx_l, dy_l), \quad l = 1, 2, \cdots, N \tag{2.2}$$

The propagation direction vector α is therefore

$$\boldsymbol{\alpha} = -\frac{1}{c}(\sin\theta, \cos\theta)^T \tag{2.3}$$

As a result, a narrow band signal received by the lth sensor can be written as

$$x_l(t) = \mathbf{s}(\mathbf{r}_l, t) = \mathbf{s}(t - \boldsymbol{\alpha}^T\mathbf{r}_l)e^{j\omega_0(t-\boldsymbol{\alpha}^T\mathbf{r}_l)} = \mathbf{s}(t)e^{-j\omega_0\boldsymbol{\alpha}^T\mathbf{r}_l}, \tag{2.4}$$

where

$$\omega_0\boldsymbol{\alpha}^T\mathbf{r}_l = -\frac{\omega_0}{c}(dx_l\sin\theta + dy_l\cos\theta) = -\frac{2\pi}{\lambda}(dx_l\sin\theta + dy_l\cos\theta)$$

Equation (2.4) has then a form

$$x_l(t) = \mathbf{s}(t)e^{j\frac{2\pi}{\lambda}(dx_l\sin\theta + dy_l\cos\theta)}, \quad l = 1, 2, \cdots, N \tag{2.5}$$

Hence, the observation vector of the general array at time t is given by

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \begin{bmatrix} e^{j\frac{2\pi}{\lambda}(dx_1 \sin\theta + dy_1 \cos\theta)} \\ e^{j\frac{2\pi}{\lambda}(dx_2 \sin\theta + dy_2 \cos\theta)} \\ \vdots \\ e^{j\frac{2\pi}{\lambda}(dx_N \sin\theta + dy_N \cos\theta)} \end{bmatrix} \mathbf{s}(t) = \mathbf{a}(\theta)\mathbf{s}(t) \tag{2.6}$$

where the vector $\mathbf{a}(\theta)$ of the exponentials is termed the *steering vector* of the array. It is a function of sensor positions $\{dx_l, dy_l\}$, DOA θ, and the wavelength λ.

Linear array For a typical linear array (LA) placed on the x-axis, the steering vector is simplified as (with $dx_n = d_n$, and $dy_n = 0$),

$$\mathbf{a}(\theta) = [e^{j2\pi\frac{d_1}{\lambda}\sin(\theta)}, e^{j2\pi\frac{d_2}{\lambda}\sin(\theta)}, \cdots, e^{j2\pi\frac{d_N}{\lambda}\sin(\theta)}]^T \tag{2.7}$$

Regular array signal model So now, consider P signals with directions of arrival θ_p $(p = 1, 2, \cdots, P)$, impinging on a linear array of N sensors. After demodulation, the data vector received by the array at time t can be expressed as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{2.8}$$

where, $\mathbf{s}(t)$ is a vector consisting of multiple narrowband signals at time instance t, $\mathbf{n}(t)$ is the Gaussian noise vector, and $\mathbf{A}$ is the array manifold matrix, consisting of columns of steering vectors at different signal directions $\{\theta_p\}$,

$$\mathbf{a}(\theta_p) = [e^{j2\pi\frac{d_1}{\lambda}\sin(\theta_p)}, e^{j2\pi\frac{d_2}{\lambda}\sin(\theta_p)}, \cdots, e^{j2\pi\frac{d_N}{\lambda}\sin(\theta_p)}]^H \tag{2.9}$$

where H represents the Hermitian transpose of a matrix, d_i $(i = 1, 2, \cdots, N)$ is the element spacing of the ith sensor, and λ denotes the carrier wavelength.

3. Rank-one correlated linear array signal model for noise suppression

Evidently, for each time instance, (2.8) is a typical sparse frame representation where the matrix $\mathbf{A}$ as in (2.9) consists of columns of a frame of the same steering vector (see *Section 5*). The cruel reality is that in DOA applications, SNR is often very low. Direct sparse signal recovery from (2.8) often fails badly under low SNR.

Traditional techniques like MUSIC [36–38] uses techniques such as correlation and eigen-space analysis for a robust and relatively good results. The effectiveness of these techniques depends on the success of the correlation and eigen-space analysis. To this end, a large number of sampling instances/snapshots (over a set of time instances $\{t_n\}_n$) are required. The large number of snapshots can be available in some applications, and difficult to obtain in some other applications [39].

One other issue with these techniques such as MUSIC is that the source number must be assumed known. As such, a pre-estimation of the source number must be carried out first.

On the other hand, sparse recovery model and techniques can recover the source numbers and source locations simultaneously. It has also the super resolution characteristic. However, the low SNR issue in a typical linear observation model is an unforgiven problem.

We introduce a rank-one correlated operation to cope with the noise in the linear observation model. By averaging the correlation between one of the element measurement and the vector consisting of the rest of $(N-1)$ sensor measurements in M snapshots, a new linear observation model particularly robust to noises and unique to the DOA estimation in array signal processing is then obtained.

Proposition 3.1. *Consider an linear array of N elements. Let $\{s_p(t)\}_{p=1}^P$ be a set of source/target signals. Assume that source signal vector $\mathbf{s} = \{s_p\}$ and the noise vector $\mathbf{n}$ are statistically uncorrelated, and that the noise has a mean $E_n(\mathbf{n}) = 0$. Denote by $\mathbf{x}(2:N,t)$ the measurement vector from the 2nd to the Nth sensor of the elements, and by ${x_1}^*$ the complex conjugate of the measurement from the 1st sensor. Define*

$$\mathbf{y} = E_n\left\{\mathbf{x}(2:N,t){x_1}^*(t)\right\},$$

where E_n is the probabilistic expectation with respect to the noise ensemble. Then $\mathbf{y}$ has a (SNR enhanced) linear observation model

$$\mathbf{y} = \Phi\mathbf{r} + \sigma, \tag{3.1}$$

where $\Phi = A(2:N,:)$ is the array manifold matrix A sampled from the second row to the Nth row, and $\mathbf{r} = [r_1, r_2, \cdots, r_P]^H$ and $\sigma = [\sigma_1, \sigma_2, \cdots, \sigma_N]^H$ are the (new) source and noise vectors, respectively, where

$$r_p = \sum_{q=1}^{P} (s_p s_q^*) \cdot e^{-j2\pi \frac{d_1}{\lambda} \sin\theta_q}, \qquad p = 1,\ 2, \cdots, P$$

and

$$\sigma_n = E_n\{\mathbf{n}(2:N)n_1^*\}. \qquad n = 1,\ 2, \cdots, N.$$

Proof. In consideration of the structure of the matrix A as given in (2.9). Direct computations show

$$\begin{aligned}
\mathbf{y} &= E_n\left\{\mathbf{x}(2:N,\cdot){x_1}^*(\cdot)\right\} \\
&= E_n\left\{\left(A(2:N,:)\mathbf{s} + \mathbf{n}(2:N)\right)\left(\mathbf{s}^*A^*(1,:) + n_1^*\right)\right\} \\
&= \Phi\left(\mathbf{s}\,\mathbf{s}^*\right)A^*(1,:) + \mathbf{s}^*A^*(1,:)E_n\left\{\mathbf{n}(2:N)\right\} + \\
&\quad A(2:N,:)\mathbf{s}E_n\left\{n_1^*\right\} + E_n\left\{\mathbf{n}(2:N)\,n_1^*\right\} \\
&= \Phi\left(\mathbf{s}\,\mathbf{s}^*\right)A^*(1,:) + E_n\left\{\mathbf{n}(2:N)\,n_1^*\right\} \\
&= \Phi\mathbf{r} + \sigma,
\end{aligned}$$

where $\mathbf{r} \equiv \mathbf{s}\,\mathbf{s}^*A^*(1,:)$ and $\sigma \equiv E_n\left\{\mathbf{n}(2:N)\,n_1^*\right\}$. The result then follows. □

Remarks. *a*. It is quite unique that the new measurement $\mathbf{y}$ as in (3.1) for the DOA formulation is also a linear model. The only difference is that the (new) signal variable $\mathbf{r}$ is of quite different form than that of the original signal, and it is generally complex even if the inpinging signal is real.

More importantly, we will show that this linear model is equivalent to the original signal model in at least the DOA applications. See *Section 3.1*.

b. Denoising capacity within a linear model: The substantial meaning of this rank-one correlation model of the DOA formulation lies in its denoising capacity. If, following the traditional assumption that noise measurements at different array elements are statistically independent, the noise component σ is clearly 0, which perfectly serves the purpose of denoising while staying as a linear model.

Note, however, array elements are physically closely placed. The assumption that noisy contaminations received among array elements are statistically independent can be invalid. It is certainly true that there are applications where noise or error in measurements cannot be assumed statistically independent. The rank-one correlation model will be appropriate to handle noise treatment when no independency is assumed. This can be another useful feature.

Numerically, the expectation E_n may only be substituted by a stationary assumption and a time-averaging approximation. So, instead of the expectation, the rank-one correlated observation is defined by the following time-averaging:

$$\mathbf{y} = \tfrac{1}{M}\sum_{i=1}^{M} \mathbf{x}(2:N,t_i)\mathbf{x}^*(t_i), \tag{3.2}$$

where M is known as the number of *snapshots* or the number of samples taken at the time instances $\{t_i\}_{i=1}^{M}$.

Completing the computation of the time-averaging, the rank-one correlation linear observation model emerges as well.

Corollary 3.2. *Let* $\mathbf{y}$ *be the time-average defined in (3.2). Then the rank-one correlation observation* $\mathbf{y}$ *has the following linear model, without changing the variable for the (new) signal source vector* $\mathbf{r}$ *and the noise vector* σ*,*

$$\mathbf{y} = \Phi\mathbf{r} + \sigma, \tag{3.3}$$

where

$$\mathbf{r} = \tfrac{1}{M}\left(\sum_{i=1}^{M} \mathbf{s}(t_i)\mathbf{s}^*(t_i)\right) A^*(1,:),$$

and

$$\sigma = A(2:N,:)\mathbf{S_n} + \mathbf{N_s}A^*(1,:) + \tfrac{1}{M}\sum_{i=1}^{M} \mathbf{n}(2:N,t_i)n_1^*(t_i),$$

and $\mathbf{s_n}$ *and* $\mathbf{N_s}$ *are given by*

$$\mathbf{s_n} = \tfrac{1}{M}\sum_{i=1}^{M} \mathbf{s}(t_i)n_1^*(t_i),$$

and

$$\mathbf{N_s} = \tfrac{1}{M}\sum_{i=1}^{M} \mathbf{n}(2:N,t_i)\mathbf{s}^*(t_i).$$

Remark. Note that $\mathbf{s}$ and $\mathbf{n}$ are typically assumed uncorrelated or statistically independent. As a result, with the assumption that the mean of noise is zero, the time-average values of $\mathbf{s_n}$ and $\mathbf{N_s}$ are quite small. Certainly, the last term $\frac{1}{M}\sum_{i=1}^{M} \mathbf{n}(2:N,t_i)n_1^*(t_i)$ gets suppressed by the time-averaging, with or without independent assumption to the noise components, as long as the noise is assumed to have zero mean.

Discussion 3.3. We comment that in the rank-one correlation model, much smaller number of snapshots M are needed for the purpose of signal recovery and noise suppression.

Traditionally, a full rank correlation is calculated such as the MUSIC algorithm for the purpose of denoising and a reliable estimation of the eigenvalues and eigenvectors of the signal and the noise subspaces.

While the denoising mechanisms are similar (with these traditional techniques), the needs for the number of snapshots M are very different between our rank-one correlation linear model, and the full-rank correlation in MUSIC in their fundamental motives and subsequent algorithms.

In MUSIC and other related approaches, full-rank correlation is required because of the essential need for the eigen-space analysis. Denoising is merely a consequence of the correlation operation.

In the rank-one correlation model, eigen-space analysis is not in the picture, denoising is the full purpose. Consequently, the requirement for the large snapshots M in our rank-one correlation model is not as important as in MUSIC and related algorithms.

As a result, solely for denoising purposes, it is seen that the number of snapshots can be substantially smaller than that required by MUSIC algorithms. As will be presented later, for similar SNR levels, our rank-one model can function extremely well under very small number of snapshots M, while MUSIC and related algorithms will fail at the same small M value.

3.1. The validation of the rank-one correlation model

The next step is an assurance that the rank-one correlation model is truly equivalent to the regular linear observation model as in (2.8). Specifically, for the purpose of direction of arrival estimation, we need to have an assurance that the pth component of the source vector $\mathbf{r}$, $r_p \neq 0$ whenever $s_p \neq 0$. Needless to say, of course, whenever $s_p = 0$, $r_p = 0$, implying that $\mathbf{r}$ will not introduce artificial fake signals.

To quantify the justification of such a validation, let us simplify the notation by writing

$$\rho_{pq} = \tfrac{1}{M} \sum_{i=1}^{M} s_p(t_i) s_q^*(t_i),$$

and

$$\begin{aligned} r_p &= \sum_{q=1}^{P} \rho_{pq} e^{-j2\pi \frac{d_1}{\lambda} \sin\theta_q} \\ &= \rho_{pp} e^{-j2\pi \frac{d_1}{\lambda} \sin\theta_p} + \sum_{q=1, q\neq p}^{P} \rho_{pq} e^{-j2\pi \frac{d_1}{\lambda} \sin\theta_q}. \end{aligned} \tag{3.4}$$

The following is clear from the equation (3.4), but needed to understand the essence why $s_p \neq 0$ implies that $r_p \neq 0$.

Lemma 3.4. *Suppose that the source signals are statistically uncorrelated. Then* $r_p = \rho_{pp} e^{-j2\pi \frac{d_1}{\lambda} \sin\theta_p}$.

It is true that, when strict statistical correlation is not available in actual computation, the time-average in the estimation of ρ_{pq} for $p \neq q$ may not be exactly zero. But ρ_{pq}, for $p \neq q$, is indeed zero under the theoretical un-correlation assumption.

Lemma 3.5. *For correlated source signals, consider an extreme case where* $s_q(t) = s_p(t - \tau_q)$ *and* $s_p(t) = e^{2\pi j f_0 t}$. *Suppose that the delay parameters* $\{\tau_q f_0\}_{q=1}^{P}$ *(assuming* $\tau_p = 0$*) does not form a root of unity. Assume also that* $d_1 = 0$ *by setting the corresponding element as the reference. Then* $r_p \equiv \sum_i \rho_{pq} \neq 0$.

Proof. Setting the reference element with $d_1 = 0$ is common. It will allow us to see the essence of completely correlated cases.

With the plane wave assumption, $s_p(t) = e^{j2\pi f_0 t}$ is generally what the signals are, except for possibly time windows applied to the exponential. As a result, $s_q(t) = e^{-j2\pi f_0 \tau_q} s_p(t)$ in such an extreme correlation case. Then $\rho_{pq} = e^{j2\pi f_0 \tau_q} \rho_{pp}$, and

$$r_p = \rho_{pp}\left(1 + \sum_{q,q\neq p} e^{j2\pi f_0 \tau_q}\right).$$

Consequently, if $\{\tau_q f_0\}_{q=1}^P$ does not form a root of unity, $r_p \neq 0$. □

We note that in practice, $\{\tau_q f_0\}_{q=1}^P$ can hardly ever form a root of unity on the unit circle, because source signals can hardly ever be exactly shifted by a uniformly placed delay parameters, however correlated they are. It is rather practical to say that with probability almost surely 1 that the delay parameters $\{\tau_q f_0\}_{q=1}^P$ are not exactly evenly placed on the interval of $[-\frac{1}{f_0}, \frac{1}{f_0}]$. Consequently, the assurance that $r_p \equiv \sum_q \rho_{pq} \neq 0$ in the extremely coherence case as stated in *Lemma 3.5* is justified.

In a general case where $d_1 \neq 0$, we may also see that $r_p \neq 0$ with an probabilistic argument.

Theorem 3.6. *Let y be the rank-one correlated model as given in (3.3). Let r_p be given in (3.4). Suppose that the source signal angles $\{\theta_p\}$ follows a uniform distribution in a range $[-\alpha, \alpha]$, $\alpha < \pi/2$ (typically $\alpha \leq 70°$). Suppose $s_p \neq 0$ or $\rho_{pp} \neq 0$, and as demonstrated in Lemma 3.5, that the source correlation $\sum_q \rho_{pq} \neq 0$. Then for all α values such that $2\pi\frac{d_1}{\lambda}\sin\alpha < \pi$, the expected value of r_p, $E_\theta\{r_p\} \neq 0$.*

Proof. Let the probability distribution of θ_q be given by

$$P_{\theta_q}(\theta) = \tfrac{1}{2\alpha}\chi_{[-\alpha,\alpha]},$$

where $\chi_{[-\alpha,\alpha]}$ is the characteristic function on $[-\alpha, \alpha]$. Then

$$r_p = \sum_{q=1}^{P} \rho_{pq}\left(\cos(2\pi\tfrac{d_1}{\lambda}\sin\theta_q) - j\sin(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)\right).$$

And,

$$E_\theta(r_p) = \sum_{q=1}^{P} \rho_{pq} E_\theta\left(\cos(2\pi\tfrac{d_1}{\lambda}\sin\theta_q\right) - j\sum_{q=1}^{P} \rho_{pq} E_\theta\left(\sin(2\pi\tfrac{d_1}{\lambda}\sin\theta_q\right).$$

$$E_\theta\left(\sin(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)\right) = \tfrac{1}{2\alpha}\int_{-\alpha}^{\alpha} \sin(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)d\theta_q = 0$$

by the symmetry of the range of the θ_q distribution, and by the fact that $\sin(2\pi\frac{d_1}{\lambda}\sin\theta_q)$ is an odd function. Meantime,

$$E_\theta\left(\cos(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)\right) = \tfrac{1}{2\alpha}\int_{-\alpha}^{\alpha} \cos(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)d\theta_q$$

$$= \tfrac{1}{\alpha}\int_{0}^{\alpha} \cos(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)d\theta_q.$$

Consequently, as long as $2\pi\frac{d_1}{\lambda}\sin\alpha < \pi$, the integral will be greater than 0 by the area interpretation, and hence, $E_\theta(r_p) > 0$. □

Remarks. *a*. The θ_q range is typically strictly less than $\pi/2$ by the physics of the array. A typical linear observation range $[-\alpha, \alpha]$ may not exceed $[-70^\circ, 70^\circ]$ by its physical property. It is also true that typically $d_1/\lambda = 1/2$. As a result, the parameter $2\pi\frac{d_1}{\lambda}\sin\alpha$ is indeed quite smaller than π. Consequently, $E_\theta(r_p) > 0$ is very typical.

b. One may also calculate the probability of $|r_p| \geq 0$. Since

$$|r_p|^2 = \sum_q \rho_{pq}^2 \left(\cos^2(2\pi\tfrac{d_1}{\lambda}\sin\theta_q) + \sin^2(2\pi\tfrac{d_1}{\lambda}\sin\theta_q)\right),$$

one may compute the probability distribution of $\sin\theta_q$ and $\cos^2(2\pi\frac{d_1}{\lambda}\sin\theta_q)$, and $\sin^2(2\pi\frac{d_1}{\lambda}\sin\theta_q)$. Then, $P\left(|r_p|^2 \geq 0\right)$ can be computed numerically (as strictly analytical answer is not possible), and the value is almost surely 1.

4. Array physics and model discretization

In the area of array signal processing, array aperture (size of the array) determines the array physics, and further determines the resolution of the radar system. In general, the synthesized array observation (a passive array and/or receiving only array) has a range of effective observation angles.

Measured by degrees, such an observation range is often termed as *beamwidth*. Such a beam is the magnetic field that array emits (or could observe) in space.

Array beamwidth is typically defined to be the angle range of the array beam measured between 50% of the peak energy levels. In dB values, from the peak energy point to the 50% energy level, the dB change is $10\log\frac{1}{2} \approx 3$ dB. As a result, array beamwidth is typically termed *3 dB beamwidth*.

Array's 3 dB-beamwidth is inversely proportional to the array aperture size. For a linear array of an aperture length D-times of the carrier wavelength, the 3 dB beamwidth is approximately $51^0/D$ degree. For instance, if an array has an aperture of $D = 50\lambda$, the 3 dB array beamwidth is slightly great than 1^0.

On the one hand, $\theta_p \in \mathbb{R}$ is unknown and it is supposedly real valued. Whenever possible, a model and/or algorithm that can handle the real-valued estimation of θ_p is apparently of immense value.

On the other hand, any algorithm needs some reasonable discretization for computations. We shall also employee a natural discretization to derive a natural sparse frame representation, and to follow up with a grid refinement technique to improve the DOA estimation precision. It is termed alternating grid optimization (AGO) *Section 8*. The technique of AGO is a step forward mitigating the non-real-valued solutions to the DOA problems. The end results are much finer in resolution than just the model discretization grid.

As part of the comments in array physics, it is worth mentioning that in case of concerns about the loss of the track of targets due to discretization (when targets are not exactly on the (refined) grids), a practical answer is that mission critical target tracking, such as weapon firing systems, uses a (or multiple) different target locking mechanism(s). For instance, techniques that utilize the nulling of the radar beams to trap the target in are commonly used. In such and relevant cases, DOA estimations often serve as the initial source/target localization, and then followed by other "locking" strategies.

5. Sparse frame representations of DOA

Let us consider a natural discretization of the model in (3.4) with a much refined angle grid $\{\theta_i\}_{i=1}^N$ with an assumption that $\{\theta_p\}_{p=1}^P \subset \{\theta_i\}_{i=1}^N$. Here $N >> P$ typically.

Define now a refined matrix $\tilde{\Phi}$ by the array manifold matrix Φ at the refined angle grid $\{\theta_i\}_{i=1}^N$. Namely,

$$\tilde{\Phi} = \begin{bmatrix} e^{j2\pi\frac{d_2}{\lambda}\sin\theta_1} & e^{j2\pi\frac{d_2}{\lambda}\sin\theta_2} & \cdots & e^{j2\pi\frac{d_2}{\lambda}\sin\theta_N} \\ e^{j2\pi\frac{d_3}{\lambda}\sin\theta_1} & e^{j2\pi\frac{d_3}{\lambda}\sin\theta_2} & \cdots & e^{j2\pi\frac{d_3}{\lambda}\sin\theta_N} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi\frac{d_N}{\lambda}\sin\theta_1} & e^{j2\pi\frac{d_N}{\lambda}\sin\theta_2} & \cdots & e^{j2\pi\frac{d_N}{\lambda}\sin\theta_N} \end{bmatrix}$$

Then equation (3.3) can also be casted as follows:

$$\mathbf{y} = \tilde{\Phi}\tilde{\mathbf{r}} + \sigma$$

where $\tilde{\Phi}$ is an over-complete dictionary matrix in terms of all possible source angles under a natural discretization as shown above; The signal field is represented by the $\tilde{\mathbf{r}}$ vector, where the ith element is r_p if the ith source comes from the angle θ_p and zero otherwise, i.e., for $i = 1, \cdots, N$,

$$\tilde{r}_i = \begin{cases} r_p, & \text{if } \theta_i = \theta_p, \quad p = 1, \cdots, P \\ 0, & \text{otherwise.} \end{cases}$$

Note that as long as in equation (3.3) $r_p \neq 0$, the pth source can be recovered regardless if the sources are coherent or not.

6. The null space tuning algorithms and the DOA estimations

Besides the rank-one correlation model that is critical in extremely low SNR applications, another characteristic of this article is the application and modification of an null space tuning algorithm (NST) to the solution of the DOA problem.

If a general underdetermined system is given by, with n being the noise,

$$y = Ax + n,$$

the NST algorithm is an iterative algorithm with a thresholding (to estimate the supporting index T_k of the solution x at the step k), and a feedback of x on the complement index T_k^c to the index set T_k for "tail" contribution to y. Specifically, the NST + HT + FB (null space tuning + hard thresholding + feedback) algorithm can be described as

$$\begin{cases} u^k_{T_k} = x^k_{T_k} + (A^*_{T_k}A_{T_k})^{-1}A^*_{T_k}A_{T^c_k}x^k_{T^c_k}, \\ u^k_{T^c_k} = 0, \\ x^{k+1} = x^k + \mathbb{P}(u^k - x^k) \end{cases}$$

Here T_k is the index set obtained by a thresholding operation identifying a specified/estimated number of non-zeros in x, and θ_{T_k} in general stands for the components of the vector (or columns of the matrix) θ indexed by T_k. $\mathbb{P}$ is an orthogonal projection onto the kernel of A. The term of *null space tuning* is a reflection of the fact that the updating at each iteration is tuned within the null space of A. Consequently, as long as x^0 is feasible, all x^k's are feasible.

The feedback of the "tail" contribution to y is the second term of the first equation updating $u^k_{T_k}$. Instead of only keeping the x^k components in the support of T_k, the NST + FB algorithm seeks for the contribution of the "tail" of x^k (namely $x^k_{T^c_k}$) by feeding them back to the support T_k, and thus becomes more feasible

than simply thresholding. It is this very feedback adjustment that greatly improves the rate of convergence. In fact, the feedback mechanism is shown in [28] to converge in finitely many steps.

The underpinning principle can be quite clearly seen from the following simple example. One can see that for $x \in \mathbb{R}^2$, the NST + HT + FB algorithm stops in just 1 iteration.

In $\mathbb{R}^2$, an underdetermined system is given by $a_1x_1 + a_2x_2 = b$. Assume $a_1 > a_2 > 0$. Then an exact sparse solution is $x = [0\ \frac{b}{a_2}]^T$. The initial solution x^0 is typically taken as the least square solution $x^0 = [\frac{a_1 b}{a_1^2+a_2^2}\ \frac{a_2 b}{a_1^2+a_2^2}]^T$.

The feedback operation is to find an approximation to η in the equation $a_2\eta = a_1x_1^0$, i.e., to feed the contribution of $a_1x_1^0$ to b back to the second component of the solution vector. Evidently, in $\mathbb{R}^2$ cases, it has an exact solution $\eta^0 = a_2^{-1}a_1x_1^0$ as long as $a_2 \neq 0$. This feedback in turn gives rise to $u^0 = [0\ \frac{b}{a_2}]^T$, which is the exact solution, and the iteration stops in 1 step. By the same principle, the NST + HT + FB algorithm generally converges in finitely many steps.

What is also substantial in the feedback mechanism is a greatly speeded algorithm for large size and/or very coherent systems, by replacing the feedback inverse computation $(A_{T_k}^* A_{T_k})^{-1}$ by an estimated eigenvalue inverse.

This resulted in a suboptimal feedback NST scheme, termed NST + HT + subFB. The NST + hard-thresholding + suboptimal feedback algorithm can be summarized by the following equations.

$$\begin{cases} u_{T_k}^k = x_{T_k}^k + \lambda^k A_{T_k}^* A_{T_k^c} x_{T_k^c}^k, \\ u_{T_k^c}^k = 0, \\ x^{k+1} = x^k + \mathbb{P}(u^k - x^k). \end{cases}$$

Here the choice for λ can be, for instance, $\lambda^k \leq 1/\|A_{T_k}^* A_{T_k}\|_2$.

Remark 6.1. The reasons that we choose to use the NST + FB algorithms in DOA estimations is based on the following features of the algorithms [28].

- The thresholding nature resembles the minimum ℓ_0 "norm" solution to an underdetermined system, which is capable to resolve sparse recovery problems better than ℓ_1 minimization procedures.
- The thresholding + feedback mechanism allows for sparse signal recovery for much larger number of non-zero components than all known methods (see [28]). This feature, for fixed sparsity s, can also be seen as less stringent restriction/requirement on the matrix A among different algorithms.
- The algorithm has the finite-step convergence guarantee. It is therefore exceedingly fast, suitable for real-time system implementations and for mission critical applications.
- The suboptimal feedback mechanism is not only a viable scheme for very large systems. It can also be very useful when the matrix A (or $\tilde{\Phi}$) is very coherent among columns. Replacing the computation of the pseudo-inverse of a submatrix by a constant (the inverse of an estimated eigenvalue) can greatly improve the stability of the procedure.

Let us apply the NST + FB algorithms to the DOA problem. We comment that the matrix $\tilde{\Phi}$ is generally a non-linearly sampled Fourier frame (because of the $\sin\theta$ in the expression). It may not easily possess a RIP constant δ_s for large s.

It is worth of mentioning, however, that in DOA problems the number of source signals are typically small for a given number of snapshots. This is because the source signals correspond to just the observations within one antenna beam.

As it is mentioned earlier, depending on the antenna aperture size D (times of the wavelength), the effective 3 dB-antenna beamwidth is only about $51/D$ degree. Within such a relatively narrow angle range, usually the source target number is very limited.

Consequently, even thought the matrix $\tilde{\Phi}$ is not random, but the number of non-zeros s is extremely limited (to often times only about 2 or 3). The NST algorithm is still capable to produce satisfactory results.

As a result, with the extremely small s value, the following theorem provides a convergence guarantee. If needed, please refer to [28] for details about the definitions of RIP (restricted isometry property) and P-RIP (pre-restricted isometry property).

Theorem 6.2. *(See [28].) Let* $\mathbf{y} = \tilde{\Phi}\tilde{\mathbf{r}} + \sigma$ *be the rank-one correlation model of the array. Let* $\tilde{\mathbf{r}}^\sharp$ *be the best* s *term approximation of the true solution* $\tilde{\mathbf{r}}$*. Suppose that the matrix* $\tilde{\Phi}$ *has the RIP and P-RIP constants* σ_{2s} *and* γ_{3s} *satisfying* $\sigma_{2s} + \sqrt{2}\gamma_{3s} < 1$*, then applying the NST + HT + FB algorithm, the DOA solutions satisfy the following*

$$\|\tilde{\mathbf{r}}^k - \tilde{\mathbf{r}}^\sharp\|_2 \le \rho^k \|\tilde{\mathbf{r}}^0 - \tilde{\mathbf{r}}^\sharp\|_2 + \tfrac{\tau}{1-\rho}\|\mathbf{e}\|_2,$$

where $\rho = \frac{\sqrt{2}\gamma_{3s}}{1-\sigma_{2s}}$, $\tau = \frac{\sqrt{2}+\sqrt{1+\sigma_s}}{1-\sigma_{2s}}$, *and* $\mathbf{e} \equiv \tilde{\Phi}(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}^\sharp + \sigma)$.

As in all sparse representation or compressed sensing problems, the sufficient condition given here is perhaps far away from being optimal. But it nevertheless serve as a rational explanation why numerical examples show that the NST algorithm applied to DOA problem is exceedingly effective.

7. Dynamic source number estimations through thresholding differences

In this section, we introduce the dynamic source number estimation through thresholding differences based on the NST + HT + subFB algorithm. The estimation of the source number is integrated in the detection of DOA in the NST + FB algorithm. The iteration is over the source number one-by-one, while estimating the DOA's. The pseudocode for the proposed method is given as follows:

Algorithm 1: Dynamic source number estimations through thresholding differences.

Input: $\tilde{\mathbf{\Phi}}$, $\mathbf{y}$, $\tilde{\theta}$, ϵ_1, ϵ_2;
Output: $\tilde{\mathbf{r}}$, $\hat{P}$, $\hat{\theta}$;
$s = 0$; $l = k = 0$; $H_1 = 1$;
$\epsilon_1 = 0.05$; $\epsilon_2 = 1e-3$;
$\tilde{\mathbf{r}}^0 = \tilde{\mathbf{\Phi}}^{\mathrm{H}}(\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^{\mathrm{H}})^{-1}\mathbf{y}$;
while $H_1 > \epsilon_1$ **do**
{ $H_2 = 1$; $s = s + 1$;
  while $H_2 > \epsilon_2$ **do**
  {
  $[\tilde{\mathbf{r}}^{k+1}, \tilde{\mathbf{u}}^k] = \mathrm{NST{+}HT{+}subFB}(\tilde{\mathbf{\Phi}}, \mathbf{y}, s)$
  $H_2 = \|\tilde{\mathbf{r}}^{k+1} - \tilde{\mathbf{r}}^k\|_2 / \|\tilde{\mathbf{r}}^k\|_2$;
  $k = k + 1$;
  }
$\gamma^{l+1} = \|\tilde{\mathbf{\Phi}}\tilde{\mathbf{u}}^k - \mathbf{y}\|_2 / \|\mathbf{y}\|_2$;
$H_1 = |\gamma^{l+1} - \gamma^l|$;
$l = l + 1$;
}
$[\mathbf{p}_V, \mathbf{p}_I] = \mathrm{findpeaks}(|\tilde{\mathbf{r}}^{k+1}|, \text{`}descend\text{'})$
return $\tilde{\mathbf{r}} = \tilde{\mathbf{r}}^{k+1}, \hat{P} = s, \hat{\theta} = \tilde{\theta}(\mathbf{p}_I(1:\hat{P}))$;

Here, $\tilde{\theta}$ is the DOA grid sequence. s is the dynamic source number; l and k are the outer and inner loop variables, respectively; H_1 and H_2 are the outer and inner loop thresholdings, respectively; The convergence parameters of outer and inner loop are denoted by ϵ_1 and ϵ_2, respectively. $\tilde{\mathbf{r}}^{\mathbf{k}}$ is the spatial recovery vector in the kth iteration; We have chosen $\lambda^k = 0.95/\|\tilde{\mathbf{\Phi}}_{T_k}^H \tilde{\mathbf{\Phi}}_{T_k}\|_2$ for stability consideration of the algorithm. $\mathbf{P}$ is the orthogonal projection matrix on the kernel subspace of $\tilde{\mathbf{\Phi}}$, i.e.

$$\mathbf{P} = \mathbf{I} - \tilde{\mathbf{\Phi}}^{\mathrm{H}}(\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^{\mathrm{H}})^{-1}\tilde{\mathbf{\Phi}}$$

$\tilde{\mathbf{r}}$ is the final recovery vector; $\hat{P}$ is the source number estimation; $\mathbf{p}_V$ and $\mathbf{p}_I$ are the magnitude and index sets corresponding to the local peaks of $\tilde{\mathbf{r}}$ which are sorted in the descending order, respectively; $\hat{\theta}$ is the DOA estimate vector.

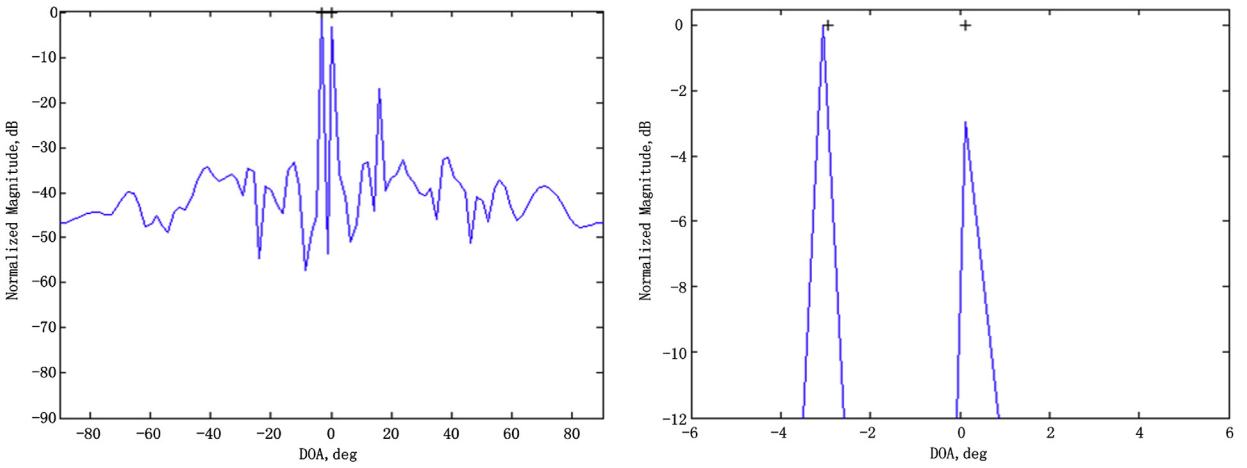
8. Alternating grid optimization

As discussed earlier, grid-based estimation is typical in practical algorithms. We outline in this section a local grid refinement technique for precision tuning. It is termed *alternating grid optimization*. The basic principle is to estimate the DOA of one source signal at a time with grid refinements [40]. The pseudocode of the algorithm is as follows:

Algorithm 2: Alternating grid optimization.

Input: $\tilde{\mathbf{\Phi}}$, $\mathbf{y}$,$\tilde{\theta}$,δ_θ,ϵ_θ;
Output: $\hat{\mathbf{r}}$, $\hat{\theta}$;
$E_\theta = 1$;
$[\tilde{\mathbf{r}}, \hat{P}, \tilde{\theta}] = \mathrm{DSETTD{+}NST{+}HT{+}subFB}(\tilde{\mathbf{\Phi}}(\tilde{\theta}), \mathbf{y}, \epsilon_1, \epsilon_2)$;
$\tilde{\theta}_0 = \mathrm{zeros}(1, \hat{P})$;
while $E_\theta > \epsilon_\theta$ **do**
{
  for $s = 1 : \hat{P}$ **do**
  $\{[\mathbf{p}_V, \mathbf{p}_I] = \mathrm{findpeaks}(|\tilde{\mathbf{r}}|, \text{`}descend\text{'})$;
  $\delta = \delta_\theta$;$p_V^{(M)} = \mathbf{p}_V(s)$; $p_I^{(M)} = \mathbf{p}_I(s)$;
  $F^{(M)} = \frac{\sum_{s=1}^{\hat{P}} \mathbf{p}_V(s)}{\sum_{i \notin I_p(1:\hat{P})} |\tilde{\mathbf{r}}(i)|}$;
    while $\delta > \epsilon_\theta$ **do**
    {
    $\delta = 0.5\delta$;
    $\tilde{\theta}^{(L)}(p_I^{(M)}) = \tilde{\theta}(p_I^{(M)}) - \delta$;
    $\tilde{\mathbf{r}}^{(L)} = \mathrm{NST{+}HT{+}subFB}(\tilde{\mathbf{\Phi}}(\tilde{\theta}^{(L)}), \mathbf{y}, \hat{P})$;
    $[\mathbf{p}_V^{(L)}, \mathbf{p}_I^{(L)}] = \mathrm{findpeaks}(|\tilde{\mathbf{r}}^{(L)}|, \text{`}descend\text{'})$;
    $F^{(L)} = \frac{\sum_{s=1}^{\hat{P}} \mathbf{p}_V^{(L)}(s)}{\sum_{i \notin \mathbf{p}_I^{(L)}(1:\hat{P})} |\tilde{\mathbf{r}}^{(L)}(i)|}$;
    $\tilde{\theta}^{(R)}(p_I^{(M)}) = \tilde{\theta}(p_I^{(M)}) + \delta$;
    $\tilde{\mathbf{r}}^{(R)} = \mathrm{NST{+}HT{+}subFB}(\tilde{\mathbf{\Phi}}(\tilde{\theta}^{(R)}), \mathbf{y}, \hat{P})$;
    $[\mathbf{p}_V^{(R)}, \mathbf{p}_I^{(R)}] = \mathrm{findpeaks}(|\tilde{\mathbf{r}}^{(R)}|, \text{`}descend\text{'})$;
    $F^{(R)} = \frac{\sum_{s=1}^{\hat{P}} \mathbf{p}_V^{(R)}(s)}{\sum_{i \notin \mathbf{p}_I^{(R)}(1:\hat{P})} |\tilde{\mathbf{r}}^{(R)}(i)|}$;
      if $F^{(L)} > F^{(M)}$ *and* $F^{(L)} > F^{(R)}$ **then**
      $\{\tilde{\theta} = \tilde{\theta}^{(L)}(\mathbf{p}_I^{(L)}(s))$;$\tilde{\mathbf{r}} = \tilde{\mathbf{r}}^{(L)}$;$F^{(M)} = F^{(L)}$;$\}$
      if $F^{(R)} > F^{(M)}$ *and* $F^{(R)} > F^{(L)}$ **then**
      $\{\tilde{\theta} = \tilde{\theta}^{(R)}(\mathbf{p}_I^{(R)}(s))$;$\tilde{\mathbf{r}} = \tilde{\mathbf{r}}^{(R)}$;$F^{(M)} = F^{(R)}$;$\}$
    }
  $\tilde{\theta}(s) = \tilde{\theta}(p_I^{(M)})$
  }
$E_\theta = \max(|\tilde{\theta} - \tilde{\theta}_0|)$;
$\tilde{\theta}_0 = \tilde{\theta}$;
}
return $\hat{\mathbf{r}} = \tilde{\mathbf{r}}, \hat{\theta} = \tilde{\theta}$;

In Algorithm 2, $\tilde{\theta}$ is the coarse DOA grid sequence with the gap δ_θ may be half of beamwidth; ϵ_θ represents DOA precision; $\hat{\theta}$ and $\hat{\mathbf{r}}$ are the final DOA estimation vector and the spatial spectrum, respectively; E_θ is the maximum difference between the last DOA estimation vector $\tilde{\theta}_0$ and the current DOA estimation vector $\tilde{\theta}$; δ is the adjustable grid spacing; $F^{(L)}$, $F^{(R)}$ and $F^{(M)}$ represents the left, the right and the mid cost



(a) Global spatial spectrum (b) Local spatial spectrum

Fig. 2. Spatial spectrum of AGO methods for two close sources under low SNR circumstance with limited snapshots by AGO.

function values, respectively; All superscripts L, R and M stand for the same meaning; Other variables, such as s, $\hat{P}$ and $\tilde{\mathbf{\Phi}}$ are the same as those in Algorithm 1.

9. Numerical examples

Extensive numerical experiments are performed. We present in this section a summary of the results. Note that, to take advantages of the sparse recovery technique, a randomly selected sparse linear array of $N = 20$ sensors is considered on an aperture of 16λ, in which the middle $N - 2$ sensor coordinates are selected randomly following an uniform distribution.

For clarity, case studies are presented in the following several subsections.

9.1. Two close-up sources placed within 3 dB beamwidth

Two narrowband source signals are considered, which are independent with low SNRs of 3 dB and 0 dB in the far-field impinge on the sparse linear array from the true DOAs of $-2.95°$ and $0.13°$. The DOA difference is $3.08°$, less than the 3 dB beam width of $3.17°$. The snapshot number is chosen as $M = 8$ which is less than the number of sensors, and typically considered very small. Testing results using techniques of AGO with initial coarser grids, MUSIC, CAPON, CVX, ADM, IAA and l_1-SVD with fine grids $0.1°$ are shown from Fig. 4 to Fig. 6. The sign "+" indicates actual locations of the DOAs in the figures.

9.1.1. DOA estimation via AGO with coarser initial grids

To alleviate the high correlation among frame columns under the fine grid of dictionary matrix, a coarser grid is used such as half of beamwidth at the beginning. Fig. 2 shows that the first peak and second peaks of alternating grid optimization spectrum are located at $-3.0490°$ and $0.1184°$ with the estimation biases $-0.0993°$ and $-0.0116°$.

The refinement process of the AGO technique can be seen in Fig. 3. Two peaks move towards the actual locations of the two source DOAs, as demonstrated by the labeled iterations.

9.1.2. DOA estimation via the methods of MUSIC and CAPON on fine grids

In comparison, Fig. 4 shows both MUSIC and CAPON methods fail quite notably in this case. The primary reason is due to the small number of snapshots $M = 8$, which results in clearly poor covariance matrix estimation, as is discussed earlier.

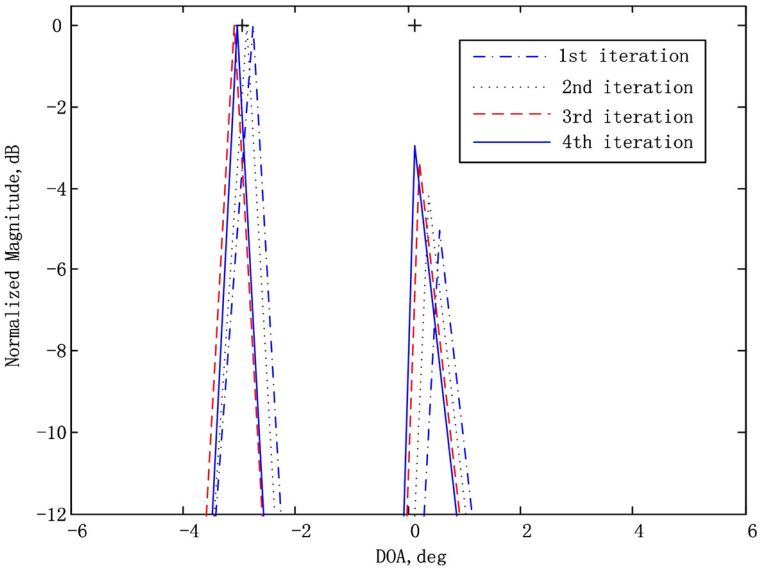
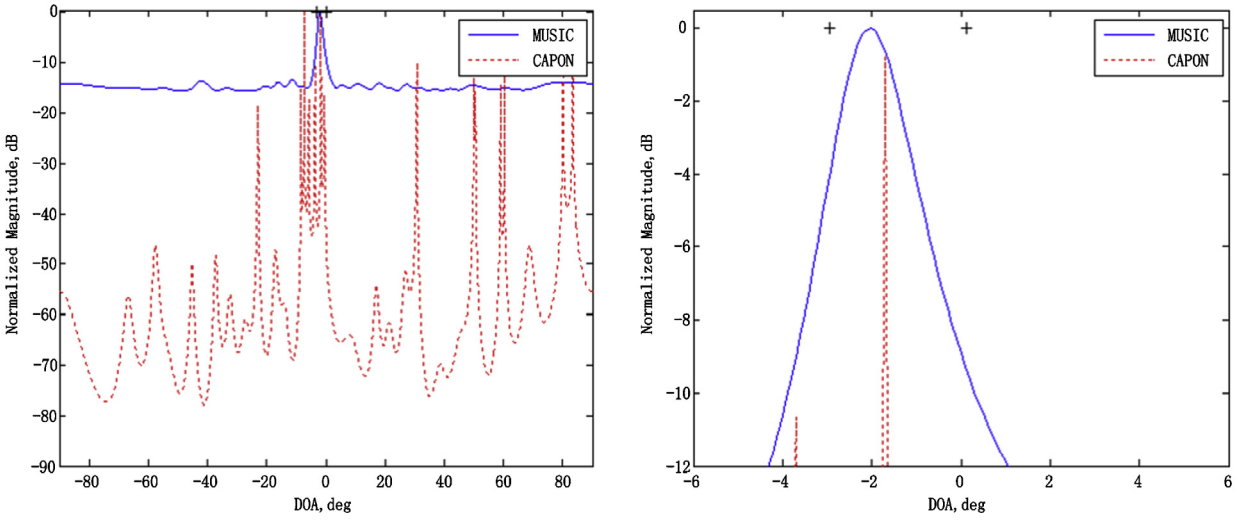


Fig. 3. Spatial Spectral of AGO method in iteration procedure.



(a) Global spatial spectra (b) Local spatial spectra

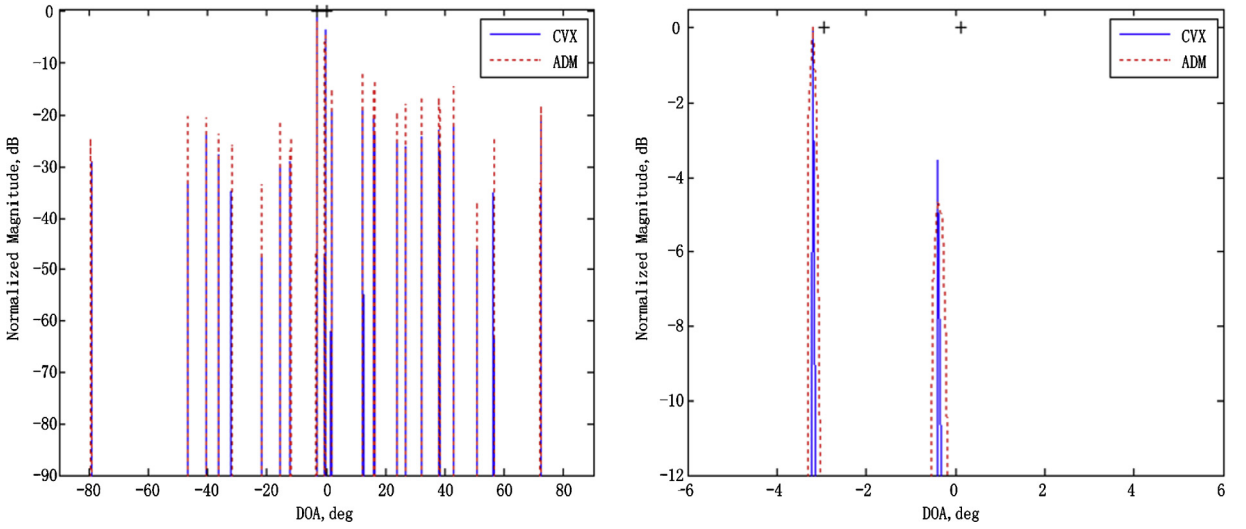
Fig. 4. Spatial spectra of MUSIC and CAPON methods for two close sources under low SNR circumstance with limited snapshots.

9.1.3. DOA estimation via CVX and ADM on fine grids

Fig. 5 is also for comparison, the first and the second peaks resulting from the CVX optimization and from the ADM technique both locate at $-3.2000°$ and $-0.4000°$ with the estimation biases $-0.2500°$ and $-0.5300°$, respectively. Although both methods of ADM and CVX find the DOAs relatively well, the absolute errors are still larger, and there are very high pseudo/fake peaks in the solution spectral of these two methods, see Fig. 5(a).

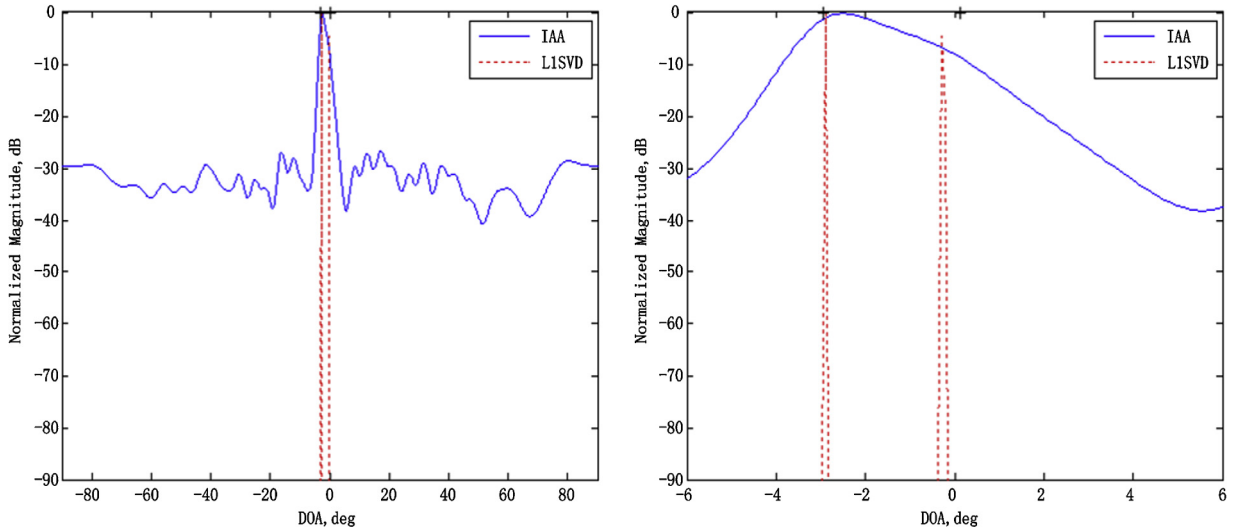
9.1.4. DOA estimation via IAA and l_1-SVD on fine grids

Fig. 6 is for comparison with the results by these two prominent techniques. One sees that only one peak of IAA spectrum appears between the true DOAs. Meantime, the first and second peaks of l_1-SVD



(a) Global spatial spectra (b) Local spatial spectra

Fig. 5. Spatial spectra of CVX and ADM methods for two close sources under low SNR circumstance with limited snapshots.



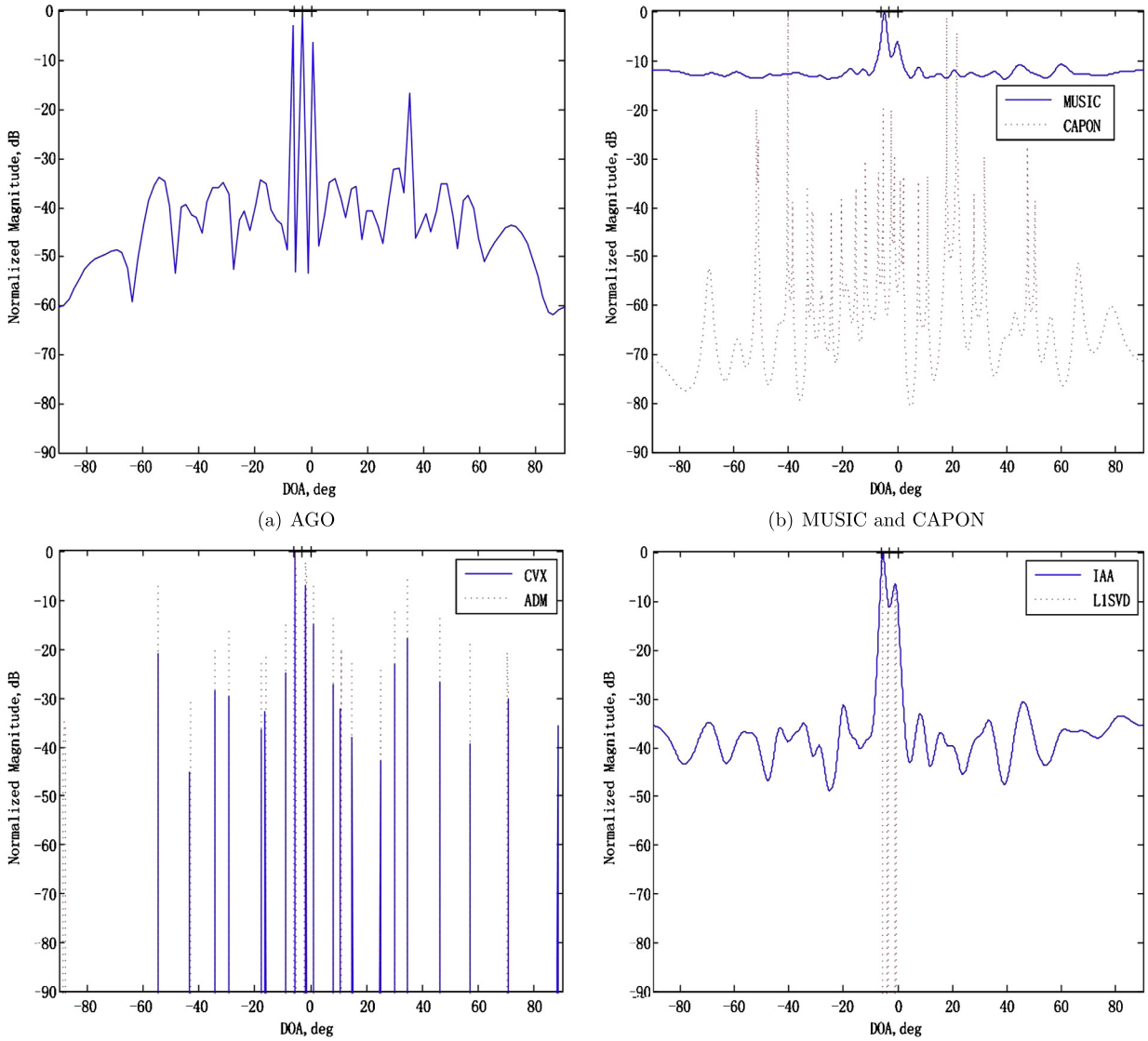
(a) Global spatial spectra (b) Local spatial spectra

Fig. 6. Spatial spectra of IAA and l_1-SVD methods for two close sources under low SNR circumstance with limited snapshots.

spectrum are located at $-2.9000°$ and $-0.3000°$ with the estimation biases $0.0500°$ and $-0.4300°$. With the right error tolerance, one could say that l_1-SVD succeeds in DOA estimation, but IAA fails. It is worth of commenting, however, that the technique of l_1-SVD is drastically more time consuming than our technique presented here.

9.2. Three close-up target sources

To verify the effectiveness of our technique for estimating DOAs of compactly placed sources, an independent source coming from $-6.03°$ with SNR 5 dB is added to the previous two targets. All other simulation conditions are identical to those in Section 9.1. Fig. 7(a) shows that our technique succeeds in resolving the three close-up target sources. Both MUSIC and CAPON fail as seen Fig. 7(b). Fig. 7(c) shows that CVX achieves relative success in resolving the three sources, but ADM fails. Fig. 7(d) demon-



(a) AGO (b) MUSIC and CAPON

(c) CVX and ADM (d) IAA and l_1-SVD

Fig. 7. Spatial spectra for compactly placed three sources with $M = 8$.

strates that l_1-SVD succeeds in resolving three sources with a much longer execution time, but IAA fails.

For a line-by-line comparison, detailed DOA estimation results of all methods are presented in Table 1.

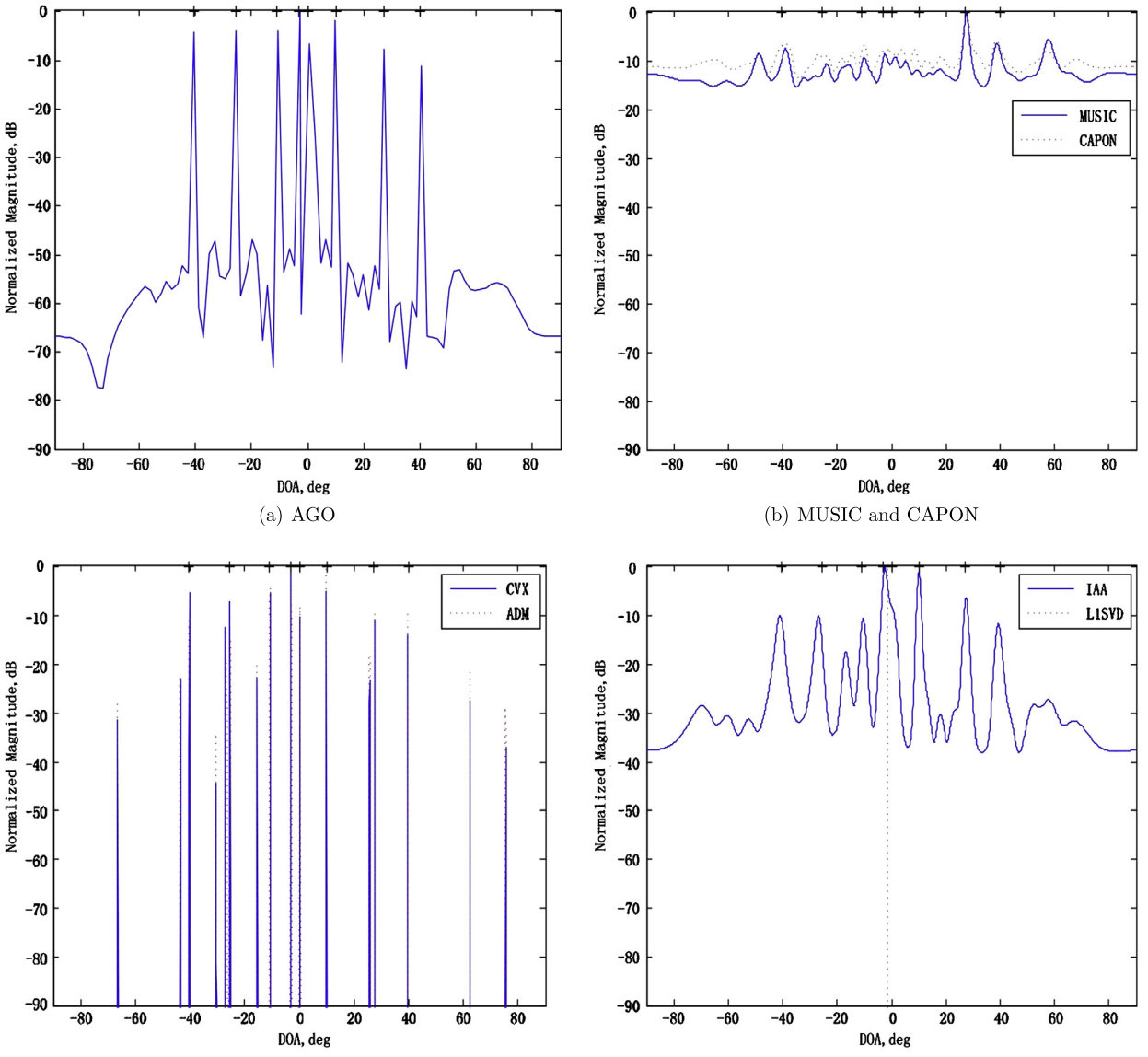
9.3. Multiple targets with coherent source signals

To further demonstrate the effectiveness of our technique, eight sources are considered. By their inter-coherence, these source signals are divided into three groups. The first group is composed of the first, the second and the third signals. The second group consists of the forth, the fifth and the sixth sources. The third group includes the seventh and the eighth sources.

The groups are independent, but the sources in the same group are coherent.

Table 1
DOA estimation results of three compactly placed sources.

Source ID	1	2	3
SNR (dB)	5	0	3
Actual DOA (°)	−6.0300	−2.9500	0.1300
MUSIC DOA (°)	−4.8000	−0.1000	60.1000
CAPON DOA (°)	−40.2000	−18.0000	21.6000
CVX DOA (°)	−5.5000	−1.8000	1.0000
ADM DOA (°)	−5.5000	−1.8000	34.7000
IAA DOA (°)	−5.5000	−1.0000	46.0000
l_1-SVD DOA estimation (°)	−5.6000	−3.7000	−0.8000
AGO DOA estimation (°)	−6.5724	−3.2566	0.5329



(a) AGO (b) MUSIC and CAPON

(c) CVX and ADM (d) IAA and l_1-SVD

Fig. 8. Spatial spectra for eight sources with coherent signals.

Fig. 8(a) to 8(d) are the results. Evidently, only our technique succeeds in resolving all sources. Detailed DOA estimation results are reported in Table 2.

Table 2
DOA estimation results of eight sources existing coherent cases.

Source ID	1	2	3	4	5	6	7	8
SNR (dB)	4.4915	4.5808	4.3227	8.0302	1.9717	7.1455	4.4310	1.7235
Actual DOA (°)	−40.5000	−25.7000	−10.8300	−2.9500	0.1300	10.0000	27.0800	40.1000
MUSIC DOA (°)	−48.6000	−38.9000	−10.0000	−2.4000	1.3000	27.6000	38.9000	57.8000
CAPON DOA (°)	−38.6000	−10.1000	−1.8000	3.8000	8.5000	27.7000	38.7000	57.4000
NST DOA (°)	−40.0000	−27.6000	−16.1000	−11.7000	−5.2000	−1.0000	11.0000	28.6000
ADM DOA (°)	−40.2000	−25.2000	−10.7000	−3.1000	0.2000	9.8000	27.6000	39.6000
IAA DOA (°)	−41.0000	−26.9000	−16.8000	−10.4000	−2.5000	10.0000	27.5000	39.4000
LISVD DOA (°)	−40.6000	−27.9000	−10.0000	−1.5000	9.4000	18.3000	27.9000	38.6000
AGO DOA (°)	−40.5296	−25.6678	−10.7467	−2.8717	0.7697	9.8882	26.9704	40.2928

9.4. Error performance evaluation

In this section, error performances are evaluated. The methods of CVX, ADM, l_1-SVD and our technique are chosen to study because they (more or less) succeed in the "two close-up sources placed within 3 dB beamwidth" simulation. Monte Carlo trial number is 100. For simplicity, only the first source from $-2.9500°$ is retained. The numerical results of root mean square error (RMSE) of DOA estimation varying with SNR, snapshot number, maximum phase mismatch and maximum gain mismatch, respectively, are shown in Fig. 9(a) to 9(d).

Fig. 9(a) is a plot of RMSE vs SNR. One sees that RMSEs of the four methods all decrease as SNR increases. The RMSEs of our technique are superior than those of other three methods from 15 dB to 25 dB and similar to that of l_1-SVD from 0 dB to 10 dB.

Fig. 9(b) is an evaluation of RMSE vs the number of snapshots M. The figure indicates that the performance of our technique AGO is consistently better than that of CVX, ADM, and better than that of l_1-SVD too for $M \geq 50$. The performance of AGO is very similar to that of l_1-SVD performance when the snapshot numbers $M < 50$, though it is seen that the RMSE is slightly above that of l_1-SVD result.

Fig. 9(c) is a plot of RMSE vs phase mismatches. There are large jumps for all four methods when $\Delta_p > 15°$. AGO is more robust than the three other method when $\Delta_p < 5°$ and the RMSE of AGO is slightly higher than that of l_1-SVD when $\Delta_p = 10°$ from the test.

Fig. 9(d) is the result of the evaluation of RMSE vs gain mismatches. Our technique AGO obtains lower RMSE than the three other methods except that AGO is slightly worse than l_1-SVD when $\Delta_g = 1.5$ dB. Without loss of generality, the phase mismatch and gain mismatch of the first sensor are assumed to be $0°$ and 0, respectively. The phase mismatch vector $\mathbf{\Gamma_p}$ and the gain mismatch vector $\mathbf{\Gamma_g}$ obey uniform distribution and standard normal distribution, respectively, which are calculated according to (9.1) and (9.2).

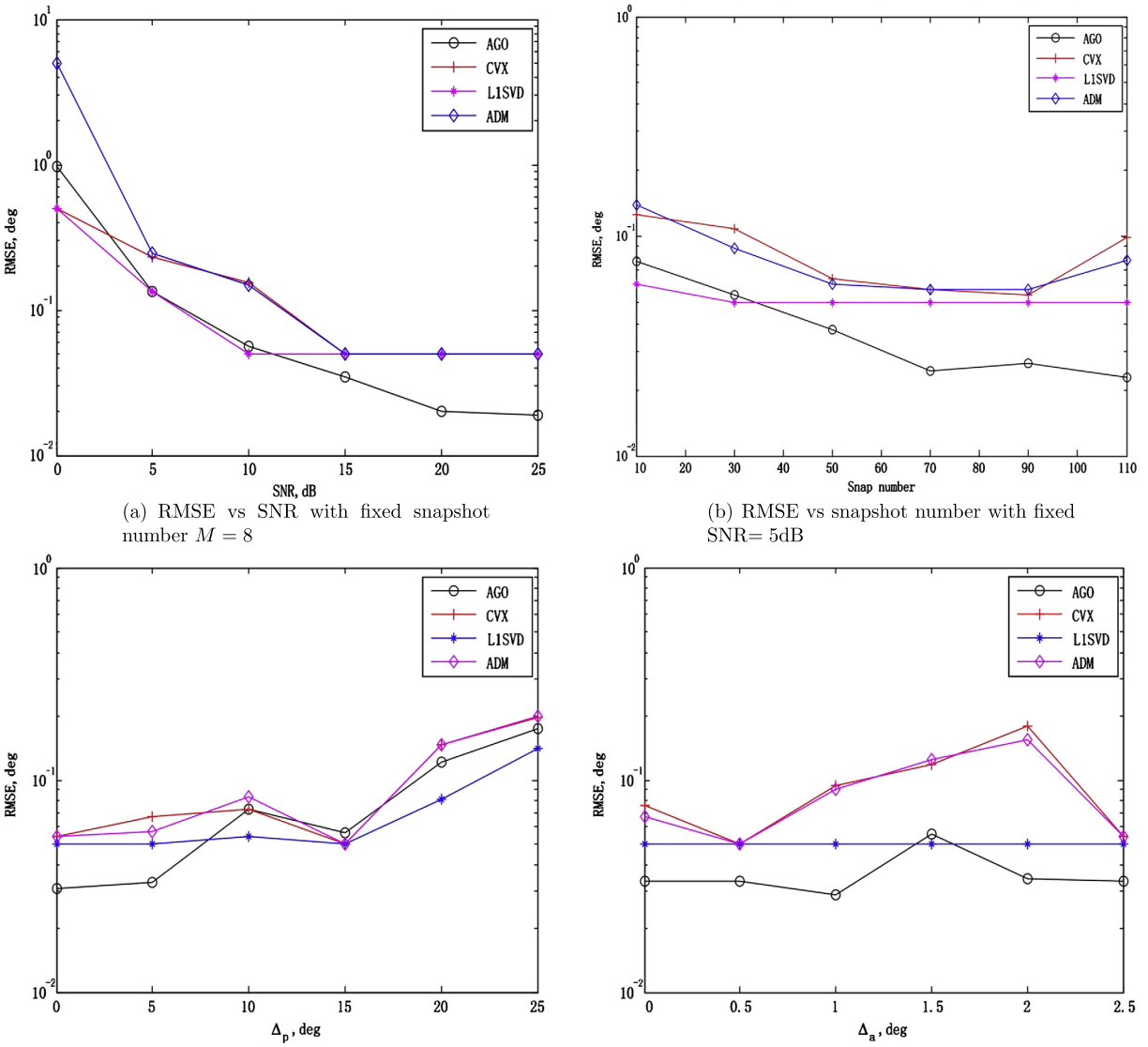
$$\mathbf{\Gamma_p} = [0; 2(rand(N-1,1) - 1) \times \Delta_p], \tag{9.1}$$

$$\mathbf{\Gamma_g} = [1; 1 + (10^{\frac{\Delta_g}{20}-1} \times (randn(N-1,1)))]. \tag{9.2}$$

The gain and phase mismatch vector $\mathbf{\Gamma} = \mathbf{\Gamma_g} \odot \mathbf{\Gamma_p}$.

9.5. Computational complexity evaluation

In order to compare the computation complexities of the four methods, simulation of execution time versus sensor number have been done on Dell workstation "PRECISION M6800" with i7-4910MQ CPU@3.90 GHz and Memory 32.0 GB. Sensor number varies by the numbers $N = 2^{k+2}$, $k = 1, 2, \cdots, 5$, other simulation conditions are the same as those of the simulation in the section "Two close-up sources placed within 3 dB beamwidth".



(a) RMSE vs SNR with fixed snapshot number $M = 8$

(b) RMSE vs snapshot number with fixed SNR= 5dB

(c) RMSE vs phase mismatch with fixed SNR= 5dB,$M = 64$ and maximum gain mismatch $\Delta_g = 0.5$dB

(d) RMSE varying with gain mismatch with fixed SNR= 5dB,$M = 64$ and maximum phase mismatch $\Delta_p = 5°$

Fig. 9. Performance comparison results based on sparse linear array consists of 20 sensor elements.

Fig. 10 demonstrates that the execution time of our technique AGO is the smallest among the four methods and increase slightly with the increase of the number of sensors. Detailed execution times of four methods are listed in Table 3. Evidently, the ascending order of the computation complexity is given by AGO, CVX, ADM and l_1-SVD. Our technique AGO spent a negligible amount of execution time than that all other techniques, particularly for larger array. It is also important to note that AGO can still be run on our workstation, while l_1-SVD is out of memory when $N = 256$.

10. Conclusion

A computationally efficient and super-resolution method of DOA estimation via sparse representation and the algorithm of null space tuning is proposed. Several advantages over existing DOA estimation meth-

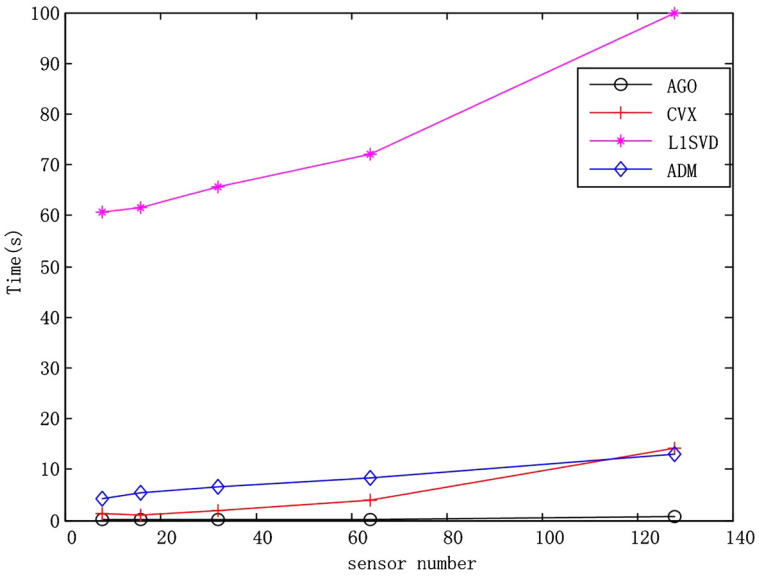


Fig. 10. Sparse linear array geometry.

Table 3
Cost times varying with sensor number.

sensor number	8	16	32	64	128
Execution time of AGO (s)	0.0701	0.0526	0.1979	0.2383	0.8134
Execution time of CVX (s)	1.2468	1.0513	1.8877	4.0500	14.1180
Execution time of l_1-SVD (s)	60.8174	61.6122	65.6252	72.0442	99.8086
Execution time of ADM (s)	4.2128	5.3126	6.5444	8.3508	13.1571

ods have been identified. These include the automatic and dynamic source number estimation, improved robustness to noise and to limited number of snapshots, and the much improved resolution.

The computational efficiency of our technique in terms of both precision and execution time is notably superior in general.

Acknowledgments

This work was supported by National Basic Research Program of China (2011CB707001), State Key Program of National Natural Science Foundation of China (61231027), Natural Science Foundation of China (61101243, 61201295), the Natural Science Foundation of Shanxi Province (2014JQ8304), and the Program for Changjiang Scholars and Innovative Research Teams in Universities (IRT 0954). The third author is also partially supported by National Science Foundation of USA (DMS-1313490, DMS-1010058), and by AFOSR grant (FA9550-11-1-0245).

References

[1] H.L. VanTrees, Optimum Array Processing, John Wiley and Sons, Inc., New York, 2002.
[2] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas and Propagation 34 (3) (1986) 276–280.
[3] R. Roy, T. Kailath, ESPRIT-a subspace rotation approach to estimation of parameters of sinusoids in noise, IEEE Trans. Acoust. Speech Signal Process. 34 (5) (1986) 1340–1342.
[4] W. Roberts, J. Li, Iterative adaptive approaches to MIMO radar imaging, IEEE J. Sel. Top. Signal Process. 4 (1) (2010) 5–20.
[5] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (1) (2008) 143–168.

[6] S.G. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.
[7] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, Found. Comput. Math. 9 (3) (2009) 317–334.
[8] J.A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (10) (2004) 2231–2242.
[9] D.L. Donoho, Y. Tsaig, I. Drori, J.-L. Starck, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, Tech. rep., Stanford Univ., 2006.
[10] D. Needell, J. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.
[11] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inform. Theory 55 (5) (2009) 2230–2249.
[12] J.M. Bioucas-Dias, M.A.T. Figueiredo, A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration, IEEE Trans. Image Process. 16 (12) (2007) 2992–3004.
[13] K. Bredies, D.A. Lorenz, Iterated hard shrinkage for minimization problems with sparsity constraints, SIAM J. Sci. Comput. 30 (2) (2008) 657–683.
[14] K. Bredies, D.A. Lorenz, Linear convergence of iterative soft-thresholding, J. Fourier Anal. Appl. 14 (5–6) (2008) 813–837.
[15] E.J. Candès, J. Romberg, Signal recovery from random projections, Proc. SPIE 5674 (2005) 76–86.
[16] I. Daubechies, M. Defrise, C.D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. 57 (11) (2004) 1413–1457.
[17] M. Ehler, Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation, J. Inverse Ill-Posed Probl. 19 (4–5) (2011) 593–614.
[18] M. Fornasier, H. Rauhut, Iterative thresholding algorithms, Appl. Comput. Harmon. Anal. 25 (2) (2008) 187–208.
[19] M. Fornasier, H. Rauhut, Recovery algorithms for vector-valued data with joint sparsity constraints, SIAM J. Numer. Anal. 46 (2) (2008) 577–613.
[20] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[21] M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J. Sel. Top. Signal Process. 1 (4) (2007) 586–597.
[22] S.J. Wright, R.D. Nowak, M.A.T. Figueiredo, Sparse reconstruction by separable approximation, IEEE Trans. Signal Process. 57 (7) (2009) 2479–2493.
[23] J.-F. Cai, S. Osher, Z. Shen, Linearized Bregman iterations for compressed sensing, Math. Comp. 78 (1) (2009) 1515–1536.
[24] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (1) (2008) 143–168.
[25] T. Blumensath, M.E. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (5–6) (2008) 629–654.
[26] K. Lange, Optimization, Springer, 2004.
[27] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (6) (2011) 2543–2563.
[28] S. Li, Y. Liu, T. Mi, Fast thresholding algorithms with feedbacks for sparse signal recovery, Appl. Comput. Harmon. Anal. 37 (1) (2014) 69–88.
[29] J. Yang, Y. Zhang, A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 288–297.
[30] C. Micchelli, L. Shen, Y. Xu, X. Zeng, Proximity algorithms for image models: II. L1/TV denoising, Adv. Comput. Math.
[31] C. Micchelli, L. Shen, Y. Xu, X. Zeng, Proximity algorithms for the L1/TV image denoising model, Adv. Comput. Math. (2011) 1–26.
[32] C. Micchelli, L. Shen, Y. Xu, Proximity algorithms for image models: denoising, Inverse Probl. 27 (4) (2011) 045009.
[33] D. Malioutov, M. Cetin, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[34] E. Candes, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Comm. Pure Appl. Math. LXVII (2014) 0906–0956.
[35] S. Li, Y. Liu, T. Mi, Fast thresholding algorithms with feedbacks for sparse signal recovery, Appl. Comput. Harmon. Anal. 37 (1) (2014) 69–88.
[36] M.D. Iordache, J.M. Bioucas-Dias, A. Plaza, B. Somers, MUSIC-CSR: hyperspectral unmixing via multiple signal classification and collaborative sparse regression, IEEE Trans. Geosci. Remote Sens. 52 (7) (2014) 4364–4382.
[37] X.Z. Ren, Y. Qin, L. Tian, Three-dimensional imaging algorithm for tomography sar based on multiple signal classification, in: 2014 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC, 2014, pp. 120–123.
[38] H.K. Hao, H. Liang, Y. Liu, Particle methods for real-time sound source localization based on the multiple signal classification algorithm, in: 2014 International Conference on Intelligent Green Building and Smart Grid, IGBSG, 2015, pp. 1–5.
[39] M.A. Richards, Fundamentals of Radar Signal Processing, McGrawHill, Inc., New York, 2005.
[40] C. Zeng, S. Li, Methods and apparatus of DOA estimations via alternating grid optimization, patent pending, 2015.