Signal Processing 105 (2014) 98–108

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

DOA and power estimation using a sparse representation of second-order statistics vector and ℓ_0-norm approximation

Ye Tian [a,*], Xiaoying Sun [a], Shishun Zhao [b]

[a] *College of Communication Engineering, Jilin University, Changchun, Jilin 130022, China*
[b] *Institute of Mathematics, Jilin University, Changchun, Jilin 130012, China*

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form
8 April 2014
Accepted 16 May 2014
Available online 26 May 2014

Keywords:
Direction-of-arrival and power estimation
Sparse representation
ℓ_0-norm approximation
TLP
DC decomposition

ABSTRACT

This paper focuses on the problem of joint direction-of-arrival (DOA) and power estimation based on sparse signal reconstruction. In this scheme, we utilize the second-order statistics (SOS) domain data of array output to construct a kind of special column vector, which also contains sufficient information on DOA and power parameters. Our aim is to transform the multiple measurement vectors (MMV) or "group sparsity" problem to the virtual single measurement vector (VSMV) problem in sparse signal representation framework. Concerning accuracy and complexity of estimation, we exploit a surrogate-TLP (truncated ℓ_1 function) to approximate ℓ_0-norm, and successively demonstrate how the nonconvex minimization problem can be treated by the DC (Difference of Convex functions) decomposition and the iterative approach. Theoretically, we prove that the proposed reconstruction algorithm can provide a stable and satisfactory performance, provided that the tuning parameter is selected properly and the noise is bounded. In addition, we also introduce an appropriate parameter selection strategy to make the algorithm robust. Numerical simulations show that the proposed algorithm not only has high resolution and good robustness to noise, but also provides an almost unbiased power estimation.

1. Introduction

Source parameter estimation is an important problem in array signal processing fields such as smart antennas, mobile communication systems, radar, sonar and seismic exploration applications [1]. Many methods for source parameter estimation, like beamforming [2], Capon's method [3], MUSIC [4] and ESPRIT [5], have been proposed over the years, based on some parameterized array data models with some deterministic and stochastic assumptions. However, the performance of these methods is generally not satisfactory in low SNR and closely spaced sources. It should be pointed out that the maximum likelihood estimation (MLE) [6] enjoys excellent statistical properties, but the corresponding cost function is non-convex, thus an accurate initialization is required to converge to a global minimum.

In recent years, a novel approach, termed sparse signal reconstruction, has been introduced for source parameter estimation, which brings some superiorities on resolution and robustness to noise. The sparse signal reconstruction process can be regarded as the ℓ_0-norm optimization problem. However, it has been stated in the literature [7,8] that the direct ℓ_0-norm optimization problem is NP-hard, and sensitive to noise (because any small amount of noise can change the number of zero components significantly).

Several alternatives have been proposed. In [9], a recursive weighted least-squares algorithm called FOCUSS is addressed for source parameter estimation. In [10,11], a

* Corresponding author.
E-mail address: tiany11@mails.jlu.edu.cn (Y. Tian).

http://dx.doi.org/10.1016/j.sigpro.2014.05.014
0165-1684/

whitened sparse covariance-based representation model is presented for source parameter estimation by applying the global matched filter (GMF). So far, the most representative sparse recovery algorithm for source parameter estimation is ℓ_1-SVD [12], which exploits ℓ_1-norm penalty to enforce sparsity and singular value decomposition (SVD) to reduce the computational complexity and the sensitivity to noise. Although the ℓ_1-norm minimization is a convex problem and the global minima can be guaranteed easily, it is not without drawbacks. An important weakness of the ℓ_1-norm minimization is their undemocratic penalization for larger coefficients, which incurs the degradation of signal recovery performance. To conquer this problem, the iterative reweighted ℓ_1 minimization [13] was designed, where the large weights could be used to discourage nonzero entries in the recovered signals. Based on [12,13], Zheng et al. [14] and Xu et al. [15] propose the weighted ℓ_1-norm penalty utilizing the property of noise subspace and Capon spectrum function to improve the performance of ℓ_1-norm minimization, respectively. In contrast, Yin and Chen [16] use the concept of a sparse representation of array covariance vectors for source parameter estimation, Stoica et al. [17] propose a sparse iterative covariance-based estimation (SPICE) approach for array signal processing by the minimization of a covariance matrix fitting criterion. However, the computational complexity of the methods addressed in [12,14–17] grows proportionally with the number of sources or sensors. In view of this, Zheng et al. [18] present a sparse spectral fitting method to provide low complexity method for DOA and power estimation, which can be regarded as a case of general LASSO. Instead of approximating ℓ_0-norm with the ℓ_1-norm, Hyder and Mahata [19] exploit a class of Gaussian functions to deal with the $\ell_{2,0}-$norm approximation problem, and further propose an alternative strategy named JLZA-DOA. Although it can provide higher DOA resolution performance with very few measurements, the convergence to global minima is not guaranteed.

In this paper, we present a new algorithm for direction-of-arrival and power estimation based on a sparse representation of second-order statistics vector and ℓ_0-norm approximation (SRSSV-ℓ_0). We transform the multiple measurement vectors (MMV) or "group sparsity" problem to the virtual single measurement vector (VSMV) problem in second-order statistics domain. In the implementation process, we use TLP (truncated ℓ_1 function) [20] to approximate ℓ_0-norm, and DC (Difference of Convex functions) decomposition [21,22] to treat nonsmooth and nonconvex minimization problem. Finally, the ℓ_0-norm penalty optimization problem is converted into a general weighted LASSO problem with nonzero weights, then the global minimum and asymptotic selection consistency are guaranteed without the need for an accurate initialization (see Section 3.1 for details). Note that the TLP and DC decomposition algorithm have been considered in high-dimensional data analysis and feature selection, which have exhibited several other advantages, such as capability of handing low-resolution coefficients, good approximation to ℓ_0-norm and efficient computation, see [20,22] and references therein. In the following, we extend TLP and DC decomposition to deal with far-field DOA and power estimation problem arising in array signal processing. We also provide an appropriate parameter selection strategy to make the algorithm robust at each iteration. Owing to the sparse representation model and the reconstruction method we adopt, the proposed approach shows several salient advantages, such as high resolution, improved estimation accuracy and robustness to noise, which are verified in comparison with several existing sparse signal reconstruction based methods as well as Cramér–Rao lower bound (CRLB).

The remainder of this paper is organized as follows: in Section 2, we describe the background involving signal model, the DOA and power estimation problem in the sparse signal reconstruction perspective and some existing representative reconstruction methods. In Section 3, a new DOA and power estimation algorithm using a sparse representation of second-order statistics vector and ℓ_0-norm approximation is introduced. Numerical simulations are conducted in Section 4 to evaluate the effectiveness of the proposed algorithm. Conclusions are given in Section 5.

2. Background

2.1. Signal model

Consider K far-field, narrow-band, uncorrelated signals with power $\{P_k\}_{k=1}^K$ impinging on a uniform linear array (ULA) consisting of L omnidirectional sensors from directions $\boldsymbol{\theta}=[\theta_1,\theta_2\ldots,\theta_K]$, which are corrupted by additive circular complex Gaussian white noise. Let the first sensor be the phase reference point, the array output at the tth snapshot can be expressed as

$$\mathbf{y}(t)=\sum_{k=1}^{K}\mathbf{a}(\theta_k)s_k(t)+\mathbf{n}(t)=\mathbf{A}\mathbf{s}(t)+\mathbf{n}(t),\quad t=1,2,\ldots,T \tag{1}$$

where $\mathbf{s}(t)=[s_1(t),s_2(t),\ldots,s_K(t)]^{\mathrm{T}}$ is the $K\times 1$ zero-mean signal vector, $(\cdot)^{\mathrm{T}}$ is the transpose operation, $\mathbf{n}(t)$ is the $L\times 1$ noise vector with zero mean and the power of each entry equals to σ^2, T is the number of data snapshots. By assumption, the entries of $\mathbf{n}(t)$ are uncorrelated with each other and the signals. $\mathbf{A}\triangleq[\mathbf{a}(\theta_1),\mathbf{a}(\theta_2),\ldots,\mathbf{a}(\theta_K)]$ is the $L\times K$ steering matrix, whose kth column is the $L\times 1$ steering vector corresponding to the DOA of the kth source, i.e., θ_k, and can be expressed as

$$\begin{aligned}\mathbf{a}(\theta_k)&=[1,e^{-j2\pi d\ \sin(\theta_k)/\lambda},\ldots,e^{-j2\pi(L-1)d\ \sin(\theta_k)/\lambda}]^{\mathrm{H}}\\&=[1,e^{-j\pi\ \sin(\theta_k)},\ldots,e^{-j\pi(L-1)\ \sin(\theta_k)}]^{\mathrm{H}}\end{aligned} \tag{2}$$

with $d=\lambda/2$, where λ and d denote the carrier wavelength and intersensor spacing, respectively. In a more concise expression, (1) can be rewritten as

$$\mathbf{Y}=\mathbf{A}(\theta)\mathbf{S}+\mathbf{N} \tag{3}$$

where $\mathbf{Y}=[\mathbf{y}(1),\ldots,\mathbf{y}(T)]\rightarrow L\times T$ array data matrix, $\mathbf{S}=[\mathbf{s}(1),\ldots,\mathbf{s}(T)]\rightarrow K\times T$ source waveform matrix and $\mathbf{N}=[\mathbf{n}(1),\ldots,\mathbf{n}(T)]\rightarrow L\times T$ sensor noise matrix.

From (1), we can also obtain the array covariance matrix

$$\mathbf{R} = E\{\mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)\} = \mathbf{A}\mathbf{P}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}_L \tag{4}$$

where $\mathbf{P} = E\{\mathbf{s}(t)\mathbf{s}^{\mathrm{H}}(t)\} = \mathrm{diag}\{P_1, P_2, \ldots, P_K\}$ is the signal covariance matrix, $P_k = E\{s_k(t)s_k^*(t)\}$ is the power of the kth signal, $\mathbf{I}_L$ is an $L \times L$ identity matrix, $E\{\cdot\}$, $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ denote the expectation, the conjugate and the conjugate transpose respectively, and the symbol $\mathrm{diag}\{z_1, z_2\}$ represents a diagonal matrix with diagonal entries z_1 and z_2.

2.2. DOA and power estimation as a sparse signal reconstruction problem

To formulate the DOA and power estimation as a sparse signal reconstruction problem, one should divide the whole area of interest into some discrete set of potential locations, such as a set of pixels in near-field case or a grid of direction-of-arrival angles in far-field case. Here, we mainly consider the far-field scenario, let $\boldsymbol{\Theta} = \{\tilde{\theta}_1, \ldots, \tilde{\theta}_Q\}$ be a sampling grid of all potential directions in spatial domain. In general, $Q \gg L > K$. By collecting all steering vectors that correspond to the element in $\boldsymbol{\Theta}$, we can obtain an overcomplete basis, which is given by

$$\boldsymbol{\Phi} = [\mathbf{a}(\tilde{\theta}_1), \ldots, \mathbf{a}(\tilde{\theta}_Q)]. \tag{5}$$

Note that $\boldsymbol{\Phi}$ is known and contains all the information and elements of $\mathbf{A}$, therefore, (1) and (4) can also be modeled as

$$\mathbf{y}(t) = \boldsymbol{\Phi}\mathbf{x}(t) + \mathbf{n}(t), \quad t = 1, 2, \ldots, T, \tag{6}$$

$$\mathbf{R} = \boldsymbol{\Phi}\mathbf{P}_Q\boldsymbol{\Phi}^{\mathrm{H}} + \sigma^2\mathbf{I}_L \tag{7}$$

respectively, where $\mathbf{x}(t)$ is $Q \times 1$ vector, $\mathbf{P}_Q$ is $Q \times Q$ matrix. The qth element of $\mathbf{x}(t)$ and (q, q)th element of $\mathbf{P}_Q$ are nonzero and equal to $s_k(t)$ and P_k respectively if source k comes from $\tilde{\theta}_q$ for some k and zero otherwise. It means that we can estimate the DOA and power parameters as long as we find the locations and the size of nonzero elements in $\mathbf{x}(t)$ or $\mathbf{P}_Q$. That is, the problem of DOA and power estimation is converted into one of sparse signal reconstruction problem from Eq. (6) or (7).

2.3. Brief analysis of existing representative reconstruction methods

The sparse signal reconstruction process can be regarded as the ℓ_0-norm optimization problem. However, it is computationally infeasible to minimize a nonconvex and discontinuous cost function involving the ℓ_0 function. An alternative strategy is to use ℓ_1–norm penalty instead of ℓ_0-norm penalty to enforce sparsity, which leads to the following optimization problems:

$$\min \|\tilde{\mathbf{x}}^{(\ell_2)}\|_1 \quad \text{s.t. } \|\mathbf{Y}_{SV} - \boldsymbol{\Phi}\mathbf{X}_{SV}\|_F^2 \le \beta_1^2, \tag{8}$$

$$\min \|\tilde{\mathbf{b}}^{(\ell_2)}\|_1 \quad \text{s.t. } \|\mathbf{W}^{-1/2}\mathrm{vec}[\hat{\mathbf{R}} - \boldsymbol{\Phi}\mathbf{B} - \sigma^2\mathbf{I}_L]\|_2^2 \le \beta_2^2 \tag{9}$$

where $\|\boldsymbol{\Lambda}\|_F^2 = \|\mathrm{vec}(\boldsymbol{\Lambda})\|_2^2$ is the Frobenius norm of $\boldsymbol{\Lambda}$. $\mathbf{Y} = \mathbf{U}\mathbf{L}\mathbf{V}^{\mathrm{H}}$, $\mathbf{Y}_{SV} = \mathbf{Y}\mathbf{V}\mathbf{D}_K$, $\mathbf{X}_{SV} = \mathbf{X}\mathbf{V}\mathbf{D}_K$, $\mathbf{D}_K = [\mathbf{I}_K, \mathbf{0}]^{\mathrm{T}}$, and $\mathbf{0}$ is a $K \times (T-K)$ zero matrix. $\hat{\mathbf{R}}$ is the estimated result of array covariance matrix $\mathbf{R}$, i.e. $\hat{\mathbf{R}} = (1/T)\sum_{t=1}^{T}\mathbf{y}(t)\mathbf{y}^{\mathrm{H}}(t)$. $\mathbf{W}^{-1/2} = \sqrt{T}\hat{\mathbf{R}}^{-T/2} \otimes \hat{\mathbf{R}}^{T/2}$ ($\otimes$ is the Kronecker matrix product), $\mathbf{B} = \mathbf{P}_Q\boldsymbol{\Phi}^{\mathrm{H}}$. $\tilde{\mathbf{x}}^{(\ell_2)}$ and $\tilde{\mathbf{b}}^{(\ell_2)}$ are the estimated spectrum whose entries are defined as to be the ℓ_2-norm of the corresponding rows of $\mathbf{X}_{SV}$ and $\mathbf{B}$ respectively. β_1 and β_2 are regularization parameters.

Formulations (8) and (9) are just the ℓ_1–SVD and L_1-SRACV methods addressed in [12] and [16] respectively. Note that this class of methods exploits the feature of "group sparsity" for parameter estimation, so they can be regarded as "group LASSO". However, the ℓ_1–norm penalty associated to LASSO and group LASSO has been proved to produce biased estimates for large coefficient [23,24]. This will incur the degradation of signal reconstruction performance. In addition, the computational complexity of these methods grows proportionally with the number of sources or sensors, which will restrict their practical applications when the number of sources is large. In the following, we will overcome these drawbacks from two aspects, i.e. new data model and new reconstruction algorithm.

3. DOA and power estimation

In this section, we first demonstrate how the ℓ_0-norm penalty minimization problem can be cast as a weighted LASSO problem using TLP and DC decomposition, and give theoretical evidence that the combination of TLP and DC decomposition can lead to a good reconstruction performance. Subsequently, we extend the new reconstruction algorithm to deal with the virtual single measurement vector (VSMV) problem arising in DOA and power estimation.

3.1. Theory on TLP and DC decomposition

Consider the problem of finding the sparsest representation possible in an overcomplete basis $\boldsymbol{\Phi}_1$. As a measure of sparsity of a vector $\mathbf{x}$, we take the so-called ℓ_0-norm $\|\mathbf{x}\|_0$, which is simply the number of nonzero elements in $\mathbf{x}$. The sparsest representation is then the solution to the optimization problem

$$\min \|\mathbf{y} - \boldsymbol{\Phi}_1\mathbf{x}\|_2 + h\|\mathbf{x}\|_0 \tag{10}$$

where $\mathbf{x} \in \mathbb{R}^Q$, $h > 0$ is the regularization parameter that controls the tradeoff between ℓ_2 term and ℓ_0 term. As stated above, this seems to be a general combinatorial optimization problem, which is hard to solve.

As a surrogate, we seek a good approximation of the ℓ_0 function by the TLP, which is defined as

$$J(|z|) = \min(|z|/\tau, 1) \tag{11}$$

where $\tau > 0$ is a tuning parameter controlling the degree of approximation. In practical applications, the parameter τ must be tuned such that $\tau < P_{\min} = \min\{P_k : 1 \le k \le K\}$ to make the approximation error of the TLP function to the ℓ_0 function become zero. To treat nonconvex minimization, we replace the ℓ_0 function by its surrogate $J(\cdot)$ to construct

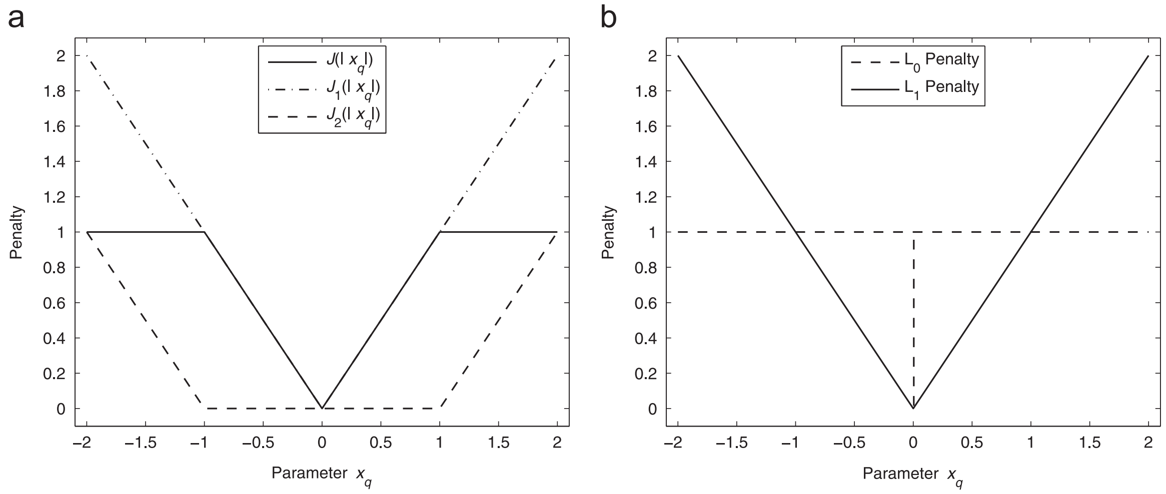

Fig. 1. Illustration of different penalty functions. (a) Truncated ℓ_1 function J with $\tau=1$ and its DC decomposition into a difference of two convex functions J_1 and J_2. (b) ℓ_0 penalty and ℓ_1 penalty.

an approximation of (10), i.e.

$$\min\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{x}\|_2+h\sum_{q=1}^{Q}J(|x_q|). \tag{12}$$

Just like (10), formulation (12) is also a noncovex minimization procedure. We develop an iterative procedure known as DC decomposition for solving this problem.

The core idea of DC decomposition [21,22] is to decompose a nonconvex (possible non-smooth) criterion $S(\cdot)$ as

$$S(\mathbf{x})=S_1(\mathbf{x})-S_2(\mathbf{x}) \tag{13}$$

where $S_1(\mathbf{x})$ and $S_2(\mathbf{x})$ are lower semi-continuous, proper convex functions on $\mathbb{R}^Q$. Here, we define them as

$$S_1(\mathbf{x})=\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{x}\|_2+h\sum_{q=1}^{Q}J_1(|x_q|) \tag{14}$$

$$S_2(\mathbf{x})=h\sum_{q=1}^{Q}J_2(|x_q|) \tag{15}$$

$$J_1(|x_q|)=|x_q|/\tau \tag{16}$$

$$J_2(|x_q|)=\max(|x_q|/\tau-1,0). \tag{17}$$

The truncated ℓ_1 function $J(|x_q|)$ with $\tau=1$, its DC decomposition into a difference of two convex functions $J_1(|x_q|)$ and $J_2(|x_q|)$ are depicted in Fig. 1(a). As a comparison, the ℓ_0 penalty and ℓ_1 penalty (LASSO) are depicted in Fig. 1(b). We point out that the ℓ_1 penalty is unbounded and this may lead to estimation bias [23]. In contrast, the truncated ℓ_1 function penalty is bounded and can correct the LASSO bias efficiently through adaptive shrinkage combining shrinkage with thresholding.

Given this DC decomposition, a sequence of upper approximation of $S(\mathbf{x})$ is constructed iteratively with $\nabla S_2(\mathbf{x})$ a subgradient of $S_2(\mathbf{x})$ in $|\mathbf{x}|$. That is, at the mth iteration, $S^{(m)}(\mathbf{x})=S_1(\mathbf{x})-(S_2(\hat{\mathbf{x}}^{(m-1)})+(|\mathbf{x}|-|\hat{\mathbf{x}}^{(m-1)}|)^{\mathrm{T}}\nabla S_2(|\hat{\mathbf{x}}^{(m-1)}|))$, by successively replacing $S_2(\mathbf{x})$ by its minorization. After ignoring $S_2(\hat{\mathbf{x}}^{(m-1)})-(h/\tau)\sum_{q=1}^{Q}|\hat{x}_q^{(m-1)}|I(|\hat{x}_q^{(m-1)}|>\tau)$ from $S^{(m)}(\mathbf{x})$ which is independent of $\mathbf{x}$, the problem (10) reduces to

$$\min\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{x}\|_2+\frac{h}{\tau}\sum_{q=1}^{Q}|x_q|I\left(|\hat{x}_q^{(m-1)}|\le\tau\right) \tag{18}$$

where the indicator function $I(\cdot)$ is given by

$$\begin{cases} I(|\hat{x}_q^{(m-1)}|>\tau)=0\\ I(|\hat{x}_q^{(m-1)}|\le\tau)=1\end{cases}. \tag{19}$$

In fact, if there exist zero-valued components in $\mathbf{x}$ which are not shrunk to the components that satisfy $|\mathbf{x}^{(0)}(k)|\le\tau$ in the initial estimation, the corresponding weight will be set to 0, which means that they will not be penalized any more in the subsequent iterations. To provide stability and to ensure that a component satisfying $|\mathbf{x}^{(0)}(k)|>\tau$ in initial estimation does not strictly prohibit a zero estimate at the end, we introduce a parameter $\epsilon>0$ in $I(\cdot)$, i.e.

$$\begin{cases} I(|\hat{x}_q^{(m-1)}|>\tau)=\epsilon\\ I(|\hat{x}_q^{(m-1)}|\le\tau)=1\end{cases}. \tag{20}$$

However, the parameter ϵ should be set much smaller than 1 to achieve a good approximation of formulation (18).

Further, the formulation (18) can be rewritten as the following weighted form:

$$\min\|\mathbf{y}-\boldsymbol{\Phi}_1\mathbf{x}\|_2+h\|\mathbf{W}^{(m-1)}\mathbf{x}\|_1 \tag{21}$$

where $\mathbf{W}^{(m-1)}=\mathrm{diag}\{\omega_1,\ldots,\omega_Q\}$, and its (q,q)th element ω_q satisfies

$$\omega_q=\begin{cases}\dfrac{1}{\tau}, & |\hat{x}_q^{(m-1)}|\le\tau\\ \dfrac{\epsilon}{\tau} & \text{otherwise}\end{cases}. \tag{22}$$

After the above processing, we observe that the ℓ_0-norm penalty optimization problem (10) reduces to a general weighted LASSO problem (21) with nonzero weights. Therefore, it does not suffer from the multiple local minimal issue, and its global minimizer can be efficiently solved using some convex-type software packages such as SeDuMi [25] and CVX [26].

Equivalently, (21) can also be written as a constraint form via Lagrange multipliers for appropriate η, i.e.,

$$\min \|\mathbf{W}^{(m-1)}\mathbf{x}\|_1 \quad \text{s.t. } \|\mathbf{y}-\mathbf{\Phi}_1\mathbf{x}\|_2 \le \eta \tag{23}$$

where η is a new regularization parameter related to h.

Theorem 1. *Assume a DC decomposition of the function S as $S=S_1-S_2$ with $S_1, S_2: \mathbb{R}^Q \to \mathbb{R}\cup\{\infty\}$ lower semi-continuous, proper convex functions such as $\mathrm{dom}\, S_1 \subset \mathrm{dom} S_2$. It holds that* (i) *the sequence $\{\mathbf{x}^{(m)}\}_{m\in\mathbb{N}}$ is well defined or equivalently $\mathrm{dom}\,\partial S_1 \subset \mathrm{dom}\,\partial S_2$.* (ii) *The objective value sequence $\{S_1(\mathbf{x}^{(m)})-S_2(\mathbf{x}^{(m)})\}_{m\in\mathbb{N}}$ is monotonically decreasing.* (iii) *If the minimum of S is finite and the sequence $\{\mathbf{x}^{(m)}\}_{m\in\mathbb{N}}$ is bounded, every limit point $\hat{\mathbf{x}}'$ is a critical point of S_1-S_2.*

Theorem 1 ensures a decrease of S from an iteration to the next. Consequently, a convergence to a stationary point of the objective function is guaranteed. Theorem 1 has been proved by some authors and we refer the reader to [22, Theorem 3.7] or [27, Theorem 1] for details.

Theorem 2. *Consider a noisy sparse signal representation system $\mathbf{y}=\mathbf{\Phi}_1\mathbf{x}_0+\mathbf{n}$. Let $M(\mathbf{\Phi}_1)$ and $\hat{x}^{(m)}$ be the mutual coherence of the overcomplete basis $\mathbf{\Phi}_1$ and the estimated result from (23) respectively, where $M=M(\mathbf{\Phi}_1)=\max_{i\ne j}|\mathrm{Re}\{\mathbf{G}(i,j)\}|$, $\mathbf{G}=\mathbf{\Phi}_1^{\mathbf{H}}\mathbf{\Phi}_1$. Assuming that the tuning parameter is selected properly and the noise is bounded by ε, i.e. $0<\tau<\min\{\mathbf{x}_0(\mathfrak{I}^c)\}$, $\|\mathbf{n}\|_2\le\varepsilon$, and $\|\mathbf{y}-\mathbf{\Phi}_1\mathbf{x}\|_2\le\eta$, $\eta\ge\varepsilon$. If the estimated result corresponds to the $(m-1)$th iteration of formulation* (23) *satisfies $\mathfrak{I}=\mathfrak{I}'$ (or $P(\mathfrak{I}=\mathfrak{I}')\to 1$), where $\mathfrak{I}$ is the support of the zeros in $\mathbf{x}_0$ with complement $\mathfrak{I}^c$, $\mathfrak{I}': \{q_1 | |\hat{\mathbf{x}}^{(m-1)}(q_1)| \le \tau, q_1\in\{1,2,\ldots,Q\}\}$. Then the estimation error of the mth iteration can be bounded by*

$$\|\hat{\mathbf{x}}^{(m)}-\mathbf{x}_0\|_2^2 \le \frac{\Delta^2}{1+M(1-K(1+\epsilon)^2)} \tag{24}$$

provided that $K<(1/M+1)/4$, where $K=\|\mathbf{x}_0\|_0$, $\Delta=\varepsilon+\eta$.

Proof. See Appendix.

Some remarks on Theorem 2: (i) As stated in [20], the truncated ℓ_1–function can lead to a good approximation of ℓ_0–function, and the corresponding estimator is selection consistent (i.e., $P(\mathfrak{I}=\mathfrak{I}'')\to 1$), provided that a good initial estimate and the adequate samples are given, where $\mathfrak{I}''$ is the support of the zeros in the final estimation result of (18). Typically, LASSO or weighted LASSO is utilized to supply the good initial estimate. That is, if the initial estimate is provide by LASSO or weighted LASSO, formulation (23) can be regarded as an asymptotic consistent estimator with a very small ϵ, then $P(\mathfrak{I}=\mathfrak{I}')\to 1$ holds with appropriate iteration m. In fact, if the initial value $\mathbf{x}^{(0)}$ is not provided by LASSO or weighted LASSO, but fixed to be $\mathbf{0}$, then the first iteration corresponding to (23) is converted into LASSO. This means that the global minimum and asymptotic selection consistency of (23) are still guaranteed without the need for an accurate initialization.

(ii) If ϵ is set to be 1 or the tuning parameter is not selected properly such that $\infty>\tau\ge\max\{\mathbf{x}_0\}$, formulation (23) will also be transformed into the general LASSO problem. In this case, the iterative approach is not necessary and cannot lead to an improved estimation performance. Particularly, if ϵ is set properly, while the tuning parameter satisfies $\min\{\mathbf{x}_0(\mathfrak{I}^c)\}\le\tau<\max\{\mathbf{x}_0\}$, then the estimation performance can also be improved in some degrees compared with general LASSO.

(iii) There will be zero error in finding the unique sparse representation using TLP and DC decomposition in the noiseless case with $\Delta=0$. That is, solving the optimization problem (23) solves the ℓ_0 problem (10). Moreover, since formulations (21) and (23) belong to convex optimization framework, even if the signal is noisy, a better performance compared with LASSO can also be assured with a small parameter ϵ. Empirically, $\epsilon=0.01$ is a good compromise.

(iv) We can decrease the noise level and the correlation coefficient M to achieve improved reconstruction performance. In practical applications, the most direct way is to increase the SNR and the number of rows of $\mathbf{\Phi}_1$.

Theorem 2 tells us that the combination of TLP and DC decomposition can lead to a good approximation of ℓ_0 function, and the proposed algorithm performs reasonably well in reconstructing performance.

3.2. Construction of a kind of special SOS column vector

From the above analysis, we can see that the TLP and DC decomposition algorithm is better suited for dealing with sparse vector model, and the reconstruction performance becomes better when the noise is eliminated efficiently. Thus, in this paper, second-order statistics (SOS) vector model is considered to reduce the computational complexity and to improve the performance accuracy of the desired parameter estimation. Observe the cross-correlation coefficient of the i_1th and i_2th sensor output in uniform linear array, which is just the (i_1,i_2)th element of the array covariance matrix $\mathbf{R}$, and given by

$$\begin{aligned} r_{i_1,i_2} &\triangleq E\{y_{i_1}(t)y_{i_2}^*(t)\} \\ &= \sum_{k=1}^{K} a_{i_1}(\theta_k)a_{i_2}^*(\theta_k)P_k+\sigma_{i_1}^2\delta(i_1-i_2) \\ &= \sum_{k=1}^{K} \exp(-j\pi(i_1-i_2)\sin(\theta_k))P_k+\sigma_{i_1}^2\delta(i_1-i_2) \end{aligned} \tag{25}$$

where $\delta(\cdot)$ is the Dirac delta function. We find that $r_{i_1,i_2}=r_{i_3,i_4}$ with $i_1-i_2=i_3-i_4$, $i_1,i_2,i_3,i_4\in[1,L]$. It means that we can provide an improved statistical property in some degrees by sum-average arithmetic, whose $\overline{m}$th $(1\le\overline{m}\le 2L-1)$ element is given by

$$\mathbf{y}_1(\overline{m}) = \begin{cases} \dfrac{1}{\overline{m}}\displaystyle\sum_{\bar{l}=1}^{\overline{m}} \mathbf{R}(\bar{l}, L+\bar{l}-\overline{m}) & \text{for } \overline{m}=1,\ldots,L \\ \dfrac{1}{2L-\overline{m}}\displaystyle\sum_{\bar{l}=1}^{2L-\overline{m}} \mathbf{R}\left(L+1-\bar{l}, 2L-\overline{m}+1-\bar{l}\right) & \text{else.} \end{cases} \tag{26}$$

The clear relationship between $\mathbf{R}$ and the column vector $\mathbf{y}_1$ is illustrated in Fig. 2, and the matrix form of

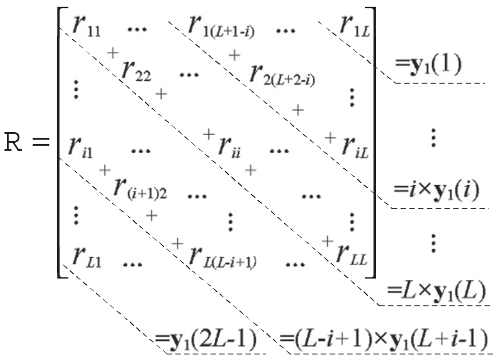

Fig. 2. Relationship between $\mathbf{R}$ and column vector $\mathbf{y}_1$.

$\mathbf{y}_1$ can be written as

$$\mathbf{y}_1 = \mathbf{B}\widehat{\mathbf{P}} + \overline{\sigma}^2 \mathbf{I}^{(L)} \tag{27}$$

where

$$\mathbf{B} = [\mathbf{b}(\theta_1), \ldots, \mathbf{b}(\theta_K)] \tag{28}$$

$$\mathbf{b}(\theta_k) = [e^{j\pi(L-1)\sin\theta_k}, \ldots, 1, \ldots, e^{-j\pi(L-1)\sin\theta_k}]^{\mathrm{T}} \tag{29}$$

$$\widehat{\mathbf{P}} = [P_1, \ldots, P_K]^{\mathrm{T}} \tag{30}$$

$$\overline{\sigma}^2 = \frac{1}{L}(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_L^2) \tag{31}$$

and $\mathbf{I}^{(L)}$ is a $(2L-1)\times 1$ vector, whose Lth element is one and the others are zeros.

Note that the formed vector data $\mathbf{y}_1$ has the similar form with the output vector of array given in (1). Therefore, it can be regarded as virtual array output in second-order statistics domain, which also contains sufficient information on DOA and power parameters. Moreover, it can be easily observed that the MMV problems composed of T snapshots in time-domain have been successfully transformed into VSMV problem in second-order statistics domain.

In sparse signal representation framework, (27) can be equivalently expressed as

$$\mathbf{y}_1 = \mathbf{\Phi}_2 \widehat{\mathbf{P}}_Q + \overline{\sigma}^2 \mathbf{I}^{(L)} \tag{32}$$

where $\mathbf{\Phi}_2 = [\mathbf{b}(\tilde{\theta}_1), \ldots, \mathbf{b}(\tilde{\theta}_Q)]$. $\widehat{\mathbf{P}}_Q = [\tilde{P}_1, \ldots, \tilde{P}_Q]^{\mathrm{T}}$ is a K-sparse vector, whose ith element is nonzero and equal to P_k if source k come from $\tilde{\theta}_i$ and zeros otherwise.

In the following, we will provide a modified algorithm for joint estimation of DOA and power parameters using this kind of vector model, designated as SRSSV-ℓ_0.

3.3. The proposed algorithm

The proposed algorithm is given in Table 1. Here, saa($\cdot$) denotes the sum-average arithmetic as described in (26). *Max* denotes the maximal number of iterations. Simulation results show that the termination condition, in general, is satisfied after one or two iterations, so it is acceptable for setting *Max*=3. $\widehat{\mathbf{P}}_Q^{(m)}$ denotes the value of $\widehat{\mathbf{P}}_Q$ updated at the mth iteration. $\hat{\sigma}^2$ is obtained by the average of the $L-K$ smallest eigenvalues of $\hat{\mathbf{R}}$. The initial estimate $\widehat{\mathbf{P}}_Q^{(0)}$ is provided by the weighted LASSO, and the weight $\hat{w}_q$ is given by

Table 1
SRSSV-ℓ_0 algorithm.

Initialization
1. $\hat{\mathbf{y}}_1 = \mathrm{saa}(\hat{\mathbf{R}})$, *Max*=3, $\varepsilon_1 = \epsilon = 0.01$, $m=1$.
2. $\widehat{\mathbf{P}}_Q^{(0)} = \arg\min \sum_{q=1}^{Q} \hat{w}_q |\tilde{P}_q|$ s.t. $\|\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2\widehat{\mathbf{P}}_Q - \hat{\sigma}^2\mathbf{I}^{(L)}\|_2 \le \eta_1$

Iteration
3. $\mathbf{W}^{(m-1)} = \mathrm{diag}\{\omega_1, \ldots, \omega_Q\}$,
$\omega_q = \begin{cases} 1/\tau, & |\tilde{P}_q^{(m-1)}| \le \tau \\ \epsilon/\tau & \text{otherwise.} \end{cases}$
4. $\widehat{\mathbf{P}}_Q^{(m)} = \arg\min \sum_{q=1}^{Q} \omega_q |\tilde{P}_q|$ s.t. $\|\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2\widehat{\mathbf{P}}_Q - \hat{\sigma}^2\mathbf{I}^{(L)}\|_2 \le \eta_1$

Termination
If $|\widehat{\mathbf{P}}_Q^{(m-1)} - \widehat{\mathbf{P}}_Q^{(m)}| \le \varepsilon_1$ or $m \ge$ *Max*.

$$\hat{w}_q = \mathbf{a}(\tilde{\theta}_q)^{\mathrm{H}} \mathbf{U}_n \mathbf{U}_n^{\mathrm{H}} \mathbf{a}(\tilde{\theta}_q) \tag{33}$$

where $\mathbf{U}_n$ denotes the $L\times(L-K)$ noise-subspace matrix of $\hat{\mathbf{R}}$, which is corresponding to the $(L-K)$ small singular values.

3.4. Parameter selection

The regularization parameter plays an important role in the final performance, where a large value may cause wrong source parameters estimation, while a small value may produce many spurious peaks [12,15]. Then an intelligent selection strategy is very necessary. As stated in [16,28], the vectorized form of the estimate error of array covariance matrix, i.e. $\Delta\mathbf{R} = \hat{\mathbf{R}} - \mathbf{R}$, satisfies

$$\mathrm{vec}(\Delta\mathbf{R}) \sim \mathrm{AsN}\left(\mathbf{0}, \frac{1}{T}\mathbf{R}^{\mathrm{T}} \otimes \mathbf{R}\right) \tag{34}$$

where $\mathrm{AsN}(\mu, \sigma^2)$ denotes the asymptotic normal distribution with mean μ and variance σ^2.

As demonstrated above, the second SOS vector $\mathbf{y}_1$ is formed by sum-average arithmetic using the data of array covariance matrix. Therefore, there exists a linear relationship between $\mathbf{y}_1$ and $\mathrm{vec}(\mathbf{R})$, i.e.,

$$\mathbf{y}_1 = \mathbf{\Gamma} \times \mathrm{vec}(\mathbf{R}) \tag{35}$$

where $\mathbf{\Gamma}$ is a $(2L-1)\times L^2$ linear transformation matrix, and $\mathrm{rank}(\mathbf{\Gamma}) = 2L-1$.

Since the linear combination of multivariate normal distribution is also normal distribution, we can approximately obtain

$$\left(\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2\widehat{\mathbf{P}}_Q - \overline{\sigma}^2\mathbf{I}^{(L)}\right) \sim \mathrm{AsN}\left(\mathbf{0}, \frac{1}{T}\mathbf{\Gamma}\left(\mathbf{R}^{\mathrm{T}} \otimes \mathbf{R}\right)\mathbf{\Gamma}^{\mathrm{T}}\right). \tag{36}$$

Further, (36) can be deduced that

$$\mathbf{\Gamma}_1 = \mathbf{D}^{-1/2}(\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2\widehat{\mathbf{P}}_Q - \overline{\sigma}^2\mathbf{I}^{(L)}) \sim \mathrm{AsN}(\mathbf{0}, \mathbf{I}_{2L-1}) \tag{37}$$

where $\mathbf{D}^{-1/2}$ is the Hermitian square root of the inverse of the asymptotic covariance matrix $(1/T)\mathbf{\Gamma}(\mathbf{R}^{\mathrm{T}} \otimes \mathbf{R})\mathbf{\Gamma}^{\mathrm{T}}$. (37) directly leads to

$$\|\mathbf{D}^{-1/2}(\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2\widehat{\mathbf{P}}_Q - \overline{\sigma}^2\mathbf{I}^{(L)})\|_2^2 \sim \mathrm{As}\chi^2(2L-1) \tag{38}$$

where $\mathrm{As}\chi^2(\mathrm{H})$ denotes the asymptotic chi-square distribution with H degrees of freedom. Combining the above derivation and analysis, we can get a statistically robust and more tractable formula for DOA and power estimation, which is given by

$$\widehat{\mathbf{P}}_Q^{(m)} = \arg\min \sum_{q=1}^{Q} \omega_q |\tilde{P}_q| \quad \text{s.t. } \|\mathbf{D}^{-1/2}(\hat{\mathbf{y}}_1 - \mathbf{\Phi}_2 \widehat{\mathbf{P}}_Q - \hat{\sigma}^2 \mathbf{I}^{(L)})\|_2 \le \eta_2 \tag{39}$$

where $m = 0, \ldots, Max$, η_2 is a new regularization parameter that can be determined properly by the chi-square distribution function with probability $1-p$ [its Matlab-based calculation is via the function chi2inv$(1-p,\ 2L-1)$]. In general, $p=0.01$ is satisfied.

The tuning parameter τ is another important parameter, where the inappropriate selection will result in some numerical instabilities. Simulation results suggest that $\hat{\tau} = 0.6\hat{P}_{\min}$ is a good choice for 0–20 dB SNR. While in lower SNR cases, the tuning parameter can be determined reasonably by a statistical technique called V-fold cross-validation (V-CV) [29,30]. The key point of V-CV is to divide the data $\hat{\mathbf{y}}_1$ into V roughly equal parts including training set and validation set. For each $v=1,\ldots,V$, fit the model with parameter τ to the other $V-1$ parts, giving $\widehat{\mathbf{P}}_Q(\tau)$ and compute its error in predicting the vth part,

$$E_v(\tau) = \|\hat{\mathbf{y}}_1^v - \mathbf{\Phi}_2^v \widehat{\mathbf{P}}_Q(\tau)\|_2^2 \tag{40}$$

where $\hat{\mathbf{y}}_1^v$ and $\mathbf{\Phi}_2^v$ denote the vth part of $\hat{\mathbf{y}}_1$ and $\mathbf{\Phi}_2$, respectively. Then, the cross-validation error is given by

$$CV(\tau) = \frac{1}{V} \sum_{v=1}^{V} E_v(\tau). \tag{41}$$

Repeat this operation for many values of τ and select the value of τ that makes $CV(\tau)$ smallest. Typically, we use 2-fold cross-validation or leave-one-out (i.e., L-fold) cross-validation in SRSSV-ℓ_0. Note that the application of cross-validation will bring considerable computational burden when the number of the predicted τ is large. For acceleration, in practical implementations, we first provide an empirical parameter $\hat{\tau}$ such as $\hat{\tau} = 0.6\hat{P}_{\min}$, and then successively select a more suitable parameter τ around $\hat{\tau}$.

3.5. Discussions

In this section, we discuss the proposed algorithm from three aspects, i.e., number of required array sensors, computational complexity, and ability to localize coherent sources.

3.5.1. Number of required array sensors

According to the unique representation theory [31], the vector sparse representation model, such as Eq. (32), admits a unique K-sparse solution only if $K <$ Spark$[\mathbf{\Phi}_2]/2$, where Spark$[\cdot]$ denotes the smallest integer of columns of $\mathbf{\Phi}_2$ that are linearly dependent. Since any set of $2L-1$ columns of $\mathbf{\Phi}_2$ is linearly independent, then Spark$[\mathbf{\Phi}_2] = 2L$. Therefore, as far as the number of sources to be processed is considered, the proposed algorithm can localize $L-1$ sources using a ULA with of L sensors.

3.5.2. Computational complexity

Regarding the computational complexity, we consider the major part. The calculation of $\hat{\mathbf{R}}$, its eigenvalue decomposition (EVD), and $\mathbf{D}^{-1/2}$ require $O(TL^2)+O(L^3)$. The construction of $\mathbf{\Gamma}_1$ via $\mathbf{D}^{-1/2}$ and $\hat{\mathbf{W}}$ requires $O(L^3)+O(L^2Q)$, where $\hat{\mathbf{W}} = [\hat{\omega}_1, \ldots, \hat{\omega}_Q]$ is the weight matrix formed by exploiting the orthogonality of the signal and noise subspace. Implementing one sparse reconstruction or one cross-validation requires $O(Q^3)$. Typically assuming $K < L \ll Q$, therefore, the computational cost of SRSSV-ℓ_0 is mainly in sparse signal reconstruction and cross-validation process, which is not influenced by the number of sources K. While for the ℓ_1-SVD and L_1-SRACV methods, the main computational cost is also the sparse signal reconstruction process, which require $O(K^3Q^3)$ and $O(L^3Q^3)$, respectively. Thus, the SRSSV-ℓ_0 will have lower complexity than ℓ_1-SVD and L_1-SRACV when source number K is large or dominant. Note that the computational complexity of the proposed algorithm is larger than that of the sparse spectral fitting method addressed in [18] (its main complexity is $O(Q^3)$) since the application of iterative approach and cross-validation, and is also larger than many subspace-based methods, such as MUSIC and ESPRIT. However, it is important to notice that the SRSSV-ℓ_0 can provide an improved performance (see the related experiment for details).

3.5.3. Ability to localize coherent sources

Some successful sparse signal reconstruction based methods, such as ℓ_1-SVD, JLZA-DOA, estimate the source parameters with the time-domain data of array output, which are not sensitive to the coherent sources. While the second-order statistics vectors formed in this paper are based on the assumption that the source is uncorrelated, when this assumption does not hold, Eq. (27) cannot be obtained. Therefore, the proposed SRSSV-ℓ_0 is not suitable for dealing with coherent or correlated sources.

4. Simulations

In this section, the performance of the SRSSV-ℓ_0 is investigated, and compared with those of MUSIC [4], MLE [6], ℓ_1-SVD [12], L_1-SRACV [16], reweighted ℓ_1-norm method (designated as REW-L_1 [13]), weighted ℓ_1-norm methods (designated as NSW-L_1 [14] and CSW-L_1 [15]), LASSO [18], and the Cramér–Rao lower bound (CRLB). Assume $P_1 = \cdots = P_K$. The input SNR of the kth signal is defined as $10\log(P_k/\sigma^2)$. An uniform linear array with half-wavelength element spacing is considered. By assumption, the number of sources K is known or correctly estimated by the Akaike information criterion (AIC) or the minimum description length (MDL) detection criterion [32]. For demonstrating the superiority of the SRSSV-ℓ_0 clearly and also for fair comparison, the REW-L_1 algorithm is implemented using our constructed vector data model and the regularization parameter related to LASSO is selected by chi-square distribution function with 0.99 confidence interval. The estimation precision is restricted by the resolution of the grid set, but too fine grid will result in large computation time. In the simulations, we first use a coarse gird in the range of -90° to 90° with 1° interval and then set a finer grid around the estimated angles to improve the estimation precision. Root mean square error (RMSE) of DOA estimation and the mean error (ME) of power estimation by 400 independent Monte Carlo trials are defined as

$$\mathrm{RMSE}=\sqrt{\frac{1}{400K}\sum_{p=1}^{400}\sum_{k=1}^{K}(\hat{\theta}_{k,p}-\theta_k)^2} \tag{42}$$

$$\mathrm{ME}=\frac{1}{400K}\sum_{p=1}^{400}\sum_{k=1}^{K}|\hat{P}_{k,p}-P_k| \tag{43}$$

where $\hat{\theta}_{k,p}$ and $\hat{P}_{k,p}$ are the estimate of DOA and power of the kth source in pth simulation.

In the first experiment, we show the capability of SRSSV-ℓ_0 to characterize the maximum number of sources. Four uncorrelated signals impinging from $[-30^\circ,-10^\circ,15^\circ,45^\circ]$ are considered. The number of sensors, snapshots and SNR are set to be 5, 500 and 5 dB, respectively. The simulation result is shown in Fig. 3, which confirms that the proposed SRSSV-ℓ_0 algorithm can estimate $L-1$ sources via an array of L sensors.

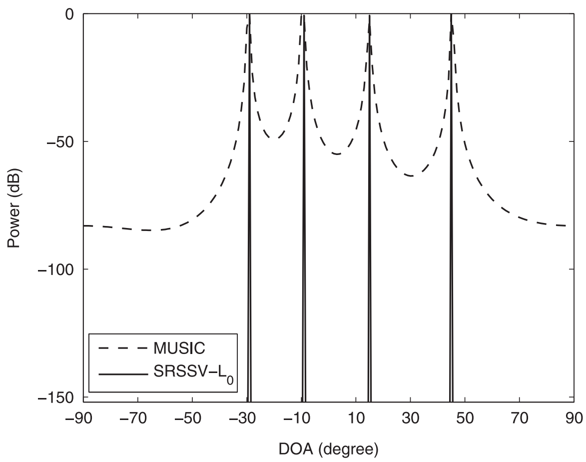

Fig. 3. Resolving $L-1$ sources: $L=5$ sensors, four sources, SNR=5 dB, $T=500$, DOAs=$[-30^\circ, -10^\circ, 15^\circ, 45^\circ]$.

In the second experiment, we compare the spectrum obtained by different algorithms using ten sensors, i.e. $L=10$. Two closely spaced sources are located at $[0^\circ, 5^\circ]$. The number of snapshots and the SNR are set to be 500 and 5 dB, respectively. As seen in Fig. 4(a), all the techniques are able to resolve the two sources. While in Fig. 4(b), we decrease the SNR to -5 dB, our approach is still able to resolve the two sources, whereas MUSIC merges the two peaks. This demonstrates the super-resolution ability of the proposed algorithm. Although other sparse-reconstruction-based methods are also able to resolve the two sources with -5 dB SNR, they detect the sources with higher estimation bias than the SRSSV-ℓ_0, and some of them even detect spurious peaks.

In the third experiment, we further examine the bias of the SRSSV-ℓ_0, NSW-L$_1$ and CSW-L$_1$ in terms of the angular separation between two sources. The first source is fixed at $\theta_1=-30^\circ$, while that of the second varies from $\theta_2=-26^\circ$ to $\theta_2=30^\circ$ in 1° steps. The number of snapshots is again $T=500$, and SNR=0 dB. As indicated in Fig. 5, the bias of the SRSSV-ℓ_0 is lower than that of NSW-L$_1$ and CSW-L$_1$ when the angle separation is less than 13°. Meanwhile, all the methods tend to become unbiased when sources are more than about 13° apart. In cases when the angular separation is less than 13°, one can decrease the bias with some appropriate manner, such as increase SNR or the number of sensors.

In the fourth experiment, we evaluate the RMSE of DOA estimates produced by SRSSV-ℓ_0 in different SNRs, different number of snapshots and different number of sensors. Two well-separated sources located at $\{\theta_1=-10.24^\circ, \theta_2=23.16^\circ\}$ are considered. The adaptive grid refinement strategy proposed in [12] is employed here to mitigate the effect of limiting estimates to a grid of spatial locations. The number of iterations in grid refinement is set to 4, where a 31-point locally uniform grid is distributed symmetrically around each spectral peak, with grid resolution $\Delta\theta=(0.3/4^{m_{ite}})^\circ$ at iteration m_{ite}, $m_{ite}\in\{1,2,3,4\}$. In Fig. 6, we vary the SNR from -20 dB to 10 dB in 2-dB steps with $T=500$, $L=10$. In Fig. 7, we vary the number of snapshots from 100 to 1000 in steps of 100 with

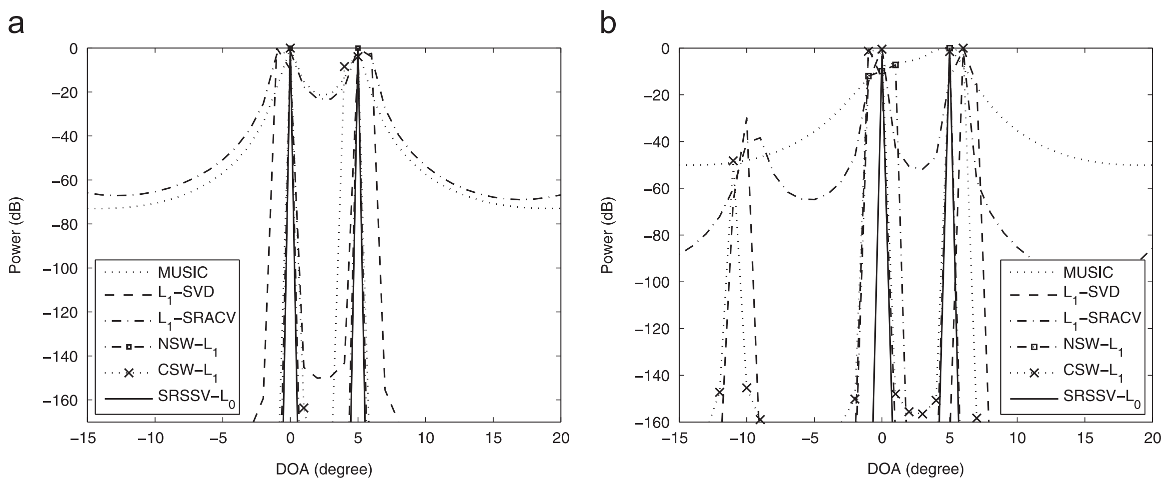

Fig. 4. Spatial spectrum obtained by different algorithms for two closely spaced sources. $L=10$, $T=500$, DOAs=$[0^\circ, 5^\circ]$. (a) SNR=5 dB. (b) SNR=-5 dB.

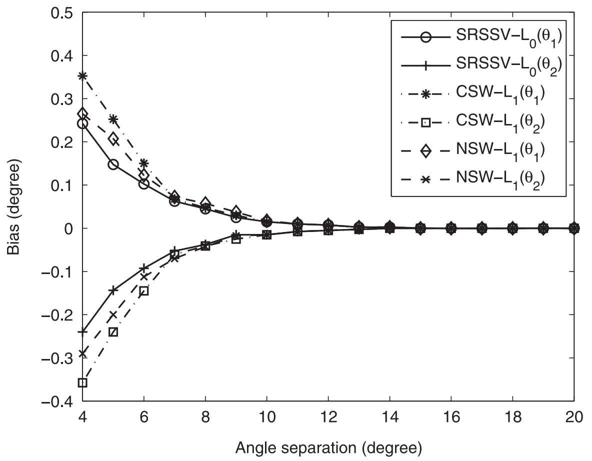

Fig. 5. Estimation bias of the SRSSV-ℓ_0, NSW-ℓ_1 and CSW-ℓ_1 versus angular separation, SNR=0 dB, T=500, L=10.

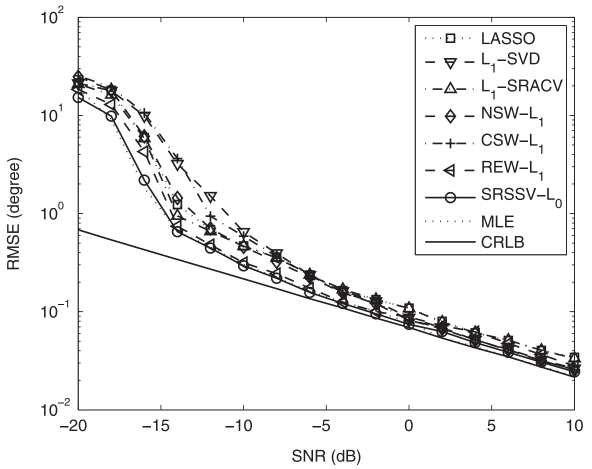

Fig. 6. RMSE of the DOA estimates for two uncorrelated sources versus SNR. T=500, L=10, $\theta_1=-10.24^\circ$, $\theta_2=23.16^\circ$.

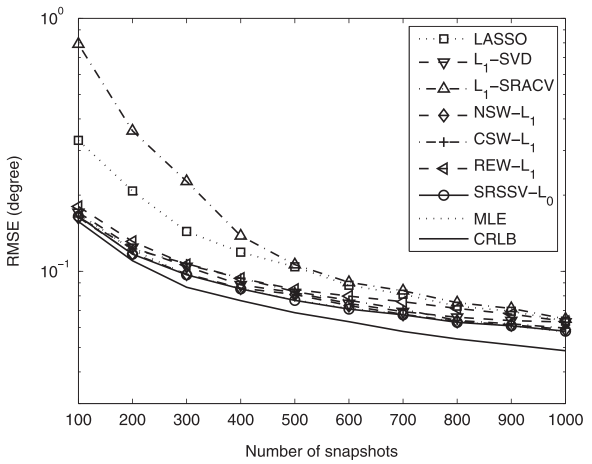

Fig. 7. RMSE of the DOA estimates for two uncorrelated sources versus the number of snapshots. SNR=0 dB, L=10, $\theta_1=-10.24^\circ$, $\theta_2=23.16^\circ$.

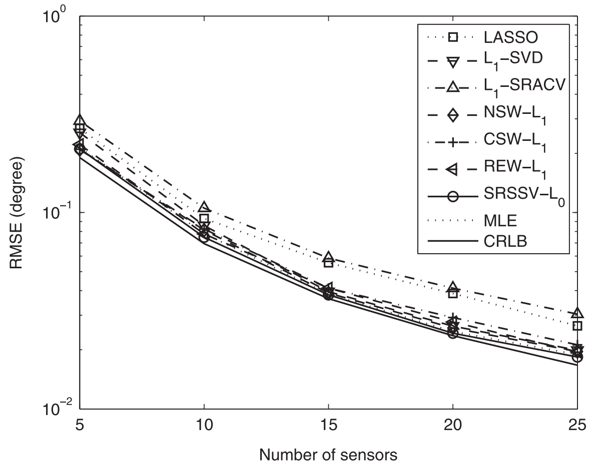

Fig. 8. RMSE of the DOA estimates for two uncorrelated sources versus the number of sensors. SNR=0 dB, T=500, $\theta_1=-10.24^\circ$, $\theta_2=23.16^\circ$.

SNR=0 dB, L=10. Whereas in Fig. 8, we fix the SNR and T to be 0 dB and 500 respectively, and vary the number of sensors from 5 to 25 in steps of 5. The nonconvex optimization in MLE is initialized to the true DOAs of the sources. Fig. 6 shows that our approach performs better than the other methods (especially in low SNR), and the threshold region occurs at lower SNR. Moreover, the proposed SRSSV-ℓ_0 algorithm is the closest one in performance to the intuitive bound provided by the MLE curve. Note that the general LASSO and L_1-SRACV are also based on a sparse representation of SOS vector, which are sensitive to small number of snapshots. However, owing to the sparse representation model, the recovery algorithm as well as sum-average arithmetic we adopt, the proposed SRSSV-ℓ_0 can reduce the sensitivity to small number of snapshots effectively, see Fig. 7. Fig. 8 shows that the RMSE of SRSSV-ℓ_0 decreases monotonically with the number of sensors, which validates that increasing the number of sensors is an efficient way to improve the estimation performance.

In the last experiment, we compare the ME of power estimation of the proposed SRSSV-ℓ_0, general LASSO and REW-L_1. Two uncorrelated signals located at $\{\theta_1=-10^\circ, \theta_2=30^\circ\}$ with $P_1=P_2=1$ W are considered. In Fig. 9, we vary the SNR from 0 dB to 20 dB in 2-dB steps with T=500, L=10. Whereas in Fig. 10, we fixed the SNR and T to be 0 dB and 500 respectively, and vary the number of sensors from 5 to 25 in 5 steps. As seen in Fig. 9, the power estimation performance of SRSSV-ℓ_0 outperforms the compared methods in the whole SNR region. Besides, it must be noted that the power estimation accuracy is satisfactory (ME is less than 0.01 W) with SNR > 4 dB. That is, the SRSSV-ℓ_0 can provide an almost unbiased power estimation. It is also seen from Fig. 10 that the performance of power estimation can also be improved efficiently by increasing the number of sensors.

5. Conclusion

In this paper, we have presented a new algorithm, namely SRSSV-ℓ_0 for DOA and power estimation in sparse

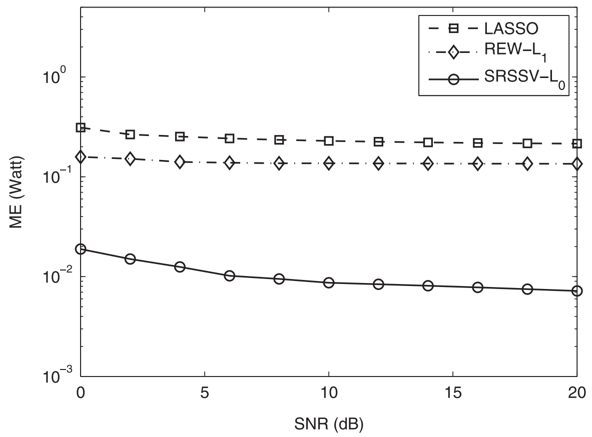

Fig. 9. ME of the power estimates for two uncorrelated sources versus SNR. $T=500$, $L=10$, $\theta_1=-10^\circ$, $\theta_2=30^\circ$, $P_1=P_2=1$ W.

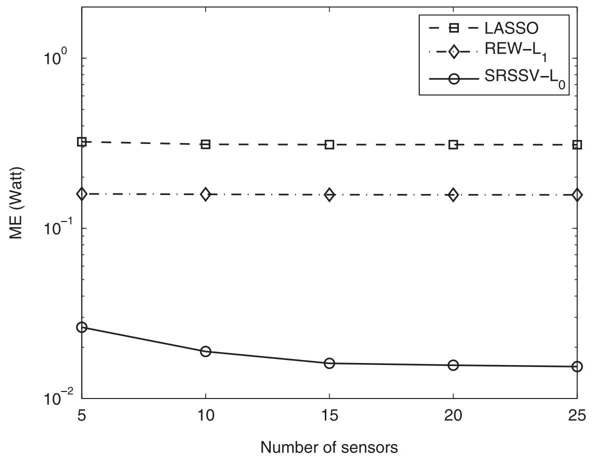

Fig. 10. ME of the power estimates for two uncorrelated sources versus the number of sensors. SNR=0 dB, $T=500$, $\theta_1=-10^\circ$, $\theta_2=30^\circ$, $P_1=P_2=1$ W.

signal representation framework. We transform a MMV or "group sparsity" problem to a VSMV problem using second-order statistics data and further exploit TLP and DC decomposition to effectively obtain the DOA and power parameters. We proved that the reconstruction algorithm we adopted can provide a reasonable good performance both by theoretical analysis and numerical simulations. We also provide an appropriate strategy on how to select regularization parameter and tuning parameter properly. The simulation results show that the proposed algorithm can resolve closely spaced sources, provide increased estimation accuracy, improved robustness to noise and almost unbiased power estimation.

Acknowledgment

The author would like to thank anonymous reviewers for their valuable comments and suggestions. This paper is supported by the National Nature Science Foundation of China (Grant no. 61171137) and New Century Excellent Talents in University (NCET-09-0426) in China.

Appendix A. Proof of Theorem 2

The estimation result $\hat{\mathbf{x}}^{(m)}$ at iteration m can be derived by the following optimization problem:

$$\left\{ \begin{array}{l} \hat{\mathbf{x}}^{(m)} = \arg\min_{\mathbf{x}} \|\mathbf{W}^{(m-1)}\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y}-\mathbf{\Phi}_1\mathbf{x}\|_2 \le \eta \\ \mathbf{y} = \mathbf{\Phi}_1\mathbf{x}_0 + \mathbf{n},\ \|\mathbf{n}\|_2 \le \varepsilon,\ \|\mathbf{x}_0\|_0 \le K. \end{array} \right\}. \tag{44}$$

Define error term $\vartheta = \hat{\mathbf{x}}^{(m)} - \mathbf{x}_0$, and similarly $\nu = \mathbf{x}-\mathbf{x}_0$, then we can rewrite the above problem as

$$\left\{ \begin{array}{l} \vartheta = \arg\min_{\nu} \|\mathbf{W}^{(m-1)}(\nu+\mathbf{x}_0)\|_1 \\ \text{s.t. } \|\mathbf{\Phi}_1\nu - \mathbf{n}\|_2 \le \eta,\ \ \|\mathbf{n}\|_2 \le \varepsilon,\ \|\mathbf{x}_0\|_0 \le K. \end{array} \right\}. \tag{45}$$

Note that if ϑ is the minimizer of $\|\mathbf{W}^{(m-1)}(\nu+\mathbf{x}_0)\|_1$ under the above constraints, then $\|\mathbf{W}^{(m-1)}(\vartheta+\mathbf{x}_0)\|_1 \le \|\mathbf{W}^{(m-1)}\mathbf{x}_0\|_1$, provided that $\eta \ge \varepsilon$ since otherwise $\nu = \mathbf{0}$ is not a feasible solution. Consequently, we consider

$$\left\{ \vartheta \left| \begin{array}{l} \|\mathbf{W}^{(m-1)}(\vartheta+\mathbf{x}_0)\|_1 \le \|\mathbf{W}^{(m-1)}\mathbf{x}_0\|_1, \\ \|\mathbf{\Phi}_1\vartheta - \mathbf{n}\|_2 \le \eta,\ \ \|\mathbf{n}\|_2 \le \varepsilon,\ \|\mathbf{x}_0\|_0 \le K. \end{array} \right. \right\}. \tag{46}$$

We next relax the constraint by eliminating the noise vector term $\mathbf{n}$

$$\{\vartheta \,|\, \|\mathbf{\Phi}_1\vartheta-\mathbf{n}\|_2 \le \eta \,\&\, \|\mathbf{n}\|_2 \le \varepsilon\} \subseteq \{\vartheta \,|\, \|\mathbf{\Phi}_1\vartheta\|_2 \le \eta+\varepsilon\}. \tag{47}$$

Define $\Delta = \eta+\varepsilon$, $\mathbf{G} = \mathbf{\Phi}_1^{\mathrm{H}}\mathbf{\Phi}_1$. The feasible set can be expanded as

$$\|\mathbf{\Phi}_1\vartheta\|_2^2 = \vartheta^{\mathrm{T}}\mathbf{G}\vartheta \le \Delta^2. \tag{48}$$

Successively, the above constraint can be relaxed

$$\begin{aligned} \Delta^2 &\ge \vartheta^{\mathrm{T}}\mathbf{G}\vartheta = \|\vartheta\|_2^2 + \vartheta^{\mathrm{T}}(\mathbf{G}-\mathbf{I}_Q)\vartheta \\ &= \|\vartheta\|_2^2 + \vartheta^{\mathrm{T}}(\mathrm{Re}\{\mathbf{G}\}-\mathbf{I}_Q)\vartheta \\ &\ge \|\vartheta\|_2^2 - |\vartheta|^{\mathrm{T}}|\mathrm{Re}\{\mathbf{G}\}-\mathbf{I}_Q||\vartheta| \\ &\ge \|\vartheta\|_2^2 - M|\vartheta|^{\mathrm{T}}|\mathbf{1}-\mathbf{I}_Q||\vartheta| \\ &= (1+M)\|\vartheta\|_2^2 - M\|\vartheta\|_1^2. \end{aligned} \tag{49}$$

where $\mathbf{1}$ is the $Q\times Q$ matrix of all ones, and $\mathbf{I}_Q$ is a $Q\times Q$ identity matrix.

As an equivalent form, the optimization problem (44) can be posed as

$$\left\{ \begin{array}{l} \mathbf{z}^{(m)} = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ s.t. } \|\mathbf{y}-\mathbf{\Phi}_1(\mathbf{W}^{(m-1)})^{-1}\mathbf{z}\| \le \eta \\ \mathbf{y} = \mathbf{\Phi}_1(\mathbf{W}^{(m-1)})^{-1}\mathbf{z}_0 + \mathbf{n},\ \|\mathbf{n}\|_2 \le \varepsilon,\ \|\mathbf{z}_0\|_0 \le K. \end{array} \right\} \tag{50}$$

with $\mathbf{x} = (\mathbf{W}^{(m-1)})^{-1}\mathbf{z}$, $\mathbf{x}_0 = (\mathbf{W}^{(m-1)})^{-1}\mathbf{z}_0$.

Since the uniqueness of ℓ_1-norm penalty is associated to the convex optimization, we must have

$$\|(\vartheta'+\mathbf{z}_0)\|_1 - \|\mathbf{z}_0\|_1 < 0 \tag{51}$$

unless $\vartheta'=0$, where $\vartheta' = (\mathbf{W}^{(m-1)})\vartheta$. Expanding (51), we have

$$\sum_{k_1\in\Im}|\vartheta'(k_1)| + \sum_{k_2\in\Im^c}(|\vartheta'(k_2)+\mathbf{z}_0(k_2)| - |\mathbf{z}_0(k_2)|) < 0 \tag{52}$$

where $\Im$ is the support of the zeros in $\mathbf{x}_0$ with complement $\Im^c$.

Note that $|\vartheta'(k_2)+\mathbf{z}_0(k_2)| - |\mathbf{z}_0(k_2)| \ge -|\vartheta'(k_2)|$, so the following relationship holds:

$$\sum_{k_1\in\Im}|\vartheta'(k_1)| < \sum_{k_2\in\Im^c}|\vartheta'(k_2)|. \tag{53}$$

Define $\mathfrak{I}': \{q_1 | |\hat{\mathbf{x}}^{(m-1)}(q_1)| \le \tau, q_1 \in \{1, 2, \ldots, Q\}\}$. If the estimated result corresponds to the $(m-1)$th iteration of formulation (23) satisfies $\mathfrak{I} = \mathfrak{I}'$ (or $P(\mathfrak{I} = \mathfrak{I}') \to 1$), the constraint (53) can be relaxed

$$\sum_{k_1 \in \mathfrak{I}} |\vartheta(k_1)| < \sum_{k_2 \in \mathfrak{I}^c} \epsilon |\vartheta(k_2)|. \tag{54}$$

Consequently, the estimation error can be bounded by

$$\left\{ \begin{array}{l} (1+M)\|\vartheta\|_2^2 - M\|\vartheta\|_1^2 \le \Delta^2 \\ \sum_{k_1 \in \mathfrak{I}} |\vartheta(k_1)| < \sum_{k_2 \in \mathfrak{I}^c} \epsilon |\vartheta(k_2)| \\ \#\mathfrak{I} \le K \end{array} \right\}. \tag{55}$$

The problem is invariant under permutations of the entries in $\mathbf{x}_0$, for the sake of simplicity, we assume that all nonzeros in $\mathbf{x}_0$ are concentrated in the initial slots of the vector, i.e., $\mathfrak{I} = \{1, \ldots, K\}$.

Let $\vartheta = [\vartheta_0; \vartheta_1]$, where ϑ_0 gives the first K entries in ϑ, and ϑ_1 the remaining $Q-K$ entries of ϑ. Obviously, we have

$$\|\vartheta\|_2^2 = \|\vartheta_0\|_2^2 + \|\vartheta_1\|_2^2, \quad \|\vartheta\|_1 = \|\vartheta_0\|_1 + \|\vartheta_1\|_1. \tag{56}$$

Using Cauchy inequality, we have

$$\|\vartheta_0\|_1 \ge \|\vartheta_0\|_2 \ge \frac{\|\vartheta_0\|_1}{\sqrt{K}}, \quad \|\vartheta_1\|_1 \ge \|\vartheta_1\|_2 \ge \frac{\|\vartheta_1\|_1}{\sqrt{Q-K}}. \tag{57}$$

We define $c_0 = (\|\vartheta_0\|_2 / \|\vartheta_0\|_1)^2$, $c_1 = (\|\vartheta_1\|_2 / \|\vartheta_1\|_1)^2$, $A = \|\vartheta_0\|_1$, $B = \|\vartheta_1\|_1$, $B = \rho A$, where $0 \le \rho < 1$. Then, (55) is given by

$$\left\{ \begin{array}{l} (1+M)(c_0 A^2 + c_1 B^2) - M(A+B)^2 \le \Delta^2 \\ 0 \le B < \epsilon A, \frac{1}{K} \le c_0 \le 1, 0 < c_1 \le 1 \end{array} \right\} \tag{58}$$

Simplifying (58) leads to

$$(1+M)\|\vartheta\|_2^2 - M\mu_1 \|\vartheta\|_2^2 \le \Delta^2 \tag{59}$$

where

$$\mu_1 = (1+\rho\epsilon)^2 / (c_0 + c_1 \rho^2 \epsilon^2) \le (1+\epsilon)^2 K. \tag{60}$$

Hence, the estimation error of the mth iteration is bounded by

$$\|\hat{\mathbf{x}}^{(m)} - \mathbf{x}_0\|_2^2 \le \frac{\Delta^2}{1 + M(1 - K(1+\epsilon)^2)}. \tag{61}$$

Note that the error bound of general LASSO is

$$\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2 \le \frac{\Delta^2}{1 + M(1-4K)}. \tag{62}$$

Therefore, if the initial estimate $\hat{x}^{(0)}$ is supplied by general LASSO, the number of nonzero coefficients K must satisfy $K < (/M+1)/4$.

This concludes the proof of Theorem 2.

References

[1] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, IEEE Signal Process. Mag. 13 (4) (1996) 67–94.

[2] B. Veen, K. Buckley, Beamforming: a versatile approach to spatial filtering, IEEE ASSP Mag. (1988) 4–24.

[3] J. Capon, High resolution frequency-wavenumber spectrum analysis, in: Proceedings of the IEEE, August 1969, pp. 1408–1418.

[4] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.

[5] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust., Speech, Signal Process. 37 (7) (1989) 984–995.

[6] P. Stoica, K. Sharman, Maximum likelihood methods for direction-of-arrival estimation, IEEE Trans. Acoust., Speech, Signal Process. 38 (1990) 1132–1143.

[7] D. Donoho, M. Elad, V. Temlyakvo, Stable recovery of sparse overcomplete representation in the presence of noise, IEEE Trans. Inf. Theory 52 (2006) 6–18.

[8] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (2005) 4203–4215.

[9] I. Gorodnitsky, B. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Trans. Signal Process. 45 (3) (1997) 600–616.

[10] J. Fuchs, On the use of the global matched filter for DOA estimation in the presence of correlated waveforms, in: Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2008, pp. 269–273.

[11] J. Fuchs, Identification of real sinusoids in noise, the global matched filter approach, in: Proceedings of the 15th IFAC Symposium on System Identification, 2009, pp. 1127–1132.

[12] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.

[13] E. Candès, B. Wakin, P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl. 14 (5–6) (2008) 877–905.

[14] C. Zheng, G. Li, H. Zhang, X. Wang, An approach of DOA estimation using noise subspace weighted ℓ_1 minimization, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May, 2011, pp. 2856–2859.

[15] X. Xu, X. Wei, Z. Ye, DOA estimation based on sparse signal recovery utilizing weighted ℓ_1−norm penalty, IEEE Signal Process. Lett. 19 (3) (2012) 155–158.

[16] J. Yin, T. Chen, Direction-of-arrival estimation using a sparse representation of array covariance vectors, IEEE Trans. Signal Process. 59 (9) (2011) 4489–4493.

[17] P. Stoica, B. Prabhu, L. Jian, SPICE: a sparse covariance-based estimation method for array processing, IEEE Trans. Signal Process. 59 (2) (2011) 629–638.

[18] J. Zheng, M. Kaveh, H. Tsuji, Sparse spectral fitting for direction of arrival and power estimation, in: Proceedings of the 15th IEEE Workshop on Statistics Signal Processing, September 2009, pp. 429–432.

[19] M. Hyder, K. Mahata, Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation, IEEE Trans. Signal Process. 58 (9) (2010) 4646–4655.

[20] X. Shen, W. Pan, Y. Zhu, Likelihood-based selection and sharp parameter estimation, J. Am. Stat. Assoc. 107 (497) (2012) 223–232.

[21] R. Horst, N. Thoai, Dc programming: overview, J. Optim. Theory Appl. 103 (1999) 1–41.

[22] P. Tao, L. An, Dc optimization algorithms for solving the trust region subproblem, SIAM J. Optim. 8 (2) (1998) 476–505.

[23] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc. 96 (456) (2001) 1348–1360.

[24] H. Wang, C. Leng, A note on adaptive group lasso, Comput. Stat. Data Anal. 52 (2008) 5277–5286.

[25] J. Sturm, Using SeDuMi 1.02: a MATLAB toolbox for optimization over symmetric cones, Optim. Methods Softw 11 (1–4) (1999) 625–653.

[26] M. Grant, S. Boyd. CVX: MATLAB Software for Disciplined Convex Programming, April 2010, Available Online at: ⟨http://cvxr.com/cvx⟩.

[27] G. Gasso, A. Rakotomamonjy, S. Canu, Recovering sparse signal with a certain family of nonconvex penalties and DC programming, IEEE Trans. Signal Process. 57 (12) (2009) 4686–4698.

[28] B. Ottersten, P. Stoica, R. Roy, Covariance matching estimation techniques for array signal processing applications, Digit. Signal Process. 8 (1998) 185–210.

[29] J. Shao, Linear model selection by cross-validation, J. Am. Stat. Assoc. 88 (422) (1993) 486–494.

[30] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Stat. Surv. 4 (2010) 40–79.

[31] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via $_1$ minimation, in: Proceedings of National Academy of Sciences of the United States of America, December, 2003, pp. 2197–2202.

[32] M. Wax, andT. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust., Speech, Signal Process. ASSP-33 (1985) 387–392.