



Communication

An Adaptive ℓ_p Norm Minimization Algorithm for Direction of Arrival Estimation

Lutao Liu and Zejing Rao *

The College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; liulutao@hrbeu.edu.cn

* Correspondence: raozejing@hrbeu.edu.cn; Tel.: +86-199-1762-8582

Abstract: In this paper, we propose a new direction of arrival (DOA) estimation algorithm, in which DOA estimation is achieved by finding the sparsest support set of multiple measurement vectors (MMV) in an over-complete dictionary. The proposed algorithm is based on ℓ_p norm minimization, which belongs to non-convex optimization. Therefore, the quasi-Newton method is used to converge the iterative process. There are two advantages of this algorithm: one is the higher possibility and resolution of distinguishing closely spaced sources, and the other is the adaptive regularization parameter adjustment. Moreover, an accelerating strategy is applied in the computation, and a weighted method of the proposed algorithm is also introduced to improve the accuracy. We conducted experiments to validate the effectiveness of the proposed algorithm. The performance was compared with several popular DOA estimation algorithms and the Cramer–Rao bound (CRB).

Keywords: DOA estimation; quasi-Newton method; adaptive regularization parameter



Citation: Liu, L.; Rao, Z. An Adaptive ℓ_p Norm Minimization Algorithm for Direction of Arrival Estimation. *Remote Sens.* **2022**, *14*, 766. <https://doi.org/10.3390/rs14030766>

Academic Editors: Konrad Jędrzejewski and Matthias Weiß

Received: 29 December 2021

Accepted: 4 February 2022

Published: 7 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With extensive applications, including radar, sonar, wireless communication and remote sensing, DOA estimation has traditionally been a popular branch in the field of array signal processing [1,2]. As a key technology in passive radar, DOA estimation is being extensively developed to locate targets in complex electromagnetic environments without being perceived.

Considering problems, such as distinguishing coherence sources and separating spatial closely sources, sparse reconstruction theory plays an important role in DOA estimation [3,4]. In dividing the whole spatial domain of interests into a discrete set of potential grids, a traditional estimation module can be converted into an ill-posed inverse problem. The specific physical model will be introduced in Section 2.

The MMV data module of a M sensor array takes the following form:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{N} \quad (1)$$

where $\mathbf{Y} \in \mathbb{C}^{M \times L}$ and $\mathbf{N} \in \mathbb{C}^{M \times L}$ are matrices of measurements and noise. The dictionary matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$ consists of N steering vectors. These vectors are determined by N different angles. In addition, $\mathbf{X} \in \mathbb{C}^{N \times L}$ is a K -row sparse matrix that represents a K sources spatial spectrum.

Note that the dictionary matrix \mathbf{A} is fat; hence, the equation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ has infinitely many exact solutions. To reconstruct the sparse signal, a simple and intuitive measure of sparsity is ℓ_0 -norm. The Lagrangian relaxation problem with the $\ell_{2,0}$ mixed norm is as follows:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_{2,0} \quad (2)$$

Note that the mixed-norm $\|\mathbf{X}\|_{2,0}$ means a sparse spatial domain but closely related time domain. That is to say, ℓ_2 norm gathers the snapshots of $\mathbf{X} \in \mathbb{C}^{N \times L}$ to obtain a spatial spectrum $\mathbf{X}_{vec} \in \mathbb{C}^{N \times 1}$, and ℓ_0 norm measures the sparsity of \mathbf{X}_{vec} . The general formula is:

$$\|\mathbf{X}\|_{p,q} = \sum_{i=1}^N \left(\sum_{j=1}^L |\mathbf{X}_{i,j}|^p \right)^{q/p} \quad (3)$$

Considering solving (2) directly is an NP-hard problem, alternative approaches are often studied. One alternative method is the greedy algorithm, OMP [5] pursued the current optimal value in every iteration. Another method is an approximate ℓ_0 norm method, such as SL0 [6].

A common convex approximation of the $\ell_{2,0}$ -pseudo-norm that is known to promote sparse solution is the $\ell_{2,1}$ -norm.

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_{2,1} \quad (4)$$

Convex optimization algorithms are always convergent to the global optimal solution. This feature makes convex optimization algorithms well-studied. Malioutov et al. proposed a heuristic approach, named ℓ_1 -SVD [7], using a reduction of the dimension of measurement matrix \mathbf{Y} by SVD and adaptive grid refinement. However, for a large number of snapshots L or large number of candidate frequencies K , the problem becomes computationally intractable.

Different from other sparse reconstruction fields, in the DOA estimation, atoms of dictionary \mathbf{A} are completely determined and fixed by the discrete spatial grid. Highly correlated atoms in the dictionary makes it difficult to separate adjusted vectors. This is far from the "approximate orthogonality" condition, which was introduced in the Restricted Isometry Property (RIP) [8]. However, ℓ_p norm optimization has a greater possibility of obtaining a sparse solution, which is usually used in imagery target detection [9]. Thereby, ℓ_p norm optimization also can be used to distinguish closely spaced sources.

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_{2,p} \quad (5)$$

Hence, in the $\ell_{2,p}$ -constrained problem we search for an N -row sparse solution that minimizes the fitting error. Algorithms, such as iterative reweighted least squares (IRLS) [10] and FOCUSS [11] are popular in DOA estimation. They both approximate the ℓ_p norm with a weighted ℓ_2 norm.

The goal of this paper is to find a practical sparse reconstruction method that fills the missing application scenario of traditional DOA estimation algorithm. The main contribution of the paper is a method for the automatic selection of the regularization parameter.

In contrast to other existing methods, the proposed method works completely in the iterative process, which means better adaptability to different situations. The key ingredient of the proposed method is the use of the quasi-Newton algorithm. Without losing accuracy, the iterative process is optimized, and the computation is reduced by a matrix inverse lemma, even with dense grids. In addition, we introduce the improvement of the weighted algorithm. Last but not least, the array does not have to be linear, and the sources can be coherent.

The rest of the paper is organized as follows. The processing of the proposed algorithm and innovative details are considered in Section 2. Potential application scenarios are discussed. Section 3 provides experimental results to compare the proposed algorithm with other DOA estimators. Our conclusions are contained in Section 4.

2. The Proposed Algorithm

The physical meaning of all parameters in the DOA estimation mathematical model will be explained here briefly. The array with all omnidirectional sensors receives plane

waves from the far field of the narrow-band signal source. All signals $s_k(k = 1, 2, \dots, K)$ are impinging with different angles $\theta_k(k = 1, 2, \dots, K)$. The phase difference between two adjacent sensors is defined as

$$\Delta\varphi_k = 2\pi \frac{d \sin \theta_k}{\lambda_s} \tag{6}$$

where λ_s denotes the wavelength. Let the sensor at the coordinate origin be the reference. The direction vector can be expressed as $\mathbf{a}(\theta_k) = [1, \exp(-j\Delta\varphi_k), \dots, \exp(-j(M-1)\Delta\varphi_k)]^T$, and thus the manifold matrix can be expressed as $\mathbf{A}_s = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]^T$.

Figure 1 shows the whole processing of DOA estimation. In order of the subplots, first, the signal space can be divided to N grids; therefore, the dictionary matrix \mathbf{A} can be expressed as $\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_N)]^T$. Then, the sensor data matrix $\mathbf{Y} \in \mathbb{C}^{M \times L}$ is the input data of the proposed algorithm. M denotes the number of sensors, and L denotes the number of samples. Finally, with a proper coefficient, linear combinations of atoms in the dictionary can reconstruct the signal. As long as the sparsest coefficient vector can be found, the direction of sources can be known from the index of the non-zero support set.

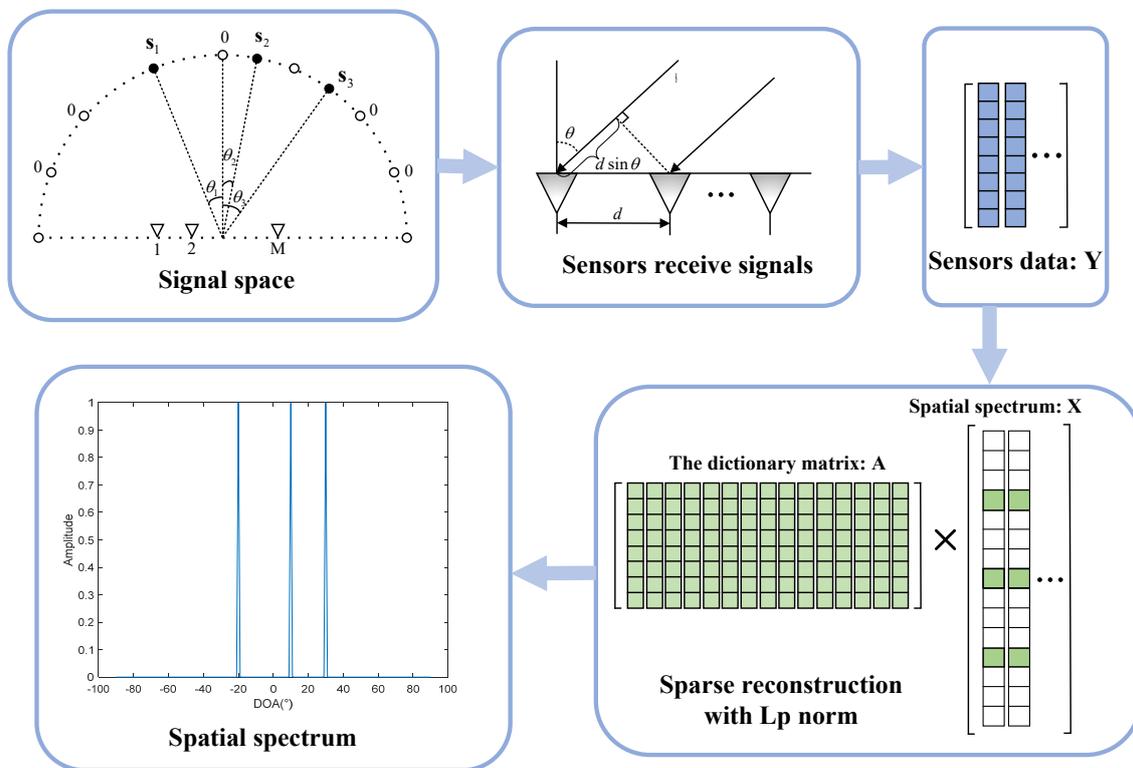


Figure 1. Schematic of the DOA estimation process.

2.1. The ℓ_p Norm Minimization Algorithm

Due to $\ell_p(0 < p < 1)$, cost function (5) is not differentiable at $\mathbf{0}$. The authors in [12] proved a differentiable approximation suiting our purpose well:

$$\|\mathbf{X}\|_{2,p} \approx \sum_{i=1}^N \left(\sum_{j=1}^L |x_{i,j}| + \epsilon \right)^{\frac{p}{2}} \tag{7}$$

where $\epsilon \geq 0$ is the smoothing parameter that controls the trade-off between smoothness and approximation. Usually the value of ϵ is set to 10^{-6} in the DOA estimation background.

If ϵ is too large, the approximation is not good, and if ϵ is too small, the inverse of (11) will be singular.

Substituting (7) for $\|\mathbf{X}\|_{2,p}$ in (5):

$$F(\mathbf{X}) \approx F_\epsilon(\mathbf{X}) = \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^N \left(\sum_{j=1}^L |X_{i,j}|^2 + \epsilon \right)^{\frac{p}{2}} \quad (8)$$

The Jacobian matrix of the objective function is

$$\mathbf{J} = (2\mathbf{A}^H\mathbf{A} + \lambda\mathbf{D}_1)\mathbf{X} - 2\mathbf{A}^H\mathbf{Y} \quad (9)$$

where the superscript $[\bullet]^H$ represents the conjugate transpose operation and

$$\mathbf{D}_1 = \text{diag}_{(i)} \left\{ \frac{p}{\left(\sum_{j=1}^L |X_{i,j}|^2 + \epsilon \right)^{1-\frac{p}{2}}} \right\} \quad (10)$$

The Hessian matrix of the objective function is

$$\mathbf{H} = 2\mathbf{A}^H\mathbf{A} + \lambda\mathbf{D}_1 + \lambda \text{diag}_{(i)} \left\{ \frac{p(p-2) \sum_{j=1}^L |X_{i,j}|^2}{\left(\sum_{j=1}^L |X_{i,j}|^2 + \epsilon \right)^{2-\frac{p}{2}}} \right\} \quad (11)$$

However, the Newton method is not convergent due to the non-positive definiteness of \mathbf{H} . A positive definite approximation of \mathbf{H} can be obtained by simply removing the negative part in (11).

$$\tilde{\mathbf{H}} = 2\mathbf{A}^H\mathbf{A} + \lambda(\mathbf{D}_1 + \mathbf{D}_2) \quad (12)$$

Note that

$$\mathbf{D}_2 = \text{diag}_{(i)} \left\{ \frac{p^2 \sum_{j=1}^L |X_{i,j}|^2}{\left(\sum_{j=1}^L |X_{i,j}|^2 + \epsilon \right)^{2-\frac{p}{2}}} \right\} \quad (13)$$

In DOA estimation, the dimension of $\tilde{\mathbf{H}}$ is affected by the number of grids, usually $N \geq M$, which causes heavy computation. Considering $\tilde{\mathbf{H}}$ consists of $\mathbf{A}^H\mathbf{A}$ and a diagonal matrix $\lambda(\mathbf{D}_1 + \mathbf{D}_2)$, the matrix inversion lemma is introduced to optimize the inversion

$$\tilde{\mathbf{H}}^{-1} = \left(\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}^H \left(\mathbf{I}_k + \mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^H \right)^{-1} \mathbf{A}\mathbf{Q}^{-1} \right) \quad (14)$$

Note that

$$\mathbf{Q} = \frac{\lambda}{2}(\mathbf{D}_1 + \mathbf{D}_2) \quad (15)$$

Applying (14) can reduce the dimension of the inverse matrix from $N \times N$ to $M \times M$; thus, the computational complexity decreases to $O(NM^2 + M^3 + NM)$ from $O(MN^2 + N^3 + NM)$. In order to facilitate the subsequent measurement of residuals and introduce (14) to reduce the computational burden, the Newton direction \mathbf{N}_d in the proposed algorithm can be expressed as:

$$\mathbf{N}_d = \tilde{\mathbf{H}}^{-1}\mathbf{J} = \mathbf{T}_4 - \mathbf{T}_3\mathbf{T}_4 - \mathbf{T}_2\mathbf{T}_1 + \mathbf{T}_3\mathbf{T}_2\mathbf{T}_1 \quad (16)$$

where $\mathbf{T}_{1\sim 4}$ are repeated parts in calculating and \mathbf{T}_1 will be used to obtain the noisy parameter in section B.

$$\begin{aligned}
 \mathbf{T}_1 &= \mathbf{Y} - \mathbf{A}\mathbf{X} \\
 \mathbf{T}_2 &= \mathbf{Q}^{-1}\mathbf{A}^H \\
 \mathbf{T}_3 &= \mathbf{T}_2(\mathbf{I}_k + \mathbf{A}\mathbf{T}_2)^{-1}\mathbf{A} \\
 \mathbf{T}_4(i) &= \frac{\sum_{j=1}^L |X_{i,j}|^2 + \epsilon}{(1+p)\sum_{j=1}^L |X_{i,j}|^2 + \epsilon}, i = 1 \dots N
 \end{aligned}
 \tag{17}$$

2.2. Adaptive Method for p

The ability to estimate the DOAs of two closely spaced sources is always an important indicator of the DOA estimation algorithm. Figure 2 is a norm ball of \mathbb{R}^2 space. The curves show that, the lower the p value, the easier it is to find a sparser solution. When faced with closely spaced sources, ℓ_p norm optimization has a greater possibility of obtaining a sparse solution, thereby, distinguishing two sources.

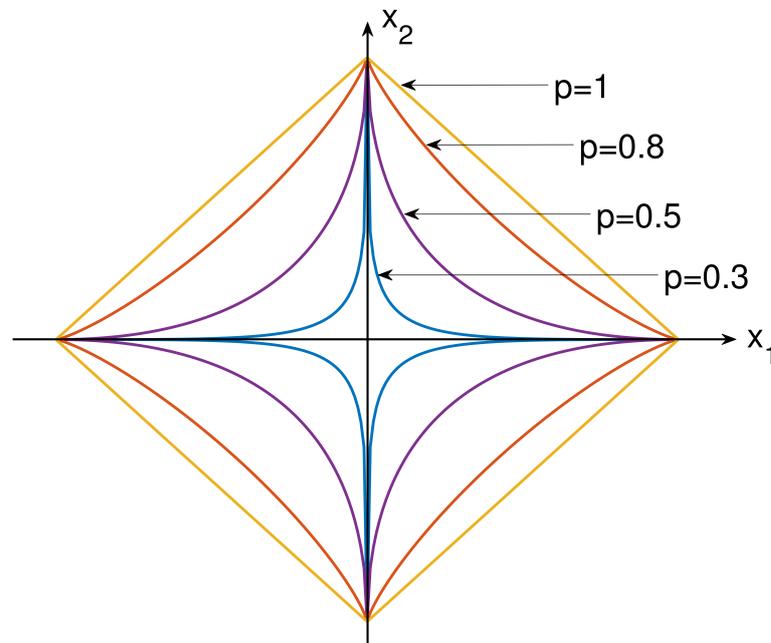


Figure 2. Unit norm ball.

As stated in the RIP condition, the uniqueness solution can be characterized in terms of the δ_K of matrix \mathbf{A} . Recall the smallest constants $0 < \delta_K \leq 1$ for which the matrix \mathbf{A} satisfies the RIP of order K , that is

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2
 \tag{18}$$

It is clear that any K -sparse vector \mathbf{x} is recovered via the ℓ_0 norm minimization if and only if the strict inequality $\delta_{2K} < 1$ holds. Define a new parameter to represent the RIP condition of matrix \mathbf{A}

$$\gamma_{2K} = \frac{(1 + \delta_{2K})^2}{(1 - \delta_{2K})^2} \geq 1
 \tag{19}$$

The quantity γ_{2K} can be made arbitrarily close to 1 by when the \mathbf{A} is orthogonal column by column. As stated in [12], given $0 < p \leq 1$, if (20) for some integer $t \geq K$, then every K -sparse vector is exactly recovered by solving the ℓ_p norm minimization.

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left(\frac{t}{K}\right)^{1/p-1/2}
 \tag{20}$$

When $t = K + 1$, condition (20) becomes

$$\gamma_{2K+2} < 1 + 4\left(\sqrt{2} - 1\right)\left(1 + \frac{1}{K}\right)^{1/p-1/2} \quad (21)$$

From (21), when $p \rightarrow 0$, the right side of the inequality becomes $+\infty$, which means the RIP condition number $\delta_{2K+2} \rightarrow 1$. That is to say, the smaller the value of p , the lower the 'approximate orthogonal' requirement of sparse reconstruction for matrix \mathbf{A} . From the perspective of DOA estimation, as long as p is small enough, the ability of the algorithm to overcome the high correlation between adjacent columns in over-complete basis \mathbf{A} is stronger, and the ability of separating spatial closely sources is stronger.

When $p \rightarrow 0$ the penalty of ℓ_p norm is close to ℓ_0 norm. However, too small p may lead the algorithm to converge to a local optimization. The common ℓ_p norm optimization algorithms, such as FOCUSS usually choose $p = 0.8$ based on experience [13]. In order to make p sensitive to the spectrum convergence and find a sparser solution, this section defines a decreasing processing of p by variation of the spectrum until $p < 0.1$.

$$p^{(k+1)} = p^{(k)} - v^{(k)} \quad (22)$$

where $v^{(k)} = \|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\|_F^2 / \|\mathbf{X}^{(k+1)}\|_F^2$. When $v^{(k)} > p^{(k)}$, proves that the spectrum variation of this iteration is too large, then we directly reduce the p value to half of the previous time. Thus, the stopping criteria of the proposed algorithm is $v^{(k)} < \delta$, and $\delta > 0$ is a small constant.

2.3. Adaptive Regularization Parameter Adjustment

According to (5), the regularization parameter λ has a significant influence on the optimization process. However, balancing the relationship between residuals and sparsity is a difficult problem. At present, the methods for selecting regularization parameters include Generalized Cross Validation (GCV) [14], L-curve [15] etc. However, the above methods usually require massive data for prior computation and are, thus, not suitable for real-time DOA estimation. Another computational algorithm L1-SRACV [16], which considers the covariance matrix distribution, can also avoid the regularization parameters.

This section will provide a practical procedure for adjusting λ in iteration. This method balances residuals and sparsity adaptively without massive prior simulation and working in the iteration process.

The covariance matrix of receiving data:

$$\mathbf{R} = \mathbf{Y}\mathbf{Y}^H / L \quad (23)$$

The ascending sorted eigenvalues $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_M]$ can be obtained by eigenvalue decomposition of R . Therefore, the expected value of the noise energy P_N is:

$$P_N = \frac{M \sum_{i=1}^{M-K} \gamma_i}{(M-K)L} \quad (24)$$

If source number K is unknown, there are two alternative methods. Under a low-SNR scenario, P_N can be expressed by the minimal eigenvalue $P_N = \frac{M\gamma_{min}}{L}$. Under a high-SNR scenario, the source number K can be roughly estimated by the slope mutation of γ .

In a convergent and stationary iterative process with accurate estimation results, the power of residual $P_R = \|\mathbf{T}_1\|_F^2 = \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2$ must converge to the noise power. Therefore, P_R can be the measurement indicator of proposed algorithm, which should be controlled within a certain range. Considering the calculation error of P_N and residual, the range of P_R is determined to be $[0.75P_N, 1.25P_N]$ after many simulation experiments.

The details are: in the k th iteration, if $P_R > 1.25P_N$, this indicates an overemphasis on the influence of sparsity on the optimal value—the model is poorly fitting—thus, we reduce

λ . If $P_R < 0.75P_N$, this indicates an overemphasis on the influence of noise—the model is over fitting—thus we enlarge λ . If $0.75P_N < P_R < 1.25P_N$, the descent direction works well, and nothing has to change. When the support set of the spectrum is determined, P_R will not have obvious fluctuations, which indicates the convergence of this computation.

Proper initialization can considerably reduce the running time of the quasi-Newton method, and it has a much lower possibility to become trapped in a local minimum. The product of the dictionary and receiving data can be the initialization

$$\mathbf{X}^{(1)} = \mathbf{A}^H \mathbf{Y} \quad (25)$$

The proposed Algorithm 1 is described as follows.

Algorithm 1 The improved ℓ_p norm DOA estimator with adaptive regularization parameter adjustment.

- 1: **Input:** \mathbf{A}, \mathbf{Y}
 - 2: **Initial:**
 $\mathbf{X}^{(1)} = \mathbf{A}^H \mathbf{Y}, p^{(1)} = 1, \epsilon = 10^{-6}, \delta = 10^{-6}, P_N, \lambda^{(1)} = 1$
 - 3: **repeat**
 - 4: Calculate $\mathbf{T}_{1 \sim 4}^{(k)}$, update
 $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \mathbf{N}_d^{(k)}$
 $\mathbf{N}_d^{(k)} = \left(\mathbf{T}_4^{(k)} - \mathbf{T}_3^{(k)} \mathbf{T}_4^{(k)} - \mathbf{T}_2^{(k)} \mathbf{T}_1^{(k)} + \mathbf{T}_3^{(k)} \mathbf{T}_2^{(k)} \mathbf{T}_1^{(k)} \right)$
 - 5: Calculate $v^{(k)} = \|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\|_F^2 / \|\mathbf{X}^{(k+1)}\|_F^2$
 - 6: **if** $P_R^{(k)} > 1.25P_N$, **then** $\lambda^{(k+1)} = 0.7\lambda^{(k)}$
else if $0.75P_N < P_R^{(k)} < 1.25P_N$, **then** $\lambda^{(k+1)} = \lambda^{(k)}$
else $\lambda^{(k+1)} = 1.5\lambda^{(k)}$
end
 - 7: **if** $p^{(k)} > 0.1$
if $v^{(k)} \geq p^{(k)}$, **then** $p^{(k+1)} = p^{(k)} / 2$, go to 4
else $p^{(k+1)} = p^{(k)} - v^{(k)}$, go to 4
end
else $p^{(k+1)} = p^{(k)}$, go to 4
end
 - 8: **until** $v^{(k)} < \delta$
 - 9: **Output:** reconstruct spectrum \mathbf{X}
-

2.4. Weighted Method

The weighted algorithm adds a weight in front of the spatial spectrum, which is a simple strategy to accelerate convergence and obtain higher resolution.

$$F(\mathbf{X}) \approx F_\epsilon(\mathbf{X}) = \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^N \left(w_i \sum_{j=1}^L |X_{i,j}|^2 + \epsilon \right)^{\frac{p}{2}} \quad (26)$$

Set a small weight in the place where the signal may occur, and set a large weight in the place where there is no signal so that the small value of the spatial spectrum is smaller, the large value is larger and the overall solution is more sparse. Normally, the spatial spectrum of CAPON algorithm is used as the weight.

$$w_i = \frac{1}{a(\theta_i)^H \mathbf{R}^{-1} a(\theta_i)}, i = 1, \dots, N \quad (27)$$

The weighted method can not only accelerate the iteration but also improve the accuracy of the algorithm. Although the weighted method will increase the computational complexity, it can improve the estimated success rate even in worse situations.

In Figure 3, the simulation is based on a eight-sensor uniform linear array with half-wavelength sensor spacing. The results show that the weighted method can distinguish two sources separated only by 3° with 200 snapshots and $\text{SNR} = 5$ dB.

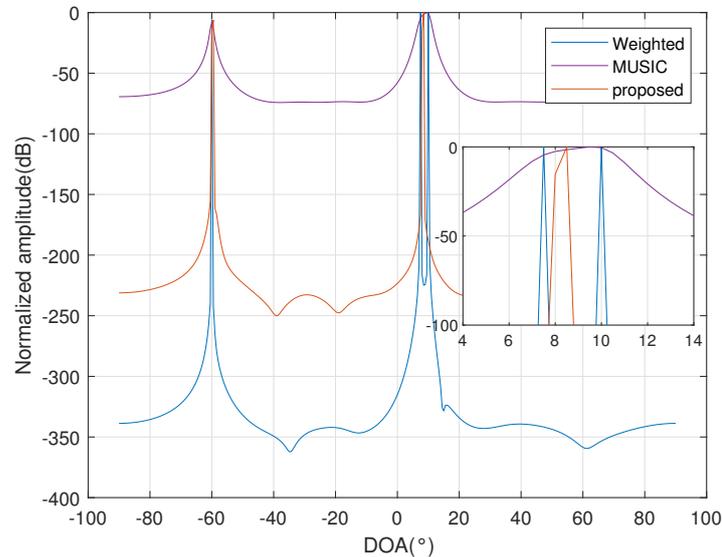


Figure 3. The spectrum of distinguishing closely spaced sources by the weighted method.

3. Experimental Results

In this section, the empirical performance of the proposed algorithm is investigated and compared with those of MUSIC, FOCUSS and L1-SVD (ℓ_1 norm optimization by CVX toolbox [17]). The simulations are based on a 12-sensor uniform linear array with half-wavelength sensor spacing. All methods here use a grid of 1801 points in the range of $[-90^\circ, 90^\circ]$, which means 0.1° .

Considering the stability and good performance of all algorithms, the regularization parameter λ of FOCUSS and L1-SVD was selected by the empirical value and $p = 0.8$ for the FOCUSS algorithm. For the proposed method, we used adaptive λ and p . The signal waveform was modeled as a complex circular Gaussian random variable $\mathbf{S}(t) \sim \mathcal{N}_C(\mathbf{0}_M, \mathbf{1}_M)$. The variance of Gaussian white noise $\mathbf{N}(t)$ is determined by the SNR. According to $\mathbf{X}(t) = \mathbf{A}_s \mathbf{S}(t) + \mathbf{N}(t)$, sensor data modeled as random variable $\mathbf{X}(t) \sim \mathcal{N}_C(\mathbf{0}_M, \mathbb{E}\{\mathbf{X}(t)\mathbf{X}^H(t)\})$.

First, the highlight of the proposed algorithm is adaptive adjustment of parameters in the iterative process. The proposed algorithm automatically adjusts λ according to the expectation power of residual P_R and adjusts p according to the variation of spatial spectrum as shown in Figure 4 with $\text{SNR} = 0$ dB. The top subplot shows the variation of λ , the middle subplot shows the variation of p , and the bottom subplot indicates the convergence procedure of the residual.

In Figure 4, we examine how the proposed method can adjust λ and p to provide a convergence result. From the beginning, the result is too close to the least squares solution, which usually has a small residual. In order to obtain a sparser solution, λ is increased to emphasize the importance of sparsity in iterations. For the whole process, the adaptive method maintains the residuals within expectations as much as possible until iteration convergence.

The adaptive methods can not only reduce the influence of the artificial selection of parameters on the results but also speed up the algorithm convergence to a certain extent. The convergence speed of the proposed algorithm was improved under the combined effect of acceleration strategy and adaptive methods. Figure 5 shows that the improved ℓ_p method will converge before 18 iterations, and the original method of ℓ_p and the FOCUSS algorithm requires nearly 24 iterations to converge.

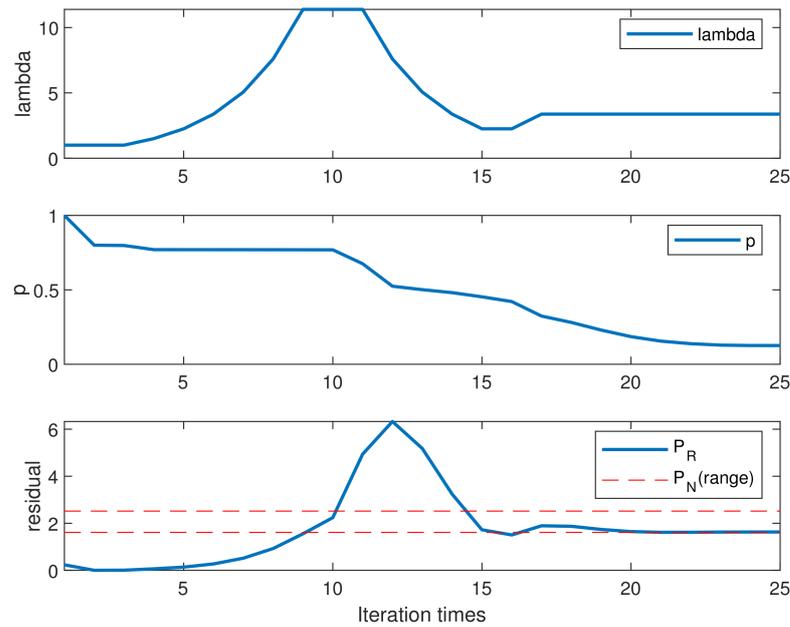


Figure 4. The change of the λ , p and residual in the iteration process.

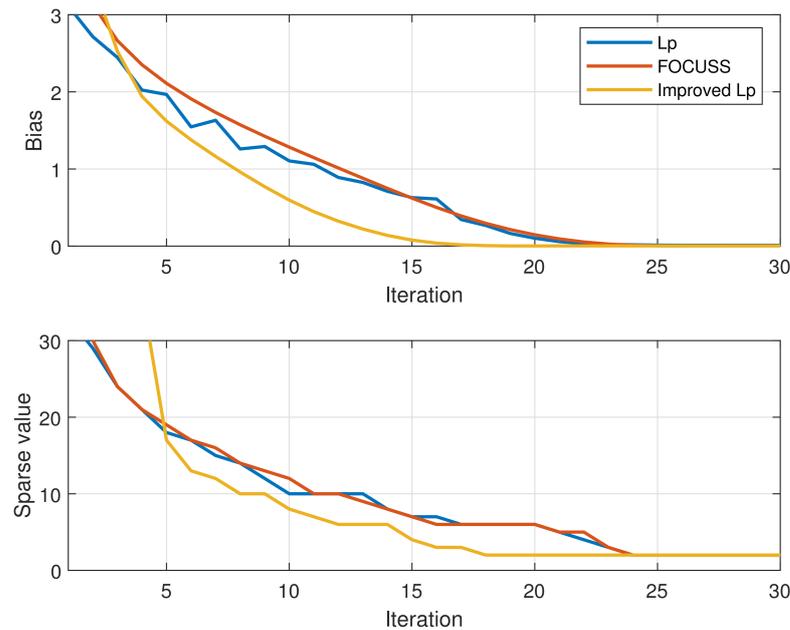


Figure 5. The change of the spectrum bias and sparsity in the iteration process.

Assume two equal-power uncorrelated signals arrive from 0° and $0^\circ + \Delta\theta$, respectively, where $\Delta\theta$ is varied from 0.1° to 4° . The SNR is set to 0 dB, and the number of snapshots is 200. The statistical result obtained via 200 multi-snapshot trials is shown in Figure 6 with the correct rate versus angular separation. As is clear from the figure, the proposed algorithm had a better ability to distinguish two closely spaced signals.

The bias curves in Figure 7 were obtained via 200 independent trials. Every multi-snapshots trial had a different DOA, θ_1 and θ_2 , drawn within the intervals $[-11^\circ, -9^\circ]$ and $[19^\circ, 21^\circ]$, and the snapshot number was 200. The root mean square error (RMSE) curves for different SNRs show that the resolution of the proposed algorithm meets the existing sparse reconstruction algorithm standards. Since the set grid size is 0.1° , when the SNR is high, the resolution of these algorithms is limited by the grid, and the curves will gradually become flat.

In conventional DOA estimation, this degree of error can generally be regarded as a correct estimation. The RMSE of the algorithm does not fully reflect the ability of DOA estimation. Sometimes, it is more meaningful to distinguish two sources in a certain spatial range with low resolution than to obtain only one estimation result with high resolution, which means that the latter option loses a target.

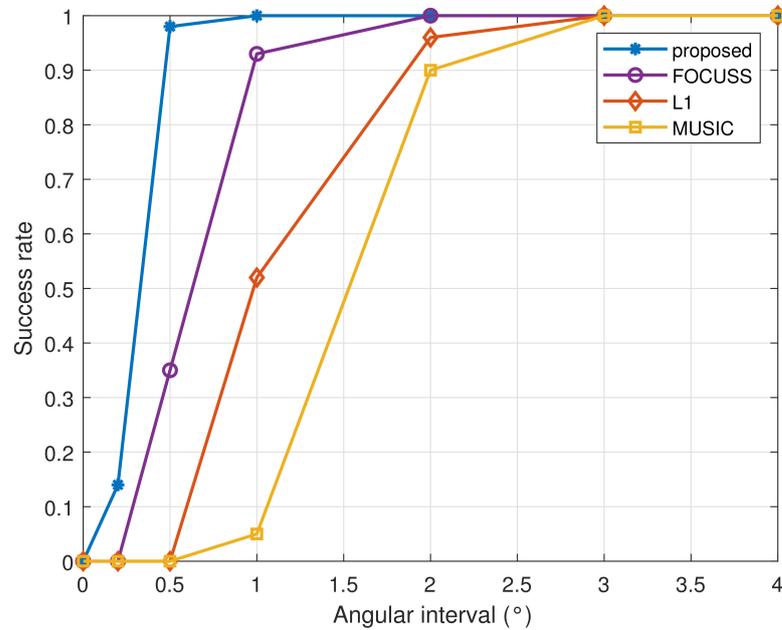


Figure 6. The probability of separating two closely spaced sources as a function of the angular interval.

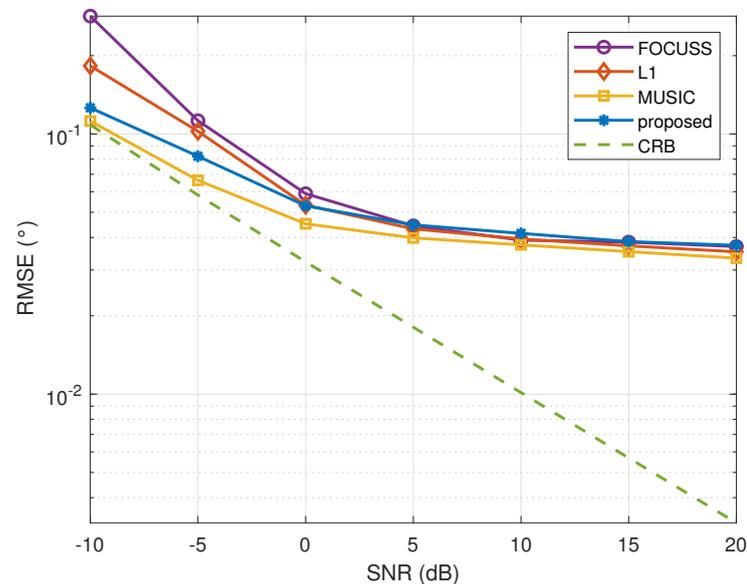


Figure 7. The RMSE of the DOA estimates as a function of the SNR.

This section will close by examining the impact of the size of the data snapshot on the error-suppression criterion of the adaptive method. Consider two uncorrelated signals arriving from $[-11^\circ, -9^\circ]$ and $[19^\circ, 21^\circ]$, where the SNR is fixed at 0 dB, the number of data snapshots is varied from 2 to 1000, and the remaining conditions are identical to those above. Figure 8 shows that, compared to the other methods, the proposed algorithm is stable in small and large snapshots.

As the focus of this paper is to improve the ability of the algorithm to distinguish closely spaced sources, the proposed algorithm and the conventional sparse reconstruction algorithms are essentially maintained in terms of the RMSE.

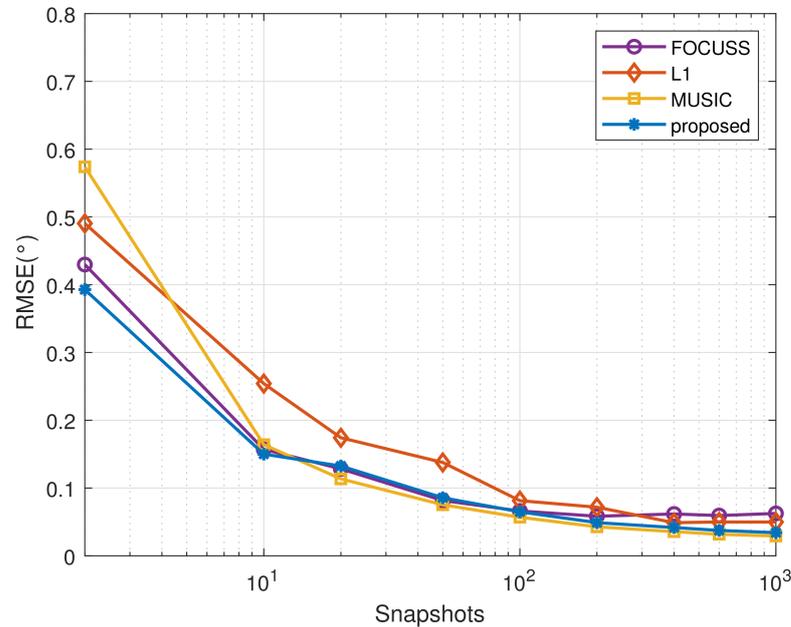


Figure 8. The RMSE of the DOA estimates as a function of the snapshots.

4. Conclusions

In this paper, an improved ℓ_p norm minimization DOA estimation algorithm was proposed. The proposed algorithm has two advantages: one is the greater ability to separate two closely spaced sources, and the other is the adaptive regularization parameter adjustment. The latter can reduce the impact of manual selection of λ on the results and can accelerate convergence. The main motivation of this work comes from the desire to increase the practicability of sparse reconstruction algorithms.

By applying ℓ_p norm minimization with the quasi-Newton method and acceleration strategies, the proposed algorithm was faster than L1-SVD, the convex optimization algorithm. In addition, the adaptive method of the regularization parameter can run stably in most non-extreme situations. Compared to other seriously time-consuming methods or by optimizing the inequality-constrained minimization problem directly, the proposed method is more practical.

Author Contributions: This work described in this article is the collaborative development of all authors. L.L. and Z.R. contributed to the idea of result processing and designed the experiment. Z.R. completed the first draft. L.L. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by “Aviation Foundation”, grant number 201901012005.

Acknowledgments: The authors would like to thank the support of the laboratory and university.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krim, H.; Viberg, M. Two decades of array signal processing research: The parametric approach. *IEEE Signal Process. Mag.* **1996**, *13*, 67–94. doi:10.1109/79.526899. [CrossRef]
2. Del-Rey-Maestre, N.; Mata-Moya, D.; Jarabo-Amores, M.P.; Gómez-del Hoyo, P.J.; Bárcena-Humanes, J.L.; Rosado-Sanz, J. Passive Radar Array Processing with Non-Uniform Linear Arrays for Ground Target’s Detection and Localization. *Remote Sens.* **2017**, *9*, 756. doi:10.3390/rs9070756. [CrossRef]

3. Yang, Z.; Li, J.; Stoica, P.; Xie, L. Chapter 11—Sparse methods for direction-of-arrival estimation. In *Academic Press Library in Signal Processing, Volume 7*; Chellappa, R., Theodoridis, S., Eds.; Academic Press: New York, NY, USA, 2018; pp. 509–581. doi:10.1016/B978-0-12-811887-0.00011-0. [[CrossRef](#)]
4. Paik, J.W.; Hong, W.; Lee, J.H. Direction-of-Departure and Direction-of-Arrival Estimation Algorithm Based on Compressive Sensing: Data Fitting. *Remote Sens.* **2020**, *12*, 2773. doi:10.3390/rs12172773. [[CrossRef](#)]
5. Tropp, J.A.; Gilbert, A.C. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. doi:10.1109/TIT.2007.909108. [[CrossRef](#)]
6. Shah, J.A.; Haider, H.; Kadir, K.A.; Khan, S. Sparse signal reconstruction of compressively sampled signals using smoothed ℓ_0 -norm. In Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, 12–14 September 2017; pp. 61–65. doi:10.1109/ICSIPA.2017.8120580. [[CrossRef](#)]
7. Malioutov, D.; Cetin, M.; Willsky, A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **2005**, *53*, 3010–3022. doi:10.1109/TSP.2005.850882. [[CrossRef](#)]
8. Donoho, D. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. doi:10.1109/TIT.2006.871582. [[CrossRef](#)]
9. Zhao, X.; Li, W.; Zhang, M.; Tao, R.; Ma, P. Adaptive Iterated Shrinkage Thresholding-Based Lp-Norm Sparse Representation for Hyperspectral Imagery Target Detection. *Remote Sens.* **2020**, *12*, 3991. doi:10.3390/rs12233991. [[CrossRef](#)]
10. Chartrand, R.; Yin, W. Iteratively reweighted algorithms for compressive sensing. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 3869–3872. doi:10.1109/ICASSP.2008.4518498. [[CrossRef](#)]
11. Gorodnitsky, I.; Rao, B. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* **1997**, *45*, 600–616. doi:10.1109/78.558475. [[CrossRef](#)]
12. Foucart, S.; Lai, M.J. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 395–407. doi:10.1016/j.acha.2008.09.001. [[CrossRef](#)]
13. Cotter, S.; Rao, B.; Engan, K.; Kreutz-Delgado, K. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* **2005**, *53*, 2477–2488. doi:10.1109/TSP.2005.849172. [[CrossRef](#)]
14. Golub Gene, M.H.; Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **1979**, *21*, 215–223. [[CrossRef](#)]
15. Rashno, A.; Tabataba, F.S.; Sadri, S. Regularization convex optimization method with l-curve estimation in image restoration. In Proceedings of the 2014 4th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 29–30 October 2014; pp. 221–226. doi:10.1109/ICCKE.2014.6993358. [[CrossRef](#)]
16. Yin, J.; Chen, T. Direction-of-Arrival Estimation Using a Sparse Representation of Array Covariance Vectors. *IEEE Trans. Signal Process.* **2011**, *59*, 4489–4493. doi:10.1109/TSP.2011.2158425. [[CrossRef](#)]
17. Grant, M.; Boyd, S. CVX: MATLAB Software for Disciplined Convex Programming, Version 2.0 Beta. Available online: <http://cvxr.com/cvx> (accessed on 25 September 2013).