HiLL PUBLISHING

Journal of Applied Mathematics and Computation, 2022, 6(3), 347-355
https://www.hillpublisher.com/journals/JAMC/
ISSN Online: 2576-0653
ISSN Print: 2576-0645

Multi-Source Localization and Signal Extraction Using a Proximal Gradient-based Compressed Sensing Approach

Chun-Shian Tao[1], Yu-An Chen[1], Yi-Cheng Hsu[1,*], Mingsian R. Bai[1,2]

[1]Department of Power Mechanical Engineering, National Tsing Hua University, Hsinchu, Taiwan, China.
[2]Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, China.

How to cite this paper: Chun-Shian Tao, Yu-An Chen, Yi-Cheng Hsu, Mingsian R. Bai. (2022) Multi-Source Localization and Signal Extraction Using a Proximal Gradient-based Compressed Sensing Approach. *Journal of Applied Mathematics and Computation*, 6(3), 347-355.
DOI: 10.26855/jamc.2022.09.007

Received: July 10, 2022
Accepted: August 8, 2022
Published: September 15, 2022

***Corresponding author:** Yi-Cheng Hsu, Department of Power Mechanical Engineering, National Tsing Hua University, Hsinchu, Taiwan, China.
Email: shane.ychsu@gmail.com

Abstract

This paper presents a computationally efficient algorithm for multiple source localization and signal extraction (SLSE). Posed as an underdetermined system, a novel compressed sensing (CS) algorithm is proposed to address SLSE problems in one stage. A Least Absolute and Selection Operator (LASSO) problem is first formulated and solved jointly for the source locations and signal amplitudes. A computationally efficient and noise-resilient algorithm is developed on the basis of the complex Proximal Gradient (Proxgrad) method. It follows that the nonzero entries of the optimal solutions give rise to the amplitudes and directions of sound sources. To further enhance the separation quality, soft thresholds based on W-disjoint orthogonality is exploited. Experiments are conducted to compare the proposed SLSE method with several baselines in terms of localization and separation metrics. The results showed that the proposed LASSO-Proxgrad algorithm yielded superior localization and signal extraction performance with the minimal processing time compared to the baselines.

Keywords

Source Localization, Signal Extraction, Compressed Sensing

1. Introduction

Source localization and signal extraction (SLSE) tasks have received much research interest in the field of acoustic signal processing. SLSE has found many applications in modern-day life, e.g., voice assistants [1], smart homes [2], video conferencing [3], acoustic scene analysis [4] and event detection [5], etc. SLSE methods fall into two categories: the one-stage methods [6], [7] and the two-stage methods [8], [9], [10]. One-stage methods conduct source localization and signal extraction in one step, whereas two-stage methods localize the sources first and separate the signals emitted from the sources in a subsequent step based on the information gained in the first stage. This paper focuses on the one-stage methods.

The one-stage method is based on the notion of Compressed Sensing (CS), also known as Sparse Coding [6], [7]. Although CS can be applied to many problems, relatively few application examples in acoustic array signal processing problems are found in literature. Bai and Chen applied the CS technique to various noise source identification problems by using nearfield and farfield arrays [11]. Gerstoft et al. proposed a compressive beamforming technique for multiple and single data snapshots [12]. As a key step for CS approaches, a dictionary needs to be constructed from a set of pre-selected steering vectors. In practice, a sufficiently fine mesh is required to prevent the solution from basis mismatch when the preselected dictionary fails to include the true source directions [13]. This usually results in an underdetermined system of

linear equations, where more source directions are pre-selected than the microphone elements. Algorithms such as relaxation methods and greedy methods have been suggested in the past to solve the CS problems [14]. Relaxation methods attempt to approximate the l_0-constraint function by using a convex optimization problem with l_1-penalty.Alternatively, greedy methods such as Orthogonal Matching Pursuit (OMP) [15] applies thresholds in solving CS problems with relatively low computational cost. In relaxation methods, the CS problems can be formulated in the least absolute shrinkage and selection operator (LASSO) form for which numerous optimization algorithms can be used, e.g., LASSO-Coordinate Descent (LASSO-CD), LASSO-Steepest Descent (LASSO-SD) [14], [16], etc.

The contributions of this paper are as follows. We approach the SLSE problem in the CS paradigm. We reformulate the CS problem in the LASSO form and develop a computationally efficient and noise-resilient algorithm on the basis of a complex Proximal Gradient (Proxgrad) method. A LASSO cost function comprised of a convex and differentiable least-square residual error term and a convex but non-differentiable l_1-penalty [17] is employed in the derivation. The proximal operator [18] can be used to map the convex non-differentiable term to a complex-valued soft-threshold operator. This leads to a gradient-descent iteration procedure. Furthermore, soft thresholding based on W-disjoint orthogonality [19] can be applied for improved separation quality.

The remainder of the paper is structured as follows. In Section II, the proposed LASSO-Proxgrad algorithm and the W-disjoint orthogonality-based postfilter are presented. Sections III shows the experimental results, followed by the conclusions in Section IV.

2. The Proposed SLSE Algorithm

In the following, we describe a one-stage method that accomplishes localization and signal extraction of sources jointly in one single step.

2.1 The CS Approach

By assuming the time-harmonic dependence, $e^{j\omega t}$, where $j=\sqrt{-1}$, ω denotes the angular frequency, and t is the continuous-time variable, the multi-source array signal model is formulated in the frequency domain as

$$\mathbf{p}(\omega) = \mathbf{A}_1(\omega)\mathbf{s}(\omega) + \mathbf{n}(\omega), \tag{1}$$

where $\mathbf{p}(\omega) \in C^M$ is the Fourier-transformed sound pressure vector received at M microphones. By assuming N candidate target directions, the steering matrix $\mathbf{A}_1(\omega) \in \mathbb{C}^{M\times N}$, $M \ll N$, serves as the "dictionary" for the CS model with its columns as the "atoms" designated according the pre-specified angular grid. As a fundamental assumption, the atoms associated with the candidate source directions are spatially independent. Typically, (1) is an underdetermined system of equations ($M \ll N$) for such a CS problem. The vector $\mathbf{s}(\omega) \in \mathbb{C}^N$ contains the Fourier transform of the candidate source signals and $\mathbf{n}(\omega) \in C^N$ denotes the additive noise vector.

In the far-field, where the plane-wave model can be applied, the steering matrix $\mathbf{A}_1(\omega) = [\mathbf{a}_0\ \mathbf{a}_1 \cdots \mathbf{a}_{N-1}] \in \mathbb{C}^{M\times N}$ consists of free-field steering vectors, $\mathbf{a}_n = [\mathrm{e}^{-j\mathbf{k}_n\cdot\mathbf{r}_1}\ \mathrm{e}^{-j\mathbf{k}_n\cdot\mathbf{r}_2}\ \cdots\ \mathrm{e}^{-j\mathbf{k}_n\cdot\mathbf{r}_M}]^T$, $n=0,\ldots,N-1$, where "$\cdot$" denotes the inner product, $\mathbf{r}_m, m=1,\ldots,M$, is the position vector of the mth microphone, $\mathbf{k}_n = -k\boldsymbol{\kappa}_n = -(\omega/c)\boldsymbol{\kappa}_n$ is the wave vector, $\boldsymbol{\kappa}_n = [\cos\theta_n\cos\phi_n\ \ \sin\theta_n\cos\phi_n\ \ \sin\phi_n]^T$ denotes the direction of arrival (DOA) vector that is a unit vector pointing to the look direction at the azimuth angle θ_n and the elevation angle ϕ_n associated with the nth candidate source direction, k denotes the wave number, and c is the speed of sound. In general, the number of physical sources (D), i.e., the nonzero entries of $\mathbf{s}(\omega)$, is smaller than the number of microphones and much smaller than the dictionary size, i.e., $D < M \ll N$. In order to recover the D-sparse solution reliably, the Restricted Isometry Property (RIP) and the incoherence property must be fulfilled [20]. These two properties can be regarded analogously as the matrix condition number that dictates the sensitivity of the errors of the sparse solution with respect to the errors of the observed data.

By assuming that the sources are spatially isolated, the constrained CS optimization problem can be posed as [14]

$$\min_{\mathbf{s}(\omega)\in\mathbb{C}^N} \left\|\mathbf{s}(\omega)\right\|_0 \ \text{ subject to } \ \left\|\mathbf{A}_1(\omega)\mathbf{s}(\omega) - \mathbf{p}(\omega)\right\|_2 \le \varepsilon, \tag{2}$$

where ε is a parameter to accommodate the additive noise level. Since an ℓ_0-norm objective function is not convex, the constrained optimization problem is reformulated with an ℓ_1-norm function via the relaxation approach [14]

$$\min_{\mathbf{s}(\omega)\in\mathbb{C}^N} \|\mathbf{s}(\omega)\|_1 \quad \text{subject to} \quad \|\mathbf{A}_1(\omega)\mathbf{s}(\omega)-\mathbf{p}(\omega)\|_2 \le \varepsilon, \tag{3}$$

which can also be regarded as a least-squares problem with ℓ_1-norm penalty. The regularization parameter λ weights the sparsity against the residual error.

2.2 The LASSO-Proxgrad Approach

This section describes the LASSO-Proxgrad method which solves the LASSO problem (4) iteratively. We denote the cost function in (4) by two scalar functions $g(\mathbf{s})$ and $h(\mathbf{s})$, i.e.,

$$\frac{1}{2}\|\mathbf{A}_1(\omega)\mathbf{s}(\omega)-\mathbf{p}(\omega)\|_2^2 + \lambda\|\mathbf{s}(\omega)\|_1 = g(\mathbf{s}(\omega)) + h(\mathbf{s}(\omega)), \tag{5}$$

where $g(\mathbf{s}(\omega)) = \frac{1}{2}\|\mathbf{A}_1(\omega)\mathbf{s}(\omega)-\mathbf{p}(\omega)\|_2^2$ is a convex and differentiable function and $h(\mathbf{s}(\omega)) = \lambda\|\mathbf{s}(\omega)\|_1$ is a convex but non-differentiable function. It can be shown that the solution of the preceding two-term optimization problem, $\min_{\mathbf{s}(\omega)\in\mathbb{C}^N} g(\mathbf{s}(\omega)) + h(\mathbf{s}(\omega))$, can be updated recursively by using the proximal gradient as [17]

$$\mathbf{s}_{d+1}(\omega) = \text{prox}_{\mu_d h}\left(\mathbf{s}_d(\omega) - \mu_d \nabla g(\mathbf{s}_d(\omega))\right), \tag{6}$$

where $\mathbf{s}_d(\omega)$, μ_d, and $\nabla g(\mathbf{s}_d(\omega))$ denote the solution of $\mathbf{s}(\omega)$, the stepsize, and the complex gradient at the dth iteration. The proximal operator applied to $h(\mathbf{s}(\omega))$ is defined as [21]

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega)) = \arg\min_{\mathbf{u}(\omega)\in\mathbb{C}^N}\left(h(\mathbf{u}(\omega)) + \frac{1}{2\mu_d}\|\mathbf{u}(\omega)-\mathbf{s}(\omega)\|_2^2\right). \tag{7}$$

An explicit expression for (6) in the context of the LASSO optimization problem (4) can be obtained by considering the facts that we are concerned with complex-valued vectors and $\|\mathbf{s}(\omega)\|_1$ is not differentiable at $\mathbf{s}(\omega) = \mathbf{0}$. It is shown in APPENDIX A that the complex subgradient set of $\|\mathbf{s}(\omega)\|_1$ is

$$\partial\|\mathbf{s}(\omega)\|_1 = \begin{cases} e^{j\boldsymbol{\theta}_s(\omega)}, & \mathbf{s}(\omega)\neq\mathbf{0}\in\mathbb{C}^N \\ \boldsymbol{\rho}(\omega)\odot e^{j\boldsymbol{\theta}_s(\omega)}, & \mathbf{s}(\omega)=\mathbf{0}, \end{cases} \tag{8}$$

where $\boldsymbol{\theta}_s(\omega)$ is the element-wise phase vector of $\mathbf{s}(\omega)$. The magnitude vector, $\boldsymbol{\rho}(\omega) = [\rho_1(\omega) \quad \cdots \quad \rho_n(\omega)]^T$, $\rho_i(\omega)\in[0,1]$, and " $\odot$ " denotes the Hadamard product. Using this fact, it can be shown in APPENDIX B that the complex-valued proximal operator for LASSO is

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega)) = \left(|\mathbf{s}(\omega)| - \mu_d\lambda\mathbf{1}_N\right)_+ \odot e^{j\boldsymbol{\theta}_s(\omega)} = S_{\mu_d\lambda}(\mathbf{s}(\omega)), \tag{9}$$

where $|\mathbf{s}(\omega)|$ denotes the element-wise complex modulus of the vector $\mathbf{s}(\omega)$, $\mathbf{1}_N = [1\cdots1]^T \in \mathbb{R}^N$, and $e^{j\boldsymbol{\theta}_s(\omega)}$ is a vector of complex exponential functions with the element-wise phase components of $\mathbf{s}(\omega)$ as their arguments. Hereby,

$$S_{\mu_d\lambda}(\mathbf{s}(\omega)) = \left(|\mathbf{s}(\omega)| - \mu_d\lambda\mathbf{1}_N\right)_+ \odot \exp(j\boldsymbol{\theta}_s(\omega))$$

represents a complex soft-threshold operator with

$$\left[\left(|\mathbf{s}(\omega)| - \mu_d\lambda\mathbf{1}_N\right)_+\right]_n = \max\left[|s_n(\omega)| - \mu_d\lambda, 0\right], \ \mu_d\lambda \ge 0, n = 1,\ldots,N.$$

From (9), it becomes clear that the proximal operator for a LASSO problem is a soft threshold.

To summarize, the complex proximal gradient algorithm for LASSO can be written as

$$\mathbf{s}_{d+1}(\omega) = S_{\mu_d\lambda}\left(\mathbf{s}_d(\omega) - \mu_d\nabla g(\mathbf{s}_d(\omega))\right), \tag{10}$$

Where

$$\nabla g(\mathbf{s}(\omega)) = \nabla\left(\frac{1}{2}\|\mathbf{A}_1(\omega)\mathbf{s}(\omega)-\mathbf{p}(\omega)\|_2^2\right) = \mathbf{A}_1^H(\omega)\left[\mathbf{A}_1(\omega)\mathbf{s}(\omega)-\mathbf{p}(\omega)\right]. \tag{11}$$

The convergence criterion for (10) used in this paper is $\left\|\mathbf{s}_{d+1}(\omega)-\mathbf{s}_d(\omega)\right\|_2 < \varepsilon_s \left\|\mathbf{s}_d(\omega)\right\|_2$. The parameter ε_s is selected to be 0.001 in the following simulation and experiment.

To conclude, the nonzero entries of the converged $\mathbf{s}_d(\omega)$ yield jointly the information of the source number, the source directions (from the associated column of the dictionary), and the associated source amplitudes at the frequency ω. The complex LASSO-Proxigrad algorithm is summarized in the following pseudocode.

Algorithm I The LASSO-Proxgrad algorithm

1: Construct a dictionary from a set of pre-selected steering vectors $\mathbf{A}_1(\omega) = \left[\mathbf{a}_0(\omega)\ \mathbf{a}_1(\omega)\cdots\ \mathbf{a}_{N-1}(\omega)\right] \in \mathbb{C}^{M\times N}$
2: Measured sound pressures $\mathbf{p}(\omega)$
3: **for** all frequency bins ω
4: $\mathbf{s}_0(\omega) \leftarrow 0$
5: **while** $\left\|\mathbf{s}_{d+1}(\omega)-\mathbf{s}_d(\omega)\right\|_2 < \varepsilon_s \left\|\mathbf{s}_d(\omega)\right\|_2$ **do**
6 : $\nabla g(\mathbf{s}_d(\omega)) \leftarrow \mathbf{A}_1(\omega)^H\left[\mathbf{A}_1(\omega)\mathbf{s}_d(\omega)-\mathbf{p}(\omega)\right]$
7: Update solution $\mathbf{s}_{d+1}(\omega) = S_{t_d\lambda}\left(\mathbf{s}_d(\omega)-t_d\nabla g(\mathbf{s}_d(\omega))\right)$;
8: **end while**
9: **end for**
10: Output the source signal vector $\mathbf{s}(\omega)$.
11: Identify nonzero entries to yield source directions and reconstruct the time-domain signals by using overlap (50%)-and-save process.

2.3 Wiener-based Postfiltering

By assuming the W-disjoint orthogonality [19], each time-frequency bin is dominated by one speaker and the signals from other speakers are treated as interference. Thus, a Wiener postfilter can be cascaded with the LASSO-Proxgrad algorithm to further improve signal quality, denoted as LASSO-Proxgrad-PF hereafter. The postfilter is based on the Wiener filter

$$W_i(\omega) = \frac{G_i^{ss}(\omega)}{G_i^{pp}(\omega)} = \frac{G_i^{ss}(\omega)}{G_i^{ss}(\omega)+G_i^{vv}(\omega)} = \frac{1}{1+\zeta_i(\omega)}, \quad i=1,\ldots,D \tag{12}$$

where $G_i^{ss}(\omega)$, $G_i^{vv}(\omega)$, and $G_i^{pp}(\omega)$ denote the power spectral density functions of the clean signal, the interference (from other sources), and the mixture signal associated with the *i*th source. Here, we assume that the source signal and the interference are uncorrelated. The factor, $\zeta_i(\omega) = G_i^{vv}(\omega)/G_i^{ss}(\omega)$, being the Noise-to-Signal Ratio. Next, the maximum power spectrum of signals $\hat{G}_i(\omega)$ is estimated , where $\hat{G}_i(\omega),\ i=1,\ldots,N$ is the first *N*greatest values in $\mathbf{s}_d(\omega)$ extracted by the LASSO-Proxgrad algorithm.

$$G_{\max}(\omega) = \max_{i=1,\ldots,D}\left[\hat{G}_i(\omega)\right]. \tag{13}$$

For simplicity, let us assume that interference is approximately Gaussian white noise with constant power spectral density $G_i^{vv}(\omega) \approx \alpha G_{\max}(\omega)$, where $0 \le \alpha \le 1$. Thus, we can rewrite $\zeta_i(\omega)$ as

$$\zeta_i(\omega) = \frac{\alpha}{\max\left[\left(G_i^{pp}(\omega)/G_{\max}(\omega)\right)-\alpha, \eta\right]} \tag{14}$$

The small number η is used to prevent (14) from being divided by zero. The denominator of (14) serves as a thresholding operator. With the Wiener filter in (12), the extracted source signals can be further enhanced.

$$s_i'(\omega) = s_i(\omega)W_i(\omega), \quad i=1,\ldots,D \tag{15}$$

3. Experimental Study

To validate the proposed algorithms in realistic conditions, experiments are undertaken in a 3.6m × 3.6m × 2.48m listening room (reverberation time, T_{60} = 240 ms). Eight 1/4" PCB® condenser microphones are configured as a uniform circular array with radius 12.0 cm. Data acquisition is performed in a sample rate of 16 kHz. The center of the array is designated as the origin. Two loudspeakers are placed at 1.5 m away from the array center with directions,

$(\theta,\phi) = (45^\circ, 90^\circ)$ and $(\theta,\phi) = (135^\circ, 90^\circ)$, respectively. Five-second clips of male and female speech signals serve as two source signals. A fan noise signal and seven independent sensor noise signals are mixed into spherically isotropic noise and added to the microphone signals with SNR=30, 20, 10 dB, by using the technique suggested in [22].

In the following experiments, the proposed LASSO-Proxgrad-PF, with $\alpha = 0.4$ and $\eta = 0.01$, is employed. LASSO solved using the convex optimization software CVX [23], denoted as LASSO-CVX hereafter, is adopted as the baseline for the localization and signal extraction tasks. Orthogonal Matching Pursuit (OMP) is used as another baseline.

3.1 Localization Performance

The normalized localization results are shown in Fig. 1, where the LASSO-CVX, LASSO-Proxgrad, and OMP are compared. The regularization parameters in these algorithms are selected by the L-curve method [24]. The localization results are illustrated by plotting the normalized source magnitude versus azimuthal angles in one-degree steps, with a fixedpolar angle 90°. Likewise, an angular grid with 1° resolution is employed to construct the dictionary of the sensing matrix for CS computation. The discrete-time signals are transformed to the STFT domain, based on 1024-point frames with 50% overlap and a Hanning window. Localization is performed using band-limited signals in 765-4672 Hz to avoid spatial aliasing. The localization errors (estimated angle - target angle) and required runtime are summarized for the proposed LASSO-Proxgrad, CVX, and OMP in Table1. In Fig. 1, OMP and LASSO-Proxgrad yields two clear maxima. In contrast, LASSO-CVX yields sporadic peaks by which the source directions are difficult to interpret. LASSO-Proxgrad and OMP yield comparable localization accuracy (Table1).

3.2 Signal Extraction Performance

With the same settings as in the preceding localization tests, the runtime for all algorithms to extract 5-second audio clips is listed in TABLE II. To quantify the signal extraction performance, Source-to-Interference Ratio (SIR), Source-to-Distortion Ratio (SDR), and Perceptual Evaluation of Speech Quality (PESQ) [25] are adopted. SDR and SIR correlate well with noise reduction performance and signal separation quality from our past experience. The perceptual signal quality is assessed using PESQ. The improvements of these three performance indices, denoted as ΔSIR, ΔSDR, and $\Delta PESQ$, are summarized in Tables II- IV. The proposed LASSO-Proxgrad attains significantly larger improvement than LASSO-CVX. LASSO-CVX suffers from pronounced artifacts. Furthermore, LASSO-Proxgrad yields extracted signals in much better quality, while requiring much shorter processing time than LASSO-CVX. Although OMP is also computationally efficient, the SDR improvement achieved by OMP is rather limited, as compared to LASSO-Proxgrad. Thanks to the post processing, LASSO-Proxgrad-PF performs the best among all methods with much improved extracted signal quality, as reflected in the SIR, SDR, and PESQ improvements.

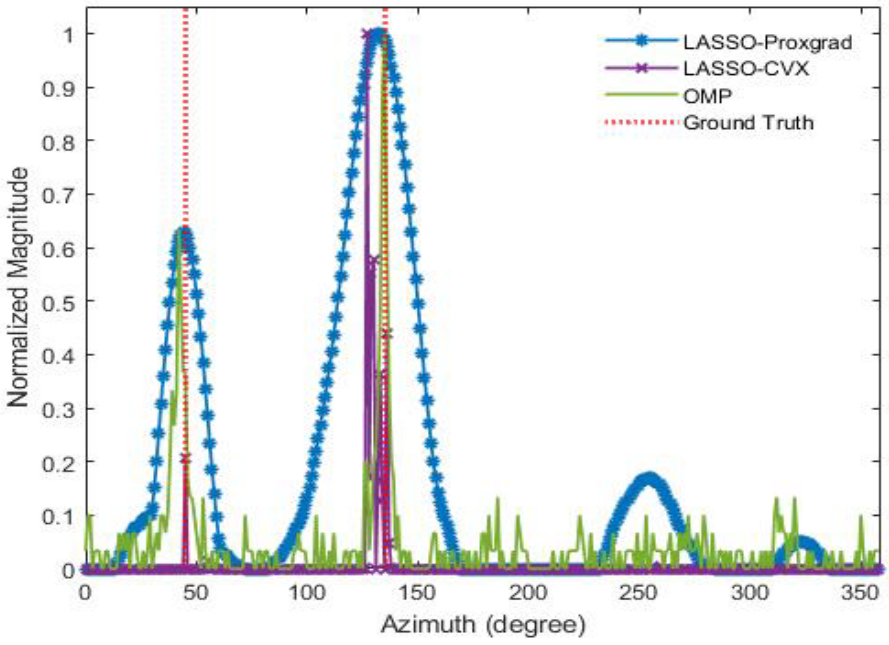

Figure 1. Normalized localization results.

Table 1. Comparison of localization performance

		LASSO-CVX	LASSO-Proxgrad	OMP
Localization Error (deg) (source 1, source 2)	without noise	(0, -9)	(-1, -2)	(-3, 0)
	SNR=30 dB	(0, -8)	(-1, -2)	(-3, 0)
	SNR=20 dB	(0, -6)	(-1, -2)	(-3, 0)
	SNR=10 dB	(0, -8)	(-1, -2)	(-3, 0)
Runtime (sec)		113.190	11.090	**0.032**

Table 2. Comparison of signal extraction performance (ΔSIR)

	Source	LASSO-CVX	OMP	LASSO-Proxgrad	LASSO-Proxgrad-PF
without noise	1	17.09	5.39	6.81	**20.34**
	2	-2.35	7.11	5.12	**20.13**
SNR=30 dB	1	12.68	5.44	6.80	**20.35**
	2	-2.67	7.11	5.11	**20.12**
SNR=20 dB	1	17.70	5.34	6.78	**20.34**
	2	-2.65	7.08	5.10	**20.10**
SNR=10 dB	1	16.96	5.30	6.70	**20.42**
	2	-2.77	7.06	5.06	**19.89**
Runtime (sec)		29864.1	**8.8**	383.7	

Table 3. Comparison of signal extraction performance (ΔSDR)

	Source	LASSO-CVX	OMP	LASSO-Proxgrad	LASSO-Proxgrad-PF
without noise	1	-7.10	-2.55	6.64	**13.02**
	2	-12.11	4.84	5.07	**13.33**
SNR=30 dB	1	-7.25	-2.49	6.62	**12.94**
	2	-7.30	4.86	5.06	**13.29**
SNR=20 dB	1	-7.18	-2.38	6.53	**12.66**
	2	-12.20	-4.89	5.01	**13.03**
SNR=10 dB	1	-9.28	-3.00	5.89	**11.32**
	2	-12.32	5.10	4.57	**10.75**

Table 4. Comparison of signal extraction performance (ΔPESQ)

	Source	LASSO-CVX	OMP	LASSO-Proxgrad	LASSO-Proxgrad-PF
without noise	1	-0.35	0.06	0.56	**1.07**
	2	-1.10	0.39	0.38	**0.91**
SNR=30 dB	1	-0.91	0.10	0.56	**1.07**
	2	-0.87	0.40	0.39	**0.91**
SNR=20 dB	1	-0.35	0.11	0.57	**1.06**
	2	-0.73	0.31	0.44	**0.92**
SNR=10 dB	1	-0.44	0.12	0.58	**0.98**
	2	-1.06	0.42	0.40	**0.79**

4. Concluding Remarks

Proximal gradient-based CS algorithms have been examined in the context of SLSE problems in this paper. LASSO-Proxgrad exhibits superior performance in source localization and signal extraction with little computational cost. Wiener based postfiltering proves useful in enhancing speech quality of the signal extracted by the LASSO-Proxgrad algorithm. On the whole, the combined LASSO-Proxgrad-PF leads to the best SLSE performance with the minimal processing time among all approaches.

APPENDIX A

Theorem 1:

The complex subgradient set of $\|\mathbf{s}(\omega)\|_1$ can be written as

$$\partial\|\mathbf{s}(\omega)\|_1 = \begin{cases} e^{j\theta_s(\omega)}, & \mathbf{s}(\omega) \neq \mathbf{0} \in \mathbb{C}^N \\ \boldsymbol{\rho}(\omega) \odot e^{j\theta_s(\omega)}, \ \rho_i(\omega) \in [0,1], & \mathbf{s}(\omega) = \mathbf{0} \end{cases} \tag{A1}$$

Proof: First we prove the complex gradient $\nabla\|\mathbf{s}(\omega)\|_1 = e^{j\theta_s(\omega)}$ for $\mathbf{s}(\omega) \neq 0 \in \mathbb{C}^N$.

Let the Fourier transform of source signal $s_i(\omega) = x_i(\omega) + y_i(\omega),\ i = 1,\ldots N$.

Then the gradient can be written as

$$\nabla\|\mathbf{s}(\omega)\|_1 = \nabla\left[\sum_{i=1}^{N}\sqrt{x_i^2(\omega)+y_i^2(\omega)}\right]$$

$$= \begin{bmatrix} \left(\frac{\partial}{\partial x_1(\omega)} + j\frac{\partial}{\partial y_1(\omega)}\right)\sum_{i=1}^{N}\sqrt{x_i^2(\omega)+y_i^2(\omega)} \\ \vdots \\ \left(\frac{\partial}{\partial x_N(\omega)} + j\frac{\partial}{\partial y_N(\omega)}\right)\sum_{i=1}^{N}\sqrt{x_i^2(\omega)+y_i^2(\omega)} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1(\omega)}\sqrt{x_1^2(\omega)+y_1^2(\omega)} + j\frac{\partial}{\partial y_1(\omega)}\sqrt{x_1^2(\omega)+y_1^2(\omega)} \\ \vdots \\ \frac{\partial}{\partial x_N(\omega)}\sqrt{x_N^2(\omega)+y_N^2(\omega)} + j\frac{\partial}{\partial y_N(\omega)}\sqrt{x_N^2(\omega)+y_N^2(\omega)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x_1(\omega)}{\sqrt{x_1^2(\omega)+y_1^2(\omega)}} + j\frac{y_1(\omega)}{\sqrt{x_1^2(\omega)+y_1^2(\omega)}} \\ \vdots \\ \frac{x_N(\omega)}{\sqrt{x_N^2(\omega)+y_N^2(\omega)}} + j\frac{y_N(\omega)}{\sqrt{x_N^2(\omega)+y_N^2(\omega)}} \end{bmatrix} = \begin{bmatrix} \frac{s_1(\omega)}{|s_1(\omega)|} \\ \vdots \\ \frac{s_N(\omega)}{|s_N(\omega)|} \end{bmatrix} = \begin{bmatrix} e^{j\theta(s_1(\omega))} \\ \vdots \\ e^{j\theta(s_N(\omega))} \end{bmatrix} = e^{j\theta_s(\omega)}$$

Next, we prove $\partial\|\mathbf{s}(\omega)\|_1 = \rho_i(\omega)e^{j\theta_i(\omega)},\ \rho_i(\omega) \in [0,1],\ s_i(\omega) = 0$, where $s_i(\omega)$ is the ith entry of $\mathbf{s}(\omega)$. Let $s_i(\omega) = x_i(\omega) + jy_i(\omega) = R_i(\omega)e^{j\theta_i(\omega)},\quad i = 1,\ldots,N.$

The function $f(s_i(\omega)) = \|s_i(\omega)\|_1 = |s_i(\omega)|$ is a convex but non-differentiable function at $s_i(\omega) = 0$. For an $s_i(\omega)$ located in the neighborhood of $s_i(\omega) = 0$, a subgradient $g_i(\omega)$ must satisfy the following inequality

$$f(s_i(\omega)) - f(0) \geq g_i^*(s_i(\omega) - 0). \tag{A2}$$

Indeed,

$$f(s_i(\omega)) - f(0) = |s_i(\omega)| - 0 = \left|R_i(\omega)e^{j\theta_i(\omega)}\right| = R_i(\omega) \geq \rho_i(\omega)e^{-j\theta_i(\omega)}R_i(\omega)e^{j\theta_i(\omega)} = \rho_i(\omega)R_i(\omega),\ \rho_i(\omega) \in [0,1]$$

In summary, the subgradient set $\partial\|s_i(\omega)\|_1$ can be written as

$$\partial\left\|s_i(\omega)\right\|_1=\begin{cases}e^{j\theta_i(\omega)}, s_i(\omega)\neq 0\\ \rho_i(\omega)e^{j\theta_i(\omega)}, \rho_i(\omega)\in[0,1], s_i(\omega)=0\end{cases} \tag{A3}$$

It is straightforward to generalize the result to the vector case, $f(\mathbf{s}(\omega))=\|\mathbf{s}(\omega)\|_1$.
Thus,

$$\partial\left\|\mathbf{s}(\omega)\right\|_1=\begin{cases}e^{j\boldsymbol{\theta}_s(\omega)}, & \mathbf{s}(\omega)\neq\mathbf{0}\in\mathbb{C}^N\\ \boldsymbol{\rho}(\omega)\odot e^{j\boldsymbol{\theta}_s(\omega)}, \rho_i(\omega)\in[0,1], & \mathbf{s}(\omega)=\mathbf{0}\end{cases}$$

This concludes the proof. The sufficient condition of the existence for a stationary point is $0\in\partial f(\mathbf{s}(\omega))$.

APPENDIX B

Theorem 2:

The proximal gradient operator of a complex vector can be reduced to a soft threshold. That is,

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega))=\left(\left|\mathbf{s}(\omega)\right|-\mu_d\lambda\mathbf{1}_N\right)_+\odot e^{j\boldsymbol{\theta}_s(\omega)},$$

$$\text{where } \text{prox}_{\mu_d h}(\mathbf{s}(\omega))=\arg\min_{\mathbf{u}(\omega)\in\mathbb{C}^N}\left(h(\mathbf{u}(\omega))+\frac{1}{2\mu_d}\left\|\mathbf{u}(\omega)-\mathbf{s}(\omega)\right\|_2^2\right),\ h(\mathbf{u}(\omega))=\lambda\left\|\mathbf{u}(\omega)\right\|_1$$

Proof. The proximal gradient is defined as

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega))=\arg\min_{\mathbf{u}\in\mathbb{C}^N}\left(h(\mathbf{u}(\omega))+\frac{1}{2\mu_d}\left\|\mathbf{u}(\omega)-\mathbf{s}(\omega)\right\|_2^2\right),\ h(\mathbf{u}(\omega))=\lambda\left\|\mathbf{u}(\omega)\right\|_1. \tag{A4}$$

or, equivalently,

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega))=\arg\min_{\mathbf{u}\in\mathbb{C}^N}\left(\left\|\mathbf{u}(\omega)\right\|_1+\frac{1}{2\mu_d\lambda}\left\|\mathbf{u}(\omega)-\mathbf{s}(\omega)\right\|_2^2\right)$$

The minimizer can be found by imposing that $0\in\partial f(\mathbf{s}(\omega))$, where $f(\mathbf{s}(\omega))=\left\|\mathbf{u}(\omega)\right\|_1+\frac{1}{2\mu_d\lambda}\left\|\mathbf{u}(\omega)-\mathbf{s}(\omega)\right\|_2^2$. Using (A1),

$$e^{j\boldsymbol{\theta}_u(\omega)}+\frac{1}{\mu_d\lambda}(\mathbf{u}(\omega)-\mathbf{s}(\omega))=\mathbf{0} \tag{A5}$$

By writing $\mathbf{u}(\omega)=\left|\mathbf{u}(\omega)\right|\odot e^{j\boldsymbol{\theta}_u(\omega)}$ and $\mathbf{s}(\omega)=\left|\mathbf{s}(\omega)\right|\odot e^{j\boldsymbol{\theta}_s(\omega)}$ (element-wise operations), we can rearrange (A5) into

$$(\mu_d\lambda\mathbf{1}_N+\left|\mathbf{u}(\omega)\right|)\odot e^{j\boldsymbol{\theta}_u(\omega)}=\left|\mathbf{s}(\omega)\right|\odot e^{j\boldsymbol{\theta}_s(\omega)},\ \text{where } \mathbf{1}_N=[1\cdots 1]^T\in\mathbb{R}^N$$

It follows that

$$\text{prox}_{\mu_d h}(\mathbf{s}(\omega))=\left(\left|\mathbf{s}(\omega)\right|-\mu_d\lambda\mathbf{1}_N\right)_+\odot e^{j\boldsymbol{\theta}_s(\omega)}, \tag{A6}$$

where $\left[\left(\left|\mathbf{s}(\omega)\right|-\mu_d\lambda\mathbf{1}_N\right)_+\right]_i=\max\left[\left|s_i\right|(\omega)-\mu_d\lambda,0\right], \tau\geq 0, i=1,\ldots,N$.

References

[1] Rascon, C., & Meza, I. (2017). Localization of sound sources in robotics: A review. Robotics and Autonomous Systems, 96, 184-210. https://doi.org/10.1016/j.robot.2017.07.011.

[2] Bian, X., Abowd, G. D., & Rehg, J. M. (2005, May). Using sound source localization in a home environment. In *Int ernational Conference on Pervasive Computing*, Springer, Berlin, Heidelberg, 19-36. https://doi.org/10.1007/11428572_2.

[3] Marti, A., Cobos, M., & Lopez, J. J. (2011, May). Real time speaker localization and detection system for camera ste ering in multiparticipant videoconferencing environments. In *2011 IEEE International Conference on Acoustics, Speech*

and Signal Processing (ICASSP), 2592-2595.

[4] Brown, G. J., & Wang, D. (2005). Separation of speech by computational auditory scene analysis. In *Speech enhancement*, Springer, Berlin, Heidelberg, 371-402.

[5] Kotus, J., Lopatka, K., & Czyzewski, A. (2014). Detection and localization of selected acoustic events in acoustic field for smart surveillance applications. *Multimedia Tools and Applications*, *68*(1), 5-21. https://doi.org/10.1007/s11042-012-1183-0.

[6] Yang, M., & De Hoog, F. (2015). Orthogonal matching pursuit with thresholding and its application in compressive sensing. *IEEE Transactions on Signal Processing*, *63*(20), 5479-5486. https://doi.org/10.1109/TSP.2015.2453137.

[7] Li, J., Wu, Z., Feng, H., Wang, Q., & Liu, Y. (2014, May). Greedy orthogonal matching pursuit algorithm for sparse signal recovery in compressive sensing. In *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 1355-1358.

[8] Dorfan, Y., Schwartz, O., Schwartz, B., Habets, E. A., & Gannot, S. (2016, November). Multiple DOA estimation and blind source separation using estimation-maximization. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 1-5.

[9] Çöteli, M. B., & Hacıhabiboğlu, H. (2021). Sparse Representations With Legendre Kernels for DOA Estimation and Acoustic Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 2296-2309. https://doi.org/10.1109/TASLP.2021.3091845.

[10] Bai, M. R., Lan, S. S., Huang, J. Y., Hsu, Y. C., & So, H. C. (2020). Audio enhancement and intelligent classification of household sound events using a sparsely deployed array. *The Journal of the Acoustical Society of America*, *147*(1), 11-24. https://doi.org/10.1121/10.0000492.

[11] Bai, M. R., & Chen, C. C. (2013). Application of convex optimization to acoustical array signal processing. *Journal of Sound and Vibration*, *332*(25), 6596-6616. https://doi.org/10.1016/j.jsv.2013.07.029.

[12] Gerstoft, P., Xenaki, A., & Mecklenbräuker, C. F. (2015). Multiple and single snapshot compressive beamforming. *The Journal of the Acoustical Society of America*, *138*(4), 2003-2014. https://doi.org/10.1121/1.4929941.

[13] Chi, Y., Scharf, L. L., Pezeshki, A., & Calderbank, A. R. (2011). Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, *59*(5), 2182-2195. https://doi.org/10.1109/TSP.2011.2112650.

[14] Elad, M. (2010). *Sparse and redundant representations: from theory to applications in signal and image processing*. New York: springer.

[15] Cai, T. T., & Wang, L. (2011). Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information theory*, *57*(7), 4680-4688. https://doi.org/10.1109/TIT.2011.2146090.

[16] Lilis, G. N., Angelosante, D., & Giannakis, G. B. (2010). Sound field reproduction using the lasso. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(8), 1902-1912. https://doi.org/10.1109/TASL.2010.2040523.

[17] Parikh, N., & Boyd, S. (2014). Proximal algorithms. *Foundations and trends® in Optimization*, *1*(3), 127-239.

[18] L. Vandenberghe, "Proximal gradient method," Ch.4, Lecture notes of ECE236C - Optimization Methods for Large-Scale Systems (Spring 2019), UCLA. http://www.seas.ucla.edu/~vandenbe/ee236c.html.

[19] Rickard, S., & Yilmaz, O. (2002, May). On the approximate W-disjoint orthogonality of speech. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASS)*, I-529-I-532.

[20] Candes, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *C. R. l'Academie des Sciences, Ser. I*, no. 346, 589-592.

[21] Chen, A. I., & Ozdaglar, A. (2012, October). A fast distributed proximal-gradient method. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing*, 601-608.

[22] Habets, E. A., Cohen, I., & Gannot, S. (2008). Generating nonstationary multisensor signals under a spatial coherence constraint. *The Journal of the Acoustical Society of America*, *124*(5), 2911-2917. https://doi.org/10.1121/1.2987429.

[23] Guimaraes, D. A., Floriano, G. H. F., & Chaves, L. S. (2015). A tutorial on the CVX system for modeling and solving convex optimization problems. *IEEE Latin America Transactions*, *13*(5), 1228-1257. https://doi.org/10.1109/TLA.2015.7111976.

[24] Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM review*, *34*(4), 561-580. https://doi.org/10.1137/1034115.