T-SVD BASED BROADBAND NON-SYNCHRONOUS MEASUREMENTS

*Long Chen** *Weize Sun*† *Lei Huang*‡ *Guitong Chen*

Shenzhen University
College of Electronics and Information Engineering
Shenzhen, Guangdong, China

ABSTRACT

It is a challenge for the state-of-the-art non-synchronous measurements beamforming method to localize multiple broadband sources due to the difficulty in selecting an appropriate operating frequency without any prior information about the target signals. In this paper, we propose a tensor singular value decomposition based non-synchronous measurements method for broadband multiple sound source localization. By adopting the proposed method, the working frequency range of the microphone array is no longer limited by the array geometry. While the proposed tensor completion approach via alternating direction method of multipliers algorithm could provide a sound map with distinct global view of three different speech signal sources with high accuracy in both simulation and experimental validations.

***Index Terms*—** Sound source localization, beamforming, non-synchronous measurements, alternating direction method of multipliers, tensor singular value decomposition

1. INTRODUCTION

In practical sound source localization (SSL), the working frequency range of a microphone array is always limited by the array size and the distance between two adjoining microphones critically via acoustical beamforming [1, 2]. As a new method to break through this limitation, the non-synchronous measurements (NSM) method is firstly proposed for the acoustic imaging in near-field acoustical holography (NAH) [3]. By sequentially moving a small prototype array during the measurements, both the size of microphone array and microphone density become flexible so that the working frequency range of the array could be extended [4, 5]. Unlike the synthetic aperture radar (SAR) or synthetic aperture sonar (SAS), the microphone array is a passive array [6, 7]. Therefore we need a matrix or tensor completion approach to obtain the missing information, which is not necessary in SAR and SAS.

The NSM method is implemented for acoustic beamforming with a cross-spectral matrix (CSM) completion approach by Yu et al. [8]. Two different algorithms including augmented Lagrange multiplier (ALM) and alternative direction method of multipliers (ADMM) are applied to solve the raised optimization problem with comparison and discussions. Based on this work, the NSM approach is further improved by block Hermitian matrix completion (BHMC), variational Bayesian (VB) inference and coprime positions (CP-NSM) for more accurate and robust source localization results [9, 10, 11].

Even though the working frequency range of acoustic beamforming could be broadened by the NSM method by a virtual larger and denser array, the frequency bin selection problem for broadband source localization is still unsolved without the frequency-domain prior information of the target sound signals. To solve the frequency bin selection problem for broadband signals in real-time practical applications, we propose a tensor singular value decomposition (T-SVD) based NSM approach for broadband multiple sound source localization (BM-SSL) by considering the complexity and multidimensional structure features of received signals in frequency, time and space domain [12]. In the proposed algorithm, the observed tensor structure could be employed and fully utilized to complete the target data by minimizing the defined tensor rank. At last, the broadband multiple sound sources could be localized by operating the NSM approach in a reasonable frequency range.

2. ALGORITHM DESCRIPTION

2.1. Problem formulation

Consider a near-field NSM model as shown in Fig. 1. Assuming that p_m^k is one of the measured sound pressure in the K measurements by the array, and r_m is the distance between the sound source and the mth receiver microphone. The acoustic pressure field could be expressed in vector form as:

$$\mathbf{p}^k(r_m) = [p_1^k(r_m) \quad p_2^k(r_m) \quad \cdots \quad p_M^k(r_m)]^T, \quad (1)$$

*Thanks to National Natural Science Foundation of China (NSFC) for Grant 62101335.

†Corresponding author: Weize Sun, proton198601@hotmail.com.
Thanks to Guangdong Basic and Applied Basic Research Foundation for Grant 2021A1515011706 and Foundation of Shenzhen for Grant JCYJ20190808122005605.

‡Thanks to National Natural Science Foundation of China (NSFC) for Grants U1713217 and 61925108.

and the CSM $\mathbf{R}^k \in \mathbb{C}^{M\times M}$ is

$$\mathbf{R}^k(r_m) = E[\mathbf{p}^k(r_m)\mathbf{p}^k(r_m)^H], \tag{2}$$

where $E[\cdot]$ denotes the mathematical expectation, $(\cdot)^T$ is the transpose, $(\cdot)^H$ is the conjugate transpose, $m = (1, 2, \cdots, M)$, and $k = (1, 2, \cdots, K)$. The location of sound source in the measurement plane could be obtained from the narrowband multiple signal classification (MUSIC) method with an eigendecomposition of $\mathbf{R}^k$:

$$\beta_N(r_m) = \frac{1}{\mathbf{w}(r_m)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{w}(r_m)}, \tag{3}$$

in which $\mathbf{U}_n$ and $\mathbf{D}_n$ are from the eigenvalues in noise subspace [2, 13].

Since the whole procedure of the NSM consists K times measurements, the K CSMs $\mathbf{R}^k \in \mathbb{C}^{M\times M}$ could be regarded as the diagonal blocks in a full CSM $\mathbf{R} \in \mathbb{C}^{MK\times MK}$ obtained by a synchronous measurement, which consists all the microphone array in the different positions. With this understanding, the observed $\mathbf{R}_\Omega$ by NSM could be expressed as:

$$\mathbf{R}_\Omega = \begin{bmatrix} \mathbf{R}^1 & & \\ & \ddots & \\ & & \mathbf{R}^K \end{bmatrix}, \tag{4}$$

where Ω is a sampling operator that extracts the element from diagonal blocks. To seek for a full estimated CSM $\hat{\mathbf{R}} \in \mathbb{C}^{MK\times MK}$ from N CSMs $\mathbf{R}^k$, the SSL problem is reformulated to a low rank diagonal block matrix completion problem. From the point of optimization view, a low rank matrix completion problem is equivalent to minimizing the matrix rank, which could be relaxed to a nuclear norm minimization problem since the rank minimization problem is NP-hard [14, 15]. In the former works, ALM and ADMM are applied to solve the optimization problem with similar results [8]. For comparison, we all adopt the ADMM algorithm in the following simulations and experiments [16, 17, 18].

Since the localization approach by the ADMM-based NSM method is conducted at one specific frequency, an improper selection of the frequency bin would lead to an inaccurate or even wrong source location in the beamforming maps. A straight forward solution is to estimate the CSM $\hat{\mathbf{R}}$ at every frequency bin f. Then the broadband NSM output could be obtained from [19]:

$$\beta_B(r_m) = \frac{1}{\sum_{f=f_0}^{f_S} \mathbf{w}(f, r_m)^H \mathbf{U}(f)_n \mathbf{U}(f)_n^H \mathbf{w}(f, r_m)}. \tag{5}$$

2.2. Proposed solution

2.2.1. T-SVD and tensor multi-norm

For ease of understanding of the proposed method, we will first introduce the concept of T-SVD and tensor multi-norm

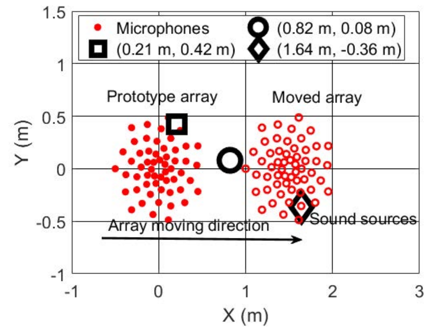


Fig. 1. Locations of the sound sources to be estimated by NSM.

(TMN). The T-SVD of a 3-D tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{D_1\times D_2\times D_3}$ is defined as [12]:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^H = \sum_{d=1}^{D_r} \boldsymbol{\mathcal{U}}(:, d, :) * \boldsymbol{\mathcal{S}}(d, d, :) * \boldsymbol{\mathcal{V}}(:, d, :)^H, \tag{6}$$

where $D_r \leq \min\{D_1, D_2\}$, $*$ is the tensor product, and $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ are orthogonal tenors satisfying

$$\boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{U}}^H = \boldsymbol{\mathcal{U}}^H * \boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{I}} \in \mathbb{C}^{D_1\times D_1\times D_3}, \tag{7}$$

and

$$\boldsymbol{\mathcal{V}} * \boldsymbol{\mathcal{V}}^H = \boldsymbol{\mathcal{V}}^H * \boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{I}} \in \mathbb{C}^{D_2\times D_2\times D_3}. \tag{8}$$

$\boldsymbol{\mathcal{I}}$ represents the identity tensor. We say a tensor is F-diagonal if each frontal slice is diagonal [12, 20]. $\boldsymbol{\mathcal{S}}$ is an F-diagonal tensor whose frontal slices are all diagonal matrices, and it contains all the D_r singular value tubes of tensor $\boldsymbol{\mathcal{X}}$ [21]. The TMN of $\boldsymbol{\mathcal{X}}$ is defined as the number of non-zero entries of the tensor $\boldsymbol{\mathcal{S}}$:

$$\|\boldsymbol{\mathcal{X}}\|_{TMN} = \sum_{d=1}^{D_r} r_d, \tag{9}$$

where r_d is the number of non-zero elements of the dth frontal slice $\boldsymbol{\mathcal{S}}(:, :, d)$ for $d = 1, 2, \cdots, D_3$ [12].

2.2.2. T-SVD based NSM

We define the $\boldsymbol{\mathcal{R}}_\Omega \in \mathbb{C}^{MK\times MK\times S}$ as the observed tensor which is composed of $K \times S$ known sub-matrices $\mathbf{R}^k \in \mathbb{C}^{M\times M}$ at every frequency bin f, and $\boldsymbol{\mathcal{R}} \in \mathbb{C}^{MK\times MK\times S}$ is the target cross-spectral tensor (CST). $\boldsymbol{\mathcal{R}}_\Omega$ could be expressed as:

$$\boldsymbol{\mathcal{R}}_\Omega = \boldsymbol{\mathcal{R}}_{1\Omega} \sqcup_3 \boldsymbol{\mathcal{R}}_{2\Omega} \sqcup_3 \cdots \sqcup_3 \boldsymbol{\mathcal{R}}_{S\Omega}, \tag{10}$$

where $f = 1, 2, \cdots, S$ is the number of frequency slices, and $\sqcup_3$ is the concatenate operation along the 3rd dimension. Similarly, to seek for a full estimated CST $\hat{\boldsymbol{\mathcal{R}}}$ from $K \times S$ CSMs $\mathbf{R}^k$, the BM-SSL problem is reformulated to a tensor completion problem, which could be solved by the minimization of

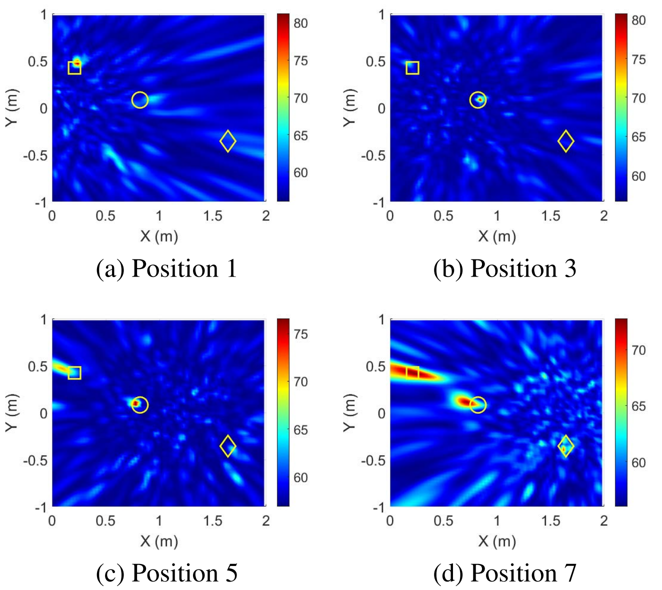


Fig. 2. Direct localization results by broadband MUSIC at different positions.

the TMN:

$$\min_{\hat{\mathcal{R}},\mathcal{M}} \quad \lambda\|\hat{\mathcal{R}}\|_{TMN} + \frac{1}{2}\|\mathcal{M}_\Omega - \mathcal{R}_\Omega\|_F^2$$
$$s.t. \quad \hat{\mathcal{R}} = \mathcal{M}, \quad \hat{\mathcal{R}}_f \succeq 0, \tag{11}$$

where λ is a regularization parameter [22], and $\|\cdot\|_F$ denotes the Frobenius norm. Its augmented Lagrangian function is:

$$L(\hat{\mathcal{R}},\mathcal{M},\mathcal{Y}) = \lambda\|\hat{\mathcal{R}}\|_{TMN} + \frac{1}{2}\|\mathcal{M}_\Omega - \mathcal{R}_\Omega\|_F^2 + \langle\mathcal{Y},\hat{\mathcal{R}} - \mathcal{M}\rangle + \frac{\mu}{2}\|\hat{\mathcal{R}} - \mathcal{M}\|_F^2. \tag{12}$$

Here we define $\langle\mathcal{A},\mathcal{B}\rangle = vec\{\mathcal{A}\}^H vec\{\mathcal{B}\}$, where $vec\{\mathcal{A}\}$ is the vectorization operation of the tensor $\mathcal{A}$. $\mathcal{Y} \in \mathbb{C}^{MK\times MK\times S}$ is the Lagrange multiplier, and μ is a positive penalty parameter. Then the $\hat{\mathcal{R}}$, $\mathcal{M}$, and $\mathcal{Y}$ can be updated iteratively using ADMM. By fixing $\mathcal{M}$ and $\mathcal{Y}$, $\hat{\mathcal{R}}$ can be updated by:

$$\hat{\mathcal{R}}_{k+1} = \arg\min_{\hat{\mathcal{R}}_f\succeq 0} \lambda\|\hat{\mathcal{R}}\|_{TMN} + \frac{1}{2}\|\hat{\mathcal{R}} - \mathcal{M}_k + \frac{1}{\mu}\mathcal{Y}_k\|_F^2 = \mathcal{U} * \max\{\mathcal{W} - \frac{\lambda}{\mu}, 0\} * \mathcal{U}_k^H, \tag{13}$$

where $\mathcal{M}_k - \frac{1}{\mu}\mathcal{Y}_k = \mathcal{U}_k * \mathcal{W}_k * \mathcal{U}_k^H$ is the T-SVD decomposition. The $\mathcal{M}$ and $\mathcal{Y}$ can be updated as follows:

$$\mathcal{M}_{k+1} = \arg\min_{\mathcal{M}} \frac{1}{2}\|\mathcal{M}_\Omega - \mathcal{R}_\Omega\|_F^2 + \frac{\mu}{2}\|\mathcal{M} - \frac{1}{\mu}\mathcal{Y}_k - \hat{\mathcal{R}}_{k+1}\|_F^2, \tag{14}$$

$$\mathcal{M}_{(k+1)\Omega} = \frac{1}{\mu+1}(\mathcal{R}_\Omega + \mu\hat{\mathcal{R}}_{(k+1)\Omega} + \mathcal{Y}_{k\Omega}), \tag{15}$$

$$\mathcal{M}_{(k+1)\bar{\Omega}} = \hat{\mathcal{R}}_{(k+1)\bar{\Omega}} + \frac{1}{\mu}\mathcal{Y}_{k\bar{\Omega}}, \tag{16}$$

$$\mathcal{Y}_{k+1} = \mathcal{Y}_k + \gamma\mu(\hat{\mathcal{R}}_{k+1} - \mathcal{M}_{k+1}), \tag{17}$$

where $\bar{\Omega}$ is a sampling operator that extracts the elements from non-diagonal blocks, and γ is a relaxation parameter [23]. And the stopping criterion is $\frac{\|\hat{\mathcal{R}}_\Omega - \mathcal{R}_\Omega\|_F}{\|\mathcal{R}_\Omega\|_F} \le \varepsilon$.

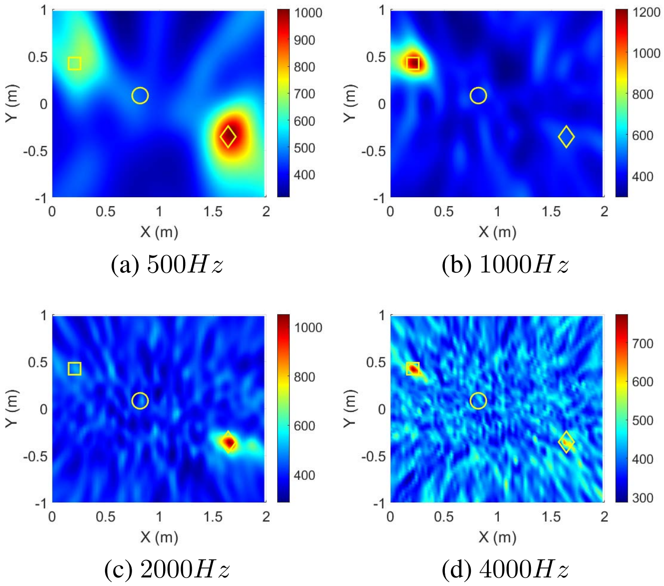
In order to get a good TMN optimization, the shrinkage factor $\frac{\lambda}{\mu}$ should be decrease as the number of iteration increases. Therefore, we set a reducing parameter $0 < \alpha < 1$ to update the λ as $\lambda_{k+1} = \lambda_k\alpha$ in each iteration. In the ADM-M optimization process, all the singular values less than the thresholding $\frac{\lambda}{\mu}$ would be rejected. The noise introduced by the incomplete CST would be reduced, and then the sidelobe levels could be reduced in the sound maps. Besides, to ensure the spatial continuity of the acoustic field, for each frequency slice, $\hat{\mathcal{R}}_f = \mathbf{\Psi}_f\hat{\mathcal{R}}_f\mathbf{\Psi}_f^H$ is set in each iteration by introducing a proposed projection basis $\mathbf{\Psi} \in \mathbb{C}^{MK\times MK}$ [8]. Therefore, the TC-based NSM problem can be solved by the ADMM algorithm. At last, by utilizing the completed CST $\hat{\mathcal{R}}$ along the frequency dimension from f_0 to f_S, the BM-SSL problem could be soved by equation (5).

3. SIMULATIONS

As shown in Fig. 1, a 56-channel spiral microphone array is considered. The number of measurement is $K = 7$. Three different speech signal sources marked by square, circle and diamond are set in the sound field with the coordinates $(0.21m, 0.42m)$, $(0.82m, 0.08m)$ and $(1.64m, -0.36m)$, respectively. The distance between the microphone array and the measurement plane is retained at $h = 0.6m$. The operating frequency range is 1000-6000 Hz with a 100 Hz step length. The sampling rate is $Fs = 16384Hz$.

Firstly, the localization results from different positions are shown in Fig. 2. The prototype array moves from left to right side along the x-axis. Clearly, the sound sources at farside could not be located at the singlehanded position. With the limitation of the size of the prototype array, it is rather difficult to find a proper position with all three accurate locations of the sources. Then the localization results of the narrowband NSM approach at different frequency are shown in Fig. 3. Since the sound sources are broadband speech signals, the selection of the operating frequency would significantly affect the localization results in this approach as they are shown.

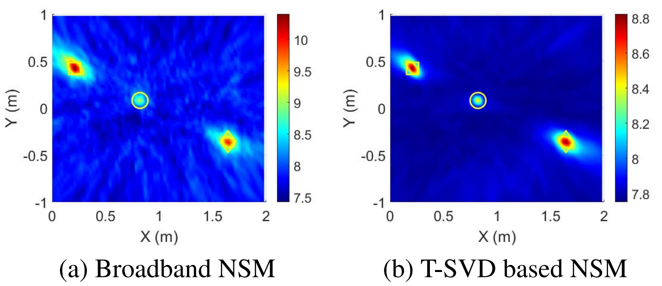
Compared with the results shown in Fig. 2 and Fig. 3, the two broadband NSM approaches virtually utilizes all the $M \times K = 392$ microphones and the three sound sources can be well located by this two broadband methods with a global view in Fig. 4. Furthermore, in Fig. 4 (b), it is shown that the proposed method could achieve a more distinct image of the three sound sources with less sidelobes.



(a) $500Hz$ (b) $1000Hz$

(c) $2000Hz$ (d) $4000Hz$

Fig. 3. Localization results by narrowband NSM at different frequencies.



(a) Broadband NSM (b) T-SVD based NSM

Fig. 4. Localization results by broadband NSM.

4. EXPERIMENTS

Experimental validation was carried out in a semi-anechoic chamber. The background noise level is $15.6dB(A)$, and the cut-off frequency is $100Hz$. As it is shown in Fig. 5, a 56-channel spiral array was placed $0.6m$ in front of the measurement plane in the experiments. The array was moved from left to right side along the x-axis. Three Philips BT25 loudspeakers were arranged in the measurement plane as sound sources with the coordinates $(-0.31m, 0.95m)$, $(-0.92m, 0.25m)$ and $(-1.71m, 0.65m)$, respectively. The acoustic signals were captured by a Mueller-BBM MKII sound measurement system with 56 $Br\ddot{u}el\&Kj\ae r4944A$ microphones, which were calibrated by a $Br\ddot{u}elKj\ae r4231$ $94dB$ calibrator before the measurements.

The direct localization results from different positions are shown in Fig. 6(a), (b) and (c) by broadband MUSIC. Obviously, the prototype array is too small to localize all the three sound sources. As a contrast, the T-SVD based NSM approach in Fig. 7(d) could provide a distinct and accurate global view of the three speech sources.

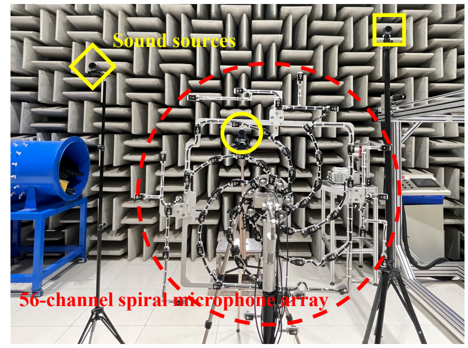
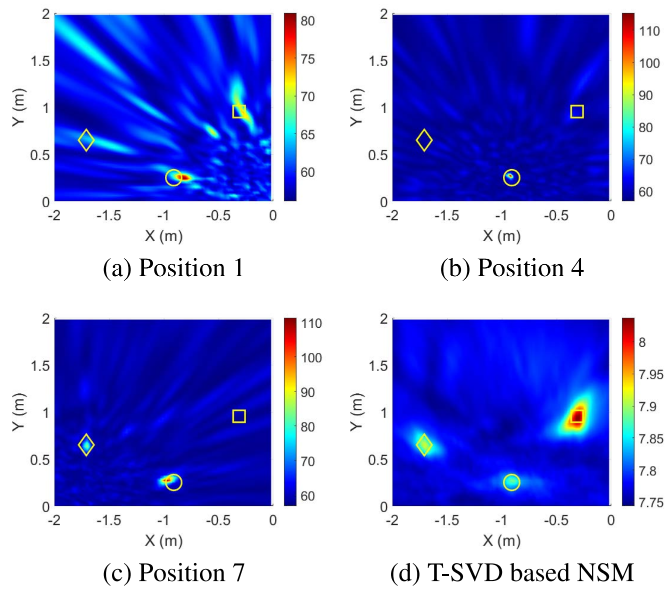


Fig. 5. Experimental setup: three loudspeakers and the 56-channel spiral microphone array in a semi-anechoic chamber.



(a) Position 1 (b) Position 4

(c) Position 7 (d) T-SVD based NSM

Fig. 6. Localization results of the experimental study.

5. CONCLUSION

It is always a trade-off between the aperture of a microphone array and the spatial separation of the microphones in microphone array design. The NSM method provides a potential solution for this problem. In this paper, we extend it and present a conceivable approach for BM-SSL. The problem of the selection of operating frequency for BM-SSL by NSM is well solved. The proposed method has the potential applications in imaging of the large-scale broadband sound radiation with a movable array (i.e., the service robot, vehicle, and industrial machinery and equipment).

6. REFERENCES

[1] Paolo Chiariotti, Milena Martarelli, and Paolo Castellini, "Acoustic beamforming for noise source localization–reviews, methodology and applications," *Mech. Syst. Signal. Process.*, vol. 120, pp. 422–448, 2019.

[2] Yang-Hann Kim and Jung-Woo Choi, *Sound visualization and manipulation*, John Wiley & Sons, 2013.

[3] J Antoni, "Synthetic aperture acoustical holography," in *International Conference on Noise and Vibration Engineering and International, Conference on Uncertainty in Structural Dynamics, ISMA2012, Leuven, Belgium*, 2012.

[4] Liang Yu, *Acoustical source reconstruction from non-synchronous sequential measurements*, Ph.D. thesis, INSA de Lyon, 2015.

[5] Dan Ding, *Implementation of Microphone Array Processing Techniques on A Synthetic Array for Fluid Power Noise Diagnostics*, Ph.D. thesis, Purdue University Graduate School, 2019.

[6] Bo Zhao, Lei Huang, and Weimin Bao, "One-bit sar imaging based on single-frequency thresholds," *IEEE Trans. Geosci. Remote. Sens.*, vol. 57, no. 9, pp. 7017–7032, 2019.

[7] Angeliki Xenaki and Yan Pailhas, "Compressive synthetic aperture sonar imaging with distributed optimization," *J. Acoust. Soc. Am.*, vol. 146, no. 3, pp. 1839–1850, 2019.

[8] Liang Yu, Jerome Antoni, Haijun Wu, Quentin Leclere, and Weikang Jiang, "Fast iteration algorithms for implementing the acoustic beamforming of non-synchronous measurements," *Mech. Syst. Signal. Process.*, vol. 134, pp. 106309, 2019.

[9] Ning Chu, Yue Ning, Liang Yu, Qian Huang, and Dazhuan Wu, "A fast and robust localization method for low-frequency acoustic source: Variational bayesian inference based on nonsynchronous array measurements," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–18, 2020.

[10] Fangli Ning, Jiahao Song, Jinglong Hu, and Juan Wei, "Sound source localization of non-synchronous measurements beamforming with block hermitian matrix completion," *Mech. Syst. Signal. Process.*, vol. 147, pp. 107118, 2021.

[11] Ning Chu, Qin Liu, Liang Yu, Yue Ning, and Peng Hou, "Non-synchronous measurements of a microphone array at coprime positions," *IEEE Signal. Process. Lett.*, vol. 28, pp. 1420–1424, 2021.

[12] Misha E Kilmer and Carla D Martin, "Factorization strategies for third-order tensors," *Linear. Algebra. Appl.*, vol. 435, no. 3, pp. 641–658, 2011.

[13] Ralph Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas. Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[14] Emmanuel J Candès and Terence Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory.*, vol. 56, no. 5, pp. 2053–2080, 2010.

[15] Lieven Vandenberghe and Stephen Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[16] Zhouchen Lin, Minming Chen, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[17] Euhanna Ghadimi, André Teixeira, Iman Shames, and Mikael Johansson, "Optimal parameter selection for the alternating direction method of multipliers (admm): quadratic problems," *IEEE Trans. Automat. Contr.*, vol. 60, no. 3, pp. 644–658, 2014.

[18] Fangfang Xu and Peng Pan, "A new algorithm for positive semidefinite matrix completion," *J. Appl. Math.*, vol. 2016, 2016.

[19] Long Chen, Yat-Sze Choy, Kai-Chung Tam, and Cheng-Wei Fei, "Hybrid microphone array signal processing approach for faulty wheel identification and ground impedance estimation in wheel/rail system," *Appl. Acoust.*, vol. 172, pp. 107633, 2021.

[20] Wei Z Sun, Peng Zhang, and Bo Zhao, "Rank revealing-based tensor completion using improved generalized tensor multi-rank minimization," *IET Signal. Process.*, vol. 15, no. 8, pp. 483–499, 2021.

[21] Anh-Huy Phan, Andrzej Cichocki, Petr Tichavskỳ, Gheorghe Luta, and Austin Brockmeier, "Tensor completion throughmultiple kronecker product decomposition," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 3233–3237.

[22] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[23] Xiaoming Yuan and Junfeng Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *preprint*, vol. 12, no. 2, 2009.