JAZZ: A COMPANION TO MUSIC FOR FREQUENCY ESTIMATION WITH MISSING DATA

Qiuwei Li,[1] *Shuang Li,*[1] *Hassan Mansour,*[2] *Michael B. Wakin,*[1] *Dehui Yang,*[1] *and Zhihui Zhu*[1]

[1] Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO USA
[2] Mitsubishi Electric Research Laboratories, Cambridge, MA USA

ABSTRACT

Frequency estimation is a classical problem in signal processing, with applications ranging from sensor array processing to wireless communications and structural health monitoring. Modern algorithms based on atomic norm minimization can cope with missing data but incur a high computational cost. To recover missing data from an ensemble of frequency-sparse signals, we propose a computationally efficient low-rank tensor completion algorithm that exploits the fact that each signal in the ensemble can be associated with a Toeplitz matrix. We name our algorithm JAZZ in the spirit of the classical MUSIC algorithm for frequency estimation and in tribute to the random, improvisational nature of jazz music.

***Index Terms*—** low-rank, tensors, Toeplitz matrices, array processing, structural health monitoring, JAZZ, MUSIC

1. INTRODUCTION

1.1. Motivation

Many tasks in signal processing involve estimating frequency parameters of signals [1, 2, 3]. In applications such as narrowband direction-of-arrival (DOA) estimation in sensor array processing [4, 5, 6], wireless communications [7], and structural health monitoring [8], signals are often assumed to be a superposition of a few sinusoids. For example, in sensor array processing, the received signal is a linear combination of sinusoids which correspond to a few pre-specified signal point sources. Although many classical algorithms for parameter estimation [9] have been developed in the signal processing literature, they usually cannot handle some of the new challenges in modern data collection. For instance, due to anomalies such as sensor failure, the acquired signal may be not only noisy but also incomplete. Therefore, new approaches for robust parameter estimation are needed.

Compressive sensing (CS) has been a powerful tool for dealing with the recovery of sparse signals from compressive measurements. Its success relies on a low-dimensional prior for the underlying signal, namely, sparsity. Frequency-domain models in CS typically assume sparsity on a discrete frequency grid. However, frequency parameters in sensor arrays are continuous and can be arbitrary. Therefore, traditional CS suffers from a possibly severe basis mismatch problem [10] and is not directly relevant to the frequency estimation problems in sensor arrays. Based on this observation, [11] proposes an atomic norm minimization framework for frequency estimation from incomplete measurements, which leads to state-of-the-art results. However, one of the main drawbacks for the atomic norm formulation is that one has to solve a computationally expensive semidefinite program (SDP), which does not scale well for signals of large dimension. Furthermore, in many applications such as sensor arrays and structural health monitoring, one has to perform parameter estimation from an ensemble of signals [12, 13], which makes atomic norm minimization cumbersome.

1.2. Main contributions

In this work, we propose a scalable frequency estimation framework that can be applied on an ensemble of signals, each of which can be written as a linear combination of K complex exponentials sharing a common set of frequencies. Furthermore, the proposed method can cope with missing data. Our method is based on the fact that, in the full data case, each signal in the ensemble can be written as a low-rank Toeplitz matrix. Hence, a low-rank Toeplitz tensor can be formed when the full ensemble of signals is available. Thus, when data is missing we seek to find a low-rank tensor that supports a Toeplitz structure while obeying the measurement constraint. Based on the proposed framework, we propose a computationally efficient algorithm using alternating minimization to solve the recovery problem, which scales to high dimensional observations. Our framework can be considered as a natural generalization of the recent work [14] from the single measurement vector (SMV) case to multiple measurement vectors (MMV).

After the signal ensemble has been recovered with our algorithm, one can use a classical algorithm such as Prony's method [15] or MUSIC [16] to estimate the spectral parameters. We name our algorithm JAZZ, as it thus provides a companion to classical MUSIC and also in tribute to the random, improvisational nature of jazz music.

The rest of the paper is organized as follows. In Section 2, we formulate frequency estimation from the ensemble of signals, each of which is a superposition of K complex ex-

ponentials with missing data, as a low-rank Toeplitz tensor recovery problem. Based on our formulation, in Section 3, we propose to use an efficient alternating minimization procedure to solve the corresponding recovery problem. In Section 4, a series of numerical simulations are conducted to support our proposed approach. These experiments show that our MMV-based method outperforms the previous SMV-based algorithm.

2. PROBLEM STATEMENT

We consider the following observation model:

$$\boldsymbol{x}_m^\star = \sum_{k=1}^{K} c_{mk} \boldsymbol{e}_{f_k}$$

which is a superposition of K length-N sampled complex sinusoids, where $\boldsymbol{e}_f = [e^{j2\pi f0}\ e^{j2\pi f1}\ \cdots\ e^{j2\pi f(N-1)}]^\top$ and c_{mk} are the corresponding coefficients. The superscript $^\top$ denotes the non-conjugate transpose. Here, each signal $\boldsymbol{x}_m^\star$, $m = 1, 2, \ldots, M$, shares the same frequencies. Conceptually, we can stack the signals $\boldsymbol{x}_m^\star$ into an $N \times M$ data matrix:

$$\boldsymbol{X}^\star = \left[\boldsymbol{x}_1^\star\ \boldsymbol{x}_2^\star\ \cdots\ \boldsymbol{x}_M^\star\right],$$

which is known as the multiple measurement vector (MMV) setup in the literature [12].

In the following, we assume that a subset entries of $\boldsymbol{X}^\star$ are observed at random. Associate the set of observation indices with an $N \times M$ matrix Ω containing the value 1 at each observed position and 0 at the unobserved positions. Then, the partially observed data matrix can be written as the pointwise product of the true data matrix $\boldsymbol{X}^\star$ and the random mask Ω, i.e.,

$$\boldsymbol{Y} = \Omega \odot \boldsymbol{X}^\star,$$

or, equivalently

$$\boldsymbol{Y} = \mathcal{P}_\Omega(\boldsymbol{X}^\star),$$

where $\mathcal{P}_\Omega$ is a projection operator that zeros out entries that are not in Ω. Our goal is to use an efficient algorithm to recover the data matrix $\boldsymbol{X}^\star$ from the measurements $\boldsymbol{Y}$, after which classical algorithms may be used to estimate the frequencies $\{f_k\}_{k=1}^K$.

The main idea of this work is to leverage the underlying low-rank Toeplitz structure of the MMV problem and then to exploit this special structure when recovering the data using a structured matrix completion algorithm. To reveal this structure, suppose we rearrange each column $\boldsymbol{x}_m^\star$ of the data matrix $\boldsymbol{X}^\star$ into a Toeplitz matrix to get a third-order Toeplitz tensor $\mathcal{T}^\star \in \mathbb{C}^{N\times N\times M}$ with the m-th slice given by

$$\mathcal{T}^\star(:,:,m) = \begin{bmatrix} x_{1m} & x_{2m}^* & \cdots & x_{Nm}^* \\ x_{2m} & x_{1m} & \cdots & x_{(N-1)m}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{Nm} & x_{(N-1)m} & \cdots & x_{1m} \end{bmatrix}$$

with $1 \le m \le M$. Here a^* denotes the complex conjugate of a. By combining the celebrated Vandermonde decomposition theorem [11] and the fact that each column $\boldsymbol{x}_m^\star$ shares the same frequencies, we can see that each Toeplitz matrix slice $\mathcal{T}^\star(:,:,m)$ has a Vandermonde decomposition sharing the same Vandermonde matrix $\boldsymbol{U}$:

$$\mathcal{T}^\star(:,:,m) = \boldsymbol{U}\boldsymbol{D}_m\boldsymbol{U}^H,$$

where

$$\boldsymbol{U} = \left[\boldsymbol{e}_{f_1}\ \cdots\ \boldsymbol{e}_{f_K}\right]$$

is an $N \times K$ matrix and $\boldsymbol{D}_m$ is a $K \times K$ diagonal matrix with positive numbers $d_{1m}, \ldots, d_{Km}$ along the diagonal for all $m \in \{1, 2, \ldots, M\}$. With some simple manipulations, we obtain a CANDECOMP/PARAFAC (CP) decomposition of our Toeplitz tensor

$$\mathcal{T}^\star = \sum_{k=1}^{K} \boldsymbol{e}_{f_k} \otimes \boldsymbol{e}_{f_k} \otimes \boldsymbol{d}_k$$

where

$$\boldsymbol{d}_k = \begin{bmatrix} d_{k1} & \cdots & d_{km} \end{bmatrix}^\top.$$

This implies that the formulated Toeplitz tensor has CP rank of exactly K.

In the next section, we will leverage this structure of the formulated Toeplitz tensor to develop an efficient algorithm for recovering the data. We will use the following tensor notations. Define $\mathbb{T} : \mathbb{C}^{N\times M} \to \mathbb{C}^{N\times N\times M}$ as the operator that maps the data matrix to a Toeplitz tensor, i.e.,

$$\mathbb{T}(\boldsymbol{X}^\star) := \mathcal{T}^\star.$$

Similarly, $\mathbb{T}^* : \mathbb{C}^{N\times N\times M} \to \mathbb{C}^{N\times M}$ is defined as an operator that transforms a Toeplitz tensor into a matrix; that is, for any Toeplitz tensor $\mathcal{T} \in \mathbb{C}^{N\times N\times M}$

$$\mathbb{T}^*(\mathcal{T}) = \boldsymbol{X},$$

where the m-th column of $\boldsymbol{X}$ is obtained by taking the first column of the Toeplitz matrix $\mathcal{T}(:,:,m)$.

The tensor Frobenius norm is defined as the square root of the sum of all the squared entries of the tensor, i.e., for any $\mathcal{T} \in \mathbb{C}^{N\times N\times M}$

$$\|\mathcal{T}\|_F = \sqrt{\sum_{m=1}^{M} \|\mathcal{T}(:,:,m)\|_F^2}.$$

Given the partial observations $\boldsymbol{Y}$, we propose the following minimization program to recover the original data matrix $\boldsymbol{X}^\star$:

$$\begin{aligned} \text{minimize}_{\mathcal{L},\mathcal{T}}\ & \|\mathcal{L} - \mathcal{T}\|_F^2 \\ \text{subject to}\ & \operatorname{rank}(\mathcal{L}) \le K, \\ & \mathcal{T} \text{ is a Toeplitz tensor}, \\ & \mathcal{P}_\Omega(\mathbb{T}^*(\mathcal{T})) = \boldsymbol{Y}. \end{aligned} \tag{1}$$

Here, the rank constraint is used to enforce the low-rankness of the Toeplitz tensor and the last constraint is utilized to ensure consistency with the measurements.

In the next section, we provide an efficient algorithm to solve the low rank Toeplitz tensor completion problem.

3. THE JAZZ ALGORITHM

Inspired by [14], we utilize an alternating minimization algorithm to iteratively update $\mathcal{L}$ and $\mathcal{T}$ in solving (1). In each iteration, we update $\mathcal{T}$ (or $\mathcal{L}$) using a projected gradient descent strategy while $\mathcal{L}$ (or $\mathcal{T}$) is fixed.

Let $\mathcal{P}_K : \mathbb{C}^{N\times N\times M} \to \mathbb{C}^{N\times N\times M}$ denote an operator which finds the best rank-K approximation to a tensor. Throughout this paper, we use alternating least squares [17] to compute a best K-rank tensor approximation. Also let $\mathcal{P}_{\mathcal{T}\Omega} : \mathbb{C}^{N\times N\times M} \to \mathbb{C}^{N\times N\times M}$ denote an operator that returns the Toeplitz tensor that is closest (with respect to Ω) to an arbitrary tensor. To be precise, for any $\mathcal{T} \in \mathbb{C}^{N\times N\times M}$, $\mathcal{P}_{\mathcal{T}\Omega}(\mathcal{T})$ is equivalent to the following loop:

for $m = 1, \ldots, M$

$$(\mathcal{P}_{\mathcal{T}\Omega}(\mathcal{T}))(:,:,m) = \text{toeplitz}(\boldsymbol{z}),$$

end

where

$$z_i = \begin{cases} \mathcal{T}(i,1,m), & \Omega(i,m) = 1, \\ \frac{1}{2}\,\text{mean}(\mathcal{T}(a,b,m), a-b=i-1) + \\ \frac{1}{2}\,(\text{mean}(\mathcal{T}(a,b,m), b-a=i-1))^*, & \Omega(i,m) = 0. \end{cases}$$

Here $\text{toeplitz}(\boldsymbol{z})$ is the Toeplitz matrix whose first column is equal to $\boldsymbol{z}$.

Calculating a gradient based on Wirtinger derivatives [18], we use the following update rule

$$\begin{aligned} \mathcal{L}_{p+1} &= \mathcal{P}_K(\mathcal{L}_p - \alpha_1(\mathcal{L}_p - \mathcal{T}_p)), \\ \mathcal{T}_{p+1} &= \mathcal{P}_{\mathcal{T}\Omega}(\mathcal{T}_p - \alpha_2(\mathcal{T}_p - \mathcal{L}_{p+1})), \end{aligned}$$

where p is the iteration number, and α_1 and α_2 are step sizes. Note that $\mathcal{L}_p - \mathcal{T}_p$ (or $\mathcal{T}_p - \mathcal{L}_p$) can be viewed as a (conjugate) gradient of $\|\mathcal{L} - \mathcal{T}_p\|_F^2$ for $\mathcal{L}$ (or $\mathcal{T}$) at $\mathcal{L}_p$ (or $\mathcal{T}_p$). The full JAZZ algorithm is summarized in Algorithm 1.

4. SIMULATIONS

In this section, we present a few numerical experiments to illustrate the effectiveness of our proposed approach.

In the first experiment, we set the number of signals $M = 5$ and the dimension of each signal $\boldsymbol{x}_m^\star$ to $N = 100$. We generate $K = 4$ frequencies uniformly at random from the unit interval $[0, 1)$, ensuring a minimum separation of $1/N$. Then, we obtain 100 samples uniformly at random from the $N \times M$ data matrix $\boldsymbol{X}^\star$ and run the proposed projected gradient descent algorithm with step sizes $\alpha_1 = 1$ and $\alpha_2 = 1$. In all of the experiments, we set the number of iterations to $P = 100$. Figure 1 plots the relative distance $\|\mathcal{L}_p - \mathcal{T}_p\|_F / \|\mathcal{T}_p\|_F$ as a function of the iteration number p. We can see from the figure that our proposed algorithm converges quickly.

Algorithm 1 JAZZ Algorithm

input: incomplete observations $\boldsymbol{Y}$, observation index set Ω, stepsizes α_1 and α_2, number of sinusoids K, number of iterations P

initialize: $\mathcal{L}_0 = \boldsymbol{0}$, $\mathcal{T}_0 = \mathbb{T}(\boldsymbol{Y})$, $p = 0$

1: **while** $p < P$ **do**
2: $\quad \mathcal{L}_{p+1} = \mathcal{P}_K(\mathcal{L}_p - \alpha_1(\mathcal{L}_p - \mathcal{T}_p))$
3: $\quad \mathcal{T}_{p+1} = \mathcal{P}_{\mathcal{T}\Omega}(\mathcal{T}_p - \alpha_2(\mathcal{T}_p - \mathcal{L}_{p+1}))$
4: **end while**
5: **return** estimate of data matrix $\widehat{\boldsymbol{X}} = \mathbb{T}^*(\mathcal{T}_P)$

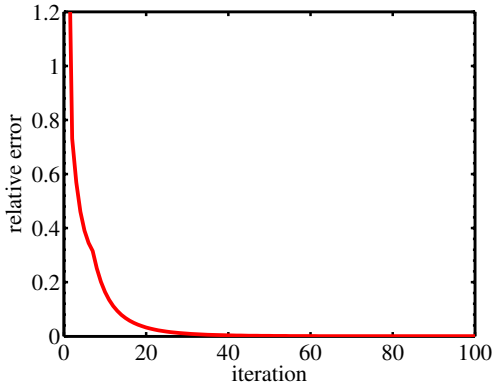

Fig. 1. Illustration of $\|\mathcal{L}_p - \mathcal{T}_p\|_F / \|\mathcal{T}_p\|_F$ against iterations.

We note that one could also recover each $\boldsymbol{x}_m^\star$ separately using the method in [14]. To show the effectiveness of our method, we compare the convergence speed of JAZZ with the algorithm in [14]. In Figure 2, we use MMV and SMV to denote our algorithm (recovering $\boldsymbol{X}^\star$ jointly) and that proposed in [14], respectively. For the purpose of this plot, each iteration of SMV corresponds to one update of all of the columns of the estimated data matrix $\boldsymbol{X}^\star$; these updates are independent from column to column. It can be seen that our joint recovery scheme converges quickly to the correct solution due to the fact that we take advantage of the joint spectrally sparse structure among all of the $\boldsymbol{x}_m^\star$; the SMV scheme fails to converge to the correct solution. In Figure 3, we show the discrete-time Fourier transform (DTFT) of one signal with full measurements, say $\boldsymbol{x}_m^\star$, and its recovered estimates using both the MMV and SMV strategies. Again, we see that the MMV performance is superior due to the fact that it takes advantage of the joint spectrally sparse structure across all of the signals.

We can ensure that the SMV algorithm converges to the

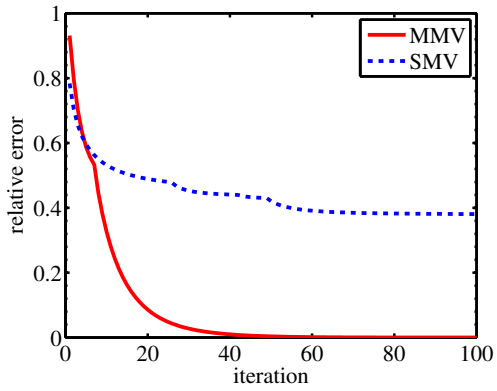

Fig. 2. Illustration of $\|\widehat{\boldsymbol{X}}_p - \boldsymbol{X}^\star\|/\|\boldsymbol{X}^\star\|$ with 100 total samples across $M = 5$ signals.

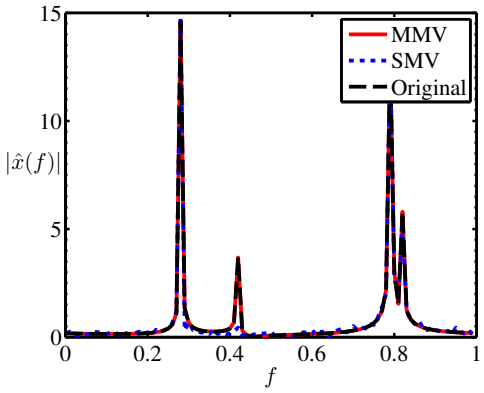

Fig. 3. The DTFT of a representative signal, and the recovered estimate using each approach.

correct solution by increasing the number of measurements. Figure 4 compares MMV and SMV when the total number of observed samples is doubled to 200. While both techniques converge, the rate is slightly faster with the MMV approach.

Finally, we study the phase transition of our proposed algorithm. In this experiment, we set $M = 5$ and $N = 30$ and vary the number of frequencies K from 1 to 10 and the total number of samples from $2M$ to MN. For each pair of K and the number of samples, 10 Monte Carlo trials are carried out with the frequencies being generated uniformly at random on the unit interval $[0, 1)$ with a minimum separation of $1/N$. For each trial, we claim signal recovery to be successful if the relative reconstruction error satisfies

$$\frac{\|\boldsymbol{X}^\star - \widehat{\boldsymbol{X}}\|_F}{\|\boldsymbol{X}^\star\|_F} \leq 0.01,$$

where we have denoted $\widehat{\boldsymbol{X}}$ as the reconstructed data matrix. Figure 5 displays the phase transition for our algorithm. We see that, as one would desire, a linear relationship exists between the number of degrees of freedom K and the number of samples when N is fixed.

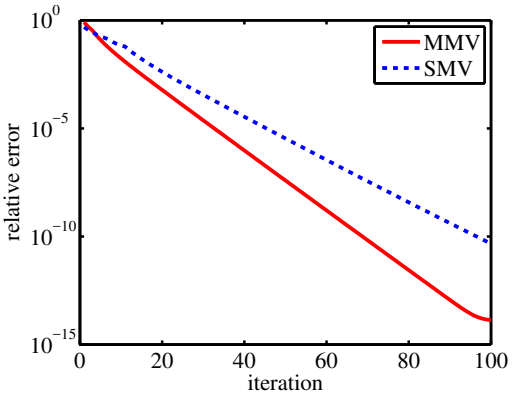

Fig. 4. Convergence comparison with 200 total samples across $M = 5$ signals.

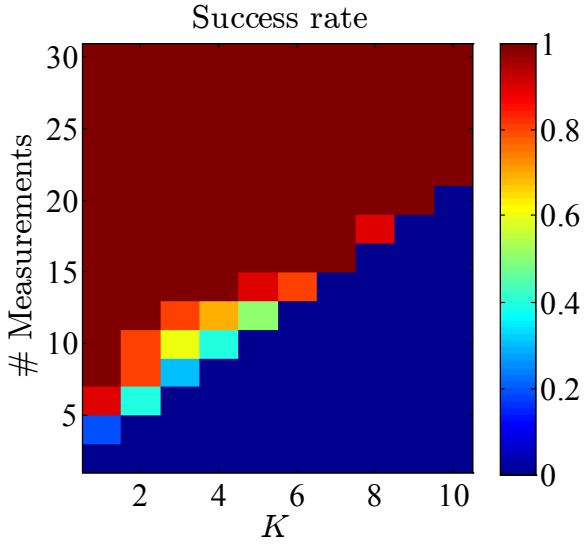

Fig. 5. Illustration of phase transition for JAZZ; the vertical axis plots the average number of measurements per column of the data matrix.

5. CONCLUSIONS

Using an iterative algorithm for recovery of low-rank Toeplitz tensors, we have extended the recent work [14] on recovery of frequency-sparse signals from the single measurement vector (SMV) case to multiple measurement vector (MMV) case. Our JAZZ algorithm serves as a companion to MUSIC, and after the signals have been reconstructed, classical spectral analysis techniques may be used for frequency estimation. In future work, we will seek to establish guarantees for the convergence of our algorithm, robustness to noise, and the requisite number of measurements for recovery.

6. REFERENCES

[1] E. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[2] B. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[3] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.

[4] H. Liao, L. Gan, and P. Wei, "A novel 2-D ESPRIT algorithm for 2-D DOA estimation," *Journal of Information and Computational Science*, vol. 8, no. 10, p. 2011, 1861.

[5] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, 1972.

[6] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[7] M. Morelli and U. Mengali, "Frequency estimation for the downlink of the UMTS-TDD component," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 554–557, 2002.

[8] F. Chang, J. F. Markmiller, J. Yang, and Y. Kim, "Structural health monitoring," *System Health Management: with Aerospace Applications*, pp. 419–428, 2011.

[9] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329–340, 1993.

[10] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[11] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[12] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[13] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *arXiv preprint arXiv:1405.6585*, 2014.

[14] J. Cai, S. Liu, and W. Xu, "A fast algorithm for reconstruction of spectrally sparse signals in super-resolution," in *SPIE Optical Engineering+ Applications*, pp. 95970A–95970A, International Society for Optics and Photonics, 2015.

[15] B. G. R. de Prony, "Essai éxperimental et analytique: Sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de lalkool, a différentes températures," *Journal de lécole polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[16] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[17] D. Nion and L. De Lathauwer, "A block factor analysis based receiver for blind multi-user access in wireless communications," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5, pp. V–V, IEEE, 2006.

[18] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.