Multidimensional Variational Line Spectra Estimation

Qi Zhang, Jiang Zhu, Ning Zhang, and Zhiwei Xu

Abstract—The fundamental multidimensional line spectral estimation problem is addressed utilizing the Bayesian methods. Motivated by the recently proposed variational line spectral estimation (VALE) algorithm, multidimensional VALE (MDVALE) is developed. MDVALE inherits the advantages of VALE such as automatically estimating the model order, noise variance and providing uncertain degrees of frequency estimates. Compared to VALE, the multidimensional frequencies of a single component is treated as a whole, and the probability density function (PDF) is projected as independent univariate von Mises distribution to perform tractable inference. Besides, for the initialization, efficient fast Fourier transform (FFT) is adopted to approximate the marginal posterior PDF of frequencies. Numerical results demonstrate the effectiveness of the MDVALE, compared to state-of-art methods.

Index Terms—Variational Bayesian inference, line spectral estimation, von Mises distribution, multidimensional frequency estimation.

I. INTRODUCTION

The problem of multidimensional frequency estimation arises in many applications. For the 2-D frequency estimation in wireless communication systems, the two dimensions correspond to the delay and angle of each path. Physical arguments and a growing body of experimental evidence suggest that utilizing the structures benefits and improves the channel reconstruction performance [1]. Another application is for the 3-D frequency estimation in radar systems, one can interpret one dimension as time delays, one dimension as Doppler shifts, and another as spatial frequencies. Since each pair of time delay, Doppler shift and spatial frequency specify the range, speed and direction of arrival (DOA) of a scatter, respectively, estimating these parameters is of great importance for target localization and tracking [2]. Traditional methods include N-D discrete Fourier transform (DFT), subspace-based approaches such as 2D MUSIC [3], 2D ESPRIT [4], matrix enhancement and matrix pencil (MEMP) [5] and the multidimensional folding (MDF) techniques [6], [7].

Manuscript received December 25, 2019; revised March 21, 2020; accepted April 16, 2020. Date of publication May 15, 2020; date of current version June 18, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61901415 and Grant 61674128. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter Jax. (Corresponding author: Jiang Zhu.)

Qi Zhang, Jiang Zhu, and Zhiwei Xu are with the Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Zhejiang University, Zhoushan 316021, China, and also with the Ocean College, Zhejiang University, Zhoushan 316021, China (e-mail: zhangqi13@zju.edu.cn; jiangzhu16@zju.edu.cn; xuzw@zju.edu.cn).

Ning Zhang is with the Nanjing Marine Radar Institute, Nanjing 211153, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: zhangn_ee@163.com).

Digital Object Identifier 10.1109/LSP.2020.2995107

Recently, compressed sensing (CS) based approaches have been proposed for N-D frequency estimation, especially for 2-D frequency estimation. Provided that the frequencies lie on the DFT grid, the signal can be recovered via efficient \(\ell_1\) minimization [8]. However, due to spectral leakage of the off-grid frequencies, the signal is not exactly sparse and conventional CS algorithms may result in significant performance degradation [9]. Consequently, grid refinement and continuous CS approaches have been developed. In [10], [11], the Newtonized orthogonal matching pursuit (NOMP) is proposed for 2-D frequency estimation, which first detects the grid and then iteratively refines the frequencies. Besides, the model order is estimated by the constant false alarm (CFAR) criterion with the knowledge of the noise variance. The computation complexity of NOMP is low, and the estimation performance is good. For the latter, [12] proposes a nuclear norm minimization over multi-fold Hankel matrices approach, [13] develops a reweighted trace minimization (RWTM) plus matrix pencil and pairing (MaPP) method, where RWTM fully exploits the multilevel Toeplitz (MLT) structure and utilizes the sparsity to obtain the covariance estimates, and MaPP calculates the Vandermonde decomposition. As [12], [13] involve solving a semidefinite programming (SDP) problem, their computation complexities are high.

A more recent approach is the variational line spectral estimation (VALE) algorithm, which uses a complete Bayesian treatment by imposing the frequencies as random variables [14]. Such method naturally allows for representing and operating with the uncertainty of the frequency estimates, in addition to only that of the weights. Numerically, accounting for the frequency uncertainty benefits the line spectral estimation performance. VALE automatically estimates the model order, noise variance, and provides the uncertain degrees of the frequency estimates. Besides, the computation is low, compared to the atomic norm minimization based approaches. It was also extended to the multi-snapshot case in array signal processing [15] and quantized setting in high-speed sampling [16]. In contrast, this work develops multidimensional VALE (MDVALE) to deal with the multidimensional line spectra estimation (LSE) problem.

The main contribution of this work is to develop the MDVALE rigorously, which is nontrivial from the following two aspects: one is that the multi frequencies are treated and updated as a whole, another aspect is that even the probability density function (PDF) of the multi frequencies may be approximated as a multivariate von Mises distribution, computing expectations of \(\prod_i e^{im_i \phi_i}\) is still hard. We novelty project the PDF of the multifrequencies as independent von Mises distribution, which allows the computation of \(\prod_i e^{im_i \phi_i}\). Besides, the N-D fast
Fourier transform (FFT) approach is adopted in the initialization to approximate the marginal posterior PDF of frequencies which can also reduce the computation complexity. Finally, numerical experiments are conducted to demonstrate the excellent performance of MDVALUE.

II. PROBLEM SETUP

Consider a measurement model for $D$ dimensional LSE problem expressed as

$$
Y = X + N, \tag{1}
$$

where $X \in \mathbb{C}^{M_1 \times \cdots \times MD}$ is the noisless line spectral composed of $K$ complex sinusoids

$$
X = \sum_{k=1}^{K} w_k A(\theta_k), \tag{2}
$$

$N$ is the additive white Gaussian noise (AWGN), i.e., $N_{MN} \sim CN(N_{MN}; 0, \nu)$ where $M = (m_1, \ldots, m_D)$. In (2), $w_k$ denotes the complex amplitude of the $k$th frequency, $A(\theta_k) \in \mathbb{C}^{M_1 \times \cdots \times MD}$ and the $M$th element is $\prod_{d=1}^D e^{j(m_{d}-1)\theta_{k,d}}$, where $\theta, \theta_k \in [-\pi, \pi]$ is the $d$th element of $\theta_k$. For simplicity, we define $\alpha(M, \theta_k) = \prod_{d=1}^D e^{j(m_{d}-1)\theta_{k,d}}$ and $\theta = [\theta_1, \ldots, \theta_K]$.

Since the number of spectral $K$ is unknown, an overcomplete model is adopted, where the number of spectral is supposed to be $N$ and $N > K$ [14]. Consequently, the binary hidden variables $s = [s_1, \ldots, s_N]^T$ are introduced to characterize whether the $i$th component is active or not. The probability mass function (PMF) is

$$
p(s; \rho) = \prod_{k=1}^N p(s_k; \rho), \quad \text{where} \ s_k \in \{0, 1\} \quad \text{and} \quad p(s_k; \rho) = \rho^{s_k}(1 - \rho)^{(1-s_k)}. \tag{3}
$$

Besides, $p(w|s; \tau) = \prod_{k=1}^N p(w_k|s_k; \tau)$, where $p(w_k|s_k; \tau)$ follows a Bernoulli-Gaussian distribution

$$
p(w_k|s_k; \tau) = (1 - \delta(s_k)) \delta(w_k) + s_k \mathcal{CN}(w_k; 0, \tau), \tag{4}
$$

where $\delta(\cdot)$ is the Dirac delta function. From (3) and (4), it can be seen that $\rho$ controls the probability of the $k$th component being active. Furthermore, uninformative prior distribution is used for the $k$th frequency, i.e., $p(\theta_k) = (1/(2\pi))^D$. For measurement model (1), the likelihood $p(Y|X; \nu)$ is

$$
p(Y|X; \nu) = \prod_{M} \mathcal{CN}(Y_M; X_M, \nu). \tag{5}
$$

Let $\beta = \{\nu, \rho, \tau\}$ and $\Phi = \{\theta, w, s\}$ be the model and estimated parameters. In general, one should first find the maximum likelihood estimates of $\beta$ by marginalizing the posterior PDF of $\Phi$, then find the maximum a posterior or minimum mean squared error estimates of $\Phi$. However, the above approach is computationally intractable and thus an iterative algorithm is developed in the ensuing section.

III. MDVALUE ALGORITHM

The variational approach tries to find a surrogate PDF $q(\Phi|Y)$ via minimizing the Kullback-Leibler (KL) divergence between $q(\Phi|Y)$ and $p(\Phi|Y)$, which is equivalent to maximize

$$
\mathcal{L}(q(\Phi|Y); \beta) = \mathbb{E}_{Q(\Phi|Y)} \left[ \ln \frac{p(Y|\Phi, \beta)}{Q(\Phi|Y)} \right], \tag{6}
$$

where

$$
p(\Phi|Y) \propto p(Y|X) \prod_{k=1}^N (p(\theta_k)p(w_k|s_k)p(s_k)) \tag{7}
$$

is the posterior PDF of variables.

Similar to [14], $q(\Phi|Y)$ is imposed to have the following structures

$$
q(\Phi|Y) = \prod_{k=1}^N q(\theta_k|Y)q(w|Y, s)q(s|Y), \tag{8}
$$

where $q(s|Y) = \delta(s - \hat{s})$.

Maximizing $\mathcal{L}(q(\Phi|Y))$ with respect to all the factors is also intractable. Let $z = [\theta_1; \ldots; \theta_N; w; s]$ be the vector of all latent variables and $z_1 = \theta_1, \ldots, z_N = \theta_N, z_{N+1} = [w; s]$. Similarly to the Gauss-Seidel method, we maximize $\mathcal{L}(q(\Phi|Y); \beta)$ (6) with respect to the posterior approximation $q(z_n|Y), n \in \{1, \ldots, N + 1\}$ with the other variables being fixed. According to [18], $q(z_n|Y)$ is

$$
\ln q(z_n|Y) = \ln p(z, z_n|Y) + \text{const}, \tag{9}
$$

where the expectation is with respect to all the variables $z$ except $z_n$ and the constant ensures normalization of the PDF.

Due to the factorization property of (8), the frequencies $\theta_k$ and $\alpha(M, \theta_k), k \in \{1, \ldots, N\}$ can be estimated from the marginal distribution $q(\theta_k|Y)$ as

$$
\hat{\theta}_{k,d} = \text{arg}[E(q(\theta_k|Y)|e^{j(m_{d}-1)\theta_{k,d}})], \quad d \in \{1, \ldots, D\}, \tag{10a}
$$

$$
\hat{\alpha}(M, \theta_k) = q(\theta_k|Y)\alpha(M, \theta_k), \quad \forall M, \tag{10b}
$$

where $\text{arg}(\cdot)$ returns the angle. Given that $q(s|Y) = \delta(s - \hat{s})$, the posterior PDF of $w$ is

$$
q(w|Y, s)\delta(s - \hat{s})ds = q(w|Y; \hat{s}). \tag{11}
$$

For the given posterior PDF $q(w|Y)$, the mean and covariance of the weights are estimated as

$$
\hat{w} = \text{E}_{q(w|Y)}[w], \tag{12a}
$$

$$
\hat{C} = \text{E}_{q(w|Y)}[ww^H] - \hat{w}\hat{w}^H. \tag{12b}
$$

where $(\cdot)^H$ denotes the Hermitian transpose operator. Let $S$ be the set of indices of the non-zero components of $s$, i.e., $S = \{k|1 \leq k \leq N, s_k = 1\}$. Let $\hat{s}$ be the estimate of $s$ and $\hat{S}$ is analogously defined based on $\hat{s}$. The model order is estimated as the cardinality of $\hat{S}$, i.e., $\hat{K} = |\hat{S}|$. According to (1), the noise-free signal is reconstructed as $\hat{X} = \sum_{k \in \hat{S}} \hat{w}_k \hat{A}_k$, where the $M$th element of $\hat{A}_k$ is $\hat{a}_{M, \theta_k}$ (10b). In the following, we detail the procedures.

A. Inferring the Frequencies

For each $k \in \{1, \ldots, N\}$, we maximize $\mathcal{L}(q(\Phi|Y); \beta)$ with respect to the factor $q(\theta_k|Y)$. For $k \notin \hat{S}$, $q(\theta_k|Y)$ remains unchanged. For $k \in \hat{S}$, substituting (10) and (12) in (9), $\ln q(\theta_k|Y)$ is obtained as

$$
\ln q(\theta_k|Y) = \ln p(\theta_k) + \sum_{M} \Re \{\eta_{k,M}a(M, \theta_k)\} + \text{const}, \tag{13}
$$

where $\eta_{k,M} = E[q(\theta_k|Y)\alpha(M, \theta_k)]$.
where \( \Re \{ \cdot \} \) returns the real part, \( f(\theta_k) \) is the exponential part of \( q(\theta_k|Y) \), and \( \eta_{k,M} \) is
\[
\eta_{k,M} = 2\nu^{-1}(\hat{Y}_M^*\hat{w}_k - \sum_{i \in S,k} \hat{a}_{M,\theta}(\hat{C}_{i,k} + \hat{w}_k\hat{w}_k^*)) \quad (14)
\]
where \((\cdot)^*\) denotes the conjugate operation. While it is hard to obtain the analytical results of (10) for the PDF (13), \( q(\theta_k|Y) \) is projected as \( D \) independent distribution, i.e.,
\[
\text{Proj}[q(\theta_k|Y)] = \prod_{d=1}^{D} f_{VM}(\theta_{k,d};\hat{\theta}_{k,d},\hat{\kappa}_{k,d}), \quad (15)
\]
where \( \text{Proj}[\cdot] \) denotes the projection operation, \( f_{VM}(\theta;\hat{\theta},\hat{\kappa}) \) denotes the von Mises distribution with mean direction \( \hat{\theta} \) and concentration parameter \( \hat{\kappa} \). In the following, the details of the parameters \( \theta_k = [\hat{\theta}_{k,1}, \ldots, \hat{\theta}_{k,D}]^T \) and \( \kappa_k = [\hat{\kappa}_{k,1}, \ldots, \hat{\kappa}_{k,D}]^T \) are presented.

Firstly, a Newton step is implemented to refine the previous estimates \( \tilde{\theta}_k^{\text{old}} \) to obtain \( \hat{\theta}_k \), i.e.,
\[
\hat{\theta}_k = \tilde{\theta}_k^{\text{old}} - (\nabla^2 f(\tilde{\theta}_k^{\text{old}}))^{-1} \nabla f(\tilde{\theta}_k^{\text{old}}), \quad (16)
\]
where \( f(\theta_k) \) is defined in (13) and \( \nabla^2 f(\theta_k) \) denotes the Hessian matrix. Then, \( \hat{\kappa}_k \) is obtained by referring to the univariate case [14]. It is shown that \( \hat{\kappa} = A^{-1} \exp\{0.5/f'(\tilde{\theta})\} \), and the detail of the inverse of \( A(\cdot) \) is given in [17]. For the multivariate case, the inverse of the Hessian matrix \( \nabla^2 f(\hat{\theta}_k) \) is obtained and \( \{\hat{\kappa}_{k,d}\}_{d=1}^{D} \) are computed separately as
\[
\hat{\kappa}_{k,d} = A^{-1} \left( \exp \left\{ \frac{1}{2} \left[ (\nabla^2 f(\hat{\theta}_k))^{-1} \right]_{d,d} \right\} \right). \quad (17)
\]
In addition, \( \tilde{a}_{M,\theta_k} \) (10b) can be approximated as
\[
\hat{a}_{M,\theta_k} \approx \prod_{d=1}^{D} \frac{I_{M,d}(\hat{\kappa}_{k,d})}{I_{0}(\hat{\kappa}_{k,d})} a(M,\hat{\theta}_k). \quad (18)
\]
In Section IV, the accuracy of the approximation (15) is demonstrated numerically.

**B. Inferring the Weights and Support**

Next \( q(\theta_k|Y), k = 1, \ldots, N \) are fixed and \( L \) is maximized w.r.t. \( q(w,s|Y) \). Define the matrix \( J \) and the vector \( h \) as
\[
J_{i,j} = \begin{cases} \prod_{d=1}^{D} M_d, & i = j, \\ \sum_{M} \tilde{a}_{M,\theta_i} \tilde{a}_{M,\theta_j}, & i \neq j, \\ \end{cases}, \quad i, j \in \{1, 2, \ldots, N\}, \quad (19a)
\]
\[
h_i = \sum_{M} Y_M \tilde{a}_{M,\theta_i}, \quad i \in \{1, 2, \ldots, N\}. \quad (19b)
\]
It can be shown that updating \( \hat{w}_S, \hat{C}_S \) and \( \hat{s} \) are the same as [14] and details are omitted here.

**C. Estimating the Model Parameters**

After updating the frequencies and weights, the model parameters \( \beta = \{\nu, \rho, \tau\} \) are estimated via maximizing the lower bound \( L(q(\Phi|Y);\beta) \) for fixed \( q(\Phi|Y) \). Plugging the postulated

**PDF (8) in (6) and setting \( \frac{\partial \nu}{\partial \nu} = 0, \frac{\partial \rho}{\partial \rho} = 0, \frac{\partial \tau}{\partial \tau} = 0 \), we have**
\[
\tilde{\nu} = \left[ \hat{w}_S^H J_S \hat{w}_S + \sum_{M} Y_M^* Y_M + \text{tr}(J_S \hat{C}_S) - 2\Re(\hat{w}_S^H h_S^*) \right],
\]
\[
\prod_{d=1}^{D} M_d \tilde{\rho} = \frac{\parallel \hat{s} \parallel_0}{N}, \quad \tilde{\tau} = \frac{\hat{w}_S^H \hat{w}_S + \text{tr}(\hat{C}_S)}{\parallel \hat{s} \parallel_0}. \quad (20)
\]

**D. Summary of MDVALSE Algorithm**

The initialization of MDVALSE algorithm is presented. For the first step of initialization, we randomly extract one dimensional data from \( Y \) and use the method in [14] to initialize \( \hat{\beta} \). For the later, we choose to initialize \( \{q(\theta_k|Y)\}_{k=1}^{N} \) in a sequential manner. For the \( k \)-th step, the marginal posterior PDF of a single sinusoid is
\[
q(\theta_k|Y) \propto \exp \left\{ \sum_{M} Y_M^* a(M,\theta_k) \right\} \left[ \left( \nu \prod_{d=1}^{D} M_d \right) \right], \quad (21)
\]
where \( Y_{r,M} = Y_{M} - \sum_{k=1}^{\hat{M}} \hat{a}_{M,\theta} \hat{w}_{i} \) is the \( M \)th element of \( Y_{r} \). To approximate \( p(\hat{\theta}_{k} | Y_{r}) \) as \( D \) independent von Mises distribution, the \( N \)-\( D \) FFT on \( Y_{r} \) is first implemented to find the peak localization of \( p(\hat{\theta}_{k} | Y_{r}) \) as the mean direction parameters, and then (17) is used to obtain the concentration parameters with \( f(\hat{\theta}_{k}) \) being replaced by the exponential part of \( p(\hat{\theta}_{k} | Y_{r}) \) (21). Now the MDVALUE is obtained and the procedures are similar to [14, Algorithm 3].

Now we analyze the computation complexity of MDVALUE. The complexity of initialization is dominated by \( D \) dimensional FFT which is \( \mathcal{O}(N \prod_{d=1}^{D} M_{d} \log(\prod_{d=1}^{D} M_{d})) \). For each iteration, the complexity is dominated by the projection of \( \{ q(\theta_k | Y) \}_{k \in B} \) as \( D \) independent von Mises distribution. The complexity is \( \mathcal{O}(K D^2 \prod_{d=1}^{D} M_{d} + \hat{K} D^3) \) which are mainly dominated by the calculation and inverse of Hessian matrix in (17). Therefore, the computational complexity of MDVALUE is \( \mathcal{O}(N \prod_{d=1}^{D} M_{d} \log(\prod_{d=1}^{D} M_{d}) + T(K D^2 \prod_{d=1}^{D} M_{d} + \hat{K} D^3)) \) where \( T \) denotes the number of whole iterations.

### IV. NUMERICAL SIMULATION

In this section, numerical experiments are conducted to evaluate the performance of MDVALUE algorithm. We define nominal signal-to-noise ratio (SNR) as \( \text{SNR} \triangleq 10 \log(\|X\|_{F}^2 / \|N\|_{F}^2) \). The Algorithm stops when \( \|\hat{X}^{(t-1)}(\cdot) - \hat{X}(\cdot)\|_{F}/\|\hat{X}^{(t-1)}(\cdot)\|_{F} < 10^{-6} \) or \( t > 500 \), where \( t \) is the number of iterations.

At first, the approximations of PDF (21) and (13) as \( D \) independent von Mises distribution are validated numerically. Note that we have approximated the PDF during initialization (21) and iteration (13). For initialization in Fig. 21(a), here the number of \( N \)-\( D \) FFT points is \( N = 1024 \). Results are shown in Fig. 1. The peak localizations along the two frequencies of the approximate PDFs are very near the true PDFs, and the peaks of the approximated PDF is a little higher than that of true PDFs. In summary, the approximated PDFs preserve the local structure of the true PDFs (13) and (21) well around the peak.

#### A. Estimation From 2D Case

In this simulation, we consider a 2D case with \( K = 8 \) signals and true frequencies are \((0.94, 1.26), (1.26, -2.51), (1.89, 1.89), (2.83, -1.26), (-2.51, 1.57), (-2.51, -2.51), (-1.57, 2.51), (-1.45, 2.76)) \). The complex weight coefficients are generated i.i.d. from complex normal distribution \( \mathcal{CN}(0, 1) \). The number of measurements is \( M_{1} = M_{2} = 10 \). As for performance comparison, reweighted trace minimization (RWTM) algorithm proposed in [13] is chosen and the performances of RWTM and MDVALUE are investigated in Fig. 2. It can be seen that for noiseless case in Fig. 2(a), MDVALUE and RWTM algorithms both estimate well. Numerically, the MSE of frequencies of RWTM is lower than that of MDVALUE. For noisy case in Fig. 2(b), MDVALUE still performs well, while RWTM overestimates the model order. This demonstrates that MDVALUE is more robust against noise than RWTM.

#### B. Estimation From 4D Case

The performance of MDVALUE algorithm for 4D case versus SNR is investigated. The normalized mean squared error (NMSE) of \( \hat{X} \), \( \text{MSE} \) of \( \hat{\theta} \) and the correct model order estimated probability \( P(\hat{K} = K) \) are adopted as the performance metrics. The frequencies estimation error is averaged over the trials in which \( \hat{K} = K \) and \( P(\hat{K} = K) \geq 0.05 \) for a given simulation point.

The magnitudes and phases of the complex weight coefficients are generated i.i.d. from normal distribution \( \mathcal{N}(1, 0.2) \) and uniform distribution \( U(-\pi, \pi) \), respectively. The number of sinusoids is \( K = 3 \). Besides, the wrap-around distance between any two frequencies is at least \( \frac{\pi}{12} \) radians for the first dimension. 300 MC trails are performed and results are shown in Fig. 3. To conduct a performance comparison, the orthogonal matching pursuit (OMP) [19] is extended to solve the multidimensional LSE problem with \( K \) being known. The algorithm is termed as MDOMP. Fig. 3(c) shows that MDVALUE works better than MDOMP in terms of frequency estimation. Also, from Fig. 3(a) the signal reconstruction error of MDVALUE is lower than that of MDOMP for \( \text{SNR} \geq 1 \text{ dB} \). As \( \text{SNR} \) increases, the successful rate of model order estimation increases and approaches to 1.

### V. CONCLUSION

This paper develops the MDVALUE for multidimensional line spectral estimation, where the multidimensional frequencies are treated as a whole and their PDFs are projected as independent von Mises distribution for tractable inference. MDVALUE inherits the advantages of automatically estimating the model order, number of nuisance parameters such as noise variance, and providing the uncertain degree of frequency estimates. Numerical results demonstrate the effectiveness of MDVALUE.
REFERENCES

[1] Y. Han, T-H. Hsu, C-K. Chao, and S. Jin, “Efficient downlink channel reconstruction for FDD multi-antenna systems,” IEEE Trans. Wireless Commun., vol. 18, no. 6, pp. 3161–3176, Jun. 2019.

[2] B. Jin, J. Zhu, Q. Wu, Y. Zhang, and Z. Xu, “One-bit LFMCW radar: Spectrum analysis and target detection,” IEEE Trans. Aerosp. Electron. Syst., to be published, doi: 10.1109/TAES.2020.2978374.

[3] Y. Hua, “A pencil-MUSIC algorithm for finding two-dimensional angles and polarizations using crossed dipoles,” IEEE Trans. Antennas Propag., vol. 41, no. 3, pp. 370–376, Mar. 1993.

[4] J. Li and R. T. Compton Jr., “Two-dimensional angle and polarization estimation using the ESPRIT algorithm,” IEEE Trans. Antennas Propag., vol. 40, no. 5, pp. 550–555, May 1992.

[5] Y. Hua, “Estimating two-dimensional frequencies by matrix enhancement and matrix pencil,” IEEE Trans. Signal Process., vol. 40, no. 9, pp. 2267–2280, Sep. 1992.

[6] X. Liu and N. D. Sidiropoulos, “Almost sure identifiability of constant modulus multidimensional harmonic retrieval,” IEEE Trans. Signal Process., vol. 50, no. 9, pp. 2366–2368, Sep. 2002.

[7] J. Liu, X. Liu, and X. Ma, “Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies,” IEEE Trans. Signal Process., vol. 55, no. 11, pp. 5179–5194, Nov. 2007.

[8] E. Candes and J. Romberg, “Sparsity and incoherence in compressive sampling,” Inverse Problems, vol. 23, no. 3, pp. 969–985, 2007.

[9] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” IEEE Trans. Signal Process., vol. 59, no. 5, pp. 2182–2195, May 2011.

[10] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive channel estimation and tracking for large arrays in mm wave picocells,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 514–527, Apr. 2016.

[11] A. Gupta, U. Madhow, and A. Arbabian, “Super-resolution in position and velocity estimation for short-range mm-wave radar,” in Proc. 50th Asilomar Conf. Signals, Syst. Comput., 2016, pp. 1144–1148.

[12] Y. Chen and Y. Chi, “Robust spectral compressed sensing via structured matrix completion,” IEEE Trans. Inf. Theory, vol. 60, no. 10, pp. 6576–6601, Oct. 2014.

[13] Z. Yang, L. Xie, and P. Stoica, “Vandermonde decomposition of multilevel toelplitz matrices with application to multidimensional super-resolution,” IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.

[14] M. A. Badiu, T. L. Hansen, and B. H. Fleury, “Variational Bayesian inference of line spectra,” IEEE Trans. Signal Process., vol. 65, no. 9, pp. 2247–2261, May 2017.

[15] J. Zhu, Q. Zhang, P. Gerstoft, M. A. Badiu, and Z. Xu, “Grid-less variational Bayesian line spectral estimation with multiple measurement vectors,” Signal Process., vol. 61, pp. 155–164, 2019.

[16] J. Zhu, Q. Zhang, and X. Meng, “Gridless variational Bayesian inference of line spectral estimation from quantized samples,” 2019, arXiv:1811.05680.

[17] K. V. Mardia and P. E. Jupp, Directional Statistics. New York, NY, USA: Wiley, 2000.

[18] K. P. Murphy, Machine Learning A Probabilistic Perspective. Cambridge, MA, USA: MIT Press, 2012.

[19] Y. C. Eldar and G. Kutyniok, Compressed Sensing: Theory and Applications. Cambridge, MA, USA: Cambridge Univ. Press, 2012.