ADMM FOR ND LINE SPECTRAL ESTIMATION USING GRID-FREE COMPRESSIVE SENSING FROM MULTIPLE MEASUREMENTS WITH APPLICATIONS TO DOA ESTIMATION

S. Semper, F. Römer

EMS Research Group, Ilmenau University of Technology

ABSTRACT

This paper is concerned with estimating unknown multi-dimensional frequencies from linear compressive measurements. This is accomplished by employing the recently proposed atomic norm minimization framework to recover these frequencies under a sparsity prior without imposing any grid restriction on these frequencies. To this end, we give a rigorous derivation of an iterative scheme called alternating direction of multipliers method, which is able to incorporate multiple compressive snapshots from a multi-dimensional superposition of complex harmonics. The key result here is how to formulate the objective function minimized by this scheme and its partial derivatives, which become hard to manage if the dimensionality of the frequencies is larger than 1. Moreover we demonstrate the performance of this approach in case of 3D line spectral estimation and 2D DOA estimation with a synthetic antenna array.

Keywords: Line Spectral Estimation, DOA Estimation, Sparse Recovery, Compressed Sensing, Optimization

I. INTRODUCTION AND SIGNAL MODEL

Direction of Arrival (DOA) estimation has been a field of active research for several decades [1] with a wide range of applications such as radar, sonar, communications, or channel sounding. Conventional techniques either exploit some algebraic structures of the underlying array manifolds or employ iterative solutions of the underlying non-convex maximum likelihood estimation problem [1].

Later, connections between the DOA estimation problem and the field of compressed sensing (CS) have been discovered [2]. Since the observed signals are sparse in the angular domain, algorithms from the field of sparse signal recovery (SSR) can be applied for DOA estimation [3], [4]. Since the angle is a continuous parameter and its discretization introduces an unwanted model mismatch [5], grid-free SSR methods are of particular interest.

However, up until recently it has not been clear how to derive recovery guarantees for compressive measurements that are more general than just randomly subselecting elements from the samples acquired in the scenario of grid-free multidimensional line spectral estimation. Moreover, these methods only allow to estimate directions of arrival from idealized arrays using measurements from uniform linear antenna arrays.

We have recently shown [6] that the DOA estimation problem with realistic antenna arrays and spatial compression can be reformulated as a line spectral estimation problem [7], where the dimensionality of the frequencies to be estimated is equal to the dimension of the angular domain. Moreover, one can formulate even higher dimensional frequency domain models, where one also aims to estimate Doppler shifts or time delays of arrival. So high dimensional frequency estimation problems with a sparsity prior arise naturally in many fields of signal processing.

A popular approach within this sparsity based framework is to formulate an optimization problem in order to retrieve the unknown frequencies, which is called atomic norm minimization (ANM). This is a direct generalization of the well studied ℓ^1 minimization, where very efficient algorithms for exact or approximate solutions exist. ANM itself is an optimization in an infinite dimensional space and as such it is infeasible to solve it directly. Instead one formulates an appropriate dual problem, which is a problem with semidefinite side constraints, for which in theory algorithms with polynomial runtime exist. In the single dimensional case, these algorithms perform well enough, but with increasing dimensions the dual problem becomes infeasibly large for these solvers, so an iterative approach, which delivers reasonable accuracy after a few steps is of high interest.

To this end an, Alternating Direction of Multipliers Method (ADMM) type [8] of optimization algorithm has already been proposed [9], [10] to solve the one dimensional line spectral estimation problem, but the derivations lack detail and explanation. Moreover, it is unclear how to generalize it to higher dimensions while also incorporating a compression step together with the multiple measurement scenario. Here we address this issue by presenting a general ADMM scheme for multi-dimensional line spectral estimation from multiple compressive measurements and we demonstrate its applicability to 2D DOA estimation.

I-A. Notation

The inner product of two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbf{C}^{n \times m}$ is defined via $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr} \boldsymbol{A}^{\mathrm{H}} \boldsymbol{B}$, the Kronecker product of two matrices $\boldsymbol{A} \in \mathbf{C}^{n \times m}$ and $\boldsymbol{B} \in \mathbf{C}^{k \times l}$ is denoted by $\boldsymbol{A} \otimes \boldsymbol{B}$, $\boldsymbol{A} \diamond \boldsymbol{B}$ on the other hand denotes the column-wise Kronecker product, $\|\boldsymbol{A}\|_F$ denotes the Frobenius norm of a matrix $\boldsymbol{A} \in \mathbf{C}^{n \times m}$ and $\|\boldsymbol{x}\|_p$ denotes the ℓ^p norm for $p \geqslant 1$ and $\boldsymbol{x} \in \mathbf{C}^n$. Moreover, we use the abbreviation for so called multi-indices $[n_1, \ldots, n_d] = \{1, \ldots, n_1\} \times \cdots \times \{1, \ldots, n_d\} \subset \mathbf{N}^d$ and $\Re(z)$ and $\Im(z)$ denote the real and imaginary part of a complex number $z \in \mathbf{C}$, whereas $\overline{z} \in \mathbf{C}$ denotes the complex conjugate, which we also use for vectors or higher order tensors element wise. For a given $\boldsymbol{x} \in \mathbf{C}^{n_1 \times \cdots \times n_d}$ we define $\operatorname{vec}(\boldsymbol{x}) \in \mathbf{C}^{n_1 \cdots n_d}$ as the rearrangement of all elements of $\boldsymbol{x}$ into a single vector.

I-B. Data Model

We observe a superposition of complex d-dimensional narrow-band harmonics with unknown amplitudes, where $d \geqslant 1$. This results in a model which read as

$$z_{\boldsymbol{k}}(t) = \sum_{i=1}^{r} s_i(t) \frac{1}{\sqrt{N_1 \cdot \cdots \cdot N_d}} \exp\left(-\jmath 2\pi \langle \boldsymbol{k}, \boldsymbol{f}_i \rangle\right)$$

for $\boldsymbol{k} \in [N_1, \ldots, N_d]$ with $N_i \geqslant 2$ for $i = 1, \ldots, d$ such that $\boldsymbol{z}(t) \in \mathbf{C}^{N_1 \times \cdots \times N_d}$. If we define

$$\boldsymbol{a}(\boldsymbol{f}) = \bigotimes_{p=1}^{d} \frac{1}{\sqrt{N_p}} \exp(-\jmath 2\pi k_p f_p)$$

$$= \frac{1}{\sqrt{N_1 \ldots N_d}} \operatorname{vec}\left(\left[\exp\left(-\jmath 2\pi \langle \boldsymbol{k}, \boldsymbol{f} \rangle\right)\right]_{\boldsymbol{k} \in [N_1, \ldots, N_d]}\right) \quad (1)$$

so that $\boldsymbol{a}(\boldsymbol{f}) \in \mathbf{C}^{M \times 1}$ we get

$$\operatorname{vec} \boldsymbol{z}(t) = \sum_{i=1}^{r} s_i(t) \boldsymbol{a}(\boldsymbol{f}_i). \quad (2)$$

If we now collect multiple measurements at K points in time $z(t_k)$, we can write
$$\boldsymbol{Z} = [\mathrm{vec}\,\boldsymbol{z}(t_1), \ldots, \mathrm{vec}\,\boldsymbol{z}(t_K)] = \boldsymbol{A}(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_r)\boldsymbol{S}(t_1, \ldots t_K)$$
for $\boldsymbol{Z} \in \mathbf{C}^{M \times K}$, $\boldsymbol{A}(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_r) \in \mathbf{C}^{M \times r}$ ans $\boldsymbol{S} \in \mathbf{C}^{r \times K}$, where $M = N_1 \cdots N_d$. In order to model a compressive sensing scenario, we assume that in stead of observing $\boldsymbol{Z}$ directly, we model our observations as a collection of linear measurements applied to $\boldsymbol{Z}$. These are represented by a combining matrix $\boldsymbol{\Phi} \in \mathbf{C}^{m \times M}$ for some $m \leqslant M$ and thus our final model reads as
$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{Z} + \boldsymbol{N} = \boldsymbol{\Phi}\boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}, \tag{3}$$
where $\boldsymbol{N} \in \mathbf{C}^{m \times K}$ accounts for additive measurement noise. For this scenario it has already been shown in [7] that sparse recovery techniques like ANM can recover the underlying frequencies from compressive measurements with high probability when imposing a separation condition and in case of a suitably generated compression matrix $\boldsymbol{\Phi}$, i.e. where the entries are drawn i.i.d. from a sub-Gaussian distribution.

I-C. Application to DOA estimation

The above model for frequency estimation can be used to cast the problem of 2-d DOA estimation with realistic antenna elements. To this end, let $\boldsymbol{r}(\boldsymbol{\theta}) : [0, 2\pi] \times [0, \pi] \to \mathbf{C}^M$ model the response of an array comprising of M antennas for a planar wave impinging from azimuth and elevation angle $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Naturally, each element $r_m(\boldsymbol{\theta})$, $m = 1, \ldots, M$ is a periodic function in θ_1 and most importantly can be measured in practice for a specific antenna array geometry. Moreover, since beam patterns are typically quite smooth functions, they can be very well approximated by a truncated Fourier series [11] given by
$$r_m(\boldsymbol{\theta}) \approx \frac{1}{\sqrt{L_1 L_2}} \sum_{\ell_1 = -\frac{L_1-1}{2}}^{\frac{L_1-1}{2}} \sum_{\ell_2 = -\frac{L_2-1}{2}}^{\frac{L_2-1}{2}} g_{m,\ell_1,\ell_2} e^{j(\theta_1 \ell_1 + \theta_2 \ell_2)}, \tag{4}$$
where we have considered odd numbers of L_1 and L_2 terms respectively. Rewriting above formula in matrix form yields
$$r_m(\boldsymbol{\theta}) \approx \boldsymbol{a}(\theta_1)^{\mathrm{T}} \cdot \boldsymbol{G}_m \cdot \boldsymbol{a}(\theta_2) = \mathrm{vec}(\boldsymbol{G}_m) \cdot (\boldsymbol{a}(\theta_1) \otimes \boldsymbol{a}(\theta_2)),$$
where the matrix $\boldsymbol{G}_m$ collects the Fourier coefficients of $r_m(\boldsymbol{\theta})$. If we now observe S plain narrowband waves from the far field of the array with unknown time varying amplitudes $s_1(t), \ldots s_S(t)$ and unknown directions of arrival $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_S$ and define
$$\boldsymbol{G} = \left[\mathrm{vec}(\boldsymbol{G}_1)^{\mathrm{T}}, \ldots, \mathrm{vec}(\boldsymbol{G}_m)^{\mathrm{T}}\right]^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{A} = \boldsymbol{A}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_S)$$
we arrive at the following model for the measurements we collect at the M ports of the antenna array
$$\boldsymbol{z}(t) = \boldsymbol{G}\boldsymbol{A}\boldsymbol{s}(t) \in \mathbf{C}^{M \times 1}.$$
If we now also take snapshots $\boldsymbol{z}(t_1), \ldots, \boldsymbol{z}(t_K)$ for $K \in \mathbf{N}$, apply spatial compression as was recently proposed in [6] for the one-dimensional case, we get
$$\boldsymbol{Y} = \boldsymbol{\Psi}\boldsymbol{Z} + \boldsymbol{N} = \boldsymbol{\Psi}\boldsymbol{G}\boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}, \tag{5}$$
where again $\boldsymbol{N}$ accounts for additive measurement noise. So we arrive at a 2-dimensional line spectral estimation problem from multiple compressed snapshots as in (3) if we set $\boldsymbol{\Phi} = \boldsymbol{\Psi}\boldsymbol{G}$ and any technique to recover the unknown frequencies in (3) is suitable to recover the unknown directions of arrival. This task and means to tackle it by using a sparse recovery approach are subject of the next section.

II. GRID-FREE RECONSTRUCTION

In the recent literature [12], [13] it has been put forward to use ANM as a generalization of the conventional ℓ^1-norm minimization in compressive sensing when one aims at recovering parameters, which are not constrained to be resided on a discrete grid but rather on a continuous manifold. The next section introduces the notion of the atomic norm in the multiple snapshot d-dimensional line spectral estimation scenario and the corresponding optimization problem for estimating these unknown frequencies from compressive measurements.

Definition II.1 (atomic norm). Let $\mathcal{A} \subset \mathbf{C}^N$ be an arbitrary set. Then, the **atomic norm** $\|\cdot\|_{\mathcal{A}} : \mathbf{C}^N \to \mathbf{R}_0^+$ is defined as
$$\boldsymbol{X} \mapsto \|\boldsymbol{X}\|_{\mathcal{A}} = \inf \left\{ \sum_{i=1}^r |c_i| \,\bigg|\, \boldsymbol{X} = \sum_{i=1}^r c_i \boldsymbol{a}_i, \quad \boldsymbol{a}_i \in \mathcal{A} \right\}. \tag{6}$$

As it turns out, this is a direct generalization of the conventional ℓ^1 norm $\|\boldsymbol{u}\|_{\mathcal{A}} = \|\boldsymbol{x}\|_1$ of a vector $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$ if one takes $\mathcal{A}$ to be the set of columns of $\boldsymbol{A}$. Next, following [14] we specialize the general atomic set from above such that it fits the structure of d-dimensional line spectral estimation from K measurements:

Definition II.2 (atomic set). Let $\boldsymbol{a} : (0,1]^d \to \mathbf{C}^{N_1 \times \cdots \times N_d}$ be defined as in (1) then the **atomic set** for K snapshots is defined as
$$\mathcal{A} = \left\{ \boldsymbol{a}(\boldsymbol{f})\boldsymbol{b}^{\mathrm{H}} \,\middle|\, \boldsymbol{b} \in \mathbf{C}^K, \|\boldsymbol{b}\|_2 = 1, \boldsymbol{f} \in (0,1]^d \right\} \subset \mathbf{C}^{N_1 \ldots N_d \times K}.$$

In some way the above set captures the model in the sense that it represent all possibly occuring harmonics in the multiple snapshot case, so our signal in (2) is comprised of a sparse and linear superposition of elements in above atomic set. To extract the parameters corresponding to these atoms we aim at solving a grid free sparse recovery problem, where we minimize the atomic norm for the above defined set $\mathcal{A}$.

So, the problem for atomic norm minimization as proposed in [10] and extended to compressive measurements in [6] reads as
$$\min_{\boldsymbol{Z} \in \mathbf{C}^{M \times K}} \|\boldsymbol{Z}\|_{\mathcal{A}} \quad \text{subject to} \quad \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{Z}\|_F^2 \leqslant \varepsilon,$$
for suitably chosen ε, where we additionally employ a compression matrix $\boldsymbol{\Phi} \in \mathbf{C}^{m \times M}$ as before, which we can choose freely. Now, still following [10] and [14] we aim at posing the equivalent dual problem, since directly minimizing the atomic norm is an infinite dimensional optimization problem. For this we need the concept of specially structured matrices, which are defined as follows.

Definition II.3 (multilevel Toeplitz matrices). Let for $d \in \mathbf{N}$ $\boldsymbol{u} \in \mathbf{C}^{n_1 \times 2 \cdot n_2 - 1 \times \cdots \times 2 \cdot n_d - 1}$ be the tensor of defining elements and $\boldsymbol{n} \in [n_1, \ldots, n_d]$ be a vector of dimension sizes. Now, the **Hermitian d-level Toeplitz matrix** $\boldsymbol{T}_{(n,d)}^{\mathrm{H}}(\boldsymbol{u})$ is recursively defined with a blockwise Toeplitz structure as
$$\boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}) = \begin{bmatrix} \boldsymbol{T}_{(\boldsymbol{m},\ell)}^{\mathrm{H}}(\boldsymbol{u}_{[1,\boldsymbol{m}]}) & \ldots & \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[n_1,\boldsymbol{m}]}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[n_1,\boldsymbol{m}]})^{\mathrm{H}} & \ldots & \boldsymbol{T}_{(\boldsymbol{m},\ell)}^{\mathrm{H}}(\boldsymbol{u}_{[1,\boldsymbol{m}]}) \end{bmatrix}$$
where $\boldsymbol{m} = [n_2, \ldots, n_d]$ and $\ell = d - 1$. Moreover a **d-level Toeplitz matrix** is defined with above notation and again $\boldsymbol{u} \in \mathbf{C}^{2 \cdot n_1 - 1 \times 2 \cdot n_2 - 1 \times \cdots \times 2 \cdot n_d - 1}$ being the tensor of defining elements via
$$\boldsymbol{T}_{(\boldsymbol{n},d)}(\boldsymbol{u}) = \begin{bmatrix} \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[1,\boldsymbol{m}]}) & \ldots & \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[2 \cdot n_1 - 1,\boldsymbol{m}]}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[n_1,\boldsymbol{m}]}) & \ldots & \boldsymbol{T}_{(\boldsymbol{m},\ell)}(\boldsymbol{u}_{[1,\boldsymbol{m}]}). \end{bmatrix}$$

With these structured matrices at hand, we can reformulate the calculation of the atomic norm in the line spectral estimation case.

Theorem II.1 ([15], [13]). *With $\mathcal{A}$ given as in Definition II.2 the following equality holds:*
$$\|\boldsymbol{X}\|_{\mathcal{A}} = \min_{\boldsymbol{W},\boldsymbol{u}} \mathrm{tr}\, \boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}) + \mathrm{tr}\, \boldsymbol{W} \tag{7}$$
$$\textit{subject to} \quad \begin{bmatrix} \boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^{\mathrm{H}} & \boldsymbol{W} \end{bmatrix} \succeq \boldsymbol{0}.$$

The above theorem transforms the infinite dimensional problem of calculating the atomic norm of multiple snapshots $\boldsymbol{X}$ into a semidefinite optimization program, which can be solved efficiently in theory. With this result at hand, we can now pose the dual problem of atomic norm minimization for generalized line spectral estimation with the model in (3) via
$$\min_{\boldsymbol{W},\boldsymbol{u},\boldsymbol{X}} \mathrm{tr}\, \boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}) + \mathrm{tr}\, \boldsymbol{W} \tag{8}$$

subject to $\begin{bmatrix} \boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}) & \boldsymbol{Z} \\ \boldsymbol{Z}^{\mathbf{H}} & \boldsymbol{W} \end{bmatrix} \succeq \boldsymbol{0}, \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{Z}\|^2_F \leqslant \varepsilon,$

which can for instance be found in [16]. The key point here is now that the resulting $\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}^*)$ for a dual optimal $\boldsymbol{u}^*$ is an estimate of the covariance of the underlying signal. So the final step would be to apply any covariance based spectral estimator, like MUSIC, ESPRIT or a Vandermonde decomposition to $\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}^*)$.

Although we succeeded in reducing the original problem complexity significantly, for larger dimensions, so $d \geqslant 2$ the explicit solvers for semidefinite programs take prohibitively many iterations to deliver feasible results if one is able to provide it with the side constraints in a reasonable way at all. So a more direct and iterative approach would be advantageous, which is also specifically tailored to the semidefinite program at hand. To this end, we aim at providing an ADMM type algorithm, which approximates a solution to above problem reasonably well and reasonably fast.

III. DERIVATION OF THE ADMM

Next, we formulate the iterative update steps of the ADMM in order to approximate a solution to (8). To this end and following [8], we make use of the augmented Lagrangian of the problem in (3), which reads as

$$\min_{\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z},\boldsymbol{V} \succeq \boldsymbol{0},\boldsymbol{\Lambda} \succeq \boldsymbol{0}} \mathcal{L}(\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z},\boldsymbol{V},\boldsymbol{\Lambda}) \qquad (9)$$

where $\mathcal{L}(\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z},\boldsymbol{V},\boldsymbol{\Lambda}) : \mathbf{C}^{K \times K} \times \mathbf{C}^{N_1 \times 2N_2-1 \cdots \times 2N_d-1} \times \mathbf{C}^{M \times K} \times \mathbf{C}^{M+K \times M+K} \times \mathbf{C}^{M+K \times M+K} \to \mathbf{R}$ and its values are defined by

$$\mathcal{L}(\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z},\boldsymbol{V},\boldsymbol{\Lambda}) = \left\langle \boldsymbol{\Lambda}, \boldsymbol{V} - \underbrace{\begin{bmatrix} \boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}) & \boldsymbol{Z} \\ \boldsymbol{Z}^{\mathbf{H}} & \boldsymbol{W} \end{bmatrix}}_{=\boldsymbol{T}} \right\rangle$$
$$+ \frac{1}{2}\|\boldsymbol{\Phi}\boldsymbol{Z} - \boldsymbol{Y}\|^2_2 + \frac{\tau}{2}(\operatorname{tr}\boldsymbol{W} + \operatorname{tr}\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}))$$
$$+ \frac{\rho}{2}\left\|\boldsymbol{V} - \begin{bmatrix} \boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}) & \boldsymbol{Z} \\ \boldsymbol{Z}^{\mathbf{H}} & \boldsymbol{W} \end{bmatrix}\right\|^2_F,$$

where $\tau > 0$ and $\rho > 0$ are suitably chosen constants. Here τ plays the role of a regularizing parameter between data fitting and the magnitude of the atomic norm of $\boldsymbol{Z}$. We also partition the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{V}$ such that they match the partitioning of the blocks in $\boldsymbol{T}$:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \hat{\boldsymbol{\Lambda}} & \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{bmatrix} \hat{\boldsymbol{V}} & \boldsymbol{V}_1 \\ \boldsymbol{V}_1 & \boldsymbol{V}_0 \end{bmatrix}.$$

The objective of the next paragraphs is to calculate the partial derivatives of $\mathcal{L}$, which has to be handled carefully, since we have to consider the fact that the variables $\mathcal{L}$ depends on are complex valued and highly structured.

III-A. Wirtinger Calculus

To calculate the partial derivatives of functions which depend on complex variables and map to $\mathbf{R}$ we make use of the so called Wirtinger calculus. To this end, let f be a function $f : \mathbf{C}^n \mapsto \mathbf{R}$, then its Wirtinger derivative is defined as

$$\frac{\partial f}{\partial \boldsymbol{x}} = \frac{1}{2}\left(\frac{\partial f}{\partial \boldsymbol{y}} - \jmath\frac{\partial f}{\partial \boldsymbol{z}}\right) \quad \text{and} \quad \frac{\partial f}{\partial \overline{\boldsymbol{x}}} = \frac{1}{2}\left(\frac{\partial f}{\partial \boldsymbol{y}} + \jmath\frac{\partial f}{\partial \boldsymbol{z}}\right),$$

where $\boldsymbol{y} = \operatorname{Re}\{\boldsymbol{x}\}$ and $\boldsymbol{z} = \operatorname{Im}\{\boldsymbol{x}\}$. Since we have specifically structured functions and matrices in $\mathcal{L}$, we need only the following three simple rules

$$\frac{\partial\langle\boldsymbol{b},\boldsymbol{a}\rangle}{\partial\overline{\boldsymbol{b}}} = \frac{\partial\boldsymbol{b}^{\mathbf{H}}\boldsymbol{a}}{\partial\overline{\boldsymbol{b}}} = \boldsymbol{a}, \quad \frac{\partial\langle\boldsymbol{a},\boldsymbol{b}\rangle}{\partial\overline{\boldsymbol{b}}} = \frac{\partial\boldsymbol{a}^{\mathbf{H}}\boldsymbol{b}}{\partial\overline{\boldsymbol{b}}} = \boldsymbol{0} \qquad (10)$$

and $\quad \dfrac{\partial\|\boldsymbol{A}\boldsymbol{b} - \boldsymbol{c}\|^2}{\partial\overline{\boldsymbol{b}}} = \boldsymbol{A}^{\mathbf{H}}(\boldsymbol{A}\boldsymbol{b} - \boldsymbol{c}). \qquad (11)$

All three of them can easily be extended to the case where $\boldsymbol{a}$ and $\boldsymbol{b}$ are matrices, since both the inner product of matrices we use here and the Frobenius norm induced by it treat matrices as if they were vectors realigned into matrices.

III-B. Special Derivatives

Taking a close look at $\mathcal{L}$ we see that the only derivatives which are not straightforward to calculate are those with respect to $\boldsymbol{u}$, since it is the defining tensor of the multilevel Toeplitz structure. In this case the expressions of interest are

$$\frac{\partial}{\partial\boldsymbol{u}}\langle\boldsymbol{A},\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})\rangle \quad \text{and} \quad \frac{\partial}{\partial\boldsymbol{u_e}}\langle\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}),\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})\rangle$$

for a given Hermitian matrix $\boldsymbol{A}$. To this end, for given $n \in \mathbf{N}$ and $p \in [n] \cup -[n] \cup \{0\}$, we define $\boldsymbol{S}^p_n$ via
$[\boldsymbol{S}^p_n]_{(k,\ell)} = 1 \quad \text{for} \quad \ell - k = p - n.$
Note that $\boldsymbol{S}^p_n$ has 1 only on a shifted diagonal and $\boldsymbol{S}^0_n = \boldsymbol{I}_n$. Now, we can rewrite
$$\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}) = \boldsymbol{T}_{\text{upper}} + \boldsymbol{T}_{\text{lower}} \qquad (12)$$
where $\boldsymbol{T}_{\text{upper}}$ and $\boldsymbol{T}_{\text{lower}}$ are constructed by explicitly unraveling the recursive definition of $\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})$ while keeping the Hermitian symmetry in mind. So they are defined as
$$\boldsymbol{T}_{\text{upper}} = \sum_{\boldsymbol{i} \in \mathcal{N}} \left(\boldsymbol{S}^{i_1-1}_{N_1} \otimes \boldsymbol{S}^{i_2-N_2}_{N_2} \otimes \cdots \otimes \boldsymbol{S}^{i_d-N_d}_{N_d}\right)u_{\boldsymbol{i}} \qquad (13)$$

$$\boldsymbol{T}_{\text{lower}} = \boldsymbol{T}^{\mathbf{H}}_{\text{upper}}, \qquad (14)$$
where $\mathcal{N} = [N_1, 2N_2 - 1, \ldots, 2N_d - 1]$. Next, we calculate for given multi-index $\boldsymbol{i} \in \mathcal{N}$
$$\frac{\partial}{\partial u_{\boldsymbol{i}}}\langle\boldsymbol{A},\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})\rangle = \qquad (15)$$
$$= \left\langle\boldsymbol{A}, \boldsymbol{S}^{i_1-1}_{N_1} \otimes \boldsymbol{S}^{i_2-N_2}_{N_2} \otimes \cdots \otimes \boldsymbol{S}^{i_d-N_d}_{N_d}\right\rangle. \qquad (16)$$
For a shorter notation we define the operator $\mathfrak{D}_{\boldsymbol{n},d} : \mathbf{C}^{M \times M} \to \mathbf{C}^{N_1 \times 2N_2-1 \cdots \times 2N_d-1}$ via
$$\boldsymbol{A} \mapsto \mathfrak{D}_{\boldsymbol{n},d}(\boldsymbol{A}) = \left(\frac{\partial}{\partial u_{\boldsymbol{i}}}\langle\boldsymbol{A},\boldsymbol{T}_{\boldsymbol{n},d}(\boldsymbol{u})\rangle\right)_{\boldsymbol{i} \in \mathcal{N}} =$$
$$= \left(\left\langle\boldsymbol{A}, \boldsymbol{S}^{i_1-1}_{N_1} \otimes \boldsymbol{S}^{i_2-N_2}_{N_2} \otimes \cdots \otimes \boldsymbol{S}^{i_d-N_d}_{N_d}\right\rangle\right)_{\boldsymbol{i} \in \mathcal{N}}.$$
This operator results in a tensor with the same dimensions as $\boldsymbol{u}$ and each entry at position $\boldsymbol{i} \in \mathcal{N}$ represents the sum of the elements in $\boldsymbol{A}$ which occur at the same position as $u_{\boldsymbol{i}}$ in $\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})$.

Now for some $\boldsymbol{i} \in \mathcal{N}$ and $\boldsymbol{u} = \boldsymbol{v} + \jmath\boldsymbol{w}$ we can also calculate (Note that here we identify the multi-index $\boldsymbol{i}$ with the the tensor of order d which has zeros everywhere except a single 1 at position $\boldsymbol{i}$.):
$$\frac{\partial}{\partial u_{\boldsymbol{i}}}\langle\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u}),\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})\rangle = \frac{\partial}{\partial u_{\boldsymbol{i}}}\left\|\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})\right\|^2_F =$$
$$= \frac{\partial}{\partial u_{\boldsymbol{i}}}\left(\sum_{\boldsymbol{i}' \in \mathcal{N}}\left[\left\|\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{i}')\right\|^2_F v^2_{\boldsymbol{i}'} - \left\|\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{i}')\right\|^2_F w^2_{\boldsymbol{i}'}\right]\right) =$$
$$= 2f_{\boldsymbol{n}}(\boldsymbol{i})\overline{u}_{\boldsymbol{i}},$$
where $f_{\boldsymbol{n}}(\boldsymbol{i})$ represents the number of occurrences of $u_{\boldsymbol{i}}$ in the Hermitian multilevel Toeplitz matrix $\boldsymbol{T}^{\mathbf{H}}_{(n,d)}(\boldsymbol{u})$.

With this intuition at hand, we can easily infer that $(f_{\boldsymbol{n}}(\boldsymbol{i}))_{\boldsymbol{i} \in \mathcal{N}} = \mathfrak{D}_{\boldsymbol{n},d}(\boldsymbol{1})$, where $\boldsymbol{1} \in \mathbf{C}^{M \times m}$ is a matrix with all entries equal to 1. Now, these rules together with (10) and (11) can be used to establish the following results about the partial derivatives of $\mathcal{L}$ which read as:
$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{W}} = \frac{\tau}{2}\boldsymbol{I}_K - \boldsymbol{\Lambda}_0 - \rho(\boldsymbol{V}_0 - \boldsymbol{W}), \qquad (17)$$
$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{u}} = \frac{\tau}{2}\boldsymbol{i}_1 - \mathfrak{D}_{\boldsymbol{n},d}(\hat{\boldsymbol{\Lambda}}) + \frac{\rho}{2}\left(\mathfrak{D}_{\boldsymbol{n},d}(\boldsymbol{1}) - 2\mathfrak{D}_{\boldsymbol{n},d}(\hat{\boldsymbol{V}})\right), \qquad (18)$$
$$\frac{\partial\mathcal{L}}{\partial\overline{\boldsymbol{Z}}} = \frac{1}{2}(\boldsymbol{\Phi}^{\mathbf{H}}\boldsymbol{\Phi}\boldsymbol{Z} - \boldsymbol{\Phi}^{\mathbf{H}}\boldsymbol{Y}) - \hat{\boldsymbol{\Lambda}} - \rho\left(\hat{\boldsymbol{V}} - \boldsymbol{Z}\right), \qquad (19)$$
where $\boldsymbol{i}_1$ is the tensor of the same dimension as $\boldsymbol{u}$ with entries all equal to 0 except at the position of $u_{[1,\ldots,1]}$. With these three derivatives at hand we can proceed to formulate the explicit update steps for the ADMM iteration.

III-C. Update Steps

This section gives the explicit updates rules for the alternating updates of the ADMM. In general and according to [10] the iteration after step k can be expressed as
$$(\boldsymbol{W}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{Z}^{k+1}) \leftarrow \operatorname*{argmin}_{\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z}}\mathcal{L}(\boldsymbol{W},\boldsymbol{u},\boldsymbol{Z},\boldsymbol{V}^k,\boldsymbol{\Lambda}^k)$$

$$\boldsymbol{V}^{k+1} \leftarrow \underset{\boldsymbol{V} \succeq 0}{\arg\min} \, \mathcal{L}(\boldsymbol{W}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{Z}^{k+1}, \boldsymbol{V}, \boldsymbol{\Lambda}^k)$$

and
$$\boldsymbol{\Lambda}^{k+1} = \boldsymbol{\Lambda}^k + \rho(\boldsymbol{V}^{k+1} - \boldsymbol{T}^{k+1})$$

By using the derivatives of the previous section in equations (17), (18) and (19) we can compute above updates in closed form by setting the respective derivatives to 0, thus satisfying the necessary conditions for optimality, and solving for the variable of interest. Finally, they read as follows

$$\boldsymbol{u}^{k+1} = \frac{1}{\rho \mathfrak{D}_{\boldsymbol{n},d}(\boldsymbol{1})} \left(\mathfrak{D}_{\boldsymbol{n},d}(\hat{\boldsymbol{\Lambda}}^k - \rho \hat{\boldsymbol{V}}^k) - \frac{\tau}{2}\boldsymbol{i}_1 \right) \tag{20}$$

$$\boldsymbol{W}^{k+1} = \frac{1}{\rho}\boldsymbol{\Lambda}_0^k + \boldsymbol{V}_0^k - \boldsymbol{I}\frac{\tau}{2\rho} \tag{21}$$

$$\boldsymbol{Z}^{k+1} = \left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} - \rho\boldsymbol{I} \right)^{-1} \cdot \left(\boldsymbol{\Phi}\boldsymbol{Y} + \boldsymbol{\Lambda}_1^k + \rho\boldsymbol{V}_1^k \right), \tag{22}$$

where $\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} - \rho\boldsymbol{I} \right)^{-1}$ can be precomputed in advance to avoid repeatedly solving a linear system. Furthermore we also update $\boldsymbol{T}$ iteratively from the current variables according to

$$\boldsymbol{T}^{k+1} = \begin{bmatrix} \boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}^{k+1}) & \boldsymbol{Z}^{k+1} \\ \boldsymbol{Z}^{\mathrm{H}\,k+1} & \boldsymbol{W}^{k+1} \end{bmatrix} \tag{23}$$

whereas we finally can also update $\boldsymbol{V}$ via

$$\boldsymbol{V}^{k+1} = \mathfrak{P}^{\succeq} \left(\boldsymbol{T}^{k+1} - \rho\boldsymbol{\Lambda}^k \right), \tag{24}$$

where $\mathfrak{P}^{\succeq}$ is the orthogonal projection onto the positive cone of the positive semi-definite matrices, which can be realized numerically by an eigendecomposition and is the computationally most expensive step during the iterate updates. With this we have completed the iterations for ADMM which now only consists of initializing $(\boldsymbol{W}^0, \boldsymbol{Z}^0, \boldsymbol{u}^0, \boldsymbol{V}^0, \boldsymbol{\Lambda}^0)$ and iteratively carrying out the steps in (20), (21), (22), (23) and (24). With the operators defined as above, which can be implemented recursively for arbitrary dimensions, one can create a very general implementation as well. Next, we proceed with the numerical evaluation.

IV. NUMERICAL SIMULATIONS

To quantify the performance of our estimator, we compare it to the deterministic Cramér-Rao Bound (CRB). With the proposed spatial compression, the deterministic CRB for the d-dimensional case with K snapshots can be computed via [17]

$$C(\boldsymbol{\theta}) = \frac{\sigma^2}{2 \cdot K} \operatorname{tr}\left(\left[\Re(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{G}}^{\perp}\boldsymbol{D} \odot (\boldsymbol{1}_{d \times d} \otimes \hat{\boldsymbol{R}})^{\mathrm{T}}) \right]^{-1} \right), \tag{25}$$

with $\boldsymbol{\Pi}_{\boldsymbol{G}}^{\perp} = \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{H}}$ and $\hat{\boldsymbol{R}} = 1/K \cdot \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}}$ being the sample covariance and we have set

$$\boldsymbol{G} = \boldsymbol{\Phi}\boldsymbol{A}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_S), \; \boldsymbol{D}_i = \frac{\partial}{\partial\boldsymbol{\theta}_i}\boldsymbol{G}, \; \boldsymbol{D} = [\boldsymbol{D}_1, \ldots, \boldsymbol{D}_d]$$

First, we carry out the atomic norm minimization for 3D line spectral estimation with uncompressed measurements, so $\boldsymbol{\Phi} = \boldsymbol{I}$ in (3) and also compressed measurements, where the entries of $\boldsymbol{\Phi}$ are drawn i.i.d. from a zero-mean Gaussian distribution and then we project the columns to the complex unit sphere in $\mathbb{C}^m$, thus normalizing the columns independently. Here, we chose $m = \lfloor \rho \cdot M \rfloor$ according to some compression rate $\rho \in (0, 1]$. In case of $\boldsymbol{\Phi} = \boldsymbol{I}$, so $\rho = 1$, we also run 3D-Standard-ESPRIT [18] directly on $\boldsymbol{Y}$ as a comparison, which is only applicable in this case, since ESPRIT is not able to deal with compressive measurements of the kind employed here. In any case, we choose $\tau = \sigma^{0.8}$ and $\rho = 0.05$ to run the ADMM and we initialized the state variables $(\boldsymbol{W}^0, \boldsymbol{u}^0, \boldsymbol{Z}^0, \boldsymbol{V}^0, \boldsymbol{\Lambda}^0)$ by sampling the real and imaginary parts from standard Gaussian distributions.

The results in Figure 1, where we plot the reconstruction error versus the noise variance σ^2, show that the derived ADMM approach is able to replicate the performance predicted by the CRB for the case $\rho = 1$, thus delivering the same performance as 3D-Standard-ESPRIT.

In the case $\rho = 0.75$ we see that the ADMM algorithm's performance highly depends on the number of steps carried out to estimate the covariance $\boldsymbol{T}_{(\boldsymbol{n},d)}^{\mathrm{H}}(\boldsymbol{u}^*)$, since the error floor decreases when iterating for 1000 steps instead of 100. In conclusion, this
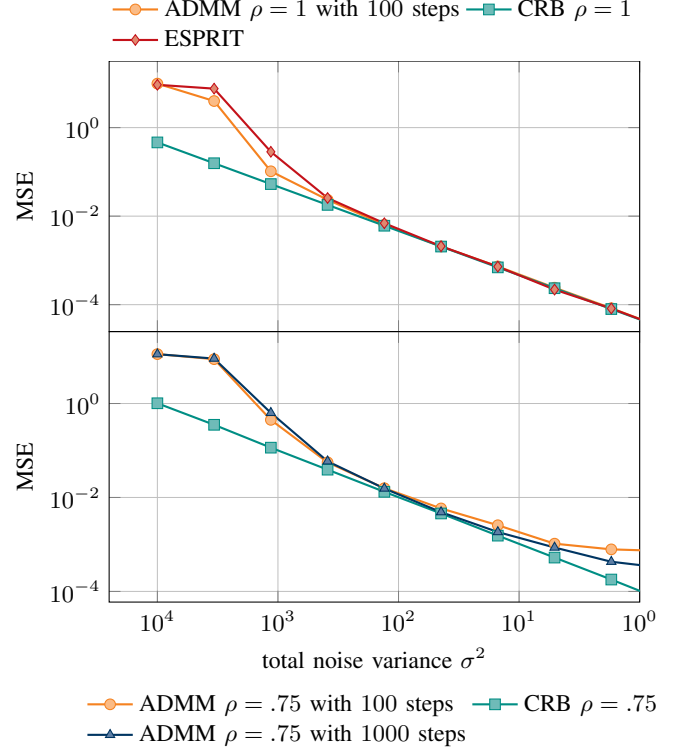

Fig. 1. ADMM reconstruction performance for line spectral estimation of $S = 3$ sources and $d = 3$-dimensional frequencies with $\boldsymbol{k} \in [3, 3, 3]$ in comparison to standard ESPRIT using $K = 100$ snapshots.
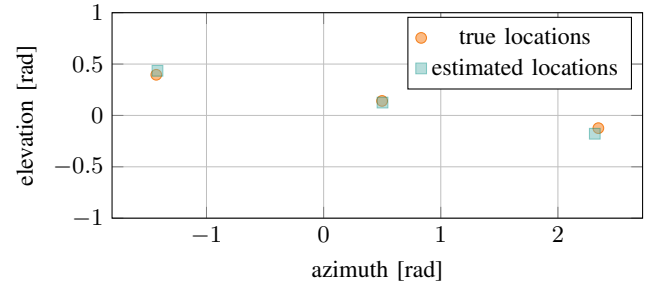

Fig. 2. Simulation results from a 2D DOA estimation scenario using a stacked uniform linear array with 3 stacks and 12 elements each with noise standard deviation of $\sigma = 0.01$.

means that the ADMM approach also achieves the CRB after a suitable amount of iterations.

Moreover, we use the derived algorithm's flexibility and apply it to the 2D DOA estimation problem with a 12×3 stacked circular array, where the stacks are aligned in the x-y-plane with distance $\mathrm{d}z = 0.375\lambda$ and diameter $12/16\lambda = 0.75\lambda$ and λ is the wavelength of the impinging wave. We use the Fourier coefficients of this (synthetic) array to formulate the DOA problem into a line spectral estimation problem as in (5) for $\boldsymbol{\Psi} = \boldsymbol{I}$. A single scenario is depicted in Figure 2 where the noise variance is $\sigma^2 = 0.001$ and we recover these locations in the 2D angular domain from $K = 100$ snapshots. It is worth noting that these results can also be obtained from realistic arrays described by measured data and also if $\boldsymbol{\Psi}$ actually carries out a compression step.

Summarizing, we have demonstrated that the derived ADMM iteration scheme is capable of recovering the unknown multidimensional frequencies from a very general model and how it can be applied to 2D DOA estimation using arbitrary antenna arrays.

V. REFERENCES

[1] H. Karim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, Jul. 1996.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[3] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[4] M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.

[5] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, May 2011.

[6] S. Semper, F. Römer, T. Hotz, and G. Del Galdo, "Grid-Free Direction-of-Arrival estimation with compressed sensing and arbitrary antenna arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2018)*, Calgary, Canada, Apr. 2018.

[7] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *IEEE Transactions on Information Theory*, vol. PP, no. 99, pp. 1–1, 2017.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

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

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

[11] M. Landmann, A. Richter, and R. S. Thomä, "DoA resolution limits in MIMO channel sounding," in *IEEE Antennas and Propagation Society Symposium*, vol. 2, Jun. 2004, pp. 1708–1711.

[12] A. Xenaki and P. Gerstoft, "Grid-free compressive beamforming," *The Journal of the Acoustical Society of America*, vol. 137, no. 4, pp. 1923–1935, 2015. [Online]. Available: https://doi.org/10.1121/1.4916269

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

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

[15] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00041-013-9292-3

[16] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3685–3701, June 2016.

[17] M. Ibrahim, V. Ramireddy, A. Lavrenko, J. König, F. Römer, M. Landmann, M. Grossmann, G. D. Galdo, and R. S. Thomä, "Design and analysis of compressive antenna arrays for direction of arrival estimation," *Elsevier Signal Processing*, vol. 138, pp. 35 – 47, Sep. 2017.

[18] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.