ON THE DESIGN OF THE MEASUREMENT MATRIX FOR COMPRESSED SENSING BASED DOA ESTIMATION

Mohamed Ibrahim, Florian Roemer, Giovanni Del Galdo

Ilmenau University of Technology, Institute for Information Technology
{mgamal.ibrahim,florian.roemer,giovanni.delgaldo}@tu-ilmenau.de

***Abstract* – In this paper we investigate the design of the measurement matrix for applying Compressed Sensing (CS) to the problem of Direction Of Arrival (DOA) estimation with antenna arrays. So far, it has been suggested to choose the coefficients randomly since this choice satisfies the restricted isometry property (RIP) with a high probability. We demonstrate that this choice may be sub-optimal since it can result in an effective array with significant sidelobes and blind spots. The sidelobes are especially problematic when we use correlation-based greedy algorithms for the sparse recovery stage as they can lead to detecting spurious peaks. To address the problem, we introduce a design methodology for constructing a measurement matrix that mitigates these unwanted effects to achieve a better DOA estimation performance. Numerical results demonstrate the usefulness of our design.**

Keywords: *Compressive Sensing, DOA Estimation, Measurement Design*

1. INTRODUCTION

Direction of arrival (DOA) estimation has been an active field of research for many decades [1]. Estimated DOAs are used in various applications like localization of the transmitting sources, beamforming to enhance the desired signal and reduce interference, channel modeling, tracking and surveillance in radar, and many others.

Recently, the application of sparse recovery, which has become popular due to its usefulness in Compressed Sensing (CS) [2, 3, 4], to direction of arrival (DOA) estimation has been considered for applications like localization of the transmitting sources [5], channel modeling [6], tracking and surveillance in radar [7], and many others. It was highlighted in [8] that, if the field is modeled as a superposition of a few planar wave-fronts, the DOA estimation problem can be expressed as a sparse recovery problem. The main focus in [8] is to use the sparse recovery algorithms that have become popular in the CS field for the DOA estimation problem as an alternative for existing parameter estimation algorithms. Many powerful sparsity-based DOA estimation algorithms have been proposed in recent years [9, 10, 11, 12]. It has been claimed that the sparsity-based DOA estimation techniques can provide some advantages over conventional parameter estimation techniques, such as, being insensitive to source correlation, allowing arbitrary array geometries, working with a single snapshot, and providing guarantees for obtaining a correct global solution via convex relaxation [13].

It has recently been proposed to apply CS to the acquisition of the RF signals that are used for DOA estimation [14, 15]. In particular, the CS paradigm can be implemented in the spatial domain by employing M passive antenna elements that are combined into a smaller number of $m < M$ channels using an analog combining network. Since only m channels need to be sampled and digitized, the hardware complexity remains comparably low while allowing us to cover a larger aperture (and thus be more selective) than a traditional, Nyquist ($\lambda/2$) spaced m-channel antenna array. Based on the fact that the underlying signal is sparse in the angular domain, CS theory suggests that it can be recovered from $m < M$ measurements, provided that the measurement kernel is appropriately chosen.

In this paper we investigate the effect of the measurement kernel on the DOA estimation performance. In particular, we show that considering a random measurement kernel may lead to an effective array with unwanted properties such as blind spots or very high side lobes. The latter are particularly problematic for correlation-based sparse recovery algorithms such as Orthogonal Matching Pursuit (OMP) [16] as they may lead to the detection of spurious peaks. We propose a design approach for the measurement kernel that can be optimized numerically and allows closed-form solutions in certain special cases. Based on the optimized measurement matrices, we investigate the achievable DOA estimation accuracy numerically.

2. DATA MODEL

Consider K transmitting narrowband sources in the far-field of an M-element antenna array. The output signal at the M antenna ports can be expressed as

$$\boldsymbol{x}(t) = \sum_{k=1}^{K} \boldsymbol{a}(\theta_k) \cdot s_k(t) + \boldsymbol{w}(t), \tag{1}$$

where $\boldsymbol{a}(\theta) \in \mathbb{C}^{M\times 1}$ represents the array manifold as a function of the azimuth angle, θ_k is the azimuth angle of arrival of the k-th source, $s_k(t)$ denotes the transmit signal of the k-th source, and $\boldsymbol{w}(t)$ represents the additive measurement noise. For simplicity, we consider an M-element half-wavelength spaced uniform linear array (ULA) such that $\boldsymbol{a}(\theta) = [1, \mathrm{e}^{\jmath\mu}, \ldots, \mathrm{e}^{\jmath(M-1)\mu}]^{\mathrm{T}}$ for $\mu = \pi\cdot\cos(\theta)$. In order to be able to resolve targets with closely spaced angles, a large aperture is required. However, since the spatial sampling theorem allows sensors to be spaced no more than half a wavelength apart, this translates into a large number of sensor elements M. Sampling a large number of channels is costly, since it requires many RF chains with costly components such as amplifiers, filters, and A/D converters.

It has therefore been suggested to apply the Compressed Sensing (CS) framework in order to reduce the number of channels that have to be sampled while maintaining a large aperture [15]. To see that the underlying model is sparse, we rewrite (1) into

$$\boldsymbol{x}(t) = \boldsymbol{A} \cdot \boldsymbol{s}(t) + \boldsymbol{w}(t), \tag{2}$$

where $\boldsymbol{A} = \left[\boldsymbol{a}(\theta_1^{(\mathrm{G})}) \quad \dots \quad \boldsymbol{a}(\theta_N^{(\mathrm{G})})\right] \in \mathbb{C}^{M\times N}$ is the array manifold sampled on a predefined N-point sampling grid and $\boldsymbol{s}(t) \in \mathbb{C}^{N\times 1}$ is K-sparse, provided that the true angles of arrival θ_k are on the sampling grid. If no prior knowledge on the true directions of arrival is available, it is advisable to choose a sampling grid that is uniform in the sense that the correlation between adjacent columns is translation-invariant since this minimizes the self-coherence of the dictionary for a given N. For a ULA, this is achieved by sampling the direction cosines uniformly, i.e., $\mu_n^{(\mathrm{G})} = \pi\cos(\theta_n^{(\mathrm{G})}) = (n-1)\cdot\Delta$ for $n = 1, 2, \dots, N$ where $\Delta = 2\pi/N$ (see [8] for details). Note that in the ULA case this implies that $\boldsymbol{A}$ is row-orthogonal, i.e., $\boldsymbol{A}\cdot\boldsymbol{A}^{\mathrm{H}} = N\cdot\boldsymbol{I}_M$.

Based on (2), CS theory states that the signal can be recovered from $m < M$ linear measurements $\boldsymbol{y}(t) = \boldsymbol{\Phi}\cdot\boldsymbol{x}(t) \in \mathbb{C}^{m\times 1}$, provided that the measurement kernel $\boldsymbol{\Phi} \in \mathbb{C}^{m\times M}$ is appropriately chosen. Inserting (1), we have

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}\cdot\boldsymbol{x}(t) = \sum_{k=1}^{K} \tilde{\boldsymbol{a}}(\theta_k)\cdot s_k(t) + \tilde{\boldsymbol{w}}(t), \tag{3}$$

where $\tilde{\boldsymbol{w}}(t) = \boldsymbol{\Phi}\cdot\boldsymbol{w}(t)$ and $\tilde{\boldsymbol{a}}(\theta) = \boldsymbol{\Phi}\cdot\boldsymbol{a}(\theta)$. Equation (3) shows that the application of CS has transformed the M-element array into an m-port array with a beam pattern given by $\tilde{\boldsymbol{a}}(\theta)$.

In [15] and its follow-up papers, it has been suggested to consider measurement kernels $\boldsymbol{\Phi}$ drawn from random distributions such as Gaussian or Bernoulli distributions. Such a choice is popular due to its simplicity and certain mathematical guarantees on the uniform support recovery, i.e., recovering arbitrary subsets of K non-zero entries in $\boldsymbol{s}(t)$. As we show below, this choice may not be suitable for the DOA estimation task since it may result in the effective array having certain blind spots (i.e., angles from which the energy is severely attenuated) or high sidelobes (which could be mistaken for spurious peaks). A design of $\boldsymbol{\Phi}$ that avoids these effects is introduced in the next section.

3. MEASUREMENT DESIGN

In this section we discuss the design of the measurement matrix $\boldsymbol{\Phi}$. Since the CS-array is built in hardware, we aim at a static design that results in an effective array having certain desired properties, e.g., uniform sensitivity and low cross-correlation. We base our design on the effective array manifold $\tilde{\boldsymbol{a}}(\theta)$ that is introduced in Section 2 and directly depends on $\boldsymbol{\Phi}$ via $\tilde{\boldsymbol{a}}(\theta) = \boldsymbol{\Phi}\cdot\boldsymbol{a}(\theta)$. An ideal generic array for direction finding would satisfy the conditions

$$\tilde{\boldsymbol{a}}(\theta_1)^{\mathrm{H}}\cdot\tilde{\boldsymbol{a}}(\theta_2) = \begin{cases} \text{const} & \theta_1 = \theta_2 \\ 0 & \theta_1 \neq \theta_2 \end{cases}, \tag{4}$$

where the first condition guarantees that the array gain is constant for all angles (to make the array uniformly sensitive) and the second condition asks for good cross-correlation properties to tell signals from different directions apart. Note that this is an example for a completely generic direction finder. For particular applications, the target may be a different one, i.e., constraining only a certain sector of angles or allowing certain values for the residual cross-correlation. We represent the target function as $T(\theta_1, \theta_2)$, where $T(\theta_1, \theta_2) = \text{const}\cdot\delta(\theta_1 - \theta_2)$ represents the example (4).

Due to the finite aperture of the M-element array, the target in (4) can only be achieved approximately. This allows us to define a criterion for optimizing $\boldsymbol{\Phi}$ according to the cost function

$$\begin{aligned} e(\boldsymbol{\Phi}, \theta_1, \theta_2) &= \left|\tilde{\boldsymbol{a}}(\theta_1)^{\mathrm{H}}\cdot\tilde{\boldsymbol{a}}(\theta_2) - T(\theta_1, \theta_2)\right| \\ &= \left|\boldsymbol{a}(\theta_1)^{\mathrm{H}}\cdot\boldsymbol{\Phi}^{\mathrm{H}}\cdot\boldsymbol{\Phi}\cdot\boldsymbol{a}(\theta_2) - T(\theta_1, \theta_2)\right|. \end{aligned} \tag{5}$$

We can eliminate the continuous variables θ_1 and θ_2 by considering the N-point sampling grid $\theta_n^{(\mathrm{G})}, n = 1, 2, \dots, N$ used for CS and define the $N\times N$ matrices $\boldsymbol{E}$ and $\boldsymbol{T}$ according to $\boldsymbol{E}_{(i,j)} = e(\boldsymbol{\Phi}, \theta_i^{(\mathrm{G})}, \theta_j^{(\mathrm{G})})$ and $\boldsymbol{T}_{(i,j)} = T(\theta_i^{(\mathrm{G})}, \theta_j^{(\mathrm{G})})$. Inserting (5) we obtain

$$\boldsymbol{E} = \left|\boldsymbol{A}^{\mathrm{H}}\cdot\boldsymbol{\Phi}^{\mathrm{H}}\cdot\boldsymbol{\Phi}\cdot\boldsymbol{A} - \boldsymbol{T}\right|. \tag{6}$$

Based on (6) the quality of $\boldsymbol{\Phi}$ can be assessed based on a suitable norm of $\boldsymbol{E}$. As a first step, let us consider the Frobenius norm, i.e., we optimize $\boldsymbol{\Phi}$ according to

$$\boldsymbol{\Phi}_{\mathrm{opt}} = \arg\min_{\boldsymbol{\Phi}} \|\boldsymbol{E}\|_{\mathrm{F}}^2. \tag{7}$$

The optimization problem in (7) admits a closed-form solution as shown in the following theorem.

Theorem 1. *Let $\boldsymbol{S} = \boldsymbol{A}\cdot\boldsymbol{T}\cdot\boldsymbol{A}^{\mathrm{H}}$ and let $\boldsymbol{S}_m$ be a rank-m-truncated version of $\boldsymbol{S}$ obtained by setting its $N-m$ smallest eigenvalues to zero. Then the set of optimal solutions to* (7) *is given by the set of matrices $\boldsymbol{\Phi}$ that satisfy $\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} = \boldsymbol{S}_m$.*

Proof: cf. Appendix A.

In other words, Theorem 1 states that we can find an optimal $\boldsymbol{\Phi}$ by computing a square-root factor of the best rank-m approximation of $\boldsymbol{S}$. Moreover, the following corollary can be found from Theorem 1:

Corollary 1. *Under the conditions of Theorem 1 any matrix $\boldsymbol{\Phi}$ is optimal in terms of the "ideal" target from* (4) *if and only if the rows of $\boldsymbol{\Phi}$ have equal norm and are mutually orthogonal.*

Proof. The sampled version of (4) is given by a scaled identity matrix, i.e., $\boldsymbol{T} = C\cdot\boldsymbol{I}_N$. Since $\boldsymbol{A}$ is row-orthogonal it follows that $\boldsymbol{S} = \boldsymbol{A}\cdot\boldsymbol{T}\cdot\boldsymbol{A}^{\mathrm{H}} = C\cdot N\cdot\boldsymbol{I}_M$. As all eigenvalues of $\boldsymbol{S}$ are equal to $C\cdot N$, its eigenvalue decomposition can be written as $\boldsymbol{S} = \boldsymbol{U}\cdot(C\cdot N\cdot\boldsymbol{I}_M)\cdot\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{U} \in \mathbb{C}^{M\times M}$ is an arbitrary unitary matrix. Truncating the $M-m$ "smallest" eigenvalues, we obtain $\boldsymbol{S}_m = C\cdot N\cdot\boldsymbol{U}_m\cdot\boldsymbol{U}_m^{\mathrm{H}}$, where $\boldsymbol{U}_m \in \mathbb{C}^{M\times m}$ contains the first m columns of $\boldsymbol{U}$. Invoking Theorem 1, we have $\boldsymbol{\Phi}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{opt}} = C\cdot N\cdot\boldsymbol{U}_m\cdot\boldsymbol{U}_m^{\mathrm{H}}$ and therefore $\boldsymbol{\Phi}_{\mathrm{opt}}$ is a scaled version of $\boldsymbol{U}_m^{\mathrm{H}}$, which proves the claim. □

Corollary 1 agrees with the intuition that the measurements (i.e., the rows of $\boldsymbol{\Phi}$) should be chosen such that they are orthogonal in order to make every observation as informative as possible. In addition, the corollary shows that this choice also minimizes $\left\|\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} - C\cdot N\cdot\boldsymbol{I}_M\right\|_{\mathrm{F}}$ which contains the correlations between all pairs of columns in $\boldsymbol{\Phi}$ as well as the deviation of the columns' norms (therefore, in a sense, this choice minimizes the "average" mutual correlation). On the other hand, this also demonstrates that the optimization in (7) is not sufficiently selective since all row-orthogonal matrices achieve the same minimum of the cost function.

The cost function (7) assigns an equal weight to the error for all pairs of grid points $\theta_1^{(\mathrm{G})}, \theta_2^{(\mathrm{G})}$, i.e., it tries to maintain a constant main lobe with the same weight as it tries to minimize sidelobes everywhere. In practice it is often desirable to have more control

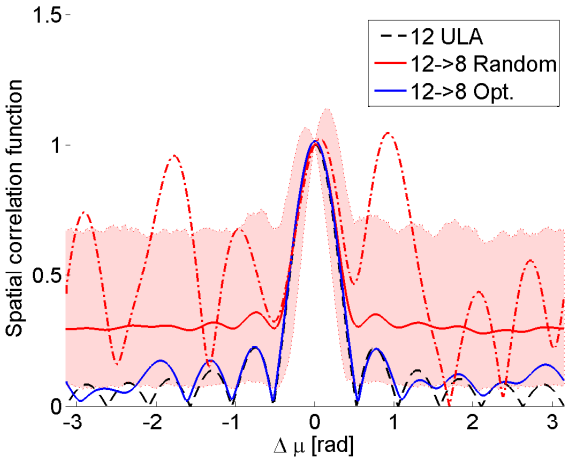

Fig. 1. Spatial correlation function $\varphi(\mu_1, \mu_1 + \Delta\mu)$ for $\mu_1 = \pi$. We compare the $M = 12$ element ULA with an 8×12 CS array. To show the variability of choosing $\mathbf{\Phi}$ randomly, we depict the average correlation (solid line), the 5-th and 95-th percentile (shaded area) and one example realization (dash-dotted line).

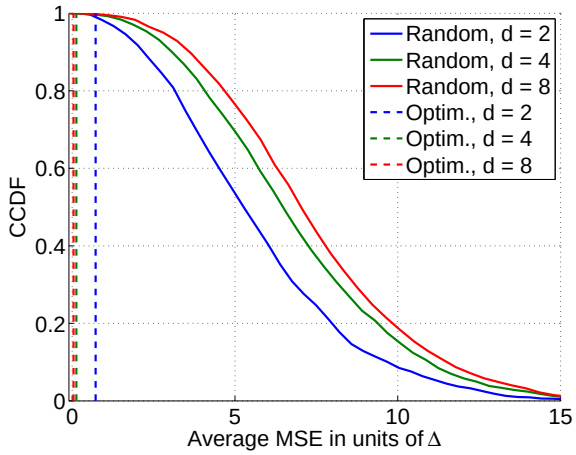

Fig. 2. Mean square estimation error for randomly drawn $\mathbf{\Phi}$, using OMP for the sparse recovery step. We consider two sources positioned at $\mu_{1,2} = (n_0 \pm d/2) \cdot \Delta$. The MSE is averaged over all values $n_0 \in [1, N]$ and a histogram over this average MSE is estimated for three values of the source distance $d = 2, 4, 8$. MSE and source distance are shown in units of the grid spacing Δ to facilitate their comparison. The average MSE of our optimized design is indicated with the dashed lines.

over the shape of the beam pattern, e.g., trading main lobe ripple against sidelobe levels or allowing for a transition region between the main lobes and side lobes that is not constrained. There are many ways such constraints could be incorporated, e.g., maximum constraints on the magnitude of cross-correlation in some region and interval constraints on the autocorrelation inside the main lobe. For numerical tractability, we follow a simpler approach by introducing a weighting matrix $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ into (7). The modified optimization problem is given by

$$\mathbf{\Phi}_{\text{opt}} = \arg\min_{\mathbf{\Phi}} \|\boldsymbol{E} \odot \boldsymbol{W}\|_{\text{F}}^2 , \tag{8}$$

where $\odot$ represents the Schur (elementwise) product. The weighting matrix allows to put more or less weight on the main diagonal (controlling how strictly the constant main lobe power shall be enforced), certain off-diagonal regions (controlling how strongly sidelobes in these regions should be suppressed), or even placing zeros for regions that remain arbitrary (such as transition regions between the main lobe and the side lobes). Thereby, more flexibility is gained and the solution can be tuned to more specific requirements.

The drawback of (8) is that it does not admit a closed-form solution in general. However, it can be solved by numerical optimization routines that are available in modern technical computing languages. Note that it is our goal to derive a generic design of $\mathbf{\Phi}$ which is computed only once (off-line). Therefore, the computational efficiency of finding the solution to (8) is not a major concern.

4. NUMERICAL RESULTS

In this section we present some numerical results to demonstrate the advantage of using a measurement matrix $\mathbf{\Phi}$ that is optimized according to our proposed methodology as compared to choosing $\mathbf{\Phi}$ randomly. To this end, we consider a $M = 12$ element ULA that is reduced to $m = 8$ channels via an 8×12 combining matrix $\mathbf{\Phi}$. This matrix is chosen according to $[\mathbf{\Phi}]_{(m,n)} = \mathrm{e}^{\jmath\varphi_{m,n}}$, where $\varphi_{m,n}$ are the optimization variables in the proposed approach and drawn from a uniform distribution for the random approach.

To find an optimized design $\mathbf{\Phi}_{\text{opt}}$ we solve the weighted optimization problem (8) via MATLAB's numerical optimization features. Since run-time is not a concern for an off-line design, in order to avoid local minima, we run `fmincon` with 100 random initializations and pick the solution with the smallest value of the cost function. As a target we set $\boldsymbol{T} = \boldsymbol{A}^{\text{H}} \cdot \boldsymbol{A}$ which is the correlation function we would achieve with an M-ULA. The weighting matrix is chosen according to $[\boldsymbol{W}]_{(n_1,n_2)} = \rho^{|n_1 - n_2|}$ where $\rho \in (0, 1]$ is a parameter that controls the decay of the weights. Essentially, smaller values of ρ put significantly more weight at the main lobe and its quick decay and less weight on the side-lobes that are far from the main lobe. The limiting value $\rho = 1$ represents the unweighted case.

Figure 1 demonstrates the normalized spatial correlation function (SCF) defined as

$$\varphi(\mu_1, \mu_1 + \Delta\mu) = \frac{\tilde{\boldsymbol{a}}(\mu_1)^{\text{H}} \tilde{\boldsymbol{a}}(\mu_1 + \Delta\mu)}{\tilde{\boldsymbol{a}}(\mu_1)^{\text{H}} \tilde{\boldsymbol{a}}(\mu_1)} \tag{9}$$

for $\mu_1 = \pi$. For reference, the black line indicates the 12-ULA whereas the blue line represents the result of the optimization for $\rho = 0.95$. We observe that the optimized design comes close to the M-ULA except for slightly higher sidelobes. Figure 1 also shows the SCF that are obtained when we draw $\mathbf{\Phi}$ randomly without any optimization. We depict the average SCF (solid line), the 5-th and the 95-th percentile (shaded area) and one example realization (dash-dotted line). As evident from the figure, randomly chosen measurement matrices lead to significantly higher spatial correlations. In particular, every realization shows sidelobes that are sometimes even higher than the main lobe. Note that these lead to erroneous decisions in greedy algorithms such as OMP. What is not shown here is that also $\tilde{\boldsymbol{a}}(\mu_1)^{\text{H}} \tilde{\boldsymbol{a}}(\mu_1)$ varies significantly over μ_1 with some very small values that correspond to regions where the array is significantly less sensitive ("blind spots").

Figure 2 demonstrates the DOA estimation performance if we use the OMP algorithm for the sparse recovery stage. We consider a noise-free scenario with two sources that are located on the $N = 48$

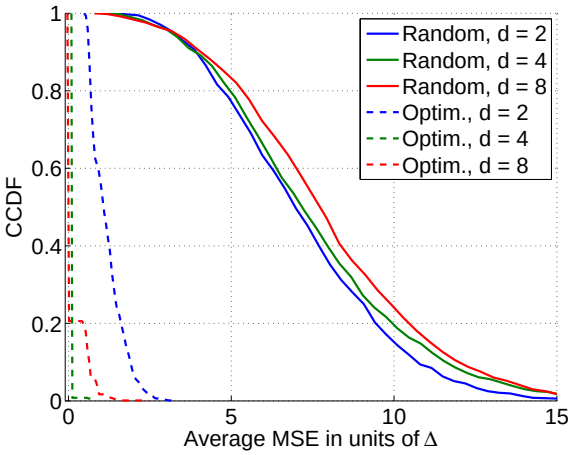

Fig. 3. The same scenario as shown in Figure 2 but this time including noise at an SNR of 10 dB.

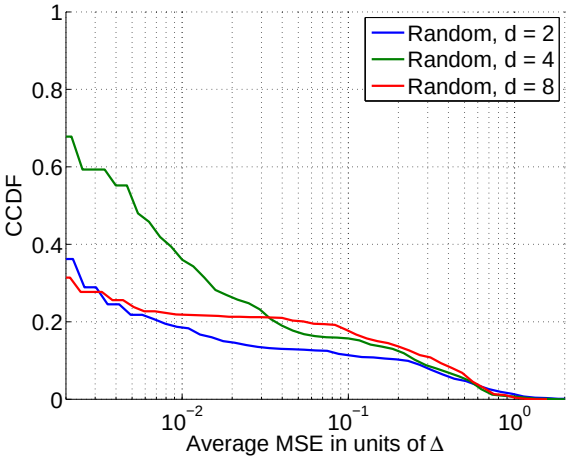

Fig. 4. The same scenario as shown in Figure 2, this time using Basis Pursuit (BP) for the recovery stage. Our proposed design achieves an average MSE of zero for $d = 2$, 4, and 8.

point uniform sampling grid, i.e., $\mu_{1,2} = (n_0 \pm d/2) \cdot \Delta$ where $n_0 \in [1, N]$ and d is the inter-source spacing in grid points. For each value of d, the mean square error $\text{MSE} = \frac{1}{2}\sum_{k=1}^{2}(\mu_k - \hat{\mu}_k)^2$ is averaged over all values of n_0. An estimate of the Complementary Cumulative Distribution Function (CCDF) of this average MSE obtained from 1000 realizations of $\boldsymbol{\Phi}$ is shown. As evident from the figure, the average MSE exceeds the source spacing d with a probability of 95 % for $d = 2$, 80 % for $d = 4$, and 30 % for $d = 8$. For the same scenario, our optimized design (choosing $w = 0.7$) achieves an MSE very close to zero (shown with the dashed lines), i.e., 0.72Δ for $d = 2$, 0.13Δ for $d = 4$, and 0.03Δ for $d = 8$.

To investigate the effect of additive noise, we have repeated the experiment from Figure 2 with additive noise. In particular, we have drawn the noise vector $\boldsymbol{w}$ (cf. (2)) from a zero mean circularly symmetric complex Gaussian distribution with a variance of 0.1, which corresponds to an SNR of 10 dB. The result is shown in Figure 3. Once more, the optimized design achieves a significantly lower average MSE compared to the randomly chosen measurement matrices.

Though the main concern for optimizing $\boldsymbol{\Phi}$ has been put on correlation-based recovery algorithms such as the OMP, we have tested it on the Basis Pursuit (BP) algorithm [17] as well and found that it also offers some advantages there. To this end, we have repeated the previous simulation with the same parameters, using BP instead of OMP for the recovery stage. The result is shown in Figure 4. The figure shows the CCDF of the average MSE over 1000 random realizations of $\boldsymbol{\Phi}$. Our proposed design is not shown since it achieves an exact reconstruction result (MSE=0) for all values of d. As expected, BP is more reliable and less sensitive to sidelobes compared to OMP. In fact, for each value of d there is a non-zero chance to draw a matrix $\boldsymbol{\Phi}$ randomly that achieves an exact reconstruction as well. However, these "lucky" choices are not very stable in the sense that changing the grid size N or the source spacing d results in estimation errors to occur.

5. CONCLUSIONS

In this paper, we have discussed the design of the measurement matrix for applying Compressive Sensing (CS) to the Direction Of Arrival (DOA) estimation problem. We have demonstrated that choosing it randomly may lead to undesirable effects in the effective CS-array such as very high sidelobes. These are particularly problematic for correlation-based sparse recovery algorithms such as the Orthogonal Matching Pursuit since they may lead to the detection of spurious peaks. We have introduced a design methodology for the measurement matrix that can avoid these issues by optimizing the resulting effective beam pattern. Numerical results demonstrate that the optimized design leads to more favorable spatial correlation functions and a significantly improved DOA estimation performance.

A. PROOF OF THEOREM 1

To prove the theorem we use the fact that for a unitary matrix $\boldsymbol{U}$ and an arbitrary square matrix $\boldsymbol{X}$ we have $\|\boldsymbol{X}\cdot\boldsymbol{U}\|_{\text{F}} = \|\boldsymbol{U}\cdot\boldsymbol{X}\|_{\text{F}} = \|\boldsymbol{X}\|_{\text{F}}$. Since $\boldsymbol{A}$ satisfies $\boldsymbol{A}\cdot\boldsymbol{A}^{\text{H}} = N\cdot\boldsymbol{I}_M$ we can find a matrix $\bar{\boldsymbol{A}} \in \mathbb{C}^{(M-N)\times N}$ such that $\boldsymbol{V} \doteq 1/\sqrt{N}\cdot[\boldsymbol{A}^{\text{T}}, \bar{\boldsymbol{A}}^{\text{T}}]^{\text{T}} \in \mathbb{C}^{N\times N}$ is a unitary matrix. Therefore, we have $\boldsymbol{V}\cdot\boldsymbol{A}^{\text{H}} = \left[\sqrt{N}\cdot\boldsymbol{I}_M, \boldsymbol{0}_{M\times N-M}\right]^{\text{T}}$. The cost function (7) can then be rewritten as

$$
\begin{aligned}
\|\boldsymbol{E}\|_{\text{F}}^2 &= \left\|\boldsymbol{V}\cdot\boldsymbol{E}\cdot\boldsymbol{V}^{\text{H}}\right\|_{\text{F}}^2 \\
&= \left\|\begin{bmatrix}\sqrt{N}\cdot\boldsymbol{I}_M \\ \boldsymbol{0}_{N-M\times M}\end{bmatrix}\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi}\left[\sqrt{N}\cdot\boldsymbol{I}_M, \boldsymbol{0}_{M\times N-M}\right] - \boldsymbol{V}\cdot\boldsymbol{T}\cdot\boldsymbol{V}^{\text{H}}\right\|_{\text{F}}^2 \\
&= \left\|\begin{bmatrix}N\cdot\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi} & \boldsymbol{0}_{M\times N-M} \\ \boldsymbol{0}_{N-M\times M} & \boldsymbol{0}_{N-M\times N-M}\end{bmatrix} - N\cdot\begin{bmatrix}\boldsymbol{A} \\ \bar{\boldsymbol{A}}\end{bmatrix}\boldsymbol{T}\left[\boldsymbol{A}^{\text{H}}, \bar{\boldsymbol{A}}^{\text{H}}\right]\right\|_{\text{F}}^2 \\
&= \left\|N\cdot\begin{bmatrix}\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi} - \boldsymbol{A}\cdot\boldsymbol{T}\cdot\boldsymbol{A}^{\text{H}} & -\boldsymbol{A}\cdot\boldsymbol{T}\cdot\bar{\boldsymbol{A}}^{\text{H}} \\ -\bar{\boldsymbol{A}}\cdot\boldsymbol{T}\cdot\boldsymbol{A}^{\text{H}} & -\bar{\boldsymbol{A}}\cdot\boldsymbol{T}\cdot\bar{\boldsymbol{A}}^{\text{H}}\end{bmatrix}\right\|_{\text{F}}^2 \\
&= N^2\cdot\left\|\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi} - \boldsymbol{S}\right\|_{\text{F}}^2 + \text{const},
\end{aligned}
\tag{10}
$$

using the short-hand notation $\boldsymbol{S} = \boldsymbol{A}\cdot\boldsymbol{T}\cdot\boldsymbol{A}^{\text{H}}$. Equation (10) demonstrates that the optimization problem is equivalent to finding the best approximation of the matrix $\boldsymbol{S}$ by the matrix $\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi}$. Since $\boldsymbol{\Phi}$ is an $m\times M$ matrix, the rank of the $M\times M$ matrix $\boldsymbol{\Phi}^{\text{H}}\cdot\boldsymbol{\Phi}$ is less than or equal to $m < M$. Therefore, (10) represents a low-rank approximation problem. According to the Eckart-Young theorem, its optimal solution is given by truncating the $M - m$ smallest eigenvalues of $\boldsymbol{S}$.

B. REFERENCES

[1] 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, Jul 1996.

[2] E. J. Candés, "Compressive sampling," International Congress of Mathematicians, Madrid, Spain, European Mathematical Society, Tech. Rep., 2006.

[3] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] V. Cevher, A. C. Gurbuz, J. H. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, Mar 2008, pp. 2497–2500.

[6] C. Feng, S. Valaee, and Z. Tan, "Multiple target localization using compressive sensing," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov 2009, pp. 1–6.

[7] J. H. Ender, "On compressive sensing applied to radar," *ELSEVIER Signal Processing Magazine*, vol. 90, no. 5, pp. 1402 – 1414, 2010.

[8] 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.

[9] A. C. Gurbuz, V. Cevher, and J. H. McClellan, "Bearing estimation via spatial sparsity using compressive sensing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1358–1369, Apr 2012.

[10] A. Gretsistas and M. D. Plumbley, "A multichannel spatial compressed sensing approach for direction of arrival estimation," in *International Conference on Latent Variable Analysis and Signal Separation*, ser. LVA/ICA'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 458–465.

[11] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, Feb 2011.

[12] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Signal Process.*, vol. 86, no. 3, pp. 624–638, Mar. 2006.

[13] M. Ibrahim, F. Römer, R. Alieiev, G. D. Galdo, and R. S. Thomä, "On the estimation of grid offsets in CS-based direction of arrival estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014.

[14] J. Gu, W. Zhu, and M. N. S. Swamy, "Compressed sensing for DOA estimation with fewer receivers than sensors," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2011, pp. 1752–1755.

[15] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *IEEE/SP Workshop on Statistical Signal Processing*, Aug 2009, pp. 626–629.

[16] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1993.

[17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, Jan. 1998.