Combining Matrix Design for 2D DoA Estimation with Compressive Antenna Arrays using Stochastic Gradient Descent

S. Pawar*, S. Semper*, F. Römer*†
*Institute for Information Technology, Technische Universität Ilmenau, Ilmenau, Germany
†Fraunhofer Institute for Nondestructive Testing (IZFP), Ilmenau, Germany
Email: sankalp-prakash.pawar@tu-ilmenau.de

***Abstract*—Recently, compressive antenna arrays have been considered for direction of arrival (DoA) estimation with reduced hardware complexity. By utilizing compressive sensing, such arrays employ a linear combining network to combine signals from a larger set of antenna elements in the analog RF domain. In this paper, we develop a design approach based on the minimization of error between spatial correlation function (SCF) of the compressive and the uncompressed array resulting in the estimation performance of the two arrays to be as close as possible. The proposed design is based on grid-free stochastic gradient descent (SGD) optimization. In addition to a low computational cost for the proposed method, we show numerically that the resulting combining matrices perform better than the ones generated by a previous approach and combining matrices generated from a Gaussian ensemble.**

I. Introduction

Determination of the direction of impinging waves using an antenna array is formulated as the DoA estimation problem [1], [2]. It is well established that a comparatively large number of elements in the receiving array is required to achieve high DoA estimation accuracy [2]. Implementation of antenna arrays for DoA estimation usually requires a radio frequency (RF) chain to process each antenna output. Such a RF chain could include components such as a low-noise amplifier (LNA), filters, down-converter, and analog-to-digital (ADC) converter. Thus, realisation of a relatively large array brings along increased hardware costs. In recent years, major focus of research has been dedicated to develop techniques that provide desirable DoA estimation performance while enabling reduction in hardware complexity. Ideas from the compressed sensing (CS) domain have been considered to reduce the complexity of hardware implementation while maintaining DoA estimation accuracy [3]–[6].

One such approach involves the application of CS paradigm in the spatial domain by employing an analog combining network to linearly combine N antenna outputs to a smaller number of $M < N$ channels [7]–[9]. Using such a CS-based combining network, only M channels need to be processed through their respective RF chains. The hardware complexity of the *compressive* array formed by the output channels of the combining network is lower compared to its uncompressed counterpart without such a network. Besides, the compressive array provides a better estimation performance owing to its larger aperture in comparison to an equivalent uncompressed array of size M.

In [10], a low-complexity design approach based on the spatial correlation function (SCF) for 1D DoA estimation is proposed while its extension for 2D DoA estimation is investigated in [11]. In [11], instead of taking the complete 4D-SCF to define the cost function for optimization, only 2D subsets are used. A method to choose these 2D subsets, and an evaluation of the DoA performance while achieving considerable reduction in computational requirements compared to a direct extension of the approach in [10] is described in [11].

Despite the effectiveness of the approach in [11] its computational requirements increase substantially with increasing size of the antenna array. In this contribution, we propose a Stochastic Gradient Descent (SGD) based approach with momentum [12] for obtaining the kernels of the combining matrix. The cost function is defined as the average difference between the compressed and the uncompressed SCF evaluated at a given set of 2D angles. The gradient of the cost function with respect to the combining matrix is analytically derived before applying the descent algorithm. We evaluate the performance of the proposed design approach for compression of a synthetic stacked uniform circular array (SUCA) in terms of the difference in the resulting spatial correlation functions as well as the Cramér-Rao Lower Bound (CRLB).

II. System model

A. Input Signal

For an N-element antenna array, the complex baseband signal received by it is given by $\boldsymbol{y}(t) \in \mathbb{C}^N$ such that,

$$\boldsymbol{y}(t) = \sum_{s=1}^{S} \boldsymbol{a}(\theta_s, \vartheta_s) x_s(t) + \boldsymbol{v}(t), \tag{1}$$

with the array receiving S far-field narrowband plane waves impinging from DoAs described by azimuth (θ) and elevation (ϑ) pair (θ_s, ϑ_s). The array steering vector is $\boldsymbol{a}(\theta_s, \vartheta_s) = [a_1(\theta_s, \vartheta_s), \ldots, a_N(\theta_s, \vartheta_s)]^{\mathrm{H}} \in \mathbb{C}^N$ for $s = 1, \ldots, S$, $x_s(t)$ is the complex transmitted signal, while $\boldsymbol{v}(t) \in \mathbb{C}^N$ is the additive noise. Determination of angles pairs (θ_s, ϑ_s) from $\boldsymbol{y}(t)$ is the goal of DoA estimation.

B. Compressive Arrays

A dedicated RF receiver chain is required for each antenna output of the array to comply with the model in (1). Hardware implementation of such an array is prohibitive with considerations of complexity, cost, and power consumption. Reduction in the number of RF channels without the compromising DoA estimation performance can be achieved by applying the *compressive* approach. A compressive array is obtained by utilizing an analog combining network to linearly combine the outputs of an array to a lower number of channels [10]. The combining network is described by the matrix $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$, such that in baseband, $[\mathbf{\Phi}]_{m,n} = \alpha_{m,n} e^{\jmath \varphi_{m,n}}$ are the complex weights that are applied to the antenna outputs, with $m = 1, \ldots M, n = 1, \ldots, N$ and $M < N$. Thus, the array output of the compressive array is given by

$$\tilde{\boldsymbol{y}}(t) = \mathbf{\Phi}\boldsymbol{y}(t) = \sum_{s=1}^{S} \tilde{\boldsymbol{a}}(\theta_s, \vartheta_s) x_s(t) + \boldsymbol{w}(t), \quad (2)$$

where $\tilde{\boldsymbol{a}}(\theta, \vartheta) = \mathbf{\Phi}\boldsymbol{a}(\theta, \vartheta) \in \mathbb{C}^{M \times 1}$ represents the effective (compressed) array steering vector after the combining, and $\boldsymbol{w}(t) \in \mathbb{C}^{M \times 1}$ is the noise vector representing additive noise sources of the system [10]. Now, the goal is to design $\mathbf{\Phi}$ in order to allow accurate estimation of (θ_s, ϑ_s) from the above compressive measurement $\tilde{\boldsymbol{y}}(t)$.

III. Combining Matrix Design based on the Spatial Correlation Function

For an array with a manifold $\boldsymbol{a}(\theta, \vartheta)$ the spatial correlation function (SCF) $\rho : \mathbb{R}^4 \to \mathbb{C}$ is given by

$$\rho(\theta_1, \theta_2, \vartheta_1, \vartheta_2) \triangleq \boldsymbol{a}^{\mathrm{H}}(\theta_1, \vartheta_1)\boldsymbol{a}(\theta_2, \vartheta_2). \quad (3)$$

Inserting $\tilde{\boldsymbol{a}}(\theta, \vartheta)$ into (3), we obtain the *effective* SCF of the compressive array as

$$\begin{aligned}\tilde{\rho}(\theta_1, \theta_2, \vartheta_1, \vartheta_2) &= \tilde{\boldsymbol{a}}^{\mathrm{H}}(\theta_1, \vartheta_1)\tilde{\boldsymbol{a}}(\theta_2, \vartheta_2) \\ &= \boldsymbol{a}^{\mathrm{H}}(\theta_1, \vartheta_1)\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\boldsymbol{a}(\theta_2, \vartheta_2) \\ &= \boldsymbol{a}^{\mathrm{H}}(\theta_1, \vartheta_1)\boldsymbol{G}_{\Phi}\boldsymbol{a}(\theta_2, \vartheta_2), \quad (4)\end{aligned}$$

where for any arbitrary matrix $\boldsymbol{B}$ we denote by $\boldsymbol{G}_{\mathrm{B}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ its Gramian.

Depending on the application, one might have different requirements to the effective SCF $\tilde{\rho}(\theta_1, \theta_2, \vartheta_1, \vartheta_2)$, such that it ensures uniform sensitivity and/or good cross-correlation properties for instance. Our particular design goal in this work is to reduce the number of receiver channels M while not compromising the DoA estimation performance compared to the original array without the combining. To evaluate the performance of a given $\mathbf{\Phi}$, we define $\delta : \mathbb{C}^{M \times N} \to \mathbb{R}$ as

$$\delta(\mathbf{\Phi}) = \iint\limits_{\theta} \iint\limits_{\vartheta} |e(\mathbf{\Phi}, \theta_1, \theta_2, \vartheta_1, \vartheta_2)|^2 \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\vartheta_1 \mathrm{d}\vartheta_2, \quad (5)$$

where $e : \mathbb{C}^{M \times N} \times \mathbb{R}^4$ is defined as

$$\begin{aligned}e(\mathbf{\Phi}, \theta_1, \theta_2, \vartheta_1, \vartheta_2) &\triangleq \tilde{\rho}(\theta_1, \theta_2, \vartheta_1, \vartheta_2) - \rho(\theta_1, \theta_2, \vartheta_1, \vartheta_2) \\ &= \boldsymbol{a}^{\mathrm{H}}(\theta_1, \vartheta_1)\boldsymbol{G}_{\Phi}\boldsymbol{a}(\theta_2, \vartheta_2) - \boldsymbol{a}^{\mathrm{H}}(\theta_1, \vartheta_1)\boldsymbol{a}(\theta_2, \vartheta_2). \quad (6)\end{aligned}$$

Note that a small value of $\delta(\mathbf{\Phi})$ implies that we can expect a DoA estimation performance close to that of the original array before the combining, as discussed in [10] for the case of 1D DoA estimation. So to this end δ serves as a suitable proxy to estimate the performance of a given $\mathbf{\Phi}$.

Consequently, in order to find a good combining matrix $\mathbf{\Phi}$ one has to minimize (5). This however implies solving an optimization problem, where the evaluation of the objective function itself is very time consuming since it has to be approximated with some numerical integration scheme. To circumvent this problem, consider an i.i.d. sequence of random vectors $(\mathbf{\Theta}^k)_{k \in \mathbb{N}} \subset ([0, 2\pi) \times (0, \pi))^L$ over some appropriate probability measure space $(\Omega, \mathcal{A}, P)$ and $L \in \mathbb{N}$, which represent L points on the unit sphere, where the first component is the angle in azimuth and the second elevation. Further, let us define

$$D(\mathbf{\Phi}, \mathbf{\Theta}) = \frac{1}{KL^2} \sum_{k=1}^{K} \sum_{\ell_1}^{L} \sum_{\ell_2}^{L} |e(\mathbf{\Phi}, \Theta^k_{\ell_1,1}, \Theta^k_{\ell_2,1}, \Theta^k_{\ell_1,2}, \Theta^k_{\ell_2,2})|^2.$$

For the following, we set

$$\boldsymbol{A}_k = [\boldsymbol{a}(\Theta^k_{1,1}, \Theta^k_{1,2}), \ldots, \boldsymbol{a}(\Theta^k_{L,1}, \Theta^k_{L,2})] \in \mathbb{C}^{M \times L}$$

and

$$\boldsymbol{E}_k(\mathbf{\Phi}) = \boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{G}_{\mathbf{\Phi}}\boldsymbol{A}_k - \boldsymbol{A}_k^{\mathrm{H}}\boldsymbol{A}_k \in \mathbb{C}^{L \times L},$$

which allows us to write

$$D(\mathbf{\Phi}, \mathbf{\Theta}) = \frac{1}{K} \sum_{k=1}^{K} \|\boldsymbol{E}_k(\mathbf{\Phi})\|_F^2$$

and

$$\nabla_{\mathbf{\Phi}} D(\mathbf{\Phi}, \mathbf{\Theta}) = \frac{1}{K} \sum_{k=1}^{K} \nabla_{\mathbf{\Phi}} \|\boldsymbol{E}_k(\mathbf{\Phi})\|_F^2.$$

The key idea now is to minimize $\delta(\mathbf{\Phi})$ by minimizing $D(\mathbf{\Phi}, \mathbf{\Theta}(\omega))$ for some fixed realization $\omega \in \Omega$. In other words we carry out the minimization of $D(\mathbf{\Phi}, \mathbf{\Theta})$ for a specific realization of $\mathbf{\Theta}(\omega)$, which is a suitable approach, if we choose K large enough and the distributions of the $\mathbf{\Theta}^k$ such that this random process explores the angular domain well enough. This stochastic approach is outlined in the following section.

IV. Stochastic Gradient Descent-based Design

Gradient Descent with momentum [13] is a popular and simple first order technique to find local minima of smooth functions. In its most simple form it minimizes a smooth function $f : \mathbb{C}^n \to \mathbb{C}$ by iterating

$$\boldsymbol{v}_{i+1} = \eta \boldsymbol{v}_i - \alpha \nabla_{\boldsymbol{x}} f(\boldsymbol{x}_i) \quad (7)$$

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \boldsymbol{v}_i \quad (8)$$

for initial velocity and state variables $\boldsymbol{v}_0, \boldsymbol{x}_0 \in \mathbb{C}^n$, a drag parameter $\eta \in [0, 1)$ and a step size $\alpha > 0$. Now, if f is of the form $f = (1/K)\sum_k^K g_k$ for smooth functions $g_k : \mathbb{C}^n \to \mathbb{C}$ above velocity can be rewritten as

$$\boldsymbol{v}_{i+1} = \eta \boldsymbol{v}_i - \alpha \frac{1}{K} \sum_{k}^{K} \nabla_{\boldsymbol{x}} g_k(\boldsymbol{x}_i).$$

Now, suppose K is prohibitively large such that the evaluation of f and its gradient becomes computationally intractable. Then a well known approach that has recently gained a lot of traction because of its use in the machine learning community [13] is stochastic gradient descent [12], which exploits the fact that

$$\mathbb{E}\frac{1}{\kappa} \sum_{k \in \mathcal{K}} \nabla_{\boldsymbol{x}} g_k(\boldsymbol{x}_i) = \frac{1}{K} \sum_{k}^{K} \nabla_{\boldsymbol{x}} g_k(\boldsymbol{x}_i)$$

for $\mathcal{K}$ selected uniformly at random from the set of all subsets of $\{1, \ldots, K\}$ with cardinality κ. This makes the above random sum depending on the random $\mathcal{K}$ a good approximation of $\nabla_{\boldsymbol{x}} f$ due to the weak law of large numbers, which states that with high probability above partial sum is close to its expectation for a single realization $\mathcal{K}(\omega)$. Then the velocity update reads as

$$\boldsymbol{v}_{i+1} = \eta \boldsymbol{v}_i - \alpha \frac{1}{\kappa} \sum_{k \in \mathcal{K}} \nabla_{\boldsymbol{x}} g_k(\boldsymbol{x}_i), \tag{9}$$

while the state variable update stays the same as in (8). So the proposed stochastic gradient descent consists of iteratively carrying out (9) and (8). Now, we outline how to apply the iterative approach explained above to the problem at hand.

A. SGD for Combining Matrix Design

For the specific problem of designing a suitable combining matrix $\boldsymbol{\Phi}$ we apply a slight modification of (9) by reformulating it as an online minimization of $D(\boldsymbol{\Phi}, \boldsymbol{\Theta})$ via

$$\boldsymbol{v}_{i+1} = \eta \boldsymbol{v}_i - \alpha \nabla_{\boldsymbol{\Phi}} \|\boldsymbol{E}_i(\boldsymbol{\Phi}_i)\|_F^2, \tag{10}$$

which allows to generate Θ^i and thus $\boldsymbol{E}_i(\boldsymbol{\Phi})$ during the iteration without the need to store or precompute them during or before the execution. This means instead of randomly subselection points from a prespecified set of points on the sphere, we keep drawing a new set of points on the fly in each iteration step.

Clearly, we are still in need of an analytical expression of the gradient of $\|\boldsymbol{E}_i(\boldsymbol{\Phi})\|_F^2$, which is provided in the following result.

Lemma 1. *For given* $\boldsymbol{\Theta}^k$ *and* $\boldsymbol{\Phi} \in \mathbb{C}^{M \times N}$ *it holds that*

$$\nabla_{\boldsymbol{\Phi}} \|\boldsymbol{E}_i(\boldsymbol{\Phi}_i)\|_F^2 = 4\boldsymbol{\Phi}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{G}_{\boldsymbol{\Phi}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}} - 4\boldsymbol{\Phi}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}.$$

Proof. First, we use the facts $\|\boldsymbol{M}\|_F^2 = \operatorname{tr}(\boldsymbol{M}^{\mathrm{H}}\boldsymbol{M})$ and $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$ to get

$$\begin{aligned}
\|\boldsymbol{E}_i(\boldsymbol{\Phi})\|_F^2 &= \operatorname{tr}\left(\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{A}_i\right) \\
&\quad - 2\operatorname{tr}\left(\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{A}_i\right) + \operatorname{tr}\left(\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}\right) \\
&= \operatorname{tr}\left(\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{A}_i\right) \\
&\quad - 2\operatorname{tr}\left(\boldsymbol{\Phi}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\right) + \operatorname{tr}\left(\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{H}}\boldsymbol{A}\right).
\end{aligned}$$

Now with two well known results from matrix calculus

$$\nabla_{\boldsymbol{X}} \operatorname{tr}\left(\boldsymbol{M}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{M}\right) = 4\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{M}\boldsymbol{M}^{\mathrm{H}}$$

and $\nabla_{\boldsymbol{X}} \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^{\mathrm{H}}\right) = 2\boldsymbol{X}\boldsymbol{M}$, we conclude the statement. □

Now we have all ingredients at hand to implement the proposed method with an analytically derived gradient calculation. The following section evaluates the performance of this approach.

V. Numerical Evaluation

In this section, we assess the performance of the compressive array designed using our proposed SGD-based approach in comparison with that of the uncompressed array, the compressive array obtained using the 2D SCF-based approach from [11] and the compressive array derived from a randomly drawn combining matrix.

In any case, we let SGD run for $K = 5000$ steps, with $L = 250$ angles per step, step size $\alpha = 10^{-2}$ and drag parameter $\eta = 0.1$, where the distribution of the $\boldsymbol{\Theta}_k$ is the uniform distribution on $(0, 2\pi] \times [\pi/4, 3\pi/4]$. Moreover we always use normalized sensing matrices, which means that the columns of any $\boldsymbol{\Phi}$ considered are normalized to unit length. As an antenna we consider a SUCA of $(\Sigma \times N_S)$ isotropic elements, so it has $\Sigma = 3$ stacks of $N_S = 11$ elements each with the total number of elements denoted by $N = \Sigma \times N_S$. The array response of the SUCA is given by $\boldsymbol{a}_{\mathrm{SUCA}}(\theta, \vartheta)$. We choose $d = 0.5\lambda$ as the distance between two consecutive stacks, $R = 0.68\lambda$ as the radius of the stacks, where $\lambda = c/f$ is the wave-length at frequency f with $c = 3 \cdot 10^8$ m/sec. For the SCF-based approach from [11], the number of grid points in azimuth and elevation used for calculation of SCF is $N_\theta = 121$ and $N_\vartheta = 61$, respectively over $\theta \in [-\pi, \pi]$ and $\vartheta \in [-\pi/2, \pi/2]$. Also, the number of reference points in elevation considered for optimization using this approach is $|\mathcal{N}| = 3$.

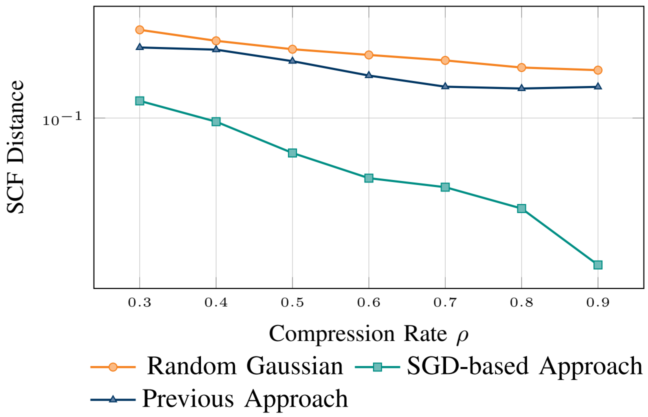

Fig. 1: Distance of the SCFs generated by the different approaches for varying levels of compression.

In Figure 1 we evaluate how the different approaches perform in terms of the overall SCF-error. To approximate the quantity in (5) we evaluate (6) on a regular 2D grid in azimuth and elevation and then sum the squared absolute values of these samples. This is done for different levels of compression rate ρ. Clearly, the proposed SGD-based method is capable of approximating the original antenna response more closely than the previous approach and the conventional combining matrices resulting from a zero-mean Gaussian ensemble.

To quantify the performance of the proposed design apporach for the case of 2D DoA estimation, we also evaluate the deterministic CRLB in Figure 2, since it serves as a proxy to assess the possible performance of any unbiased estimator. For instance, the maximum likelihood estimator always reaches this lower bound asymptotically in the effective SNR, so one can expect that any well designed estimation procedure behaves similarly in the asymptotic regime. With spatial compression, the deterministic CRLB for the 2-dimensional case with S sources and 1 snapshot can be computed via [14]

$$C(\boldsymbol{\theta}) = \frac{\sigma^2}{2} \operatorname{tr}\left(\left[\Re(\boldsymbol{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{G}}^{\perp}\boldsymbol{D} \odot (\boldsymbol{1}_{2\times 2} \otimes \hat{\boldsymbol{R}})^{\mathrm{T}})\right]^{-1}\right), \tag{11}$$

with $\boldsymbol{\Pi}_{\boldsymbol{G}}^{\perp} = \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^{\mathrm{H}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{H}}$, $\otimes$ denoting the Kronecker product, $\Re$ the real part of a complex number and $\hat{\boldsymbol{R}} = \boldsymbol{x} \cdot \boldsymbol{x}^{\mathrm{H}}$

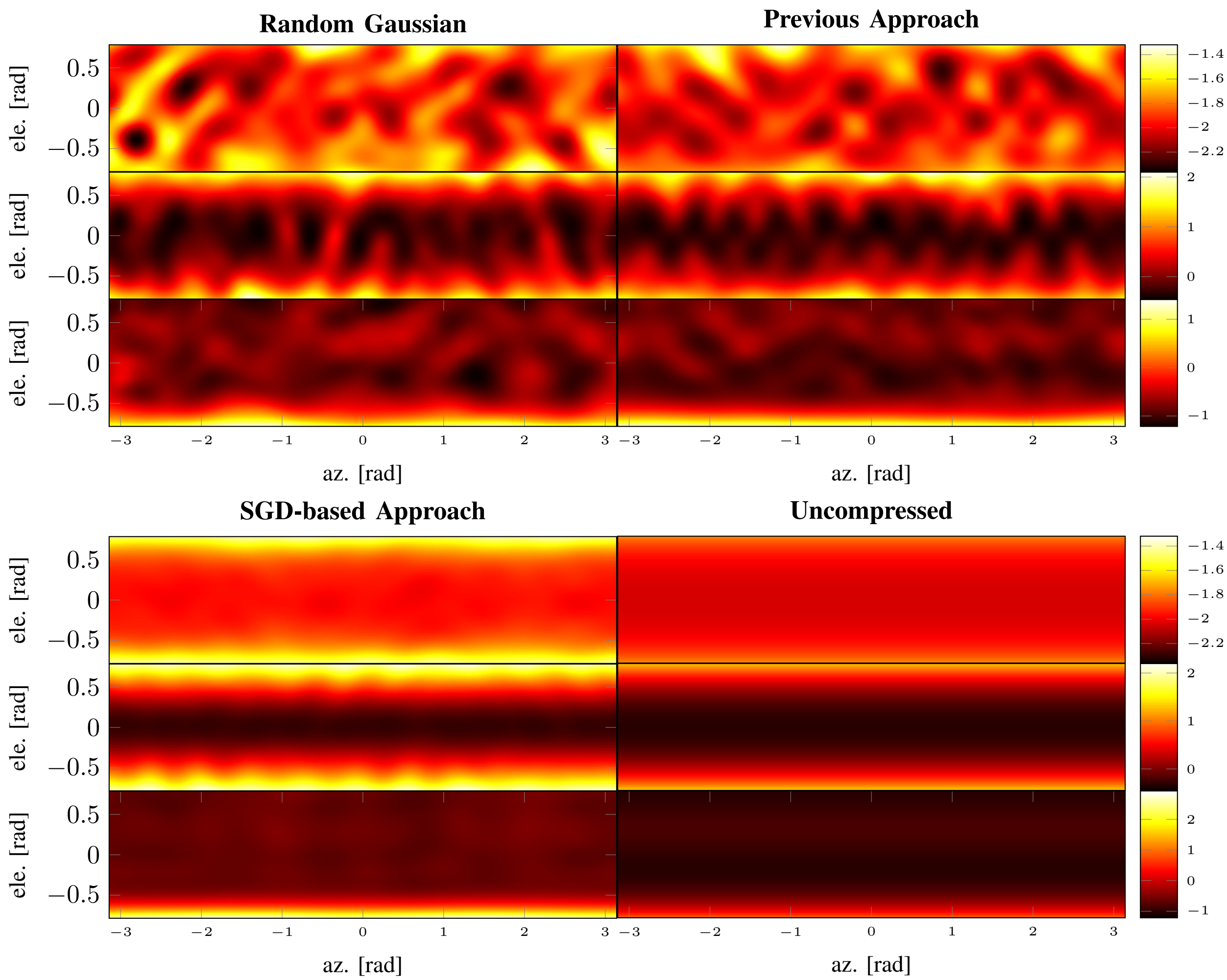

Fig. 2: Magnitude of the deterministic CRLB on a logarithmic scale for fixed noise level $\sigma^2 = 1$ dependent on azimuth and elevation. *Top*: single source, *Middle*: two sources separated in azimuth by $2\pi/10$, *Bottom*: two sources separated in elevation by $2\pi/10$. Notice the differing color bars for top, middle and bottom.

being the sample covariance and we have set

$$\boldsymbol{G} = \boldsymbol{\Phi}[\boldsymbol{a}_{\text{SUCA}}(\theta_1, \vartheta_1), \ldots, \boldsymbol{a}_{\text{SUCA}}(\theta_S, \vartheta_S)],$$
$$\boldsymbol{D}_i = \frac{\partial}{\partial \boldsymbol{\theta}_i}\boldsymbol{G},\ \boldsymbol{D} = [\boldsymbol{D}_1, \ldots, \boldsymbol{D}_d]\,.$$

The results in Figure 2 show for a fixed noise level $\sigma^2 = 1$ how the CRLB changes for the four different sensing matrix designs depending on the position of a single source (top), two sources separated in azimuth (middle) and elevation (bottom). In the two sources case, the first source is located at the position denoted in the plot and the second with angular distance $2\pi/10$ in azimuth or elevation. As one can see the random combining matrix and the previous approach introduce a highly varying sensitivity of the CRLB with respect to azimuth and elevation, rendering the resulting combining matrices hard to apply for DoA estimation because of this non-uniformity in the angular domain. The SGD-based approach results in a significantly smoother behavior of the CRLB with a uniform increase across the whole azimuth and elevation region. Thus, the compressed array resulting from SGD mimics the behavior of the uncompressed array more closely in terms of the CRLB, which ultimately was the goal of the proposed design process.

VI. Conclusion and Outlook

We have presented a flexible and computationally efficient scheme to optimize combining matrices for 2D DoA with compressive antenna arrays. First, it is not bound to a discrete dictionary for the angles, since they are selected during the gradient descent and do not have to reside on any grid. Second, it is possible to extend the model and the optimization to incorporate polarization or bistatic TX-RX setups with applications in channel sounding. Moreover, since the method is not bound to special antenna geometries, it can even cope with measured antenna patterns. Additionally, one has some level of control over the compressed array's behavior by choosing a suitable distribution for $\boldsymbol{\Theta}$. The low complexity also allows to use the proposed method as an online optimization during the measurement process itself by adaptively focusing on certain regions of the parameter space by again selecting the distribution of $\boldsymbol{\Theta}$. Ultimately, one could change the objective function in each gradient step to be something even more suitable for parameter estimation, like the deterministic CRLB.

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] H. L. Van Trees, *Detection, estimation, and modulation theory. Part IV., Optimum array processing*. New York: Wiley-Interscience, 2002.

[3] K. Han, Y. Wang, B. Kou, and W. Hong, "Parameters estimation using a random linear array and compressed sensing," in *International Congress on Image and Signal Processing (CISP)*, vol. 8, Oct 2010, pp. 3950–3954.

[4] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing in MIMO radar with random arrays," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*. IEEE, 2012, pp. 1–6.

[5] S. Shakeri, D. D. Ariananda, and G. Leus, "Direction of arrival estimation using sparse ruler array design," in *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*. IEEE, 2012, pp. 525–529.

[6] M. B. Hawes and W. Liu, "Compressive sensing-based approach to the design of linear robust sparse antenna arrays with physical size constraint," *IET Microwaves, Antennas Propagation*, vol. 8, no. 10, pp. 736–746, July 2014.

[7] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*. IEEE, 2009, pp. 626–629.

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

[9] S. Semper, F. Roemer, 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.

[10] M. Ibrahim, V. Ramireddy, A. Lavrenko, J. König, F. Römer, M. Landmann, M. Grossmann, G. Del 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, 2017.

[11] A. Lavrenko, S. Pawar, M. Ibrahim, F. Roemer, G. D. Galdo, and R. Thomae, "Combining matrix design for 2d doa estimation with compressive antenna arrays," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018, pp. 1–8.

[12] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, 09 1951.

[13] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48.

[14] 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.