

Adaptive Measurement Matrix Design in Direction of Arrival Estimation

Berkan Kilic , *Graduate Student Member, IEEE*, Alper Güngör , *Graduate Student Member, IEEE*, Mert Kalfa , *Member, IEEE*, and Orhan Arıkan , *Member, IEEE*

Abstract—Advances in compressed sensing (CS) theory have brought new perspectives to encoding and decoding of signals with sparse representations. The encoding strategies are determined by measurement matrices whose design is a critical aspect of the CS applications. In this study, we propose a novel measurement matrix design methodology for direction of arrival estimation that adapts to the prior probability distribution on the source scene, and we compare its performance over alternative approaches using both on-grid and gridless reconstruction methods. The proposed technique is derived in closed-form and shown to provide improved compression rates compared to the state-of-the-art. This technique is also robust to the uncertainty in the prior source information. In the presence of significant mutual coupling between antenna elements, the proposed technique is adapted to mitigate these mutual coupling effects.

Index Terms—Measurement matrix design, compressed sensing, direction of arrival estimation, mutual coupling, atomic norm minimization.

I. INTRODUCTION

COMPRESSED sensing (CS) enables the accurate reconstruction of signals with sampling rates below the Nyquist criterion by exploiting their sparse representation in a known dictionary or domain [1], [2], [3], [4]. The successful recovery of a signal requires the use of a sparsifying signal dictionary and a proper sampling strategy. There is a vast literature on the choice of an appropriate dictionary, which may be either learned using model-based training data or can be pre-defined by exploiting a known received signal model [5], [6]. Additionally, there is a variety of known sampling strategies that provide signal samples using a measurement matrix, that is designed to be as incoherent as possible with the dictionary [7]. Random matrices are the most commonly-used measurement matrices since they are incoherent with any dictionary with high probability. However, it

is possible to improve their performance using alternative design criteria [7], [8].

Direction of Arrival (DOA) estimation using sensor/antenna arrays is one field where CS has been successfully applied [9] to overcome the limitations of classical DOA estimation techniques such as Bartlett beamformer [10], [11], Capon's beamformer [12], Multiple Signal Classification (MUSIC) [13]. CS-based DOA estimation techniques can provide improved angular resolution and do not require accurate covariance matrix estimates of the received signal across the antenna array. Furthermore, the new perspective on the sampling theory brought by CS has also important implications for DOA estimation. In a typical phased array antenna in practice, to reduce hardware complexity, analog antenna outputs can be first combined in the analog domain and then digitized to obtain far fewer digital channels than the number of antenna elements. This operation can be equivalently expressed as multiplying the received analog signal with a measurement vector/matrix in order to generate the compressed measurements. CS-based techniques enable unambiguous DOA estimates based on these compressed measurements allowing for a significant reduction in the hardware complexity of the sensor system [14]. The performance of such a sampling strategy highly depends on the amount of relevant information captured by the measurement matrix. Hence, measurement matrix design is a crucial step in any implementation of CS, and an important research area in CS theory [15], [16], [17], [18], [19], [20], [21], [22].

In a typical DOA estimation scenario, the source DOAs are expected to be correlated across different snapshots. Therefore, previous estimates of the tracker can be used to obtain statistical information for a prior DOA distribution on the current measurements. Adaptive techniques can be designed to exploit this prior information for a more reliable estimation. In addition to their usefulness in classical DOA estimation algorithms [11], adaptive techniques have a significant potential to improve the performance of CS-based DOA estimation techniques as well [23]. There are alternative adaptation strategies that can be applied to CS-based DOA estimation. In one such approach, prior information on the signal support over the dictionary is used to formulate a weighted optimization problem in which the sparsity is promoted more on a subset of vector entries [24]. Another approach is the design of measurement matrices in order to perform adaptive measurements in the data acquisition phase [25]. However, a relationship between these two strategies has not been established so far. Here, we propose a novel,

Manuscript received 3 March 2022; revised 8 August 2022 and 20 September 2022; accepted 20 September 2022. Date of publication 26 September 2022; date of current version 7 October 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yuejia Chi. This work was supported in part by The Scientific and Technological Research Council of Turkey, TÜBİTAK under Project CoSAMS-116E006. (Corresponding author: Berkan Kilic.)

Berkan Kilic and Alper Güngör are with ASELSAN Research Center, ASELSAN, Ankara 06200, Turkey, and also with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey (e-mail: bekilic@aselsan.com.tr; alpergungor@aselsan.com.tr).

Mert Kalfa and Orhan Arıkan are with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey (e-mail: kalfa@ee.bilkent.edu.tr; oarikan@ee.bilkent.edu.tr).

Digital Object Identifier 10.1109/TSP.2022.3209880

computationally feasible, and closed-form adaptation strategy for designing measurement matrices that incorporates the prior information on the signal support over the dictionary, which are better suited for real-time applications than the state-of-the-art alternatives. Our proposed Adaptive Measurement Matrix Design (A-MMD) technique is also extended to mitigate the performance degradation due to the mutual coupling between antenna elements, and significantly reduces the necessary number of digital channels for a phased array antenna system. While the design algorithm of A-MMD requires a dictionary and hence a grid, A-MMD can also be used with gridless methods, which we formulate as an atomic norm minimization (ANM) based optimization problem [26]. Although A-MMD is demonstrated on DOA estimation problems, it can be adopted in other applications of CS as well, such as image [8], [25] and audio processing [27].

The paper is organized as follows. In Section II, the basics of CS-based DOA estimation are introduced and discussed. In Section III, various design criteria for measurement matrices and the existing approaches in the literature are reviewed. In Section IV, the proposed measurement matrix design methodology is introduced. The superior performance of this approach compared to the existing alternatives is demonstrated in Section V. Concluding remarks are given in Section VI.

Throughout this paper, uppercase bold characters denote matrices and lowercase bold characters denote vectors. The i^{th} entry of a vector \mathbf{x} is denoted by x_i , the j^{th} column of a matrix \mathbf{X} is denoted by \mathbf{x}_j , the i^{th} entry in the j^{th} column of a matrix \mathbf{X} is denoted by X_{ij} . The ℓ_q -norm of \mathbf{x} is denoted by $\|\mathbf{x}\|_q \equiv (\sum_i |x_i|^q)^{1/q}$ where $|x_i|$ is the magnitude of x_i . The ℓ_0 -norm of \mathbf{x} is a special case and is denoted by $\|\mathbf{x}\|_0$ which is the number of non-zero entries in \mathbf{x} . $\|\mathbf{X}\|_F$ is the Frobenius norm of \mathbf{X} which is defined as $\|\mathbf{X}\|_F \equiv (\sum_{i,j} |X_{ij}|^2)^{1/2}$. \mathbf{X}^T and \mathbf{X}^H denote the transpose and the conjugate transpose of \mathbf{X} , respectively. $\text{Tr}(\mathbf{X})$ denotes the trace of \mathbf{X} and $\text{diag}(\mathbf{X})$ is a vector comprised of the diagonal entries of \mathbf{X} . $\text{diag}(\mathbf{x})$ is the diagonal matrix with diagonal entries \mathbf{x} . The vectors and matrices of ones or zeros are denoted by $\mathbf{1}$ and $\mathbf{0}$, respectively. $\mathbf{X} \succeq \mathbf{0}$ means that \mathbf{X} is positive semi-definite. \mathbf{I}_N is the identity matrix of size $N \times N$.

II. COMPRESSED SENSING BASED DIRECTION OF ARRIVAL ESTIMATION

The DOA estimation problem for narrowband, far-field sources using passive antenna/sensor arrays can be modeled as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (1)$$

where \mathbf{x} denotes the received signal by the array, \mathbf{s} denotes the vector of unknown source signals, and \mathbf{n} is the measurement noise. Here, \mathbf{A} is the steering matrix that is determined by the array geometry, the carrier frequency and the DOAs of the source signals. For an array of M antenna elements and K far-field sources, we have $\mathbf{x} \in \mathbb{C}^{M \times 1}$, $\mathbf{A} \in \mathbb{C}^{M \times K}$, and $\mathbf{s} \in \mathbb{C}^{K \times 1}$.

Let θ_k , $1 \leq k \leq K$, denote the DOA of k^{th} source, then we have $\mathbf{A} = [\mathbf{a}(\theta_1) \dots \mathbf{a}(\theta_K)]$ where the steering vectors $\mathbf{a}(\theta_k)$'s are the columns of \mathbf{A} . For example, for an isotropic,

uniform linear array (ULA) with element spacing d , $\mathbf{a}(\theta_k) = [1 e^{j\omega(\theta_k)} \dots e^{j(M-1)\omega(\theta_k)}]^T / \sqrt{M}$ and $\omega(\theta) = 2\pi d \cos(\theta) / \lambda$ where λ is the wavelength of the impinging plane wave. The source DOA coordinate θ is the angle which increases clockwise and $\theta = 90^\circ$ corresponds to the boresight of the array. Finally, the real and imaginary parts of \mathbf{n} in (1) are assumed to be i.i.d. Gaussian random variables with zero mean and a standard deviation of $\sigma / \sqrt{2M}$, which implies that $\mathbb{E}[\mathbf{n}] = \mathbf{0}$, $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \sigma^2 \mathbf{I}_M$, and $\mathbb{E}[\mathbf{n}\mathbf{n}^T] = \mathbf{0}$.

A. On-Grid DOA Estimation

The CS-based, on-grid DOA estimation problem has the measurement model:

$$\bar{\mathbf{x}} = \mathbf{D}\bar{\mathbf{s}} + \mathbf{n}, \quad (2)$$

where \mathbf{D} is a predefined signal dictionary whose columns are the steering vectors corresponding to the sources that are assumed to be located on L predefined directions $\bar{\theta}_i$'s, $1 \leq i \leq L$. We define ω_i , where $\omega_i \equiv \omega(\bar{\theta}_i)$, as the grid point. Note that we use $\bar{\mathbf{x}}$ and $\bar{\mathbf{s}}$ instead of \mathbf{x} and \mathbf{s} to emphasize that the models (1) and (2) are fundamentally different. Here, $\bar{\mathbf{x}} \in \mathbb{C}^{M \times 1}$, $\bar{\mathbf{s}} \in \mathbb{C}^{L \times 1}$, and $\mathbf{D} \in \mathbb{C}^{M \times L}$ where $\mathbf{D} = [\mathbf{a}(\omega_1) \dots \mathbf{a}(\omega_L)]$. Since the sources may appear in any direction, the columns of \mathbf{A} in (1) may not be among the columns of \mathbf{D} in (2) which results in performance degradation. This is known as the off-grid problem [26] and a well-known issue also for many of the classical techniques including Bartlett beamformer, Capon's beamformer, and MUSIC. L can be increased to reduce the off-grid error, since the estimation resolution and L are related. However, both the computational requirements and the mutual coherence of \mathbf{D} , defined as the magnitude-wise maximum of its columns' normalized inner products, increase with increasing L . While increasing the mutual coherence weakens the recovery guarantees, mutual coherence bounds are generally pessimistic [28] and increasing mutual coherence does not necessarily result in performance degradation [8]. On the other hand, it has been shown that the mutual coherence related metrics are important for the measurement matrix design in CS methods since they promote receiving incoherent measurements [7], [8] and enable adaptive data acquisition strategies [21], as we discuss in Section III and demonstrate in Section V.

In CS-based reconstruction, the sparsity of $\bar{\mathbf{s}}$ in (2) is exploited to estimate the unknown DOAs in the following optimization problem:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\bar{\mathbf{s}}} \|\bar{\mathbf{s}}\|_0 \\ \text{s.t. } &\|\mathbf{x} - \mathbf{D}\bar{\mathbf{s}}\|_2 \leq \epsilon. \end{aligned} \quad (3)$$

Unfortunately, (3) is an NP-hard optimization problem requiring combinatorial search over the grid points [29]. There are computationally-efficient algorithms to approximate the solution of (3) such as Orthogonal Matching Pursuit (OMP) [30], and Basis Pursuit Denoising (BPDN) [31], [32]. While OMP is a greedy algorithm, BPDN replaces the ℓ_0 -norm with the ℓ_1 -norm to obtain a convex relaxation of the optimization problem that can be solved by linear programming [4]. Indeed, the ℓ_1 -norm

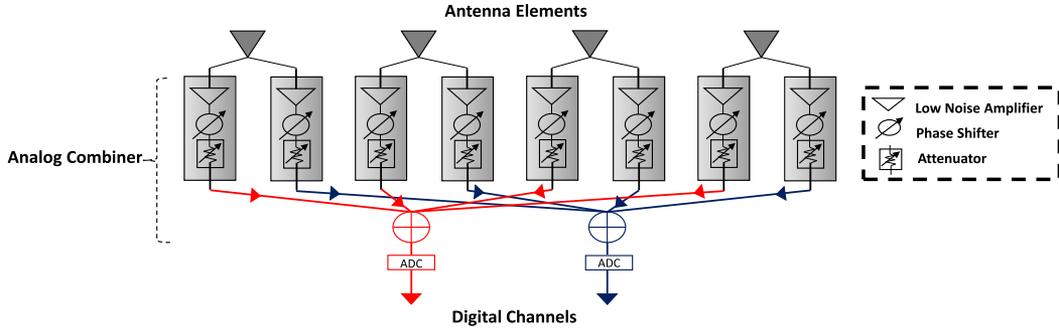


Fig. 1. A phased array system consisting of 4 antenna elements each having an in-phase and quadrature channel output that is reduced to 2 digital channels through an analog combiner. The phase shifting and attenuation determine the entries of the corresponding measurement matrix.

approach has a long history in the DOA literature, being advantageous for estimation accuracy and robustness [9].

As mentioned in Section I, CS-based techniques enable unambiguous estimation using *compressed measurements* [14]. The analog signal \mathbf{x} is undersampled by the measurement matrix Φ to obtain the compressed signal \mathbf{y} as:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \mathbf{A} \mathbf{s} + \Phi \mathbf{n}. \quad (4)$$

From an analog hardware implementation perspective, in an antenna array, as shown in Fig. 1, measurement matrices can be realized using low noise amplifiers, attenuators, and phase shifters. Following these analog operations, the resulting analog outputs are digitized to obtain a set of digital channels. Then, the measurement matrix Φ becomes an $m \times M$ complex valued matrix where $m < M$ is the number of digital channels.

With the inclusion of Φ and by relaxing the ℓ_0 -norm with the ℓ_1 -norm, the following convex optimization problem can be formulated [31], [32]:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\mathbf{s}} \|\mathbf{s}\|_1 \\ \text{s.t. } \|\mathbf{y} - \Phi \mathbf{D} \mathbf{s}\|_2 &\leq \bar{\beta}. \end{aligned} \quad (5)$$

While the noise \mathbf{n} in (1) is assumed to be spatially white [10], $\Phi \mathbf{n}$ is colored in general. Hence, the least squares estimator (LSE) and the maximum likelihood estimator (MLE) are not equivalent in this case. In [33], the MLE-based framework in the data fidelity is used leading to the replacement of $\|\mathbf{y} - \Phi \mathbf{D} \mathbf{s}\|_2$ with $\|(\Phi \Phi^H)^{-1/2}(\mathbf{y} - \Phi \mathbf{D} \mathbf{s})\|_2$ to improve the recovery performance. This approach actually whitens the measurements. The whitening matrix \mathbf{W} is chosen as $\mathbf{W} = \Sigma^{-1/2} \mathbf{U}^H$, where $\Phi \Phi^H = \mathbf{U} \Sigma \mathbf{U}^H$ with a unitary \mathbf{U} and diagonal Σ . The corresponding DOA estimates can be obtained as:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\mathbf{s}} \|\mathbf{s}\|_1 \\ \text{s.t. } \|\mathbf{W} \mathbf{y} - \mathbf{W} \Phi \mathbf{D} \mathbf{s}\|_2 &\leq \beta. \end{aligned} \quad (6)$$

There are many efficient algorithms to solve (6), including the SPGL1 [34], [35] and the alternating direction method of multipliers (ADMM) based methods [36], [37]. The results presented in Section V are obtained using an ADMM based algorithm. In (6), β determines the trade-off between sparsity and data fidelity, and can be chosen as $\sqrt{\mathbb{E}[\|\mathbf{W} \Phi \mathbf{n}\|_2^2]} = \sqrt{\mathbb{E}[\text{Tr}(\mathbf{W} \Phi \mathbf{n} (\mathbf{W} \Phi \mathbf{n})^H)]} = \sigma \sqrt{m}$. In practice, $\beta = c \sigma \sqrt{m}$ is

used where c typically resides in the interval $[0.1, 10]$ to achieve better performance.

B. Gridless DOA Estimation

Gridless CS-based techniques perform their estimation on a continuous spectrum and hence avoid the grid mismatch problem [26], [38], [39], [40], [41], [42], [43], [44], [45]. The application of gridless approaches using a measurement matrix such as the one shown in Fig. 1 is currently limited to sub-Gaussian measurement matrices [41], [42]. Although we shift our main focus on the grid-based reconstruction methods, we also formulate how the gridless techniques can be used along with our proposed methodology. To this end, we use the ANM, which poses the DOA estimation problem, first as a semi-definite programming (SDP) problem, then estimates the DOAs using the obtained solution [40], [43]. Following the convention in [43], atomic norm is defined through an optimization problem involving the infimum point of a set as:

$$\mathcal{A} = \{\mathbf{a}(\theta) \phi \mid \phi \in \mathbb{C}, |\phi| = 1, \theta \in [0, 180^\circ]\}, \quad (7)$$

$$\|\mathbf{x}\|_{\mathcal{A}} = \inf \{t > 0 \mid \mathbf{x} \in t \text{conv}(\mathcal{A})\}, \quad (8)$$

where $\text{conv}(\mathcal{A})$ is the convex hull of \mathcal{A} . For compressed sensing based gridless DOA estimation using ANM, the following problem is solved:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{\mathcal{A}} \\ \text{s.t. } \|\mathbf{W} \mathbf{y} - \mathbf{W} \Phi \mathbf{x}\|_2 &\leq \beta. \end{aligned} \quad (9)$$

Indeed, (9) is highly related with (5). In ℓ_1 -norm approach, the sparsity is promoted over a dictionary, while the sparsity is promoted over the atomic set (7) in ANM. Hence, ANM can be considered as the continuous modification of ℓ_1 -norm approach where the grid is infinitely dense [26]. The ANM problem in (9) can be recast to be solved using an SDP solver [26], [43]:

$$\begin{aligned} \min_{z, \mathbf{r}, \mathbf{x}} \frac{1}{2} z + \frac{1}{2} r_0 \\ \text{s.t. } \begin{bmatrix} z & \mathbf{x}^H \\ \mathbf{x} & \mathbf{R} \end{bmatrix} \succeq 0, \|\mathbf{W} \mathbf{y} - \mathbf{W} \Phi \mathbf{x}\|_2 \leq \beta, \end{aligned} \quad (10)$$

where \mathbf{R} is a Hermitian Toeplitz matrix characterized by the vector $\mathbf{r} \in \mathbb{C}^{M \times 1}$ as $R_{kl} = r_{l-k}$, $1 \leq k \leq l \leq M$, and r_0 is the first element of \mathbf{r} . In (10), \mathbf{x} denotes the uncompressed received

signal by a ULA and \mathbf{R} corresponds to its covariance matrix allowing the unique decomposition of \mathbf{R} using the Vandermonde decomposition of Toeplitz matrices [26] as $\mathbf{R} = \mathbf{A} \text{diag}(\mathbf{s}) \mathbf{A}^H$, where \mathbf{A} is the steering matrix comprised of columns $\mathbf{a}(\theta_i)$'s for $1 \leq i \leq K$. Since (10) is valid for any Φ , the changes in the antenna array can be reflected in (10) by replacing Φ with $\Phi \mathcal{G}$ where the array properties such as array element displacements and mutual coupling effects are embedded in \mathcal{G} as explained in [41]. In this study, we used SDPT3 solver [46] of CVX [47] to solve (10). Then, we used generalized pencil of functions to estimate the DOAs using the Vandermonde decomposition of \mathbf{R} [40], [48]. In terms of the computational complexity, solving (10) is $O(M^{3.5})$ if an interior-point method such as SDPT3 is used [49], while it can be reduced to $O(M^2)$ as proposed in [50]. On the other hand, solving (6) is $O(L \log L)$ per iteration using an ADMM-based implementation [37].

While our focus is on single-snapshot signal model, the extension to the multi-snapshot case has been extensively studied in the literature for both ℓ_1 -norm minimization and ANM. In the multi-snapshot case, dimensionality reduction methods can be applied on the received data for computational efficiency. Let N_s denote the number of snapshots, then the computational complexity becomes independent of N_s if $N_s > M$ for ANM [26], and if $N_s > K$ for ℓ_1 -norm minimization [9].

III. REVIEW OF THE MEASUREMENT MATRIX DESIGN APPROACHES

In this section, we present a comprehensive review of the measurement matrix design approaches with respect to their criterion of optimality. The restricted isometry property (RIP) [4] and the mutual coherence [51] of the effective dictionary $\mathbf{B} \equiv \Phi \mathbf{D}$ are the two most common criteria to assess the stable recovery performance of (5) or (6). Although the RIP provides a tighter bound [29], it is NP-hard to evaluate [52]. The mutual coherence μ of \mathbf{B} is defined as the maximum of $\mu_{ij}(\mathbf{B})$ for $i \neq j$ where:

$$\mu_{ij}(\mathbf{B}) = \frac{|\mathbf{b}_i^H \mathbf{b}_j|}{\|\mathbf{b}_i\|_2 \|\mathbf{b}_j\|_2} = \frac{|G_{ij}|}{\sqrt{|G_{ii}| |G_{jj}|}}. \quad (11)$$

In (11), $\mathbf{G} \equiv \mathbf{B}^H \mathbf{B}$ is the Gram matrix of \mathbf{B} . The mutual coherence does not necessarily reflect the true recovery performance of the algorithms like BPDN and OMP [8]. In [8], the following t -averaged mutual coherence metric is proposed:

$$\mu_t(\mathbf{B}) \equiv \frac{\sum_{i \neq j} \mu_{ij}(\mathbf{B}) \mathbb{I}(\mu_{ij}(\mathbf{B}), t)}{\sum_{i \neq j} \mathbb{I}(\mu_{ij}(\mathbf{B}), t)}, \quad \mathbb{I}(x, t) \equiv \begin{cases} 1, & \text{if } x \geq t \\ 0, & \text{else} \end{cases}, \quad (12)$$

with the aim of excluding smaller μ_{ij} values than t . For increasing t , the t -averaged mutual coherence converges to the mutual coherence. Even though most of the measurement matrix design techniques do not directly attempt to minimize the t -averaged mutual coherence, it provides important insights.

In adaptive approaches, the main purpose is to design a measurement matrix that compresses the analog antenna measurements adaptively for CS-based reconstructions. Indeed, linearly combining the analog antenna measurements to decrease the number of digital channels is not a new topic. Long before the use

of CS techniques in DOA applications, various techniques have been proposed for optimally preprocessing the analog antenna outputs prior to their digitization [53], [54].

We can broadly classify measurement matrix design techniques in three different categories based on their respective criterion of optimality: the mutual coherence, the Cramer Rao Lower Bound (CRLB), and the mutual information (MI).

A. Mutual Coherence Based Measurement Matrix Design

Mutual coherence based measurement matrix design can be formulated as the solution to the following optimization problem:

$$\begin{aligned} \hat{\mathbf{G}} &= \arg \min_{\mathbf{G}} \|\mathbf{G} - \mathbf{T}\|_q \\ \text{s.t. } \mathbf{G} &\succeq \mathbf{0}, \text{rank}(\mathbf{G}) \leq m, \end{aligned} \quad (13)$$

where \mathbf{T} denotes the target matrix for $\mathbf{G} \equiv (\Phi \mathbf{D})^H (\Phi \mathbf{D})$. The reason why we call (13) as *mutual coherence based* is because of the relationship given in (11). Typically, the dictionary \mathbf{D} is assumed to be fixed and Φ is the variable of minimization. Following the results presented in [7], for computational efficiency, q is typically chosen as the Frobenius norm [55], [56]. Indeed, when $\mathbf{T} = \mathbf{I}_L$, $\|\mathbf{G} - \mathbf{T}\|_F$ is closely related to the t -averaged mutual coherence [57]. In [15] and [58], \mathbf{T} is chosen as $\mathbf{D}^H \mathbf{D}$ to set \mathbf{T} as the Gram matrix of the effective dictionary when no compression is performed, but this has been shown to be unsuitable for some signal models [21]. In [21], we proposed using the element-wise magnitude of $\mathbf{D}^H \mathbf{D}$:

$$T_{ij} = |\mathbf{d}^H(\omega_i) \mathbf{d}(\omega_j)|, \quad 1 \leq i, j \leq L. \quad (14)$$

The equiangular tight frame (ETF) criterion [59], [60] can also be used in measurement matrix design [61], [62]. When $\mathbf{T} = \mathbf{I}_L$, the different columns of \mathbf{G} are forced to be orthogonal. However, there is a bound on the achievable minimum of mutual coherence, called the Welch bound [63]:

$$\mu_w = \sqrt{\frac{L-m}{m(L-1)}}, \quad (15)$$

for any matrix with dimensions $m \times L$ with $L > m$. In ETF-based designs, \mathbf{T} is typically chosen as a target Gram matrix from the following convex set [60]:

$$\mathcal{T} = \left\{ \mathbf{T} \in \mathbb{C}^{L \times L} : \mathbf{T}^H = \mathbf{T}, \text{diag}(\mathbf{T}) = \mathbf{1}, \max_{i \neq j} |T_{ij}| \leq \mu_w \right\}. \quad (16)$$

In addition to the non-adaptive methods mentioned above, there are also adaptive mutual coherence based measurement matrix design techniques [16], [21], [64]. In [16], (13) is solved after the off-diagonal entries of \mathbf{T} are set to 0's and the diagonal entries are set to 1 or 0 depending on whether a source is expected to be located on the corresponding grid point or not. For ULAs with isotropic antenna elements and uniform spacing of $\omega(\theta)$ over the grid points ω_i 's, a closed-form solution for Φ is obtained for a fixed dictionary \mathbf{D} satisfying $\mathbf{D} \mathbf{D}^H = \mathbf{I}_M$. In Section IV, we provide the closed-form solution for Φ for any \mathbf{D} and \mathbf{T} .

In [64], the objective function in (13) is replaced by $\|\mathbf{\Gamma}(\mathbf{G} - \mathbf{I}_L)\mathbf{\Gamma}\|_F^2$, where $\mathbf{\Gamma} \equiv \text{diag}(\boldsymbol{\gamma})$ and the vector $\boldsymbol{\gamma}$ is defined as:

$$\boldsymbol{\gamma} = \tau \mathbf{1}_L + (1 - \tau) \frac{\sqrt{|\tilde{\mathbf{s}}|}}{\max_i \sqrt{|\tilde{s}_i|}}, \quad (17)$$

where $\tilde{\mathbf{s}}$ reflects the prior support information and $\tau \in (0, 1)$ is a user-defined parameter. $\sqrt{|\tilde{\mathbf{s}}|}$ is a vector where we apply element-wise magnitude and square-root operators to each entry of $\tilde{\mathbf{s}}$. The resulting problem is solved using the iterative majorization minimization algorithm to find Φ for a fixed \mathbf{D} . In Section IV, a closed-form solution is provided.

In summary, in [16], a closed-form expression for Φ is achieved. This is a significant result for the applications that require real-time measurement matrix updates for each snapshot of a sensor data stream. However, the result is obtained for ULAs with isotropic antenna elements and uniform spacing of $\omega(\theta)$ over the grid points ω_i 's. In [64], the measurement matrix is obtained by the iterative majorization minimization algorithm which is both computationally more expensive and less accurate than obtaining the solution using a closed-form expression. Moreover, both of these techniques are highly sensitive to the accuracy of the prior information, as demonstrated in Section V. In [21], we addressed these issues and proposed a measurement matrix design methodology with a closed-form solution that is robust to the uncertainties in the prior information. However, we had not investigated the relationship between the prior support information and the adaptive measurement matrix design. Hence, the adaptation had been achieved by only changing the diagonal entries of \mathbf{T} as in [16]. As detailed in Section IV, the proposed A-MMD technique not only provides a closed-form solution for the measurement matrix, but has a strong justification for handling the prior information, alleviating the issues associated with the approaches in [16], [21], and [64].

B. The CRLB Based Measurement Matrix Design

The CRLB provides the minimum mean-squared error (MSE) that can be achieved by any unbiased estimator [65]. Minimization of the CRLB to find an optimal pre-processing of analog antenna measurements had also been used before the invention of CS-based DOA estimation techniques [53]. Following the development of CS approaches, the minimization of the CRLB has been used for measurement matrix design [17], [18], [66]. The general form of the optimization task in such approaches can be stated as:

$$\hat{\Phi} = \arg \min_{\Phi} \text{CRLB}(\mathbf{A}, \Phi, \mathbf{s}, \sigma), \quad (18)$$

subject to application-specific constraints. Note that the explicit form of $\text{CRLB}(\mathbf{A}, \Phi, \mathbf{s}, \sigma)$ depends on the assumed model.

In [66], by minimizing the CRLB, a closed-form expression for the measurement matrix is derived. However, the dictionary \mathbf{D} is chosen to represent the shifted unit sample sequences which is not applicable for DOA applications.

As in [66], a closed-form expression for the optimal measurement matrix is found in [17]. However, it is assumed that there is a single source with a known amplitude. This restricts

the optimality of the algorithm to single-source scenarios and the availability of reliable amplitude information on the source signals.

In [18], the proposed methodology is stated as suitable for multi-source scenarios. However, the proposed extension to multi-source scenarios is highly restricted. Furthermore, the presented design is obtained as the solution to a non-convex optimization problem that has a complicated cost surface, requiring the use of global optimization techniques, that are too slow and computationally challenging for real-time applications.

C. MI Based Measurement Matrix Design

The maximum MI criterion [67] is another criterion that has been used for measurement matrix design [19], [20], [68]. The basic idea is to maximize the MI or the conditional MI between the measurements and the vector of unknowns (the DOAs). The gradient of the MI, $\nabla_{\Phi} I(\mathbf{y}, \theta)$, can be expressed as [19], [20]:

$$\nabla_{\Phi} I(\mathbf{y}, \theta) = \nabla_{\Phi} \mathbb{E}_{\mathbf{y}, \theta} [\log f(\mathbf{y}|\theta)] - \nabla_{\Phi} \mathbb{E}_{\mathbf{y}} [\log f(\mathbf{y})], \quad (19)$$

where f is the probability density function of the compressed measurements. In the absence of a closed-form expression, $\nabla_{\Phi} I(\mathbf{y}, \theta)$ is numerically approximated and Φ is updated as:

$$\Phi^{(k+1)} = \Phi^{(k)} + \mu \nabla_{\Phi^{(k)}} I(\mathbf{y}, \theta), \quad (20)$$

where k is the iteration number and μ is the step-size [19]. By running the algorithm with several initializations, the desired $\hat{\Phi}$ is obtained [20].

The techniques in [19], [20], [68] require solutions to non-convex optimization problems over complicated cost surfaces using global optimization. Therefore, their computational requirements limit their real-time applicability.

IV. THE PROPOSED MEASUREMENT MATRIX DESIGN METHODOLOGY: A-MMD

In this section, we present the proposed A-MMD technique that provides a closed-form expression for the measurement matrix which is obtained as the solution to the following reweighted ℓ_1 minimization problem that exploits the prior information on the DOA distribution:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\bar{\mathbf{s}}} \|\mathbf{P}^{-1} \bar{\mathbf{s}}\|_1 \\ \text{s.t. } &\|\mathbf{W} \mathbf{y} - \mathbf{W} \Phi \mathbf{D} \bar{\mathbf{s}}\|_2 \leq \beta, \end{aligned} \quad (21)$$

where the diagonal entries of \mathbf{P} denote the prior probability distribution on each grid point, promoting sparsity over those grid points with less likely DOAs. Note that the minimization variable $\bar{\mathbf{s}}$ is weighted in (21), which is the only difference from (6). By defining $\mathbf{s}_p \equiv \mathbf{P}^{-1} \bar{\mathbf{s}}$, (21) can be rewritten as:

$$\begin{aligned} \hat{\mathbf{s}}_p &= \arg \min_{\mathbf{s}_p} \|\mathbf{s}_p\|_1 \\ \text{s.t. } &\|\mathbf{W} \mathbf{y} - \mathbf{W} \Phi \mathbf{D} \mathbf{P} \mathbf{s}_p\|_2 \leq \beta. \end{aligned} \quad (22)$$

By solving (22), the DOAs are found as $\hat{\mathbf{s}} = \mathbf{P} \hat{\mathbf{s}}_p$. Note that in (22), the effective dictionary changes from \mathbf{B} to $\mathbf{B}_p \equiv \Phi \mathbf{D}_p$ where $\mathbf{D}_p \equiv \mathbf{D} \mathbf{P}$. Hence, the corresponding Gram matrix becomes $\mathbf{G}_p \equiv \mathbf{B}_p^H \mathbf{B}_p$. \mathbf{G}_p does not involve the term \mathbf{W} since

whitening does not affect the data acquisition stage of a DOA estimation system. In other words, \mathbf{W} can be implemented as a decoupled operation following reception and digitization of \mathbf{y} . Similar to the mutual coherence based measurement matrix designs, the following optimization problem can be cast in terms of the Gram matrix \mathbf{G}_p :

$$\begin{aligned} \hat{\mathbf{G}}_p &= \arg \min_{\mathbf{G}_p} \|\mathbf{G}_p - \mathbf{T}\|_F^2 \\ \text{s.t. } \mathbf{G}_p &\succeq \mathbf{0}, \text{rank}(\mathbf{G}_p) \leq m. \end{aligned} \quad (23)$$

Since the dimensions of Φ are $m \times M$ and $m \leq M$, $\text{rank}(\mathbf{G}_p) \leq m$ is guaranteed to be satisfied. Furthermore, $\mathbf{G}_p \equiv \mathbf{B}_p^H \mathbf{B}_p$ also satisfies $\mathbf{G}_p \succeq \mathbf{0}$. Then, for a fixed \mathbf{D} , (23) can be written as an unconstrained optimization problem over the measurement matrix Φ as:

$$\hat{\Phi} = \arg \min_{\Phi} \|(\Phi \mathbf{D}_p)^H (\Phi \mathbf{D}_p) - \mathbf{T}\|_F^2. \quad (24)$$

Let the singular value decomposition (SVD) of \mathbf{D}_p be $\mathbf{U}_{D_p} \Sigma_{D_p} \mathbf{V}_{D_p}^H$ where Σ_{D_p} is diagonal, \mathbf{U}_{D_p} is unitary and \mathbf{V}_{D_p} is semi-unitary, i.e., $\mathbf{U}_{D_p}^H \mathbf{U}_{D_p} = \mathbf{U}_{D_p} \mathbf{U}_{D_p}^H = \mathbf{I}_M$ and $\mathbf{V}_{D_p}^H \mathbf{V}_{D_p} = \mathbf{I}_M$. Then, (24) can be written as:

$$\hat{\Phi} = \arg \min_{\Phi} \|\mathbf{V}_{D_p} \Sigma_{D_p}^H \mathbf{U}_{D_p}^H \Phi^H \Phi \mathbf{U}_{D_p} \Sigma_{D_p} \mathbf{V}_{D_p}^H - \mathbf{T}\|_F^2. \quad (25)$$

By defining $\Psi \equiv \Phi \mathbf{U}_{D_p} \Sigma_{D_p}$, the following optimization problem for Ψ is obtained:

$$\hat{\Psi} = \arg \min_{\Psi} \|\mathbf{V}_{D_p} \Psi^H \Psi \mathbf{V}_{D_p}^H - \mathbf{T}\|_F^2. \quad (26)$$

Expanding the Frobenius norm as a trace and using the cyclic property and linearity of trace operator, we get:

$$\begin{aligned} J(\Psi) &= \text{Tr}((\mathbf{V}_{D_p} \Psi^H \Psi \mathbf{V}_{D_p}^H - \mathbf{T})(\mathbf{V}_{D_p} \Psi^H \Psi \mathbf{V}_{D_p}^H - \mathbf{T})^H) \\ &= \text{Tr}(\Psi^H \Psi \Psi^H \Psi) - 2\text{Tr}(\Psi^H \Psi \mathbf{V}_{D_p}^H \mathbf{T} \mathbf{V}_{D_p}) \\ &\quad + \text{Tr}(\mathbf{T} \mathbf{T}^H) \\ &= \|\Psi^H \Psi - \mathbf{V}_{D_p}^H \mathbf{T} \mathbf{V}_{D_p}\|_F^2 + C, \end{aligned} \quad (27)$$

where $J(\Psi) \equiv \|\mathbf{V}_{D_p} \Psi^H \Psi \mathbf{V}_{D_p}^H - \mathbf{T}\|_F^2$, and the constant C includes those terms that do not depend on Ψ . Then, (24) reduces to:

$$\hat{\Psi} = \arg \min_{\Psi} \|\Psi^H \Psi - \mathbf{V}_{D_p}^H \mathbf{T} \mathbf{V}_{D_p}\|_F^2, \quad (28)$$

which is a quadratic optimization problem that can be recast in terms of the Gram matrix $\mathbf{G}_\Psi = \Psi^H \Psi$, and $\mathbf{Z} = \mathbf{V}_{D_p}^H \mathbf{T} \mathbf{V}_{D_p}$ as:

$$\begin{aligned} \hat{\mathbf{G}}_\Psi &= \arg \min_{\mathbf{G}_\Psi} \|\mathbf{G}_\Psi - \mathbf{Z}\|_F^2 \\ \text{s.t. } \mathbf{G}_\Psi &\succeq \mathbf{0}, \text{rank}(\mathbf{G}_\Psi) \leq m. \end{aligned} \quad (29)$$

For a real-valued \mathbf{G}_Ψ , the optimization problem in (29) is solved iteratively in [69] and analytically in [70]. Following [70], the optimal solution of (29) is obtained as:

$$\hat{\mathbf{G}}_\Psi = \sum_{i=1}^{\min\{m, z\}} \lambda_i \mathbf{q}_i \mathbf{q}_i^H, \quad (30)$$

where \mathbf{Z} has the eigendecomposition $\mathbf{Z} = \mathbf{Q} \Lambda \mathbf{Q}^H = \sum_{i=1}^M \lambda_i \mathbf{q}_i \mathbf{q}_i^H$ with $\lambda_1 \geq \dots \geq \lambda_M$, and z denotes the number of non-negative eigenvalues of \mathbf{Z} . Once the optimal positive semi-definite matrix $\hat{\mathbf{G}}_\Psi$ of rank at most m is obtained by (30), $\hat{\Psi}$ that satisfies $\hat{\Psi}^H \hat{\Psi} = \hat{\mathbf{G}}_\Psi$ is found by the eigendecomposition of $\hat{\mathbf{G}}_\Psi$:

$$\hat{\Psi} = \begin{cases} \Lambda_m \mathbf{Q}_m^H, & \text{if } m \leq z \\ [\mathbf{Q}_z \Lambda_z^H \mathbf{0}]^H, & \text{else} \end{cases}, \quad (31)$$

where \mathbf{Q}_m (or \mathbf{Q}_z) is found by taking the first m (or z) columns of \mathbf{Q} , and Λ_m (or Λ_z) is found by taking the first m (or z) columns and rows of Λ . Then, the optimal measurement matrix $\hat{\Phi}$ that solves (29) is found as $\hat{\Phi} = \hat{\Psi} \Sigma_{D_p}^{-1} \mathbf{U}_{D_p}^H$.

So far, for a given \mathbf{P} and \mathbf{T} , a closed-form expression, $\hat{\Phi}$, for the optimal adaptive measurement matrix is obtained. A preferred choice for \mathbf{T} is similar to (14), which we proposed in [21]. One motivation for this choice is to capture the correlation values of the original dictionary \mathbf{D} . Moreover, we reported its robust performance in [21]. However, \mathbf{D} is replaced by \mathbf{D}_p here, and (14) is modified as:

$$T_{ij} = |\mathbf{d}_p^H(\omega_i) \mathbf{d}_p(\omega_j)|, 1 \leq i, j \leq L, \quad (32)$$

where $\mathbf{d}_p(\omega_i)$'s are the columns of \mathbf{D}_p .

The matrix \mathbf{P} that is formed based on the prior probability distribution can be initialized as $\mathbf{P} = \mathbf{I}_L$ if the sources are equally likely to emerge from any direction. However, if the sources are more likely to appear in certain angular sectors, \mathbf{P} can be initialized accordingly. This prior information can be provided by a tracker operating on the previous source detections obtained from the sensor array. Following the DOA estimations, there are sources for which we have prior information, which are called the *tracked sources*. In addition to the tracked sources, the emergence of new sources is also a possibility. These sources are called the *emerging sources*. While the probabilistic angular distribution of the tracked sources is available with some level of certainty, a uniform probability distribution can be assigned to the emerging sources. Specifically, assume that there are K sources and Q of them are tracked sources each having probability mass function (PMF) of $p_q(\omega_i)$, $1 \leq q \leq Q$, on the grid point ω_i . Then, the probability distribution of the overall source scene is closely related to the convex combination of $p_q(\omega_i)$'s and the uniform PMF [21]:

$$p(\omega_i) = \sum_{q=1}^Q \alpha_q p_q(\omega_i) + \frac{(1 - Pr_t)}{L}, 1 \leq i \leq L, \quad (33)$$

where $Pr_t \equiv \sum_{q=1}^Q \alpha_q \leq 1$, and α_q 's can be chosen based on the importance levels of the sources, power of the received source signals that are impinging on the antenna array, and Q/K ratio. While α_q 's for different targets are chosen the same in Section V, it is also possible to assign different α_q values to different sources. For example, if the detection of the i^{th} source is more crucial than the the detection of the j^{th} source, $\alpha_i > \alpha_j$ should be chosen. In another scenario, the received signal power of the i^{th} source might be significantly lower than the received signal power of the j^{th} source. In this case, the

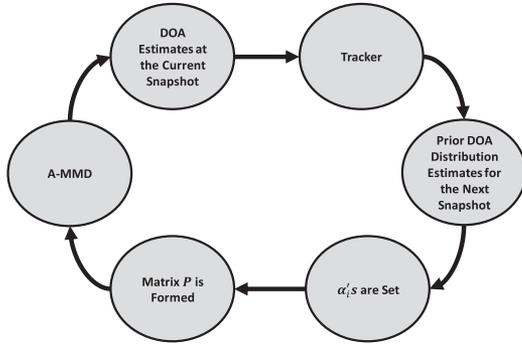


Fig. 2. Application of the proposed A-MMD technique in the presence of a tracker.

Algorithm 1: A-MMD.

Input: D, T, P, m where $m < M < L$

Output: $\hat{\Phi}$

- 1: $D_p \equiv DP$ {Matrix multiplication: $O(ML)$ }
 - 2: $D_p = U_{D_p} \Sigma_{D_p} V_{D_p}^H$ {SVD: $O(M^2 L)$ }
 - 3: $Z \equiv V_{D_p}^H T V_{D_p}$ {Matrix multiplication: $O(ML^2)$ }
 - 4: $Z = Q \Lambda Q^H = \sum_{i=1}^M \lambda_i \mathbf{q}_i \mathbf{q}_i^H$
{Eigendecomposition: $O(M^3)$ }
 - 5: Compute $\hat{\Psi}$ according to (31)
{Matrix multiplication: $O(\min\{m, z\}^2 M)$ }
 - 6: $\hat{\Phi} = \hat{\Psi} \Sigma_{D_p}^{-1} U_{D_p}^H$
{Matrix multiplication: $O(\min\{m, z\} M^2)$ }
-

imbalance between two received signal powers can be alleviated by assigning $\alpha_i > \alpha_j$. After the selection of α parameters, the diagonal matrix \mathbf{P} is determined by $P_{ii} = p(\omega_i)$. The schematic of the overall procedure is given in Fig 2. Note that the tracker implementation is out of the scope of this work, and reliable prior distribution estimates from the tracker are assumed to be available. However, the performance of algorithms with various levels of uncertainty in the prior information is investigated in Section V. The steps of A-MMD and their respective computational complexities are presented in Algorithm 1. Computational complexity is dominated by the computation of \mathbf{Z} in Step 3, with a cost of $O(ML^2)$. \mathbf{D} is a discrete Fourier transform matrix if the array elements are isotropic in a ULA configuration with the uniform spacing of $\omega(\theta)$ over the grid points ω_i 's, $1 \leq i \leq M$. For this case, Steps 2 and 3 are not necessary since $U_{D_p} = \mathbf{D}$, $\Sigma_{D_p} = \mathbf{P}$, $V_{D_p} = \mathbf{I}$ and $\mathbf{Z} = \mathbf{T}$.

So far, we have not considered mutual coupling between the antenna array elements in our design methodology. Mutual coupling can be a significant source of error in DOA applications, especially for most wideband antenna array designs that depend on mutual coupling to reduce active reflection coefficients, e.g., in tapered slot antennas [71] and in connected dipole antennas [72]. If the S-parameters (i.e., network parameters expressing mutual coupling across elements) are known (via simulations or measurements), the mutual coupling effects can be mitigated using a slightly modified measurement matrix design. Let \mathbf{S} denote the matrix whose entries are the S-parameters, then the mutual

coupling matrix can be formed as $\mathbf{M} \equiv \mathbf{I}_M + \mathbf{S}$. The signal model can then be updated as $\mathbf{x} = \mathbf{M}\mathbf{A}\mathbf{s} + \mathbf{n}$. This update can be reflected on the dictionary matrix using $\mathbf{M}\mathbf{D}$ as the input to A-MMD. The resulting design is called Adaptive Measurement Matrix Design with Mutual Coupling Effects (A-MMD-MC). Even when the mutual coupling matrix is not known precisely, which may be the case for some practical applications, DOA estimation algorithms should be robust against mutual coupling effects.

A-MMD allows for joint optimization over both measurement matrix and dictionary. In [73], we introduced a simple but useful adaptive dictionary design. Although the proposed approach is heuristic, it is used to show that A-MMD is suitable for any \mathbf{D} . The resulting method is called Adaptive Measurement Matrix Design with Updated Dictionary (A-MMD-UD). The key idea in [73] is to use a denser grid where the tracked sources are more likely to appear. For this purpose, the following mixture PMF is proposed, as the convex combination of the PMFs of the tracked sources and the uniform PMF:

$$\bar{p}(\omega_i) = \sum_{q=1}^Q \bar{\alpha}_q p_q(\omega_i) + \frac{(1 - \bar{P}r_t)}{L}, \quad 1 \leq i \leq L, \quad (34)$$

whose parameters are different from those in (33). After forming (34), the cumulative distribution function (CDF) F is computed by integrating \bar{f} , which is achieved by interpolating \bar{p} . Then:

$$\bar{\omega}_i = F^{-1} \left(\frac{i}{L-1} \right), \quad 0 \leq i \leq L-1, \quad (35)$$

are the adapted, non-uniformly spaced grid points.

While the A-MMD is derived starting from a reweighted ℓ_1 minimization problem, it does not imply that the reconstruction must be performed by solving (21). The designed measurement matrix changes only the data acquisition strategy, and makes the DOA system more focused on particular regions depending on the prior information. Once \mathbf{y} is received as in (4), the reconstruction can also be performed using the ANM approach given in (9), which is numerically validated in Section V. Note that the ANM approach demonstrated in Section II-B does not include any reweighting operation. On the other hand, reweighted ANM has already been proposed in [74], which can be used along with our proposed A-MMD as a future extension.

V. NUMERICAL RESULTS

In this section, a thorough performance comparison of the measurement matrix design techniques is conducted using Monte Carlo simulations over a large set of scenarios. First, the performances of the various techniques are investigated for a range of Signal-to-Noise Ratio (SNR) levels. Second, we choose the best-performing techniques and continue to investigate their performances by changing other parameters, namely the number of grid points (L), the compression rate (M/m), and the uncertainty in the prior information for a particular SNR level. Third, the mutual coupling effects for an antenna array are included in the simulations. In all of the simulations described above, a conventional $\lambda/2$ spaced ULA structure is used and multipath effects are ignored. Fourth, a random linear

TABLE I
LIST OF THE COMPARED TECHNIQUES

Name and Abbreviation	Brief Explanation	Parameters
Random Gaussian Design (RGD)	Entries of the measurement matrix are drawn from a normal distribution.	-
Identity Matrix Design (IMD)	(24) is solved for $\mathbf{T} = \mathbf{P} = \mathbf{I}_L$.	-
Equiangular Tight Frame based Design (ETFD)	(24) is solved with \mathbf{T} chosen from the set (16) and for $\mathbf{P} = \mathbf{I}_L$.	-
Li, Zhang, Kirubarajan, Rajan (LZKR)	The technique proposed in [64].	τ in (17)
Ibrahim, Roemer, Galdo (IRG)	The technique proposed in [16].	-
Kilic, Güngör, Kalfa, Arıkan (KGKA)	Our previously proposed technique [21].	$\alpha_1, \dots, \alpha_Q$
Adaptive Measurement Matrix Design (A-MMD)	The currently proposed technique.	$\alpha_1, \dots, \alpha_Q$
Adaptive Measurement Matrix Design with Updated Dictionary (A-MMD-UD)	Described in Section IV.	$\alpha_1, \dots, \alpha_Q$ and $\bar{\alpha}_1, \dots, \bar{\alpha}_Q$
Adaptive Measurement Matrix Design with Mutual Coupling Effects (A-MMD-MC)	Described in Section IV.	$\alpha_1, \dots, \alpha_Q$

array (RLA) with multipath fading on the incoming signals is modeled and DOA accuracy is compared to the ULA case. Next, the performance of A-MMD is investigated using the phase transition phenomenon [29], [75], where the effects of adaptive reconstruction and A-MMD are also separately evaluated. Finally, A-MMD is used along with ANM to show its applicability when gridless methods are used in the reconstruction. For the sake of clarity, the techniques which are used in the performance comparison are abbreviated as given in Table I.

Our performance criterion is the root mean squared error (RMSE) between the actual and the estimated DOAs:

$$\text{RMSE} \equiv \sqrt{\frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K (\hat{\theta}_{nk} - \theta_{nk})^2}, \quad (36)$$

where θ_{nk} and $\hat{\theta}_{nk}$ denote the actual and the estimated DOAs (in degrees) for the k^{th} source and the n^{th} Monte Carlo iteration among a total of N trials. For a more statistically reliable comparison, bootstrapping is applied [76]. Specifically, let the n^{th} entry of the vector $\mathbf{h} \in \mathbb{C}^{N \times 1}$ be defined as:

$$h_n = \frac{1}{K} \sum_{k=1}^K (\hat{\theta}_{nk} - \theta_{nk})^2, \quad (37)$$

where h_n is the mean squared error for the n^{th} Monte Carlo run and the square of the RMSE in (36) is the mean of \mathbf{h} . For bootstrapping, \mathbf{h} is randomly sampled N times with replacement, and a new vector is formed. This procedure is repeated J times to form \mathbf{h}^j , $1 \leq j \leq J$, where J is chosen as 10,000. The bootstrap means are then computed as:

$$\bar{\mu}^j = \sqrt{\frac{1}{N} \sum_{n=1}^N h_n^j}. \quad (38)$$

TABLE II
DOA DISTRIBUTION FOR CASE 1

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(-0.5455, 0.2^2)$	$\theta = 100^\circ$ ($\omega = -0.5455$)
2	$\mathcal{N}(1.0745, 0.2^2)$	$\theta = 70^\circ$ ($\omega = 1.0745$)

After the calculation of (38), $\bar{\mu}^j$'s for $1 \leq j \leq J$ are sorted in an ascending manner, and their 2.5th and 97.5th percentile entries are assigned as the lower and upper RMSEs [76]. In our results, we include the lower and upper RMSEs in addition to the original RMSE value, which is calculated using the original data as given in (36). In all the simulations, the DOAs are realized from a continuous distribution across the horizon; hence, off-grid values are allowed and multiple sources can be arbitrarily close to one another.

As shown in Table I; LZKR, KGKA, A-MMD, A-MMD-UD, and A-MMD-MC are all parameter dependent techniques. For LZKR, $\tau \in [0.1, 0.9]$ [64]; for KGKA $\alpha \in [0, 0.95]$ [21], for the dictionary optimization procedure described in Section IV, $\bar{\alpha} \in [0.01, 0.05]$ [73] are used and the best results we could achieve using the parameters in the given ranges are reported. The choice of α used in A-MMD is explained in Section V-B after the presentation of the first set of numerical results. Note that A-MMD-MC and A-MMD-UD always use the same α parameters since they are variants of A-MMD.

A. Performance Comparison At Different SNRs

To compare the performances of the techniques at different SNR levels, we use a ULA with $M = 50$ antenna elements with $\lambda/2$ spacing and digitize the analog output of these elements in $m \in \{5, 10\}$ digital channels. For this array configuration, $\omega(\theta) = \pi \cos(\theta)$. SNR is defined as $10 \log_{10}(mP_s/\sigma^2)$ where P_s is the received power of the source, and is assumed to be the same for all K sources. Note that the array processing gain is already incorporated into the SNR definition. In the simulations, we have $K \in \{2, 3\}$ and $Q \in \{1, 2\}$, so that there are up to three sources and at least one of them is tracked. Except for A-MMD-UD that adaptively assigns the grid points, $L = 3 \times M = 150$ uniformly spaced grid points are chosen. 100,000 Monte Carlo runs, each with independent DOAs and noise realizations, are conducted to produce the tabulated results where the lower, mean, and upper RMSEs are shown. The lower and upper RMSEs are computed using the bootstrap analysis. For the sake of clarity, the mean RMSE of the best performing technique is printed in bold.

Case 1 - $K = 2, Q = 2, m = 10, L = 150$: We first start with a relatively simple scenario where all the techniques are expected to perform well. There exist two tracked sources in the environment and they are close to the boresight direction of the antenna array ($\theta = 90^\circ$). These sources have the DOA distributions given in Table II. The interpretation of Table II is as follows: the DOAs of both sources are chosen at random by sampling a Gaussian distribution with means $-0.5455, 1.0745$; and standard deviation $\sigma_s = 0.2$ (in the ω -domain). The value $\sigma_s = 0.2$ is achieved by $5 \times 2/M = 0.2$ for $M = 50$ antenna elements where $2/M$ is the approximate half-power beamwidth

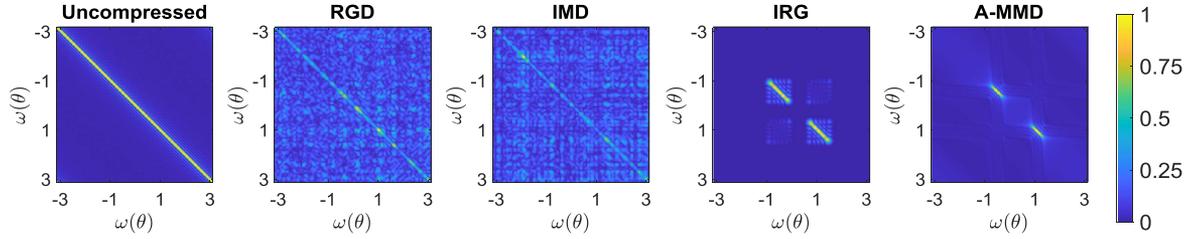


Fig. 3. Magnitudes of the entries in $(\Phi \mathbf{D})^H(\Phi \mathbf{D})$, where Φ is designed according to the respective algorithms. For the leftmost figure $\Phi = \mathbf{I}_L$, i.e., no compression is performed. The same color axis is used for all results (see the colorbar).

TABLE III

THE RMSE [DEG] RESULTS FOR CASE 1. THE LOWER AND UPPER RMSES ACHIEVED BY BOOTSTRAP ANALYSIS ARE ALSO GIVEN

Methods	SNR		
	10 (dB)	25 (dB)	40 (dB)
RGD	30.0/30.1/30.2	12.4/12.6/12.7	10.8/10.9/11.1
IMD	24.0/24.1/24.3	5.13/5.28/5.42	4.17/4.32/4.64
ETFD	22.8/23.0/23.1	5.33/5.47/5.61	4.64/4.78/4.92
LZKR	5.51/5.58/5.65	3.61/3.67/3.72	3.44/3.49/3.55
IRG	3.08/3.16/3.24	1.81/1.89/1.97	1.72/1.81/1.90
KGKA	5.59/5.66/5.73	0.39/0.39/0.41	0.24/ 0.24 /0.24
A-MMD, $\alpha = 0.05$	4.16/4.26/4.36	0.34/ 0.37 /0.41	0.24/ 0.24 /0.24
A-MMD-UD, $\alpha = 0.05$	5.68/5.82/5.96	0.40/0.42/0.44	0.25/0.26/0.26
A-MMD, $\alpha = 0.20$	2.99/ 3.06 /3.13	0.75/0.79/0.81	0.73/0.76/0.79
A-MMD-UD, $\alpha = 0.20$	3.08/3.15/3.22	0.69/0.73/0.76	0.67/0.70/0.73

The lowest mean RMSEs are printed in bold.

of a $\lambda/2$ spaced ULA [77]. For Source 1, the DOAs are chosen randomly over the interval $[84^\circ, 116^\circ]$ and for Source 2, the DOAs are chosen randomly over the interval $[51^\circ, 87^\circ]$. The results are given in Table III. As expected, all the given techniques provide accurate results. Among the non-adaptive techniques, IMD and ETFD outperform RGD, while their performances are far behind all the adaptive techniques. Among the adaptive techniques, A-MMD has the best performance. In the low-SNR regime, a high α improves the performance of A-MMD, since the prior information is more reliable than the noisy measurements and it is available for both sources. In this scenario, the use of an adaptive dictionary does not increase the estimation performance.

We also plot $|\mathbf{d}^H(\omega_i)\Phi^H\Phi\mathbf{d}(\omega_j)|$ for $1 \leq i, j \leq L$ in Fig. 3. When IRG and A-MMD are used, the highest values are located around $\omega = -0.5455$ and $\omega = 1.0745$, which are consistent with the prior information that is embedded in the proposed optimization. Moreover, the (unnormalized) cross-correlation of the columns of $\Phi \mathbf{D}$ are successfully suppressed by these methods, while this is not the case for the results obtained with RGD and IMD. Note that A-MMD does not perform normalization on the columns of the effective dictionary since \mathbf{P} is used to scale those columns.

Case 2 - $K = 2, Q = 1, m = 10, L = 150$: Source 2 in Table II is replaced with an emerging source having a uniform DOA distribution between -0.9π and 0.9π (in the ω -domain) resulting in the DOA distribution given in Table IV. Hence, the DOAs of Source 2 are chosen randomly over the interval $[25^\circ, 155^\circ]$. The results for that scenario are given in Table V. Compared

TABLE IV

DOA DISTRIBUTION FOR CASE 2

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(-0.5455, 0.2^2)$	$\theta = 100^\circ$ ($\omega = -0.5455$)
2	$\mathcal{U}(-0.9\pi, 0.9\pi)$	$\theta = 90^\circ$ ($\omega = 0$)

TABLE V

THE RMSE [DEG] RESULTS FOR CASE 2. THE LOWER AND UPPER RMSES ACHIEVED BY BOOTSTRAP ANALYSIS ARE ALSO GIVEN

Methods	SNR		
	10 (dB)	25 (dB)	40 (dB)
RGD	34.0/34.1/34.3	11.9/12.0/12.2	10.3/10.5/10.7
IMD	29.9/30.1/30.2	8.67/8.86/9.04	7.69/7.88/8.06
ETFD	29.8/30.0/30.1	7.90/8.07/8.24	6.73/6.90/7.07
LZKR	24.9/24.9/25.1	24.9/25.0/25.0	25.0/25.0/25.2
IRG	24.9/25.0/25.1	24.6/24.7/24.8	24.7/24.8/24.9
KGKA	19.4/19.6/19.7	3.66/3.78/3.91	1.52/1.62/1.71
A-MMD	19.4/19.6/19.7	2.31/ 2.43 /2.55	1.19/ 1.30 /1.42
A-MMD-UD	17.2/ 17.3 /17.5	2.65/2.78/2.90	1.22/1.34/1.47

The lowest mean RMSEs are printed in bold.

to Case 1, there is a performance degradation for all the techniques. For adaptive techniques, the reason for this degradation is the lack of prior information about the emerging source. IRG and LZKR almost completely miss the emerging source which can be deduced by observing that their performance does not change much with increasing SNR. A-MMD-UD improves the performance of A-MMD at SNR = 10 dB and SNR = 40 dB. The reason for the performance degradation of the non-adaptive techniques is the DOA of Source 2. In Case 1, where the better results are obtained, Source 2 is located around 70° , where the DOA estimation performance is expected to be more accurate in general. In Case 2, the emerging source may appear anywhere between $\omega = -0.9\pi$ and $\omega = 0.9\pi$. As Source 2 may emerge at DOAs that are far from the array boresight, a degradation in performance is expected. For A-MMD and A-MMD-UD, $\alpha = 0.15$ was chosen.

Case 3 - $K = 2, Q = 1, m = 10, L = 150$: In Case 2, the tracked source is located around the boresight of the antenna array. Now, the mean position of the DOA distribution of the tracked source (Source 1) in Table IV is changed to $\theta = 30^\circ$ ($\omega = 2.7207$). The DOA distributions of the sources are given in Table VI. The target DOAs of Source 1 are chosen randomly over the interval $[0.5^\circ, 54^\circ]$. Compared to Case 1 and Case 2, there is a performance degradation for all the techniques as shown in Table VII. The conventional RGD outperforms both

TABLE VI
DOA DISTRIBUTION FOR CASE 3

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(2.7207, 0.2^2)$	$\theta = 30^\circ$ ($\omega = 2.7207$)
2	$\mathcal{U}(-0.9\pi, 0.9\pi)$	$\theta = 90^\circ$ ($\omega = 0$)

TABLE VII
THE RMSE [DEG] RESULTS FOR CASE 3. THE LOWER AND UPPER RMSES ACHIEVED BY BOOTSTRAP ANALYSIS ARE ALSO GIVEN

Methods	SNR		
	10 (dB)	25 (dB)	40 (dB)
RGD	32.9/33.0/33.2	11.8/12.0/12.2	11.3/11.5/11.7
IMD	37.3/37.5/37.6	14.1/14.3/14.5	13.0/13.2/13.4
ETFD	36.8/36.9/37.1	15.4/15.6/15.8	14.1/14.3/14.5
LZKR	45.1/45.2/45.4	42.4/42.5/42.6	42.4/42.5/42.6
IRG	45.6/45.7/45.8	41.8/41.9/42.1	40.6/40.7/40.9
KGKA	20.0/20.2/20.4	9.20/9.41/9.63	8.25/8.47/8.69
A-MMD	32.0/32.2/32.4	6.90/7.12/7.33	6.41/6.64/6.86
A-MMD-UD	20.5/ 20.6 /20.8	5.06/ 5.27 /5.47	3.28/ 3.49 /3.70

The lowest mean RMSEs are printed in bold.

TABLE VIII
THE RMSE [DEG] RESULTS FOR CASE 4. THE LOWER AND UPPER RMSES ACHIEVED BY BOOTSTRAP ANALYSIS ARE ALSO GIVEN

Methods	SNR		
	10 (dB)	25 (dB)	40 (dB)
RGD	37.0/37.2/37.3	33.3/33.5/33.6	33.0/33.1/33.3
IMD	42.1/42.2/42.3	39.2/39.3/39.5	38.9/39.1/39.2
ETFD	37.6/37.7/37.9	33.7/33.8/34.0	33.4/33.5/33.7
LZKR	24.8/24.9/24.9	24.8/24.9/25.0	24.8/24.9/25.0
IRG	24.8/24.8/24.9	24.8/24.8/24.9	24.7/24.8/24.9
KGKA	20.2/20.4/20.5	7.14/7.25/7.37	4.17/4.27/4.37
A-MMD	16.9/17.0/17.2	4.75/4.86/4.97	3.04/3.13/3.23
A-MMD-UD	16.3/ 16.4 /16.5	4.71/ 4.81 /4.91	3.03/ 3.12 /3.21

The lowest mean RMSEs are printed in bold.

IMD and ETFD implying that these designs are not robust to the changes in the source DOAs. Only KGKA, A-MMD, and A-MMD-UD can provide DOA estimates with RMSE $< 10^\circ$. Among these three, KGKA is worse than the other two and A-MMD-UD performs the best. For A-MMD and A-MMD-UD, $\alpha = 0.15$ was chosen.

Case 4 - $K = 2, Q = 1, m = 5, L = 150$: The same scenario given in Table IV is repeated for $m = 5$. Hence, the compression rate is now $M/m = 10$. Examination of such a scenario is important since reducing the number of digital channels drastically simplifies the hardware and software complexity. The results are demonstrated in Table VIII. As the number of compressed measurements is decreased, the RMSEs of all the techniques increase compared to those in Table V. As before, only KGKA, A-MMD, and A-MMD-UD can obtain RMSE $< 10^\circ$ with A-MMD-UD providing the best performance. For A-MMD and A-MMD-UD, $\alpha = 0.025$ was chosen.

B. Selection of Parameters

Note that $\alpha = 0.15$ is used in Case 2 while $\alpha = 0.025$ is used in Case 4, although the same scenario is simulated except for decreasing m from 10 to 5. α serves as a parameter adjusting the *resource allocation*, so less resources are allocated for the

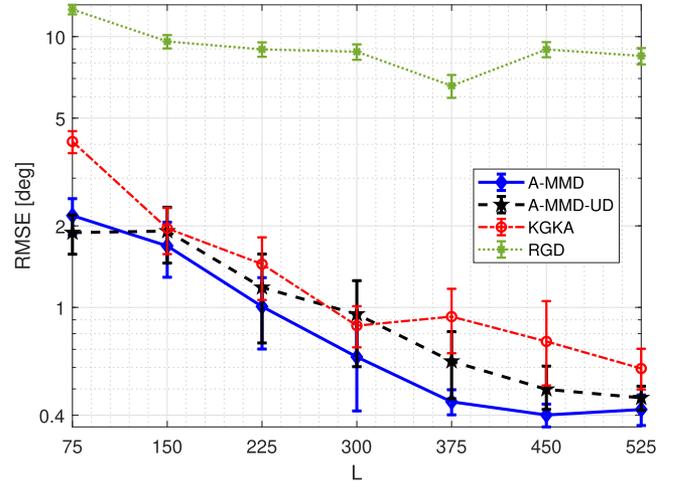


Fig. 4. RMSE vs the number of grid points.

emerging sources when it is increased. When m is decreased for a fixed M , the total allocation of resources decreases and setting $\alpha = 0.15$ makes it difficult to find the DOA of the emerging source. Hence, the m/M ratio is important to determine α . Moreover, α must be inversely proportional to K . Therefore, $\alpha = m/(MK)$ is always used for A-MMD, A-MMD-UD, and A-MMD-MC in the remainder of this manuscript. Note that this choice does not directly consider the noise power, which is shown to affect the optimal parameter choice in Case 1. Furthermore, different array configurations can also affect the optimal parameter choice, which is not comprehensively investigated in this study. An improved parameter selection algorithm for α can be developed, which is left as a future study.

C. Effects of the Number of Grid Points, the Compression Rate, and the Uncertainty in the Prior Information

In this section, we provide the results of the simulations conducted at a fixed SNR level for different choices of the number of grid points (L), the number of digital channels (m), and the uncertainty in the prior information (σ_s). We also report the results of RGD in addition to KGKA, A-MMD, and A-MMD-UD as a reference design. The reported results are obtained by performing 10,000 Monte Carlo iterations each with independent noise and DOA realizations. The plots with error bars are used to present both the upper and lower bounds of the obtained RMSEs.

1) Effect of the Number of Grid Points: To investigate the effect of L , we again use a ULA with $M = 50$ antenna elements spaced $\lambda/2$ apart and $m = 10$ digital channels. The DOA distribution given in Table IV is assumed and the SNR is fixed at 40 dB. The results are given in Fig. 4. Although the mutual coherence of the effective dictionary increases with increasing L , the performances of A-MMD, A-MMD-UD and KGKA improve. Hence, lower mutual coherence does not necessarily imply higher performance in CS based DOA estimation. This result also demonstrates that the proposed algorithm can work with larger L values and provide higher resolution. Since A-MMD-UD forms an adaptive, non-uniform grid, for $L = 75$,

TABLE IX
DOA DISTRIBUTION OF THE SOURCES TO INVESTIGATE THE m DEPENDENCY

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(-0.5455, 0.1^2)$	$\theta = 100^\circ$ ($\omega = -0.5455$)
2	$\mathcal{U}(-0.9\pi, 0.9\pi)$	$\theta = 90^\circ$ ($\omega = 0$)

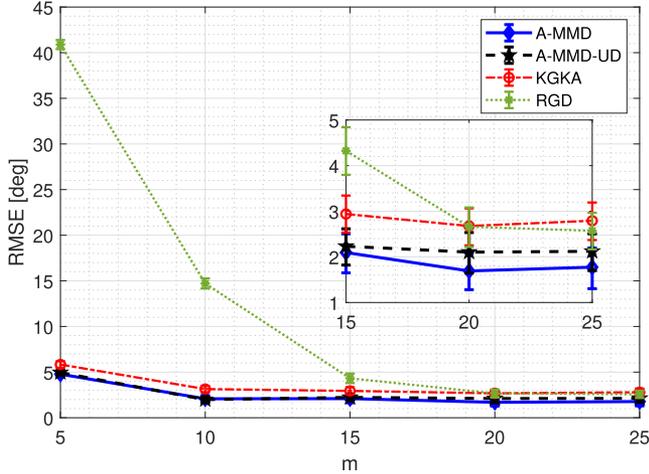


Fig. 5. RMSE vs the number of digital channels. The inset is used to show the details better.

A-MMD-UD is the best performing technique as expected. However, as L increases, A-MMD becomes the best performing technique. Hence, when the computational requirements of a system cannot support the use of large number of grid points, A-MMD-UD offers computational advantages over A-MMD.

2) *Effect of the Compression Rate:* The performance of a technique with varying M/m represents its compression ability. As M/m increases, both the hardware and software complexity of the system are reduced. For this case, a ULA with $M = 100$ antenna elements spaced $\lambda/2$ apart and $m \in \{5, 10, 15, 20, 25\}$ digital channels are investigated to check the performances of the techniques at up to 20 times the compression rate. We fix SNR to be 25 dB. The noise power is increased with increasing m for this scenario since $\text{SNR} \equiv 10 \log_{10}(mP_s/\sigma^2)$. $L = 3 \times M = 300$ uniform grid points are used, and the DOA distribution scheme given in Table IX is assumed. The standard deviation is again found by $5 \times 2/M$. The target DOAs of Source 1 are chosen randomly over the interval $[93^\circ, 108^\circ]$. The results are given in Fig. 5. As expected, increasing m improves the performances of all the techniques. However, when the compression rate is large, RGD cannot perform reliable estimations. The other techniques can perform reliable estimations up to a compression rate of at least 20. Among them, A-MMD has the best performance for all m values while the performance of A-MMD-UD is a close second.

3) *Effect of the Prior Information Uncertainty:* In all the simulations discussed before, σ_s is chosen as $5 \times 2/M$. In certain cases, σ_s can be set to a different value depending on the rate of change of the dynamic sources or the reliability of the tracker output. To see the effect of σ_s on the estimation performance, we use a ULA with $M = 50$ antenna elements spaced $\lambda/2$ apart and $m = 10$ digital channels. We fix SNR = 25 dB,

TABLE X
DOA DISTRIBUTION OF THE SOURCES TO INVESTIGATE THE σ_s DEPENDENCY

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(-0.5455, \sigma_s^2)$	$\theta = 100^\circ$ ($\omega = -0.5455$)
2	$\mathcal{N}(1.5708, \sigma_s^2)$	$\theta = 60^\circ$ ($\omega = 1.5708$)
3	$\mathcal{U}(-0.9\pi, 0.9\pi)$	$\theta = 90^\circ$ ($\omega = 0$)

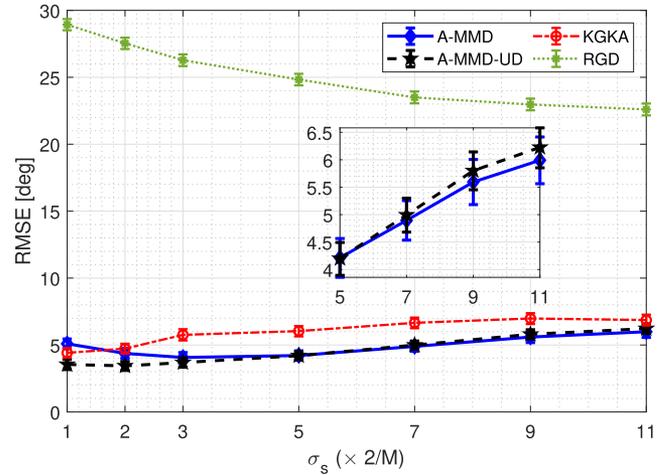


Fig. 6. RMSE vs the uncertainty in the prior information. The inset is used to show the details better.

and use $L = 3 \times M = 150$ grid points. The DOA distribution given in Table X is assumed on the source scene. In Table X, $\sigma_s \in \{1, 2, 3, 5, 7, 9, 11\} \times 2/M$. When $\sigma_s = 2/M$, the DOAs of Source 1 are chosen randomly over the interval $[97^\circ, 103^\circ]$, and the DOAs of Source 2 are chosen randomly over the interval $[56^\circ, 63^\circ]$. When $\sigma_s = 11 \times 2/M$, the DOAs of Source 1 are chosen randomly over the interval $[69^\circ, 133^\circ]$, and the DOAs of Source 2 are chosen randomly over the interval $[11^\circ, 90^\circ]$. Hence, a wide variation on the prior distribution of DOAs is considered. The results are presented in Fig. 6. When $\sigma_s < 5 \times 2/M$, A-MMD-UD outperforms A-MMD and KGKA. For other σ_s values, A-MMD performs the best. As σ_s gets larger, the performances of the adaptive techniques tend to degrade as anticipated, since the uncertainty in the prior information increases. These results further imply that the adaptive dictionary design strategies may cause performance degradation when there is high uncertainty in the prior information.

D. Performance Comparison for Mutual Coupling Effects

We consider a simulated electromagnetic model of a ULA at 10 GHz with $M = 64$ antenna elements and an element spacing of $d = 0.5\lambda = 0.015m$. Electromagnetic coupling effects on the array are analyzed via a Finite Element Method (FEM) solver [78], and its S-parameters are exported as a matrix \mathcal{S} which is shown in Fig. 7. The source scenario given in Table X is used with $\sigma_s = 5 \times 2/M = 0.1563$. The DOAs of Source 1 are chosen randomly over the interval $[86^\circ, 113^\circ]$, and the DOAs of Source 2 are chosen randomly over the interval $[46^\circ, 74^\circ]$. We set $L = 3 \times M = 192$, and $m = 12$. The same scenario is simulated with and without including the mutual coupling

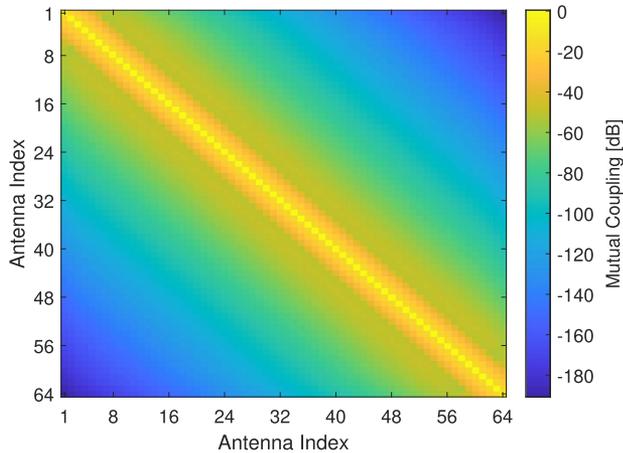


Fig. 7. Mutual coupling matrix: The mutual coupling values are given in dB scale.

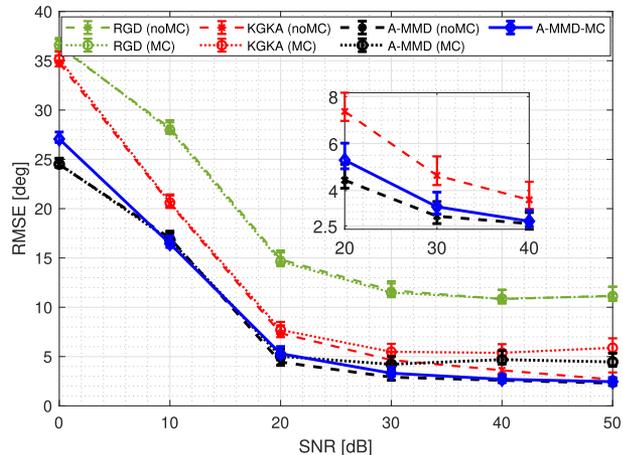


Fig. 8. RMSE vs SNR under mutual coupling effects. The inset is used to show the details better.

effects. The achieved results are shown Fig. 8. RGD (noMC), KGKA (noMC), A-MMD (noMC) abbreviations are used to label the performances of those techniques when the mutual coupling effects are not included in the simulation. For RGD (MC), KGKA (MC), A-MMD (MC); the mutual coupling effects are included in the simulation. Fig. 8 shows that the existence of mutual coupling decreases the performance of KGKA and A-MMD. Note that RGD has already low performance; therefore, the mutual coupling effects do not cause a noticeable degradation in its performance. By modifying A-MMD, A-MMD-MC algorithm is designed as described in Section IV. Fig. 8 shows that the mutual coupling effects are successfully mitigated by A-MMD-MC since A-MMD (noMC) and A-MMD-MC perform very close to each other. Moreover, A-MMD-MC performs significantly better than A-MMD (MC).

E. Performance Comparison Using a Random Linear Array in Multipath Environments

Here we investigate the performance of A-MMD on a non-uniform linear array. An RLA with $M = 100$ antenna elements

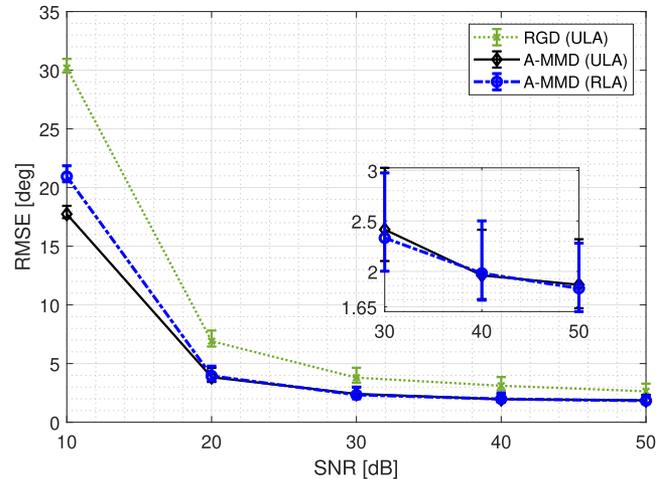


Fig. 9. RMSE vs SNR using an RLA in a multipath environment. The inset is used to show the details better.

and $m = 20$ digital channels is used and the number of grid points is selected as $L = 3 \times M = 300$. The multipath effects are also included in the simulation. The element spacing between two adjacent antenna elements is randomly sampled from the uniform distribution $\mathcal{U}(0.4\lambda, 0.6\lambda)$. The source scenario given in Table X is used with $\sigma_s = 5 \times 2/M = 0.1$ except that Source 2 is now the multipath of Source 1 with the received signal power $P_s/2$. The DOAs of Source 1 are chosen randomly over the interval $[93^\circ, 107^\circ]$, and the DOAs of Source 2 are chosen randomly over the interval $[51^\circ, 68^\circ]$. Fig. 9 shows that A-MMD performs very close for both array configurations with $\text{RMSE} < 2^\circ$ at the high-SNR regime. It shows that A-MMD can provide reliable results with array structures other than ULA in a multipath environment.

F. Phase Transition Results

So far, we have considered A-MMD with the reweighted ℓ_1 -norm optimization (21), i.e., by exploiting prior information both in reconstruction and data acquisition stages. Since they both affect the estimation performance, it is important to evaluate the performances of adaptive data acquisition (A-MMD) and adaptive reconstruction (reweighted ℓ_1 -norm minimization) separately. To do that, A-MMD can be used with the standard ℓ_1 optimization (6) and RGD can be used with the reweighted ℓ_1 optimization (21). In this experiment, we investigate the RMSE performances of A-MMD and RGD for variable number of sources and channels using phase transition phenomenon. For the simulation, a ULA with $M = 64$ antenna elements spaced $\lambda/2$ apart is used, and $\text{SNR} = 40$ [dB], $L = 3 \times M = 192$, $\sigma_s = 5 \times 2/M$ are fixed. We obtain the results for different m and K values, where m is changed from 8 to 32 with a step-size of 4, and K is changed from 1 to 6 with a step-size of 1. All the sources are assumed to be tracked sources, whose DOAs are drawn from the distributions given in Table XI. The DOAs of Sources 1–6 are chosen randomly over the intervals $[87^\circ, 111^\circ]$, $[68^\circ, 91^\circ]$, $[108^\circ, 135^\circ]$, $[47^\circ, 74^\circ]$, $[125^\circ, 163^\circ]$, $[19^\circ, 55^\circ]$, respectively. When $K = 1$, there is only Source 1 in

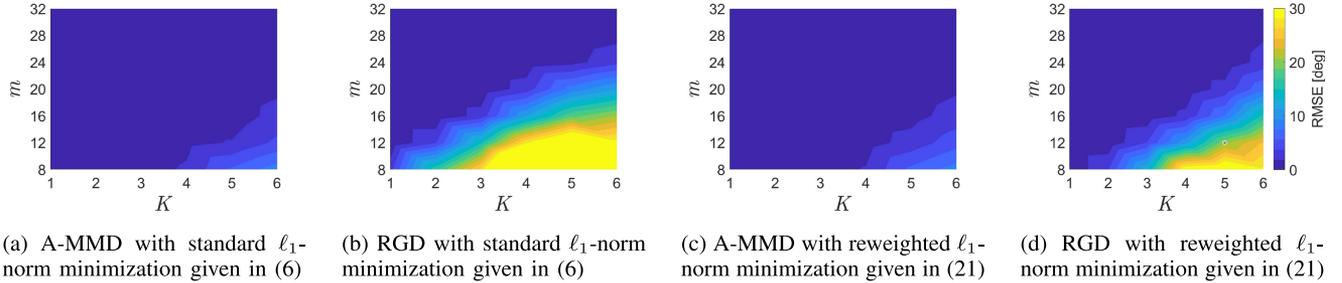


Fig. 10. RMSE [deg] results achieved by A-MMD and RGD using standard and reweighted ℓ_1 -norm optimization, depending on m and K . The same color axis is used for all results (see the colorbar).

TABLE XI
DOA DISTRIBUTION OF THE SOURCES TO COMPARE RGD AND A-MMD USING PHASE TRANSITION PHENOMENON

Source	DOA Distribution	Mean Position
1	$\mathcal{N}(-0.5455, 0.1563^2)$	$\theta = 100^\circ$ ($\omega = -0.5455$)
2	$\mathcal{N}(0.5455, 0.1563^2)$	$\theta = 80^\circ$ ($\omega = 0.5455$)
3	$\mathcal{N}(-1.5708, 0.1563^2)$	$\theta = 120^\circ$ ($\omega = -1.5708$)
4	$\mathcal{N}(1.5708, 0.1563^2)$	$\theta = 60^\circ$ ($\omega = 1.5708$)
5	$\mathcal{N}(-2.4066, 0.1563^2)$	$\theta = 140^\circ$ ($\omega = -2.4066$)
6	$\mathcal{N}(2.4066, 0.1563^2)$	$\theta = 40^\circ$ ($\omega = 2.4066$)

the environment. When $K = 2$, Source 2 is also included in the source environment and so on. The obtained phase transition results are demonstrated in Fig. 10. By comparing Fig. 10(b) and Fig. 10(d), we observe that the reweighted ℓ_1 minimization significantly improves the performance of RGD. On the other hand, by comparing Fig. 10(a) and Fig. 10(c), we observe that A-MMD provides high performance regardless of the reconstruction technique. In all cases, A-MMD performs significantly better than RGD, especially when m is low and K is high.

G. Gridless Reconstruction

As explained in Section II-B, A-MMD can also be used with ANM. To demonstrate that, the same array configuration described in Case 1 is used with $L = 500$. We assume the source scenario given in Table X, where $\sigma_s = 5 \times 2/M = 0.2$. The DOAs of Source 1 and Source 2 are chosen randomly over the intervals $[86^\circ, 114^\circ]$ and $[42^\circ, 74^\circ]$, respectively. The results are given in Fig. 11. A-MMD (ANM), A-MMD-UD (ANM), and RGD (ANM) denote that the reconstruction is performed using ANM for the corresponding measurement matrix design. A-MMD (ℓ_1), A-MMD-UD (ℓ_1), and RGD (ℓ_1) are obtained by using ℓ_1 -norm minimization with the corresponding measurement matrix design, i.e., (21) is used with A-MMD and A-MMD-UD, and (6) is used with RGD. A-MMD and A-MMD-UD perform significantly better than RGD, which shows that A-MMD based algorithms can also be used in conjunction with a gridless reconstruction algorithm. Furthermore, higher resolution can be obtained by ANM since its resolution is not limited by the grid resolution unlike ℓ_1 -norm minimization. Moreover, A-MMD (ANM) and A-MMD-UD (ANM) perform differently, which shows that the grid selection for measurement matrix design changes the data acquisition strategy even though the reconstruction is gridless. As explained in Section II, the

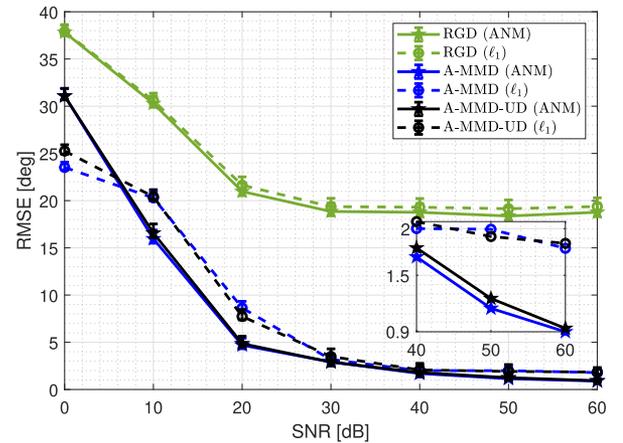


Fig. 11. RMSE vs SNR when the DOAs are reconstructed using ANM and ℓ_1 -norm minimization. The inset is used to show the details better.

main disadvantage of ANM is its higher computational time. For this scenario, our ANM implementation takes approximately 12 times longer compared to our implementation of ℓ_1 -norm minimization. If $L = 3 \times M = 150$ was used, the difference would increase up to 40 times. As discussed in Section II-B, there are more efficient algorithms for ANM. Adopting a more computationally efficient methodology as in [50] can drastically decrease this time difference.

Numerical results demonstrate that A-MMD and its variations, A-MMD-UD and A-MMD-MC, outperform their alternatives under various scenarios. Comparison between A-MMD and A-MMD-UD indicates that their performance ordering depends on the scenario. However, when L has to be kept small, A-MMD-UD is shown to be advantageous due to its adaptive grid selection strategy. The results further imply that the adaptive dictionary design is not necessary when the computational resources allow selection of large L values. Furthermore, when the prior information uncertainty is high, A-MMD-UD should not be the preferred technique. It is also empirically demonstrated that A-MMD based algorithms can compensate for the hardware implementation issues such as mutual coupling, and can also work with non-uniform array structures in multipath environments. Phase transition results also validate the superior performance of A-MMD for a variety of compression rates and number of sources. When we present the phase transition results, we also investigate the performance of A-MMD independently

of the reweighted ℓ_1 optimization, where we show that A-MMD provides high performance also in that case. Finally, we show that gridless algorithms like ANM can be used in conjunction with A-MMD and its variations.

VI. CONCLUSION

Measurement matrix design is among the most important aspects of CS-based sensor processing applications. In this paper, a novel adaptive measurement matrix design methodology that provides accurate results on the sensor data was proposed. A computationally feasible, closed-form expression for the measurement matrix was derived enabling online updates of the measurement matrix at a high compression rate for each snapshot of the sensor data. Thus, the proposed methodology allows for efficient hardware and software implementations. Over an extensive set of array signal processing scenarios, the superior performance of the proposed measurement matrix design methodology over the alternative online design techniques was demonstrated. The proposed technique successfully mitigates the mutual coupling effects between array elements and provides accurate results even in multipath environments with non-uniform array configurations. We further demonstrated the applicability of our proposed algorithm along with a gridless reconstruction method, namely, atomic norm minimization. While the atomic norm minimization based algorithm improves the estimation performance, it has higher time complexity compared to ℓ_1 -norm minimization.

The proposed technique allows for joint optimization over the measurement matrix and the signal dictionary. In the simulation results, it was demonstrated that adaptive dictionary design may improve performance. As a future work, alternative dictionary optimization algorithms can be investigated. Moreover, the parameter selection methodology followed in this study can be improved by defining the parameter as a function of not only the number of digital channels, the number of sensors, and the number sources; but also including other aspects such as the noise power and array configuration. Decreasing the computational complexity of atomic norm minimization is another research topic. Although the main focus of this study was DOA estimation, the proposed methodology can be applied to other areas of sensor processing including detection of sparse signals in video and audio streams.

ACKNOWLEDGMENT

The authors would like to thank Tutku Bakan for computing the mutual coupling matrix of a practical linear phased array antenna via full-wave electromagnetic simulations.

REFERENCES

- [1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [4] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 19–60, 2010.
- [7] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [8] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [9] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [10] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [11] B. D. V. Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [12] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [14] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Proc. IEEE/SP 15th Workshop Stat. Signal Process.*, 2009, pp. 626–629.
- [15] M. Ibrahim, F. Roemer, and G. D. Galdo, "On the design of the measurement matrix for compressed sensing based DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 3631–3635.
- [16] M. Ibrahim, F. Roemer, and G. D. Galdo, "An adaptively focusing measurement design for compressed sensing based DOA estimation," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 859–863.
- [17] B. Ozer, A. Lavrenko, S. Gezici, F. Römer, G. Del Galdo, and O. Arikan, "Adaptive measurement matrix design for compressed DOA estimation with sensor arrays," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2015, pp. 1769–1773.
- [18] M. Ibrahim et al., "Design and analysis of compressive antenna arrays for direction of arrival estimation," *Signal Process.*, vol. 138, pp. 35–47, 2017.
- [19] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 3181–3185.
- [20] M. Guo, Y. D. Zhang, and T. Chen, "DOA estimation using compressed sparse array," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4133–4146, Aug. 2018.
- [21] B. Kılıç, A. Güngör, M. Kalfa, and O. Arikan, "Adaptive measurement matrix design in compressed sensing based direction of arrival estimation," in *Proc. Eur. Signal Process. Conf.*, 2021, pp. 1881–1885.
- [22] B. Kılıç, "Adaptive techniques in compressed sensing based direction of arrival estimation," M.S. thesis, Bilkent Univ., Ankara, Turkey, 2021.
- [23] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf, "Constrained adaptive sensing," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5437–5449, Oct. 2016.
- [24] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.
- [25] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
- [26] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing*, vol. 7. Amsterdam, The Netherlands: Elsevier, 2018, pp. 509–581.
- [27] M. G. Christensen, J. Østergaard, and S. H. Jensen, "On compressed sensing and its application to speech and audio signals," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2009, pp. 356–360.
- [28] L. Carin, D. Liu, and B. Guo, "Coherence, compressive sensing, and random sensor arrays," *IEEE Antennas Propag. Mag.*, vol. 53, no. 4, pp. 28–39, Aug. 2011.
- [29] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.
- [30] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

- [31] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [32] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [33] G. Li, Z. Zhu, X. Wu, and B. Hou, "On joint optimization of sensing matrix and sparsifying dictionary for robust compressed sensing systems," *Digit. Signal Process.*, vol. 73, pp. 62–71, 2018.
- [34] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," Dec. 2019. [Online]. Available: <https://friedlander.io/spgl1>
- [35] E. Van Den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2009.
- [36] S. Boyd, N. Parikh, and E. Chu, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Delft, The Netherlands: Now Publishers Inc, 2011.
- [37] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2010.
- [38] M. Wagner, Y. Park, and P. Gerstoft, "Gridless DOA estimation and root-music for non-uniform linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 2144–2157, 2021.
- [39] Y. Park and P. Gerstoft, "Alternating projections gridless covariance-based estimation for DOA," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 4385–4389.
- [40] M. Kalfa and H. E. Güven, "Fast 2-D direction of arrival estimation using two-stage gridless compressive sensing," in *Proc. IEEE Int. Conf. Radar*, 2018, pp. 1–5.
- [41] S. Semper, F. Roemer, T. Hotz, and G. D. Galdo, "Grid-free direction-of-arrival estimation with compressed sensing and arbitrary antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 3251–3255.
- [42] S. Semper and F. Römer, "ADMM for ND line spectral estimation using grid-free compressive sensing from multiple measurements with applications to DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 4130–4134.
- [43] Z. Yang, J. Tang, Y. C. Eldar, and L. Xie, "On the sample complexity of multichannel frequency estimation via convex optimization," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2302–2315, Apr. 2018.
- [44] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries—just discretize," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2013, pp. 1043–1047.
- [45] V. Duval and G. Peyré, "Sparse spikes super-resolution on thin grids ii: The continuous basis pursuit," *Inverse Problems*, vol. 33, no. 9, 2017, Art. no. 095008.
- [46] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Math. Program.*, vol. 95, no. 2, pp. 189–217, 2003.
- [47] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014. [Online]. Available: <http://cvxr.com/cvx>
- [48] Y. Hua and T. Sarkar, "Generalized pencil-of-function method for extracting poles of an EM system from its transient response," *IEEE Trans. Antennas Propag.*, vol. 37, no. 2, pp. 229–234, Feb. 1989.
- [49] K. Krishnan and T. Terlaky, "Interior point and semidefinite approaches in combinatorial optimization," in *Graph Theory and Combinatorial Optimization*. Berlin, Germany: Springer, 2005, pp. 101–157.
- [50] Y. Wang and Z. Tian, "IVDST: A fast algorithm for atomic norm minimization in line spectral estimation," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1715–1719, Nov. 2018.
- [51] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [52] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014.
- [53] A. J. Weiss and B. Friedlander, "Preprocessing for direction finding with minimal variance degradation," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1478–1485, Jun. 1994.
- [54] J. Sheinvald and M. Wax, "Direction finding with fewer receivers via time-varying preprocessing," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 2–9, Jan. 1999.
- [55] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. Eur. Signal Process. Conf.*, 2010, pp. 427–431.
- [56] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
- [57] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.
- [58] N. Cleju, "Optimized projections for compressed sensing via rank-constrained nearest correlation matrix," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 3, pp. 495–507, 2014.
- [59] M. A. Sustik, J. A. Tropp, I. S. Dhillon, and R. W. Heath Jr, "On the existence of equiangular tight frames," *Linear Algebra Appl.*, vol. 426, no. 2/3, pp. 619–635, 2007.
- [60] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.
- [61] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–8, 2010.
- [62] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "A gradient-based alternating minimization approach for optimization of the measurement matrix in compressive sensing," *Signal Process.*, vol. 92, no. 4, pp. 999–1009, 2012.
- [63] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 397–399, May 1974.
- [64] B. Li, L. Zhang, T. Kirubarajan, and S. Rajan, "Projection matrix design using prior information in compressive sensing," *Signal Process.*, vol. 135, pp. 36–47, 2017.
- [65] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [66] T. Huang, Y. Liu, H. Meng, and X. Wang, "Adaptive compressed sensing via minimizing Cramer-Rao bound," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 270–274, Mar. 2014.
- [67] T. M. Cover, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.
- [68] R. Obermeier and J. A. Martinez-Lorenzo, "Sensing matrix design via capacity maximization for block compressive sensing applications," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 27–36, Mar. 2019.
- [69] X. Duan, J. Li, Q. Wang, and X. Zhang, "Low rank approximation of the symmetric positive semidefinite matrix," *J. Comput. Appl. Math.*, vol. 260, pp. 236–243, 2014.
- [70] A. Dax, "Low-rank positive approximants of symmetric matrices," *Adv. Linear Algebra Matrix Theory*, vol. 4, no. 03, 2014, Art. no. 172.
- [71] T.-H. Chio and D. H. Schaubert, "Parameter study and design of wide-band widescan dual-polarized tapered slot antenna arrays," *IEEE Trans. Antennas Propag.*, vol. 48, no. 6, pp. 879–886, Jun. 2000.
- [72] W. H. Syed, D. Cavallo, H. T. Shivamurthy, and A. Neto, "Wideband, wide-scan planar array of connected slots loaded with artificial dielectric superstrates," *IEEE Trans. Antennas Propag.*, vol. 64, no. 2, pp. 543–553, Feb. 2016.
- [73] B. Kılıç, M. Kalfa, and O. Arıkan, "Prior based grid selection algorithm for compressed sensing based direction of arrival estimation methods," in *Proc. IEEE Int. Symp. Phased Array Syst. Technol.*, 2019, pp. 1–5.
- [74] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb. 2016.
- [75] J. A. Tropp, "A mathematical introduction to compressive sensing [book review]," *Bull. Amer. Math. Soc.*, vol. 54, no. 1, pp. 151–165, 2017.
- [76] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL, USA: CRC Press, 1994.
- [77] M. A. Richards, *Fundamentals of Radar Signal Processing*. New York, NY, USA: McGraw-Hill Education, 2014.
- [78] ANSYS "ANSYS Electronics Desktop, Release 19.0," ANSYS, Inc., Canonsburg, PA, USA. [Online]. Available: <https://ansys.com>