DOA Estimation via Shift-Invariant Matrix Completion

Vaibhav Garg[a], Pere Giménez-Febrer[b], Alba Pagès-Zamora[b], Ignacio Santamaria[a]

[a] *Dept. of Communications Engineering, Universidad de Cantabria, Spain*
[b] *SPCOM Group, Universitat Politècnica de Catalunya-Barcelona Tech, Spain*

Abstract

In this paper, we propose an energy-efficient approach to estimate directions of arrival (DOAs) of multiple uncorrelated sources received by a uniform linear array (ULA) with many antennas. To reduce energy consumption and hardware costs, the receiving array uses antenna switching techniques so that at every time instant or snapshot only the radio-frequency (RF) signals received by a randomly selected subset of antennas is downconverted to baseband and sampled. Low-rank matrix completion (MC) techniques are then used to reconstruct the missing entries of the signal data matrix to keep the angular resolution of the original large-scale array. The proposed MC algorithm exploits not only the low-rank structure of the signal subspace, but also the shift-invariance property of ULAs, which results in a better estimation of the signal subspace. Further, the effect of MC on DOA estimation is discussed under the perturbation theory framework. The simulation results suggest that the proposed method provides accurate DOA estimates even in the small-sample regime with a significant reduction in the number of RF chains required for a given spatial resolution.

Keywords: Direction of arrival (DOA), Uniform Linear Array, Massive MIMO, Matrix Completion, Shift Invariance

1. Introduction

The need of large bandwidths in 5G networks motivates to operate in mm-Wave bands, which require large-scale antenna arrays to compensate for the path loss [1, 2]. Indeed, research in wireless communication systems has shifted towards the use of large antenna arrays as in massive multiple-input multiple-output (MIMO) systems [3]. This poses new challenges not only to antenna

⋆This work was supported by the Ministerio de Ciencia e Innovación (MICINN) of Spain, and AEI/FEDER funds of the E.U., under grants TEC2016-75067-C4-4-R /2-R (CARMEN), PID2019-104958RB-C43/C41 (ADELE) and BES-2017-080542.

Email addresses: vaibhav.garg@unican.es (Vaibhav Garg), p.gimenez@upc.edu (Pere Giménez-Febrer), alba.pages@upc.edu (Alba Pagès-Zamora), i.santamaria@unican.es (Ignacio Santamaria)

calibration and complexity issues associated with channel state information acquisition and precoding [4], but also to energy consumption. It is acknowledged that power consumption requirements in 5G networks increase by about 3 times over 4G, and that the signal processing in massive MIMO systems can represent up to 40% of the total power consumption for below-6 GHz bands, and even larger in mm-Wave bands [5].

A classical problem when processing multiple signals received by a uniform linear array (ULA) is that of estimating their directions-of-arrival (DOAs). DOA estimation has a long and rich history in array processing [6], and numerous high-resolution direction finding algorithms have been proposed over the last decades. As representative examples it is worth mentioning subspace-based methods such as the multiple signal classification (MUSIC) algorithm [7] and the estimation of signal parameters via rotational invariance technique (ESPRIT) [8], which provide high angular resolution. However, using MUSIC or ESPRIT with a large-scale fully-digital receive antenna array can be challenging due to their computational complexity and high energy consumption requirements. A possible solution is to reduce the number of radio frequency (RF) transceiver chains by performing antenna selection at the receiving array (cf. Fig. 1). At every time instant a random switch selects a subset of antennas whose RF signals are downconverted and further processed. Since the number of targets or sources is typically much smaller than the number of antennas, it is feasible to reconstruct (or at least to approximate) the low-rank signal data matrix using matrix completion (MC) algorithms as if it had been received by the full array, as long as we sample a sufficiently large fraction of the sensors [9].

Low-rank MC methods for DOA estimation are used in [10] for scenarios in which the number of sources exceeds the number of sensors, and in [11] in the presence of diffuse noise. An iterative reweighted nuclear norm minimization method is used in [12] for DOA estimation with nested arrays. When a sparse coprime array is used, array interpolation techniques can be applied to improve the DOA estimation performace. In [13] the authors consider this scenario and apply MC techniques to reconstruct the Toeplitz virtual array covariance matrix. In [14], a DOA estimation algorithm based on virtual array interpolation and MC techniques is developed for coherent sources in coprime arrays. A Toeplitz reconstruction algorithm based on nuclear norm minimization is proposed in [15] for uniform and sparse linear arrays. A different approach is proposed in [16], where MC algorithms are used to reconstruct the entries of the sample covariance matrix (SCM) along its diagonal, which are deliberately set to zero. In [17] MC is used for order estimation in the presence of noise with a diagonal spatial covariance matrix. Whereas most of existing MC methods in array signal processing target the reconstruction of the signal covariance matrix, it is the data matrix itself that needs to be reconstructed when only a subset of sensors is sampled.

In this paper, an energy-efficient approach to DOA estimation is proposed based on the recovery of the data matrix by means of a MC method. We consider that only a randomly chosen subset of sensors are sampled at each time instant. By reducing the number of RF chains of the receiver, the overall

hardware cost and energy consumption are reduced as well. In our approach, the matrix completion problem is tailored to enforce the shift-invariance property of ULAs by including an additional regularization term in the MC cost function. Then, the Optimal Subspace Estimation (OSE) technique proposed by Vaccaro and Ding in [18] is used to estimate the signal subspace and, finally, the DOAs are estimated using ESPRIT as a high-resolution technique. The simulations show that the number of RF chains can be largely reduced without significant performance loss in DOA estimation accuracy.

The rest of the paper is organized as follows. Section 2 presents the signal model assuming an array architecture with random antenna switching, and formulates the problem. The proposed Shift-Invariant Matrix Completion (SIMC) method is described in Section 3. A direct application of the Davis-Kahan theorem [19] allows us to analyze in Section 4 the chordal distance between the true signal subspace and the signal subspace of the sparse and reconstructed matrices. The simulation results are discussed in Section 5, and concluding remarks are provided in Section 6.

Notation. Bold lowercase letters denote vectors and bold uppercase matrices; $\mathbf{B}(i,j)$ is the entry in the i-th row and j-th column of matrix $\mathbf{B}$. Superscripts $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ denote transpose, complex conjugate and Hermitian, respectively. $|z|$ denotes the modulus of a complex number z, and $||\mathbf{x}||_2$ is the l_2-norm of vector $\mathbf{x}$. The trace, nuclear, spectral, Frobenius and infinity norms of a matrix are denoted, respectively, as $\mathrm{tr}(\cdot)$, $||\cdot||_*$, $||\cdot||_2$, $||\cdot||_F$ and $||\cdot||_\infty$. The k-th largest singular value is denoted as $\lambda_k(\cdot)$. Furthermore, $\mathbf{x} \sim \mathcal{CN}_n(\mathbf{0}, \mathbf{\Sigma})$ denotes a proper Gaussian random vector in $\mathbb{C}^n$ with zero mean and covariance $\mathbf{\Sigma}$.

2. Observed Data Matrix and Problem Statement

Let us consider K narrowband signals impinging on a large half-wavelength ULA with M antennas. For a fully digital receiver with M RF-branches, the received signal at time instant or snapshot n is

$$\mathbf{z}[n] = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_K)]\, \mathbf{s}[n] + \mathbf{e}[n] = \mathbf{A}\, \mathbf{s}[n] + \mathbf{e}[n], \tag{1}$$

where $\mathbf{e}[n]$ is the noise vector, $\mathbf{s}[n] = [s_1[n], \ldots, s_K[n]]^T$ is the signal vector with complex gains $s_k[n]$ and $\mathbf{a}(\theta_k) = \begin{bmatrix} 1 & e^{-j\theta_k} & e^{-j\theta_k(M-1)} \end{bmatrix}^T$ is the $M \times 1$ complex array response to the k-th source with electrical angle θ_k, which is unknown; and $\mathbf{A} = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_K)]$ is the steering matrix. In the case of narrowband sources, free space propagation, and a ULA with inter-element spacing d, the spatial frequency or electrical angle is

$$\theta_k = \frac{2\pi}{\lambda} d \sin(\phi_k),$$

where λ is the wavelength and ϕ_k is the direction-of-arrival (DOA). We will refer to θ_k as the DOA of source k for simplicity. Note that for a half-wavelength

ULA $\theta_k = \pi \sin(\phi_k)$, and the spatial frequency varies between $-\pi$ and π when ϕ_k varies between $-\pi/2$ and $\pi/2$, with 0° being the broadside direction.

The signal and noise vectors are assumed uncorrelated and modeled as $\mathbf{s}[n] \sim \mathcal{CN}_K(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_K^2)$, and $\mathbf{e}[n] \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I})$, respectively. Using the signal model in (1), the full $M \times M$ covariance matrix is

$$\mathbf{R} = E\left[\mathbf{z}[n]\mathbf{z}^H[n]\right] = \mathbf{R}_s + \sigma^2\mathbf{I} \tag{2}$$

where $\mathbf{R}_s = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^H$.

After collecting N snapshots, the full data matrix $\mathbf{Z} = \left[\mathbf{z}[1] \ldots \mathbf{z}[N]\right]$ can be written as

$$\mathbf{Z} = \mathbf{X} + \mathbf{E} \tag{3}$$

where $\mathbf{E} = \left[\mathbf{e}[1] \ldots \mathbf{e}[N]\right]$, and $\mathbf{X} = \mathbf{AS}$ is the noiseless signal component with $\mathbf{S} = \left[\mathbf{s}[1] \ldots \mathbf{s}[N]\right]$. A simplified receiver architecture composed of an $M \times L$ RF switching network is considered such that, at each snapshot, it randomly selects L out of the M antennas to be downconverted and sampled at baseband (see Fig. 1). Multi-switch antenna selection techniques for massive MIMO have been studied and experimentally validated in [20]. After downconversion and sampling, the $L \times N$ samples are arranged in a $\mathbf{Z}_d \in \mathbb{C}^{M \times N}$ matrix so that missing entries are replaced with zeros. The sampling process can be compactly expressed as

$$\mathbf{Z}_d = P_\Omega(\mathbf{Z}), \tag{4}$$

where $\Omega \subseteq \{1, \ldots, M\} \times \{1, \ldots, N\}$ is the set of observed (antenna, time) indexes, and P_Ω is a projection operator that sets to zero the missing entries and leaves the observed ones unchanged.

The problem addressed in this paper is, given the observed data matrix $\mathbf{Z}_d$ in (4), to estimate the rank-K noiseless signal matrix, denoted by $\hat{\mathbf{X}}$, and use it to further estimate the DOAs $\{\theta_k\}_{k=1}^K$. We assume that the number of sources K is known and satisfies $K \ll L < M$.

3. Shift-Invariant Matrix Completion (SIMC)

3.1. Matrix Completion

The problem of estimating the low-rank signal matrix $\mathbf{X}$ from $\mathbf{Z}_d \in \mathbb{C}^{M \times N}$ can be solved using MC techniques. According to [9], we can recover $\mathbf{X}$ by solving

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{C}^{M \times N}} \quad & ||\mathbf{X}||_* \\ \text{subject to} \quad & ||P_\Omega(\mathbf{X} - \mathbf{Z}_d)||_F \leq \eta \end{aligned} \tag{5}$$

where $||\mathbf{X}||_*$ denotes the nuclear norm of $\mathbf{X}$, $\Omega \subseteq \{1, \ldots, M\} \times \{1, \ldots, N\}$ the set of observed entries of $\mathbf{Z}_d$, and $\eta > 0$ is a tolerance parameter that limits the fitting error.

The main assumption for a successful recovery in low-rank MC is that of incoherence, which means that each singular vector of the unknown matrix

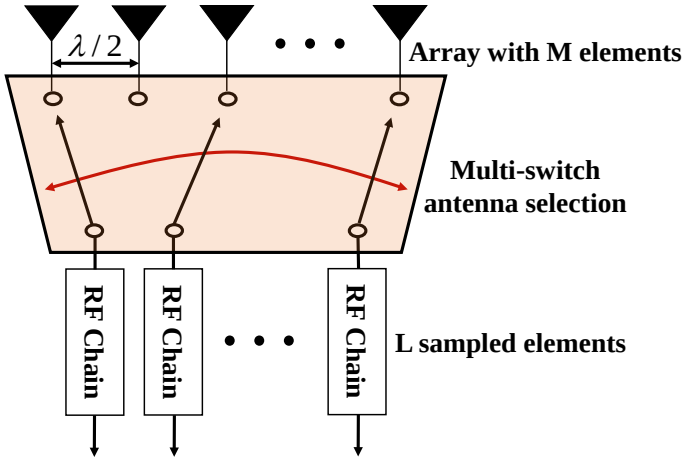


Figure 1: Simplified large-scale multi-switch array architecture where L out of M sensors are randomly selected and sampled at each time instant.

must be evenly spread across its coordinates instead of having a few entries with large value i.e., singular vectors are not too sparse. Intuitively, this implies that there is no underlying matrix structure and that all entries have similar importance. Formally, the coherence of the column space of a rank-K matrix $\mathbf{Y} \in \mathbb{C}^{M \times N}$ is defined as

$$\tau(\mathbf{Y}) = \frac{M}{K} \max_{1 \leq i \leq M} ||\mathbf{P}_{\mathbf{Y}} \mathbf{e}_i||_2$$

where $\mathbf{P}_{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H$ is the orthogonal projection matrix onto the column space of $\mathbf{Y}$, and $\mathbf{e}_i$ is the i-th vector of the Euclidean basis.

As shown in [21], in array processing $\tau(\mathbf{X})$ and $\tau(\mathbf{X}^H)$ are small enough that the complete matrix $\mathbf{X}$ can be recovered via (5). Indeed, in the noiseless case an exact recovery is possible with high probability provided that we observe $|\Omega| \geq DKN^{\frac{6}{5}} \log N$ for a constant D assuming a random uniform sampling distribution and $N > M$ [9]. In our problem we have $|\Omega| = NL$, therefore $L \geq DKN^{\frac{1}{5}} \log N$ antenna elements need to be sampled for successful recovery. In the noisy case, $\mathbf{X}$ is recovered with an error proportional to η as long as $||P_\Omega(\boldsymbol{E})||_F \leq \eta$ [9].

While standard MC assumes uniform random sampling, this scheme does not exactly match the multi-switch array architecture in Fig. 1. In the proposed architecture, exactly L sensors, chosen at random, are sampled per snapshot, which is termed as uniform spatial sampling in [21] and does not correspond to uniform random sampling across $\mathbf{Z}$. Nevertheless, as it proved in [21], the uniform spatial sampling scheme satisfies the coherence conditions for matrix recovery and hence it can be used in array processing problems.

When the number of sources K is known, $\mathbf{X}$ can be factored as $\mathbf{X} = \mathbf{W}\mathbf{H}^H$, where $\mathbf{W} \in \mathbb{C}^{M\times K}$ and $\mathbf{H} \in \mathbb{C}^{N\times K}$. Then, using the identity

$$||\mathbf{X}||_* = \min_{\mathbf{X}=\mathbf{W}\mathbf{H}^H} \frac{1}{2}\left(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2\right),$$

$\mathbf{X}$ can be estimated by solving the optimization problem [22]

$$\{\hat{\mathbf{W}}, \hat{\mathbf{H}}\} = \underset{\substack{\mathbf{W}\in\mathbb{C}^{M\times K}\\ \mathbf{H}\in\mathbb{C}^{N\times K}}}{\operatorname{argmin}} \left\|P_\Omega\left(\mathbf{Z}_d - \mathbf{W}\mathbf{H}^H\right)\right\|_F^2 + \mu\left(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2\right) \tag{6}$$

where μ is a regularization parameter. In the next subsection, we modify (6) to exploit the shift-invariance property of the steering matrix $\mathbf{A}$.

3.2. Shift-invariant Matrix Completion

In addition to being a low-rank matrix, $\mathbf{X}$ has additional structure inherited from the array geometry that can be exploited by the MC method. Specifically, when ULAs are employed, the shift-invariance property holds. According to this property, each row of the steering matrix $\mathbf{A}$ is related to the previous one as follows

$$\mathbf{a}_i^H\mathbf{Q} = \mathbf{a}_{i-1}^H \qquad i = 2,\dots,M \tag{7}$$

where $\mathbf{a}_i^H$ is the i-th row of $\mathbf{A}$ and $\mathbf{Q} = \operatorname{diag}(e^{j\theta_1},\dots,e^{j\theta_K})$, as it can be readily verified from (1). From the shift-invariance property, it follows that the column span of $\mathbf{X}^\uparrow$, formed by the first $M-1$ rows of $\mathbf{X}$, and the column span of $\mathbf{X}^\downarrow$, formed by the last $M-1$ rows of $\mathbf{X}$, are identical. In other words, the K-dimensional signal subspaces of $\mathbf{X}^\uparrow$ and $\mathbf{X}^\downarrow$ are the same.

It is then clear that the factor $\mathbf{W}$ in

$$\mathbf{X} = \mathbf{W}\mathbf{H}^H \tag{8}$$

should satisfy the shift-invariance property as well. Since the factorization (8) is not unique, we use a relaxed version of (7) to enforce the following relation between the rows of $\mathbf{W}$, i.e.,

$$\mathbf{w}_i^H\mathbf{T} = \mathbf{w}_{i-1}^H \qquad i = 2,\dots,M \tag{9}$$

where $\mathbf{w}_i^H$ is the i-th row of $\mathbf{W}$ and $\mathbf{T} \in \mathcal{D}$ where $\mathcal{D}$ is the set of $K\times K$ diagonal complex matrices not necessarily unitary.

To enforce (9), the shift-invariant matrix completion (SIMC) problem (6) includes an additional regularization term:

$$\begin{aligned}\{\hat{\mathbf{W}}, \hat{\mathbf{H}}, \hat{\mathbf{T}}\} = \underset{\substack{\mathbf{W}\in\mathbb{C}^{M\times K}\\ \mathbf{H}\in\mathbb{C}^{N\times K}\\ \mathbf{T}\in\mathcal{D}}}{\operatorname{argmin}} & \sum_{(i,j)\in\Omega} \left|\mathbf{Z}_d(i,j) - \mathbf{w}_i^H\mathbf{h}_j\right|^2 + \mu\Big(\sum_{i=1}^M \|\mathbf{w}_i\|_2^2 + \sum_{j=1}^N \|\mathbf{h}_j\|_2^2\Big)\\ & + \alpha\sum_{i=2}^M \|\mathbf{w}_i^H\mathbf{T} - \mathbf{w}_{i-1}^H\|_2^2\end{aligned} \tag{10}$$

where $\mathbf{h}_j^H$ is the j-th row of $\mathbf{H}$ and α is an additional regularization parameter.

The solution $\hat{\mathbf{X}} = \hat{\mathbf{W}}\hat{\mathbf{H}}^H$ can be obtained by iteratively optimizing (10) over each $\mathbf{w}_i^H$, $\mathbf{h}_j^H$ and $\mathbf{T}$ until convergence. To optimize (10) for $\mathbf{w}_i^H$, we take the derivative with respect to $\mathbf{w}_i^H$, assuming $\mathbf{H}$ and $\mathbf{T}$ fixed, and equate it to zero, which provides the following solution

$$\mathbf{w}_i^H = \begin{cases} \Big(\mathbf{g}_1^H + \mathbf{g}_2^H\Big)\Big(\mathbf{Y}_1 + \alpha\mathbf{I}\Big)^{-1} & \text{if } i = 1 \\ \Big(\mathbf{g}_1^H + \mathbf{g}_2^H + \mathbf{g}_3^H\Big)\Big(\mathbf{Y}_1 + \mathbf{Y}_2 + \alpha\mathbf{I}\Big)^{-1} & \text{if } 1 < i < M \\ \Big(\mathbf{g}_1^H + \mathbf{g}_3^H\Big)\Big(\mathbf{Y}_1 + \mathbf{Y}_2\Big)^{-1} & \text{if } i = M \end{cases} \tag{11}$$

where

$$\begin{aligned} \mathbf{g}_1^H &= \sum_{j \in \mathcal{J}_i} \mathbf{Z}_d(i,j)\mathbf{h}_j^H \\ \mathbf{g}_2^H &= \alpha \mathbf{w}_{i+1}^H \mathbf{T} \\ \mathbf{g}_3^H &= \alpha \mathbf{w}_{i-1}^H \mathbf{T}^H \\ \mathbf{Y}_1 &= \sum_{j \in \mathcal{J}_i} \mathbf{h}_j \mathbf{h}_j^H + \mu\mathbf{I} \\ \mathbf{Y}_2 &= \alpha \mathbf{T}\mathbf{T}^H \end{aligned}$$

and $\mathcal{J}_i$ is the set of observed indices of the i-th row of $\mathbf{Z}_d$. Similarly, (10) can be optimized for $\mathbf{h}_j^H$ to find the solution as

$$\mathbf{h}_j^H = \Big(\sum_{i \in \mathcal{I}_j} \mathbf{Z}_d(i,j)^* \mathbf{w}_i^H\Big)\Big(\sum_{i \in \mathcal{I}_j} \mathbf{w}_i \mathbf{w}_i^H + \mu\mathbf{I}\Big)^{-1} \tag{12}$$

where $\mathcal{I}_j$ is the set of observed indices of the j-th column of $\mathbf{Z}_d$. Since $\mathbf{T} = \text{diag}(t_1, \ldots, t_K)$ is a diagonal matrix, (10) can be optimized for each diagonal element t_k individually. To this end, the third term in the right hand side of (10) can be rewritten in terms of t_k as

$$\sum_{i=2}^{M} \|\mathbf{w}_i^H \mathbf{T} - \mathbf{w}_{i-1}^H\|_2^2 = \sum_{i=2}^{M} \sum_{k=1}^{K} \big|t_k \mathbf{W}(i,k) - \mathbf{W}(i-1,k)\big|^2, \tag{13}$$

which can be optimized with respect to t_k to get

$$t_k = \frac{\sum_{i=2}^{M} \mathbf{W}(i-1,k)\mathbf{W}^*(i,k)}{\sum_{i=2}^{M} \big|\mathbf{W}(i,k)\big|^2} \tag{14}$$

3.3. Post-processing via Optimal Subspace Estimation (OSE)

As the shift-invariance property is enforced through a regularization term, the solution of (10) provides a low-rank data matrix, $\hat{\mathbf{X}}$, which has the required

structure only in an approximate fashion. This motivates applying the Optimal Subspace Estimation (OSE) technique as a final post-processing step of our algorithm. The OSE algorithm takes $\hat{\mathbf{X}}$ as input and provides an estimate of the underlying noise-free signal subspace with the required shift-invariant structure. Let $\mathbf{U}_{ose} \in \mathbb{C}^{M \times K}$ be a basis for this subspace, and let $\mathbf{P}_{ose} = \mathbf{U}_{ose}\mathbf{U}_{ose}^H$ be its orthogonal projection matrix. Then, the output of the OSE algorithm is

$$\hat{\mathbf{X}}_{ose} = \mathbf{P}_{ose}\hat{\mathbf{X}} \tag{15}$$

For a full account of the OSE method the reader is referred to [18, 23, 24].

A summary of the shift-invariant matrix completion method, denoted as SIMC, is provided in Algorithm 1. Once $\hat{\mathbf{X}}_{ose}$ is obtained, any subspace-based method can be used to estimate the DOAs. We choose ESPRIT, as it effectively exploits the shift-invariance property of ULAs.

Algorithm 1: Shift-Invariant Matrix Completion (SIMC)

Input: $\mathbf{Z}_d$, μ, K, $itr_{\max}$
Output: $\hat{\mathbf{R}}_{ose}$
Initialization: $\hat{\mathbf{T}} = \mathbf{I}$, $itr = 1$
Compute the SVD of $\mathbf{Z}_d = \mathbf{F}\mathbf{\Lambda}\mathbf{G}^H$ and initialize $\hat{\mathbf{W}} = \mathbf{F}_K\mathbf{\Lambda}_K^{1/2}$ and $\hat{\mathbf{H}} = \mathbf{G}_K\mathbf{\Lambda}_K^{1/2}$, using the K largest singular vectors and singular values of $\mathbf{Z}_d$ (best K-rank approximation of $\mathbf{Z}_d$)
Set α as in (17)
REPEAT
Compute $\hat{\mathbf{W}}$, $\hat{\mathbf{H}}$ and $\hat{\mathbf{T}}$ using (11), (12) and (14), respectively
$itr = itr + 1$
Until Convergence = true or $itr = itr_{\max}$
Compute $\hat{\mathbf{X}} = \hat{\mathbf{W}}\hat{\mathbf{H}}^H$
Apply OSE algorithm to estimate $\hat{\mathbf{X}}_{ose} = \mathbf{P}_{ose}\hat{\mathbf{X}}$

The SIMC algorithm has a computational cost of $\mathcal{O}((M+N)K^3)$ per iteration, which is basically the cost of standard MC algorithms based on alternating least squares, since the extra cost due to (14) is negligible. The OSE post-processing step, has a computational complexity of $\big(\mathcal{O}(M^2N) + 2\mathcal{O}((MK)^3)\big)$. Finally, the proposed initialization step performs a compact SVD with cost $\mathcal{O}(MK^2)$. Note that for this problem $K \ll M$.

3.4. Selection of regularization parameters

The values of α and μ in (10) control the trade-off among the fulfillment of the shift-invariance property, the fitting to the observed data and the nuclear norm of the solution. Since α enforces the shift-invariance property into $\hat{\mathbf{X}}$, its value should depend on some measure that quantifies the compliance of the shift-invariance property by the original sparse matrix $\mathbf{Z}_d$. As we know, given a rank-K matrix for which the shift-invariance property holds, the subspaces spanned by the first and the last $M-1$ rows are identical. Thus, the regularization

parameter α is chosen to be a function of the chordal subspace distance [25] between $\mathbf{Z}_d^{\uparrow}$ and $\mathbf{Z}_d^{\downarrow}$, which are formed by the first and the last $M-1$ rows of the sparse $\mathbf{Z}_d$, respectively.

Specifically, let $\mathbf{U}_1 \in \mathbb{C}^{(M-1)\times K}$ and $\mathbf{U}_2 \in \mathbb{C}^{(M-1)\times K}$ be the K largest left singular vectors (that is, those associated to the K largest singular values) of $\mathbf{Z}_d^{\uparrow}$ and $\mathbf{Z}_d^{\downarrow}$, respectively. Then, the chordal subspace distance between $\mathbf{Z}_d^{\uparrow}$ and $\mathbf{Z}_d^{\downarrow}$ is

$$d_{cs} = \|\mathbf{U}_1\mathbf{U}_1^H - \mathbf{U}_2\mathbf{U}_2^H\|_F. \tag{16}$$

A large value of d_{cs} implies that the K-dimensional subspaces extracted from $\mathbf{Z}_d^{\uparrow}$ and $\mathbf{Z}_d^{\downarrow}$ are far apart from each other and, consequently, the shift-invariance property does not hold. This in turn implies that a large α must be used in the reconstruction process. According to our simulations, a value that provides good performance for a wide range of scenarios is

$$\alpha = d_{cs}\mu, \tag{17}$$

where $\mu = \frac{M}{20}$.

4. Perturbation analysis

The main factor impacting the performance of the random multi-switch sampling scheme is how well the signal subspace is preserved. The SIMC algorithm aims at estimating an improved signal subspace by leveraging its shift-invariant low-rank structure. This section analyzes how DOA estimation is impacted when performed after MC.

Since the DOA estimates are essentially determined by the singular vectors of the signal subspace, we want to assess how much the principal directions change after each processing step of the original sparse data matrix. To do so, we will analyze the problem from a matrix perturbation standpoint. A perturbed matrix is a matrix which has its singular values and vectors altered after an addition with a second matrix. Thus, $\mathbf{Z}_d$ in (4) is a perturbed version of $\mathbf{X}$, with the perturbation being caused by the missing entries and noise. The Davis-Kahan theorem is a useful tool to measure the angular difference between the singular vectors of two matrices. We show below Theorem 1 in [26] adapted to our use-case.

Theorem 1. Davis-Kahan sin theorem. *[26] Let $\mathbf{U}_X$ and $\mathbf{U}_{\tilde{X}}$ denote the first K left singular vectors of $\mathbf{X}$ and the perturbed $\tilde{\mathbf{X}}$, respectively, and $\Theta(\mathbf{U}_X, \mathbf{U}_{\tilde{X}})$ be the $K\times K$ diagonal matrix containing the principal angles $cos^{-1}(\xi_i)_{i=1}^K$, where ξ_i is the i-th singular value of $\mathbf{U}_X^H\mathbf{U}_{\tilde{X}}$. Then,*

$$||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{\tilde{X}})||_F \leq \frac{2\sqrt{K}(2||\mathbf{X}||_2 + ||\mathbf{X}-\tilde{\mathbf{X}}||_2)\min(||\mathbf{X}-\tilde{\mathbf{X}}||_2, \frac{1}{\sqrt{K}}||\mathbf{X}-\tilde{\mathbf{X}}||_F)}{\lambda_K(\mathbf{X})} \tag{18}$$

Theorem 1 shows that the subspace distance between the singular vectors $\mathbf{U}_X$ and $\mathbf{U}_{\tilde{X}}$ scales with the norm difference between $\mathbf{X}$ and $\tilde{\mathbf{X}}$. Interestingly, it also shows that the larger the K-th singular value is, the smaller the subspace distance will be. Below, we leverage the Davis-Kahan theorem to compare the signal space of $\mathbf{X}$ firstly with that of the sparse matrix $\mathbf{Z}_d$, and secondly with the recovered estimate $\hat{\mathbf{X}}_{MC}$ in (6) obtained through MC.

Clearly, due to the missing data, $\mathbf{Z}_d$ is a poor approximation to $\mathbf{X}$. Nevertheless, the K first singular vectors of the sampled matrix are often used as a crude estimate or initialization point for iterative algorithms [27]. Let $P_K(\mathbf{Z}_d)$ denote the projection of $\mathbf{Z}_d$ onto the subspace spanned by its first K left singular vectors, which is obtained by setting $\lambda_k(\mathbf{Z}_d) = 0,\ \forall\, k > K$. Moreover, let us assume a uniform random sampling scheme where each entry in $\mathbf{Z}$ is sampled with probability $q = L/M$. From [28], we have the bound

$$||\mathbf{X} - \frac{1}{q}P_K(\mathbf{Z}_d)||_2 \leq C||\mathbf{X}||_\infty \frac{N^{\frac{3}{4}}}{M^{\frac{1}{4}}\sqrt{q}} + C\sigma\sqrt{\frac{N\log M}{q}} \tag{19}$$

which is satisfied with probability greater than $1 - \frac{1}{M^3}$ for some constant C. Note the scaling $\frac{1}{q}$ of $P_K(\mathbf{Z}_d)$ in (19), which compensates for the norm loss due to the missing entries. Thus, since $\frac{1}{q}P_K(\mathbf{Z}_d)$ and $\mathbf{Z}_d$ share the first K left singular vectors, then

$$||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{Z_d})||_F = ||\sin\Theta(\mathbf{U}_X, \mathbf{U}_{\frac{1}{q}P_K(\mathbf{Z}_d)})||_F$$

and we can use (19) in conjunction with Theorem 1 to bound the subspace distance.

With regard to $\hat{\mathbf{X}}_{MC}$, assuming that $N \geq M$ the recently developed bounds in [29] show that

$$||\mathbf{X} - \hat{\mathbf{X}}_{MC}||_2 \leq ||\mathbf{X}||_2 \frac{\sigma}{\lambda_K(\mathbf{X})}\sqrt{\frac{N}{q}}. \tag{20}$$

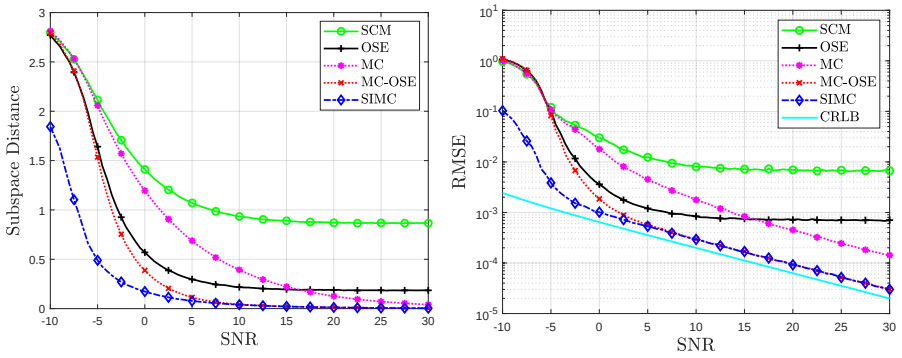
with probability exceeding $1 - \frac{1}{N^3}$.

Assuming constant $q = L/M$ and M, and bounded $||\mathbf{X}||_\infty$, we have that the bound for $P_K(\mathbf{Z}_d)$ in (19) grows as $\mathcal{O}(N^{\frac{3}{4}})$. Therefore, comparing it to that of $\hat{\mathbf{X}}_{MC}$ in (20), we observe that the bound for $\hat{\mathbf{X}}_{MC}$ grows as $\mathcal{O}(\sqrt{N})$. Therefore, we can conclude that MC will improve the DOA estimates.

5. Simulation Results

In this section we illustrate the performance of the proposed SIMC algorithm by means of Monte Carlo simulations. For comparison, we include the performance of the following methods:

- SCM: The sample covariance matrix without MC is estimated as $\hat{\mathbf{R}}_d = \frac{1}{N}\mathbf{Z}_d\mathbf{Z}_d^H$.



(a) (b)

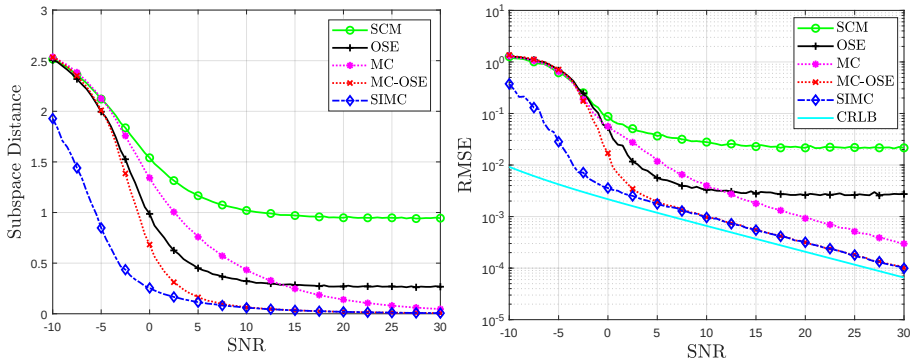
Figure 2: Subspace distance (a) and RMSE (b) vs. SNR for $M = 100$, $N = 80$, $K = 5$, $\Delta_\theta = 10^\circ$ and $L = 50$.

- OSE: The shift-invariance property is enforced by applying OSE to $\mathbf{Z}_d$ (without MC).
- MC: The standard MC algorithm solution given by (6) is used to reconstruct $\mathbf{X}$ from $\mathbf{Z}_d$.
- MC-OSE: OSE is applied as a post-processing step to the previous method.
- SIMC: The proposed method.

ESPRIT is used to compute the DOAs for all competing methods. As figures of merit we use: i) the chordal subspace distance between the true signal subspace and the estimated signal subspace, and ii) the Root Mean Squared Error (RMSE) for the DOA estimates in radians. The chordal distance between the true signal subspace or column span of $\mathbf{X}$, and the estimated signal subspace or column span of $\hat{\mathbf{X}}$ is shown to assess how different these subspaces are. Note that this distance is different from the chordal distance in (16) used to select the regularization parameter α.

For all simulations we assume that K uncorrelated narrowband signals with a separation of Δ_θ (in electrical angle) are impinging on a ULA with M half-wavelength separated antennas. Unless stated otherwise, sources have equal power. For both SIMC and MC, we use $\mu = M/20$ and $itr_{max} = 200$. SNR $= 10 \log \frac{\mathrm{tr}(\mathbf{R}_s)}{M\sigma^2}$, where $\mathbf{R}_s$ is the signal covariance matrix and σ^2 is the noise variance. L denotes the number of randomly sampled sensors per snapshot. The Cramer-Rao lower bound (CRLB) [30] when the full data matrix $\mathbf{Z}$ is available is included as a reference benchmark.

In the first example, we consider a sample-poor scenario with $M = 100$ antennas, $N = 80$ snapshots, $K = 5$ sources and $\Delta_\theta = 10^\circ$. At each time instant the multi-switch network randomly selects $L = 50$ out of the $M = 100$ antennas. Fig. 2 shows the subspace distance (left plot) and the RMSE



(a) (b)

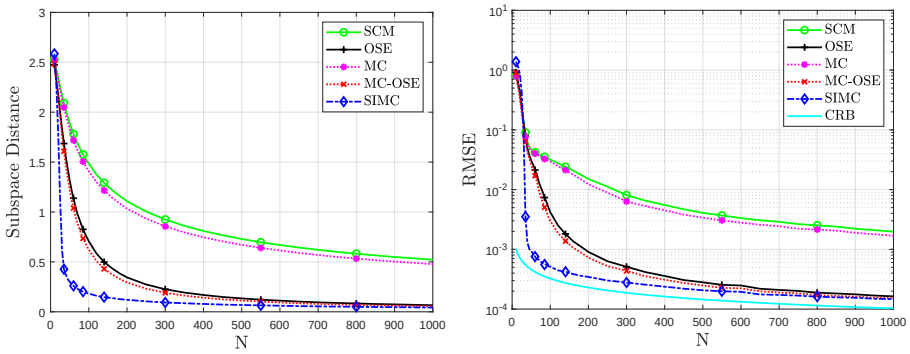
Figure 3: Subspace distance (a) and RMSE (b) vs. SNR for $M = 50$, $N = 50$, $K = 4$, $\Delta_\theta = 10°$, $\mathbf{\Sigma} = \sigma_s^2 \text{diag}(1, 0.8, 0.6, 0.5)$ and $L = 25$.

(right plot) vs. the SNR. The performance of SCM and OSE without MC saturates at high SNR due to the relatively high fraction of missing entries. The benefits of using MC techniques in combination with enforcing the shift-invariance property are evident, specially at low or moderate SNRs. In fact, even with 50 % of missing data and SNR$\approx$ 0, we observe that SIMC is close to the CRLB (which gives us the performance limit when all data are available). At high SNRs MC-OSE and SIMC have identical performances, which suggest that the post-processing OSE step is sufficient to enforce in the solution the required invariance to displacements.

The second example considers a scenario with $M = 50$ antennas, $K = 4$ sources with $\Delta_\theta = 10°$, $N = 50$ snapshots and $L = 25$ (i.e., 50% of missing entries in $\mathbf{Z}_d$). The sources in this example have unequal power with signal covariance matrix $\mathbf{\Sigma} = \text{diag}(1, 0.8, 0.6, 0.5)$. A similar behavior to the previous example is observed in Fig. 3, with SIMC providing satisfactory performance over a large range of SNR values.

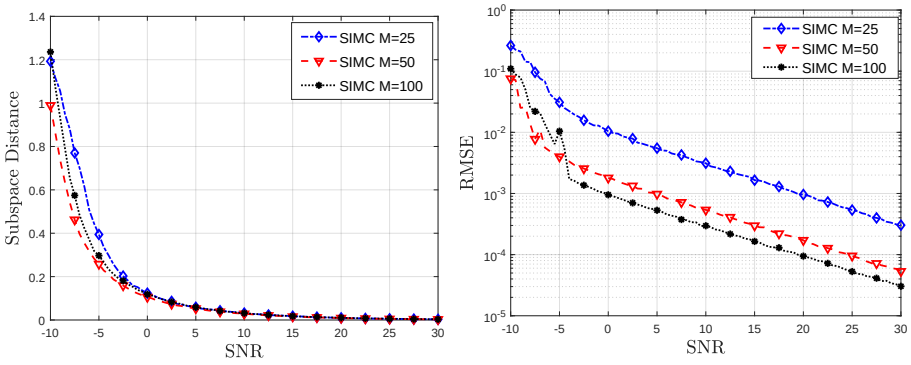
The third example compares the performance of the methods with respect to N for $M = 200$, $K = 4$, SNR$= -5$ dB, $\Delta_\theta = 5°$, and $L = 100$. We can observe in Fig. 4 that if N is large enough, SIMC, MC-OSE and OSE provide very similar results. However, SIMC outperforms the rest of methods when N is small. This example demonstrates a clear advantage of the proposed method in the small-sample regime where $N \leq M$.

The next example compares the performances of arrays of different number of antennas when the number of sampled sensors L is fixed. Therefore, the spatial sampling ratio L/M decreases as M increases. We consider ULAs with $M = 25$, $M = 50$ and $M = 100$ antennas using a fixed value of $L = 25$ so that at every snapshot the percentages of sampled sensors are 100%, 50% and 25%, respectively. For all three cases, the number of snapshots is $N = 100$ and $K = 3$ sources with $\Delta_\theta = 10°$ of separation impinge on each array. Since L and N are fixed, the energy consumption will be roughly the same for all



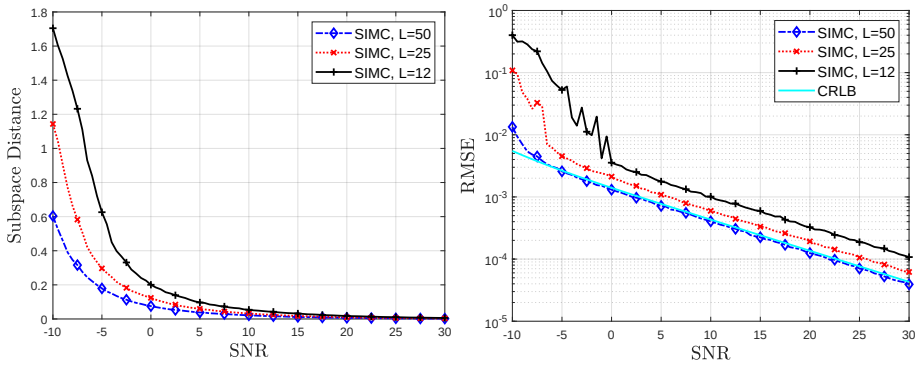
(a) (b)

Figure 4: Subspace distance (a) and RMSE (b) vs. N for $M = 200$, $K = 4$, SNR$= -5$ dB, $\Delta_\theta = 5°$, and $L = 100$.



(a) (b)

Figure 5: Subspace distance (a) and RMSE (b) vs. SNR for ULAs with different number of antennas when $K = 3$, $\Delta_\theta = 10°$, $N = 100$ and $L = 25$.



(a) (b)

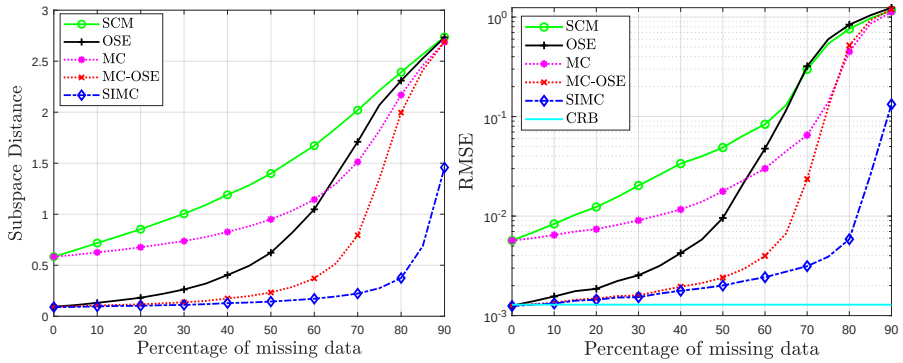
Figure 6: Subspace distance (a) and RMSE (b) vs. SNR when $M = 50$, $N = 80$, $K = 3$ and $\Delta_\theta = 10°$ for $L = (50, 25, 12)$.

array architectures. However, the effective spatial resolution is improved as M increases, as it is observed in Fig. 5. In this way, the proposed SIMC algorithm allows us to increase the spatial resolution of an array with a fixed number of RF chains. In other words, we can trade-off spatial resolution by energy consumption.

The following experiments analyze the impact of the percentage of missing data on the methods under comparison. We consider a scenario with $M = 50$ antennas, $N = 80$, snapshots and $K = 3$ sources with a separation of $\Delta_\theta = 10°$. Fig. 6 shows the subspace distance and the RMSE curves vs. the SNR when the number of sampled antennas is $L = 50$, $L = 25$, or $L = 12$. Obviously, the best performance is achieved when all sensors are sampled. Nevertheless, performance degrades smoothly with L and hence both the hardware costs and energy consumption can be substantially reduced with only a minor performance degradation. As we increase L, we observe more entries of the data matrix and the MSE of the SIMC method approaches the CRLB.

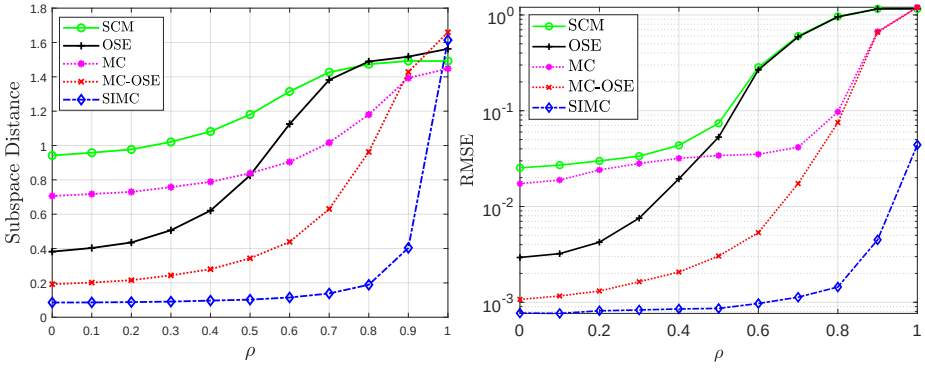
Fig. 7 shows results for the same scenario when the number of sampled sensors is $L = \lfloor \frac{M(100-P_s)}{100} \rfloor$, where P_s is the percentage of missing data and $\lfloor \cdot \rfloor$ is the floor function. It can be observed in Fig. 7 that the performance of SIMC is robust against missing data, providing satisfactory performance for $P_s < 70\%$. The results of Fig. 7 allow us to conclude that to obtain accurate signal subspace and DOA estimates it is important to exploit in the reconstruction of the data matrix both its low-rank structure and its shift-invariant structure. When exploited independently, the shift-invariant structure (OSE) provides more benefits than the low-rank structure (MC) for $P_s < 50\%$.

In the last experiment, we evaluate the impact of having correlated sources. We consider a scenario with $K = 2$ correlated sources when $M = 100$, $N = 80$, SNR= 0 dB, $\Delta_\theta = 5°$, and $L = 25$. The correlation coefficient between the two sources, ρ, varies from 0 (uncorrelated) to 1 (fully correlated). As Fig. 8 shows, SIMC outperforms the rest of methods and provides accurate DOA estimates



(a) (b)

Figure 7: Subspace distance (a) and RMSE (b) vs. Percentage of missing data when $M = 50$, $N = 50$, $K = 5$, $\Delta_\theta = 10^\circ$, $SNR = 5$ dB and sources have equal power



(a) (b)

Figure 8: Subspace distance (a) and RMSE (b) vs. ρ for $M = 100$, $N = 80$, $K = 2$, SNR$= 0$ dB, $\Delta_\theta = 5^\circ$, $L = 25$.

even for highly correlated sources $\rho < 0.8$. Nevertheless, the performance of SIMC under correlated sources needs additional theoretical analysis.

6. Conclusion

The high hardware complexity and energy consumption of massive MIMO systems is a challenge for its fully-digital implementation. A solution is to reduce the number of RF chains by performing random antenna selection techniques, which result in a data matrix with multiple missing entries. In this paper we have proposed a matrix completion technique tailored to this array processing architecture. The reconstruction algorithm exploits both the low-rank structure of the partially observed matrix and the shift-invariance property of uniform linear arrays. After reconstruction, standard high-resolution subspace-based techniques can be used for DOA estimation. As long as the number of RF chains is sufficiently larger than the number of sources, the proposed shift-invariant matrix completion (SIMC) method provides a substantial reduction of hardware costs and energy consumption without significant performance loss in resolution or DOA estimation accuracy.

References

[1] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Prentice Hall, 2015.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, Jan. 2013.

[4] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[5] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal of Selected Areas in Communications*, vol. 34, no. 4, pp. 697–709, 2016.

[6] H. L. Van Trees, *Detection, estimation and modulation theory: Optimum array processing (part IV)*. John Wiley & Sons, 2002.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, pp. 276–280, 1986.

[8] R. Roy and T. Kailath, "ESPRIT- estimation of signal parameters via rotational invariance techniques," *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[9] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[10] P. Pal and P. P. Vaidyanathan, "A grid-less approach to underdetermined direction of arrival estimation via low rank matrix denoising," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 737–741, 2014.

[11] N. Ito, E. Vincent, N. Ono, and S. Sagayama, "Robust estimation of directions-of-arrival in diffuse noise based on matrix-space sparsity. [research report]," *RR-8120, INRIA. hal-00746271*, 2012.

[12] W. Tan and X. Feng, "Covariance matrix reconstruction for direction finding with nested arrays using iterative reweighted nuclear norm minimization," *Int. Journal of Antennas and Propagation*, 2019.

[13] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Transactions on Signal Processing*, vol. 66, pp. 5956–5971, 2018.

[14] Z. Zheng, Y. Huang, W. Q. Wang, and H. C. SO, "Direction-of-arrival estimation of coherent signals via coprime array interpolation," *IEEE Signal Processing Letters*, vol. 27, pp. 585–589, 2020.

[15] X. Wu, W. Zhu, and J. Yan, "A Toeplitz covariance matrix reconstruction approach for direction-of-arrival estimation," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 8223–8237, 2017.

[16] B. Liao, C. Guo, L. Huang, and J. Wen, "Matrix completion based direction-of-arrival estimation in nonuniform noise," in *2016 IEEE International Conference on Digital Signal Processing (DSP).*, Beijing, China, Oct. 2016.

[17] V. Garg, P. Giménez-Febrer, A. Pagès-Zamora, and I. Santamaria, "Source enumeration via Toeplitz matrix completion," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 6004–6008.

[18] R. J. Vaccaro and Y. Ding, "Optimal subspace-based parameter estimation," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Proc. (ICASSP)*, Minneapolis, MN, USA, Apr. 1993, pp. IV368–IV371.

[19] C. Davis and W. M. Kahan, "The rotation of eigenvectors by a pertubation," *SIAM J. Numer. Anal.*, vol. 7, pp. 1–46, 1970.

[20] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Multi-switch for antenna selection in massive MIMO," in *IEEE Global Communications Conference (GLOBECOM)*, San Diego (CA), USA, Dec. 2015.

[21] Z. Weng and X. Wang, "Low-rank matrix completion for array signal processing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2697–2700.

[22] N. Srebro, J. Rennie, and T. Jaakkola, "Maximum-margin matrix factorization," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, Dec 2005, pp. 1329–1336.

[23] R. J. Vaccaro, "The role of subspace estimation in sensor array signal processing," in *Proc. 2017 Underwater Acoust. Signal Process. Work.*, University of Rhode Island, Oct. 2017.

[24] C. Grayson, "Applications of optimal subspace estimation," *Open Access Master's Theses. Paper 959 http://digitalcommons.uri.edu/theses/959*, 2016.

[25] A. Srivastava and E. Klassen, "Monte Carlo extrinsic estimators of manifold-valued parameters," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 299–308, 2002.

[26] Y. Yu, T. Wang, and R. J. Samworth, "A useful variant of the Davis-Kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2015.

[27] P. Giménez-Febrer and A. Pagès-Zamora, "Matrix completion of noisy graph signals via proximal gradient minimization," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 4441–4445.

[28] R. Keshavan, A. Montanari, and S. Oh, "Matrix Completion from Noisy Entries," Jun 2009. [Online]. Available: http://arxiv.org/abs/0906.2027

[29] Y. Chen, Y. Chi, J. Fan, C. Ma, and Y. Yan, "Noisy Matrix Completion: Understanding Statistical Guarantees for Convex Relaxation via Nonconvex Optimization," 2019. [Online]. Available: https://arxiv.org/pdf/1902.07698.pdf

[30] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, 1990.