



High-Dimensional MVDR Beamforming: Optimized Solutions Based on Spiked Random Matrix Models

Liusha Yang, Matthew R. McKay, Romain Couillet

► To cite this version:

Liusha Yang, Matthew R. McKay, Romain Couillet. High-Dimensional MVDR Beamforming: Optimized Solutions Based on Spiked Random Matrix Models. IEEE Transactions on Signal Processing, 2018, 66 (7), pp.1933-1947. 10.1109/tsp.2018.2799183 . hal-01957672

HAL Id: hal-01957672

<https://hal.science/hal-01957672>

Submitted on 19 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

High-Dimensional MVDR Beamforming: Optimized Solutions based on Spiked Random Matrix Models

Liusha Yang*, Matthew R. McKay*, Romain Couillet†

Abstract—Minimum variance distortionless response (MVDR) beamforming (or Capon beamforming) is among the most popular adaptive array processing strategies due to its ability to provide noise resilience while nulling out interferers. A practical challenge with this beamformer is that it involves the inverse covariance matrix of the received signals, which must be estimated from data. Under modern high-dimensional applications, it is well-known that classical estimators can be severely affected by sampling noise, which compromises beamformer performance. Here we propose a new approach to MVDR beamforming which is suited to high-dimensional settings. In particular, by drawing an analogy with the MVDR problem and the so-called “spiked models” in random matrix theory, we propose robust beamforming solutions which are shown to outperform classical approaches (e.g., matched filters and sample matrix inversion techniques), as well as more robust solutions, such as methods based on diagonal loading. The key to our method is the design of an optimized inverse covariance estimator which applies eigenvalue clipping and shrinkage functions that are tailored to the MVDR application. Our proposed MVDR solution is simple, in closed form, and easy to implement.

Index Terms—MVDR beamforming, spiked covariance model, random matrix theory.

I. INTRODUCTION

Adaptive beamforming is widely used in wireless communications, radar, sonar, microphone array speech processing, medical imaging and other areas [1–5]. A general goal is to adaptively steer a beam towards a desired signal, while placing nulls at interference directions. The well-known minimum variance distortionless response (MVDR) adaptive beamformer [6] is designed to linearly combine the outputs of the sensors in order to minimize the array output power, while maintaining a fixed response towards the desired signal. A desirable feature of the MVDR beamformer is that it yields the maximum signal-to-interference-plus-noise ratio (SINR) at the array output [7].

A key practical issue with the MVDR beamformer is that its construction involves the inverse covariance matrix of the received signals, which must be estimated from observations. The standard approach is to use the inverse of the sample

covariance matrix (SCM), a strategy often referred to as “sample matrix inversion” (SMI). However, while this performs well in classical application settings in which the number of observations greatly exceeds the number of sensors of the receive array, in modern data-limited scenarios or high-dimensional applications (with large numbers of sensors), estimation errors in the SCM can significantly compromise beamformer performance [8–13]. The severity of this degradation is also known to depend strongly on whether the signal of interest (SoI) is present or absent from the sample data used for covariance estimation [11–13].

Numerous approaches have been proposed to design more robust beamforming solutions which aim to overcome this problem (e.g. [8, 14–17]). The most popular are algorithms that apply diagonal loading, which form a linear combination of the SCM and a scaled identity matrix [15–22]. (Note that this estimator is also commonly used in other fields, including mathematical finance [22].) The idea of diagonal loading is to regularize the SCM in order to provide resilience against inaccuracies, such as sampling noise. Numerous recent studies have proposed different solutions for optimizing the regularization parameter of diagonal loading [17, 20–24]. Extensions have also been proposed which replace the SCM with a robust covariance matrix estimator (such as Tyler’s estimator) to provide resilience against outliers [25, 26], and optimized solutions for the loading parameter have been obtained under different scenarios [25, 27]. An important point, however, is that the diagonal loading construction (even with an optimized diagonal loading factor) is not the most “natural” construction for the MVDR beamforming problem, since it does not exploit basic structural features of the received signal covariance matrix under high-dimensional settings, as we indicate below. Moreover, as we will demonstrate numerically, even when given the (unobservable) theoretically-optimal parameter for the MVDR beamformer, its performance can still be quite poor under data-limited or high-dimensional scenarios.

Despite the fact that MVDR is a classical construction which has been well studied, the problem of designing beamformers which yield high performance at low and high signal-to-noise ratios (SNRs), under high-dimensional settings, and in the face of data limitations, has still not been fully resolved. The key challenge is to develop covariance estimation solutions which are appropriately optimized for the MVDR application. In this paper, we propose a novel beamforming solution aimed at addressing this problem.

We consider a high-dimensional setting for which the number of array sensors is assumed large, and the number of interferers is small in comparison. Relevant large-array application

*L. Yang and M. R. McKay are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. (email: lyangag@connect.ust.hk; m.mckay@ust.hk).

†R. Couillet is with the Laboratoire de Signaux et Systèmes (L2S, UMR8506), CNRS-CentraleSupélec-Université Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France (email: romain.couillet@centralesupelec.fr).

Yang and McKay’s work was supported by the Hong Kong RGC General Research Fund under grant numbers 16206914 and 16203315. Couillet’s work was supported by the ANR Project RMT4GRAPH (ANR-14-CE28-0006).

examples include [28–31]. Under such setting, our proposed method is based on the fact that the receive covariance matrix (with or without the SoI) has a specific structure reminiscent of the so-called “spiked models” in random matrix theory (RMT). Specifically, this construction comprises a low-rank perturbation of a scaled identity matrix, with the scaling factor corresponding to the noise variance σ^2 . Here, the leading few eigenmodes represent the subspace spanned by the signal and interferers, while the bulk of the eigenmodes (with identical eigenvalues σ^2) represent the noise subspace. This structure allows for the design of robust estimation solutions which are suitable for the MVDR problem, as opposed to generic solutions, such as the SCM or diagonal loading. Specifically, we consider a class of estimators which apply transformations (or “shrinkage” functions) to the eigenvalues of the SCM. In accordance with the spiked structure, the bulk of these eigenvalues are hard-thresholded (or “clipped”) to the noise variance σ^2 , while the transformations for the few leading eigenvalues are carefully optimized to maximize the MVDR performance. Specifically, this optimization is done under the so-called random matrix asymptotic regime, which considers both the number of sensors N and the number of observations n growing together such that N/n approaches a constant.

To the best of our knowledge, this approach, which exploits the spike-model construction, has not been considered previously for the problem of MVDR beamforming despite the fact that the strategy appears natural. We point out however, that there is some related and very recent work [32] that proposed optimized spike-model-based covariance estimators for a range of different objective functions. A defining feature of these solutions is that they were designed for a class of objective functions having a certain rotational-invariance property, which does not hold for the MVDR problem. Moreover, the derivation approach relied heavily on this property, and is not easily adapted to our problem. As such, here we applied a different technique to identify the MVDR-optimal eigenvalue shrinkage functions yielding our proposed inverse covariance estimator, and consequently, our proposed MVDR beamforming solution.

Our proposed method is simple to implement, and it is shown through numerical examples to yield exceptional beamformer performance for high-dimensional, data-limited scenarios, at high and low SNRs, and in scenarios where the interferers are substantially stronger than the SoI. The robustness of our approach is calibrated against the traditional benchmarks, including the SMI, as well as more robust solutions including diagonal loading and eigen-subspace beamforming [8, 14]. Notably, we find that the diagonal loading method—as a widely used modern MVDR approach—yields degraded performance compared with our proposed solution, even when fed the best possible “oracle” parameterization (which is normally unknown in practice). The importance of our optimized solution is also calibrated against beamformers constructed based on other relevant spike-model-based covariance estimators from [32]; again, demonstrating appreciable performance gains.

II. BACKGROUND

A. Array signal model and optimal MVDR beamforming

Consider a uniform linear array consisting of N sensor elements which receives $m < N$ narrowband signals over an observation period of n snapshots. At snapshot $j \in \{1, \dots, n\}$, the array observation vector $\mathbf{x}(j) \in \mathbb{C}^N$ is given by

$$\mathbf{x}(j) = \sqrt{p_1} \mathbf{a}(\theta_1) z_1(j) + \sum_{i=2}^m \sqrt{p_i} \mathbf{a}(\theta_i) z_i(j) + \mathbf{n}(j),$$

where for $i = 1, \dots, m$, $\mathbf{a}(\theta_i) \in \mathbb{C}^N$ is the unit norm steering vector of the i -th source, parameterized by its direction of arrival (DoA) $\theta_i \in (-\pi, \pi]$, whereas $p_i \in \mathbb{R}^+$ is the corresponding signal power. We consider the case where $z_1(j)$ denotes a SoI, while $\{z_i(j)\}_{i=2}^m$ denote interferers. All signals, $z_i(j) \in \mathbb{C}$, are independent complex Gaussian with zero mean and variance one. Similarly, the noise samples $\mathbf{n}(j) \in \mathbb{C}^N$ are independent, zero mean and spatially-white complex Gaussian with variance σ^2 (i.e., $\text{cov}(\mathbf{n}(j)) = \sigma^2 \mathbf{I}_N$), assumed independent of $z_i(j)$ for all i and j . We allow the DoAs and the powers to possibly vary with N , though this will not be shown explicitly.

Let us drop the time index j for convenience. The interference-plus-noise covariance matrix is given by

$$\begin{aligned} \mathbf{C}_{i+n} &= \mathbb{E} \left[\left(\sum_{i=2}^m \sqrt{p_i} \mathbf{a}(\theta_i) z_i + \mathbf{n} \right) \left(\sum_{i=2}^m \sqrt{p_i} \mathbf{a}(\theta_i) z_i + \mathbf{n} \right)^H \right] \\ &= \sum_{i=2}^m p_i \mathbf{a}(\theta_i) \mathbf{a}^H(\theta_i) + \sigma^2 \mathbf{I}_N. \end{aligned}$$

For simplicity, we will assume that σ^2 is known; noting that numerous methods have been proposed for estimating this quantity consistently (see, e.g., [33–36]). We will also assume initially that m is known; but we will explore a simple method for estimating this parameter in our numerical results.

The received signals are linearly combined using a beamformer $\mathbf{h} \in \mathbb{C}^N$, transforming each vector output sample $\mathbf{x}$ to a scalar $y = \mathbf{h}^H \mathbf{x}$. The aim of the beamformer is to enhance the SoI while attenuating interferers originating from other directions. To this end, we consider the classical MVDR beamformer [6], which seeks to minimize the output interference-plus-noise power

$$P_{i+n}(\mathbf{h}) = \mathbb{E} \left[\left| \mathbf{h}^H \left(\sum_{i=2}^m \sqrt{p_i} \mathbf{a}(\theta_i) z_i + \mathbf{n} \right) \right|^2 \right] = \mathbf{h}^H \mathbf{C}_{i+n} \mathbf{h}$$

while constraining unity beam response along a look direction of the SoI. Specifically, for look direction θ_1 (assumed known), this beamformer solves

$$\min_{\mathbf{h} \in \mathbb{C}^N} P_{i+n}(\mathbf{h}) \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{a}(\theta_1) = 1,$$

which takes the well-known form [6, 37]

$$\mathbf{h}_{\text{MVDR}} = \frac{\mathbf{C}_{i+n}^{-1} \mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1) \mathbf{C}_{i+n}^{-1} \mathbf{a}(\theta_1)}. \quad (1)$$

We note that in addition to minimizing the total output power,

the MVDR beamformer also maximizes the output SINR [7]

$$\text{SINR} = \frac{p_1 \mathbf{h}^H \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) \mathbf{h}}{\mathbf{h}^H \mathbf{C}_{i+n} \mathbf{h}}. \quad (2)$$

Hence, the beamformer performance may be also quantified in terms of its achieved SINR.

In practice, the optimal MVDR beamformer is challenging to construct, since the inverse covariance matrix $\mathbf{C}_{i+n}^{-1}$ is unknown, and it is difficult to estimate without access to dedicated “SoI free” training samples (i.e., comprising only of interference and noise). To address this problem, an equivalent representation of the optimal beamformer is often used, which involves the *total* array covariance matrix

$$\mathbf{C}_N = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = p_1 \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) + \mathbf{C}_{i+n}. \quad (3)$$

From the matrix inversion lemma [38], it readily follows that (1) equivalently expresses as

$$\mathbf{h}_{\text{MVDR}} = \frac{\mathbf{C}_N^{-1} \mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1) \mathbf{C}_N^{-1} \mathbf{a}(\theta_1)}. \quad (4)$$

Moreover, in terms of the total power at the beamformer output, $P(\mathbf{h}) = \mathbb{E}[\|\mathbf{h}^H \mathbf{x}\|^2] = \mathbf{h}^H \mathbf{C}_N \mathbf{h}$, this construction is also seen to solve

$$\min_{\mathbf{h} \in \mathbb{C}^N} P(\mathbf{h}) \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{a}(\theta_1) = 1;$$

that is, it is the beamformer which yields the minimum total output power (subject to unity SoI response):

$$P(\mathbf{h}_{\text{MVDR}}) = \frac{1}{\mathbf{a}^H(\theta_1) \mathbf{C}_N^{-1} \mathbf{a}(\theta_1)}.$$

In the following, for notational convenience, we will consider the normalized total output power, $\rho(\mathbf{h}) = P(\mathbf{h})/\sigma^2$.

B. Sample-based implementation of MVDR beamforming

The advantage of the alternative beamformer specification given in (4) is that it involves $\mathbf{C}_N^{-1}$ (which we refer to as the “precision matrix”), which may be estimated from data, even if SoI-free observations are not available. We focus on this scenario throughout the paper, while simply noting that our results could also be applied for applications with SoI-free observations. Denoting any given estimate by $\hat{\mathbf{C}}_N^{-1}$, the beamformer is constructed as

$$\hat{\mathbf{h}}_{\text{MVDR}} = \frac{\hat{\mathbf{C}}_N^{-1} \mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1} \mathbf{a}(\theta_1)}. \quad (5)$$

The performance of this beamformer, measured by its normalized total output power, is now a function of $\hat{\mathbf{C}}_N^{-1}$ and admits

$$\rho(\hat{\mathbf{h}}_{\text{MVDR}}) = \frac{P(\hat{\mathbf{h}}_{\text{MVDR}})}{\sigma^2} = \frac{1}{\sigma^2} \frac{\mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1} \mathbf{C}_N \hat{\mathbf{C}}_N^{-1} \mathbf{a}(\theta_1)}{\left(\mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1} \mathbf{a}(\theta_1)\right)^2}. \quad (6)$$

This coincides with $\rho_{\min} = \rho(\mathbf{h}_{\text{MVDR}})$ only when $\hat{\mathbf{C}}_N^{-1} = \mathbf{C}_N^{-1}$, otherwise it is larger, which reflects the penalty due to imperfect inverse covariance estimation.

The consequence of estimation errors can be severe, particularly when the number of training samples available for

estimation n is not substantially larger than the signal dimensionality N . The classical estimation approach is simply to employ the SCM,

$$\mathbf{S}_N = \frac{1}{n} \sum_{j=1}^n \mathbf{x}(j) \mathbf{x}^H(j).$$

The beamformer obtained by plugging $\mathbf{S}_N^{-1}$ for $\hat{\mathbf{C}}_N^{-1}$ in (6) is commonly referred to as SMI. It is well-known, however, that the SMI-based beamformer $\hat{\mathbf{h}}_{\text{SMI}} = \mathbf{S}_N^{-1} \mathbf{a}(\theta_1) / (\mathbf{a}^H(\theta_1) \mathbf{S}_N^{-1} \mathbf{a}(\theta_1))$ can yield significantly higher normalized total output power than the theoretical minimum $\rho_{\min}$, especially when N and n have a similar order of magnitude, and for the case of interest with the desired signal being present in the received sample data [11, 17, 39]. It is shown later (see Fig. 1) that $\rho(\hat{\mathbf{h}}_{\text{SMI}})$ and $\rho_{\min}$ can have a large and non-vanishing gap as n and N increase in proportion to one another. In fact, from prior work [40–42], it follows that as $N, n \rightarrow \infty$ with $c_N = N/n \rightarrow c \in (0, 1)$, then $\rho(\hat{\mathbf{h}}_{\text{SMI}})/\rho_{\min} \rightarrow 1/(1-c)$, demonstrating that even if one has twice as many samples as receive sensors (with both sufficiently large), the estimation errors can result in a significant penalty of 3 dB in terms of total output power. Moreover, the penalty can even become unbounded as n approaches N .

In this work, our goal is to design a robust estimation strategy which can overcome this problem by exploiting inherent *structure* of the MVDR beamforming problem. In particular, we will consider the practically relevant scenario for which the number of sensors N and n are reasonably large, while the number of signal plus interferers m is relatively small and is assumed fixed. In this case, the covariance matrix $\mathbf{C}_N$ is identified to be a low-rank (i.e., rank m) perturbation of the identity matrix which, in the language of RMT, conforms to the so-called “spiked” covariance construction. There is indeed a wide and growing recent literature focused on the statistical analysis of spiked covariance models [43–48], and these results may be leveraged to develop novel estimation strategies which significantly overcome the sampling noise issues which fundamentally plague the SCM estimator. To our knowledge, the exploitation of theoretical spike model properties for developing optimized MVDR beamforming solutions has not been considered previously, and this presents a key contribution of our work. In particular, by utilizing the prior knowledge that $\mathbf{C}_N$ has a spiked covariance structure, we will propose an optimized precision matrix estimator $\hat{\mathbf{C}}_{\text{MVDR}}^{-1}$, and consequently an optimized beamformer $\hat{\mathbf{h}}_{\text{MVDRopt}}$, which is designed to minimize (6).

III. OPTIMIZED HIGH-DIMENSIONAL BEAMFORMER DESIGN

A. MVDR beamforming and spiked covariance models

The aim is to estimate the precision matrix $\mathbf{C}_N^{-1}$, with $\mathbf{C}_N$ defined in (3). To this end, we start by noting that the matrix $\mathbf{C}_N$ may be expressed through eigen-decomposition as

$$\mathbf{C}_N = \sigma^2 \left(\mathbf{I}_N + \sum_{i=1}^m t_i \mathbf{v}_i \mathbf{v}_i^H \right), \quad (7)$$

which has eigenvalues $(\sigma^2(t_1+1), \dots, \sigma^2(t_m+1), \sigma^2, \dots, \sigma^2)$, where $t_i > 0$ for $i = 1, \dots, m$, and with $\mathbf{v}_1, \dots, \mathbf{v}_m$ denoting the eigenvectors corresponding to the largest m eigenvalues. The eigenvalues $\sigma^2(t_1+1), \dots, \sigma^2(t_m+1)$ will be referred to as “spike” eigenvalues.

Our proposed estimator will be based on appropriately manipulating the eigenvalue spectrum of the SCM. Specifically, introducing the eigen-decomposition

$$\mathbf{S}_N = \sum_{i=1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^H, \quad (8)$$

with λ_i the i -th largest eigenvalue and $\mathbf{u}_i$ the corresponding eigenvector, we look for estimators of the form

$$\hat{\mathbf{C}}_N^{-1}(\mathbf{S}_N) = \sum_{i=1}^N \eta_i \mathbf{u}_i \mathbf{u}_i^H \quad (9)$$

where $\eta_i > 0$ are appropriate eigenvalue “shrinkage” functions to be designed.

We note that this general construction has been proposed in extensive literature [20, 47, 49–53], and is most relevant when one lacks prior knowledge about the eigenvector structure of $\mathbf{C}_N$ (sparsity, etc.). Moreover, choosing the eigenvectors of $\hat{\mathbf{C}}_N^{-1}$ to coincide with those of the SCM has the desirable property of making it equivariant with respect to rotations of the observed data [52]. Given the spiked model covariance construction, it is natural to apply “hard clipping” to the weakest $N - m$ sample eigenvalues, so that $\eta_{m+1} = \dots = \eta_N = 1/\sigma^2$. This, in effect, accounts for the *eigenvalue spreading* phenomenon [54] caused by finite sampling, in which the non-spike sample eigenvalues (i.e., the $N - m$ smallest) are known to be distributed asymptotically in accordance with the Marcenko-Pastur law, a well-known result in RMT [55]. Introducing the notation $w_i = \sigma^2 \eta_i - 1$, this then leads to the estimator construction

$$\hat{\mathbf{C}}_N^{-1}(\mathbf{S}_N) = \frac{1}{\sigma^2} \left(\mathbf{I}_N + \sum_{i=1}^m w_i \mathbf{u}_i \mathbf{u}_i^H \right). \quad (10)$$

The challenge is to design appropriate values for the shrinkage functions η_i (equivalently w_i), for $i = 1, \dots, m$. Specifically, our aim is to find the optimal $\mathbf{w}^* = [w_1^*, \dots, w_m^*]^T$ which produces a beamformer of the form (5) leading to the minimization of ρ in (6).

By plugging (10) and (7) into (6), the normalized total output power ρ is seen as a function of $\mathbf{w} = [w_1, \dots, w_m]^T$, which we will make explicit as $\rho(\mathbf{w})$ in the following (i.e., we will drop the beamformer dependency, as specified in (6)). In particular, our optimization problem now becomes

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathcal{L}^m}{\operatorname{argmin}} \rho(\mathbf{w}) \quad (11)$$

where $\rho(\mathbf{w})$ is defined in (12) at the top of the next page. The parameter range is specified as¹ $\mathcal{L} = [-1 + \xi, q]$, for some small $\xi > 0$ and large $q > 0$.

The difficulty here is twofold. One is that it is challenging

¹Note that we restrict $\mathbf{w}$ to a bounded set $\mathcal{L}^m$, which is a technical condition that will be employed subsequently for establishing uniform convergence results.

to find the optimal $\mathbf{w}^*$ that minimizes $\rho(\mathbf{w})$ as specified in (12) in closed-form. The other is that even if such an optimal $\mathbf{w}^*$ were obtained, this could not be used in practice, since it involves unobservable quantities (i.e., the t_i and $\mathbf{v}_i$), which are properties of $\mathbf{C}_N$, the object we are trying to estimate. To solve these problems, we will appeal to results on the asymptotic properties of spiked covariance matrices to give a simplified asymptotic representation for (12). As we will show, this may be readily optimized, and the optimal solution may be consistently estimated with observations. We will employ the following assumptions:

Assumption 1.

- As $N, n \rightarrow \infty$, $N/n = c_N \rightarrow c$ for a certain $c > 0$.
- The number of spikes m is fixed, independently of N and n , while $t_1 > \dots > t_m$ with $t_m > \sqrt{c}$ for all large N .

Remark 1. Assumption 1.b requires some explanation. The quantity $\sqrt{c}$ represents a fundamental “phase transition” point; that is, for each $i \in (1, \dots, m)$ such that for large N , $t_i > \sqrt{c}$, there is a deterministic one-to-one mapping between t_i and λ_i which holds under the asymptotic regime specified by Assumption 1.a (see (33) in Appendix C). This is important, since the i -th sample eigenvalue λ_i then informs of the unobservable quantity t_i , which can thus be estimated. For the alternative case $t_i \leq \sqrt{c}$ for large N , the relation no longer holds, and the value of t_i can no longer be estimated. A similar comment also applies for eigenvectors (see (13) below for the case $t_i > \sqrt{c}$ for large N). As a result, part of the information about $\mathbf{C}_N$ cannot be retrieved if some $t_i \leq \sqrt{c}$ for large N .

Under these assumptions, in the following we will introduce a procedure to obtain $\hat{\mathbf{C}}_{\text{MVDR}}^{-1}$, yielding the optimized beamformer, $\hat{\mathbf{h}}_{\text{MVDRopt}}$. The approach is summarized as follows:

- 1) We first characterize an asymptotic deterministic equivalent $\bar{\rho}(\mathbf{w})$ of $\rho(\mathbf{w})$, when $N, n \rightarrow \infty$, $N/n = c_N \rightarrow c > 0$.
- 2) Based on $\bar{\rho}(\mathbf{w})$, we determine the oracle $\bar{\mathbf{w}}^*$ that minimizes $\bar{\rho}(\mathbf{w})$ in terms of $t_1, \dots, t_m$ and the corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_m$.
- 3) Finally, we develop a sample-based consistent estimator $\hat{\mathbf{w}}^*$ of $\bar{\mathbf{w}}^*$, involving the sample eigenvalues $\lambda_1, \dots, \lambda_m$ and eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_m$. This yields the desired $\hat{\mathbf{C}}_{\text{MVDR}}^{-1}$, and consequently $\hat{\mathbf{h}}_{\text{MVDRopt}}$.

B. Deterministic equivalent $\bar{\rho}(\mathbf{w})$ and the optimal $\bar{\mathbf{w}}^*$

To derive the deterministic characterization $\bar{\rho}(\mathbf{w})$, we will use the fact that, under Assumption 1, the projection of each of the leading m sample eigenvectors $\mathbf{u}_j$ onto the leading m population eigenvectors $\mathbf{v}_i$ is consistently estimated by a deterministic function. Specifically, it is known that [55]

$$|\mathbf{v}_i^H \mathbf{u}_j \mathbf{u}_j^H \mathbf{v}_i - s_i \delta_{ij}| \xrightarrow{\text{a.s.}} 0, \quad i, j = 1, \dots, m \quad (13)$$

where $s_i = \frac{1 - c_N/t_i^2}{1 + c_N/t_i}$, with δ_{ij} the kronecker-delta function. Define also the deterministic quantities, $k_i = \mathbf{a}^H(\theta_1) \mathbf{v}_i \mathbf{v}_i^H \mathbf{a}(\theta_1)$, $i = 1, \dots, m$. We have the following result:

$$\rho(\mathbf{w}) = \frac{\mathbf{a}^H(\theta_1) (\mathbf{I}_N + \sum_{i=1}^m w_i \mathbf{u}_i \mathbf{u}_i^H) \left(\mathbf{I}_N + \sum_{j=1}^m t_j \mathbf{v}_j \mathbf{v}_j^H \right) (\mathbf{I}_N + \sum_{h=1}^m w_h \mathbf{u}_h \mathbf{u}_h^H) \mathbf{a}(\theta_1)}{(\mathbf{a}^H(\theta_1) (\mathbf{I}_N + \sum_{l=1}^m w_l \mathbf{u}_l \mathbf{u}_l^H) \mathbf{a}(\theta_1))^2} \quad (12)$$

Theorem 1. [Deterministic equivalent] Let Assumption 1 hold. As $N, n \rightarrow \infty$, $\sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \bar{\rho}(\mathbf{w})| \xrightarrow{\text{a.s.}} 0$ where

$$\bar{\rho}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{B} \mathbf{w} + 2\mathbf{w}^T \mathbf{d} + a}{(1 + \mathbf{w}^T \mathbf{e})^2} \quad (14)$$

with

$$\begin{aligned} \mathbf{B} &= \text{diag}[s_1 k_1 (1 + t_1 s_1), \dots, s_m k_m (1 + t_m s_m)], \\ \mathbf{d} &= [s_1 k_1 (1 + t_1), \dots, s_m k_m (1 + t_m)]^T, \\ \mathbf{e} &= [s_1 k_1, \dots, s_m k_m]^T, \quad a = 1 + \sum_{i=1}^m t_i k_i. \end{aligned}$$

Proof: See Appendix A.

We now seek the value $\bar{\mathbf{w}}^* = [\bar{w}_1^*, \dots, \bar{w}_m^*]^T$ that minimizes $\bar{\rho}(\mathbf{w})$. This is given by the following result:

Theorem 2. [Optimal weights] Under the setting of Theorem 1,

$$\bar{\mathbf{w}}^* = \underset{\mathbf{w} \in \mathcal{L}^m}{\text{argmin}} \bar{\rho}(\mathbf{w})$$

where, for $i = 1, \dots, m$,

$$\bar{w}_i^* = \frac{t_i + c_N}{t_i^2 + t_i} (\psi - t_i), \quad \text{with } \psi = \frac{\sum_{j=1}^m \frac{k_j}{t_j}}{\sum_{j=1}^m \frac{k_j}{t_j^2}}. \quad (15)$$

Proof: See Appendix B.

Remark 2. For the special case $m = 1$ (i.e., a SoI with no interferers), $\bar{w}_1^* = 0$, giving $\hat{\mathbf{C}}_{\text{MVDR}}^{-1} = \frac{1}{\sigma^2} \mathbf{I}_N$. Hence, $\hat{\mathbf{h}}_{\text{MVDRopt}} = \mathbf{a}(\theta_1)$, which represents the classical matched filter. This is in fact the optimal beamformer, as seen from (1) and the fact that $\mathbf{C}_{i+n} = \sigma^2 \mathbf{I}_N$.

C. Estimated optimal weights $\hat{\mathbf{w}}^*$ and proposed algorithm

The optimal weights specified in the theorem above involve unobservable quantities, i.e., t_i and $\mathbf{v}_i$, and are therefore not directly useful in practice. To address this issue, in the following we provide consistent estimators of these optimal weights. These are based on the sample eigenvalues λ_i and sample eigenvectors $\mathbf{u}_i$, for $i = 1, \dots, m$, and hence are computable from sample observations.

Theorem 3. [Estimated optimal weights] Under the setting of Theorem 1, for all large n with probability one, $\lambda_i > \sigma^2(1 + \sqrt{c_N})^2$, $i = 1, \dots, m$, and we have

$$|\hat{w}_i^* - \bar{w}_i^*| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{w}_i^* = \frac{\hat{t}_i + c_N}{\hat{t}_i^2 + \hat{t}_i} (\hat{\psi} - \hat{t}_i), \quad (16)$$

Algorithm 1 Proposed MVDR beamformer construction

- 1) Compute the optimized shrinkage parameters $\hat{w}_i^*$, $i = 1, \dots, m$ in accordance with (16).
- 2) Form the precision matrix estimator (leading scale factor can be omitted):

$$\hat{\mathbf{C}}_{\text{MVDR}}^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I}_N + \sum_{i=1}^m \hat{w}_i^* \mathbf{u}_i \mathbf{u}_i^H \right).$$

- 3) Construct the MVDR beamformer:

$$\hat{\mathbf{h}}_{\text{MVDRopt}} = \frac{\hat{\mathbf{C}}_{\text{MVDR}}^{-1} \mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_{\text{MVDR}}^{-1} \mathbf{a}(\theta_1)}.$$

in which $\hat{\psi} = \frac{\sum_{j=1}^m \frac{k_j}{\hat{t}_j}}{\sum_{j=1}^m \frac{k_j}{\hat{t}_j^2}},$

$$\hat{t}_i = \frac{\lambda_i/\sigma^2 + 1 - c_N + \sqrt{(\lambda_i/\sigma^2 + 1 - c_N)^2 - 4\lambda_i/\sigma^2}}{2} - 1,$$

$$\hat{k}_i = \frac{1 + c_N/\hat{t}_i}{1 - c_N/(\hat{t}_i)^2} \mathbf{a}^H(\theta_1) \mathbf{u}_i \mathbf{u}_i^H \mathbf{a}(\theta_1).$$

Proof: See Appendix C.

This leads immediately to the specification of our MVDR beamformer, summarized in Algorithm 1. A pertinent point is the simplicity of the proposed approach; the shrinkage functions are in closed form, and are just simple functions of the eigenvalues and eigenvectors of the SCM. Despite this simplicity, we will demonstrate in the following that the proposed beamformer exhibits excellent performance compared with classical solutions (e.g., the SMI beamformer) as well as more advanced techniques.

D. Scenario with some t_i satisfying $0 < t_i < \sqrt{c}$

The previous results are based on the assumption that for large N , $t_i > \sqrt{c}$ for all $i = 1, \dots, m$. In practice, we would typically not know if some of the spikes are smaller than this threshold (i.e., if $t_i < \sqrt{c}$ for some i). In this subsection, we explore this situation and investigate how our algorithm will be applied under such scenarios. Specifically, consider $m_2 < m$, such that for all large N ,

$$t_1 > t_2 > \dots > t_{m-m_2} > \sqrt{c} > t_{m-m_2+1} > \dots > t_m > 0.$$

In this case, under Assumption 1.a, $\mathbf{u}_i$ no longer asymptotically carries information about $\mathbf{v}_i$, while λ_i is no longer informative of t_i , for $i = m-m_2+1, \dots, m$. A key distinction is that now [55, Corollary 9.2]

$$\mathbf{v}_i^H \mathbf{u}_i \mathbf{u}_i^H \mathbf{v}_i \xrightarrow{\text{a.s.}} 0 \quad (17)$$

rather than obeying (13), while

$$\lambda_i/\sigma^2 \xrightarrow{\text{a.s.}} (1 + \sqrt{c})^2 \quad (18)$$

$$\bar{\rho} = \frac{\mathbf{w}^{(m_2)T} \mathbf{B}^{(m_2)} \mathbf{w}^{(m_2)} + 2\mathbf{w}^{(m_2)T} \mathbf{d}^{(m_2)} + a^{(m_2)}}{(1 + \mathbf{w}^{(m_2)T} \mathbf{e}^{(m_2)})^2} + \frac{\sum_{i=m-m_2+1}^m t_i k_i}{(1 + \mathbf{w}^{(m_2)T} \mathbf{e}^{(m_2)})^2} \quad (19)$$

rather than obeying (33).

Writing the population covariance matrix as

$$\mathbf{C}_N = \sigma^2 \left(\mathbf{I}_N + \sum_{i=1}^{m-m_2} t_i \mathbf{v}_i \mathbf{v}_i^H + \sum_{j=m-m_2+1}^m t_j \mathbf{v}_j \mathbf{v}_j^H \right)$$

and the corresponding precision matrix estimator as

$$\hat{\mathbf{C}}_N^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I}_N + \sum_{i=1}^{m-m_2} w_i \mathbf{u}_i \mathbf{u}_i^H + \sum_{j=m-m_2+1}^m w_j \mathbf{u}_j \mathbf{u}_j^H \right),$$

mirroring the same steps as in the proof of Theorem 1 (while also using (17) and (18)), the deterministic equivalent of $\rho(\mathbf{w})$ under the same setting as Theorem 1 becomes (19) at the top of the page, where all terms with the superscript $\cdot^{(m_2)}$ coincide with those in Theorem 1, but with m replaced by $m - m_2$.

There are two important points to note regarding this result:

- 1) The deterministic equivalent (19) does not contain w_i ($i = m - m_2 + 1, \dots, m$); hence, these weights can be chosen arbitrarily. They can be set to zero without loss of generality.
- 2) The first term in (19) coincides exactly with the deterministic equivalent obtained when $t_{m-m_2+1} = \dots = t_m = 0$. The effect of these being non-zero (but below the transition point $\sqrt{c}$) is captured by the second term in (19). However, the values t_i and k_i , $i = m - m_2 + 1, \dots, m$ involved in this expression cannot be typically estimated from data, unlike for the case $i = 1, \dots, m - m_2$, and therefore the exact weights $w_1, \dots, w_{m-m_2}$ minimizing (19), which will involve these quantities, cannot be consistently estimated in practice.

As a consequence of the above arguments, our algorithm will simply act as if the number of signal plus interferers is the number of “strong” spike eigenvalues $m - m_2$. And so, it will provide a consistent estimator of the minimizer of the first right-hand side term in (19), which is given through Algorithm 1 but with m replaced by $m - m_2$. We point out, however, that this is a suboptimal solution, since it (necessarily) ignores the second right-hand side term in (19).

IV. NUMERICAL SIMULATIONS

For our simulations, we assume a uniform linear array with N identical omnidirectional sensors placed at half-wavelength spacing. For each data point, we average over 200 independent Monte-Carlo trials. Incident upon the sensor array are $m = 6$ uncorrelated narrow-band signals from the far field. Unless otherwise specified, the SoI impinges from $\theta_1 = 0^\circ$, while five interferers impinge from $\theta_2 = 5^\circ$, $\theta_3 = 10^\circ$, $\theta_4 = 30^\circ$, $\theta_5 = 50^\circ$ and $\theta_6 = 70^\circ$. The steering vector is $\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} [1, e^{j\pi \sin(\theta)}, \dots, e^{j\pi \sin(\theta)(N-1)}]^T$. The noise is complex Gaussian with mean zero and variance one.

We note that unless otherwise specified, the number of spikes m is assumed known. For simplicity, we also assume

that the noise variance σ^2 is known, while simply noting that the incorporation of an existing estimation approach [36] into our experiments revealed qualitatively similar results.

A. ρ -Performance and the deterministic equivalent

We first investigate convergence of our proposed algorithm in terms of the function ρ . Define $\text{SNR} = \frac{p_1}{\sigma^2}$ and $\text{INR} = \frac{p_i}{\sigma^2}$ (taken to be the same for all $i = 2, \dots, m$). In Fig. 1, for $\text{SNR} = 5$ dB and $\text{INR} = 30$ dB and different values of $n = 2N$ and $n = 10N$, we compare the expectation $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$ (computed empirically) with our proposed $\hat{\mathbf{w}}^*$ in Theorem 3 against the asymptotic deterministic equivalent $\bar{\rho}(\bar{\mathbf{w}}^*)$ based on Theorem 1. Note that $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$ converges to $\bar{\rho}(\bar{\mathbf{w}}^*)$ with the increase of N and n , as expected. For additional comparison, we also plot the theoretical minimum (oracle) $\rho_{\min} = 1/(\sigma^2 \mathbf{a}^H(\theta_1) \mathbf{C}_N^{-1} \mathbf{a}(\theta_1))$ and the expectation $\mathbb{E}[\rho(\hat{\mathbf{h}}_{\text{SMI}})]$ with the SMI. These results demonstrate that $\bar{\rho}(\bar{\mathbf{w}}^*)$ is close to $\rho_{\min}$, indicating that our proposed approach leads to near-optimal performance. Comparison with $\mathbb{E}[\rho(\hat{\mathbf{h}}_{\text{SMI}})]$, on the other hand, demonstrates that the total output power is about $\frac{1}{1-c_N}$ times as large as that achieved with our algorithm (i.e., $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$), consistent with the discussion in Section II-B.

B. Beamformer performance and comparison against previous methods

Next we compare the performance of our proposed method $\hat{\mathbf{h}}_{\text{MVDRopt}}$ against alternative beamforming methods, considering both the beampatterns and the output SINR (2). Note that the beampattern reflects the response of the beamformer to signals originating from DoA α , given by $\mathbb{E} \left[20 \log_{10} |\hat{\mathbf{h}}^H(\theta_1) \mathbf{a}(\alpha)| \right]$, which is computed by averaging over 200 simulation runs, with α swept from -90° to 90° . For these comparisons, we fix $\text{INR} = 30$ dB. We first compare the performance of $\hat{\mathbf{h}}_{\text{MVDRopt}}$ against the optimal MVDR beamformer (4) with perfect knowledge of $\mathbf{C}_N$, considering $n = 2N = 200$. Results are shown in Fig. 2(a). Despite the fact that SNR is low (i.e., 5 dB), as for the optimal solution, our proposed beamformer correctly places nulls in the directions of each of the interferers. We also compare with two traditional beamformers: $\hat{\mathbf{h}}_{\text{SMI}}$, the SMI beamformer, and $\hat{\mathbf{h}}_{\text{I}}$, the phased array beamformer (or matched filter), which is the steering vector $\mathbf{a}(\theta_1)$. These results are shown in Fig. 2(b) and Fig. 2(c) respectively. Clearly, our method $\hat{\mathbf{h}}_{\text{MVDRopt}}$ has a significant smaller “noise gain” compared with $\hat{\mathbf{h}}_{\text{SMI}}$, in addition to placing more accurate interference nulls. The phased array beamformer, $\hat{\mathbf{h}}_{\text{I}}$, demonstrates no interference rejection capability, as expected. The SINR performance of the different beamformers is shown in Fig. 2(d), for different SNRs. The performance advantage of the proposed beamformer $\hat{\mathbf{h}}_{\text{MVDRopt}}$ over the classical strategies is again evident. For further comparison, we also plot the SINR performance of

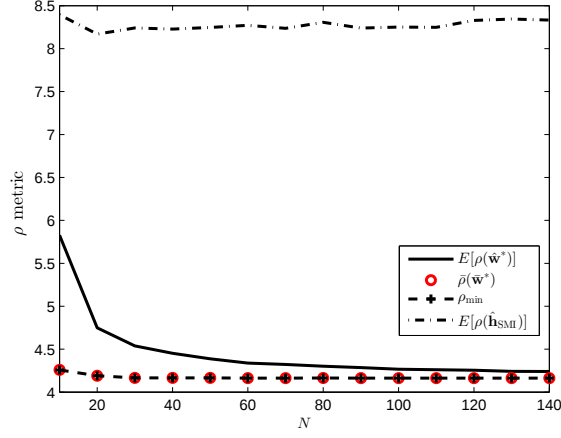
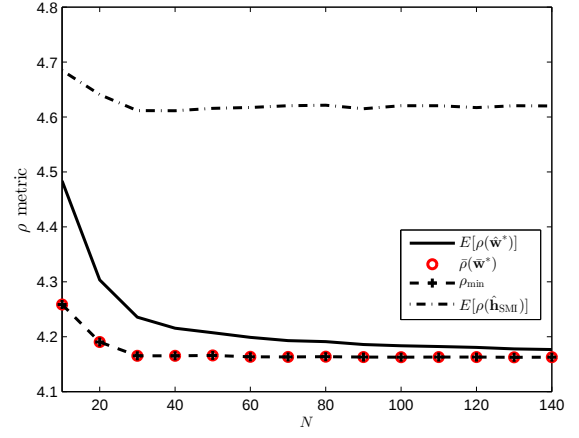
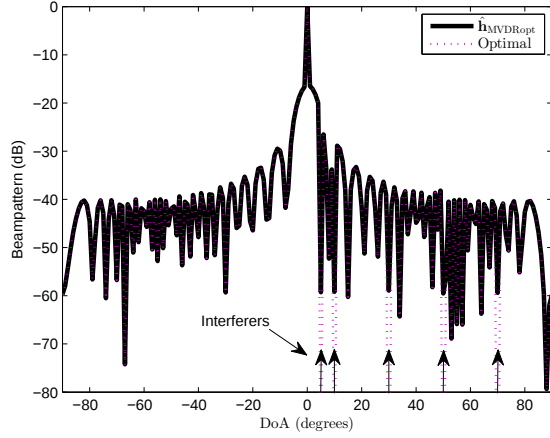
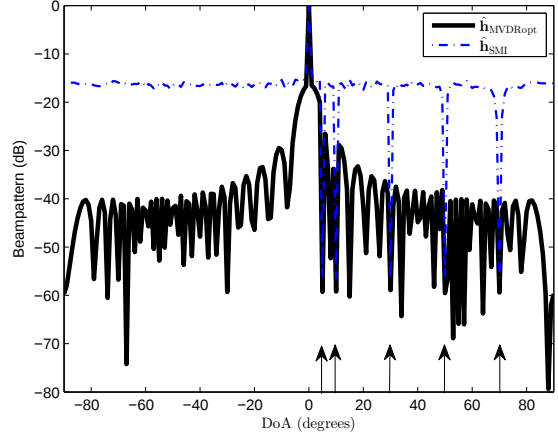
(a) ρ -performance when $n = 2N$ (b) ρ -performance when $n = 10N$

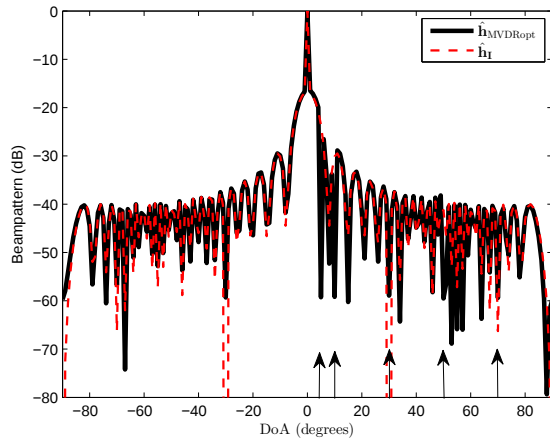
Fig. 1. Convergence of $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$ to its deterministic equivalent $\bar{\rho}(\hat{\mathbf{w}}^*)$ and the oracle $\rho_{\min}$ for different values of $n = 2N$ and $n = 10N$. The expectation $\mathbb{E}[\rho(\hat{\mathbf{h}}_{\text{SMI}})]$ is also shown, which is about $\frac{1}{1-c_N}$ times as large as $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$. The expectations $\mathbb{E}[\rho(\hat{\mathbf{w}}^*)]$ and $\mathbb{E}[\rho(\hat{\mathbf{h}}_{\text{SMI}})]$ are estimated by averaging over 200 realizations.



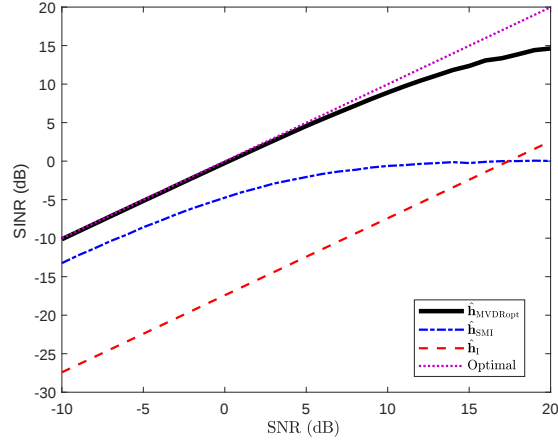
(a) Beampatterns (SNR = 5 dB). The arrows show the directions of the interferers.



(b) Beampatterns (SNR = 5 dB)



(c) Beampatterns (SNR = 5 dB)

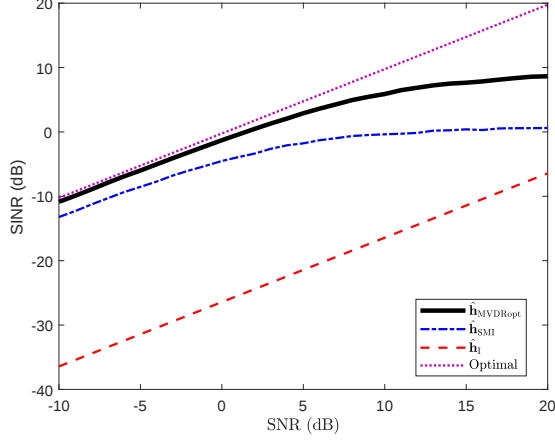
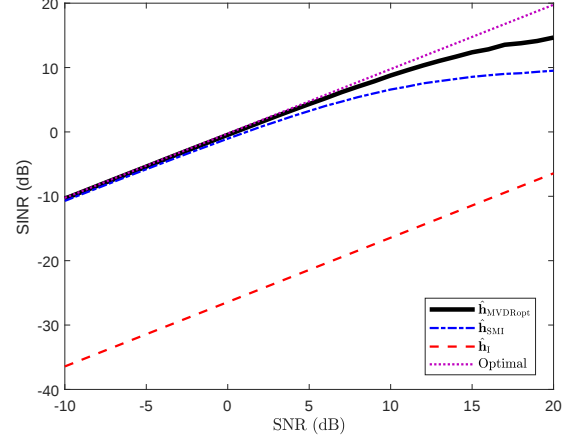


(d) Output SINR

Fig. 2. Beampattern and SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, $\hat{\mathbf{h}}_{\text{SMI}}$ and $\hat{\mathbf{h}}_{\text{I}}$ ($N = 100, n = 200$).

each beamformer with different choices of n and N in Fig. 3, yielding qualitatively similar results.

In addition to the traditional SMI and phased array beamformers, we also compared with more robust strategies, which

(a) Output SINR ($N = 20, n = 40$)(b) Output SINR ($N = 20, n = 200$)Fig. 3. SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, $\hat{\mathbf{h}}_{\text{SMI}}$ and $\hat{\mathbf{h}}_{\text{I}}$.

are designed to at least partially overcome the problem of sample insufficiency which plagues accurate inverse covariance matrix estimation, and is the cause of the poor performance of the SMI. First, we considered the popular diagonal loading strategy $\hat{\mathbf{h}}_{\text{DL}}$ (see, e.g., [15–18, 20]), which employs the construction (5) with the covariance matrix estimator

$$\hat{\mathbf{C}}_N(\varphi) = (1 - \varphi)\mathbf{S}_N + \varphi\mathbf{I}_N$$

where $\varphi \in (0, 1)$ is a parameter to be specified. While numerous methods have been developed to specify φ (e.g., [17, 20]), here we compare with the empirically computed “oracle” solution, in which φ is chosen to minimize the normalized total output power in (6). Note that this optimal solution is typically unknown in practice, but it provides an upper bound on the performance achievable with any diagonal loading method. The beam pattern of this oracle beamformer is shown in Fig. 4(a) for SNR = 5 dB. Strikingly, despite the fact that the optimal diagonal loading parameter is employed, it is seen that $\hat{\mathbf{h}}_{\text{MVDRopt}}$ has a significantly improved response, particularly in terms of the noise suppression away from the DoA of interest (i.e., away from 0°). We also compared with the eigen-subspace beamformer $\hat{\mathbf{h}}_{\text{Eigsub}}$ (see [8, 14]), which constructs a beamformer based on first projecting $\mathbf{a}(\theta_1)$ onto the estimated signal-plus-interference subspace (where it is supposed to lie), giving

$$\mathbf{a}_{\text{Sub}} = \sum_{i=1}^m \mathbf{u}_i \mathbf{u}_i^H \mathbf{a}(\theta_1)$$

and then specifying

$$\hat{\mathbf{h}}_{\text{Eigsub}} = \frac{\mathbf{S}_N^{-1} \mathbf{a}_{\text{Sub}}}{\mathbf{a}_{\text{Sub}}^H \mathbf{S}_N^{-1} \mathbf{a}_{\text{Sub}}} = \frac{(\sum_{i=1}^m 1/\lambda_i \mathbf{u}_i \mathbf{u}_i^H) \mathbf{a}(\theta_1)}{\mathbf{a}^H(\theta_1) (\sum_{i=1}^m 1/\lambda_i \mathbf{u}_i \mathbf{u}_i^H) \mathbf{a}(\theta_1)}.$$

Note that this scheme can be seen as a special case of the construction (9), for which $\eta_i = 0$ for $i = m+1, \dots, N$, while $\eta_i = 1/\lambda_i$ for $i = 1, \dots, m$. That is, it completely removes all noise dimensions, and performs no eigenvalue shrinkage to the sample eigenvalues. The performance of this eigen-subspace scheme is shown in Fig. 4(a), where it is seen to produce a

similar response to the diagonal loading method, albeit with slightly improved noise suppression.

Fig. 4(b) compares the output SINR achieved by the different methods for different SNRs. As evident, $\hat{\mathbf{h}}_{\text{MVDRopt}}$ uniformly displays the highest SINR, performing comparable to the diagonal loading method at low SNR, and with the eigen-subspace method at high SNR. Similar conclusions are also reached for different choices of n and N , such as $n = 2N = 40$ and $n = 10N = 200$ (results not shown due to space limitations).

When the signals are more closely-spaced (e.g., with angular separation smaller than $360^\circ/N$), similar performance comparisons are also obtained, as demonstrated in Fig. 5. Here, two scenarios are considered: one in which the DoAs are separated by 3° ; another in which they are separated by 1° . In both cases, our proposed solution performs very closely to the optimal beamformer (which we recall is the MVDR beamformer constructed based on perfect knowledge of $\mathbf{C}_N$), and offers significantly enhanced noise resilience with respect to the diagonal loading and eigen-subspace methods.

The results in Fig. 5, together with those in Fig. 4, demonstrate the uniformly superior performance of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, regardless of whether the array sensors are widely-spaced or close to one another (or some combination thereof). Interestingly, this performance difference is quite distinct to that which has been observed previously for RMT-based Multiple Signal Classification (MUSIC) algorithms as applied to the (albeit different) problem of DoA estimation [56–62]. Specifically, for that application, it was shown in [57] that the benefits of RMT-optimized approaches over classical approaches (e.g., conventional MUSIC or spatial periodograms based on the SCM) are typically only revealed for large- N under “closely-spaced” scenarios, for which the DoA separations are $O(1/N)$.

C. Robustness to under or overestimation of the number of signals, m

Our proposed beamformer $\hat{\mathbf{h}}_{\text{MVDRopt}}$, as well as $\hat{\mathbf{h}}_{\text{Eigsub}}$, requires knowledge of the number of signals m . In practice,

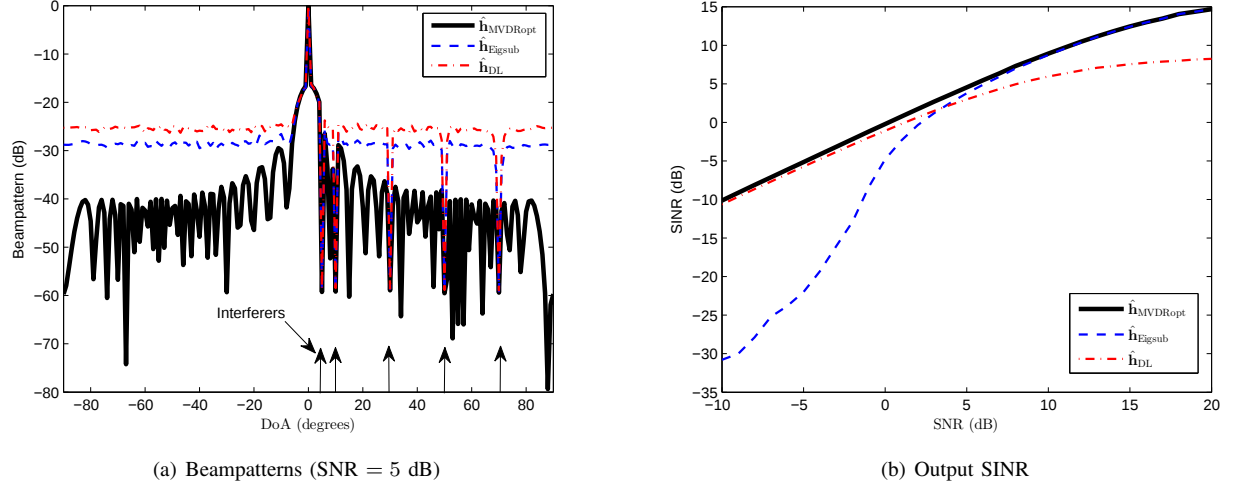


Fig. 4. Beampattern and SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, $\hat{\mathbf{h}}_{\text{Eigsub}}$ and $\hat{\mathbf{h}}_{\text{DL}}$ ($N = 100, n = 200$).

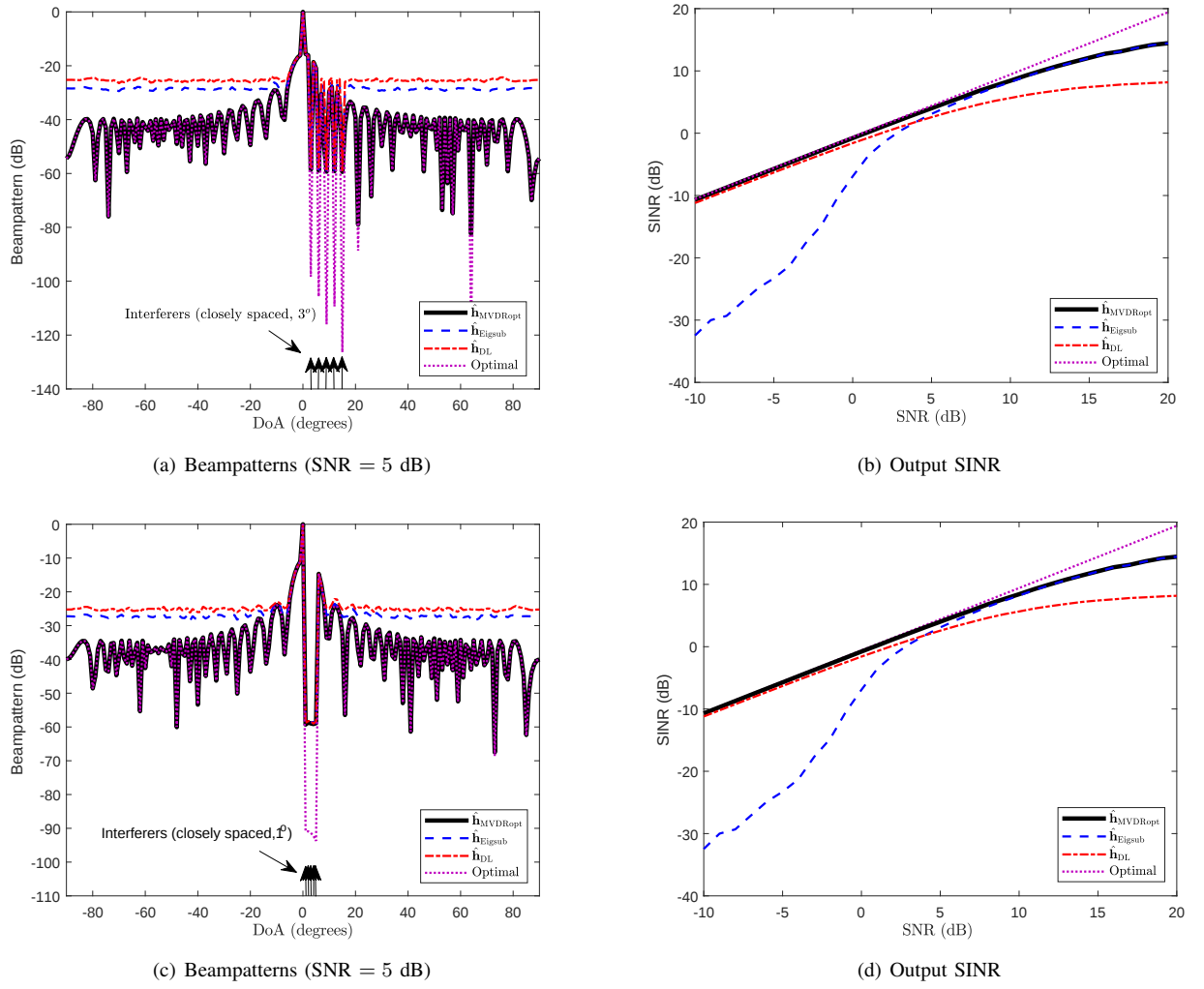


Fig. 5. Beampattern and SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, $\hat{\mathbf{h}}_{\text{Eigsub}}$ and $\hat{\mathbf{h}}_{\text{DL}}$ ($N = 100, n = 200$). In (a) and (b), the DoA spacing is 3° . In (c) and (d) the DoA spacing is 1° .

this often needs to be estimated, and inaccurate estimation may lead to performance degradation. To assess this, we first quantified the performance loss of each beamformer when m is over or underestimated by a fixed amount. The results are shown in Fig. 6(a) under a similar parameterization to Section IV-B. The most noteworthy point is that our proposed method is robust to both under and overestimation of m . For the eigen-subspace beamformer, on the other hand, while it also shows robustness to overestimation, its performance degrades severely when m is underestimated. This may be an important issue in practice, when underestimation may be difficult to avoid.

Beyond studying the effect of systematic under or overestimation, we also investigated the beamformer performance when a data-based method is used to estimate m . Numerous algorithms have been classically proposed for this purpose (e.g., [63–66]), though they were often developed for scenarios with $n \gg N$. Recent contributions have proposed algorithms for high-dimensional settings (i.e., with n not substantially exceeding N), which include [34, 36, 67]. In this line, a popular and very simple method is to estimate m as the number of eigenvalues of the SCM that exceed the right-hand edge of the Marcenko-Pastur law, $\sigma^2(1 + \sqrt{c})^2$. This method, which is found to perform well through simulations, is theoretically justified based on results in [43, 44].

Applying this method to estimate m , Fig. 6(b) compares the performance of $\hat{\mathbf{h}}_{\text{MVDRopt}}$ and $\hat{\mathbf{h}}_{\text{Eigsub}}$. Importantly, for our proposed MVDR beamformer, the performance with m estimation is nearly identical to that with m known perfectly. The performance of the eigen-subspace method, on the other hand, degrades substantially, which is due to the mis-classification (and particularly underestimation) as indicated previously.

D. Comparison with MVDR constructions based on other spike-model covariance estimators

Optimal spike-model-based covariance (and inverse covariance) estimation strategies have been proposed recently in [32]. These methods share the same general structure (10), but with the weights $\{w_i\}$ designed based on different loss functions. Nonetheless, in all cases, the optimal weights w_i ($i = 1, \dots, m$) were found to depend in a one-to-one manner on specific sample eigenvalue and eigenvector pairs, λ_i and $\mathbf{u}_i$ ($i = 1, \dots, m$), which differs from the optimal MVDR weights $\hat{w}_i^*$, specified in (16), which are *jointly* dependent on λ_i and $\mathbf{u}_i$ for all $i = 1, \dots, m$. We compare with beamformers constructed from two optimized covariance estimation solutions from [32]: (i) $\hat{\mathbf{h}}_{\text{Op}}$, the MVDR beamformer constructed from inverting a covariance matrix estimator with weights designed to minimize (asymptotically) the operator loss $\|\hat{\mathbf{C}}_{\text{Op}} - \mathbf{C}_N\|_{\text{Op}}$, and (ii) $\hat{\mathbf{h}}_{\text{Froinv}}$, the MVDR beamformer constructed from the inverse covariance matrix estimator with weights designed to minimize (again asymptotically) the Frobenius loss $\|\hat{\mathbf{C}}_{\text{Froinv}}^{-1} - \mathbf{C}_N^{-1}\|_{\text{F}}$. As a simple baseline method, we also consider the “eigenvalue clipping” strategy, denoted $\hat{\mathbf{h}}_{\text{Clip}}$, which corresponds to the beamformer obtained by setting $w_i = \sigma^2/\lambda_i - 1$ in (10). This represents the naive estimator obtained with no optimization of the weights $\{w_i\}$. We consider the same parameter setting as in Section IV-B.

Beampatterns of these three beamformers are compared with our MVDR approach in Fig. 7(a) for the case $n = 2N = 200$. While each strategy succeeds in placing nulls at the correct interference directions, our proposed beamformer achieves significantly enhanced noise suppression. These results demonstrate the importance of not only employing the general structure in (10), but in designing the weights $\{w_i\}$ which are suitably optimized for the MVDR problem. This gain is also reflected in the SINR performance (whenever SNR is not too small), as shown in Fig. 7(b). Similar conclusions are reached for other choices of n and N also, e.g. $n = 2N = 40$ and $n = 10N = 200$, which are not shown due to space limitations.

V. CONCLUSION

We have proposed a new MVDR beamformer which is optimally designed for high-dimensional applications in which the number of samples does not greatly exceed the number of array elements. Our technique is based on constructing an optimized estimator of the inverse received signal covariance matrix, which employs eigenvalue clipping as well as carefully designed shrinkage functions. Leveraging results from RMT and spiked models in particular, our solution exploits the structure of a relatively low dimensional signal subspace buried in a high dimensional noise space in the MVDR design problem.

There are various avenues to be explored for future work. For example, it would be interesting to conduct a second-order analysis of our proposed MVDR beamformer (i.e., deriving a central limit theorem, as in [19, 26]), which would help to provide insights into the fluctuations of the output SINR. In addition, our discussion has assumed that the DoA of the desired signal component is known. While in some applications this is reasonable, in others it may need to be estimated, and estimation errors may lead to a performance degradation. An important challenge, falling beyond the scope of the current work, is to quantify the effect of such errors on beamformer performance, and to extend our results to produce DoA-robust methods. This may involve, for example, following similar lines of reasoning to the contributions [39, 68–70]. We also remark that the MVDR beamforming problem is closely related to that of designing minimum-variance portfolios in mathematical finance, and our approach could naturally be applied to that problem. Further motivating this direction, it is noteworthy that an empirical analysis of sample correlation matrices of financial stock return data has demonstrated spectral properties which are closely reminiscent of those expected for spiked models [71].

APPENDIX A PROOF OF THEOREM 1

For a specific $\mathbf{w} \in \mathcal{L}^m$, the result $|\rho(\mathbf{w}) - \bar{\rho}(\mathbf{w})| \xrightarrow{\text{a.s.}} 0$ as $N, n \rightarrow \infty$ and $c_N = N/n \rightarrow c > 0$ follows immediately from (12) and the following properties [55]:

$$\begin{aligned} |\mathbf{a}^H(\theta_1)\mathbf{u}_i\mathbf{u}_i^H\mathbf{a}(\theta_1) - s_i k_i| &\xrightarrow{\text{a.s.}} 0 \\ |\mathbf{a}^H(\theta_1)\mathbf{u}_i\mathbf{u}_i^H\mathbf{v}_i\mathbf{v}_i^H\mathbf{a}(\theta_1) - s_i k_i| &\xrightarrow{\text{a.s.}} 0 \end{aligned} \quad (20)$$

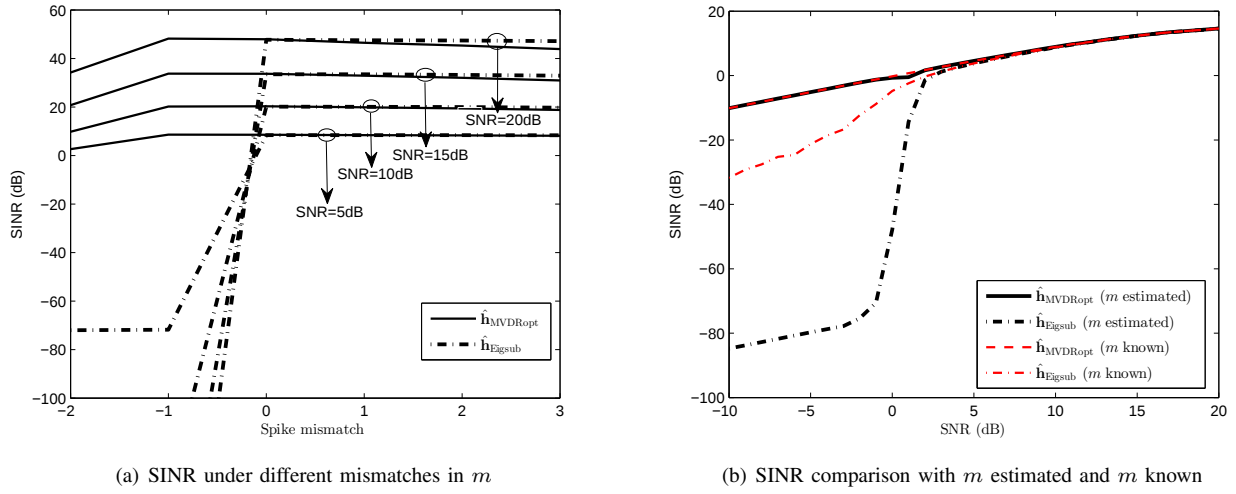


Fig. 6. SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$ and $\hat{\mathbf{h}}_{\text{Eigsub}}$ in two scenarios ($N = 100, n = 200$). (a) m is over or underestimated by a fixed amount. (b) m is estimated empirically with a sample-based algorithm.

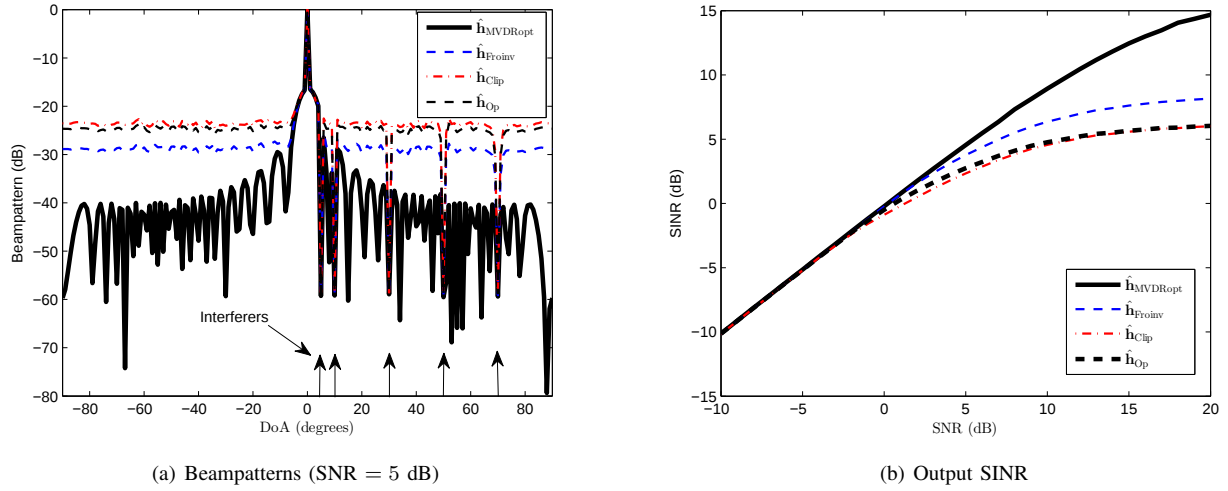


Fig. 7. Beampattern and SINR performance comparison of $\hat{\mathbf{h}}_{\text{MVDRopt}}$, $\hat{\mathbf{h}}_{\text{Clip}}$, $\hat{\mathbf{h}}_{\text{Op}}$ and $\hat{\mathbf{h}}_{\text{Froinv}}$ ($N = 100, n = 200$).

$$|\mathbf{a}^H(\theta_1) \mathbf{u}_i \mathbf{u}_i^H \mathbf{v}_i \mathbf{v}_i^H \mathbf{u}_i \mathbf{u}_i^H \mathbf{a}(\theta_1) - s_i^2 k_i| \xrightarrow{\text{a.s.}} 0.$$

The main difficulty is to establish the *uniform* convergence

$$\sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \bar{\rho}(\mathbf{w})| \xrightarrow{\text{a.s.}} 0. \quad (21)$$

From the definition of uniform convergence, this amounts to showing that for some $K > 0$ and any given $\delta > 0$,

$$\sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \bar{\rho}(\mathbf{w})| < K\delta \quad (22)$$

for all large n almost surely. We start by noting that since $\mathcal{L}$ is bounded, for any given $\delta > 0$, we can always create a lattice of $\mathbf{w}_1, \dots, \mathbf{w}_J \in \mathcal{L}^m$ with J finite such that, for each $\mathbf{w} = [w_1, \dots, w_m] \in \mathcal{L}^m$, there exists $\mathbf{w}' = [w'_1, \dots, w'_m] \in (\mathbf{w}_1, \dots, \mathbf{w}_J)$ for which $\max_{i \in (1, \dots, m)} |w_i - w'_i| < \delta$. With this, we may write:

$$\begin{aligned} & \sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \bar{\rho}(\mathbf{w})| \\ & \leq \sup_{\mathbf{w} \in \mathcal{L}^m} \{|\rho(\mathbf{w}) - \rho(\mathbf{w}')| + |\bar{\rho}(\mathbf{w}') - \bar{\rho}(\mathbf{w})| + |\rho(\mathbf{w}') - \bar{\rho}(\mathbf{w}')|\} \end{aligned}$$

$$\begin{aligned} & \leq \sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \rho(\mathbf{w}')| + \sup_{\mathbf{w} \in \mathcal{L}^m} |\bar{\rho}(\mathbf{w}') - \bar{\rho}(\mathbf{w})| \\ & \quad + \max_{\mathbf{w}'' \in (\mathbf{w}_1, \dots, \mathbf{w}_J)} |\rho(\mathbf{w}'') - \bar{\rho}(\mathbf{w}'')|. \end{aligned} \quad (23)$$

Hence, it follows that the relation (22) would be established upon proving that, for certain $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$, we have $\sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \rho(\mathbf{w}')| < K_1\delta$, $\sup_{\mathbf{w} \in \mathcal{L}^m} |\bar{\rho}(\mathbf{w}') - \bar{\rho}(\mathbf{w})| < K_2\delta$, and $\max_{\mathbf{w}'' \in (\mathbf{w}_1, \dots, \mathbf{w}_J)} |\rho(\mathbf{w}'') - \bar{\rho}(\mathbf{w}'')| < K_3\delta$ for all large n almost surely.

To establish the first bound, we start by using (12) to write

$$\begin{aligned} & |\rho(\mathbf{w}) - \rho(\mathbf{w}')| \\ & = \left| \frac{D(\mathbf{w}') [N(\mathbf{w}) - N(\mathbf{w}')] + [D(\mathbf{w}') - D(\mathbf{w})] N(\mathbf{w}')}{D(\mathbf{w}) D(\mathbf{w}')} \right| \end{aligned} \quad (24)$$

with the correspondences

$$N(\mathbf{w}) = \sigma^2 \mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) \mathbf{C}_N \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) \mathbf{a}(\theta_1), \quad (25)$$

$$D(\mathbf{w}) = \left(\sigma^2 \mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) \mathbf{a}(\theta_1) \right)^2, \quad (26)$$

where we now explicitly show the dependence of $\hat{\mathbf{C}}_N^{-1}$ on $\mathbf{w}$.

Next, we bound the difference terms on the numerator. Using the definition of the spectral norm, we trivially obtain

$$|N(\mathbf{w}) - N(\mathbf{w}')| \leq \sigma^2 \left\| \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) \mathbf{C}_N \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) - \hat{\mathbf{C}}_N^{-1}(\mathbf{w}') \mathbf{C}_N \hat{\mathbf{C}}_N^{-1}(\mathbf{w}') \right\|.$$

Next, plugging in (10) for $\hat{\mathbf{C}}_N^{-1}$, applying simple algebra, and invoking the following spectral norm properties [72]

$$\|\mathbf{U}\mathbf{V}\| \leq \|\mathbf{U}\| \|\mathbf{V}\|, \quad \|\mathbf{U} + \mathbf{V}\| \leq \|\mathbf{U}\| + \|\mathbf{V}\| \quad (27)$$

for $\mathbf{U}$ and $\mathbf{V}$ two arbitrary square complex matrices, this is further bounded as

$$\begin{aligned} |N(\mathbf{w}) - N(\mathbf{w}')| &\leq 2 \left\| \frac{\mathbf{C}_N}{\sigma^2} \right\| \left\| \sum_{i=1}^m (w_i - w'_i) \mathbf{u}_i \mathbf{u}_i^H \right\| \\ &\quad + \left\| \sum_{i=1}^m (w_i^2 - w'^2_i) \mathbf{u}_i \mathbf{u}_i^H \frac{\mathbf{C}_N}{\sigma^2} \mathbf{u}_i \mathbf{u}_i^H \right\| \\ &\leq 2 \left\| \frac{\mathbf{C}_N}{\sigma^2} \right\| \max_{i \in (1, \dots, m)} |w_i - w'_i| \\ &\quad + \left\| \frac{\mathbf{C}_N}{\sigma^2} \right\| \max_{j \in (1, \dots, m)} |w_j + w'_j| \max_{h \in (1, \dots, m)} |w_h - w'_h| \\ &< g_1 \delta \end{aligned}$$

where $g_1 = 2 \left\| \frac{\mathbf{C}_N}{\sigma^2} \right\| + 2q \left\| \frac{\mathbf{C}_N}{\sigma^2} \right\| = 2(t_1 + 1)(q + 1)$. (This last line is established by recalling that $w_i \in \mathcal{L} = [-1 + \xi, q]$, for small $\xi > 0$ and large $q > 0$.) The second difference term in (24) can be bounded similarly, leading to

$$|D(\mathbf{w}) - D(\mathbf{w}')| < g_2 \delta$$

where $g_2 = 2q$. Hence, combining these results leads to the following bound for (24):

$$|\rho(\mathbf{w}) - \rho(\mathbf{w}')| < \varepsilon_1 \delta,$$

where

$$\varepsilon_1 = \frac{D(\mathbf{w}')g_1 + N(\mathbf{w}')g_2}{D(\mathbf{w})D(\mathbf{w}')}.$$

We now need to determine that ε_1 is bounded. To this end, regarding $D(\mathbf{w})$, given in (26), we note that

$$\frac{1}{\|\hat{\mathbf{C}}_N(\mathbf{w})\|} \leq \mathbf{a}^H(\theta_1) \hat{\mathbf{C}}_N^{-1}(\mathbf{w}) \mathbf{a}(\theta_1) \leq \|\hat{\mathbf{C}}_N^{-1}(\mathbf{w})\|$$

which yields $D(\mathbf{w}) \in [\xi^2, (q + 1)^2]$, and the same for $D(\mathbf{w}')$. Moreover, using (25) and (27), we readily obtain

$$N(\mathbf{w}') < (t_1 + 1)(q + 1)^2.$$

These results yield $\varepsilon_1 < \frac{(q+1)^2 g_1 + (t_1+1)(q+1)^2 g_2}{\xi^4} \triangleq K_1$. Therefore, we have established the desired property

$$\sup_{\mathbf{w} \in \mathcal{L}^m} |\rho(\mathbf{w}) - \rho(\mathbf{w}')| < K_1 \delta. \quad (28)$$

We now turn to deriving the analogous result for the second term in (23). To this end, similar to before, we start with

$$|\bar{\rho}(\mathbf{w}) - \bar{\rho}(\mathbf{w}')|$$

$$= \left| \frac{\bar{D}(\mathbf{w}') [\bar{N}(\mathbf{w}) - \bar{N}(\mathbf{w}')] + [\bar{D}(\mathbf{w}') - \bar{D}(\mathbf{w})] \bar{N}(\mathbf{w}')}{\bar{D}(\mathbf{w})\bar{D}(\mathbf{w}')} \right| \quad (29)$$

where now

$$\begin{aligned} \bar{N}(\mathbf{w}) &= 1 + \sum_{i=1}^m t_i k_i + \sum_{i=1}^m s_i k_i (1 + t_i s_i) w_i^2 \\ &\quad + 2 \sum_{i=1}^m s_i k_i (1 + t_i s_i) w_i, \end{aligned} \quad (30)$$

$$\bar{D}(\mathbf{w}) = \left(1 + \sum_{i=1}^m s_i k_i w_i \right)^2.$$

For the first difference term in the numerator of (29), we obtain

$$\begin{aligned} |\bar{N}(\mathbf{w}) - \bar{N}(\mathbf{w}')| &= \left| \sum_{i=1}^m s_i k_i (1 + t_i s_i) (w_i^2 - w'^2_i) \right. \\ &\quad \left. + \sum_{i=1}^m 2s_i k_i (1 + t_i s_i) (w_i - w'_i) \right|. \end{aligned}$$

Now, substituting for k_i and s_i (defined prior to the theorem statement), and using the definition of spectral norm and (27), this is readily bounded as

$$\begin{aligned} |\bar{N}(\mathbf{w}) - \bar{N}(\mathbf{w}')| &\leq \left\| \sum_{i=1}^m \frac{(t_i^2 - c_N)(t_i + 1)}{(t_i + c_N)^2} (w_i^2 - w'^2_i) \mathbf{v}_i \mathbf{v}_i^H \right\| \\ &\quad + \left\| \sum_{i=1}^m 2 \frac{(t_i^2 - c_N)(t_i + 1)}{(t_i + c_N)^2} (w_i - w'_i) \mathbf{v}_i \mathbf{v}_i^H \right\| \\ &\leq \max_{i \in (1, \dots, m)} \frac{(t_i^2 - c_N)(t_i + 1)}{(t_i + c_N)^2} \\ &\quad \times \left(\max_{j \in (1, \dots, m)} |w_j - w'_j| \max_{h \in (1, \dots, m)} |w_h + w'_h| + 2 \max_{l \in (1, \dots, m)} |w_l - w'_l| \right). \end{aligned}$$

Further, since $\frac{(t_i^2 - c_N)(t_i + 1)}{(t_i + c_N)^2}$ is increasing in t_i , we have $\max_{i \in (1, \dots, m)} \frac{(t_i^2 - c_N)(t_i + 1)}{(t_i + c_N)^2} = \frac{(t_1^2 - c_N)(t_1 + 1)}{(t_1 + c_N)^2}$, leading to

$$|\bar{N}(\mathbf{w}) - \bar{N}(\mathbf{w}')| < g_3 \delta$$

where $g_3 = \frac{(t_1^2 - c_N)(t_1 + 1)}{(t_1 + c_N)^2} (2q + 2)$. The second difference term in the numerator of (29) can be bounded using similar manipulations, which leads to

$$\begin{aligned} |\bar{D}(\mathbf{w}) - \bar{D}(\mathbf{w}')| &\leq \max_{i \in (1, \dots, m)} \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} \max_{j \in (1, \dots, m)} |w_j - w'_j| \\ &\quad \times \left(2 + \max_{h \in (1, \dots, m)} \frac{t_h^2 - c_N}{t_h^2 + c_N t_h} \max_{l \in (1, \dots, m)} |w_l + w'_l| \right). \end{aligned}$$

Like before, $\frac{t_i^2 - c_N}{t_i^2 + c_N t_i}$ is increasing in t_i , giving $\max_{i \in (1, \dots, m)} \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} = \frac{t_1^2 - c_N}{t_1^2 + c_N t_1}$, and thus leading to

$$|\bar{D}(\mathbf{w}) - \bar{D}(\mathbf{w}')| \leq g_4 \delta$$

where $g_4 = \frac{t_1^2 - c_N}{t_1^2 + c_N t_1} \left(2 + 2q \frac{t_1^2 - c_N}{t_1^2 + c_N t_1} \right)$. Combining these results produces the following bound for (29):

$$|\bar{\rho}(\mathbf{w}) - \bar{\rho}(\mathbf{w}')| < \varepsilon_2 \delta$$

where

$$\varepsilon_2 = \frac{\bar{D}(\mathbf{w}')g_3 + \bar{N}(\mathbf{w}')g_4}{\bar{D}(\mathbf{w})\bar{D}(\mathbf{w}')}.$$

Next, note that $\bar{D}(\mathbf{w})$ writes explicitly as $\bar{D}(\mathbf{w}) = \left(1 + \mathbf{a}^H(\theta_1) \sum_{i=1}^m \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} w_i \mathbf{v}_i \mathbf{v}_i^H \mathbf{a}(\theta_1)\right)^2$. As we have

$$\begin{aligned} \min_{i \in (1, \dots, m)} \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} w_i &\leq \mathbf{a}^H(\theta_1) \sum_{i=1}^m \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} w_i \mathbf{v}_i \mathbf{v}_i^H \mathbf{a}(\theta_1) \\ &\leq \max_{i \in (1, \dots, m)} \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} w_i, \end{aligned}$$

together with $0 < \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} < 1$ and $w_i \in [-1 + \xi, q]$, we have $-1 + \xi < \mathbf{a}^H(\theta_1) \sum_{i=1}^m \frac{t_i^2 - c_N}{t_i^2 + c_N t_i} w_i \mathbf{v}_i \mathbf{v}_i^H \mathbf{a}(\theta_1) < q$. Thus, we obtain $\bar{D}(\mathbf{w}) \in (\xi^2, (q+1)^2)$. Similarly, we have $\bar{D}(\mathbf{w}') \in (\xi^2, (q+1)^2)$. As for $\bar{N}(\mathbf{w})$, using (30) and (27), we obtain $\bar{N}(\mathbf{w}) \leq 1 + t_1 + \frac{(t_1^2 - c_N)(t_1 + 1)}{(t_1 + c_N)^2} q^2 + 2q \frac{(t_1^2 - c_N)(t_1 + 1)}{(t_1 + c_N)^2} \triangleq g_5$,

which then yields $\varepsilon_2 < \frac{(q+1)^2 g_3 + g_5 g_4}{\xi^4} \triangleq K_2$. Therefore, we have established the desired property

$$\sup_{\mathbf{w} \in \mathcal{L}^m} |\bar{\rho}(\mathbf{w}) - \bar{\rho}(\mathbf{w}')| < K_2 \delta. \quad (31)$$

Finally, we turn to deriving the analogous result (for all large n , almost surely) for the third term in (23). Since, as already established, for each $\mathbf{w}_i \in (\mathbf{w}_1, \dots, \mathbf{w}_J)$, as $N, n \rightarrow \infty$ such that $c_N = N/n \rightarrow c > 0$, $|\rho(\mathbf{w}_i) - \bar{\rho}(\mathbf{w}_i)| \xrightarrow{\text{a.s.}} 0$, we have that for each $\mathbf{w}_i \in (\mathbf{w}_1, \dots, \mathbf{w}_J)$, $|\rho(\mathbf{w}_i) - \bar{\rho}(\mathbf{w}_i)| < \delta$ for all large n almost surely. Thus,

$$\max_{\mathbf{w}'' \in (\mathbf{w}_1, \dots, \mathbf{w}_J)} |\rho(\mathbf{w}'') - \bar{\rho}(\mathbf{w}'')| < \sum_{i=1}^J |\rho(\mathbf{w}_i) - \bar{\rho}(\mathbf{w}_i)| < J\delta$$

for all large n almost surely. This, combined with (31) and (28) completes the proof that (22) holds, hence establishing the desired uniform convergence (21).

APPENDIX B PROOF OF THEOREM 2

By rewriting the numerator of (14) as

$$1 + \sum_{i=1}^m \left[(t_i s_i^2 k_i + s_i k_i) \left(w_i + \frac{t_i + 1}{t_i s_i + 1} \right)^2 + t_i k_i - \frac{(s_i k_i + t_i s_i k_i)^2}{t_i s_i^2 k_i + s_i k_i} \right],$$

and the denominator as

$$\left(\sum_{i=1}^m \left[s_i k_i \left(w_i + \frac{t_i + 1}{t_i s_i + 1} \right) - \frac{t_i s_i k_i + s_i k_i}{t_i s_i + 1} \right] + 1 \right)^2,$$

the optimization problem of interest can be recast as

$$\mathbf{w}^* = \underset{\mathbf{w} \in \mathcal{L}^m}{\operatorname{argmin}} \frac{\sum_{i=1}^m a_i (w_i + b_i)^2 + d}{\left(\sum_{i=1}^m e_i (w_i + b_i) + f \right)^2} \quad (32)$$

where, for $i = 1, \dots, m$,

$$a_i = t_i s_i^2 k_i + s_i k_i, \quad b_i = \frac{t_i + 1}{t_i s_i + 1}, \quad e_i = s_i k_i,$$

and

$$\begin{aligned} d &= 1 + \sum_{j=1}^m \left(t_j k_j - \frac{(s_j k_j + t_j s_j k_j)^2}{t_j s_j^2 k_j + s_j k_j} \right) \\ f &= 1 - \sum_{j=1}^m \frac{t_j s_j k_j + s_j k_j}{t_j s_j + 1}. \end{aligned}$$

To solve this, we first note that application of the Cauchy-Schwarz inequality to the denominator of (32) yields

$$\begin{aligned} \left(\sum_{i=1}^m e_i (w_i + b_i) + f \right)^2 &= \left(\sum_{i=1}^m \sqrt{a_i} (w_i + b_i) \frac{e_i}{\sqrt{a_i}} + \sqrt{d} \frac{f}{\sqrt{d}} \right)^2 \\ &\leq \left(\sum_{i=1}^m a_i (w_i + b_i)^2 + d \right) \left(\sum_{i=1}^m \frac{e_i^2}{a_i} + \frac{f^2}{d} \right) \end{aligned}$$

and therefore the objective function on the right-hand side of (32) satisfies

$$\frac{\sum_{i=1}^m a_i (w_i + b_i)^2 + d}{\left(\sum_{i=1}^m e_i (w_i + b_i) + f \right)^2} \geq \frac{1}{\sum_{i=1}^m \frac{e_i^2}{a_i} + \frac{f^2}{d}}.$$

Here, the “=” is reached when $\frac{\sqrt{a_i}(w_i + b_i)}{\frac{e_i}{\sqrt{a_i}}} = \frac{\sqrt{d}}{\frac{f}{\sqrt{d}}}$, $i = 1, \dots, m$.

As such, the the optimal weights $\bar{w}_i^*$, $i = 1, \dots, m$ admit

$$\bar{w}_i^* = \frac{e_i d}{a_i f} - b_i = \frac{t_i + c_N}{t_i^2 + t_i} (\tau - t_i)$$

with $\tau = \frac{\sum_{j=1}^m \frac{c_N k_j}{t_j}}{1 - \sum_{j=1}^m k_j + \sum_{j=1}^m \frac{c_N k_j}{t_j^2}}$.

Additionally, since the signal steering vector $\mathbf{a}(\theta_1)$ lies entirely in the signal-plus-interference subspace, it follows that $\sum_{j=1}^m k_j = 1$, giving the further simplification

$$\bar{w}_i^* = \frac{t_i + c_N}{t_i^2 + t_i} (\psi - t_i), \quad \text{where } \psi = \frac{\sum_{j=1}^m \frac{k_j}{t_j}}{\sum_{j=1}^m \frac{k_j}{t_j^2}},$$

which is the stated result.

APPENDIX C PROOF OF THEOREM 3

First, we recall the following RMT result [43]: For $t_i > \sqrt{c}$, for all large n with probability one, $\lambda_i > \sigma^2(1 + \sqrt{c_N})^2$. When $N, n \rightarrow \infty$, with $c_N \rightarrow c$,

$$\left| \lambda_i / \sigma^2 - 1 - t_i - \frac{c_N(1 + t_i)}{t_i} \right| \xrightarrow{\text{a.s.}} 0. \quad (33)$$

By direct inversion, we obtain consistent estimators of t_i . Specifically, we obtain

$$|\hat{t}_i - t_i| \xrightarrow{\text{a.s.}} 0, \quad i = 1, \dots, m$$

where

$$\hat{t}_i = \frac{\lambda_i / \sigma^2 + 1 - c_N + \sqrt{(\lambda_i / \sigma^2 + 1 - c_N)^2 - 4\lambda_i / \sigma^2}}{2} - 1.$$

Together with (20), it also yields consistent estimators of k_i . Specifically,

$$\left| \frac{1}{\hat{s}_i} \mathbf{a}^H(\theta_1) \mathbf{u}_i \mathbf{u}_i^H \mathbf{a}(\theta_1) - k_i \right| \xrightarrow{\text{a.s.}} 0$$

$$\text{with } \hat{s}_i = \frac{1 - c_N/(\hat{t}_i)^2}{1 + c_N/\hat{t}_i}.$$

Therefore, we obtain consistent estimators $\hat{w}_i^*$ of the asymptotically-optimal weights $\bar{w}_i^*$ by substituting $\hat{t}_i$ and $\hat{k}_i$ for t_i and k_i respectively.

REFERENCES

- [1] L. Brennan, J. Mallett, and I. Reed, "Adaptive arrays in airborne MTI radar," *IEEE Trans. Antennas Propag.*, vol. 24, no. 5, pp. 607–615, Sept. 1976.
- [2] J. L. Krolik, "The performance of matched-field beamformers with mediterranean vertical array data," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2605–2611, Oct. 1996.
- [3] A. B. Gershman, V. I. Turchin, and V. A. Zverev, "Experimental results of localization of moving underwater signal by adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 43, no. 10, pp. 2249–2257, Oct. 1995.
- [4] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, no. 6, pp. 1391–1400, Dec. 1986.
- [5] L. C. Godara, "Application of antenna arrays to mobile communications Part II: Beam-forming and direction-of-arrival consideration," *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.
- [6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*. New York: Wiley, 2002.
- [8] D. D. Feldman and L. J. Griffiths, "A projection approach for robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 867–876, Apr. 1994.
- [9] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. and Electron. Syst.*, no. 6, pp. 853–863, Nov. 1974.
- [10] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. and Electron. Syst.*, no. 2, pp. 115–127, Mar. 1986.
- [11] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.
- [12] D. M. Boroson, "Sample size considerations for adaptive arrays," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 24, no. 4, pp. 446–451, Jul. 1980.
- [13] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997, vol. 1.
- [14] L. Chang and C.-C. Yeh, "Performance of DMI and eigenspace-based beamformers," *IEEE Trans. Antennas Propag.*, vol. 40, no. 11, pp. 1336–1347, Nov. 1992.
- [15] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 24, no. 4, pp. 397–401, July 1988.
- [16] M. W. Ganz, R. Moses, and S. Wilson, "Convergence of the SMI and the diagonally loaded SMI algorithms with weak interference (adaptive array)," *IEEE Trans. Antennas Propag.*, vol. 38, no. 3, pp. 394–399, Mar. 1990.
- [17] X. Mestre and M. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.
- [18] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.
- [19] F. Rubio, X. Mestre, and W. Hachem, "A CLT on the SNR of diagonally loaded MVDR filters," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4178–4195, May 2012.
- [20] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
- [21] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Modified GLRT and AMF framework for adaptive detectors," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 43, no. 3, Nov. 2007.
- [22] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.
- [23] Y. Kim, S. U. Pillai, and J. R. Guerci, "Optimal loading factor for minimal sample support space-time adaptive radar," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1998, pp. 2505–2508.
- [24] N. Ma and J. T. Goh, "Efficient method to determine diagonal loading value," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 2003, pp. V–341.
- [25] L. Yang, R. Couillet, and M. R. McKay, "A robust statistics approach to minimum variance portfolio optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6684–6697, Aug. 2015.
- [26] K. Elkhailil, A. Kammoun, T. Y. Al-Naffouri, and M.-S. Alouini, "Fluctuations of the SNR at the output of the MVDR with regularized Tyler estimators," *Signal Process.*, vol. 135, pp. 1–8, May 2017.
- [27] R. Couillet and M. McKay, "Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators," *J. Multivar. Anal.*, vol. 131, pp. 99–120, Aug. 2014.
- [28] P. J. Napier, A. R. Thompson, and R. D. Ekers, "The very large array: Design and performance of a modern synthesis radio telescope," *Proc. IEEE*, vol. 71, no. 11, pp. 1295–1320, Nov. 1983.
- [29] S. D. Somasundaram, N. H. Parsons, P. Li, and R. C. De Lamare, "Reduced-dimension robust Capon beamforming using Krylov-subspace techniques," *IEEE Trans. Aerosp. and Electron. Syst.*, vol. 51, no. 1, pp. 270–289, Apr. 2015.
- [30] E. Y. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, "Deep-water acoustic coherence at long ranges: Theoretical prediction and effects on large-array signal processing," *IEEE J. Ocean. Eng.*, vol. 24, no. 2, pp. 156–171, Apr. 1999.
- [31] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [32] D. L. Donoho, M. Gavish, and I. M. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," to appear in *Ann. Statist.*, *arXiv preprint arXiv:1311.0851*, 2017.
- [33] I. M. Johnstone and A. Y. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *J. Amer. Stat. Assoc.*, vol. 104, no. 486, pp. 682–693, 2009.
- [34] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics Intell. Lab. Syst.*, vol. 94, no. 1, pp. 19–32, Nov. 2008.
- [35] D. Passemier, Z. Li, and J. Yao, "On estimation of the noise variance in high dimensional probabilistic principal component analysis," *J. R. Stat. Soc.*, vol. 79, no. 1, pp. 51–67, Jan. 2017.
- [36] M. O. Ulfarsson and V. Solo, "Dimension estimation in noisy PCA with SURE and random matrix theory," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5804–5816, Sept. 2008.
- [37] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [38] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley, 1999.
- [39] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
- [40] S. Pafka and I. Kondor, "Noisy covariance matrices and portfolio optimization II," *Physica A*, vol. 319, pp. 487–494, 2003.
- [41] —, "Estimated correlation matrices and portfolio optimization," *Physica A*, vol. 343, pp. 623–634, 2004.
- [42] G. Papp, S. Pafka, M. A. Nowak, and I. Kondor, "Random matrix filtering in portfolio optimization," *Acta Physica Polonica B*, vol. 36, no. 9, pp. 2757–2765, 2005.
- [43] J. Baik, G. B. Arous, and S. Pécché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.
- [44] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivar. Anal.*, vol. 97, no. 6, pp. 1382–1408, 2006.
- [45] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, pp. 1617–1642, Oct. 2007.
- [46] N. E. Karoui, "Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices," *Ann. Probab.*, vol. 35, no. 2, pp. 663–714, 2007.

- [47] —, “Spectrum estimation for large dimensional covariance matrices using random matrix theory,” *Ann. Statist.*, vol. 36, no. 6, pp. 2757–2790, 2008.
- [48] B. Nadler, “Finite sample approximation results for principal component analysis: A matrix perturbation approach,” *Ann. Statist.*, vol. 36, no. 6, pp. 2791–2817, 2008.
- [49] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, “Shrinkage algorithms for MMSE covariance estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Jun. 2010.
- [50] R. Daniels, Michael Jand Kass, “Shrinkage estimators for covariance matrices,” *Biometrics*, vol. 57, no. 4, pp. 1173–1184, 2001.
- [51] J. Fan, Y. Fan, and J. Lv, “High dimensional covariance matrix estimation using a factor model,” *J. Economet.*, vol. 147, no. 1, pp. 186–197, Nov. 2008.
- [52] O. Ledoit and M. Wolf, “Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions,” *J. Multivar. Anal.*, vol. 139, pp. 360–384, Jul. 2015.
- [53] —, “Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks,” *Rev. Finan. Stud.*, Jun. 2017.
- [54] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Math. USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.
- [55] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. U.K., Cambridge: Cambridge Univ. Press, 2011.
- [56] X. Mestre and M. Á. Lagunas, “Modified subspace algorithms for DoA estimation with large arrays,” *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–614, Jan. 2008.
- [57] P. Vallet, X. Mestre, and P. Loubaton, “Performance analysis of an improved MUSIC DoA estimator,” *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6407–6422, Aug. 2015.
- [58] R. Couillet and A. Kammoun, “Robust G-MUSIC,” in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. IEEE, Sep. 2014, pp. 2155–2159.
- [59] R. Couillet, “Robust spiked random matrices and a robust G-MUSIC estimator,” *J. Multivar. Anal.*, vol. 140, pp. 139–161, Sep. 2015.
- [60] B. A. Johnson, Y. I. Abramovich, and X. Mestre, “MUSIC, G-MUSIC, and maximum-likelihood performance breakdown,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3944–3958, Jul. 2008.
- [61] S. Kritchman and B. Nadler, “Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory,” *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, May 2009.
- [62] P. Vallet, P. Loubaton, and X. Mestre, “Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case,” *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1043–1068, Feb. 2012.
- [63] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
- [64] G. Schwarz *et al.*, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.
- [65] J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [66] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
- [67] R. R. Nadakuditi and A. Edelman, “Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, Apr. 2008.
- [68] J. Li, P. Stoica, and Z. Wang, “On robust Capon beamforming and diagonal loading,” *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
- [69] R. G. Lorenz and S. P. Boyd, “Robust minimum variance beamforming,” *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, Apr. 2005.
- [70] K. L. Bell, Y. Ephraim, and H. L. Van Trees, “A Bayesian approach to robust adaptive beamforming,” *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.
- [71] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, “Random matrix approach to cross correlations in financial data,” *Phys. Rev. E*, vol. 65, no. 6, p. 066126, Jun. 2002.
- [72] R. A. Horn and C. R. Johnson, *Matrix Analysis*. U.K., Cambridge: Cambridge Univ. Press, 2012.