Threshold Effects in Parameter Estimation From Compressed Data

Pooria Pakrooh, *Student Member, IEEE*, Louis L. Scharf, *Life Fellow, IEEE*, and Ali Pezeshki, *Member, IEEE*

***Abstract*—In this paper, we investigate threshold effects associated with the swapping of signal and noise subspaces in estimating signal parameters from compressed noisy data. The term threshold effect refers to a sharp departure of mean-squared error from the Cramér–Rao bound when the signal-to-noise ratio falls below a threshold SNR. In many cases, the threshold effect is caused by a subspace swap event, when the measured data (or its sample covariance) is better approximated by a subset of components of an orthogonal subspace than by the components of a signal subspace. We derive analytical lower bounds on the probability of a subspace swap in compressively measured noisy data in two canonical models: a first-order model and a second-order model. In the first-order model, the parameters to be estimated modulate the mean of a complex multivariate normal set of measurements. In the second-order model, the parameters modulate the covariance of complex multivariate measurements. In both cases, the probability bounds are tail probabilities of F-distributions, and they apply to any linear compression scheme. These lower bounds guide our understanding of threshold effects and performance breakdowns for parameter estimation using compression. In particular, they can be used to quantify the increase in threshold SNR as a function of a compression ratio C. We demonstrate numerically that this increase in threshold SNR is roughly $10\log_{10}\ C$ dB, which is consistent with the performance loss that one would expect when measurements in Gaussian noise are compressed by a factor C.**

***Index Terms*—Co-prime sampling, Cramér-Rao bound, maximum likelihood estimation, mean squared error, random compression, subspace swap, threshold effects.**

I. Introduction

THE performance of many high resolution parameter estimation methods, including subspace and maximum likelihood methods, may suffer from performance breakdown, where the mean squared error (MSE) departs sharply from the Cramér-Rao bound at low signal-to-noise ratio (SNR). Performance breakdown may happen when either the sample size or SNR falls below a certain threshold [1]. The main reason for this threshold effect is that in low SNR or sample size regimes, parameter estimation methods lose their capability to resolve signal and noise subspaces. As a result of this, one or more components in the orthogonal (noise) subspace better approximate the data than at least one component of the signal subspace, which in turn leads to a large error in parameter estimation [2]. This phenomenon is called a subspace swap.

A number of papers have studied subspace swap events and threshold effects in the past. The first insights into subspace swap events were established in [1], where the authors present an analytical study of threshold effects in linear prediction methods that use singular value decomposition (SVD) of a data matrix for rank reduction. They present a procedure for estimating the threshold SNR, and associate the threshold effect with the probability of the event that a subset of components in the orthogonal (noise) subspace better approximate the data than at least one component of the signal subspace. Following [1], the authors of [2] revisited the problem, introduced the term subspace swap in describing the event observed in [1], and derived analytical lower bounds on the probability of a subspace swap in the data matrix. The lower bounds are tail probabilities of a weighted sum of chi-squared random variables, and are obtained by finding the exact probability distribution of subevents of the event considered in [1]. In both of these papers, the data follows a multivariate normal model with an unknown parameterized mean vector, formed by a superposition of complex exponential modes.

In [3], the authors predict the probability of a subspace swap in the eigendecomposition of the sample covariance matrix of multivariate normal measurements. The prediction is asymptotic in the number of snapshots, but the dimension of the data vector itself is fixed. As the number of snapshots grows without bound, this prediction can be calculated as the tail probability of a multivariate normal. Another asymptotic study is presented in [4], where the authors study performance breakdowns of maximum likelihood, MUSIC and G-MUSIC algorithms for DOA estimation. In this study, both the array dimension and the number of snapshots grow without bound (at the same rate). Under these assumptions, the authors predict the number of snapshots for which a subspace swap occurs. Their numerical examples show that their predictions remain valid even in relatively modest snapshot regimes. They show that different mechanisms are responsible for the breakdown of maximum likelihood estimation and MUSIC. While a subspace swap is the main source of performance breakdown in maximum likelihood, even a small subspace leakage can lead to performance breakdown of MUSIC in resolving closely-spaced sources.

Manuscript received April 29, 2015; revised October 15, 2015 and December 22, 2015; accepted January 07, 2016. Date of publication January 25, 2016; date of current version March 30, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Tabrikian. This work is supported in part by NSF under grants CCF-1018472 and CCF-1422658. A preliminary version of a subset of the results reported here was presented at the IEEE Global Conference on Signal and Information Processing (GlobalSIP), Austin, Texas, USA, December 3–5, 2013.

P. Pakrooh and A. Pezeshki are with the Department of Electrical and Computer Engineering and the Department of Mathematics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: Pooria.Pakrooh@colostate.edu; Ali.Pezeshki@colostate.edu).

L. L. Scharf is with the Department of Mathematics and the Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: Louis.Scharf@colostate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2521617

Over the past decade or so, following first the introduction of compressed sensing (see, e.g., [5]–[7]) and later co-prime sampling (see, e.g., [8]–[12]), parameter estimation from compressed data has become a topic of enormous interest in the signal processing community. A large number of papers have so far investigated both identifiability conditions (noise-free) and performance bounds (in noise) in estimating parameters from compressed noisy data, for a variety of linear compression schemes. The analyses include bounds on ℓ_p norm of the error (see, e.g., [5]–[7], [13], and [14]) as well as bounds on loss of Fisher Information and increase in Cramér-Rao Bound due to compression (see, e.g., [15]–[17]).

In this paper, we address the effect of compression on the probability of a subspace swap. In other words, we ask what effect compression has on the threshold SNR at which performance breaks down. To answer this question, we follow the approach of [2] to derive analytical lower bounds on the probability of a subspace swap in parameter estimation from compressed noisy measurements. These bounds provide the first set of results on the impact of compression on subspace swaps and threshold SNRs. We consider two measurement models. In the first-order model, the parameters to be estimated modulate the mean of a complex multivariate normal set of measurements. In the second-order model, the parameters modulate the covariance of complex multivariate measurements. In both cases, the probability bounds are tail probabilities of F-distributions, and they can be used with any linear compression scheme. Moreover, these bounds are not asymptotic and are valid in finite snapshot regimes. The bounds guide our understanding of threshold effects and performance breakdowns for parameter estimation using compression. In particular, they can be used to quantify the increase in threshold SNR as a function of a compression ratio C. We demonstrate numerically that this increase in threshold SNR is roughly $10\log_{10} C$ dB, which is consistent with the performance loss that one would expect when measurements in Gaussian noise are compressed by a factor C. This can in turn be used to decide whether or not compression to a particular dimension fits performance constraints in a desired application. As a case study, we investigate threshold effects in maximum likelihood (ML) estimation of directions of arrival of two closely-spaced sources using co-prime subsampling and uniformly at random subsampling. Our MSE plots validate the increase in threshold SNR.

Remark 1: There are a number of other papers that study the performance breakdown regions of high resolution parameter estimation methods. These studies are based on perturbation analyses of SVDs and do not directly analyze or bound probabilities of subspace swap events. For example, in [18]–[20], the authors carry out perturbation analyses of SVDs to study the performance of subspace based methods for parameter estimation, when *subspace leakage* happens between signal and noise subspaces, and in [21] a method to reduce the subspace leakage for the direction of arrival (DOA) estimation problem using root-MUSIC algorithm has been proposed. Performance breakdown of maximum likelihood has been studied in [22]–[26] by perturbation analysis using an asymptotic assumption on the number of snapshots. Other relevant perturbation analysis papers include [27]–[29]. For problems with large arrays, large random matrix theory (see, e.g., [30] and [31] for surveys of relevant literature) provides powerful tools for analysis of performance breakdowns and threshold effects. For example, in [32], the performance breakdown of subspace based methods is attributed to the breakdown point of Principal Component Analysis (PCA) methods of signal subspace estimation, and tools for studying the behavior of eigenvalues and eigenvectors of low rank perturbations of large random matrices from [33], have been used to predict this breakdown point. Although this study does not provide exact lower bounds, it offers valuable insights and great predictions for studying threshold effects in high dimensional settings.

II. Measurement Model

In the following subsections, we consider two models for the random measurement vector $\mathbf{y} \in \mathbb{C}^n$. In the first-order model, the parameters to be estimated nonlinearly modulate the mean of a complex multivariate normal vector, and in the second-order model the parameters nonlinearly modulate the covariance of a complex multivariate normal vector.

A. Parameterized Mean Case

Let $\mathbf{y} \in \mathbb{C}^n$ be a complex measurement vector in a signal plus noise model $\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) + \mathbf{n}$. Here, we assume that $\mathbf{n}$ is a proper complex white Gaussian noise with covariance $\sigma^2\mathbf{I}$ and $\mathbf{x}(\boldsymbol{\theta})$ is parameterized by $\boldsymbol{\theta} \in \mathbb{C}^p$, $p \leq n$. We assume that the parameters are nonlinearly embedded in $\mathbf{x}(\boldsymbol{\theta})$ as $\mathbf{x}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}$, where the columns of $\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{k}(\boldsymbol{\theta}_1)\ \mathbf{k}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{k}(\boldsymbol{\theta}_p)]$ define the signal subspace, and $\boldsymbol{\alpha} \in \mathbb{C}^p$ is a deterministic vector of complex amplitudes α_i, $i = 1, 2, \ldots, p$, of the modes. Therefore, $\mathbf{y}$ is distributed as $\mathcal{CN}_n(\mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}, \sigma^2\mathbf{I})$, and the parameters $\boldsymbol{\theta} \in \mathbb{C}^p$ to be estimated nonlinearly modulate the mean of a complex multivariate normal vector. Assume we compress the measurement vector $\mathbf{y}$ by a unitary compression matrix $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1/2}\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$, $p \leq m < n$. Then, we obtain $\mathbf{w} = \boldsymbol{\Psi}\mathbf{y}$ which is distributed as $\mathcal{CN}_m(\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I})$, where $\mathbf{z}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\mathbf{x}(\boldsymbol{\theta})$. We form the data matrix $\mathbf{W} = [\mathbf{w}_1\mathbf{w}_2\cdots\mathbf{w}_M]$, where $\mathbf{w}_i$'s are independent realizations of $\mathbf{w}$. To specify a basis for the signal subspace and the orthogonal subspace in our problem, we define $\mathbf{H}(\boldsymbol{\theta}) = \boldsymbol{\Psi}\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{h}(\boldsymbol{\theta}_1)\ \mathbf{h}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{h}(\boldsymbol{\theta}_p)]$, with $\mathbf{h}(\boldsymbol{\theta}_i) = \boldsymbol{\Psi}\mathbf{k}(\boldsymbol{\theta}_i)$. The singular value decomposition of $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \tag{1}$$

where

$$\begin{aligned}
&\mathbf{U} \in \mathbb{C}^{m\times m} : \mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}\\
&\mathbf{V} \in \mathbb{C}^{p\times p} : \mathbf{V}\mathbf{V}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}\\
&\boldsymbol{\Sigma} \in \mathbb{C}^{m\times p} : \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_p \\ \mathbf{0} \end{bmatrix}\\
&\boldsymbol{\Sigma}_p = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p), \quad \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p.
\end{aligned} \tag{2}$$

Now we can define the basis vectors from $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p|\mathbf{u}_{p+1}, \ldots, \mathbf{u}_m] = [\mathbf{U}_p|\mathbf{U}_0]$, where $\langle\mathbf{U}_p\rangle$ and $\langle\mathbf{U}_0\rangle$ represent signal and orthogonal subspaces, respectively. The columns of $\mathbf{U}_p$ and $\mathbf{U}_0$ can be considered as basis vectors for the signal and orthogonal subspaces, respectively.

B. Parameterized Covariance Case

Assume in the signal plus noise model $\mathbf{y} = \mathbf{x} + \mathbf{n}$, the signal component $\mathbf{x}$ is of the form $\mathbf{x} = \mathbf{K}(\boldsymbol{\theta})\boldsymbol{\alpha}$, where the columns of $\mathbf{K}(\boldsymbol{\theta}) = [\mathbf{k}(\boldsymbol{\theta}_1)\ \mathbf{k}(\boldsymbol{\theta}_2)\ \cdots\ \mathbf{k}(\boldsymbol{\theta}_p)]$ are the modes and $\boldsymbol{\alpha} \in \mathbb{C}^p$

is the random vector of complex amplitudes of the modes. We assume $\boldsymbol{\alpha}$ is distributed as $\mathcal{CN}_p(0, \mathbf{R}_{\boldsymbol{\alpha}\boldsymbol{\alpha}})$. Therefore, $\mathbf{R}_{xx}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\alpha}\mathbf{K}^H(\boldsymbol{\theta})$ is parameterized by $\boldsymbol{\theta} \in \mathbb{C}^p$. We assume $\mathbf{n}$ is a proper complex white Gaussian noise with covariance $\sigma^2\mathbf{I}$, and $\mathbf{x}$ and $\mathbf{n}$ are independent. Therefore, $\mathbf{y}$ is distributed as $\mathcal{CN}_n(\mathbf{0}, \mathbf{R}_{\mathbf{yy}}(\boldsymbol{\theta}))$, where $\mathbf{R}_{\mathbf{yy}}(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\alpha}\mathbf{K}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}$. Such a data model arises in many applications such as direction of arrival and spectrum estimation.

Assume we compress the measurement vector $\mathbf{y}$ by a unitary compression matrix $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1/2}\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}(m < n)$. Then, we obtain $\mathbf{w} = \boldsymbol{\Psi}\mathbf{y}$ which is distributed as

$$\mathbf{w} \sim \mathcal{CN}_m(\mathbf{0}, \mathbf{R}_{\mathbf{ww}}) \tag{3}$$

where $\mathbf{R}_{\mathbf{ww}} = \boldsymbol{\Psi}\mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\alpha}\mathbf{K}^H(\boldsymbol{\theta})\boldsymbol{\Psi}^H + \sigma^2\mathbf{I}$. We form the data matrix $\mathbf{W} = [\mathbf{w}_1\mathbf{w}_2\cdots\mathbf{w}_M]$, where $\mathbf{w}_i$'s are independent realizations of $\mathbf{w}$. Each of these i.i.d. realizations consists of an i.i.d. realization of $\mathbf{y}_i$, compressed by a common compressor $\boldsymbol{\Psi}$ for all $i = 1, 2, \ldots, M$. We may define the signal covariance matrix after compression as

$$\begin{aligned}\mathbf{R}_{\mathbf{zz}} &= \boldsymbol{\Psi}\mathbf{K}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\alpha}\mathbf{K}^H(\boldsymbol{\theta})\boldsymbol{\Psi}^H \\ &= \mathbf{H}(\boldsymbol{\theta})\mathbf{R}_{\boldsymbol{\alpha}\alpha}\mathbf{H}^H(\boldsymbol{\theta}),\end{aligned} \tag{4}$$

where $\mathbf{H}(\boldsymbol{\theta}) = [\mathbf{h}(\boldsymbol{\theta}_1)\mathbf{h}(\boldsymbol{\theta}_2)\cdots\mathbf{h}(\boldsymbol{\theta}_p)]$, and $\mathbf{h}(\boldsymbol{\theta}_i) = \boldsymbol{\Psi}\mathbf{k}(\boldsymbol{\theta}_i)$. Now, we can write the singular value decomposition of $\mathbf{R}_{\mathbf{zz}}$ and $\mathbf{R}_{\mathbf{ww}}$ as

$$\begin{aligned}\mathbf{R}_{\mathbf{zz}} &= \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H \\ \mathbf{R}_{\mathbf{ww}} &= \mathbf{U}(\boldsymbol{\Lambda} + \sigma^2\mathbf{I})\mathbf{U}^H\end{aligned} \tag{5}$$

where $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are defined as

$$\begin{aligned}&\mathbf{U} \in \mathbb{C}^{m\times m} : \mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I} \\ &\boldsymbol{\Lambda} \in \mathbb{C}^{m\times m} : \boldsymbol{\Lambda} = \begin{bmatrix}\boldsymbol{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \\ &\boldsymbol{\Lambda}_p = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p), \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p.\end{aligned} \tag{6}$$

Assuming $\mathbf{R}_{\mathbf{zz}}$ has rank p, the unitary matrix $\mathbf{U}$ can be written as $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p|\mathbf{u}_{p+1}, \ldots, \mathbf{u}_m] = [\mathbf{U}_p|\mathbf{U}_0]$. Here $\langle\mathbf{U}_p\rangle$ represents the signal subspace and $\langle\mathbf{U}_0\rangle$ represents the orthogonal subspace which completes $\mathbb{C}^{m\times m}$, assuming $p \leq m < n$. Fig. 1 gives a geometrical representation of (6).

III. Subspace Swap Events

We define three subspace swap events in this section. The first event, denoted by E, encompasses all possible combinations for a swap of vectors in the noise subspace and modes in the signal subspace. The other two events, denoted by F and G, are subevents of E. We use these to lower bound the probability of event E.

We define the subspace swap event E as

$$E = \cup_{q=1}^{p} E(q), \tag{7}$$

where $E(q)$ is the subevent that

$$\min_{\mathbf{A}\in\mathcal{I}_{p,q}} \mathrm{tr}\,(\mathbf{W}^H\mathbf{P}_{\mathbf{HA}}\mathbf{W}) < \max_{\mathbf{B}\in\mathbb{C}^{(n-p)\times q}} \mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0\mathbf{B}}\mathbf{W}). \tag{8}$$

In the above equation, the columns of $\mathbf{H}$ are the modes that are defined in Section II, $\mathcal{I}_{p,q}$ is the set of all $p \times q$ slices of the $p \times p$ identity matrix $\mathbf{I}_p$, and $\mathbf{B}$ is an $(n-p)\times q$ matrix

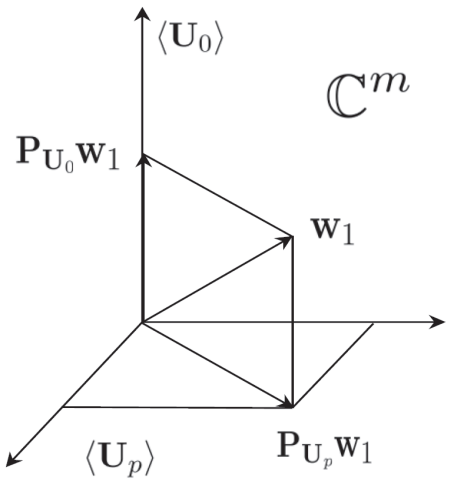


Fig. 1. Signal and noise subspaces.

of rank $q \leq (n-p)$. On the left hand side, the matrix $\mathbf{A}$ selects q of the modes in $\mathbf{H}$ and $\mathbf{P}_{\mathbf{HA}}$ is the projection onto the q-dimensional subspace of the signal subspace that is spanned by those modes. On the right hand side, the columns of $\mathbf{U}_0\mathbf{B}$ form a basis for a q-dimensional subspace of the orthogonal subspace $\langle\mathbf{U}_0\rangle$, and $\mathbf{P}_{\mathbf{U}_0\mathbf{B}}$ is the orthogonal projection matrix onto $\langle\mathbf{U}_0\mathbf{B}\rangle$. The terms $\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{HA}}\mathbf{W})$ and $\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0\mathbf{B}}\mathbf{W})$ are the energies in $\mathbf{W}$ resolved by $\langle\mathbf{HA}\rangle$ and $\langle\mathbf{U}_0\mathbf{B}\rangle$, respectively. The subevent $E(q)$ is the event that there exists a q-dimensional subspace of the orthogonal subspace $\langle\mathbf{U}_0\rangle$ that resolves more energy in $\mathbf{W}$ than a subset of q mode vectors do. The subspace swap event E is the union of subevents $\{E(q)\}_{q=1}^{p}$, which are obviously not mutually exclusive. From this definition of the subspace swap event, the event E may be stochastically simulated to stochastically evaluate $P(E)$. But our ambitions are different: we aim to analytically bound $P(E)$.

Remark 2: The definition of E is inspired by the description of the subspace swap event in [1] and [2]. In these papers, a subspace swap is described as the event that one or more modes of the orthogonal subspace resolve more energy in $\mathbf{W}$ than one or more modes of the noise-free signal subspace, but the event is not explicitly defined.

We define F as the event that

$$\frac{1}{p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_p}\mathbf{W}) \leq \frac{1}{m-p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0}\mathbf{W}). \tag{9}$$

This is the event that the average energy resolved in the orthogonal subspace $\langle\mathbf{U}_0\rangle$ is greater than the average energy resolved in the noise-free signal subspace $\langle\mathbf{U}_p\rangle$ (or equivalently $\langle\mathbf{H}\rangle$). Clearly, F is a subset of $E(1)$, and therefore a subset of E, as the following argument shows

$$\begin{aligned}\min_{1\leq i\leq p} \mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{h}_i}\mathbf{W}) &\leq \frac{1}{p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_p}\mathbf{W}) \\ &< \frac{1}{m-p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0}\mathbf{W}) \\ &\leq \max_{p+1\leq i\leq m} \mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{u}_i}\mathbf{W}) \\ &\leq \max_{\mathbf{b}\in\mathbb{C}^{(n-p)\times 1}} \mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0\mathbf{b}}\mathbf{W}),\end{aligned} \tag{10}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$.

We define G as the event that

$$\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{h}_{min}}\mathbf{W}) \leq \frac{1}{m-p}\mathrm{tr}\,(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0}\mathbf{W}). \quad (11)$$

For the parameterized mean measurement model, $\mathbf{h}_{min}$ is

$$\mathbf{h}_{min} = \underset{\mathbf{h}\in\{\mathbf{h}(\boldsymbol{\theta}_1),\mathbf{h}(\boldsymbol{\theta}_2),\ldots,\mathbf{h}(\boldsymbol{\theta}_p)\}}{\arg\min} \frac{|\mathbf{h}^H\mathbf{z}(\boldsymbol{\theta})|^2}{\|\mathbf{h}\|_2^2}, \quad (12)$$

and for the parameterized covariance measurement model $\mathbf{h}_{min}$ is

$$\mathbf{h}_{min} = \underset{\mathbf{h}\in\{\mathbf{h}(\boldsymbol{\theta}_1),\mathbf{h}(\boldsymbol{\theta}_2),\ldots,\mathbf{h}(\boldsymbol{\theta}_p)\}}{\arg\min} \frac{|\mathbf{h}^H\mathbf{R}_{\mathbf{zz}}(\boldsymbol{\theta})\mathbf{h}|}{\|\mathbf{h}\|_2^2}. \quad (13)$$

Event G is the event that the energy resolved in the apriori minimum mode $\mathbf{h}_{min}$ of the noise-free signal subspace $\langle\mathbf{H}\rangle$ (or equivalently $\langle\mathbf{U}_p\rangle$) is smaller than the average energy resolved in the orthogonal subspace $\langle\mathbf{U}_0\rangle$. Clearly, G is a subset of $E(1)$, and therefore a subset of E, as the following argument shows

$$\begin{aligned}\min_{1\leq i\leq p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{h}_i}\mathbf{W}) &\leq \mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{h}_{min}}\mathbf{W})\\ &< \frac{1}{m-p}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0}\mathbf{W})\\ &\leq \max_{p+1\leq i\leq m}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{u}_i}\mathbf{W})\\ &\leq \max_{\mathbf{b}\in\mathbb{C}^{(n-p)\times 1}}\mathrm{tr}(\mathbf{W}^H\mathbf{P}_{\mathbf{U}_0\mathbf{b}}\mathbf{W}). \quad (14)\end{aligned}$$

We use the apriori minimum mode $\mathbf{h}_{min}$ of the noise-free signal subspace in the definition of event G, because based on the measurement models established in Section II, and the definition of $\mathbf{h}_{min}$ in (12) and (13), $\mathbf{h}_{min}$ is most likely to be swapped with one of the components of the noise subspace. This gives us the tightest lower bound compared to other choices of $\mathbf{h}(\theta_i)$.

IV. Bounds on the Probability of a Subspace Swap After Compression

Because events F and G are subsets of event E, their probabilities of occurrence $P(F)$ and $P(G)$ give lower bounds on the probability of a subspace swap, $P_{ss} \triangleq P(E)$. In this section, we derive analytical expressions for $P(F)$ and $P(G)$, in terms of tail probabilities of F-distributed random variables, for the two data models given in Section II. These probabilities lower bound P_{ss}, that is, $P_{ss} \geq P(F)$ and $P_{ss} \geq P(G)$.

Let us define T_F and T_G as

$$\mathbf{T}_F = \frac{1}{m-p}\mathbf{P}_{\mathbf{U}_0} - \frac{1}{p}\mathbf{P}_{\mathbf{U}_p}. \quad (15)$$

and

$$\begin{aligned}\mathbf{T}_G &= \frac{1}{m-p}\mathbf{P}_{\mathbf{U}_0} - \mathbf{P}_{\mathbf{h}_{min}}\\ &= \frac{1}{m-p}\mathbf{P}_{\mathbf{U}_0} - \boldsymbol{\rho}_{min}\boldsymbol{\rho}_{min}^H, \quad (16)\end{aligned}$$

where $\mathbf{P}_{\mathbf{h}_{min}} = \boldsymbol{\rho}_{min}\boldsymbol{\rho}_{min}^H$ with $\boldsymbol{\rho}_{min} = \frac{\mathbf{h}_{min}}{\|\mathbf{h}_{min}\|_2}$.

Then, looking at the definitions of event F and event G in (9) and (11), we can write $P(F)$ and $P(G)$ as

$$P(F) = P(\mathrm{tr}\left[\mathbf{W}^H\mathbf{T}_F\mathbf{W}\right] > 0)$$

and

$$P(G) = P(\mathrm{tr}\left[\mathbf{W}^H\mathbf{T}_G\mathbf{W}\right] > 0).$$

A. Parameterized Mean Case

For the parameterized mean measurement model discussed in Section II.A, using event F for bounding, we have

$$\begin{aligned}P_{ss} &\geq P(F)\\ &= P(\mathrm{tr}[\mathbf{W}^H\mathbf{T}_F\mathbf{W}] > 0)\\ &= P\left(\frac{\mathrm{tr}\left[\mathbf{W}^H\mathbf{U}_p\mathbf{U}_p^H\mathbf{W}\right]/2p}{\mathrm{tr}\left[\mathbf{W}^H\mathbf{U}_0\mathbf{U}_0^H\mathbf{W}\right]/2(m-p)} < 1\right)\\ &= P\left(\frac{\sum_{i=1}^M\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/2p}{\sum_{i=1}^M\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)} < 1\right). \quad (17)\end{aligned}$$

Here, the elements of $\{\mathbf{U}_p^H\mathbf{w}_i\}_{i=1}^M$ are independent and identically distributed as

$$\mathbf{U}_p^H\mathbf{w}_i \sim \mathcal{CN}_p(\mathbf{U}_p^H\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I}) \quad \forall 1\leq i\leq M. \quad (18)$$

Therefore, using (18), and the fact that $\mathbf{z}(\boldsymbol{\theta}) \in \langle\mathbf{U}_p\rangle$, we have

$$\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/\sigma^2 \sim \chi_{2p}^2(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2) \quad \forall 1\leq i\leq M, \quad (19)$$

which is the distribution of a complex noncentral chi-squared random variable with $2p$ degrees of freedom and noncentrality parameter $\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2$. Also, since $\langle\mathbf{U}_p\rangle$ and $\langle\mathbf{U}_0\rangle$ are orthogonal, we can conclude that in (17), each $\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/\sigma^2$ is independent of $\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/\sigma^2$ and is distributed as $\chi_{2(m-p)}^2$. Hence, the term $\frac{\sum_{i=1}^M\|\mathbf{U}_p^H\mathbf{w}_i\|_2^2/2p}{\sum_{i=1}^M\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)}$ is the ratio of two independent normalized chi-squared random variables and is distributed as $F_{2pM,2(m-p)M}(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2)$, which is a noncentral F distribution with $2pM$ and $2(m-p)M$ degrees of freedom and noncentrality parameter $\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2$. Thus, the probability of a subspace swap after compression is lower bounded by the probability that a $F_{2pM,2(m-p)M}(\|\mathbf{z}(\boldsymbol{\theta})\|_2^2/\sigma^2)$ distributed random variable is less than 1. When there is no compression, this lower bound turns into the probability that a $F_{2pM,2(n-p)M}(\|\mathbf{x}(\boldsymbol{\theta})\|_2^2/\sigma^2)$ random variable is less than 1.

Using event G for bounding, we have

$$\begin{aligned}P_{ss} &\geq P(G)\\ &= P(\mathrm{tr}[\mathbf{W}^H\mathbf{T}_G\mathbf{W}] > 0)\\ &= P\left(\frac{\mathrm{tr}\left[\mathbf{W}^H\boldsymbol{\rho}_{min}\boldsymbol{\rho}_{min}^H\mathbf{W}\right]/2}{\mathrm{tr}\left[\mathbf{W}^H\mathbf{U}_0\mathbf{U}_0^H\mathbf{W}\right]/2(m-p)} < 1\right)\\ &= P\left(\frac{\sum_{i=1}^M\|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/2}{\sum_{i=1}^M\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)} < 1\right). \quad (20)\end{aligned}$$

Here, the elements of $\{\boldsymbol{\rho}_{min}^H\mathbf{w}_i\}_{i=1}^M$ are independent and identically distributed as

$$\boldsymbol{\rho}_{min}^H\mathbf{w}_i \sim \mathcal{CN}(\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta}), \sigma^2\mathbf{I}) \quad \forall 1\leq i\leq M. \quad (21)$$

Therefore, $\|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/\sigma^2 \sim \chi_2^2(|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2)$ which is the distribution of a complex noncentral chi-squared random variable with 2 degrees of freedom and noncentrality parameter $|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2$. Thus, with the same type of arguments as for event F, we can conclude that the term $\frac{\sum_{i=1}^M\|\boldsymbol{\rho}_{min}^H\mathbf{w}_i\|_2^2/2}{\sum_{i=1}^M\|\mathbf{U}_0^H\mathbf{w}_i\|_2^2/2(m-p)}$ is distributed as $F_{2M,2(m-p)M}(|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2)$, which is a noncentral F distribution with $2M$ and $2(m-p)M$ degrees of freedom and noncentrality parameter $|\boldsymbol{\rho}_{min}^H\mathbf{z}(\boldsymbol{\theta})|^2/\sigma^2$. When

there is no compression, this turns into the probability that a $F_{2M,2(n-p)M}(|\boldsymbol{\kappa}_{min}^H \mathbf{x}(\boldsymbol{\theta})|^2/\sigma^2)$ random variable is less than 1. Here, $\boldsymbol{\kappa}_{min} = \frac{\mathbf{k}_{min}}{\|\mathbf{k}_{min}\|_2}$, and $\mathbf{k}_{min}$ is the apriori minimum mode of the signal subspace before compression.

Remark 3: The lower bounds derived in this section can guide our understanding of threshold effects and performance breakdowns for parameter estimation from compressed noisy data in the measurement model of Section II.A. In particular, they can be used to quantify the increase in threshold SNR as a function of a compression ratio as we demonstrate numerically in Section V. This can in turn be used to decide whether or not compression to a particular dimension fits performance constraints in a desired application.

B. Parameterized Covariance Case

For the parameterized covariance measurement model discussed in Section II.B, the columns of the measurement matrix $\mathbf{W}$ are i.i.d. random vectors distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R_{ww}})$. Using event F for bounding, we have

$$\begin{aligned} P_{ss} &\geq P(F) \\ &= P(\mathrm{tr}[\mathbf{W}^H \mathbf{T}_F \mathbf{W}] > 0) \\ &= P\left(\frac{\mathrm{tr}\left[\mathbf{W}^H \mathbf{U}_p \mathbf{U}_p^H \mathbf{W}\right]/2p}{\mathrm{tr}\left[\mathbf{W}^H \mathbf{U}_0 \mathbf{U}_0^H \mathbf{W}\right]/2(m-p)} < 1 \right) \\ &= P\left(\frac{\sum_{i=1}^M \|\mathbf{U}_p^H \mathbf{w}_i\|_2^2/2p}{\sum_{i=1}^M \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2/2(m-p)} < 1 \right). \end{aligned} \tag{22}$$

Here, the elements of $\{\mathbf{U}_p^H \mathbf{w}_i\}_{i=1}^M$ are independent and identically distributed as

$$\mathbf{U}_p^H \mathbf{w}_i \sim \mathcal{CN}_p(\mathbf{0}, \boldsymbol{\Lambda}_p + \sigma^2 \mathbf{I}_p) \quad \forall 1 \leq i \leq M. \tag{23}$$

Therefore we can write

$$\|\mathbf{U}_p^H \mathbf{w}_i\|_2^2 = \sum_{i=1}^p (\lambda_i + \sigma^2)\rho_i, \tag{24}$$

where ρ_i's are i.i.d. random variables, each distributed as χ_2^2. Therefore,

$$\sum_{i=1}^M \|\mathbf{U}_p^H \mathbf{w}_i\|_2^2 = \sum_{i=1}^p (\lambda_i + \sigma^2)\xi_i, \tag{25}$$

where ξ_i's are i.i.d. random variables, each distributed as χ_{2M}^2. Also, we can write $\sum_{i=1}^M \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2 = \sigma^2 \nu$, where ν is distributed as $\chi_{2M(m-p)}^2$ and is independent of the ξ_i's. Therefore, following the sequence of identities in (22), we obtain

$$\begin{aligned} P_{ss} &\geq P(F) \\ &= P\left(\frac{\sum_{i=1}^p (1 + \lambda_i/\sigma^2)\xi_i/2Mp}{\nu/2M(m-p)} < 1 \right). \end{aligned} \tag{26}$$

Here, the term $\frac{\sum_{i=1}^p (1+\lambda_i/\sigma^2)\xi_i/2Mp}{\nu/2M(m-p)}$ is distributed as $GF[(1+\frac{\lambda_1}{\sigma^2}), \ldots, (1+\frac{\lambda_p}{\sigma^2}); 2M; 2M(m-p)]$, which is the distribution of a generalized F random variable [34]. Thus, the probability of a subspace swap in this case is lower bounded by the probability that a $GF[(1+\frac{\lambda_1}{\sigma^2}), \ldots, (1+\frac{\lambda_p}{\sigma^2}); 2M; 2M(m-p)]$ random variable is less than 1. Without compression, this turns into the probability that a $GF[(1+\frac{\tilde{\lambda}_1}{\sigma^2}), \ldots, (1+\frac{\tilde{\lambda}_p}{\sigma^2}); 2M; 2M(n-p)]$ random variable is less than 1. Here $\tilde{\lambda}_i$'s are the eigenvalues of the signal covariance matrix $\mathbf{R_{xx}}$ before compression.

Using event G for bounding, we have

$$\begin{aligned} P_{ss} &\geq P(G) \\ &= P(\mathrm{tr}[\mathbf{W}^H \mathbf{T}_G \mathbf{W}] > 0) \\ &= P\left(\frac{\sum_{i=1}^M \|\boldsymbol{\rho}_{min}^H \mathbf{w}_i\|_2^2/2}{\sum_{i=1}^M \|\mathbf{U}_0^H \mathbf{w}_i\|_2^2/2(m-p)} < 1 \right), \end{aligned} \tag{27}$$

Here, the elements of $\{\boldsymbol{\rho}_{min}^H \mathbf{w}_i\}_{i=1}^M$ are independent and identically distributed as

$$\boldsymbol{\rho}_{min}^H \mathbf{w}_i \sim \mathcal{CN}(0, \tau) \quad \forall 1 \leq i \leq M, \tag{28}$$

where $\tau = \boldsymbol{\rho}_{min}^H \mathbf{R_{ww}} \boldsymbol{\rho}_{min}$. Therefore,

$$\sum_{i=1}^M \|\boldsymbol{\rho}_{min}^H \mathbf{w}_i\|_2^2/\tau \sim \chi_{2M}^2. \tag{29}$$

From (27) and (29), we have

$$\begin{aligned} P_{ss} &\geq P(G) \\ &= P\left(\vartheta < \frac{\sigma^2}{\tau} \right), \end{aligned} \tag{30}$$

where ϑ is distributed as $F_{2M,2M(m-p)}$, which is a central F random variable with $2M$ and $2M(m-p)$ degrees of freedom. Without compression, this turns into the probability that a $F_{2M,2M(n-p)}$ random variable is less than $\frac{\sigma^2}{\tilde{\tau}}$, where $\tilde{\tau} = \boldsymbol{\kappa}_{min}^H \mathbf{R_{yy}} \boldsymbol{\kappa}_{min}$, $\boldsymbol{\kappa}_{min} = \frac{\mathbf{k}_{min}}{\|\mathbf{k}_{min}\|_2}$, and $\mathbf{k}_{min}$ is the apriori minimum mode of the signal subspace before compression.

Remark 4: In Sections IV.A and IV.B, we have derived lower bounds on the probability of a subspace swap for the case that $\boldsymbol{\Psi} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1/2}\boldsymbol{\Phi}$ is deterministic, as in standard or co-prime subsamplings. In the case that $\boldsymbol{\Psi}$ is random, these probability bounds would have to be integrated over the distribution of $\boldsymbol{\Psi}$ to give lower bounds on marginal probabilities of a subspace swap. For example, for random $\boldsymbol{\Psi}$ and for the subevent F we have

$$P_{ss} = \int P(E|\boldsymbol{\Psi}) P(\boldsymbol{\Psi}) d\boldsymbol{\Psi} \geq \int P(F|\boldsymbol{\Psi}) P(\boldsymbol{\Psi}) d\boldsymbol{\Psi} \tag{31}$$

where $P(F|\boldsymbol{\Psi})$ is given in Sections IV.A and IV.B for the parameterized mean and parameterized covariance measurement models, respectively. For the class of random compression matrices that have density functions of the form $g(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)$, that is, the distribution of $\boldsymbol{\Phi}$ is right orthogonally invariant, $\boldsymbol{\Psi}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_m(\mathbb{C}^n)$ [35]. The compression matrix $\boldsymbol{\Phi}$ whose elements are i.i.d. standard normal random variables is one such matrix.

Remark 5: The lower bounds derived in this section can guide our understanding of threshold effects and performance breakdowns for parameter estimation from compressed noisy data in measurement model of Section II.B. They can be used to quantify the increase in threshold SNR as a function of a compression ratio as we demonstrate numerically in Section V. This can in turn be used to decide whether or not compression to a particular dimension fits performance constraints in a desired application.

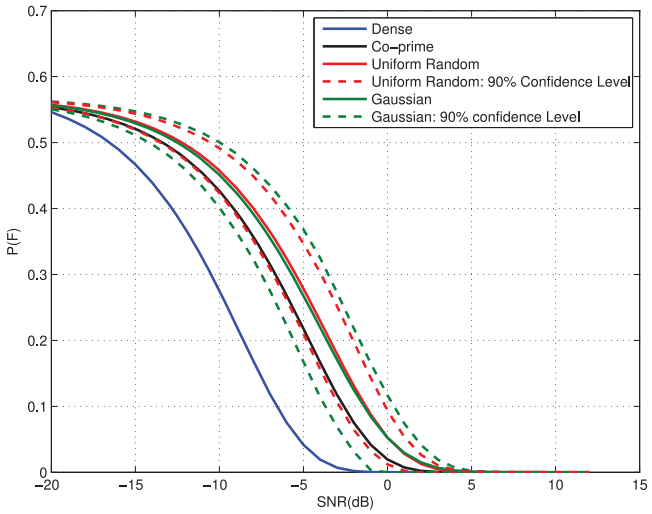
V. Simulation Results

In this section, we present numerical examples to show the impact of compression on threshold effects for estimating directions of arrival using a sensor array. We consider a dense uniform line array with n elements at half-wavelength inter-element spacings, and three different compressions of this array: Gaussian, uniform random, and co-prime. In the Gaussian compression, we compress the measurements of the dense array using an $m \times n$ left unitary matrix $\mathbf{\Psi} = (\mathbf{\Phi}\mathbf{\Phi}^H)^{-1/2}\mathbf{\Phi}$, where the elements of $\mathbf{\Phi}$ are i.i.d. random variables distributed as $\mathcal{N}(0, 1/m)$. In the uniform random compression, we consider subarrays of dimension m of the dense array, drawn uniformly at random, where the subarrays have similar apertures as the dense array. In the co-prime compression, we uniformly subsample the dense array once by a factor m_1 and once by a factor m_2, where m_1 and m_2 are co-prime. We then interleave these two subarrays to form the co-prime array of $m = m_1 + 2m_2 - 1$ elements. We choose n, m_1, and m_2 such that the aperture of the co-prime array equals the aperture of the dense array. We consider two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$. The Rayleigh limit of the dense array in electrical angle is $2\pi/n$. Therefore, in our examples the two sources are separated by half the Rayleigh limit of the dense array.

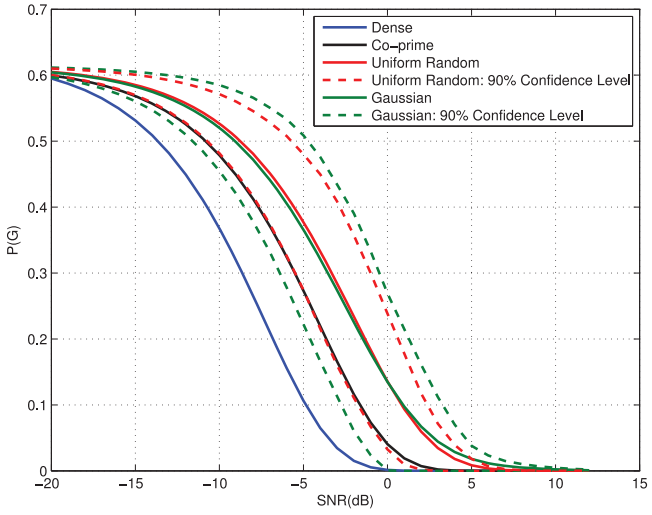
Remark 6: Gaussian compression (see, e.g., [7]) is impractical in sensor array processing as it requires laying down n uniformly spaced sensors to form $m < n$ linear combinations of them. We consider this compression scheme only for comparison with uniform random and co-prime subsamplings. Co-prime sensor array processing was introduced recently in [8]–[12] as a sparse alternative to uniform line arrays. Measurements in a co-prime array may be used directly for inference, or they may be used indirectly to fill in the entries of a Toeplitz covariance matrix in a difference co-array. The former case has been studied in [11] and [12], where the authors establish identifiability theorems for identifying undamped modes using MUSIC. The indirect methodology applies when the modes to be identified are undamped, mode amplitudes are uncorrelated, and the scene to be imaged remains stationary long enough for temporal averaging of snapshots to produce reliable estimates of covariances that may then be loaded into the Toeplitz covariance matrix for the difference co-array. There are two potential uses for co-prime arrays. In one of these, the number of sensors in the co-prime array equals the number of sensors in the ULA, but the aperture of the co-prime array is much larger than the aperture of the ULA. There are no consequences for effective SNR in this case, and the Cramér-Rao bound (CRB) and the Fisher information matrix for DOA estimation favor the co-prime array, as shown in [11] and [12]. In the other use, the aperture of the co-prime array is specified to equal the aperture of a ULA, but many fewer sensors are required for the co-prime array than would be used in the ULA. This is the case we have considered in the paper. In this case, there are degradations in the Fisher information and the CRB (see [36]), and in the threshold SNR for the co-prime array.

A. Parameterized Mean Case

Fig. 2 shows the lower bounds $P(F)$ and $P(G)$ for dense, co-prime, uniform random, and Gaussian compressed arrays.



(a)



(b)

Fig. 2. Parameterized mean case. Analytical lower bounds, (a) $P(F)$ and (b) $P(G)$, for the probabilities of subspace swaps in dense and compressed arrays. The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$, with amplitudes $\alpha_1 = \alpha_2 = 1$. The dense array has $n = 36$ elements and the compressed arrays have $m = 12$ elements, covering similar apertures as the dense array. The mean and confidence level curves for Gaussian and uniform random compressions are generated from 200 trials.

The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$, with amplitudes $\alpha_1 = \alpha_2 = 1$. The dense array has $n = 36$ elements and the compressed arrays have $m = 12$ elements, covering similar apertures as the dense array. The compression ratio is $\frac{n}{m} = 3$. For the uniform random and Gaussian compressed arrays, we show the bounds obtained by averaging our lower bounds for 200 draws of the random compression matrix, as well as their corresponding 90% confidence levels. These results indicate that bounding event G is a slightly tighter bounding event in the parameterized mean case.

Fig. 3 shows the difference in threshold SNRs, between dense and compressed arrays, at $P(F) = 0.1$ for different m

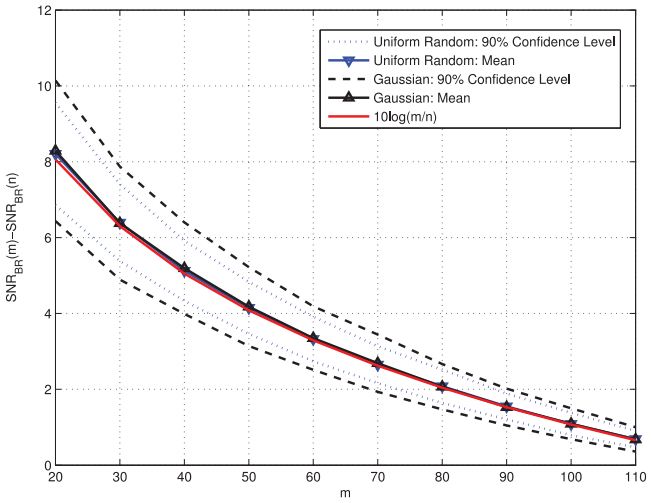


Fig. 3. Parameterized mean case. Difference in threshold SNRs, between dense and compressed arrays, at $P(F) = 0.1$ for different m values and $n = 128$. The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$, with amplitudes $\alpha_1 = \alpha_2 = 1$. The mean and confidence level curves for Gaussian and uniform random compressions are generated from 1000 trials.

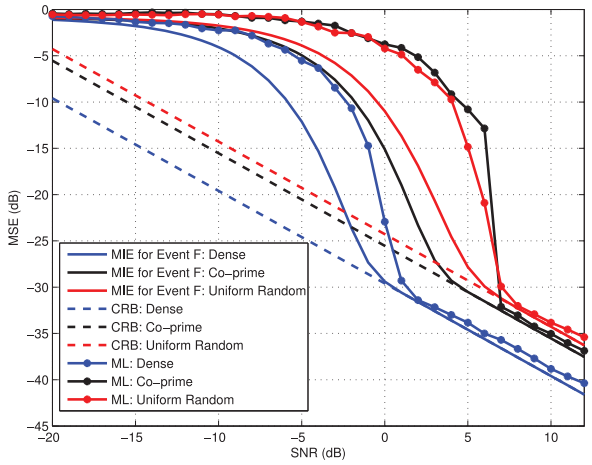


Fig. 4. Parameterized mean case. Comparisons of CRBs, MSEs for ML, and approximate MSEs (using the method of interval errors) in estimating $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/n$, using an $n = 36$ element dense array and an $m = 12$ element co-prime array. The data consists of 200 realizations.

values and $n = 128$.[1] Again, the modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$, with amplitudes $\alpha_1 = \alpha_2 = 1$. The plots show the average and the 90% confidence levels of $SNR_{Br}(m) - SNR_{Br}(n)$, obtained over 1000 trials, where $SNR_{Br}(m)$ is the SNR value at which $P(F)$ for an m-element array equals 0.1. As Fig. 3 shows, the average loss in breakdown threshold for the compressed arrays of size m relative to a dense array of size n is approximately $10\log_{10} n/m$ dB. These results hold as well for event G.

As Figs. 2 and 3 show, our analytical results for the uniform random and Gaussian compressed arrays are similar, and we expect the same MSE behavior in DOA estimation for these two arrays. Therefore, we are going to drop the Gaussian compression in our numerical experiments for the MSE.

Fig. 4 shows the MSE for the maximum likelihood estimator of the source at θ_1 in the presence of the interfering source at θ_2, using the dense and compressed arrays. The CRBs corresponding to the $(n = 36)$-element dense array and the $(m = 12)$-element compressed arrays are also shown in this figure as references for performance analysis. The MSE for the uniform random scheme is calculated by averaging the MSEs for different realizations. Fig. 4 also shows approximations to the MSE (in solid lines) obtained using the *method of interval errors* (introduced in [37] and used in [2]). At each SNR, the approximate MSE σ_T^2 is computed as

$$\sigma_T^2 = P_{ss}\sigma_0^2 + (1 - P_{ss})\sigma_{CR}^2. \tag{32}$$

Here, P_{ss} is the probability of the subspace swap as a function of SNR, which we approximate using the lower bound $P(F)$ in (17), normalized by its value at extremely low SNRs where a subspace swap occurs with probability 1; σ_{CR}^2 is the value of the CRB as a function of SNR, and σ_0^2 is the variance of the error given the occurrence of a subspace swap. The justification for using this formula is that when a subspace swap does not occur, MSE almost follows the CRB. However, given the occurrence of the subspace swap (and in the absence of any prior knowledge) the error in estimating the electrical angle θ_1 may be taken to be uniformly distributed between $(-\pi/2, \pi/2)$ and the error variance is $\sigma_0^2 = \pi^2/12$.

The utility of the lower bounds in Fig. 2 is in understanding the increase in the threshold SNR, due to compression. This is validated by the MSE curves of Fig. 4, where the SNR gap between the onset of MSE breakdown for dense (blue solid-circled) and co-prime (black solid-circled) arrays is roughly $10\log_{10} n/m$ dB. The SNR gap between the onsets of MSE breakdowns for dense and uniform random (red solid-circled) arrays is roughly the same. This is in agreement with our analytical predictions in Figs. 2 and 3. This performance loss is consistent with the performance loss (e.g., in Fisher information) that one would expect when measurements in white Gaussian noise are compressed by a factor n/m.

Fig. 5 further studies the threshold effect. It shows the empirical outlier probability versus SNR for the maximum likelihood estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$ for 200 realizations. An outlier is considered an ML estimate with a large error. Our lower bounds $P(F)$ for the probabilities of subspace swaps for dense, co-prime, and uniform random arrays are overlaid on this figure. For the uniform random array, the plotted bound is obtained by averaging the lower bounds $P(F)$ over 200 draws of the random compression matrix. We observe that our lower bounds underestimate their corresponding empirical outlier probabilities. However, they correctly predict the gaps in threshold SNRs, based on outlier probability.

Remark 7: In the parameterized mean case, the maximum likelihood estimation (MLE) of the DOAs is obtained as

$$\boldsymbol{\theta}_{ML} = \underset{\boldsymbol{\theta}=[\theta_1,\theta_2]}{\arg\max}\, \mathbf{y}^H \mathbf{P}_{\mathbf{H}(\boldsymbol{\theta})}\mathbf{y}, \tag{33}$$

[1]We use $n = 128$ to have more freedom in generating different compression ratios by changing m. Also, since co-prime arrays of fixed aperture are not realizable for arbitrary number of sensors m, we are not using co-prime arrays in this experiment.

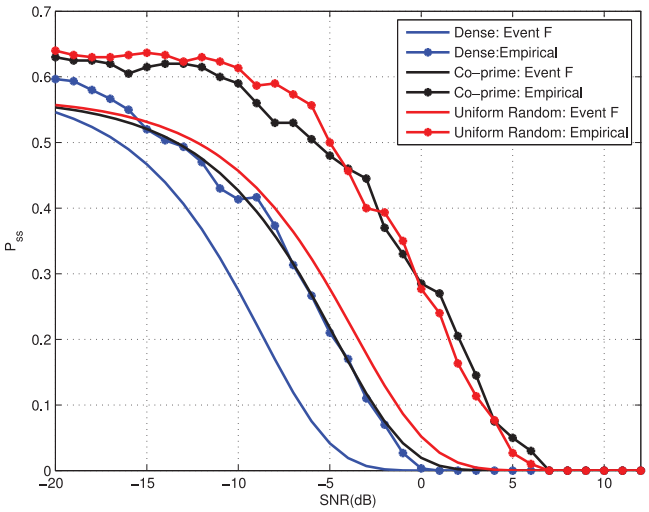


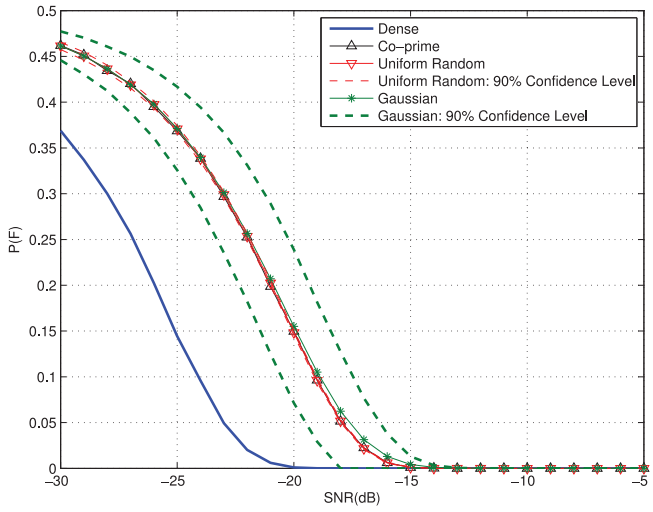
Fig. 5. Parameterized mean case. Analytical lower bounds $P(F)$ and empirical outlier probabilities for estimation of $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/36$. The dense array has $n = 36$ elements and the compressed arrays have $m = 12$ elements, covering similar apertures as the dense array.

where $\mathbf{y}$ is the measurement vector of the sensors, and $\mathbf{H}(\boldsymbol{\theta})$ is the matrix whose columns are the steering vectors associated with the DOAs θ_1 and θ_2. We are showing the MSE of the maximum likelihood estimation method because as (33) shows, the maximum likelihood estimation of the DOAs involves the maximization of the energy of the projection of sensor measurements into the noise-free signal subspace. Therefore, the occurrence of a subspace swap simply leads to a performance breakdown in the MLE method. Furthermore, as it is shown in [4], while a subspace swap is responsible for the performance breakdown of the MLE, earlier breakdown of subspace based methods can be attributed to the loss of resolution in resolving closely spaced modes. Therefore, we do not show the MSE plots for any of the subspace methods such as MUSIC and ESPRIT.

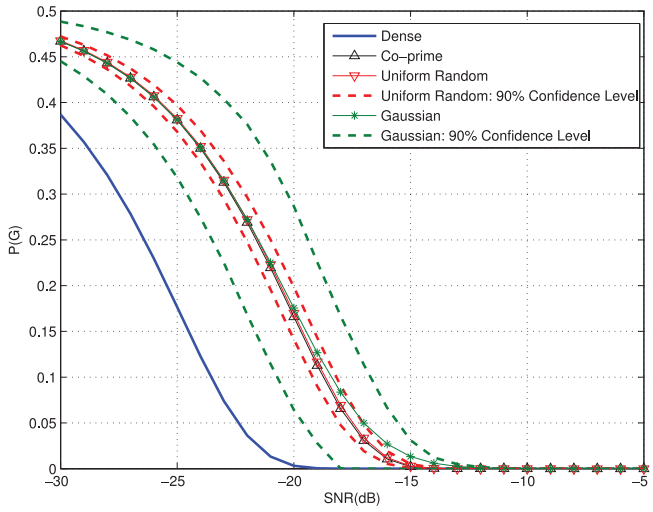
B. Parameterized Covariance Case

Fig. 6 shows the lower bounds $P(F)$ and $P(G)$ in dense, co-prime, uniform random, and Gaussian compressed arrays. The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$. The source amplitudes are assumed to be random with an identity covariance matrix. The dense array has $n = 36$ elements and the compressed arrays have $m = 12$ elements, covering similar apertures as the dense array, giving a compression ratio of $n/m = 3$. The mean and confidence level curves for Gaussian and uniform random compressions are generated from 200 trials. Again, the gaps between the threshold SNRs for the dense and the compressed arrays is approximately $10\log_{10} n/m$ dB.

Fig. 7 shows the difference in threshold SNRs, between dense and compressed arrays, at $P(G) = 0.1$ for different m values and $n = 128$. Again, we have used $n = 128$ to have more freedom in generating different compression ratios by changing m, and we have dropped co-prime arrays for the reason mentioned in the parameterized mean case. Similar to the previous plot, the modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$ and the source amplitudes are assumed to be random with an identity covariance matrix. The plots show the average and the 90% confidence levels



(a)



(b)

Fig. 6. Parameterized covariance case. Analytical lower bounds, (a) $P(F)$ and (b) $P(G)$, for the probabilities of subspace swaps in dense and compressed arrays. The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$. The source amplitudes are assumed to be random with an identity covariance matrix. The dense array has $n = 36$ elements and the compressed arrays have $m = 12$ elements, covering similar apertures as the dense array. The mean and confidence level curves for Gaussian and uniform random compressions are generated from 200 trials.

of $SNR_{Br}(m) - SNR_{Br}(n)$, obtained over 1000 trials, where $SNR_{Br}(m)$ is the SNR value at which $P(G)$ for an m-element array equals 0.1. As Fig. 7 shows, the average loss in breakdown threshold for the compressed arrays of size m relative to a dense array of size n is approximately $10\log_{10} n/m$ dB.

Fig. 8 shows the results for the MSE of the maximum likelihood estimator of the source at $\theta_1 = 0$ in the presence of the interfering source at $\theta_2 = \pi/36$. Our approximations for the MSE using the *method of interval errors* in (32) and the Cramér-Rao bound are also shown for each array. Similar to the parameterized mean case, Figs. 6, 7 and 8 show that performance loss, measured by onset of threshold effect for the compressed arrays is approximately $10\log_{10} n/m$ dB.

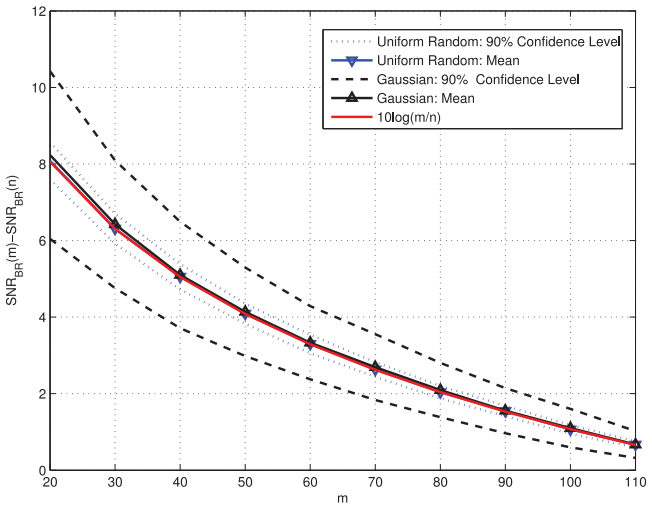


Fig. 7. Parameterized covariance case. Difference in threshold SNRs, between dense and compressed arrays, at $P(G) = 0.1$ for different m values and $n = 128$. The modes correspond to two point sources at far field at electrical angles $\theta_1 = 0$ and $\theta_2 = \pi/n$. The source amplitudes are assumed to be random with an identity covariance matrix. The mean and confidence level curves for Gaussian and uniform random compressions are generated from 1000 trials.

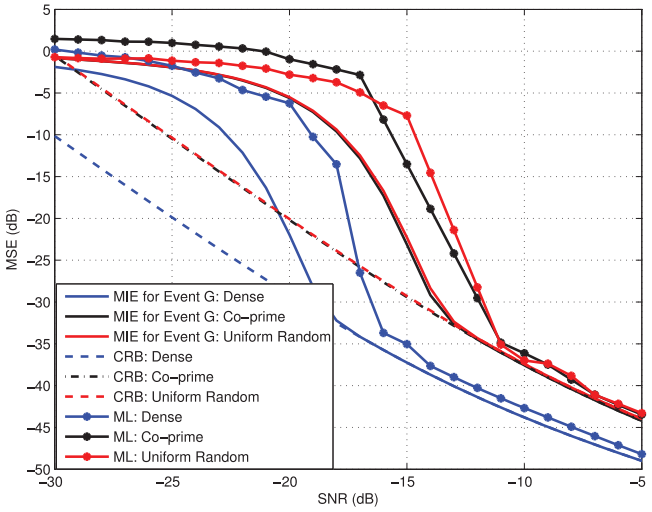


Fig. 8. Parameterized covariance case. Comparisons of CRBs, MSEs for ML, and approximate MSEs (using the method of interval errors) in estimating $\theta_1 = 0$ in the presence of an interfering source at $\theta_2 = \pi/n$, using an $n = 36$ element dense array and an $m = 12$ element co-prime array. The MSE curves for ML are averaged over 200 trails, each with 200 snapshots.

Remark 8: In [3], the authors derive predictions for the probability of a subspace swap in parameterized covariance case, based on asymptotic reasoning about the eigenvalues and eigenvectors of a sample covariance matrix. In some finite snapshot regimes, the approximations of [3] are tighter than the bounds of this paper. In other cases they are neither tighter nor do they lower bound the probability of a subspace swap.

VI. Conclusion

We have addressed the effect of compression on performance breakdown associated with swapping of signal and noise subspaces in estimating signal parameters from compressed noisy data. We have derived analytical bounds on this probability for two measurement models. In the first-order model, the parameters modulate the mean of a set of complex i.i.d. multivariate normal measurements. In the second-order model, the parameters to be estimated modulate a covariance matrix. Our lower bounds take the form of tail probabilities of F-distributions, and they can be used with any linear compression scheme. The ambient dimension n, the compressed dimension m, the number p of parameters, and the number M of snapshots determine the degrees of freedom of the F-distributions. The choice of the compression matrix affects the non-centrality parameter of the F-distribution in the parameterized mean case, and the left tail probability of a central F (or a generalized F) distribution in the parameterized covariance case. The derived bounds are not asymptotic and are valid in finite snapshot regimes. They can be used to quantify the increase in threshold SNR as a function of a compression ratio C. We have demonstrated numerically that this increase in threshold SNR is roughly $10 \log_{10} C$ dB, which is consistent with the performance loss that one would expect when measurements in Gaussian noise are compressed by a factor C. As a case study, we have investigated threshold effects in maximum likelihood (ML) estimation of directions of arrival of two closely-spaced sources using co-prime subsampling and uniformly at random subsampling. Our MSE plots validate the increases in threshold SNRs.

References

[1] D. Tufts, A. Kot, and R. Vaccaro, "The threshold effect in signal processing algorithms which use an estimated subspace," in *SVD and Signal Processing II: Algorithms, Analysis and Applications*. New York, NY, USA: Elsevier, 1991, pp. 301–320.

[2] J. K. Thomas, L. L. Scharf, and D. W. Tufts, "The probability of a subspace swap in the SVD," *IEEE Trans. Signal Process.*, vol. 43, no. 3, pp. 730–736, 1995.

[3] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: Prediction and cure," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Salt Lake City, UT, USA, May 2001, pp. 4005–4008.

[4] 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, Aug. 2008.

[5] E. J. Candés, "Compressive sampling," in *Proc. Int. Congr. Math.*, 2006, vol. 3, pp. 1433–1452.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[8] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proc. Digit. Signal Process. Workshop/IEEE Signal Process. Educ. Workshop (DSP/SPE)*, Jan. 2011, vol. 51, pp. 3289–294.

[9] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[10] P. P. Vaidyanathan and P. Pal, "Theory of sparse coprime sensing in multiple dimensions," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3592–3608, Aug. 2011.

[11] P. P. Vaidyanathan and P. Pal, "Direct-MUSIC on sparse arrays," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, Bangalore, India, Jul. 2012, pp. 1–5.

[12] P. P. Vaidyanathan and P. Pal, "Why does direct-MUSIC on sparse-arrays work?," in *Conf. Rec. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2013, pp. 2007–2011.

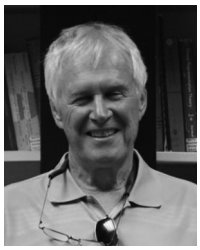
[13] E. J. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[14] E. J. Candés, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Math.*, vol. 346, no. 9, pp. 589–592, 2008.

[15] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive estimation in AWGN: General observations and a case study," in *Proc. 46th Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, Nov. 2012, pp. 953–957.

[16] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 6630–6634.
[17] P. Pakrooh, A. Pezeshki, L. L. Scharf, D. Cochran, and S. D. Howard, "Analysis of Fisher information and the Cramér-Rao bound for nonlinear parameter estimation after random compression," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6423–6428, Dec. 2015.
[18] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2820–2830, Nov. 2002.
[19] J. Liu, X. Liu, and X. Ma, "First-order perturbation analysis of singular vectors in singular value decomposition," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3044–3049, July 2008.
[20] J. Steinwandt, F. Roemer, M. Haardt, and G. Del Galdo, "R-dimensional ESPRIT-type algorithms for strictly second-order non-circular sources and their performance analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 18, pp. 4824–4838, Sep. 2014.
[21] M. Shaghaghi and S. Vorobyov, "Subspace leakage analysis and improved DOA estimation with small sample size," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3251–3265, 2015, doi: 10.1109/TSP.2015.2422675.
[22] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.
[23] A. Steinhardt and C. Bretherton, "Thresholds in frequency estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Tampa, FL, USA, Apr. 1985, pp. 1273–1276.
[24] B. G. Quinn and P. J. Kootsookos, "Threshold behaviour of the maximum likelihood estimator of frequency," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3291–3294, Nov. 1994.
[25] B. James, B. Anderson, and R. Williamson, "Characterization of threshold for single tone maximum likelihood frequency estimation," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 817–821, Apr. 1995.
[26] H. Wang and M. Kaveh, "On the performance characterization of signal-subspace processing," in *Proc. 19th Asilomar Conf. Circuits, Syst., Comput.*, Nov. 1985, pp. 73–77.
[27] R. J. Vaccaro, "A second-order perturbation expansion for the SVD," *SIAM J. Matrix Anal. Appl.*, vol. 15, no. 2, pp. 661–671, 1994.
[28] R. J. Vaccaro and A. C. Kot, "A perturbation theory for the analysis of SVD-based algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Apr. 1987, pp. 1613–1616.
[29] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.
[30] I. M. Johnstone, "High dimensional statistical inference and random matrices," in *Proc. Int. Congr. Math.*, 2007, vol. I, pp. 307–333.
[31] A. Edelman and N. R. Rao, "Random matrix theory," *Acta Numerica*, vol. 14, pp. 233–297, 2005.
[32] R. R. Nadakuditi and F. Benaych-Georges, "The breakdown point of signal subspace estimation," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Oct. 2010, pp. 177–180.
[33] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Adv. Math.*, vol. 227, no. 1, pp. 494–521, 2011.
[34] C. F. Dunkl and D. E. Ramirez, "Computation of the generalized F distribution," *Austral. New Zealand J. Statist.*, vol. 43, no. 1, pp. 21–31, 2001.
[35] Y. Chikuse, *Statistics on Special Manifolds*. New York, NY, USA: Springer, 2003.
[36] P. Pakrooh, L. L. Scharf, and A. Pezeshki, "Modal analysis using co-prime arrays," *IEEE Trans. Signal Process.*, Sep. 2015, DOI: 10.1109/TSP.2016.2521616, preprint.
[37] H. L. Van Trees , *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Melbourne, FL, USA: Krieger, 1992.

Pooria Pakrooh (S'13) received the B.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2008, and the M.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2011. He received his Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, Colorado, in 2015. He is currently a postdoctoral research associate with the Department of Electrical and Computer Engineering and the Department of Mathematics at Colorado State University. His current research interests include compressed sensing and statistical signal processing, and their applications in sensor and array signal processing.



Louis L. Scharf (S'67–M'69–SM'77–F'86–LF'07) received his Ph.D. from the University of Washington, Seattle. He is currently Research Professor of Mathematics and Emeritus Professor of Electrical and Computer Engineering, with a joint appointment in Statistics, at Colorado State University. He has served on the faculties of University of Rhode Island and University of Colorado, and held visiting positions at Ecole Superieure d'electricité, Ecole Superieure des Télécommunications, EURECOM, University of La Plata, Duke University, University of Wisconsin, University Tromsø, and University of Paderborn.

Professor Scharf's research interests are in statistical signal processing, as it applies to adaptive radar and sonar, electric power, cyber security, and wireless communication. He has authored three books: L. L. Scharf *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1991; L. L. Scharf, *A First Course in Electrical and Computer Engineering*, Addison-Wesley, 1998; and P. J. Schreier and L. L. Scharf, "*Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*, Cambridge University Press, 2010.

Professor Scharf was Technical Program Chair for the 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Tutorials Chair for ICASSP 2001; and Technical Program Chair for the Asilomar Conference on Signals, Systems, and Computers, 2002. He is past Chair of the Fellow Committee for the IEEE Signal Processing Society. He has received numerous awards for his contributions to statistical signal processing, including an IEEE Distinguished Lectureship; an IEEE Third Millennium Medal; the Donald W. Tufts Award for Underwater Acoustic Signal Processing, the Technical Achievement Award from the IEEE Signal Processing Society (SPS); and the Society Award from the IEEE SPS.

Ali Pezeshki (S'95–M'05) received the B.Sc. and M.Sc. degrees in electrical engineering from University of Tehran, Tehran, Iran, in 1999 and 2001, respectively. He received his Ph.D. degree in electrical engineering at Colorado State University in 2004. In 2005, he was a postdoctoral research associate with the Electrical and Computer Engineering Department at Colorado State University. From January 2006 to August 2008, he was a postdoctoral research associate with The Program in Applied and Computational Mathematics at Princeton University. In August 2008, he joined the faculty of Colorado State University, where he is now an Associate Professor in the Department of Electrical and Computer Engineering, and the Department of Mathematics. His research interests are in statistical signal processing, coding theory, applied harmonic analysis, and bioimaging.