Subspace Averaging and Order Determination for Source Enumeration

Vaibhav Garg, Ignacio Santamaria, David Ramírez and Louis L. Scharf

***Abstract*—In this paper we address the problem of subspace averaging, with special emphasis placed on the question of estimating the dimension of the average. The results suggest that the enumeration of sources in a multi-sensor array, which is a problem of estimating the dimension of the array manifold, and as a consequence the number of radiating sources, may be cast as a problem of averaging subspaces. This point of view stands in contrast to conventional approaches, which cast the problem as one of identifiying covariance models in a factor model. We present a robust formulation of the proposed order fitting rule based on majorization-minimization algorithms. A key element of the proposed method is to construct a bootstrap procedure, based on a newly proposed discrete distribution on the manifold of projection matrices, for stochastically generating subspaces from a function of experimentally-determined eigenvalues. In this way, the proposed subspace averaging (SA) technique determines the order based on the eigenvalues of an average projection matrix, rather than on the likelihood of a covariance model, penalized by functions of the model order. By means of simulation examples, we show that the proposed SA criterion is especially effective in high-dimensional scenarios with low sample support.**

***Keywords*—Array processing, dimension, Grassmann manifold, order estimation, source enumeration, subspace averaging**

I. Introduction

In many applications of statistical signal processing, high-dimensional data exhibits a low dimensional structure that admits a subspace representation. In pattern recognition and machine learning, for instance, discriminative features are typically obtained after a principal component analysis (PCA) stage that selects a subspace to explain a large fraction of the variance in the original data [1]. In computer vision, the set of images under different illuminations can be represented by a low-dimensional subspace [2]. And subspaces appear also as invariant representations of signals geometrically deformed under a set of affine transformations [3]. There are many more applications where low-dimensional subspaces capture the intrinsic geometry of the problem, ranging from array processing [4], motion segmentation [5], subspace clustering [6], spectrum sensing for cognitive radio [7], or noncoherent multiple-input multiple-output (MIMO) wireless communications [8], [9].

When input data are modeled as subspaces, possibly of different dimensions, a fundamental problem is to compute an average or central subspace and, more importantly, to determine the dimension of the average. When all subspaces have the same dimension, they are formally represented as points on the Grassmann manifold. Geodesic, or intrinsic, distances on the Grassmannian are measured by the arc length (l_2-norm of the vector of principal angles), and the average subspace according to this canonical distance metric is the Riemannian center of mass, also known as the Karcher mean [10]. The Karcher mean is typically found by iterative algorithms that map the subspaces to and from the tangent plane at a given point (using Exp and Log maps), which make them computationally costly [11], and in fact not computable if some subspaces lie outside the injectivity radius of a provisional average, in which case the log map is undefined. Another drawback of the intrinsic distance metric is that a unique optimal Karcher mean is not always guaranteed to exist [12].

As an alternative to the intrinsic mean of the manifold, Srivastava and Klassen proposed the extrinsic mean in [13], which uses a chordal distance metric in the ambient vector space defined as the squared Frobenius norm of the difference between projection matrices. Unlike the intrinsic mean, the extrinsic mean of a collection of subspaces is always unique, it is easy to compute, and can be used for subspaces that have different dimensions and therefore live in a union of Grassmannians. For these reasons, in this paper as in [14], we focus on the extrinsic distance to compare subspaces that might not live in the same manifold.

We address the problem of determining the optimal order of the low-dimension average subspace that minimizes an extrinsic distance measure between subspaces and their average. The solution to this problem provides the simple order fitting rule derived in [14] for minimizing the extrinsic mean-squared error between a collection of subspaces and their average. The order fitting rule uses a threshold test on the eigenvalues of the average projection matrix, and thus it is free of penalty terms or other tuning parameters commonly used by other model order estimation techniques. The proposed rule appears to be particularly well suited to problems involving

V. Garg and I. Santamaria are with the Department of Communications Engineering, University of Cantabria, Santander, Spain (e-mail: {vaibhav.garg ,i.santamaria@unican.es}@unican.es).

D. Ramírez is with the Department of Signal Theory and Communications, University Carlos III of Madrid, Leganés, Spain and with the Gregorio Marañón Health Research Institute, Madrid, Spain (e-mail: david.ramirez@uc3m.es).

L. L. Scharf is with the Departments of Mathematics and Statistics, Colorado State University, USA (e-mail: scharf@colostate.edu).

The work of V. Garg and I. Santamaria was supported by the Ministerio de Economía y Competitividad (MINECO) of Spain, and AEI/FEDER funds of the E.U., under grants TEC2016-75067-C4-4-R (CARMEN), TEC2015-69648-REDC, and BES-2017-080542. The work of D. Ramírez was supported by the Ministerio de Ciencia, Innovación y Universidades under grant TEC2017-92552-EXP (aMBITION), by the Ministerio de Ciencia, Innovación y Universidades, jointly with the European Commission (ERDF), under grants TEC2015-69868-C2-1-R (ADVENTURE) and TEC2017-86921-C2-2-R (CAIMAN), and by The Comunidad de Madrid under grant Y2018/TCS-4705 (PRACTICO-CM). The work of L. L. Scharf was supported in part by the US NSF under contract CISE-1712788. Parts of this paper were presented at the 2016 and 2018 Workshop on Statistical Signal Processing.

high-dimensional data and low sample support, such as the determination of the number of sources with a large array of sensors [15]–[17]: the so-called source enumeration problem. In this paper we generalize this result by showing that it may also be applied to minimization of the mean-squared error between subspaces and their average, even when the extrinsic distance between two subspaces is replaced by a monotone function of the distance. This makes the resulting order fitting rule robust to outliers in the sequence of projections to be averaged.

In multi-sensor signal processing, temporal snapshots are typically used to estimate a second-order spatial covariance matrix. The eigenvalues of this sample covariance matrix are used for detection and localization, using methods that are inspired variations on factor analysis. But the use of these eigenvalues for source enumeration is fraught with difficulties, as they scale with source powers and background noise levels, and this fact conflicts with the fact that the dimension of an array manifold is invariant to scale. So the fundamental problem is to extract from a sample covariance matrix a scale invariant subspace. In summary, our approach to find this subspace is this. We replace a large sensor array by an overlapping sequence of subarrays, as inspired by [18], [19], and extract a collection of subspaces from the measurements made in each such subarray. These subspaces are then averaged, and from this average the dimension of the array manifold is determined. We propose a method for extracting subspaces based on a bootstrap sampling for subspaces drawn from a distribution determined by eigenvalues. This method is quite different in philosophy from previously published methods, based on asymptotic formulas for the distribution of the eigenvalues of the sample covariance matrix.

The order fitting rule for subspace averaging (SA) was first published in [14], and an unrefined application to source enumeration was presented in [20]. In this paper, we extend and refine these papers to provide a common framework for SA and its application to source enumeration. The main contributions of this paper may be summarized as follows:

- We consider continuous and discrete distributions on the manifold of projection matrices as underlying distributions from which the measured collection of subspaces is a random draw. From this standpoint, the eigenvalues of the average projection matrix admit a probabilistic interpretation that enables a better understanding of the proposed order estimation rule.
- We propose a robust formulation of the problem to account for outliers within the set of measured subspaces. The standard extrinsic mean distance is replaced by a smooth concave function such as the l_1 norm or the Huber loss function that limits the effect of subspaces far away from the average. A majorization-minimization (MM) algorithm [21] is then used to find the minimizer of this robust distance measure, which, in turn, provides a robust order fitting rule.
- The application of SA techniques to source enumeration in [20] is enhanced by including a sampling mechanism to generate random subspaces based on the eigenstructure of the sample covariance matrix. When exploited jointly with the shift invariance property of uniform linear arrays (ULAs), this random sampling scheme enhances the performance of SA in high-resolution scenarios in comparison to the preliminary results presented in [20]. Further, the method is proven to provide a consistent estimate of the number of sources as the number of samples or the signal-to-noise-ratio grow.

The structure of the paper is as follows. In Section II we derive the order estimation rule for the average subspace using an extrinsic distance measure. As reported in [12], [14], an average of projection matrices, not itself a projection matrix, is the key quantity summarizing all information needed to solve this problem. We also present in this section a robust version of the SA problem and solve it using MM algorithms. Sec. III reviews uniform and non-uniform distributions on the Grassmannian, and proposes a new discrete distribution motivated by interpretating the eigenvalues of the average projection matrix as probabilities. The application of the SA technique to source enumeration in large uniform linear arrays is discussed in detail in Section IV. Section V evaluates the performance of the order fitting rule through numerical simulations, paying special attention to the application to source enumeration. Finally, the main conclusions are summarized in Section VI.

Notation: In this paper we use $\langle \mathbf{A} \rangle$ to denote a subspace of the complex vector space $\mathcal{C}^n$ spanned by the unitary frame $\mathbf{A}$, and $\mathbf{P}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^H$ denotes the orthogonal projection onto $\langle \mathbf{A} \rangle$. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian, respectively. The trace and Frobenius norm of a matrix $\mathbf{B}$ will be denoted, respectively, as $\mathrm{tr}(\mathbf{B})$ and $||\mathbf{B}||_F$. $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix whose diagonal is $\mathbf{a}$, and $\mathbf{I}_n$ denotes the identity matrix of size n. $\mathcal{CN}(0,1)$ denotes a complex Gaussian distribution with zero mean and unit variance, $\mathbf{x} \sim \mathcal{CN}_n(\mathbf{0}, \mathbf{R})$ denotes a complex Gaussian vector in $\mathcal{C}^n$ with zero mean and covariance $\mathbf{R}$. $\mathbb{S}(q,n)$ denotes the complex Stiefel manifold of orthonormal q-frames in $\mathcal{C}^n$, $\mathbb{G}(q,n)$ denotes the complex Grassman manifold of q-dimensional linear subspaces of the n-dimensional complex vector space $\mathcal{C}^n$, and $\mathbb{P}(q,n)$ denotes the set of all projection matrices of rank q.

II. Order Estimation via Subspace Averaging

A. Distances Between Subpaces

Let us consider two subspaces $\langle \mathbf{V} \rangle \in \mathbb{G}(q_V, n)$ and $\langle \mathbf{U} \rangle \in \mathbb{G}(q_U, n)$. Let $\mathbf{V} \in \mathcal{C}^{n \times q_V}$ be a matrix whose columns form a unitary basis for $\langle \mathbf{V} \rangle$. Then $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{q_V}$, and $\mathbf{P}_{\mathbf{V}} = \mathbf{V}\mathbf{V}^H$ is the idempotent orthogonal projection onto $\langle \mathbf{V} \rangle$. Notice that $\mathbf{P}_{\mathbf{V}}$ is a unique representation of $\langle \mathbf{V} \rangle$, whereas $\mathbf{V}$ is not unique, because if $\mathbf{G}$ is an arbitrary unitary $q_V \times q_V$ matrix, then $\mathbf{V}\mathbf{G}$ will be another representation of $\langle \mathbf{V} \rangle$ with orthonormal columns. In a similar way, we define $\mathbf{U}$ and $\mathbf{P}_{\mathbf{U}}$ for the subspace $\langle \mathbf{U} \rangle$.

To measure the distance between two subspaces we need the concept of principal angles, which is introduced in the following definition [22].

Definition 2.1: Let $\langle \mathbf{V} \rangle$ and $\langle \mathbf{U} \rangle$ be subspaces of $\mathcal{C}^n$ whose dimensionality satisfy $\dim(\langle \mathbf{V} \rangle) = q_V \geq \dim(\langle \mathbf{U} \rangle) = q_U \geq$

1. The principal angles $\theta_1, \ldots, \theta_{q_U} \in [0, \pi/2]$ between $\langle \mathbf{V} \rangle$ and $\langle \mathbf{U} \rangle$ are defined recursively by

$$\begin{aligned} \cos(\theta_k) = & \max_{\mathbf{u} \in \langle \mathbf{U} \rangle} \max_{\mathbf{v} \in \langle \mathbf{V} \rangle} \mathbf{u}^H \mathbf{v} = \mathbf{u}_k^T \mathbf{v}_k \\ \text{subject to} \quad & ||\mathbf{u}|| = ||\mathbf{v}|| = 1, \\ & \mathbf{u}^H \mathbf{u}_i = 0, \quad i = 1, \ldots, k-1, \\ & \mathbf{v}^H \mathbf{v}_i = 0, \quad i = 1, \ldots, k-1, \end{aligned}$$

for $k = 1, 2, \ldots, q_U$.

Assume that $\mathbf{U}$ and $\mathbf{V}$ are unitary basis for the two subspaces, then the singular values of $\mathbf{U}^H \mathbf{V}$ are $(\cos(\theta_1), \ldots, \cos(\theta_{q_U}))$ [23]. The principal angles induce several distance metrics, from which the most widely used are the geodesic or intrinsic distance [10], [24]

$$d_{geo}\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2 = \sum_{r=1}^{q_U} \theta_r^2,$$

and the extrinsic distance, which is given by the Frobenius norm of the difference between the respective projection matrices [12], [13],

$$\begin{aligned} d\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2 &= \frac{1}{2} \left\| \mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}} \right\|_F^2 \\ &= \frac{1}{2} \left(q_V + q_U - 2 \sum_{r=1}^{q_U} \cos(\theta_r)^2 \right) \\ &= \frac{|q_V - q_U|}{2} + \sum_{r=1}^{q_U} \sin(\theta_r)^2. \end{aligned} \tag{1}$$

The second term in the right hand side of Eq. (1) measures the chordal distance defined by the principal angles, whereas the first term accounts for projection matrices of different ranks.

There are arguments in favor of the extrinsic distance (1), among them, its uniqueness and its computational simplicity in contrast to the intrinsic distance that needs to compute the singular values of $\mathbf{V}^H \mathbf{U}$. Also, the extrinsic distance is related to the squared error in resolving the standard basis for the ambient space, $\{\mathbf{e}_i\}_{i=1}^n$, onto the subspace $\langle \mathbf{V} \rangle$ as opposed to the subspace $\langle \mathbf{U} \rangle$,

$$\begin{aligned} &\sum_{i=1}^{n} \mathbf{e}_i^T (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}})^H (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}}) \mathbf{e}_i \\ &\qquad = \operatorname{tr}\left[(\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}})^H (\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}}) \right] \\ &\qquad\qquad = \left\| \mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\mathbf{V}} \right\|_F^2 = 2d\left(\langle \mathbf{U} \rangle, \langle \mathbf{V} \rangle\right)^2. \end{aligned}$$

B. Order Selection Rule [14], [25]

Let us consider a collection of measured subspaces $\{\langle \mathbf{V}_r \rangle\}_{r=1}^R$ of $\mathcal{C}^n$, each with respective dimension $\dim(\langle \mathbf{V}_r \rangle) = q_r < n$. To simplify the notation, we denote the orthogonal projection matrix onto the rth subspace as $\mathbf{P}_r$. Each subspace $\langle \mathbf{V}_r \rangle$ is a point on the Grassmann manifold $\mathbb{G}(q_r, n)$, and the collection of subspaces lives on a disjoint union of Grasmannians. Without loss of generality, the dimension of the union of all subspaces is assumed to be the ambient space dimension n.

Using the extrinsic distance metric between subspaces, an order estimation criterion for the central subspace that "best approximates" the collection is

$$(s^*, \mathbf{P}_s^*) = \underset{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}}{\arg\min} \quad \frac{1}{2R} \sum_{r=1}^{R} \left\| \mathbf{P} - \mathbf{P}_r \right\|_F^2, \tag{2}$$

where $\mathbb{P}(s, n)$ denotes the set of all idempotent projection matrices of rank s. For completeness, we also accept solutions $\mathbf{P} = \mathbf{0}$ with rank $s = 0$, meaning that there is no a central "signal subspace" shared by the collection of input subspaces.

Expanding the cost function in (2) we obtain the equivalent problem

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{2} \operatorname{tr}\left(\mathbf{P}(\mathbf{I} - 2\overline{\mathbf{P}}) + \overline{\mathbf{P}} \right), \tag{3}$$

where $\overline{\mathbf{P}}$ is an average of orthogonal projection matrices

$$\overline{\mathbf{P}} = \frac{1}{R} \sum_{r=1}^{R} \mathbf{P}_r, \tag{4}$$

with compact eigendecomposition $\overline{\mathbf{P}} = \mathbf{F}\mathbf{K}\mathbf{F}^H$, where $\mathbf{K} = \operatorname{diag}(k_1, \ldots, k_n)$ with $1 \geq k_1 \geq k_2 \geq \ldots \geq k_n$.

Now, discarding constant terms and writing the projection matrix as $\mathbf{P} = \mathbf{U}\mathbf{U}^H$, where $\mathbf{U}$ is a unitary $n \times s$ matrix, problem (3) can be rewritten as

$$\min_{\mathbf{U} \in \mathbb{S}(s,n)} \quad \operatorname{tr}\left(\mathbf{U}^H (\mathbf{I} - 2\overline{\mathbf{P}}) \mathbf{U} \right),$$

where $\mathbb{S}(s, n)$ denotes the complex Stiefel manifold of orthonormal s-frames in $\mathcal{C}^n$. Hence, the optimal order s^* is the number of negative eigenvalues of the matrix

$$\mathbf{S} = \mathbf{I} - 2\overline{\mathbf{P}},$$

or, equivalently, the number of eigenvalues of $\overline{\mathbf{P}}$ larger than 1/2, which is the order fitting rule proposed in [14]. The proposed rule can be written alternatively as

$$s^* = \underset{s \in \{0,1,\ldots,n\}}{\arg\min} \quad \sum_{i=1}^{s} (1 - k_i) + \sum_{s+1}^{n} k_i.$$

A similar rule was developed in [25] for the problem of designing optimum time-frequency subspaces with a specified time-frequency pass region.

Once the optimal order s^* is known, a basis for the average subspace can be obtained as the solution of the following optimization problem

$$\max_{\mathbf{U} \in \mathbb{S}(s^*,n)} \quad \operatorname{tr}\left(\mathbf{U}^H \mathbf{F}\mathbf{K}\mathbf{F}^H \mathbf{U} \right),$$

whose solution is given by any unitary matrix whose column space is the same as the subspace spanned by the s^* principal eigenvectors of $\mathbf{F}$

$$\mathbf{U}^* = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{s^*}) = \mathbf{F}_{s^*},$$

and $\mathbf{P}^* = \mathbf{U}^*(\mathbf{U}^*)^H$. So the average subspace is constructed by quantizing the eigenvalues of the average projection matrix at 0 or 1.

C. Properties of the average projection matrix

The average of projection matrices in (4) is not a projection matrix itself, and therefore is not idempotent. However, it has the following properties:

1) It is symmetric and positive semidefinite.
2) Its eigenvalues are real and satisfy $0 \leq k_i \leq 1$.

1) is trivially proved by noticing that $\overline{\mathbf{P}}$ is an average of symmetric and positive semidefinite projection matrices. To prove 2) let us take without loss of generality the ith eigenvalue-eigenvector pair $(k_i, \mathbf{f}_i)$, then we have

$$k_i = \mathbf{f}_i^H \overline{\mathbf{P}} \mathbf{f}_i = \frac{1}{R}\sum_{r=1}^{R} \mathbf{f}_i^H \mathbf{P}_r \mathbf{f}_i$$
$$\overset{(a)}{=} \frac{1}{R}\sum_{r=1}^{R} \mathbf{f}_i^H \mathbf{P}_r^2 \mathbf{f}_i = \frac{1}{R}\sum_{r=1}^{R} ||\mathbf{P}_r \mathbf{f}_i||^2 \leq 1,$$

where (a) holds because all $\mathbf{P}_r$ are idempotent, and the inequality follows from the fact that each term $||\mathbf{P}_r \mathbf{f}_i||^2$ is the squared norm of the projection of a unit norm vector, $\mathbf{f}_i$, onto the subspace $\langle \mathbf{V}_r \rangle$ and therefore $||\mathbf{P}_r \mathbf{f}_i||^2 \leq 1$ with equality only if the eigenvector belongs to the subspace.

Assuming that $\mathbf{V}_r = [\mathbf{v}_{r1}, \ldots, \mathbf{v}_{rq_r}]$ is a unitary basis for the rth subspace, the eigenvalues of the average projection matrix can be further expressed as

$$k_i = \frac{1}{R}\sum_{r=1}^{R}\sum_{j=1}^{q_j} ||\mathbf{v}_{rj}^H \mathbf{f}_i||^2,$$

and hence they can be interpreted as the squared norm of the average projection along the direction $\mathbf{f}_i$. It is important to remark, however, that the eigenvalues of $\overline{\mathbf{P}}$ are invariant to a common change in the basis of all subspaces. That is, we can apply an arbitrary change of basis $\mathbf{V}_r' = \mathbf{V}_r \mathbf{Q}$ for $r = 1, \ldots, R$ with $\mathbf{Q}$ unitary, and the eigenvalues k_i do not change.

D. Robust version

In some applications there is a need to account for outliers within our collection of measured or extracted subspaces. To this end, we discuss in this section a robust formulation of the proposed order fitting rule based on majorization-minimization (MM) algorithms [21].

The simplest robust formulation of Problem (2) is

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{R}\sum_{r=1}^{R} \rho\left(d_r^2(\mathbf{P})\right) \tag{5}$$

where

$$d_r^2(\mathbf{P}) = \frac{1}{2}\|\mathbf{P} - \mathbf{P}_r\|_F^2$$

and $\rho(\cdot)$ is a smooth concave increasing function that saturates so that outliers or subspaces far away from the average have a limited effect. Examples of robust functions are [26]:

- ℓ_p-norm:
$$\rho(t) = t^{p/2} \tag{6}$$
where $0 < p \leq 2$ with the nonrobust ℓ_2-norm formulation recovered for $p = 2$.
- Huber: for $T > 0$,
$$\rho_H(t) = \begin{cases} t/\sqrt{T} & t \leq T, \\ \sqrt{t} & t > T. \end{cases} \tag{7}$$
For $T = 0$ we obtain the median estimator $\rho_H(t) = \sqrt{t}$.
- Log-loss:
$$\rho_{LL}(t) = \theta \ln(\theta + t),$$
where $\theta \geq 1$.
- Logistic:
$$\rho_L(t) = \frac{1}{1 + e^{-t}}.$$
- Geman-McClure estimator [27], [28]: for $\theta > 0$,
$$\rho_{GM}(t) = \frac{t}{\theta + t}.$$

The idea of the MM algorithm [21] is, at each iteration, to find a majorizer of the objective function. Since the robust function $\rho(\cdot)$ is a smooth concave function, we can easily majorize at some point simply by linearizing:

$$\rho(t) \leq \rho(t_0) + \rho'(t_0)(t - t_0).$$

In the context of our problem, the problem at iteration k (where a central subspace $\mathbf{P}^{(k)}$ of dimension $s^{(k)}$ is available) is

$$\min_{\mathbf{P} \in \mathbb{P}(s,n)} \quad \frac{1}{R}\sum_{r=1}^{R} \rho\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right) + \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)\left(d_r^2(\mathbf{P}) - d_r^2\left(\mathbf{P}^{(k)}\right)\right)$$

or, removing unnecessary constant terms,

$$\min_{\mathbf{P} \in \mathbb{P}(s,n)} \frac{1}{R}\sum_{r=1}^{R} \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right) d_r^2(\mathbf{P}).$$

Let us now define the normalized weights that define a simplex

$$\bar{w}_r^{(k)} = \frac{\rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}{\sum_{r=1}^{R} \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}, \quad \bar{w}_r^{(k)} \geq 0, \quad \sum_r \bar{w}_r^{(k)} = 1.$$

With this definition we can obtain the next iterate $\mathbf{P}^{(k+1)}$ as the solution to

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{2}\sum_{r=1}^{R} \bar{w}_r^{(k)} \|\mathbf{P} - \mathbf{P}_r\|_F^2. \tag{8}$$

Expanding the cost function (8), we obtain

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{2}\operatorname{tr}\left(\mathbf{P}(\mathbf{I} - 2\overline{\mathbf{P}}_w^{(k)}) + \overline{\mathbf{P}}_w^{(k)}\right), \tag{9}$$

where we have used the fact that $\sum_{r=1}^{R} \bar{w}_r^{(k)} \operatorname{tr}(\mathbf{P}) = \operatorname{tr}(\mathbf{P})$ and $\overline{\mathbf{P}}_w^{(k)}$ is the weighted average projection matrix given by

$$\overline{\mathbf{P}}_w^{(k)} = \sum_{r=1}^{R} \bar{w}_r^{(k)} \mathbf{P}_r.$$

Writing the projection matrix in (9) as $\mathbf{P} = \mathbf{U}\mathbf{U}^H$ and discarding constant terms, the optimization problem can be rewritten as

$$\min_{\mathbf{U} \in \mathbb{S}(s,n)} \quad \operatorname{tr}\left(\mathbf{U}^H (\mathbf{I} - 2\overline{\mathbf{P}}_w^{(k)}) \mathbf{U}\right).$$

Then, the optimal order at iteration $k+1$, $s^{(k+1)}$, is the number of negative eigenvalues of the matrix

$$\mathbf{S}^{(k)} = \mathbf{I} - 2\overline{\mathbf{P}}_w^{(k)}. \tag{10}$$

While the non-robust average projection matrix in (4) equally weights all subspaces in the collection by $\bar{w}_r^{(k)} = 1/R$, the robust version uses different weights at each iteration. It is also clear that $\overline{\mathbf{P}}_w^{(k)}$ is symmetric with real eigenvalues bounded above by one, like its non-robust version $\overline{\mathbf{P}} = \sum_r \mathbf{P}_r / R$. A unitary basis for the central subspace at iteration $k+1$ is given by the $s^{(k+1)}$ largest eigenvectors of $\overline{\mathbf{P}}_w^{(k)}$, where recall that $s^{(k+1)}$ is the number of non-negative eigenvalues of $\mathbf{S}^{(k)}$ in (10).

Notice that the objective function (5) is bounded below and that the sequence of objective values at each iteration is non-increasing. Then, the convergence of the sequence of robust order estimates $s^{(1)}, s^{(2)}, \ldots,$ to a stationary point s^* is guaranteed. For a more detailed study of the convergence of MM algorithms, the reader is referred to [21].

III. Distributions on the manifold of projection matrices

In many problems it is useful to assume that the measured subspaces $\{\langle \mathbf{V}_r \rangle\}_{r=1}^{R}$ are random samples drawn from an underlying distribution. Uniform and non-uniform distributions on the Grassmann manifold $\mathbb{G}(k,n)$ or, equivalently, on the manifold of projection matrices of $\operatorname{rank} = k$, $\mathbb{P}(k,n)$, have been extensively discussed in [29]. For uniform distributions the basic experiment is this: generate $\mathbf{X}$ as a random $n \times k$ matrix with i.i.d $\mathcal{CN}(0,1)$ random variables. Perform a QR decomposition of this random matrix as $\mathbf{X} = \mathbf{Q}\mathbf{R}$, then, $\mathbf{Q}\mathbf{H}$ where $\mathbf{H} \in \mathbb{U}(k)$ is any unitary matrix independent of $\mathbf{X}$, is uniformly distributed on $\mathbb{G}(k,n)$, and $\mathbf{Q}\mathbf{Q}^H$ is uniformly distributed on $\mathbb{P}(k,n)$. Remember that points on $\mathbb{G}(k,n)$ are equivalence classes of $n \times k$ matrices, where $\mathbf{Q}_1 \sim \mathbf{Q}_2$ if $\mathbf{Q}_1 = \mathbf{Q}_2\mathbf{H}$ for some $\mathbf{H} \in \mathbb{U}(k)$.

If $\mathbf{P}$ is uniformly distributed on $\mathbb{P}(k,n)$, it is immediate to prove that (see [29], pp. 29)

$$E[\mathbf{P}] = \frac{k}{n}\mathbf{I}_n,$$

so all eigenvalues of the mean projection matrix when the subspaces are uniformly distributed are identical to $k_i = k/n$, $i = 1, \ldots, n$, indicating no preference for any particular direction. In this way, for uniformly distributed subspaces the proposed order fitting rule will return 0 if $k < n/2$, and n otherwise, in both cases suggesting there is no central low-dimensional subspace.

The matrix Langevin (or von Mises-Fisher) has been suggested as a non-uniform distribution on the Stiefel and Grassman manifolds [29]–[31]. For real matrices, the matrix Langevin, as defined by Downs in [32], has an exponential distribution of the form $\mathcal{L}(\mathbf{X}) \propto \exp\left\{\operatorname{tr}(\mathbf{B}^T\mathbf{X})\right\}$, where $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ is a matrix that parameterizes the distribution with $\mathbf{U} \in \mathbb{S}(k,n)$, $\mathbf{V} \in \mathbb{U}(k)$ and $\mathbf{D}$ a $k \times k$ diagonal matrix with positive entries. Matrices $\mathbf{U}$ and $\mathbf{V}$ are interpreted as orientations, while the diagonal elements of $\mathbf{D}$ are concentration parameters along the k directions determined by $\mathbf{U}$ and $\mathbf{V}$. The matrix Langevin $\mathcal{L}(\mathbf{X})$ is unimodal and the density is maximized at $\mathbf{U}\mathbf{V}^T$, which is the central k-frame or subspace of the distribution. Note that when $\mathbf{B} = \mathbf{0}$ we recover the uniform distribution. As suggested in [29], to generate samples from $\mathcal{L}(\mathbf{X})$ we might use a rejection sampling mechanism with the uniform as proposal density. This rejection sampling, however, can be very inefficient for large n and $k > 1$. More efficient sampling algorithms have been proposed in [33].

The uniform and the matrix Langevin are continuous distributions on the manifold of projection matrices of fixed $\operatorname{rank} = k$. To deal with subspaces or projection matrices that do not live on the same manifold we would need distributions defined over unions of Grassmannians, which, to the best of our knowledge, have not been studied. Nevertheless, it is possible to define the following discrete distribution that will be useful for the application of the proposed subspace averaging technique to array processing in Section IV.

Definition 3.1: Let $\mathbf{U} = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_n] \in \mathbb{U}(n)$ be an arbitrary unitary basis of the ambient space, and let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ with $0 \leq \alpha_i \leq 1$; the α_i are ordered from largest to smallest, but they need not sum to 1. We define a discrete distribution $\mathcal{D}$ on the set of random projection matrices $\mathbf{P} = \mathbf{V}\mathbf{V}^H$ (or, equivalently, the set of random subspaces $\langle \mathbf{V} \rangle$, or set of frames $\mathbf{V}$) with parameter vector $\boldsymbol{\alpha}$ and orientation matrix $\mathbf{U}$. The distribution of $\mathbf{P}$ will be denoted $\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ or $\mathbf{V} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$.

To shed some light on this distribution, let us explain the experiment that determines $\mathcal{D}$. Draw 1 includes $\mathbf{u}_1$ with probability α_1, and excludes it with probability $(1 - \alpha_1)$; draw 2 includes $\mathbf{u}_2$ with probability α_2, and excludes it with probability $(1 - \alpha_2)$; continue in this way until draw n includes $\mathbf{u}_n$ with probability α_n, and excludes it with probability $(1 - \alpha_n)$. We may call the string $i_1, i_2, \ldots, i_n$, the indicator sequence for the draws; that is, $i_k = 1$, if $\mathbf{u}_k$ is drawn on draw k, and $i_k = 0$ otherwise. In this way Pascal's triangle shows that the probability of drawing the subspace $\langle \mathbf{V} \rangle$ is $p(\langle \mathbf{V} \rangle) = \prod_I \alpha_i \prod_{\bar{I}} (1 - \alpha_j)$, where the index set I is the set of indices k for which $i_k = 1$ in the construction of $\mathbf{V}$. This is also the probability law on frames $\mathbf{V}$ and projections $\mathbf{P}$. For example, the probability of drawing an empty frame is $\prod_1^n (1 - \alpha_i)$, the probability of drawing the dimension-1 frame $\mathbf{u}_i \mathbf{u}_i^H$ is $\alpha_i \prod_{j \neq i} (1 - \alpha_j)$, and so on. It is clear from this pdf on the 2^n frames that all probabilities lie between 0 and 1,

and that they sum to 1.

Let $\mathbf{P}_r \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, $r = 1, \ldots, R$, be a sequence of i.i.d. draws from the distribution $\mathcal{D}$, and let $\overline{\mathbf{P}} = \sum_r \mathbf{P}_r / R$ be its sample mean with eigenvalues $(k_1, \ldots, k_n)$. Then, we have the following properties:

1) $E[\mathbf{P}_r] = \mathbf{U}\text{diag}(\boldsymbol{\alpha})\mathbf{U}^H$, that is, the mean is not generally a projection.
2) $E\left[\text{tr}(\mathbf{P}_r)\right] = \sum_{i=1}^{n} \alpha_i$.
3) $E\left[k_i\right] = \alpha_i$.

These properties follow directly from the definition of $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$.

Remark 1: The α_i's control the *concentrations* or probabilities in the directions determined by the unitary basis $\mathbf{U}$. For instance, if $\alpha_i = 1$ all random subspaces contain direction $\mathbf{u}_i$, whereas if $\alpha_i = 0$ the angle between $\mathbf{u}_i$ and all random subspaces drawn from that distribution will be $\pi/2$.

Example 1: Suppose $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ is the standard basis in $\mathcal{R}^3$ and let $\boldsymbol{\alpha} = (3/4, 1/4, 1/4)$. The discrete distribution $\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ has an alphabet of $2^3 = 8$ subspaces with the following probabilities:

- $\Pr\left(\mathbf{P} = \mathbf{0}\right) = 9/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1\mathbf{u}_1^H\right) = 27/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_2\mathbf{u}_2^H\right) = 3/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_3\mathbf{u}_3^H\right) = 3/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1\mathbf{u}_1^H + \mathbf{u}_2\mathbf{u}_2^H\right) = 9/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_1\mathbf{u}_1^H + \mathbf{u}_3\mathbf{u}_3^H\right) = 9/64$
- $\Pr\left(\mathbf{P} = \mathbf{u}_2\mathbf{u}_2^H + \mathbf{u}_3\mathbf{u}_3^H\right) = 1/64$
- $\Pr\left(\mathbf{P} = \mathbf{I}\right) = 3/64$

The distribution is unimodal with mean

$$E[\mathbf{P}] = \mathbf{U}\,\text{diag}(\boldsymbol{\alpha})\mathbf{U}^H.$$

and expected dimension $E[\text{tr}(\mathbf{P})] = 5/4$. Given R draws from the distribution $P \sim D(\mathbf{U}, \alpha)$, the eigenvalues of the sample average projection $\overline{\mathbf{P}} = \sum_{r=1}^{R} \mathbf{P}_r / R$ converge to $k_i \to \alpha_i$ as R grows, and the proposed order fitting rule will return $s^* = 1$ as the dimension of the central subspace for this example. It is easy to check that the probability of drawing a dimension-1 subspace for this example is $33/64$.

IV. Subspace Averaging for Source Enumeration

In this section we apply the proposed order fitting rule for subspace averaging to the problem of estimating the number of signals received by a sensor array, which is referred to as source enumeration. This is a classic and well-researched problem in radar, sonar, and communications [34], [35], and numerous criteria have been proposed over the last decades to solve this problem, most of which are given by functions of the eigenvalues of the sample covariance matrix [15], [36]–[42]. These methods tend to underperform when the number of antennas is large and the number of snapshots is relatively small in comparison to the number of antennas, the so-called small-sample regime [17], which is the situation of interest to this paper.

For instance, [43] showed that the penalty term used by the MDL criterion is effective in preventing overestimation of the number of sources, but it causes a considerable increase in the probability of underestimation when the number of snapshots (in relation to the number of antennas) is low. Other perfomance studies of classic information-theoretic criteria can be found in [44], [45], where similar conclusions are drawn.

To overcome this limitation of classic information-theoretic criteria, our method is to construct a collection of subspaces based on the array geometry and a random sampling procedure from a specifically designed distribution $\mathcal{D}$, and then use the order fitting rule for averages of projections to enumerate the sources. This SA method is particularly effective when the dimension of the input space is large (high-dimensional data) and we have few snapshots, which is when the sample eigenvalues are poorly estimated and all classic methods fail.

The conventional approach to order determination is to compute likelihoods for a sequence of rank-k plus diagonal covariance models to multivariate normal measurements [55], [37], and then to penalize likelihoods for large ranks k. The resulting formulas use sums and products of sub-dominant eigenvalues of a sample covariance matrix in what amount to tests of whiteness of the trailing sequence of eigenvalues. In these methods, the scale of the rank-k and diagonal components are implicitly estimated in the estimation of the covariance model. In our approach the eigenvalues of the sample covariance matrix are used only to determine a distribution on a space of subspaces that could have produced the sample covariance. A bootstrap draws subspaces from this distribution, averages them, and returns an order for the average. This is the estimated number of sources. The procedure is scale-invariant.

A. Problem Statement

Let us consider K narrowband signals impinging on a large, uniform, half-wavelength linear array with M antennas (cf. Fig. 1). The received signal is

$$\mathbf{x}[t] = [\mathbf{a}(\theta_1) \quad \cdots \quad \mathbf{a}(\theta_K)]\,\mathbf{s}[t] + \mathbf{e}[n] = \mathbf{A}\mathbf{s}[t] + \mathbf{e}[t], \quad (11)$$

where $\mathbf{s}[t] = [s_1[t], \ldots, s_K[t]^T$ is the vector of complex gains $s_k[t]$ for $M \times 1$ complex array response $\mathbf{a}(\theta_k) = \begin{bmatrix}1 & e^{-j\theta_k} & e^{-j\theta_k(M-1)}\end{bmatrix}^T$ to the kth source whose direction-of-arrival (DOA) θ_k is unknown. The signal and noise vectors are modeled as $\mathbf{s}[t] \sim \mathcal{CN}_K(\mathbf{0}, \mathbf{S})$ and $\mathbf{e}[t] \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I})$, respectively. The dimensions are these: $\mathbf{x} \in \mathcal{C}^M, \mathbf{A} \in \mathcal{C}^{M\times K}, \mathbf{s} \in \mathcal{C}^K$, and $\mathbf{e} \in \mathcal{C}^M$. From the signal model (11), the theoretical covariance matrix is

$$\mathbf{R} = E\left[\mathbf{x}[t]\mathbf{x}^H[t]\right] = \mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I}.$$

We assume there are N snapshots collected in the data matrix matrix $\mathbf{X} = [\mathbf{x}[1] \quad \ldots \quad \mathbf{x}[N]]$. The source enumeration problem consists of estimating K from $\mathbf{X}$.

B. Subspace Generation

A key ingredient of the proposed technique is the method of generating the collection of subspaces from which to estimate the average projection matrix and its dimension. In some applications, such as image or video processing, the collection of subspaces might be given, but in array processing the signal subspace is an array manifold, to be estimated from array

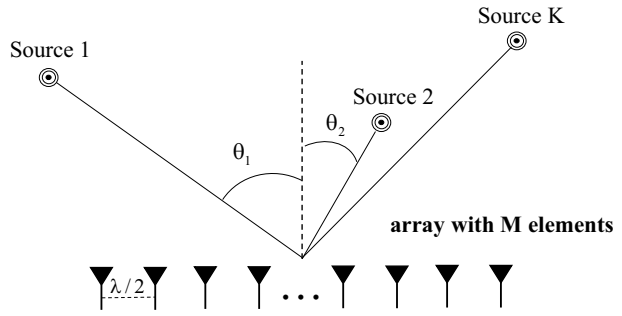

Fig. 1: Source enumeration problem in large scale arrays: estimating the number of sources K in a ULA with a high number of antenna elements M.

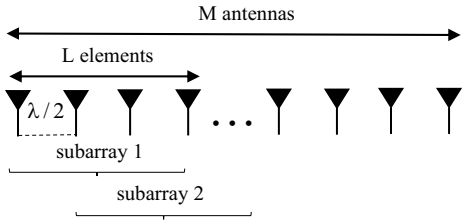

Fig. 2: L-dimensional subarrays extracted from a uniform linear array with $M > L$ elements.

snapshots. For instance, when a uniform linear array (ULA) is used, we can exploit the shift-invariance property to estimate a subspace from the data acquired by a subarray of the sensors. The subspaces could also be estimated from subsets of $L < N$ snapshots randomly selected from the original dataset, which would be appropriate for cyclostationary snapshots, or using any other bootstrapping scheme.

For the SA method to be effective, it is important that the subspaces to average overlap as much as possible with the true signal subspace. In fact, as long as each extracted subspace contains a large common portion of the signal subspace and (more or less) independent portions of the noise subspace, then, the averaging procedure enhances signal coordinates while averaging out noise coordinates. As we will see, this translates into a better performance for the proposed order estimation rule compared to the state-of-the-art.

In the following, we describe a subspace generation procedure that has proven to be effective for this particular application. It generates random subspaces by randomly sampling from the distribution $D(\mathbf{U}, \boldsymbol{\alpha})$, whose orientations $\mathbf{U}$ and concentrations $\boldsymbol{\alpha}$ are determined by the eigenvectors and eigenvalues of the sample covariance matrix. Moreover, it exploits the shift invariance property of ULAs. A preliminary version of this method that only exploited the shift-invariance property was presented in [20].

1) Shift invariance: When uniform linear arrays are used, a property called shift invariance holds, which forms the basis of the ESPRIT method [46], [47] and its many variants. Let $\mathbf{A}_s$ be the $L \times K$ matrix with rows $s, \ldots, s+L-1$ extracted from the steering matrix $\mathbf{A}$. This steering matrix for the sth subarray is illustrated in Fig.2.

Then, from (11) it is readily verified that

$$\mathbf{A}_s \operatorname{diag}(e^{-j\theta_1}, \ldots, e^{-j\theta_K}) = \mathbf{A}_{s+1}, \;\; s = 1, \ldots, M-L+1,$$

which is the shift invariance property. In this way, $\mathbf{A}_s$ and $\mathbf{A}_{s+1}$ are related by a nonsingular rotation matrix,

$$\mathbf{Q} = \operatorname{diag}(e^{-j\theta_1}, \ldots, e^{-j\theta_K}),$$

and therefore they span the same subspace. That is, $\langle \mathbf{A}_s \rangle = \langle \mathbf{A}_{s+1} \rangle$, with $\dim(\langle \mathbf{A}_s \rangle) = K < L$. In ESPRIT, two subarrays of dimension $L = M-1$ are considered, and thus we have $\mathbf{A}_1\mathbf{Q} = \mathbf{A}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ select, respectively, the first and the last $M-1$ rows of $\mathbf{A}$.

There is an interesting characterization of the shift invariance property. Let $\mathbf{x}_s[t]$ be an $L \times 1$ vector containing the noise-free observations acquired by sensors $s, \ldots, s+L-1$ of $\mathbf{x}[t]$, and let $\mathcal{S}^r$ denote a shift operator, so that $\mathcal{S}^r \mathbf{x}_s[t] = \mathbf{x}_{s+r}[t]$. Then, in the noise-free model $\mathbf{x}_s[t] = \mathbf{A}_s \mathbf{s}[t]$, this shift invariance produces

$$\mathcal{S}^r \mathbf{x}_s[t] = \mathcal{S}^r \mathbf{A}_s \mathbf{s}[t] = \mathbf{A}_{s+r} \mathbf{s}[t] = \mathbf{A}_s \mathbf{Q}^r \mathbf{s}[t].$$

The source signal $\mathbf{Q}^r \mathbf{s}[t]$ is distributed as $\mathbf{s}[t]$ is distributed, provided $\mathbf{s}[t]$ is complex normal with diagonal covariance matrix, making $\mathbf{Q}$ a measure-preserving transformation. So shift on $\mathbf{x}_s[t]$ is measure-preserving on $\mathbf{s}$. This property leaves second-order matrices invariant.

When noise is present, however, the shift-invariance property does not hold for the main eigenvectors extracted from the sample covariance matrix. The Optimal Subspace Estimation (OSE) technique proposed by Vaccaro et al. obtains an improved estimate of the signal subspace with the required structure (up to the first order) [48]. The OSE has recently been used with the subspace averaging of [14] to improve DOA estimation [19], [49], [50]. Nevertheless, the OSE technique requires the dimension of the signal subspace to be known in advance and, therefore, does not apply directly to the source enumeration problem.

From the $L \times 1$ ($L > K$) sub-array snapshots $\mathbf{x}_s[t]$ we can estimate an $L \times L$ sample covariance as

$$\hat{\mathbf{R}}_s = \frac{1}{N} \sum_{t=1}^{N} \mathbf{x}_s[t] \mathbf{x}_s^H[t].$$

Note that each $\hat{\mathbf{R}}_s$ block corresponds to an $L \times L$ submatrix of the full sample covariance $\hat{\mathbf{R}}$ extracted along its diagonal, that is, in Matlab notation $\hat{\mathbf{R}}_s = \hat{\mathbf{R}}(s : s+L-1, s : s+L-1)$.

Due to the shift invariance property of uniform linear arrays the noiseless signal subspaces of the theoretical $\mathbf{R}_s$ are identical. Since there are M sensors and we extract L-dimensional subarrays, there are $S = M-L+1$ different submatrices $\hat{\mathbf{R}}_s$, $s = 1, \ldots, S$. For each $\hat{\mathbf{R}}_s$ we compute its eigendecomposition $\hat{\mathbf{R}}_s = \mathbf{U}_s \boldsymbol{\Lambda}_s \mathbf{U}_s^H$, where $\boldsymbol{\Lambda}_s = \operatorname{diag}(\lambda_{s,1}, \ldots, \lambda_{s,L})$, $\lambda_{s,1} \geq \ldots \geq \lambda_{s,L}$.

For each $\hat{\mathbf{R}}_s$ we can define a distribution $\mathcal{D}(\mathbf{U}_s, \boldsymbol{\alpha}_s)$ from which to draw random subspaces: $\mathbf{P}_{sk}, k = 1, \ldots, K$. Obviously a key point for the success of the SA method is to determine a good distribution $\mathcal{D}(\mathbf{U}_s, \boldsymbol{\alpha}_s)$ and a good sampling procedure to draw random subspaces. This is described in the next subsection.

2) Random generation of subspaces: To describe the random sampling procedure for subspace generation, let us take for simplicity the full $M \times M$ sample covariance matrix $\hat{\mathbf{R}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_M)$, $\lambda_1 \geq \ldots \geq \lambda_M$.

Each random subspace $\langle \mathbf{V} \rangle$ has dimension $\dim(\langle \mathbf{V} \rangle) = k_{max}$, where $k_{max} < \min(M, N)$ is an overestimate of the maximum number of sources that we expect in our problem. The subspace is iteratively constructed as follows:

1) Initialize $\langle \mathbf{V} \rangle = \emptyset$
2) While $\operatorname{rank}(\mathbf{V}) \leq k_{max}$ do
   a) Generate a random draw $\langle \mathbf{G} \rangle \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, according to the sampling description in Definition 3.1
   b) $\langle \mathbf{V} \rangle = \langle \mathbf{V} \rangle \bigcup \langle \mathbf{G} \rangle$

The orientation matrix $\mathbf{U}$ of the distribution $\mathcal{D}$ is given by the eigenvectors of the sample covariance matrix. On the other hand, the concentration parameters should be chosen such that the signal subspace directions are selected more often than the noise subspace directions, and, consequently, they should be a function of the eigenvalues of the sample covariance λ_k. In this work we propose to use the following concentration parameters for $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$

$$\alpha_i = \frac{\Delta\lambda_i}{\sum_i \Delta\lambda_i}, \tag{12}$$

where

$$\Delta\lambda_i = \begin{cases} \lambda_i - \lambda_{i+1}, & i = 1, \ldots, M-1, \\ 0, & i = M. \end{cases} \tag{13}$$

Here is a motivating example for this choice. Consider the wide-sense stationary time series $\{x[n]\}$ with covariance $r[k] = E[x[n]x^*[n+k]] = \sin(\beta\pi k)/(\pi k)$ for all k, where $0 < \beta < 1$ determines the signal bandwidth. A snapshot $\mathbf{x} = [x[0], \ldots, x[M-1]]$ has symmetric Toeplitz covariance matrix $\mathbf{R}$ with entries $r[k]$ on its k-th diagonal. This covariance matrix may be written

$$\mathbf{R} = E\left[\int_{-\beta\pi}^{\beta\pi} \frac{d\theta}{2\pi} S(\theta)\boldsymbol{\psi}(\theta)\boldsymbol{\psi}^H(\theta)\right]$$

where $\boldsymbol{\psi} = [1, e^{j\theta}, \ldots, e^{j\theta(M-1)}]^T$, and $S(\theta) = 1$. The trace of this matrix is βM, so the sum of its eigenvalues is this trace. The eigenvalue decomposition of the covariance $\mathbf{R}$ is $\mathbf{R} = \mathbf{F}\mathbf{K}\mathbf{F}^T$, where the columns of $\mathbf{F}$ are the discrete prolate spheroidal wave functions, or Slepian sequences [51]. The first $\lfloor \beta M \rfloor$ eigenvalues λ_i are approximately 1, and the trailing $M - \lfloor \beta M \rfloor$ are approximately 0. Moreover, the matrix $\mathbf{F}\mathbf{K}\mathbf{F}^T$ is approximately a rank $\lfloor \beta M \rfloor$ projection onto the subspace spanned by the first $\lfloor \beta M \rfloor$ columns of the matrix $\mathbf{F}\mathbf{K}\mathbf{F}^T$. An estimator of this rank is $\operatorname{argmax}_i(\lambda_{i+1} - \lambda_i)$, which returns the integer part of βM. This suggests that the function $\alpha_i = \lambda_{i+1} - \lambda_i$, after proper normalization, would be a suitable function of the eigenvalues to use in a sequence of stochastic draws of subspaces, as outlined previously.

Algorithm 1: Generation of a Random Subspace.

Input: $\hat{\mathbf{R}} = \mathbf{U}_0\boldsymbol{\Lambda}\mathbf{U}_0^H$, k_{max}
Output: Unitary basis for a random subspace $\mathbf{V}$
Initialization: $\mathbf{U} = \mathbf{U}_0$, $\boldsymbol{\lambda} = \operatorname{diag}(\boldsymbol{\Lambda})$, and $\mathbf{V} = \emptyset$
while $\operatorname{rank}(\mathbf{V}) \leq k_{max}$ **do**
  `/* Generate concentration parameters` $\boldsymbol{\alpha}$ `*/`
  $M = |\boldsymbol{\lambda}|$
  $\alpha_i = \frac{\Delta\lambda_i}{\sum_i \Delta\lambda_i}$, $i = 1, \ldots, M$ with $\Delta\lambda_i$ given by (13)
  `/* Sample from` $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ `*/`
  $\mathbf{g} = (g_i, \ldots, g_M)$ with $g_i \sim \mathcal{U}(0, 1)$
  $\mathcal{I} = \{i \mid g_i \leq \alpha_i\}$
  $\mathbf{G} = \mathbf{U}(:, \mathcal{I})$
  `/* Append new subspace */`
  $\mathbf{V} = [\mathbf{V} \quad \mathbf{G}]$
  `/* Eliminate selected directions */`
  $\mathbf{U} = \mathbf{U}(:, \overline{\mathcal{I}})$
  $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\overline{\mathcal{I}})$

With this choice for $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$, the probability of picking the ith direction from $\mathbf{U}$ is proportional to $\lambda_i - \lambda_{i+1}$, thus placing more probability on jumps of the eigenvalue profile. Notice also that whenever $\lambda_i = \lambda_{i+1}$ then $\alpha_i = 0$, which means that $\mathbf{u}_i$ will never be chosen in any random draw. We take the convention that if $\Delta\lambda_i = 0$, $\forall i$, then we do not apply the normalization in Eq. (12) and hence the concentration parameters are also all zero: $\alpha_i = 0$, $\forall i$.

A summary of the proposed algorithm is shown in Algorithm 1.

3) Subspace averaging (SA) criterion: For each subarray sample covariance matrix we can generate T random subspaces according to the procedure described in the previous section. Since we have S subarray matrices, we get a total of $R = ST$ subspaces. The SA approach simply finds the average projection matrix

$$\overline{\mathbf{P}} = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \mathbf{P}_{st},$$

to which the order estimation method described in Sec. II can be applied. Notice that the only parameters in the method are the dimension of the subarrays, L, the dimension of the extracted subspaces, k_{max}, and the number T of random subspaces extracted from each subarray. For large-scale arrays ($M \geq 50$), we have found that $L = M - 5$, $k_{max} = \lfloor M/5 \rfloor$, and $T = 20$ provide in general good performance for many scenarios.

A summary of the proposed algorithm is shown in Algorithm 2.

Regarding the computational complexity of the method, the proposed SA technique requires $(S+1)\mathcal{O}(L^3)$ operations since we need to perform $S+1$ EVDs of $L \times L$ matrices: S for the

Algorithm 2: Subspace Averaging (SA) Criterion.

Input: $\hat{\mathbf{R}}$, L, T and k_{max};
Output: Order estimate $\hat{k}_{SA}$
for $s = 1, \ldots, S$ **do**
  Extract $\hat{\mathbf{R}}_s$ from $\hat{\mathbf{R}}$ and obtain $\hat{\mathbf{R}}_s = \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{U}_s^H$
  Generate T random subspaces from $\hat{\mathbf{R}}_s$ using Algorithm 1
  Compute the projection matrices $\mathbf{P}_{st} = \mathbf{V}_{st}\mathbf{V}_{st}^H$
Compute $\overline{\mathbf{P}}$ and its eigenvalues $(k_1, \ldots, k_L)$
Estimate $\hat{k}_{SA}$ as the number of eigenvalues of $\overline{\mathbf{P}}$ larger than 1/2

subarray sample covariance matrices $\hat{\mathbf{R}}_s$, $s = 1, \ldots, S$ and one for the average projection matrix $\overline{\mathbf{P}}$. For large scale arrays with $L \approx M$ so that resolution is not significantly reduced, the computational complexity of the SA method is roughly $S+1$ times higher than that of standard methods, which have a complexity of $\mathcal{O}(M^3)$ due to the EVD of $\hat{\mathbf{R}}$. Thus, with the aforementioned choice for L, the number of subarrays is $S = 6$.

C. Consistency of the SA method

In this section we show that the SA criterion equipped with the proposed subspace generation procedure is consistent in the classic asymptotic regime (fixed M, $N \to \infty$).

Theorem 1: Let $\mathbf{R} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I}$ be an $M \times M$ theoretical covariance matrix with eigenvalues $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K > \sigma_{K+1} = \cdots = \sigma_M$, corresponding to an scenario with K uncorrelated sources. Let $\hat{\mathbf{R}}$ be the sample covariance matrix formed by N snapshots and let us denote its eigenvalues as λ_i, $i = 1, \ldots, M$. Then, if $M \geq L > k_{max} \geq K$, $\hat{k}_{SA}$ is a consistent estimator of K as $N \to \infty$.

Proof: Let us take without loss of generality the case $M = L$, that is, we only generate subspaces from the full sample covariance matrix. In the classic fixed system-size M, large sample-size asymptotic regime $N \to \infty$, the eigenvalues of the sample covariance estimated from independent snapshots converges to the theoretical ones $\lambda_i \to \sigma_i$, $i = 1, \ldots, M$. Then, in the first draw to construct each random subspace we sample from $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$ where the orientation matrix $\mathbf{U}$ contains the eigenvectors of $\mathbf{R}$ and the concentration parameters are

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K, 0, \ldots, 0)\,. \tag{14}$$

In (14), the $M - K$ trailing concentration values are zero because they are constructed from (normalized) differences of eigenvalues, whose $M-K$ smallest values are identical in the asymptotic regime: $\sigma_{K+1} = \ldots = \sigma_M$.

Therefore, the first draw samples exclusively from the signal directions. Since we sample until the dimension of the subspace is k_{max}, or until all concentration parameters are zero, then, as long as $k_{max} \geq K$, all random subspaces will be the true signal subspace. Consequently, for any number of generated subspaces the average projection matrix has exactly K eigenvalues equal to 1 and $M - K$ eigenvalues equal to zero, and the SA criterion returns $\hat{k}_{max} = K$, thus proving the result. ∎

Notice that this consistency results also holds as the noise variance $\sigma^2 \to 0$, unlike the minimum description length (MDL) criterion which, as shown in [52], is inconsistent with increasing signal-to-noise ratio.

V. Simulation Results

In this section we evaluate the performance of the proposed order fitting rule by means of numerical examples. Firstly, we study the performance of the order fitting rule, as well as its robust version. Secondly, we consider the application of subspace averaging techniques as a method of enumerating sources in large linear arrays, under conditions of low sample support.

A. Performance of the order fitting rule

Experiment 1: In the first example, we generate a collection of R subspaces, $\langle \mathbf{V}_r \rangle \in \mathbb{G}(k, n)$, $r = 1, \ldots, R$, as follows: we first generate

$$\mathbf{G}_r = \left[\mathbf{V}_0 \mid \mathbf{0}_{n\times(n-k)}\right] + \sigma\mathbf{Z}_r, \qquad r = 1, \ldots, R$$

where $\mathbf{V}_0 \in \mathbf{C}^{n\times k}$ is a matrix whose columns form an orthonormal basis for a central subspace $\langle \mathbf{V}_0 \rangle$, $\mathbf{0}_{n\times(n-k)}$ is an $n \times (n-k)$ zero matrix, and $\mathbf{Z}_r \in \mathbf{C}^{n\times n}$ is a matrix whose entries are independent and identically distributed complex Gaussian random variables with zero mean and variance $\sigma^2 = 1/n$. The value of σ determines the signal-to-noise-ratio, which determines the spread of the subspaces around its mean and is defined as $\mathrm{SNR} = 10\log_{10}\left(\frac{k}{n\sigma^2}\right)$.

An orthogonal basis for the rth subspace, $\mathbf{V}_r$, is then constructed from the first k orthonormal vectors of the QR decomposition of $\mathbf{G}_r$. For this example all subspaces in the collection have exactly the same dimension.

Fig. 3 shows the estimated order as a function of the SNR for different values of (k, n) and a total number of $R = 200$ subspaces. The curves represent averaged results of 500 independent simulations. As we observe, the estimated order coincides with the true one as the SNR grows, thus showing the consistency of the method with increasing SNR. Also, there is phase-transition behavior between $s^* = 0$ (no central subspace) and the correct order $s^* = k$. Similar phase-transition phenomena have been reported for the rank estimation problem of a sample covariance matrix in the asymptotic regime [53]. Based on random matrix arguments, [53] provides a threshold, which depends on the SNR and the ratio of sensors to snapshots, below which the sample eigenvalues are unrelated to the true eigenvalues and hence any order estimation rule based on sample eigenvalues provides inconsistent results. The results of Fig. 3 seem to suggest that a similar concentration-of-measure phenomenon happens for the sample eigenvalues of an average projection matrix. This phase transition apparently depends also on the SNR as well as on the ratio k/n.

Experiment 2: In the second experiment we evaluate the robust order fitting rule proposed in II-D. To this end, we create

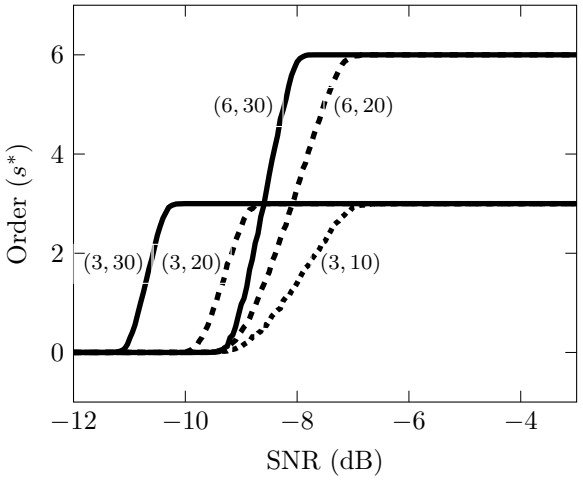

Fig. 3: Estimated order as a function of the SNR for different values of (k, n). In all examples the number of measured subspaces is $R = 200$.

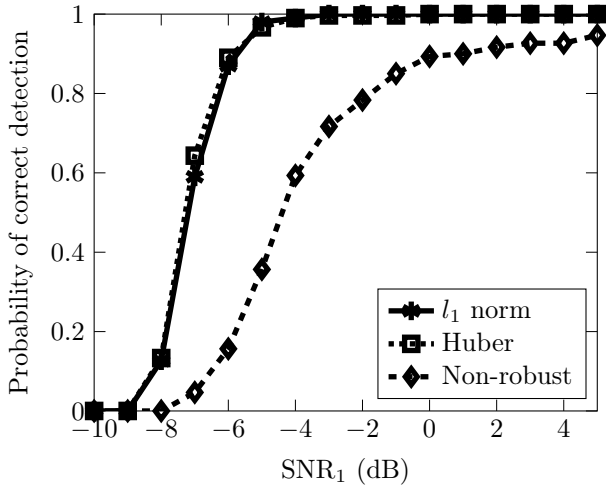

Fig. 4: Probability of correct order detection for robust and non-robust methods ($M = 100$, $k = 3$, $n = 40$, $\mathrm{SNR}_2 = -20$ dB and $\epsilon = 0.5$).

a collection of subspaces contaminated by outliers as follows: we first generate

$$\mathbf{G}_r = \left[\mathbf{V}_0 \mid \mathbf{0}_{n\times(n-k)}\right] + \mathbf{Z}_r, \qquad r = 1, \ldots, R$$

where now the elements of $\mathbf{Z}_r$ are drawn from a Gaussian mixture $\mathbf{Z}_r(i,j) \sim (1-\epsilon)\mathcal{CN}(0, \sigma_1^2/n) + \epsilon\mathcal{CN}(0, \sigma_2^2/n)$, where $\sigma_2^2 \gg \sigma_1^2$. In words, with probability $(1-\epsilon)$ the central subspace is additively perturbed by a random matrix whose entries are i.i.d. zero-mean Gaussians random variables with variance σ_1^2/n, whereas with probability ϵ the entries of the noise matrix are drawn from a Gaussian distribution with variance $\sigma_2^2/n \gg \sigma_1^2/n$. Again, an orthogonal basis for the rth subspace, $\mathbf{V}_r$, is constructed from the first k orthonormal vectors of the QR decomposition of $\mathbf{G}_r$. In this way, we emulate a Gaussian mixture model for this problem. For low values of σ_1^2, with probability $1-\epsilon$ the subspaces are well clustered around $\mathbf{V}_0$. On the other hand, with probability ϵ the subspaces are generated with a much higher variance $\sigma_2^2 \gg \sigma_1^2$ and hence they can be interpreted as outliers.

The signal-to-noise-ratio for the normal data (inliers) and the outliers is defined as $\mathrm{SNR}_i = 10\log_{10}\left(\frac{k}{n\sigma_i^2}\right)$ for $i = 1, 2$. For this example, we estimate the order of the central subspace using the extrinsic mean squared error distance, and the robust versions using the l_1 norm (6) and the Huber loss function with $T = 0.5$ (7). In all simulations, the MM algorithm converged in less than 5 iterations. We consider a set of $R = 100$ subspaces of dimension $k = 3$ in an ambient space of dimension $n = 20$. The proportion of outliers is $\epsilon = 0.5$, and its signal-to-noise ratio is $\mathrm{SNR}_2 = -20$ dB. Fig. 4 shows the probability of correct order estimation as the signal-to-noise-ratio for the inliers, SNR_1, varies, where increased robustness of both the l_1 and Huber cost functions are evident.

B. Application to source enumeration

We consider a scenario with K narrowband incoherent unit-power signals and DOAs separated Δ_θ impinging on a uniform linear array with M antennas and half-wavelength element separation (cf. Fig. 1). The number of snapshots is N. The proposed SA method uses subarrays of size $L = M - 5$, so the total number of subarrays is $S = 6$. From the sample covariance matrix of each subarray we generate $T = 20$ random subspaces of dimension $k_{max} = \lfloor M/5 \rfloor$, which gives us a total of $R = 120$ subspaces on the Grassmann manifold $\mathbb{G}(k_{max}, L)$ to compute the average projection matrix $\overline{\mathbf{P}}$. For the examples in this section, we define the signal-to-noise-ratio as $\mathrm{SNR} = 10\log_{10}(1/\sigma^2)$, which is the input or per-sample SNR. The SNR at the output of the array is $20\log_{10}(M)$ dBs higher.

Some representative methods for source enumeration with high-dimensional data and few snaphots have been selected for comparison. They exploit random matrix results and are specifically designed to operate in this regime. Further, all of them are functions of the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_M$ of the sample covariance matrix $\hat{\mathbf{R}}$. We now present a brief description of the methods under comparison.

- LS-MDL criterion in [54]: The standard MDL method proposed by Wax and Kailath in [37], based on a fundamental result of Anderson [55], is

$$\hat{k}_{MDL} = \underset{0\leq k\leq M-1}{\operatorname{argmin}} \; (M-k)N\log\left(\frac{a(k)}{g(k)}\right) + \frac{1}{2}k(2M-k)\log N, \quad (15)$$

  where $a(k)$ and $g(k)$ are the arithmetic and the geometric mean, respectively, of the $M-k$ smallest eigenvalues of $\hat{\mathbf{R}}$. When the number of snapshots is smaller than the

number of sensors or antennas ($N < M$), the sample covariance becomes rank-deficient and (15) can not be applied directly.
The LS-MDL method proposed by Huang and So in [54] replaces the noise eigenvalues λ_i in the MDL criterion by a linear shrinkage (LS), calculated as

$$\rho_i^{(k)} = \beta^{(k)} a(k) + (1 - \beta^{(k)})\lambda_i, \quad i = k+1, \ldots, M,$$

where $\beta^{(k)} = \min(1, \alpha^{(k)})$, with

$$\alpha^{(k)} = \frac{\sum_{i=k+1}^{M} \lambda_i^2 + (M-k)^2 a(k)^2}{(N+1)\left(\sum_{i=k+1}^{M} \lambda_i^2 - (M-k)a(k)^2\right)}.$$

- NE criterion in [17]: The method proposed by Nadakuditi and Edelman in [17], which we refer to as the NE criterion, is given by

$$\hat{k}_{NE} = \underset{0 \le k \le M-1}{\operatorname{argmin}} \left\{ \frac{1}{2}\left(\frac{N t_k}{M}\right)^2 \right\} + 2(k+1),$$

where

$$t_k = \left[\frac{\sum_{i=k+1}^{M} \lambda_i^2}{a(k)^2 (M-k)} - \left(1 + \frac{M}{N}\right) \right] M.$$

- BIC method for large-scale arrays in [16]: The variant of the Bayesian Information Criterion (BIC) [39] for large-scale arrays proposed in [16] is

$$\hat{k}_{BIC} = \underset{0 \le k \le M-1}{\operatorname{argmin}} \; 2(M-k)N \log\left(\frac{a(k)}{g(k)}\right) + P(k, M, N),$$

where

$$P(k, M, N) = Mk\left(\log(2N) - \frac{1}{k}\sum_{i=1}^{k} \log\left(\frac{\lambda_i}{a(k)}\right)\right).$$

Experiment 3: In this example we consider an array with $M = 100$ antennas receiving $K = 3$ sources separated $\Delta_\theta = 10°$, and $N = 60$ snapshots, thus yielding a rank-deficient sample covariance matrix. The Rayleigh limit for this scenario is $2\pi/M \approx 3.6°$, so in this example the sources are well separated.

Fig. 5 shows the probability of correct detection vs. the signal-to-noise-ratio (SNR) for all methods under comparison. Increasing the number of snapshots to $N = 150$ and keeping fixed the rest of the parameters, we obtain the results shown in Fig. 6. For this scenario, where source separations are roughly 3 times the Rayleigh limit, the SA method outperforms competing methods.

Experiment 4: To analyze the impact of the separation between sources, we consider again a scenario with $M = 100$ antennas, and $K = 3$ sources but now separated at angles of $\Delta_\theta = 2°$, which is within the Rayleigh limit of approximately $3.6°$. The results for $N = 150$ and $N = 60$ snapshots are shown in Figs. 7 and 8, respectively. In comparison to the previous examples with sources separated $\Delta_\theta = 10°$, for closely separated sources (below, or close to, the Rayleigh limit), there is a requirement for higher SNRs and/or larger sample support for the methods to perform satisfactorily. Again, the SA method provides the best performance. Also, the LS-MDL method seems to be more robust than NE and BIC for very low sample support scenarios ($N = 60$).

Experiment 5: In this experiment we compare the performance for an increasing number of snapshots when the number of antennas is fixed to $M = 100$ antennas, the signal-to-noise-

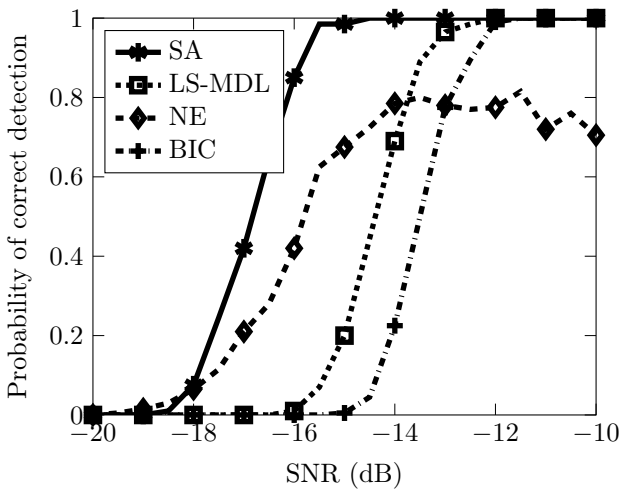

Fig. 5: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources separated $\Delta_\theta = 10°$, the number of antennas is $M = 100$, the number of snapshots is $N = 60$ and $L = \lfloor M - 5 \rfloor$.

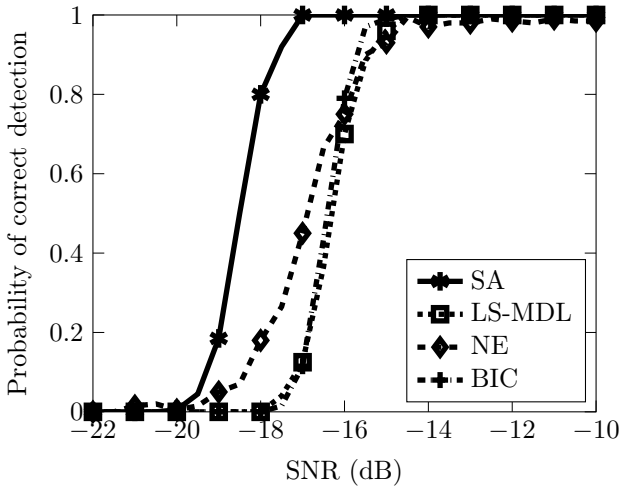

Fig. 6: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources separated $\Delta_\theta = 10°$, the number of antennas is $M = 100$, the number of snapshots is $N = 150$ and $L = \lfloor M - 5 \rfloor$.

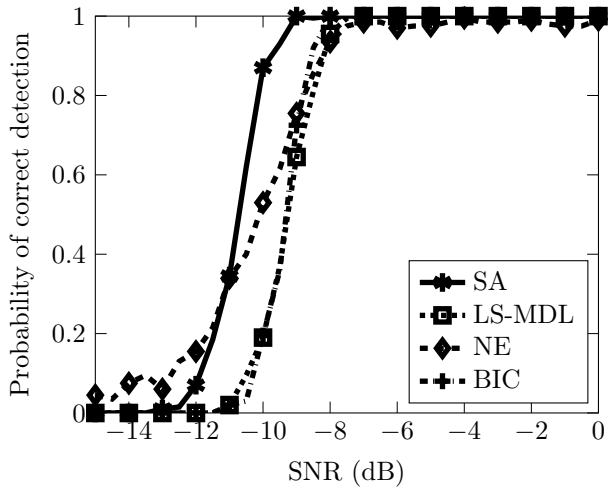

Fig. 7: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources separated $\Delta_\theta = 2°$, the number of antennas is $M = 100$, the number of snapshots is $N = 150$ and $L = \lfloor M - 5 \rfloor$.

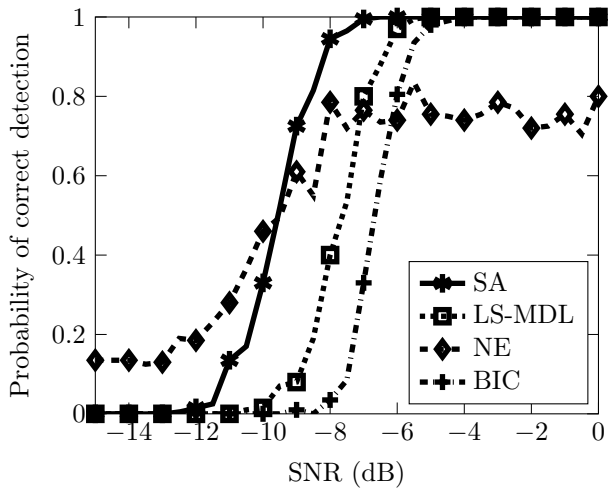

Fig. 8: Probability of correct detection vs. SNR for all methods. In this experiment, there are $K = 3$ sources separated $\Delta_\theta = 2°$, the number of antennas is $M = 100$, the number of snapshots is $N = 60$ and $L = \lfloor M - 5 \rfloor$.

ratio is SNR $= -16$ dB, and there are $K = 3$ uncorrelated sources separated $\Delta_\theta = 10°$. As Fig. 9 shows, the SA method provides very competitive results with only a few snapshots, while the other methods require a much higher number of snapshots to consistently estimate the right number of sources.

Experiment 6: In the last experiment we consider an array with $M = 120$ antennas, the signal-to-noise-ratio is SNR $= -16$ dB, and there are $K = 6$ uncorrelated sources now with separation of $\Delta_\theta = 12°$. As Fig. 10 shows, the SA method starts performing well with very few snapshots.

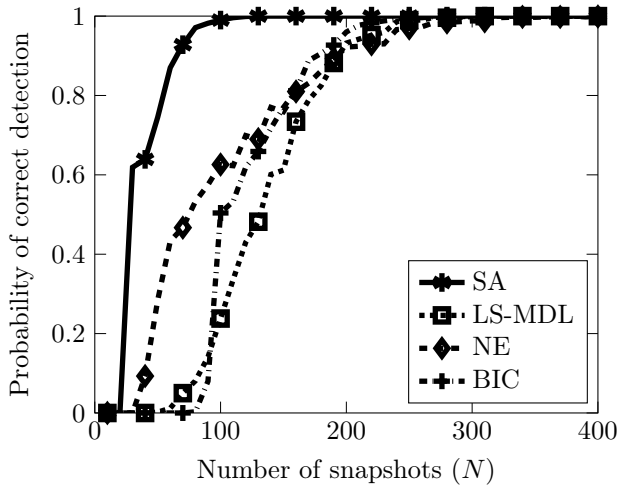

Fig. 9: Probability of correct detection vs. number of snapshots for all methods. In this experiment, there are $K = 3$ sources separated $\Delta_\theta = 10°$, the number of antennas is $M = 100$, $SNR = -16$ dB and $L = \lfloor M - 5 \rfloor$.

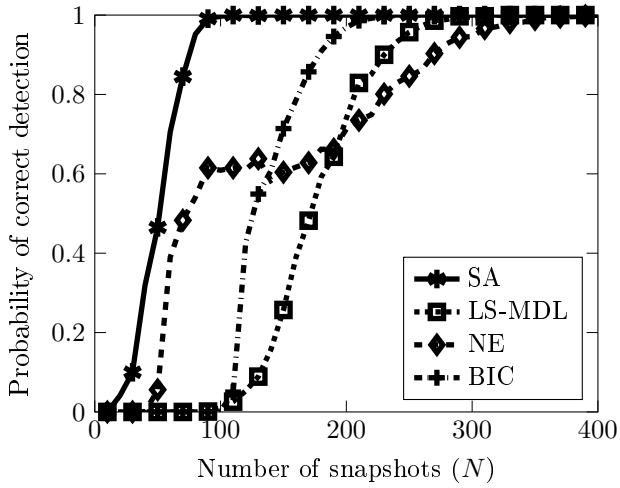

Fig. 10: Probability of correct detection vs. number of snapshots for all methods. In this experiment, there are $K = 6$ sources separated $\Delta_\theta = 12°$, the number of antennas is $M = 120$, $SNR = -16$ dB and $L = \lfloor M - 5 \rfloor$.

Discussion: It may be said that the method NE uses asymptotic results, based on large random matrix theory, to derive an order fitting rule. The rule is then applied to randomly generated eigenvalues computed from finite samples of finite matrices, as if these eigenvalues behaved as the eigenvalues for a large random matrix. The methods LS-MDL (based on MDL), and BIC use geometric and arithmetic means of *sub-dominant eigenvalues*, derived from likelihood formulas, to determine the likelihood of a factor model of a fixed order.

In fact, in the computation of likelihood, it is the *likelihood of a signal covariance matrix* $\mathbf{F}\mathbf{\Lambda}\mathbf{F}^H$*, of order* k*, plus a diagonal noise covariance* $\sigma^2\mathbf{I}$ *of unknown variance* σ^2 *that is computed.* So, in a very real sense, all these methods are based on the likelihood of a *full covariance model for multivariate normal data*, and likelihoods of different models are rank- ordered after penalties for large order are applied. This rank ordering depends critically on the scales of the components $\mathbf{F}\mathbf{\Lambda}\mathbf{F}^H$ and σ^2 that are identified in the likelihood computation. The method SA, treats eigenvalues computed from finite samples of finite matrices as variables that only indicate which model *subspace* could have produced these eigenvalues as the eigenvalues of a corresponding covariance matrix, consisting of a rank k component. Importantly, all eigenvalues are used in a bootstrap, and not only the sub-dominant eigenvalues. Perhaps more importantly, scale is removed from consideration. That is, the methods NE, LS-MDL, BIC account for scale in the fitting of a covariance model to the data, whereas the method of SA is entirely scale-invariant, as it computes scale-invariant probabilities from the eigenvalues and then produces draws of randomly-generated, scale-invariant, subspaces. Subspace modeling seems better matched to the problem of order determination for an array manifold than does covariance modeling. The experiments in this section indicate that this normalization with respect to scale is useful for estimating model order in experiments where the scales of the signal covariance and the noise covariance are unknown, and the SNR and/or sample support are small.

VI. Conclusions

In this paper we have studied the problem of source enumeration from measurements in a uniform linear array. The approach is to extract a subspace from each of several subarrays, and then average these subspaces for a subspace whose dimension is the estimated number of far-field sources. A key element of the method is the automatic order-fitting rule for extracting the dimension of the average subspace that minimizes the mean-squared error between the average and each individual subspace. The net of this procedure is that eigenvalues of a sample covariance matrix determine a distribution on subspaces that could have produced the measured covariance matrix. This procedure normalizes scale by replacing scale-dependent covariance models by scale-invariant subspace models. The method requires no penalty terms for controlling the estimated order.

Simulations indicate performance that is superior to other published methods, over a range of signal-to-noise ratios, sample supports, and source separations. The results suggest that the problem of source enumeration may be viewed as a problem of identifying an approximating subspace, and its dimension, from a set of subspaces estimated from measurements. This point of view stands in contrast to methods that compute likelihood for covariance models, where scale is retained, and then penalize these likelihoods for large dimension.

Acknowledgment

The authors would like to thank Prof. Daniel P. Palomar for his comments and help with the robust formulation of the order estimation problem discussed in Sec. II-D. The authors also wish to acknowledge fruitful discussions of this problem with Prof. Michael Kirby and Prof. Chris Peterson.

References

[1] C. J. C. Burges, "Dimension reduction: a guided tour," *Foundations and Trends in Machine Learning*, vol. 2, no. 4, pp. 275–365, 2009.

[2] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 25, pp. 218–233, 2003.

[3] R. R. Hagege and J. M. Francos, "Universal manifold embedding for geometric deformed functions," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3676–3684, 2016.

[4] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, 1994.

[5] R. Vidal, R. Tron, and R. Hartley, "Multiframe motion segmentation with missing data using power factorization and GPCA," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 85–105, 2008.

[6] R. Vidal, "Subspace clustering," *IEEE Signal Proc. Magazine*, vol. 28, no. 2, pp. 52–68, 2011.

[7] D. Ramirez, G. Vázquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, 2011.

[8] R. H. Gohary and T. N. Davidson, "Noncoherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, 2009.

[9] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[10] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. Pure Appl. Math.*, vol. 5, no. 30, pp. 509–541, 1977.

[11] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2273–2286, 2011.

[12] R. Marrinan, J. R. Beveridge, B. Draper, M. Kirby, and C. Peterson, "Finding the subspace mean or median to fit your need," in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 1082–1089.

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

[14] I. Santamaria, L. L. Scharf, M. Kirby, C. Peterson, and J. Francos, "An order fitting rule for optimal subspace averaging," in *Proc. IEEE Work. Stat. Signal Process. (SSP)*, Palma de Mallorca, Spain, Jun. 2016.

[15] Z. Zhu and S. Kay, "On Bayesian exponentially embeded family for model order selection," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 933–943, 2018.

[16] L. Huang, Y. Xiao, H. C. So, and J.-K. Zhang, "Bayesian information criterion for source enumeration in large-scale adaptive antenna array," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3018–3032, 2016.

[17] R. R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise with relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2625–2638, 2008.

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

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

[20] I. Santamaria, D. Ramirez, and L. L. Scharf, "Subspace averaging for source enumeration in large arrays," in *Proc. IEEE Work. Stat. Signal Process. (SSP)*, Freiburg, Germany, Jun. 2018.

[21] Y. Sun, P. Babu, and P. Palomar, "Majorization-Minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Processing*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, USA: The John Hopkins University Press, 1996.

[23] A. Bjrk and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 37, no. 123, pp. 579–594, 1973.

[24] A. Edelman, T. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[25] F. Hlawatsch and W. Kozek, "Time-frequency projection filters and time-frequency signal expansions," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3321–3334, 1994.

[26] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *arXiv:1803.01013v1*, 2018.

[27] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[28] J. T. Barron, "A more general robust loss function," *arXiv:1701.03077v5*, 2017.

[29] Y. Chikuse, *Statistics on Special Manifolds*. New York, USA: Springer-Verlag, 2003.

[30] C. G. Khatri and K. V. Mardia, "The von Mises-Fisher matrix distribution in orientation statistics," *J. Roy. Statist. Soc. Ser. B*, vol. 39, no. 1, pp. 95–106, 1977.

[31] Y. Chikuse, "Distributions of orientations on Stiefel manifolds," *J. Multivariate Analysis*, vol. 33, pp. 247–264, 1990.

[32] T. D. Downs, "Orientation statistics," *Biometrika*, vol. 59, pp. 665–676, 1972.

[33] P. D. Hoff, "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate data," *arXiv:0712.4166*, 2013.

[34] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[35] H. V. Trees, *Detection, Estimation and Modulation Theory Part IV: Optimum Array Processing*. New York, NJ: Wiley, 2002.

[36] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.

[37] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, 1985.

[38] J. Rissanen, "Modeling by shortest data description," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 14, pp. 465–471, 1978.

[39] Z. Lu and A. M. Zoubir, "Generalized Bayesian information criterion for source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1470–1480, 2013.

[40] S. Kay, "Exponentially embedded families: New approaches to model order estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 1, pp. 333–345, 2005.

[41] C. Xu and S. Kay, "Source enumeration via the EEF criterion," *IEEE Signal Process. Letters*, vol. 15, pp. 569–572, 2008.

[42] A. Quinlan, J. P. Barbot, P. Larzabal, and M. Haardt, "Model order selection for short data: an exponential fitting test (EFT)," *EURASIP J. Adv. Signal Process.*, vol. 34, pp. 122–148, 2007.

[43] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1557–1567, 1989.

[44] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: General asymptotic performance analysis," *IEEE Trans. Signal Processing*, vol. 50, pp. 1027–1036, 2002.

[45] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, pp. 36–47, 2004.

[46] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. of the IEEE*, vol. 74, pp. 1044–1045, 1986.

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

[48] F. Li and R. J. Vaccaro, "Unified analysis for DOA estimation algorithms in array signal processing," *Signal Process.*, vol. 25, pp. 147–169, 1991.

[49] T. A. Palka, *Bounds and algorithms for subspace estimation on Riemannian quotient submanifolds*. University of Rhode Island: Ph.D. Dissertation, 1996.

[50] T. A. Palka and R. J. Vaccaro, "A differential geometry-based framework for constrained subspace estimation," in *Proc. 2017 Underwater Acoust. Signal Process. Work.*, University of Rhode Island, Oct. 2017.

[51] D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, fourier analysis and uncertainty, i," *The Bell System Technical Journal*, vol. 40, no. 1, pp. 43–63, 1961.

[52] Q. Ding and S. Kay, "Inconsistency of the MDL: On the performance of model order selection criteria with increasing signal-to-noise ratio," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1959–1969, 2011.

[53] P. O. Perry and P. J. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 504–513, 2010.

[54] L. Huang and H. C. So, "Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4806–4821, 2013.

[55] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. J. Math. Stat.*, vol. 34, pp. 122–148, 1963.