Grid-less DOA Estimation Using Sparse Linear Arrays Based on Wasserstein Distance

Mianzhi Wang, *Student Member, IEEE,* Zhen Zhang, *Student Member, IEEE,* and Arye Nehorai, *Life Fellow, IEEE*

***Abstract*—Sparse linear arrays, such as nested and co-prime arrays, are capable of resolving $O(M^2)$ sources using only $O(M)$ sensors by exploiting their so-called difference coarray model. One popular approach to exploit the difference coarray model is to construct an augmented covariance matrix from the sample covariance matrix. By applying common direction-of-arrival (DOA) estimation algorithms to this augmented covariance matrix, more sources than the number of sensors can be identified. In this letter, inspired by the optimal transport theory, we develop a new approach to construct this augmented covariance matrix. We formulate a structured covariance estimation problem that minimizes the Bures-Wasserstein distance between the sample covariance matrix and the subsampled augmented covariance matrix, which can be either casted to a semi-definite programming problem, or directly solved using gradient-based methods. Our approach contributes to a new grid-less DOA estimation algorithm for sparse linear arrays. Numerical examples show that our approach achieves state-of-art estimation performance.**

***Index Terms*—Direction-of-arrival estimation, sparse linear arrays, co-prime and nested arrays, Wasserstein distance, convex optimization**

I. Introduction

DIRECTION-OF-ARRIVAL (DOA) estimation, which recovers the direction information of multiple incoming waves using the measurements of sensor arrays, is one of the major problems in array signal processing [1]–[3]. With uniform linear arrays (ULAs), classical DOA estimation methods, such as MUtiple SIgnal Classification (MUSIC) [4], can resolve up to $M-1$ sources using M sensors. With sparse linear arrays, such as minimum redundancy arrays (MRAs) [5], [6], nested arrays [7], [8], and co-prime arrays [9]–[12], up to $O(M^2)$ uncorrelated sources can be resolved using M sensors by exploiting the difference coarray model.

Various DOA estimation algorithms can be applied to exploit the difference coarray model. These methods can be grouped into two categories: grid-based methods and grid-less methods. Grid-based methods discretize the parameter space of interest into a fine grid and recover the DOAs using sparse enforcing techniques [13]–[18]. On the contrary, grid-less methods do not require the parameters of interest lie on a predefined grid. Therefore, their estimation errors are not limited by the grid size. Many grid-less methods rely on the construction of the covariance matrix of the difference coarray model, namely the augmented covariance matrix, whose size is usually much larger than the sample covariance matrix of the physical array model. Based on this augmented covariance matrix, more sources than the number of sources can be resolved. Various methods of constructing such an augmented covariance matrix are developed based on spatial-smoothing [7], [19], nuclear-norm minimization [20], maximum-likelihood function with Toeplitz parameterization [21], asymptotic efficient covariance fitting criterion [22], and virtual array interpolation [23].

In this letter, we explore a new way of constructing the augmented covariance matrix inspired by the Wasserstein distance between zero-mean Gaussians, which originates in the optimal transport theory [24], [25]. The Wasserstein distance between zero-mean Gaussians introduces a metric over the space of covariance matrices. Using this metric, we formulate a structured covariance matrix estimation problem to construct the augmented covariance matrix, which can be cast to a semi-definite programming (SDP) problem. We can then estimate the DOAs based on the constructed augmented covariance matrix. To our best knowledge, this is the first time optimal transport theory has been applied to sparse linear arrays, which opens up a new direction for developing DOA estimators for such arrays. Our formulation provides a new grid-less and regularization-free DOA estimation method for sparse linear arrays that achieves state-of-art performance.

Notations: We use $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote a complex circularly-symmetric Gaussian distribution whose mean is $\boldsymbol{\mu}$ and covariance is $\boldsymbol{\Sigma}$. Given a matrix $\boldsymbol{A}$, we use $\|\boldsymbol{A}\|_F$ to denote its Frobenius norm. Given a Hermitian matrix $\boldsymbol{H}$, $\boldsymbol{H} \succeq 0$ means $\boldsymbol{H}$ is positive semi-definite (PSD). We use $\otimes$, $\odot$, and $\circ$ to denote the Kronecker product, the Khatri-Rao product, and the Hadamard product, respectively. We use $\mathbb{H}^n$ and $\mathbb{T}^n$ to denote the set of $n \times n$ Hermitian and Toeplitz matrices, respectively. Given a set of $n \times n$ matrices, $\mathbb{S}^n$, we use $\mathbb{S}^n_+$ to denote its subset of PSD matrices.

II. A Review of Direction Finding Using Sparse Linear Arrays

We consider an M-sensor sparse linear array whose sensors are placed on a grid with grid size d_0. We denote the sensor locations by $\mathcal{D} = \{\bar{d}_1 d_0,\ \bar{d}_2 d_0,\ \ldots,\ \bar{d}_M d_0\}$, where $\bar{d}_i$ are integers. Without loss of generality, we assume that the first sensor is located at the origin (i.e., $\bar{d}_1 = 0$), and that $\bar{d}_1 < \bar{d}_2 < \cdots < \bar{d}_M$. Such a sparse linear array can be viewed as a thinned ULA of $M_0 = \bar{d}_M + 1$ sensors (note that $\{\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_M\} \subseteq \{\bar{d}_1, \bar{d}_1 + 1, \ldots, \bar{d}_M\}$). Assume that K far-field narrow-band sources impinge on the array from the

This work was supported by the ONR Grant No. N00014-13-1-0050 and No. N00014-17-1-2371.

M. Wang, Z. Zhang, and A. Nehorai are with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: mianzhi.wang@wustl.edu; zhen.zhang@wustl.edu; nehorai@ese.wustl.edu).

directions $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^T$. The N snapshots received by the array can be expressed as

$$\boldsymbol{y}(t) = \boldsymbol{S}\boldsymbol{A}_0(\boldsymbol{\theta})\boldsymbol{x}(t) + \boldsymbol{n}(t),\ t = 1, 2, \ldots, N. \tag{1}$$

Here $\boldsymbol{x}(t)$ denotes the source signals, and $\boldsymbol{n}(t)$ denotes the additive noise. $\boldsymbol{A}_0(\boldsymbol{\theta}) = [\boldsymbol{a}_0(\theta_1), \boldsymbol{a}_0(\theta_2), \ldots, \boldsymbol{a}_0(\theta_K)]$ is the steering matrix of an M_0-sensor ULA, where

$$\boldsymbol{a}_0(\theta_k) = \begin{bmatrix} 1 & e^{j\frac{2\pi}{\lambda}d_0 \sin\theta_k} & \cdots & e^{j\frac{2\pi}{\lambda}(M_0-1)d_0 \sin\theta_k} \end{bmatrix},$$

and λ denotes the wavelength of the carrier wave. $\boldsymbol{S} \in \mathbb{R}^{M\times M_0}$ is a sensor selection matrix encoding the thinning process, where $S_{mn} \in \{0, 1\}$ and $S_{mn} = 1$ only if the n-th sensor in the full ULA is selected as the m-th sensor in the sparse linear array. Typically, the inter-element spacing d_0 is chosen to be $\lambda/2$ to avoid grating lobes.

We consider the stochastic model [26] and assume that the sources are uncorrelated. Under this signal model,

1) The source signals, $\boldsymbol{x}(t) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{P})$, where $\boldsymbol{P} = \mathrm{diag}(p_1, p_2, \ldots, p_K)$ and p_k denotes the power of the k-th source;
2) The additive noise vectors, $\boldsymbol{n}(t)$, are independently and identically distributed as $\mathcal{CN}(\boldsymbol{0}, \sigma\boldsymbol{I}_M)$;
3) Both the source signals and the noise are temporally uncorrelated, and the source signals are uncorrelated from the additive noise.

Consequently, $\boldsymbol{y}(t)$ are i.i.d. complex circularly-symmetric Gaussians with mean zero and covariance $\boldsymbol{R}$, where

$$\boldsymbol{R} = \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^H] = \boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T = \boldsymbol{S}\boldsymbol{A}_0\boldsymbol{P}\boldsymbol{A}_0^H\boldsymbol{S}^H + \sigma\boldsymbol{I}_M. \tag{2}$$

Here $\boldsymbol{R}_0 = \boldsymbol{A}_0\boldsymbol{P}\boldsymbol{A}_0^H + \sigma\boldsymbol{I}$ is the covariance matrix of an M_0-sensor ULA. Because $\boldsymbol{A}_0$ is the steering matrix of a ULA and $\boldsymbol{P}$ is diagonal, $\boldsymbol{R}_0$ is a Toeplitz matrix.

Vectorizing $\boldsymbol{R}$ leads to

$$\boldsymbol{r} = \mathrm{vec}(\boldsymbol{R}) = (\boldsymbol{S} \otimes \boldsymbol{S})(\boldsymbol{A}_0^* \odot \boldsymbol{A}_0)\boldsymbol{P} + \sigma\,\mathrm{vec}(\boldsymbol{I}_M). \tag{3}$$

Here $(\boldsymbol{S} \otimes \boldsymbol{S})(\boldsymbol{A}_0^* \odot \boldsymbol{A}_0)$ can be viewed as the steering matrix of a virtual array whose sensor locations are given by $\mathcal{D}_{\mathrm{co}} = \{(\bar{d}_m - \bar{d}_n)d_0 | m, n = 1, 2, \ldots, M\}$. This virtual array is usually referred as "different coarray". Consequently, $\boldsymbol{r}$ resembles a measurement vector of this difference coarray. If $md_0 \in \mathcal{D}_{\mathrm{co}}\ \forall m = 0, 1, \ldots, M_0 - 1$, we say the difference coarray is "hole-free". It has been shown that, given the sample covariance matrix $\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{y}(t)\boldsymbol{y}^H(t)$, by exploiting the difference coarray structure of "hole-free" a sparse linear array, one can construct an augmented covariance matrix that provides a good estimate of $\boldsymbol{R}_0$ [7]. Because M_0 is typically much larger than M, by applying MUSIC or other spectral estimation algorithms, we can resolve many more sources than the number of sensors.

Remark 1: Some sparse linear arrays, such as co-prime arrays, are not "hole-free", and only a smaller submatrix of $\boldsymbol{R}_0$ can be estimated from $\hat{\boldsymbol{R}}$. As long as the size of this submatrix is greater than M, more sources than the number of sensors can still be resolved. For brevity, we consider only "hole-free" arrays in the following discussion, which can be easily extended to arrays that are not "hole-free".

In general, the process of recovering $\boldsymbol{R}_0$ from $\hat{\boldsymbol{R}}$ can be described with the following optimization problem:

$$\min_{\boldsymbol{R}_0}\ d(\hat{\boldsymbol{R}}, \boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T) + L(\boldsymbol{R}_0),\ \text{s.t. } \boldsymbol{R}_0 \in \mathcal{C}, \tag{P1}$$

where $d(\cdot,\cdot)$ is some similarity measure defined over $\mathbb{H}_+^M$, $L(\boldsymbol{R}_0)$ captures the regularization terms, and $\mathcal{C}$ denotes the constraints.

One simple method of estimating $\boldsymbol{R}_0$ consists of only two steps: redundancy averaging and direct augmentation [27]–[29]. Denoting this estimated $\boldsymbol{R}_0$ by $\hat{\boldsymbol{R}}_{\mathrm{DA}}$, it is straightforward to show that $\hat{\boldsymbol{R}}_{\mathrm{DA}}$ is the solution to (P1) by choosing $d(\cdot,\cdot)$ as

$$d_{\mathrm{F}}(\boldsymbol{R}_1, \boldsymbol{R}_2) = \|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_F, \tag{4}$$

removing the regularization $L(\boldsymbol{R}_0)$, and setting $\mathcal{C} = \mathbb{T}^{M_0}$. By apply MUSIC to $\hat{\boldsymbol{R}}_{\mathrm{DA}}$, we recover the DA-MUSIC algorithm [19], [29]. One issue with $\hat{\boldsymbol{R}}_{\mathrm{DA}}$ is that it is not guaranteed to be PSD. To tackle this issue, spatial smoothing (SS) instead of direct augmentation can be performed as the second step. The resulting estimate, denoted by $\hat{\boldsymbol{R}}_{\mathrm{SS}}$, is equal to $\hat{\boldsymbol{R}}_{\mathrm{DA}}\hat{\boldsymbol{R}}_{\mathrm{DA}}^H/M_0$. By apply MUSIC to $\hat{\boldsymbol{R}}_{\mathrm{SS}}$, we recover the SS-MUSIC algorithm [7]. It is also possible to introduce regularizations to further exploit the structure of $\boldsymbol{R}_0$. Because $\boldsymbol{R}_0 - \sigma\boldsymbol{I}$ is low-rank, one popular choice is the nuclear norm [20], [30]. Additional constraints, such as the ones based on correlation subspaces [31], [32], can be introduced if there exists some prior knowledge on the source directions. It should be noted that d_{F} does not capture the geometry of the space of covariance matrices. Therefore, DOA estimates obtained from $\hat{\boldsymbol{R}}_{\mathrm{DA}}$ or $\hat{\boldsymbol{R}}_{\mathrm{SS}}$ are generally statistically inefficient [29].

Because $\boldsymbol{y}(t)$ follows a zero-mean complex Gaussian distribution, another popular choice of $d(\cdot,\cdot)$ is the KL-divergence between two zero-mean complex Gaussian distributions:

$$d_{\mathrm{KL}}(\boldsymbol{R}_1, \boldsymbol{R}_2) = \mathrm{tr}(\boldsymbol{R}_2^{-1}\boldsymbol{R}_1) + \log\frac{\det\boldsymbol{R}_2}{\det\boldsymbol{R}_1} - M. \tag{5}$$

After removing the constants and the regularization term, and setting $\mathcal{C} = \mathbb{T}_+^{M_0} \cap \mathbb{H}^{M_0}$, (P1) becomes

$$\begin{aligned} &\min_{\boldsymbol{R}_0}\ \mathrm{tr}((\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T)^{-1}\hat{\boldsymbol{R}}) + \log\det(\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T), \\ &\text{s.t. } \boldsymbol{R}_0 \in \mathbb{T}_+^{M_0} \cap \mathbb{H}^{M_0}, \end{aligned} \tag{P2}$$

which corresponds to the maximum-likelihood based structured covariance estimation problem. The problem (P2) is non-trivial, whose suboptimal solutions can be obtained via fixed-point like iterations [21], [33]. To tackle the difficulty in solving (P2), the authors of [16] and [22] suggested using the following covariance fitting criterion instead of d_{KL}:

$$d_{\mathrm{SPA}}(\boldsymbol{R}_1, \boldsymbol{R}_2) = \|\boldsymbol{R}_2^{-\frac{1}{2}}(\boldsymbol{R}_1 - \boldsymbol{R}_2)\boldsymbol{R}_1^{-\frac{1}{2}}\|_F. \tag{6}$$

In [22], the authors showed that resulting optimization problem can be converted to a SDP problem, which can be numerically solved with off-the-shelf solvers such as SDPT3 [34].

III. Wasserstein Distance Inspired Covariance Matrix Estimation

In this section, we introduce our new Wasserstein distance based approach, show its SDP formulation, and derive the

gradients of the resulting optimization problem. Given two probability distributions, μ and ν, on $\mathbb{R}^n$, the Kantorovich's formulation of the optimal transport problem minimizes the following objective function:

$$\inf_{q\in\mathcal{Q}(\mu,\nu)}\int_{\mathbb{R}^n\times\mathbb{R}^n}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\mathrm{d}q(\boldsymbol{x},\boldsymbol{y}), \tag{7}$$

where the cost function is chosen to be $\|\cdot\|_2$, $\mathcal{Q}(\mu,\nu)$ is the set of joint probability distributions on $\mathbb{R}^n\times\mathbb{R}^n$ whose marginals are given by μ and ν, respectively. The square root of (7) defines the so-called 2-Wasserstein distance over the probability distributions on $\mathbb{R}^n$ [24]. When both μ and ν are proper zero-mean Gaussians[1] with covariance matrix $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, respectively, the optimization problem (7) has a closed-form solution [35], $d^2_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$, which is given by

$$d^2_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)=\mathrm{tr}\left(\boldsymbol{R}_1+\boldsymbol{R}_2-2(\boldsymbol{R}_1^{\frac{1}{2}}\boldsymbol{R}_2\boldsymbol{R}_1^{\frac{1}{2}})^{\frac{1}{2}}\right), \tag{8}$$

where $(\cdot)^{\frac{1}{2}}$ denotes the principal matrix square root[2]. It has been shown that $d_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$ introduces a metric on $\mathbb{H}^n_+$, namely the Bures-Wasserstein metric [36]. It is straightforward to verify that $d^2_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$ is generally not equal to $d^2_{\mathrm{WG}}(\boldsymbol{R}_1-\boldsymbol{R}_2,\boldsymbol{0})$. Hence, $d^2_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$ captures the non-Euclidean geometry of covariance matrices.

Combining (P1) and (8), we can formulate a structured covariance estimation problem while considering the geometry of covariance matrices:

$$\begin{aligned}&\min_{\boldsymbol{R}_0}\ \mathrm{tr}\left(\hat{\boldsymbol{R}}+\boldsymbol{R}-2(\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}})^{\frac{1}{2}}\right),\\&\text{s.t. }\boldsymbol{R}=\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T,\boldsymbol{R}_0\in\mathbb{T}_+^{M_0}\cap\mathbb{H}^{M_0}.\end{aligned} \tag{P3}$$

Note that $\boldsymbol{R}_0\succeq 0$ guarantees that $\boldsymbol{R}=\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T\succeq 0$.

We next analyze the convexity of (P3). Here the constraint set $\mathbb{T}_+^{M_0}\cap\mathbb{H}^{M_0}$ is a convex cone, and $\mathrm{tr}(\boldsymbol{R})$ is linear function of $\boldsymbol{R}_0$. Because that $\mathrm{tr}(\boldsymbol{X}^{\frac{1}{2}})$ is strictly convex when $\boldsymbol{X}\in\mathbb{H}^n_+$ [36], and that $\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}}$ is a linear function of $\boldsymbol{R}_0$, $\mathrm{tr}\left((\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}})^{\frac{1}{2}}\right)$ is convex in $\boldsymbol{R}_0$. Therefore, the optimization problem (P3) is convex. More specifically, when the array is "hole-free" and $\hat{\boldsymbol{R}}$ is positive definite, $\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T\hat{\boldsymbol{R}}^{\frac{1}{2}}$ defines a bijection. Consequently, (P3) is strictly convex and admits a unique solution.

Analytically solving (P3) turns out to be non-trivial due to the matrix square root operation. It has been shown in [35] that $d^2_{\mathrm{WG}}(\boldsymbol{R}_1,\boldsymbol{R}_2)$ can be obtained by solving the following optimization problem:

$$\min_{\boldsymbol{V}}\ \mathrm{tr}(\boldsymbol{R}_1+\boldsymbol{R}_2-\boldsymbol{V}-\boldsymbol{V}^H),\ \text{s.t. }\begin{bmatrix}\boldsymbol{R}_1&\boldsymbol{V}\\\boldsymbol{V}^H&\boldsymbol{R}_2\end{bmatrix}\succeq 0. \tag{9}$$

Therefore, we can cast (P3) into the following SDP problem:

$$\begin{aligned}&\min_{\boldsymbol{R}_0,\boldsymbol{V}}\ \mathrm{tr}(\hat{\boldsymbol{R}}+\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T-\boldsymbol{V}-\boldsymbol{V}^H)\\&\text{s.t. }\begin{bmatrix}\boldsymbol{S}\boldsymbol{R}_0\boldsymbol{S}^T&\boldsymbol{V}\\\boldsymbol{V}^H&\hat{\boldsymbol{R}}\end{bmatrix}\succeq 0,\boldsymbol{R}_0\in\mathbb{T}_+^{M_0}\cap\mathbb{H}^{M_0},\end{aligned} \tag{P4}$$

[1]An n-dimensional complex circularly-symmetric Gaussian distribution can be fully characterized by a $2n$-dimensional real Gaussian distribution. Therefore, similar results follow for complex $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ [35].

[2]Given a PSD matrix $\boldsymbol{A}$ with eigendecomposition $\boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^H$, where $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$, its principal matrix square root is $\boldsymbol{E}\,\mathrm{diag}(\lambda_1^{\frac{1}{2}},\ldots,\lambda_n^{\frac{1}{2}})\boldsymbol{E}^H$.

which can be numerically solved via the interior-point method [37], [38]. The optimization problem (P4) provides a regularization free approach to estimate $\boldsymbol{R}_0$ from $\hat{\boldsymbol{R}}$, resulting in a new grid-less method for underdetermined ($K\geq M$) DOA estimation using sparse linear arrays. When the number of snapshots, N, is less than M, $\hat{\boldsymbol{R}}$ is rank-deficient. Nevertheless, (P4) can still estimate $\boldsymbol{R}_0$ in such cases. However, the resulting estimates are generally rank-deficient.

The size of the SDP problem (P4) is similar to that of the SDP formulation of SPA in [22]. Therefore, the worst case (assuming $M_0=M$) complexity of solving (P4) is $\mathcal{O}(M_0^{6.5})$, which can be expensive for large sparse linear arrays.

Alternatively, we can directly reparameterize $\boldsymbol{R}_0$ by expressing it as the sum of $2M_0-1$ linear-independent matrices [21]: $\boldsymbol{R}_0=\sum_{i=1}^{2M_0-1}c_i\boldsymbol{Q}_i$, where

$$\boldsymbol{Q}_i=\begin{cases}\boldsymbol{I}_{M_0}, & i=1,\\ \boldsymbol{I}_{M_0}^{(i-1)}+(\boldsymbol{I}_{M_0}^{(i-1)})^T, & 2\leq i\leq M_0,\\ j\boldsymbol{I}_{M_0}^{(i-M_0)}-j(\boldsymbol{I}_{M_0}^{(i-M_0)})^T, & M_0<i\leq 2M_0-1,\end{cases}$$

where c_i are coefficients to be estimated, $\boldsymbol{I}_{M_0}^{(i)}$ denotes the $M_0\times M_0$ matrix whose elements are zeros except for the i-th upper diagonal (i.e., the (m,n)-th element of $\boldsymbol{I}_{M_0}^{(i)}$ is one only if $m-n=i$). We can also replace the hard constraint, $\boldsymbol{R}_0\succeq 0$, with a soft one using the log-barrier function, resulting in the following unconstrained convex optimization problem:

$$\min_{\boldsymbol{c}}\ \mathrm{tr}\left(\hat{\boldsymbol{R}}+\boldsymbol{R}-2(\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}})^{\frac{1}{2}}\right)-\mu\log\det\boldsymbol{R}_0, \tag{P5}$$

where $\mu>0$ controls the steepness of the log-barrier function.

To evaluate the gradient of this objective function with respect to c_i, we need to evaluate the gradients of following three terms: $\mathrm{tr}(\boldsymbol{R})$, $\mathrm{tr}\left(-2(\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}})^{\frac{1}{2}}\right)$, and $-\mu\log\det\boldsymbol{R}_0$, which we denote as $g_i^{(1)}$, $g_i^{(2)}$, and $g_i^{(3)}$, respectively. By matrix calculus [39], it is straightforward to show that:

$$g_i^{(1)}=\mathrm{tr}(\boldsymbol{S}\boldsymbol{Q}_i\boldsymbol{S}^T),\ g_i^{(3)}=-\mu\,\mathrm{tr}(\boldsymbol{R}_0^{-1}\boldsymbol{Q}_i).$$

Let $\boldsymbol{V}=(\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}})^{\frac{1}{2}}$. Differentiating both side of $\boldsymbol{V}\boldsymbol{V}=\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{R}\hat{\boldsymbol{R}}^{\frac{1}{2}}$ leads to

$$\frac{\partial\boldsymbol{V}}{\partial c_i}\boldsymbol{V}+\boldsymbol{V}\frac{\partial\boldsymbol{V}}{\partial c_i}=\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{Q}_i\boldsymbol{S}^T\hat{\boldsymbol{R}}^{\frac{1}{2}}, \tag{10}$$

which is a Sylvester equation. Because $\boldsymbol{V}$ is positive definite (note that the log-determinant term enforces positive definiteness), it admits the eigendecomposition $\boldsymbol{V}=\boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^H$. Let $\boldsymbol{X}=\boldsymbol{E}^H(\partial\boldsymbol{V}/\partial c_i)\boldsymbol{E}$. Because $\boldsymbol{E}$ is unitary, we can rewrite (10) as

$$\boldsymbol{X}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{X}=\boldsymbol{E}^H\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{Q}_i\boldsymbol{S}^T\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{E}.$$

Since $\boldsymbol{\Lambda}$ is a diagonal matrix, we can rewrite $\boldsymbol{X}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{X}$ as $(\boldsymbol{\Lambda}\mathbf{1}\mathbf{1}^T+\mathbf{1}\mathbf{1}^T\boldsymbol{\Lambda})\circ\boldsymbol{X}$, where $\mathbf{1}\mathbf{1}^T$ represents a square matrix of ones with proper size. Noting that $\boldsymbol{E}$ is unitary, we have $\mathrm{tr}(\boldsymbol{X})=\mathrm{tr}(\partial\boldsymbol{V}/\partial c_i)$. Therefore,

$$g_i^{(3)}=-2\,\mathrm{tr}\left((\boldsymbol{E}^H\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{Q}_i\boldsymbol{S}^T\hat{\boldsymbol{R}}^{\frac{1}{2}}\boldsymbol{E})\circ(\boldsymbol{\Lambda}\mathbf{1}\mathbf{1}^T+\mathbf{1}\mathbf{1}^T\boldsymbol{\Lambda})^{-1}\right)$$

The full gradient of (P5) with respect to c_i is given by $g_i^{(1)}+g_i^{(2)}+g_i^{(3)}$, and gradient-based methods can then be applied to solve (P5).

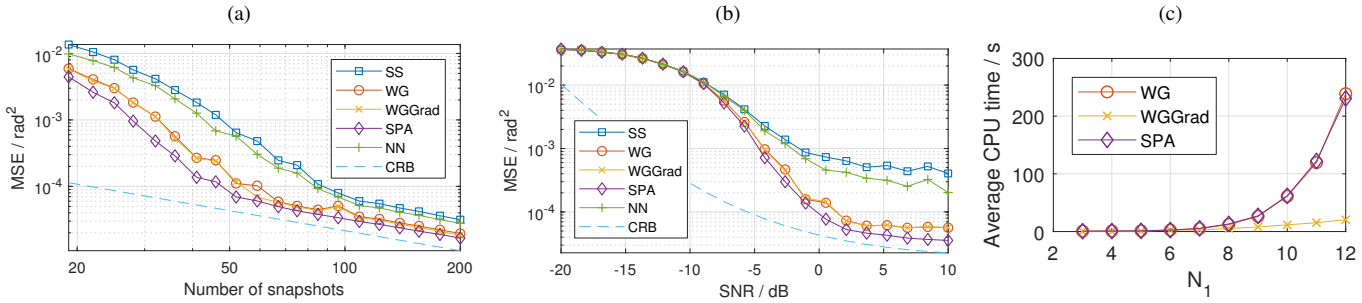

Fig. 1: (a) MSE vs. number of snapshots, N, for different DOA estimation methods, averaged from 3000 Monte Carlo trials; (b) MSE vs. SNR for different DOA estimation methods, averaged from 3000 Monte Carlo trials; (c) CPU time vs. N_1 for three different algorithms. In (a) and (b), the solid lines with markers denote the MSEs, and the dashed line denotes the CRB.

Noting that $g_i^{(1)}$ can be precomputed, the worst case computational complexity is mainly determined by $g_i^{(2)}$ and $g_i^{(3)}$. Because eigendecomposition and matrix multiplication are both $O(n^3)$ operations, $g_i^{(2)}$ and $g_i^{(3)}$ cost at most $O(M_0^3)$ operations. Because i ranges from 1 to $2M_0-1$, the worst case computational complexity of gradient-based methods is $O(kM_0^4)$, where k is the number of descent steps.

We close this section by considering an interesting case: $K=1$ and $\sigma=0$. In such a case, $\boldsymbol{R}=p\boldsymbol{a}(\theta)\boldsymbol{a}(\theta)$ is a rank one matrix parameterized by a single parameter, θ, and $\boldsymbol{R}^{\frac{1}{2}}=\sqrt{p/M}\boldsymbol{a}(\theta)\boldsymbol{a}(\theta)$. Given the sample covariance matrix, $\hat{\boldsymbol{R}}$, we have

$$\begin{aligned}\min_\theta\ d^2_{\mathrm{WG}}(\boldsymbol{R},\hat{\boldsymbol{R}}) &\iff \min_\theta\ pM-\frac{p}{M}\operatorname{tr}\left((\boldsymbol{a}\boldsymbol{a}^H\hat{\boldsymbol{R}}\boldsymbol{a}\boldsymbol{a}^H)^{\frac{1}{2}}\right)\\ &\iff \max_\theta\ (\boldsymbol{a}^H\hat{\boldsymbol{R}}\boldsymbol{a})^{\frac{1}{2}}\operatorname{tr}\left(\frac{1}{\sqrt{M}}(\boldsymbol{a}\boldsymbol{a}^H)\right)\\ &\iff \max_\theta\ \boldsymbol{a}^H\hat{\boldsymbol{R}}\boldsymbol{a},\end{aligned}$$

which is the objective function of a conventional beamformer [2]. This result gives a geometric interpretation of the conventional beamformer. By maximizing $\boldsymbol{a}^H\hat{\boldsymbol{R}}\boldsymbol{a}$, we are finding the DOA that minimizes the Bures-Wasserstein distance between the sample covariance matrix $\hat{\boldsymbol{R}}$ and the structured covariance matrix $\boldsymbol{R}=p\boldsymbol{a}(\theta)\boldsymbol{a}^H(\theta)$.

IV. Numerical Examples

We consider a nested array generated by the parameter pair $(4,4)$: $[0,1,2,3,4,9,14,19]d_0$. We uniformly place nine off-grid sources, which is greater than the number of sensors, within the range $(-70\pi/219, 76\pi/219)$. We consider the following methods to estimate $\boldsymbol{R}_0$:

1) SS: Spatial smoothing after redundancy averaging [29].
2) NN: Nuclear-norm regularized reconstruction [30]. As suggested in the paper, we set the regularization parameter to $2.5\times10^{-3}/((\log N)^2\log M)$.
3) SPA: Sparse and parametric approach [22].
4) WG: Solving the SDP formulation (P4).
5) WGGrad: Solving (P5) using gradient descent with backtracking . We set $\mu=0.01$ and stop when the relative change in the objective function is less than 10^{-7}.

Once $\boldsymbol{R}_0$ is estimated, the DOAs are obtained using root-MUSIC [40]. For reference, we included the Cramér-Rao bound (CRB) for sparse linear arrays [29], [41]–[43] for comparison.

Fig. 1(a) plots the mean-squared errors (MSEs) of the five methods and the CRB versus the number of snapshots, where the signal-to-noise ratio (SNR) is set to 0 dB in all trials. Fig. 1(b) plots the MSEs of the five methods and the CRB versus the SNR, where the number of snapshots is fixed to 50 in all trials. We observe that the performance of WGGrad well agrees with that of WG in all cases. SPA has the best overall performance because d_{SPA} asymptotically approximates the maximum-likelihood function [22], [44], [45]. The performance of our proposed methods, WS and WSGrad, closely matches that of the SPA, and is significantly better than SS and NN.

To compare the CPU time of WS, WSG, and SPA, we consider nested arrays generated by the parameter pair (N_1,N_1) and vary N_1 from 3 to 12. The number of snapshots, N, is fixed to 50 in all trials. All trials are performed on a server with an Intel Xeon E5-2650v3 CPU. We plot the results in Fig. 1(c). We observe that as the size of the array grows, SPA and WG consumes significantly more CPU time than WGGrad, as predicted by our complexity analysis in Section III. Therefore, it is computationally more efficient to use WGGrad when the sparse linear array is large.

V. Concluding Remarks

In this letter, we proposed a new grid-less DOA estimation algorithm based on the Bures-Wasserstein metric between covariance matrices. We first construct an augmented covariance matrix by minimizing the Bures-Wasserstein distance between the sample covariance matrix and the subsampled augmented covariance matrix, and then estimate the DOAs by applying root-MUSIC to the augmented covariance matrix. We showed that the minimization problem can be solved with either SDP or gradient-based methods. It is observed through numerical examples that our algorithm achieves state-of-art DOA estimation performance. Our result will inspire further research on the application of optimal transport theory to array signal processing. Possible future research includes analyzing the connection between our algorithm and the SPA algorithm as well as developing more efficient numerical methods.

REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.

[2] H. L. Van Trees, *Optimum array processing*, ser. Detection, estimation, and modulation theory / Harry L. Van Trees. New York: Wiley, 2002, no. 4.

[3] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, N.J: Pearson/Prentice Hall, 2005.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[5] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Trans. Antennas Propag.*, vol. 16, no. 2, pp. 172–175, Mar. 1968.

[6] M. Ishiguro, "Minimum redundancy linear arrays for a large number of antennas," *Radio Science*, vol. 15, no. 6, pp. 1163–1170, Nov. 1980.

[7] P. Pal and P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[8] K. Han and A. Nehorai, "Wideband gaussian source processing using a linear nested array," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1110–1113, Nov. 2013.

[9] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the music algorithm." IEEE, Jan. 2011, pp. 289–294.

[10] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5565–5576, Nov. 2014.

[11] E. BouDaher, Y. Jia, F. Ahmad, and M. G. Amin, "Multi-frequency co-prime arrays for high-resolution direction-of-arrival estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3797–3808, Jul. 2015.

[12] S. Qin, Y. D. Zhang, M. G. Amin, and A. M. Zoubir, "Generalized coprime sampling of Toeplitz matrices for spectrum estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 81–94, Jan. 2017.

[13] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[14] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[15] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4489–4493, Sep. 2011.

[16] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[17] Y. Zhang, M. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 3967–3971.

[18] P. Gerstoft, C. F. Mecklenbrauker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[19] C.-L. Liu and P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Process. Lett.*, vol. 22, no. 9, Sep. 2015.

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

[21] M. Wang, Z. Zhang, and A. Nehorai, "Direction finding using sparse linear arrays with missing data," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3066–3070.

[22] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4959–4973, Oct. 2014.

[23] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5956–5971, Nov. 2018.

[24] C. Villani, *Topics in optimal transportation*, ser. Graduate studies in mathematics. Providence, RI: American Mathematical Society, 2003, no. v. 58.

[25] Z. Zhang, M. Wang, Y. Huang, and A. Nehorai, "Aligning infinite-dimensional covariance matrices in reproducing kernel Hilbert spaces for domain adaptation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3437–3445.

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

[27] A. Gorokhov, Y. Abramovich, and J. F. Bohme, "Unified analysis of DOA estimation algorithms for covariance matrix transforms," *Signal Processing*, vol. 55, no. 1, pp. 107–115, Nov. 1996.

[28] Y. Abramovich, N. Spencer, and A. Gorokhov, "Detection-estimation of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays—Part I: Fully augmentable arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 959–971, May 2001.

[29] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2017.

[30] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

[31] M. Rahmani and G. K. Atia, "A subspace method for array covariance matrix estimation," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2016, pp. 1–5.

[32] C. Liu and P. P. Vaidyanathan, "Correlation subspaces: Generalizations and connection to difference coarrays," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5006–5020, Oct. 2017.

[33] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, Sep. 1982.

[34] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 – A Matlab software package for semidefinite programming, Version 1.3," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[35] D. C. Dowson and B. V. Landau, "The Fréchet distance between multivariate normal distributions," *Journal of Multivariate Analysis*, vol. 12, no. 3, pp. 450–455, Sep. 1982.

[36] R. Bhatia, T. Jain, and Y. Lim, "On the Bures-Wasserstein distance between positive definite matrices," *Expositiones Mathematicae*, Jan. 2018.

[37] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[38] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK; New York: Cambridge University Press, 2004.

[39] G. A. F. Seber, *A matrix handbook for statisticians*. Hoboken, N.J.: Wiley-Interscience, 2008, oCLC: 191879555.

[40] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 8, Apr. 1983, pp. 336–339.

[41] A. Koochakzadeh and P. Pal, "Cramér Rao bounds for underdetermined source localization," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 919–923, Jul. 2016.

[42] C.-L. Liu and P. P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digital Signal Processing*, vol. 61, pp. 43 – 61, 2017.

[43] M. Wang, Z. Zhang, and A. Nehorai, "Further results on the Cramér-Rao bound for sparse linear arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1493–1507, Mar. 2019.

[44] P. Stoica and T. Söderström, "On reparametrization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, no. 4, pp. 383–387, Aug. 1989.

[45] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, Jul. 1998.