

# On Interference-Rejection Using Riemannian Geometry for Direction of Arrival Estimation

Amitay Bar  and Ronen Talmon , *Senior Member, IEEE*

**Abstract**—We consider the problem of estimating the direction of arrival of desired acoustic sources in the presence of multiple acoustic interference sources. All the sources are located in noisy and reverberant environments and are received by a microphone array. We propose a new approach for designing beamformers and DoA estimation methods based on the Riemannian geometry of the manifold of Hermitian positive definite matrices. Specifically, we show theoretically that incorporating the Riemannian mean of the spatial correlation matrices into frequently-used beamformers gives rise to spatial spectra that reject the directions of interference sources and result in a higher signal-to-interference ratio. We experimentally demonstrate the advantages of our approach in designing several beamformers and a recent DoA estimation method in the presence of simultaneously active multiple interference sources.

**Index Terms**—Array signal processing, direction of arrival estimation, interference rejection, Hermitian positive definite matrices, Riemannian geometry.

## I. INTRODUCTION

ESTIMATION of the direction of arrival (DoA) of an acoustic source is prevalent in signal processing; it is an important step in many tasks, such as source localization, beamforming, source separation, spectrum sensing, and speech enhancement [1], to name but a few. Despite the large research attention it has drawn in the past decades, acoustic DoA estimation is still considered a challenging open problem. Especially in noisy and reverberant environments and in the presence of interference sources, it continues to be an active research field.

Acoustic source localization, and particularly DoA estimation, are often addressed using beamforming [2]. Many beamformers have been proposed over the years for these tasks. One class of beamformers is based on the steered response power (SRP) of a beamformer output. For example, considering the maximum likelihood criterion for a single source, the output power of the beamformer from all the directions is computed, and the DoA is identified as the direction with the maximal

Manuscript received 25 January 2023; revised 16 July 2023; accepted 11 September 2023. Date of publication 12 October 2023; date of current version 20 December 2023. This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant 802735-ERC-DIFFOP. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thomas Oberlin. (*Corresponding author: Amitay Bar.*)

The authors are with the Viterbi Faculty of Electrical and Computer Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: amitayb@campus.technion.ac.il; ronen@ef.technion.ac.il).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TSP.2023.3322779>, provided by the authors

Digital Object Identifier 10.1109/TSP.2023.3322779

power [3], [4], [5], [6]. Another example is the Minimum Variance Distortionless Response (MVDR) beamformer [7], [8], [9], which was first introduced by Capon [10]. The MVDR beamformer extracts the DoA of each of the existing sources, maintaining a unit gain at their direction while minimizing the response from other directions. An important generalization of the MVDR beamformer is the Linearly Constrained Minimum Variance (LCMV) beamformer [11], obtained by minimizing the output power under multiple linear constraints, and can be used for DoA estimation as well [12]. Another line of beamformers is derived based on a subspace approach, i.e., by identifying the subspace of the desired sources, which is assumed to contain only a small portion of the noise and the interference sources. A prominent subspace method, which is also used for DoA estimation, is MUltiple Signal Classification (MUSIC) [13], [14], [15], [16].

A notable algorithm for acoustic source localization is the Steered-Response Power Phase Transform (SRP-PHAT) algorithm proposed in [17]. In SRP-PHAT, the phase transform is used to normalize the different frequencies, such that only their phase information is considered. This allows for the fusion of the different frequencies when considering a broadband signal. A popular time-domain implementation of the SRP-PHAT is the generalized cross-correlation with phase transform (GCC-PHAT) proposed in [18], which normalizes each cross-correlation using the phase transform. In recent years, sparse signal recovery methods have been proposed for DoA estimation [19], [20], [21]. In particular, one approach for solving the sparse recovery problem is the sparse Bayesian learning approach proposed in [22], which was adopted for the DoA estimation problem as well, for example in [23], [24].

In this paper, we consider DoA estimation in a reverberant enclosure consisting of desired sources along with interference sources. We assume that the desired sources are constantly active, whereas the interference sources are only intermittently active. The number of sources, their locations, and their times of activity are all unknown. Consequently, their identification as desired or interference is unknown as well. The power of the different sources is also unknown, and the interference sources could, in fact, be stronger than the desired sources with overlapping activity periods. Our goal is to estimate the DoA of the desired sources in the presence of possibly simultaneously active, multiple interference sources.

This setting poses a major challenge to the common practice in existing methods that rely on maximal power because

estimating the DoA of the strongest sources might result in distinct beams in the direction of the interference sources rather than the desired sources. Furthermore, these beams could mask the beams pointing at the directions of desired sources.

We address this challenge from a geometric standpoint. Our approach relies on the observation that the frequently-used beamformers implicitly consider Euclidean geometry when processing sample correlation matrices. Therefore, since the sample correlation matrices are Hermitian Positive Definite (HPD) matrices, important geometric information is not fully utilized. Instead, we propose a new approach for beamforming design that is based on the Riemannian geometry of the manifold of HPD matrices [25], [26], [27]. Concretely, we analyze the received signal in short time windows and consider the Riemannian mean [25] of the sample correlation matrices in these windows. Then, we leverage particular spectral properties of the Riemannian mean. In [28], it was shown that the Riemannian mean of HPD matrices preserves shared spectral components and attenuates unshared spectral components. Consequently, the continual activity of the desired sources and the intermittent activity of the interference sources enable us to associate desired sources with shared spectral components and interference sources with unshared spectral components. By combining the above, we show that the incorporation of the Riemannian mean into the beamformer design leads to interference rejection, i.e., gives rise to a spatial spectrum that implicitly rejects the beams pointing at the interference sources and preserves the beams pointing at the desired sources. The resulting spectrum is, in turn, used for the estimation of the DoA of the desired sources. Importantly, our approach is applicable to a large number of beamformers used for DoA estimation.

By incorporating our Riemannian approach, we present new implementations of several beamformers: the Delay and Sum (DS) beamformer, subspace-based beamformers, and the MVDR beamformer, as well as the Bayesian learning method for DoA estimation proposed in [23].

The main contributions of this article are as follows. First, we observe that despite being commonly-used in array processing in general, and for DoA estimation in particular, the HPD structure of the signal correlation matrices is not fully exploited. Here, we identify that by dividing the computation of the sample correlation matrix into short-time segments, it can be viewed as the Euclidean mean of the sample correlation matrices of the segments. This observation enables the introduction of the Riemannian geometry of HPD matrices, and in particular, it allows us to introduce a rather simple approach that promotes the use of the Riemannian mean instead of the Euclidean mean and to show its multiple merits. Second, we demonstrate that the proposed approach can naturally be applied to a broad range of DoA estimation methods that use the signal sample correlation matrix, and we show that it is computationally efficient in the sense that it typically does not change the order of the computational complexity of the DoA estimation methods. Third, we theoretically analyze the proposed approach and show that in the case of the DS beamformer, it results in higher SIR values that lead to better DoA estimation accuracy. In addition, we present a noise sensitivity analysis. Fourth, we present empirical

results in adverse conditions that include simultaneously active multiple interference sources. We showcase the applicability of our approach to both classical and recent DoA estimation methods and demonstrate that the obtained performance improvement is by a large margin.

We conclude the introduction with three remarks. First, a similar setting to ours, consisting of desired sources accompanied by interference sources, was considered in [29] and [30] but in the context of signal enhancement. In [29], a single desired source and a single interference source were considered, and in [30], multiple desired sources and multiple interference sources were considered. However, in both works, it was assumed that there is at least one segment for each source, desired or interference, in which it is the only active source. Furthermore, in [30], the number of the desired sources and their activity patterns were assumed to be known. Second, in the context of radar, the Riemannian geometry of the Toeplitz HPD matrices was used in [31] and [32] for target detection by comparing Riemannian distances to a threshold. In the radar settings, [33] estimated the correlation matrix as a linear combination of correlation matrices, with weights that are based on the Riemannian distance. Third, in this article, we demonstrate the Riemannian approach for designing beamformers that reject interference sources for DoA estimation. However, other applications, e.g., signal enhancement, could also benefit from spectra that reject interference sources.

This paper is organized as follows. In Section II, we present a brief background on the HPD manifold. In Section III, we formulate the problem and the setting. In Section IV we describe the proposed approach and present the algorithm for DoA estimation. In Section V, we provide a theoretical analysis of the proposed approach for the DS beamformer. In Section VI, extensions of the approach to other DoA estimation methods are presented. Section VII shows simulation results demonstrating our Riemannian approach. Lastly, we conclude the work in Section VIII.

## II. BACKGROUND ON THE HPD MANIFOLD

An HPD matrix,  $\Gamma \in \mathbb{C}^{n \times n}$ , is a Hermitian matrix, i.e.  $\Gamma = \Gamma^H$ , where  $\Gamma^H$  is the conjugate transpose of  $\Gamma$ , whose real eigenvalues are strictly positive. Associating the space of HPD matrices with the Affine Invariant metric [34] constitutes a Riemannian manifold,  $\mathcal{M}$ . The distance between two matrices  $\Gamma_1$  and  $\Gamma_2$ , induced by the Affine Invariant metric, is given by

$$d_R^2(\Gamma_1, \Gamma_2) = \left\| \log \left( \Gamma_2^{-\frac{1}{2}} \Gamma_1 \Gamma_2^{-\frac{1}{2}} \right) \right\|_F^2, \quad (1)$$

where  $\|\cdot\|_F$  is the Frobenius norm.

The tangent space to a point on a manifold  $\Gamma \in \mathcal{M}$  is a vector space that can be viewed as a local approximation of the manifold around  $\Gamma$ . For the HPD manifold, the tangent space around each  $\Gamma \in \mathcal{M}$  is the vector space of Hermitian matrices of dimensions  $n \times n$ .

The Riemannian mean,  $\Gamma_R$ , of a set of points  $\{\Gamma_i | \Gamma_i \in \mathcal{M}\}$  is defined by the Fréchet mean as the point minimizing the sum

**Algorithm 1** Riemannian mean for the HPD manifold [36]

**Input:** a set of  $K$  HPD matrices  $\{\Gamma_j\}_{j=1}^K$   
**Output:** the Riemannian mean  $\Gamma_R$

- 
- 1: Compute  $\Gamma_R = \frac{1}{K} \sum_{j=1}^K \Gamma_j$
  - 2: **do**
  - 1) Compute the Euclidean mean in the tangent plane:  
 $\bar{\mathbf{P}} = \frac{1}{K} \sum_{j=1}^K \text{Log}_{\Gamma_R}(\Gamma_j)$
  - 2) Update  $\Gamma_R = \text{Exp}_{\Gamma_R}(\bar{\mathbf{P}})$
  - 3) Stop if  $\|\bar{\mathbf{P}}\|_F < \epsilon$
- 

of the distances from all the points in the set as follows

$$\Gamma_R \equiv \arg \min_{\Gamma \in \mathcal{M}} \sum_i d_R^2(\Gamma, \Gamma_i). \quad (2)$$

In general, there is no closed-form expression for the Riemannian mean of more than two matrices, and a solution can be found using an iterative procedure [35], described in Algorithm 1.

The computation of the Riemannian mean requires two maps. The Logarithm map, which maps an HPD matrix  $\Gamma_i \in \mathcal{M}$  to the tangent space of the HPD manifold at  $\Gamma$ , is given by

$$\text{Log}_{\Gamma}(\Gamma_i) = \Gamma^{\frac{1}{2}} \log(\Gamma^{-\frac{1}{2}} \Gamma_i \Gamma^{-\frac{1}{2}}) \Gamma^{\frac{1}{2}}. \quad (3)$$

The Exponential map, which maps a vector  $\mathbf{T}$  from the tangent space at  $\Gamma$ , is given by

$$\text{Exp}_{\Gamma}(\mathbf{T}) = \Gamma^{\frac{1}{2}} \exp(\Gamma^{-\frac{1}{2}} \mathbf{T} \Gamma^{-\frac{1}{2}}) \Gamma^{\frac{1}{2}}. \quad (4)$$

We note that there exists an efficient iterative estimator for the Riemannian mean [37]. It is described in Appendix C. A discussion about the choice of the Affine Invariant metric appears in Appendix D.

### III. PROBLEM FORMULATION

We consider the problem of localizing  $N_D$  desired sources in the presence of  $N_I$  interference sources. All the sources are static and located in a reverberant environment. The signals are received at a noisy microphone array of  $M$  microphones, which are positioned in known, but possibly arbitrary, positions. The acoustic environment between each source and each microphone is modeled by the Acoustic Impulse Response (AIR). The signal at the  $m$ th microphone, considering a far-field setting, is given by

$$z_m(n) = \sum_{j=1}^{N_D} s_j^d(n) * h_{jm}^d(n) + \sum_{j=1}^{N_I} s_j^i(n) * h_{jm}^i(n) + v_m(n), \quad (5)$$

where  $s_j^d(n)$  is the  $j$ th desired source,  $s_j^i(n)$  is the  $j$ th interference source, and  $v_m(n)$  is the  $m$ th microphone noise. We denote by  $h_{jm}^d(n)$  the AIR between the  $m$ th microphone and the  $j$ th desired source. Similarly, we denote by  $h_{jm}^i(n)$  the AIR between the  $m$ th microphone and the  $j$ th interference source.

The sources are characterized by their activation times. The desired sources are active during the entire interval. In contrast, the interference sources are only partially active, namely active

during segments of the interval. We emphasize that albeit we consider intermittent interference sources, longer signals might consist of a larger number of interference sources in addition to more occurrences of active interference sources. We do not assume there is a segment with only the desired source. Longer intervals do not necessarily increase the probability of such a segment since new interference sources could emerge in such a scenario. Consequently, even a long signal still poses a significant challenge to the DoA estimation. Additionally, we assume that the desired sources, the interference sources, and the noise are all uncorrelated. The noise is assumed to be spatially white. We note that these assumptions are made for simplicity. Empirically, the proposed approach leads to superior results for correlated signals and for noise with arbitrary covariance matrix as well.

The received signal is processed using the Short-Time Fourier Transform (STFT). We denote by  $s_j^d(l, k)$  the STFT at the  $l$ th window and the  $k$ th frequency of  $s_j^d(n)$ . The notation for  $s_j^i(l, k)$ ,  $h_{jm}^d(l, k)$ ,  $h_{jm}^i(l, k)$ , and  $v_m(l, k)$  follows similarly. Then, the STFT at the  $l$ th window and the  $k$ th frequency of the received signal is given by

$$z_m(l, k) = \sum_{j=1}^{N_D} s_j^d(l, k) h_{jm}^d(l, k) + \sum_{j=1}^{N_I} s_j^i(l, k) h_{jm}^i(l, k) + v_m(l, k), \quad (6)$$

where we assume the length of the window is much larger than the AIR length.

We stack the received signals,  $\{z_m(l, k)\}_m$ , from all the microphones to obtain a column vector  $\mathbf{z}(l, k) \in \mathbb{C}^{M \times 1}$

$$\mathbf{z}(l, k) = [z_1(l, k) \dots z_M(l, k)]^\top. \quad (7)$$

Its explicit expression is

$$\mathbf{z}(l, k) = \mathbf{H}^d(l, k) \mathbf{s}^d(l, k) + \mathbf{H}^i(l, k) \mathbf{s}^i(l, k) + \mathbf{v}(l, k), \quad (8)$$

where  $\mathbf{s}^d(l, k)$  and  $\mathbf{s}^i(l, k)$  denote the stacked STFT representations of the desired sources and the interference sources, respectively, and are given by

$$\begin{aligned} \mathbf{s}^d(l, k) &= [s_1^d(l, k) \dots s_{N_D}^d(l, k)]^\top \\ \mathbf{s}^i(l, k) &= [s_1^i(l, k) \dots s_{N_I}^i(l, k)]^\top, \end{aligned} \quad (9)$$

and the noise term is

$$\mathbf{v}(l, k) = [v_1(l, k) \dots v_M(l, k)]^\top. \quad (10)$$

The Acoustic Transfer Functions (ATFs) from the  $j$ th desired source and the  $j$ th interference source to the microphone array are

$$\begin{aligned} \mathbf{h}_j^d(l, k) &= [h_{j1}^d(l, k) \dots h_{jM}^d(l, k)]^\top \quad j = 1, \dots, N_D \\ \mathbf{h}_j^i(l, k) &= [h_{j1}^i(l, k) \dots h_{jM}^i(l, k)]^\top \quad j = 1, \dots, N_I, \end{aligned} \quad (11)$$

and in a matrix form

$$\begin{aligned} \mathbf{H}^d(l, k) &= [\mathbf{h}_1^d(l, k) \dots \mathbf{h}_{N_D}^d(l, k)] \\ \mathbf{H}^i(l, k) &= [\mathbf{h}_1^i(l, k) \dots \mathbf{h}_{N_I}^i(l, k)]. \end{aligned} \quad (12)$$

Henceforth, we focus on a single frequency bin and omit the frequency index. Throughout the article, we refer to  $\mathbf{z}(l)$  as the

received signal. Since all the sources are static the ATFs do not change over time, so in the following, we omit their STFT window index  $l$ .

In this work, we focus on a single frequency to demonstrate the Riemannian approach, which is suitable for narrowband signals. For broadband signals, this is the formulation of only a single frequency. We note that the fusion of the different frequencies is of great importance, and normalization techniques for the different beamformers have been proposed [38] and can be applied after our method.

Our goal is to estimate the direction to the desired sources, given  $\mathbf{z}(l)$ ,  $l = 1, \dots, L_{\text{STFT}}$ , where  $L_{\text{STFT}}$  is the number of STFT windows. The main challenge is the existence of interference sources, positioned at unknown locations with possibly high signal power.

#### IV. PROPOSED APPROACH

Typically, the DoA estimation of a desired source is based on the output of a beamformer. In this section, we present the proposed approach applied to the Delay-and-Sum (DS) beamformer. In Section VI, we extend the proposed approach to other DoA estimation methods.

We consider arbitrary indexing of the microphones in the array and designate the first microphone as the reference microphone. Let  $\mathbf{d}(\theta)$  denote the steering vector of the array to direction  $\theta$  relative to the first (reference) microphone, which is given by

$$\mathbf{d}(\theta) = [1, e^{j\phi_2(\theta)}, \dots, e^{j\phi_M(\theta)}]^{\top}, \quad (13)$$

where  $\phi_m(\theta)$  is the phase of the received signal at the  $m$ th microphone with respect to the first microphone. For example, for a uniform linear array and the typical microphone indexing, we have  $\phi_m(\theta) = 2\pi \cdot m \frac{\delta}{\lambda} \sin \theta$ , where  $\lambda$  is the wavelength of the received signal, and  $\delta$  is the distance between the microphones.

The SRP of the DS beamformer is given by

$$P_{\text{DS}}(\theta; \boldsymbol{\Gamma}) = \mathbf{d}^H(\theta) \boldsymbol{\Gamma} \mathbf{d}(\theta), \quad (14)$$

where

$$\boldsymbol{\Gamma} = \mathbb{E}[\mathbf{z}(l)\mathbf{z}^H(l)] \quad (15)$$

is the population covariance matrix. Since the desired source is constantly active and assumed to be at a fixed location during the entire interval, we estimate the population correlation matrix  $\boldsymbol{\Gamma}$  using the sample correlation matrix  $\widehat{\boldsymbol{\Gamma}}$  by averaging  $\mathbf{z}(l)\mathbf{z}^H(l)$  over multiple STFT windows. We divide the STFT of the signal into  $L_s$  disjoint segments, each consisting of  $L_w$  consecutive STFT windows, i.e.,  $L_{\text{STFT}} = L_s \cdot L_w$  (for more details regarding the partitioning see Section V-D). Then, the sample correlation matrix over each segment is computed as follows

$$\widehat{\boldsymbol{\Gamma}}_i = \frac{1}{L_w} \sum_{l=(i-1) \cdot L_w + 1}^{i \cdot L_w} \mathbf{z}(l)\mathbf{z}^H(l), \quad (16)$$

where  $i$  is the segment index. To obtain a full rank correlation matrix, the number of STFT windows is set to be larger than

---

**Algorithm 2** Direction estimation in the presence of multiple interference sources

---

**Input:** the received signal in the STFT domain  $\{\mathbf{z}(l)\}_{l=1}^{L_{\text{STFT}}}$

**Output:** the estimated direction of the desired source  $\hat{\theta}$

---

- 1: Divide  $\{\mathbf{z}(l)\}_{l=1}^{L_{\text{STFT}}}$  into  $L_s$  consecutive segments
  - 2: For each segment  $i$ , compute the sample correlation matrix  $\widehat{\boldsymbol{\Gamma}}_i$  using (16)
  - 3: Compute  $\boldsymbol{\Gamma}_R$  of the set  $\{\widehat{\boldsymbol{\Gamma}}_i\}_{i=1}^{L_s}$  using Algorithm 1
  - 4: Compute  $P_{\text{DS}}(\theta; \boldsymbol{\Gamma}_R)$  according to (18)
  - 5: Return  $\hat{\theta} = \operatorname{argmax}_{\theta} P_{\text{DS}}(\theta; \boldsymbol{\Gamma}_R)$
- 

the dimension of the matrix (the number of microphones), i.e.,  $L_w \geq M$ .

The incorporation of the Riemannian geometry is realized by viewing each matrix  $\widehat{\boldsymbol{\Gamma}}_i$  as a point on the HPD manifold [34] and considering their Riemannian mean, denoted as  $\widehat{\boldsymbol{\Gamma}}_R$  and given by

$$\widehat{\boldsymbol{\Gamma}}_R = \operatorname{argmin}_{\boldsymbol{\Gamma} \in \mathcal{M}} \sum_{i=1}^{L_s} d_R^2(\boldsymbol{\Gamma}, \widehat{\boldsymbol{\Gamma}}_i). \quad (17)$$

In general, there is no closed-form solution to (17) on the HPD manifold for more than two points [35]. Therefore, Algorithm 1 proposed in [36] is used to compute the Riemannian mean of the  $L_s$  correlation matrices.

Once  $\widehat{\boldsymbol{\Gamma}}_R$  is at hand, the SRP of the DS beamformer,  $P_{\text{DS}}(\theta; \widehat{\boldsymbol{\Gamma}}_R)$ , is computed by

$$P_{\text{DS}}(\theta; \widehat{\boldsymbol{\Gamma}}_R) = \mathbf{d}^H(\theta) \widehat{\boldsymbol{\Gamma}}_R \mathbf{d}(\theta). \quad (18)$$

In the case of a single desired source and assuming the direct path is dominant in the AIR, the direction to it is set as the direction achieving the maximum value of the SRP of the DS beamformer, i.e.,

$$\hat{\theta} = \operatorname{argmax}_{\theta} P_{\text{DS}}(\theta; \widehat{\boldsymbol{\Gamma}}_R). \quad (19)$$

We note that (19) is used since the DoA estimation is based on a single STFT frequency bin. In the case of  $N_D$  desired sources, the  $N_D$  directions are set according to the  $N_D$  strongest lobes in the SRP. The algorithm for a single desired source is described in Algorithm 2.

As a baseline, we consider the common practice of the computation of the SRP of the DS beamformer, which is typically based on the sample correlation matrix over the entire interval, i.e.,  $P_{\text{DS}}(\theta; \widehat{\boldsymbol{\Gamma}}_E)$ , where

$$\widehat{\boldsymbol{\Gamma}}_E = \frac{1}{L_{\text{STFT}}} \sum_{l=1}^{L_{\text{STFT}}} \mathbf{z}(l)\mathbf{z}^H(l). \quad (20)$$

We observe that computing the Euclidean mean of the sample correlation matrices per segment,  $\{\widehat{\boldsymbol{\Gamma}}_i\}_{i=1}^{L_s}$ , results in  $\widehat{\boldsymbol{\Gamma}}_E$  in (20). So,  $\widehat{\boldsymbol{\Gamma}}_E$  in (20) is the Euclidean counterpart of  $\widehat{\boldsymbol{\Gamma}}_R$ , the sample correlation matrix resulting from the Riemannian approach.

We will show that our Riemannian approach exploits the assumption that the desired source is constantly active and at a fixed location, whereas the interference sources are intermittent. More specifically, we will show both theoretically in

Section V and empirically in Section VII that the Riemannian mean attenuates the intermittent interferences while preserving the constantly active sources. In contrast, the standard Euclidean mean accumulates all the sources, and as a result, the main lobe could deviate from the direction of a desired source, and even focus on an interference source.

We show in Section V and Section VII that our proposed approach results in a SRP that rejects the interference sources, allowing the beamformer to extract the DoA of the desired sources.

We remark that the proposed approach only requires that the desired sources are the only sources active during the entire interval. The rank of the signal matrix is not known nor needs to be estimated. The number of interference sources is unknown as well. This is by virtue of the Riemannian mean. Unlike other works (e.g. [30]), we do not need to know the activation times of each interference, nor the number of interference sources. Furthermore, we do not assume that there exists a segment, at which a desired source is the only active source, namely, it could always be accompanied by interference sources.

In terms of complexity, the Riemannian approach requires the computation of the Riemannian mean of the correlation matrices, which is more complex than the Euclidean mean. However, the excess complexity depends on the number of microphones, which is typically not high compared to the complexity of computing the correlation matrix which depends on the number of signal samples. Consequently, the excess complexity is negligible. In particular, following the iterative computation of the Riemannian mean in Algorithm 3 in Appendix C, at each iteration, the computation involves the eigenvalue decomposition of two complex matrices of dimension  $M \times M$ , which is  $O(M^3)$ , and the computation of 6 matrix products with the complexity of  $O(M^2)$ . So, at each iteration, the overall excess complexity is  $O(M^3)$ . We recall that for  $N$  samples, the complexity of computing the sample correlation matrix is  $O(M^2N)$  which is much larger than  $O(M^3)$  since typically  $N \gg M$ . Empirically, we found that Algorithm 3 converges very quickly after several iterations. To summarize, the proposed approach leads to improved performance with only negligible additional computational cost. In Appendix C, we present the estimator along with an implementation for the streaming data setting.

To evaluate the performance of the proposed approach, we define the output Signal to Interference Ratio (SIR) as follows:

$$\text{SIR}_j(\widehat{\boldsymbol{\Gamma}}) = \frac{P(\theta^d; \widehat{\boldsymbol{\Gamma}})}{P(\theta_j^i; \widehat{\boldsymbol{\Gamma}})}, \quad (21)$$

where  $P(\theta; \widehat{\boldsymbol{\Gamma}})$  is the SRP computed using the sample correlation matrix  $\widehat{\boldsymbol{\Gamma}}$ ,  $\theta^d$  is the direction of a desired source, and  $\theta_j^i$  is the direction of the  $j$ th interference. When using the DS beamformer, the output SIR becomes

$$\text{SIR}_j(\widehat{\boldsymbol{\Gamma}}) = \frac{\mathbf{d}^H(\theta^d)\widehat{\boldsymbol{\Gamma}}\mathbf{d}(\theta^d)}{\mathbf{d}^H(\theta_j^i)\widehat{\boldsymbol{\Gamma}}\mathbf{d}(\theta_j^i)}. \quad (22)$$

This measure of performance is used because the main challenge in this setting is the presence of interference sources rather than the microphone noise.

## V. ANALYSIS

In this section, we analyze the proposed approach which is based on Riemannian geometry and compare it to its Euclidean counterpart. The proofs of the statements appear in the Supplementary Material (SM). In the analysis, we consider the population correlation matrix of the received signal, neglecting the estimation errors stemming from the finite sample in a segment. We note that sections IV, VI, and VII consider the sample correlation matrices, and only Section V considers the population correlation matrix.

We begin with a short derivation, demonstrating that Riemannian geometry preserves better the desired source subspace in comparison to Euclidean geometry. Consider a single desired source and assume its ATF, denoted by  $\mathbf{h}_0$ , is a common eigenvector of all the correlation matrices per segment associated with the same eigenvalue (this assumption is made formal in Assumption 1 in the sequel). In this case, according to Lemma 2 (see Appendix D),  $\mathbf{h}_0$  is an eigenvector of both means and it is associated with the same eigenvalue, namely  $\lambda_0(\boldsymbol{\Gamma}_R) = \lambda_0(\boldsymbol{\Gamma}_E)$ , where  $\lambda_i(\boldsymbol{\Gamma})$  is the  $i$ th eigenvalue of  $\boldsymbol{\Gamma}$ . In addition, all other eigenvectors span the interference and noise subspace. Under the aforementioned assumption, and by the following known property of the Riemannian mean [39]

$$\boldsymbol{\Gamma}_R \preceq \boldsymbol{\Gamma}_E. \quad (23)$$

We get that

$$\frac{\lambda_0(\boldsymbol{\Gamma}_R)}{\sum_{i=1}^{M-1} \lambda_i(\boldsymbol{\Gamma}_R)} \geq \frac{\lambda_0(\boldsymbol{\Gamma}_E)}{\sum_{i=1}^{M-1} \lambda_i(\boldsymbol{\Gamma}_E)}, \quad (24)$$

due to the equality of the numerators and (23). The inequality in (24) implies that the desired source subspace is more dominant than the subspace of the interference and noise in the Riemannian mean compared to the Euclidean mean.

In the remainder of this section, we extend this analysis and present additional results.

### A. Assumptions

To make the analysis tractable, we consider a single desired source and multiple interference sources. Therefore, we simplify the notations by omitting the superscripts  $(\cdot)^i$  and  $(\cdot)^d$  associated with the interference sources and the desired sources and setting the index of the desired source to 0.

For the purpose of analysis, we make the following assumptions.

*Assumption 1:*  $\mathbf{h}_0^H \mathbf{h}_j = 0, \forall j = 1, \dots, N_I$ .

*Assumption 2:*  $\mathbf{h}_l^H \mathbf{h}_j = 0, \forall l \neq j$ .

It follows from Assumption 1 and Assumption 2 that the ATFs, associated with the desired source and the interference sources are all uncorrelated. These are common assumptions, e.g., see [30]. The assumptions are made only for analysis purposes, whereas in the experimental results, we consider acoustic signals in a reverberant environment without any assumptions. We note that we do not assume there exists a segment at which only one of the sources is active (e.g., as in [30]). In case an interference source is only partially active during a segment, we consider it active during the entire segment.

The population correlation matrix of the  $i$ th segment is given by

$$\boldsymbol{\Gamma}_i = \sigma_0^2 \mathbf{h}_0 \mathbf{h}_0^H + \mathbf{H} \boldsymbol{\Lambda}_i \mathbf{H}^H + \sigma_v^2 \mathbf{I}_{M \times M}, \quad (25)$$

where  $\sigma_0^2$  and  $\sigma_v^2$  are the power of the desired source and the power of the noise, respectively. The diagonal matrix  $\boldsymbol{\Lambda}_i$  captures the signal power of the interference sources and is given by:

$$\boldsymbol{\Lambda}_i = \text{diag} (\sigma_1^2(i) \cdot \mathcal{I}_{i \in \mathcal{L}_1}, \dots, \sigma_{N_i}^2(i) \cdot \mathcal{I}_{i \in \mathcal{L}_{N_i}}), \quad (26)$$

where  $\mathbf{H} = \mathbf{H}^i(l, k)$  due to the omission of the indices and  $\sigma_j^2(i) = \mathbb{E}[|\mathbf{s}_j(n)|^2 | n \in i\text{th segment}]$  is the expected signal power of the  $j$ th interference source at the  $i$ th segment,  $\mathcal{L}_j$  is the set of segments at which the  $j$ th interference source is active, and  $\mathcal{I}_{i \in \mathcal{L}_j}$  is an indicator function, attaining the value of 1 when the  $j$ th interference is active during the  $i$ th segment and 0 otherwise. We denote by  $\tau_j = \frac{|\mathcal{L}_j|}{L_s}$  the relative number of segments during which the  $j$ th interference source is active. We assume the same expected power at all the segments in the interval, i.e.,  $\sigma_j^2(i) = \sigma_j^2$  for all  $j = 1, \dots, N_i$  and  $i = 1, \dots, L_s$ .

We continue with defining the Signal to Noise Ratio (SNR) at the  $m$ th microphone as

$$\text{SNR}_m = \frac{\sigma_0^2 |\mathbf{h}_0[m]|^2}{\sigma_v^2}, \quad (27)$$

where  $|\mathbf{h}_0[m]|^2$  is the attenuation of the signal due to the acoustic channel between the desired source and the  $m$ th microphone. Since we focus on a single frequency bin, (27) is the narrow-band SNR.

To capture the correlation between the steering vectors and the ATFs, we define

$$\rho_{rs} = \frac{|\langle \mathbf{d}_r, \mathbf{h}_s \rangle|^2}{\|\mathbf{d}_r\|^2 \cdot \|\mathbf{h}_s\|^2} = \frac{|\langle \mathbf{d}_r, \mathbf{h}_s \rangle|^2}{M \|\mathbf{h}_s\|^2}, \quad (28)$$

where  $r, s = 0, 1, 2, \dots, N_i$ , indicating the desired source or an interference source.

We conclude the preliminaries of the analysis with two additional assumptions.

*Assumption 3:*  $\rho_{rr}$  is fixed  $\forall r$ , and  $\rho_{rs}$  is fixed  $\forall r \neq s$ .

Assumption 3 implies that the correlation between the ATFs and the steering vectors depends only on whether they are associated with the same source or not. Following Assumption 3, henceforth we denote  $\kappa = \rho_{rr}$  and  $\rho = \rho_{rs}$  for  $r \neq s$ .

*Assumption 4:*  $\kappa > \rho$ .

Assumption 4 is typically made in the context of source localization. It implies that the correlation between a steering vector to a source and the ATF associated with that source is higher than the correlation between a steering vector to a source and the ATF associated with a different source.

## B. Main Results

Our first result states that the output SIR (22) of the Riemannian-based DS beamformer is higher than the output SIR of the Euclidean-based DS beamformer.

*Proposition 1:* For every interference source  $j$ , the following holds

$$\text{SIR}_j(\boldsymbol{\Gamma}_R) > \text{SIR}_j(\boldsymbol{\Gamma}_E), \quad (29)$$

for any number of microphones in the array.

Examining the dependency of the output SIR on the noise power  $\sigma_v^2$  leads to the following result.

*Proposition 2:* If

$$\sigma_0^2 \|\mathbf{h}_0\|^2 \geq \sigma_j^2 \tau_j \|\mathbf{h}_j\|^2, \quad \forall j, \quad (30)$$

then

$$\frac{\partial}{\partial \sigma_v^2} \text{SIR}_j(\boldsymbol{\Gamma}_R) < \frac{\partial}{\partial \sigma_v^2} \text{SIR}_j(\boldsymbol{\Gamma}_E) < 0. \quad (31)$$

Namely, the lower the noise power is the higher the output SIR is, and the improvement in  $\text{SIR}_j(\boldsymbol{\Gamma}_R)$  is greater than the improvement in  $\text{SIR}_j(\boldsymbol{\Gamma}_E)$ . Since we established that the Riemannian approach is better than the Euclidean one in terms of the SIR in Proposition 1, Proposition 2 implies that increasing the SNR further increases the gap between the two approaches. Nevertheless, it also indicates that the performance of the Riemannian approach in terms of the SIR is more sensitive to noise compared to the Euclidean counterpart. Note that this statement holds under condition (30), which implies that the received power of the desired source is stronger than the received power of each interference source, considering the attenuation stemming from the activity duration. See more details in Section I-C in the SM.

The proofs of Proposition 1 and Proposition 2 rely on the following lemma, which is important in its own right.

*Lemma 1:* The Riemannian or the Euclidean mean of the population correlation matrices of the segments (25) over the entire interval can be written in the same parametric form as

$$\boldsymbol{\Gamma} = \sigma_0^2 \mathbf{h}_0 \mathbf{h}_0^H + \sum_{j=1}^{N_i} \mu_j^2 \mathbf{h}_j \mathbf{h}_j^H + \sigma_v^2 \mathbf{I}. \quad (32)$$

The Riemannian mean  $\boldsymbol{\Gamma}_R$  is obtained by setting the parameters  $\mu_j$  to

$$\mu_j^2 = \frac{(\sigma_j^2 \|\mathbf{h}_j\|^2 + \sigma_v^2)^{\tau_j} (\sigma_v^2)^{1-\tau_j} - \sigma_v^2}{\|\mathbf{h}_j\|^2}, \quad (33)$$

and the Euclidean mean  $\boldsymbol{\Gamma}_E$  is obtained by setting

$$\mu_j^2 = \sigma_j^2 \tau_j. \quad (34)$$

We note that only assumptions 1 and 2 are necessary for this lemma to hold. In addition, we note that if the interference sources are always active, i.e.,  $|\mathcal{L}_j| = L_s, \forall j$ , it holds that  $\boldsymbol{\Gamma}_R = \boldsymbol{\Gamma}_E$ .

Lemma 1 shows that both  $\boldsymbol{\Gamma}_R$  and  $\boldsymbol{\Gamma}_E$ , i.e., the population correlation matrices of the segments in (25), can be decomposed into three terms associated with the desired source, the interference sources, and the noise. By (32), the desired source term and the noise term (the first and third terms) are the same in both the Riemannian and the Euclidean means. In contrast, the coefficients in (33) and (34) imply that the amplitude of the interference sources term (the second term) depends on the used

geometry. By further inspecting the expressions of  $\{\mu_j^2\}$ , we see that the interference attenuation using the Riemannian geometry in (33) is more involved than its Euclidean counterpart in (34), depending not only on the interference power and the duration of activity but also on the noise power and the corresponding ATF.

Furthermore, considering  $\mu_j$  in (34), the condition (30) could be viewed as the dominance of the desired source after the attenuation of the Euclidean mean. Discussion about the condition (30) for  $\mu_j$  in the Riemannian case in (33) appears in Section I-C in the SM.

Next, we examine a family of correlation matrices that pertain to the same parametric form as in (32) in Lemma 1, i.e.,

$$\boldsymbol{\Gamma}_a = \mathbf{h}_0 \mathbf{h}_0^H + \sum_{j=1}^{N_1} a_j \mathbf{h}_j \mathbf{h}_j^H + \sigma_v^2 \mathbf{I}, \quad (35)$$

for some coefficients  $\mathbf{a} = [a_1, a_2, \dots, a_{N_1}]$ . Without loss of generality, we set the coefficient of  $\mathbf{h}_0 \mathbf{h}_0^H$  to 1. We note that  $\boldsymbol{\Gamma}_a$  is in accordance with (25).

For any  $j$ , we have that

$$\boldsymbol{\Gamma}_{\text{opt}} \equiv \underset{\boldsymbol{\Gamma}_a}{\operatorname{argmax}} \text{SIR}_j(\boldsymbol{\Gamma}_a) = \mathbf{h}_0 \mathbf{h}_0^H + \sigma_v^2 \mathbf{I}, \quad (36)$$

where  $\mathbf{a} = \mathbf{0}$ . Consequently,

$$\text{SIR}_j(\boldsymbol{\Gamma}_{\text{opt}}) = \frac{\mathbf{d}_0^H (\mathbf{h}_0 \mathbf{h}_0^H + \sigma_v^2 \mathbf{I}) \mathbf{d}_0}{\mathbf{d}_j^H (\mathbf{h}_0 \mathbf{h}_0^H + \sigma_v^2 \mathbf{I}) \mathbf{d}_j}. \quad (37)$$

Considering vanishing noise, i.e., when the noise power approaches zero, the following result stems from Lemma 1 by considering the limit  $\lim_{\sigma_v^2 \rightarrow 0} \mu_j^2 = 0$  using (33).

*Corollary 1:*

$$\lim_{\sigma_v^2 \rightarrow 0} \boldsymbol{\Gamma}_R = \boldsymbol{\Gamma}_{\text{opt}}. \quad (38)$$

According to Corollary 1, the Riemannian mean approaches the optimal correlation matrix as the noise becomes negligible. By adding a condition on the presence of the interference sources, from Lemma 1 and (33) we also have the following.

*Corollary 2:* For any interference source  $j$ , if  $\tau_j < \frac{1}{2}$ , then

$$\lim_{\sigma_j^2 \rightarrow \infty, \sigma_v^2 \rightarrow 0} \boldsymbol{\Gamma}_R = \boldsymbol{\Gamma}_{\text{opt}}. \quad (39)$$

Additionally, if  $\tau_j < \frac{1}{2}$  for all  $j$ , then

$$\lim_{\sigma_j^2 \rightarrow \infty, \forall j=1, \dots, N_1, \sigma_v^2 \rightarrow 0} \boldsymbol{\Gamma}_R = \boldsymbol{\Gamma}_{\text{opt}}. \quad (40)$$

Corollary 2 implies that for vanishing noise, even when all the interference sources have infinite power, the desired source is still the dominant source in the SRP of the DS beamformer using the Riemannian mean. Following (37) it holds that  $\lim_{\sigma_j^2 \rightarrow \infty, \sigma_v^2 \rightarrow 0} \text{SIR}_j(\boldsymbol{\Gamma}_{\text{opt}}) = \frac{\kappa}{\rho} > 1$  for all  $j$ . For the Euclidean mean it holds that  $\lim_{\sigma_j^2 \rightarrow \infty, \sigma_v^2 \rightarrow 0} \text{SIR}_j(\boldsymbol{\Gamma}_E) = \frac{\rho}{\kappa} < 1 < \frac{\kappa}{\rho} = \lim_{\sigma_j^2 \rightarrow \infty, \sigma_v^2 \rightarrow 0} \text{SIR}_j(\boldsymbol{\Gamma}_{\text{opt}})$  for all  $j$ . Note that we consider noise power approaching zero rather than strictly zero, because when  $\sigma_v^2 = 0$ , the correlation matrix is singular, and therefore, lies outside the HPD manifold. Additionally, in practice, noise is always present.

To illustrate the obtained expressions for the Riemannian and the Euclidean SIR, we present the following simple example.

*Example 1:* Consider an anechoic environment without attenuation, for which  $\kappa = 1$  and  $\rho = 0$ , and two interference sources. Each interference source is active at a different segment, i.e.,  $L_s = 2$ ,  $\mathcal{L}_1 = \{1\}$ , and  $\mathcal{L}_2 = \{2\}$ . All the sources have the same power. In this setting, for the Riemannian geometry, we have

$$\text{SIR}_j(\boldsymbol{\Gamma}_R) = \sqrt{\frac{M}{\sigma_v^2} + 1}, \quad (41)$$

and for the Euclidean geometry, we have:

$$\text{SIR}(\boldsymbol{\Gamma}_E) = \frac{2(M + \sigma_v^2)}{M + 2\sigma_v^2}. \quad (42)$$

Therefore, in the limit of  $\sigma_v^2 \rightarrow 0$ , or  $M \rightarrow \infty$ , we have  $\text{SIR}(\boldsymbol{\Gamma}_R) \rightarrow \infty$ , whereas  $\text{SIR}(\boldsymbol{\Gamma}_E) \approx 2$ .

We conclude this analysis with a few remarks. First, we note that  $\boldsymbol{\Gamma}_E$  leads to the ML estimator by taking  $\hat{\theta}_0 = \operatorname{argmax}_{\theta} P_{\text{DS}}(\theta; \boldsymbol{\Gamma}_E)$  for the interference-free setting. In this case, the Riemannian mean coincides with the Euclidean mean, i.e., the proposed method coincides with the ML estimator. The main advantage of the proposed method lies in attenuating the interference sources while preserving the desired source. Second, under assumptions 1 and 2, the number of sources, both desired and interference, are limited by the number of microphones in the array, i.e.,  $N_1 + N_D < M$ . In Section V-C we alleviate Assumption 2, which removes this restriction. Third, following the same techniques in the proof of Proposition 1 and Proposition 2, similar results are derived for an alternative definition of the SIR:  $\text{SIR}_{\text{tot}}(\boldsymbol{\Gamma}) \equiv \frac{\mathbf{d}_0^H \boldsymbol{\Gamma} \mathbf{d}_0}{\sum_{j=1}^{N_1} \mathbf{d}_j^H \boldsymbol{\Gamma} \mathbf{d}_j}$ , which captures the ratio between the desired source and the sum of all interference sources. See Section II in the SM for more details.

### C. Relation to Signal Enhancement

For signal enhancement in reverberant environments, the estimation of the ATF of the desired source is typically required. In our setting, there is no segment at which the desired source is the only active source, and therefore, the ATF estimation is done in the presence of the interference sources. In such a case, the following quantity could be of interest

$$\overline{\text{SIR}}_j(\boldsymbol{\Gamma}) = \frac{\mathbf{h}_0^H \boldsymbol{\Gamma} \mathbf{h}_0}{\mathbf{h}_j^H \boldsymbol{\Gamma} \mathbf{h}_j}, \quad (43)$$

which is different than (22) in the use of the ATFs instead of the steering vectors.

Similarly to Proposition 1, the following Proposition 3 examines the performance in terms of the SIR defined in (43). Here, assumptions 2-4 are not required, and therefore, the ATFs of the interference sources could be correlated, and the number of sources is not limited by the number of microphones in the array.

*Proposition 3:* Under Assumption 1, for all  $j$  we have

$$\overline{\text{SIR}}_j(\boldsymbol{\Gamma}_R) \geq \overline{\text{SIR}}_j(\boldsymbol{\Gamma}_E). \quad (44)$$

Another interesting component in signal enhancement is the Relative Transfer Function (RTF) between different microphones [40], [41], [42]. We compute the RTFs with respect to the first microphone, i.e.,  $\frac{h_j}{h_j(1)}$ . Since the RTFs are proportional to the ATFs, assumption 1 holds for the RTFs, and therefore, all the derived results apply to the RTFs as well.

#### D. The Segments and the Interference Sources Activity

In this section, we investigate the effect of misalignment between the segments and the activity of the interference sources. We consider two interference sources and two segments. We denote by  $\alpha$  the offset between the segments and the activity of the interference sources. For simplicity, we consider alternately active interference sources. Suppose the first interference source is active during  $\alpha \in [0, 1]$  of the first segment and during  $1 - \alpha$  of the second segment, and suppose the second interference source is active during  $1 - \alpha$  and  $\alpha$  of the first and second segments, respectively. In this case, the population correlation matrices of the two segments are given by

$$\begin{aligned}\Gamma_1(\alpha) &= \sigma_0^2 \mathbf{h}_0 \mathbf{h}_0^H + \alpha^2 \sigma_1^2 \mathbf{h}_1 \mathbf{h}_1^H + (1 - \alpha)^2 \sigma_2^2 \mathbf{h}_2 \mathbf{h}_2^H + \sigma_v^2 \mathbf{I} \\ \Gamma_2(\alpha) &= \sigma_0^2 \mathbf{h}_0 \mathbf{h}_0^H + (1 - \alpha)^2 \sigma_1^2 \mathbf{h}_1 \mathbf{h}_1^H + \alpha^2 \sigma_2^2 \mathbf{h}_2 \mathbf{h}_2^H + \sigma_v^2 \mathbf{I}.\end{aligned}\quad (45)$$

The correlation matrices in (45) depend on  $\alpha$ , and as a result, their Riemannian mean  $\Gamma_R(\alpha)$  and their Euclidean mean  $\Gamma_E(\alpha)$  depend on  $\alpha$  as well.

Examining the dependency of the SIR on  $\alpha$  leads to the following result.

*Proposition 4:* For any  $\alpha \in [0, 1]$ , we have

$$\text{SIR}(\Gamma_R(\alpha)) \geq \text{SIR}(\Gamma_E(\alpha)). \quad (46)$$

Proposition 4 states that for every misalignment between the segments and the activity of the interference sources, the Riemannian mean leads to higher SIR in comparison to its Euclidean counterpart. Equality in (46) is obtained for  $\alpha = \frac{1}{2}$ , which means 50% offset. In this case, it holds that  $\Gamma_1 = \Gamma_2$ , and both means are the same.

Empirically, we found that the advantage of the Riemannian mean over the Euclidean mean decreases as the offset between the segments and the activity of the interference sources increases. We leave the question of optimal partitioning of the STFT windows into segments to future work.

## VI. EXTENSION TO OTHER DOA ESTIMATION METHODS

In this section, to broaden its applicability, we demonstrate the incorporation of the Riemannian approach in other beamformers. Each beamformer generates a spatial spectrum from which the directions to the desired sources are estimated according to the highest peaks in the spectrum.

As a subspace (SbSp) approach, we implement MUSIC [13] in the following way. Given  $N_D$  desired sources, we take the leading  $N_D$  eigenvectors of the sample correlation matrix,  $\hat{\Gamma}$ , and construct the signal subspace matrix,  $\mathbf{U}(\hat{\Gamma}) \in \mathbb{C}^{M \times N_D}$ , whose columns are the  $N_D$  eigenvectors. Then, the SbSp spectrum is defined by

$$\mathbf{P}_{\text{SbSp}}(\theta; \hat{\Gamma}) = \mathbf{d}^H(\theta) \mathbf{U}(\hat{\Gamma}) \mathbf{U}^H(\hat{\Gamma}) \mathbf{d}(\theta). \quad (47)$$

We note that the appropriate number of eigenvectors  $N_D$  (the dimension of the subspace) needs to be estimated.

Similarly to the DS beamformer based on the SRP in (18) and (19), the Riemannian and the Euclidean SbSp methods are given by  $\mathbf{P}_{\text{SbSp}}(\theta; \hat{\Gamma}_R)$  and  $\mathbf{P}_{\text{SbSp}}(\theta; \hat{\Gamma}_E)$ , respectively, according to (47). The SbSp method could also benefit from our Riemannian approach. Recalling assumptions 1 and 2 and the structure of the mean correlation matrices  $\Gamma_R$  and  $\Gamma_E$  in (32), we see that  $\mathbf{h}_0$  is an eigenvector of both  $\Gamma_R$  and  $\Gamma_E$ , spanning the signal subspace (we assume a single desired source for simplicity). Moreover, the vectors  $\{\mathbf{h}_j\}$  are also eigenvectors of  $\Gamma_R$  and  $\Gamma_E$ , spanning the subspace of the interference sources. SbSp methods typically focus on the principal eigenvector. Following (32), the principal eigenvector is determined according to the largest coefficient among  $\sigma_0$  and  $\{\mu_j\}_{j=1}^{N_1}$ . The parameter  $\mu_j$  in (33) for the Riemannian mean  $\Gamma_R$  is smaller than  $\mu_j$  in (34) for the Euclidean mean  $\Gamma_E$ , as proven in Proposition 1. As a result, the signal subspace is more dominant relative to the interference subspace when considering the Riemannian mean in comparison to the Euclidean mean, implying better results for the Riemannian SbSp approach. For an interference source with sufficiently high power, the leading eigenvector of  $\Gamma_E$  could span the interference subspace rather than the signal subspace, whereas the leading eigenvector of  $\Gamma_R$  spans the signal subspace. In Section VII, we demonstrate these SbSp methods empirically and show the advantage of the Riemannian approach.

Our approach is also applicable to the MVDR beamformer [10], whose spectrum is given by

$$\mathbf{P}_{\text{MVDR}}(\theta; \hat{\Gamma}) = \frac{1}{\mathbf{d}^H(\theta) \hat{\Gamma}^{-1} \mathbf{d}(\theta)}. \quad (48)$$

The typical spectrum of the MVDR beamformer is obtained by using  $\hat{\Gamma}_E$  in (48), namely  $\mathbf{P}_{\text{MVDR}}(\theta; \hat{\Gamma}_E)$ . We propose to use the Riemannian mean by setting  $\hat{\Gamma}$  to be  $\hat{\Gamma}_R$  in (48) to obtain  $\mathbf{P}_{\text{MVDR}}(\theta; \hat{\Gamma}_R)$ .

In principle, many DoA estimation methods that employ the sample correlation matrix could potentially benefit from the proposed approach, even if it is not based on a beamformer. For example, the Bayesian learning method for signal recovery for DoA estimation proposed in [23] employs the sample correlation matrix. Its Riemannian alternative is implemented by following their algorithm only with the Riemannian mean instead of the Euclidean mean. The empirical results appear in Section VII.

## VII. SIMULATION RESULTS

In this section, we demonstrate the performance of the proposed approach based on Riemannian geometry, and compare it to Euclidean geometry, implicitly considered by the common practice<sup>1</sup>. Additionally, we compare our approach to a heuristic method, based on the intersection of subspaces. The intersection leads to the rejection of non-common components, such as the

<sup>1</sup>The code is available at <https://github.com/amitaybar/Interference-Rejection-using-Riemannian-Geometry-for-DoA-Estimation>

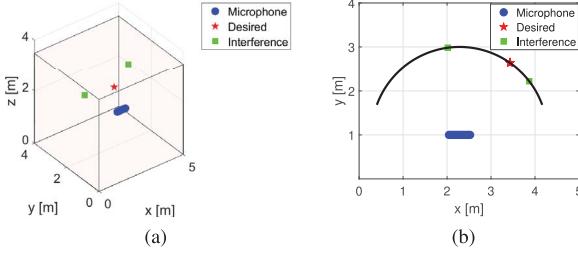


Fig. 1. The reverberant room with the microphone array (blue circles), the desired source (red star), and the interference sources (green squares). (a) A 3D view. (b) A 2D view.

interference sources subspace, and preserves common components, such as the desired source subspace. We refer to it as the intersection beamformer (see Appendix A for more details).

We consider a reverberant enclosure of dimensions  $5\text{m} \times 4\text{m} \times 3.5\text{m}$  consisting of a microphone array with  $M = 12$  microphones. The AIRs between the different sources and the array are generated based on the image method [43], as implemented by the simulator in [44]. The sampling frequency is 16KHz, and the length of the AIRs is set to 2048 samples. The received signal is transformed to the time-frequency domain using STFT with a window size of 1024 samples with 50% overlap. We test all methods using a single frequency bin, of index 250, chosen according to the microphone spacing, which is 4.36cm. Here, the correlation matrix estimation is based on 16 STFT windows, which results in a segment duration of 1.024s. The emitted signals are generated as white Gaussian noise. The desired source is constantly active, whereas the interference sources are active only intermittently.

All the sources are positioned on a  $140^\circ$  arc of radius 2m from the center of the array on the XY plane. The heights of the interference sources vary randomly, uniformly distributed between 0.5m and 3m. The height of the desired source is set to 1.8m. Fig. 1 presents the room layout, where the (two) interferences are marked by green squares, the desired source by a red star, and the microphone array by blue circles. The leftmost microphone is positioned at (2.0436m, 1m, 2m), and the rest are positioned 4.36cm apart along the x-axis. While we focus on this specific configuration, we note that we tested other configurations that yielded similar results.

We examine the performance of both the DS and the SbSp methods. Algorithm 2 is used for the proposed Riemannian DS, and the common practice is implemented by replacing step 3 in Algorithm 2 with (20). The SbSp methods require knowing the dimension of the signal space of the mean correlation matrix. For the intersection method only, the dimension of the signal space of each segment is also required. Since the desired source may not be the strongest source received at the array, we need to consider all the active sources, and not merely the strongest one when estimating the dimension of the signal subspace. To estimate the dimension, we implement a heuristic algorithm, based on the spectral gap. We consider the dimension of the signal space to be the number of eigenvalues higher than a threshold. For all methods, the threshold for the mean correlation matrix is the mean value plus the standard deviation of the eigenvalues (normalized to a unit sum). For the intersection

method, the threshold for the sample correlation matrix of each segment is 1.5 times the mean of the eigenvalues of the sample correlation matrix. Apart from this practical implementation, we also present results for an oracle implementation, assuming the dimension of the signal space is perfectly known.

For quantitative evaluation, we use the root mean square error (RMSE) and the accuracy, for which DoA estimation error that is smaller than  $3^\circ$  (arbitrarily chosen) is considered accurate. In addition, we use two other metrics. The first is the mean of the empirical output SIR with respect to all the interference sources, which is given by

$$\text{SIR} = \frac{1}{N_I} \sum_{j=1}^{N_I} \frac{P(\theta_j^d)}{P(\theta_j^i)}, \quad (49)$$

where  $P$  is the spectrum computed using the evaluated method. The second metric is the directivity [45, ch.2], which is given by

$$D(\hat{\Gamma}) = \frac{P(\theta^d)}{\frac{1}{2} \int_0^\pi P(\theta) \sin(\theta) d\theta}. \quad (50)$$

The proposed approach results in inherent interference rejection, which is its greatest merit. The attenuation of the interference sources by our approach allows for accurate DoA estimation even in the presence of strong interference sources. The measure of SIR is indicative of the amount of interference rejection.

In the first experiment, we consider two interference sources, each active at a single, but disjoint, segment, resulting in a signal of 2.048s. The reverberation time is set to 150ms, and the SNR is 20dB.

We start with an example of the SRP of the DS beamformer (see (18)), computed using  $\hat{\Gamma}_R$  and  $\hat{\Gamma}_E$ , which is presented at the top of Fig. 2(a). The SRP of the Riemannian DS beamformer is shown in solid blue and the SRP of the Euclidean DS beamformer is shown in dashed red. Both SRPs are in a dB (log) scale. The directions to the desired source and the interference sources are represented by a solid black line and a dashed black line, respectively. We see that by using  $\hat{\Gamma}_R$  the main lobe is directed towards the desired source. In contrast, the SRP using  $\hat{\Gamma}_E$  is peaked at two different directions, none of which is the direction to the desired source. The bottom of Fig. 2(b) is the same as Fig. 2(a), only with the SRP computed using the sample correlation matrix of each of the two segments, presented in different shades of orange. Even though the main lobes of the two SRPs are not pointing toward the desired source, the Riemannian mean leads to a SRP with the main lobe directed at the desired source, whereas the lobes to other directions are highly attenuated. We emphasize that viewing the SRP of the Euclidean DS beamformer in addition to the SRP of each segment separately does not allow correct identification of the direction to the desired source. We note that since the population correlation matrix is approximately low rank and the number of microphones is typically small, the estimation of the correlation matrix based on a small sample is feasible.

Next, we randomly generate 200 different pairs of positions for the interference sources. For each pair, the desired source

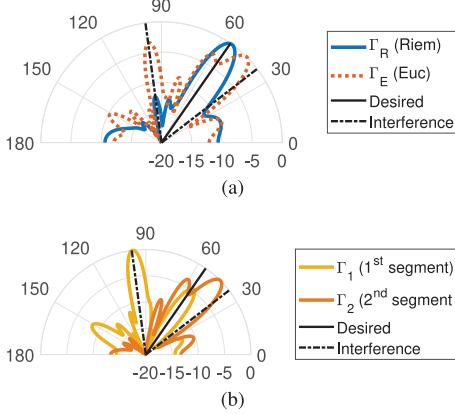


Fig. 2. The SRP of the DS beamformer using (a)  $\widehat{\Gamma}_R$  in solid blue and  $\widehat{\Gamma}_E$  in dashed red, and (b)  $\widehat{\Gamma}_1$  and  $\widehat{\Gamma}_2$  in different shades of orange. The black solid line indicates the direction of the desired source, and the dashed black lines indicate the directions to the interference sources. The Input SIR is  $-6\text{dB}$ .

is located at 20 different equally spaced directions along the arc (with the height of 1.8m). Thus, in total, 4000 different scenarios are examined.

Fig. 3 presents the mean output SIR (a) and the directivity (b) for the DS method using the correlation matrix estimates:  $\widehat{\Gamma}_R$  (based on Riemannian geometry) in blue and  $\widehat{\Gamma}_E$  (based on Euclidean geometry) in red. The box indicates the 25th and 75th percentiles, and the line marks the median. We test the different methods, Riemannian or Euclidean, in varying input SIR values, i.e. the SIR with respect to the source's power (excluding the AIRs and the beamformer processing).

We see that the Riemannian DS method attains high output SIR values, even for strong interference sources (high input SIR). In contrast, the Euclidean DS method results in relatively low output SIR values. The gap in the output SIR values between the Riemannian DS, and the Euclidean DS is up to  $10\text{dB}$ . These results coincide with Proposition 1, stating that  $\text{SIR}_j(\widehat{\Gamma}_R) > \text{SIR}_j(\widehat{\Gamma}_E)$ , for every interference source  $j$ . We emphasize that the Euclidean mean,  $\widehat{\Gamma}_E$ , is equivalent to the common practice of using the entire signal for a single correlation matrix estimation. Since both the mean output SIR and the directivity present similar trends, and due to space considerations, in the following, we only present the mean output SIR.

Fig. 4 is the same as Fig. 3, but presenting the SbSp method with the addition of the intersection method, which appears in orange. Fig. 4(a) presents the results for the practical implementation that includes estimating the dimension, whereas Fig. 4(b) presents the results for the oracle. We see that the Riemannian approach outperforms its Euclidean counterpart by approximately  $20\text{dB}$ . In addition, the oracle SbSp method is better than the practical SbSp. In comparison to Fig. 3(a), it can be seen that the Riemannian SbSp method results in higher output SIRs than the Riemannian DS method. In contrast, for the Euclidean approach, the SbSp method yields slightly lower SIRs than the DS method. The reason is that the Riemannian mean better attenuates the interference sources, allowing for a better estimation of the signal subspace than the Euclidean mean.

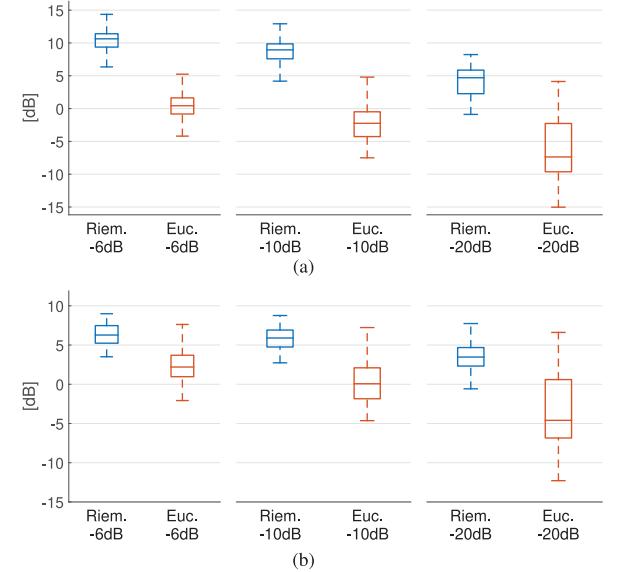


Fig. 3. (a) The mean output SIR and (b) the directivity for two interference sources, for the Riemannian and the Euclidean DS method. The x-axis indicates the input SIR, and the y-axis indicates the output SIR. The box indicates the 25th and 75th percentiles, and the central line marks the median. Several input SIR values are presented.

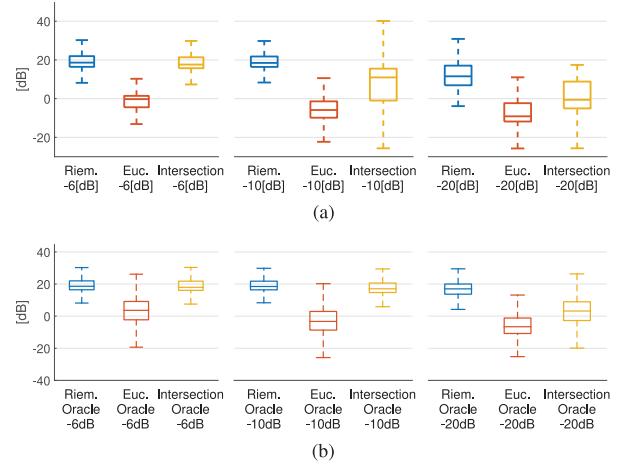


Fig. 4. The mean output SIR of the SbSp methods. (a) Practical implementation. (b) Using an oracle. The box indicates the 25th and 75th percentiles and the central line marks the median. The x-axis indicates the input SIR, and the y-axis indicates the output SIR. Several input SIR values are presented.

We continue with examining the direction estimation to the desired source. The estimated direction is defined as the direction leading to the maximal value of the SRP, namely

$$\hat{\theta}^d = \operatorname{argmax}_{\theta} P(\theta). \quad (51)$$

Fig. 5 shows the estimated direction to the desired source for the Riemannian DS method (blue square), the Euclidean DS method (red circle), and the intersection method (orange star). The black solid line marks the true location of the desired source (at 20 different positions). The dashed lines mark the fixed location of the interference sources. The results for input SIR  $-6\text{dB}$  and  $-10\text{dB}$  appear in Fig. 5(a) and Fig. 5(b), respectively.

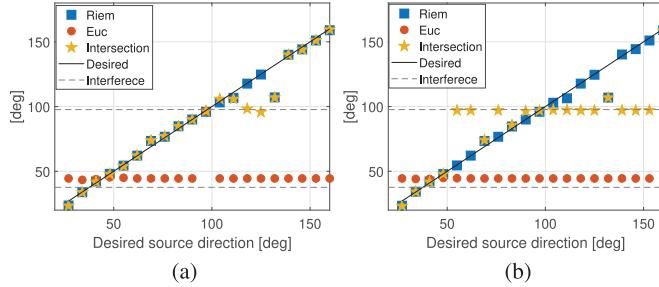


Fig. 5. Estimation of the DoA to the desired source for (a) input SIR of  $-6\text{dB}$  and (b) input SIR of  $-10\text{dB}$ .

We see that using the Riemannian mean, the direction estimation follows the desired source. In contrast, using the Euclidean mean (the entire signal) results in estimating the direction of one of the interference sources. The intersection method is also inferior to the proposed approach, resulting in direction estimation to the desired source or an interference source depending on the input SIR. We note that since the DoA is estimated using (51), only the main lobe of the SRP is considered for the DoA estimation (for simplicity, we assume a single desired source). Typically-used beamformers, based on Euclidean geometry, result in a main lobe that is pointing toward the strongest source which is an interference source. Since the number of sources is unknown, and their relative positioning is unknown as well, considering even lower lobes does not allow for estimating the DoA of the desired source using the SRP. Even if an interference source does not mask the desired source (which is often the case), we cannot differentiate between an interference source and the desired source, and cannot assign the correct direction to the desired source. In contrast, the Riemannian approach rejects the interference sources, resulting in a main lobe that points toward only the desired source only. We report that similar results as in Fig. 5 are obtained also for an SNR value of  $0\text{dB}$  and for interference sources with similar DoA. Results for speech signals from the TIMIT dataset appear in Appendix B where similar trends are demonstrated.

We repeat the experiment for 200 iterations, where at each iteration the desired source is positioned at 20 different positions as before. Fig. 6 presents the RMSE and the accuracy for a different number of microphones for the Riemannian approach (blue circles) and its Euclidean counterpart (red squares). We recall that DoA estimation error that is smaller than  $3^\circ$  (arbitrarily chosen) is considered accurate. We see that the proposed approach leads to significant improvements for all the tested number of microphones.

Next, we examine the sensitivity of the proposed approach to the SNR and the reverberation time. We repeat the setting of the two interferences, as described in the first experiment. The results are presented in Fig. 7. In Fig. 7(a), the mean output SIR for the DS method is presented as a function of the reverberation time for a fixed SNR of  $20\text{dB}$ . Several input SIR values are shown:  $0\text{dB}$  (asterisk),  $-6\text{dB}$  (circle), and  $-10\text{dB}$  (triangle). The results for the Riemannian and the Euclidean DS appear in blue and red, respectively. Fig. 7(b) is the same as Fig. 7(a), only for different SNR values for a fixed reverberation time of

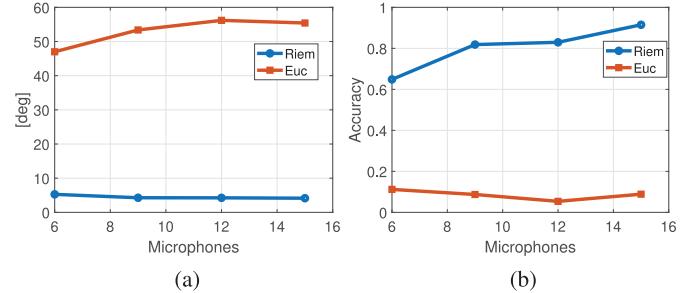


Fig. 6. (a) RMSE and (b) accuracy versus the number of microphones for two interference sources with input SIR of  $-10\text{dB}$ .

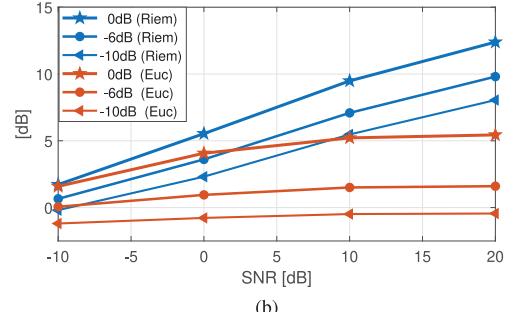
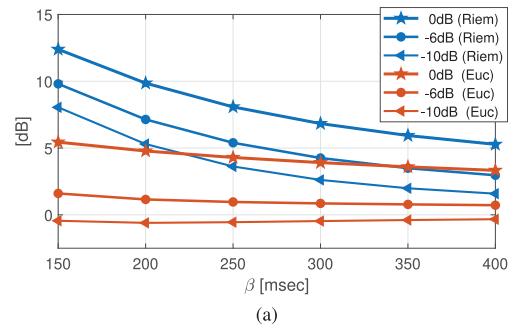


Fig. 7. The mean output SIR as a function of (a) the reverberation time and (b) the SNR, for two interference sources. The Riemannian DS appears in blue, whereas the Euclidean DS appears in red. Several input SIR values are presented:  $0\text{dB}$  (asterisk),  $-6\text{dB}$  (circle), and  $-10\text{dB}$  (triangle).

$\beta = 150\text{ms}$ . We see that the smaller the reverberation time is, the higher the output SIR is for the Riemannian DS method. In contrast, the Euclidean DS is less affected by the reverberation time, resulting in relatively low output SIRs. For all values of reverberation times, the Riemannian DS results in higher output SIR than its Euclidean counterpart. From Fig. 7(b), we see that the SNR has a large impact on the performance of the Riemannian DS; the higher the SNR is, the higher the output SIR becomes. Conversely, the Euclidean DS is moderately affected by the SNR, resulting in much lower output SIR values. A possible explanation is that the main phenomenon limiting the performance of the Euclidean approach is the existence of interference sources. In addition, as Proposition 2 predicts, the sensitivity of the Riemannian DS to the SNR is higher than the Euclidean DS, and the higher the SNR is, the larger the gap in the performance between the Riemannian and the Euclidean approaches.

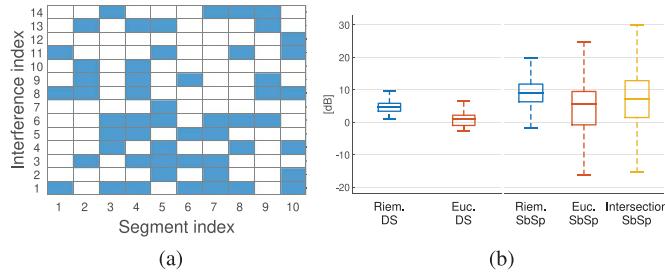


Fig. 8. (a) Activation map for the 14 interference sources during the observed 10 time segments (blue indicates ‘active’). (b) Output SIR of the different beamformers. The input SIR of each interference is  $-6\text{dB}$ .

We examine the performance of the MVDR beamformer, given by (48). The MVDR beamformer is popular when interference sources are present, thanks to its distortionless response in the direction of the desired source, and its typical narrow beams. However, in our setting, since the desired source is accompanied by interference sources, and the directions to all the sources are unknown, the DoA estimation of the desired source using the MVDR beamformer is outperformed by the DS and the SbSp methods. We also examine the case of two desired sources. The results show similar trends and appear in Appendix B, due to space considerations.

In the second experiment, we examine a multiple interference setting, by considering  $N_I = 14$  interference sources. We note that the number of interference sources is larger than the number of microphones,  $N_I > M = 12$ , which typically limits the number of interference sources that can be accommodated (e.g. see [30]). Furthermore, assumptions 1 and 2 implicitly restrict the number of interference sources to be bounded by  $M - N_D = 11$ . This limitation is only for the analysis, and, in practice, improved results are obtained even for a larger number of interference sources. The signal contains 10 segments, demonstrating the number of segments could be smaller than the number of interference sources. The duration of the emitted signal is 10.24s. The input SIR for each interference is  $-6\text{dB}$ . Each interference has a 30% probability of being active at each segment. The position of all the sources is set at random on an arc as described in the first experiment. The activation map of the interference sources at the first Monte Carlo iteration appears in Fig. 8(a), where light blue indicates ‘active’. On average, 4.2 interference sources are active at the same time at each segment. An interference source that is partially active during a segment is considered active during the entire segment (a worst case). Note that there are interference sources active continuously during more than one segment, so their activation is not necessarily related to the division of the received signal into segments. Additionally, there exists no segment in which only one source is active. The output SIRs are presented in Fig. 8(b). The Riemannian DS and SbSp methods appear in blue, the Euclidean DS and SbSp methods are in red, and the intersection method is in orange. It can be seen that the Riemannian approach is superior to the Euclidean one, resulting in higher output SIRs.

We note that Fig. 8(b) serves as an example of an increased number of segments in comparison to Fig. 3. Since there are

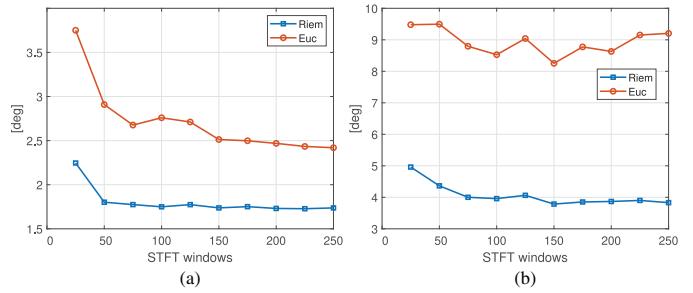


Fig. 9. The RMSE of the Bayesian learning method for 6 desired sources with and without the proposed approach in blue and red, respectively, for (a) SIR  $-6\text{dB}$  and (b) SIR  $-10\text{dB}$ .

on average 4.2 active interference sources at every segment, the results are not improved as the number of segments is increased.

We remark that the empirical results of the proposed approach applied to the typically-used beamformers are not sensitive to the number of samples used for the computation of the sample correlation matrix. Since, in the experiments, additional samples also include samples from at least one interference source, adding samples increases the effect of the interference sources, resulting in weak dependency on the number of samples.

Next, we evaluate the performance of the proposed approach applied to the Bayesian learning method proposed in [23] for sparse signal recovery for DoA estimation. This is done by replacing the typically-used (Euclidean) sample correlation matrix with the Riemannian mean of sample correlation matrices computed in short-time segments. We repeat the same experiment as in [23] with 6 desired sources, only we consider two disjointly active interference sources with SIR  $-6\text{dB}$  and  $-10\text{dB}$ . The SNR is 20dB. The number of desired sources is unknown. We evaluate the performance using 3000 experiments, where at each experiment the direction of the interference sources is generated uniformly at random between  $[-50^\circ, 50^\circ]$ . Fig. 9 presents the RMSE for a different number of STFT windows, which are equivalent to the number of samples for the sample correlation matrix computation of each segment (there are two segments in the experiment). We see that applying the proposed Riemannian approach leads to improved performance for all the tested number of samples.

## VIII. CONCLUSION

We present a Riemannian approach for the design of beamformers and DoA estimation methods for interference rejection in reverberant environments. Specifically, the Riemannian mean is incorporated instead of the sample correlation matrix for DoA estimation of the desired source, as it inherently rejects the interference sources. We analytically show that the DS beamformer, based on the Riemannian geometry of the HPD manifold, results in a higher output SIR than the typical DS beamformer, which implicitly considers the Euclidean geometry. We extend our approach to other beamformers, such as subspace-based beamformers and the MVDR, as well as a Bayesian learning method, experimentally demonstrating superior output SIR

and better DoA estimations in comparison to their Euclidean counterparts.

## APPENDIX A THE INTERSECTION BEAMFORMER

Another beamformer we examine is based on the observation that the desired signal subspace is the intersection of subspaces spanned by eigenvectors of the correlation matrix of the different segments. From each segment,  $i$ , we extract the desired signal subspace from the sample correlation matrix,  $\widehat{\Gamma}_i$ , to obtain  $\mathbf{V}(\widehat{\Gamma}_i)$  whose columns are the  $N_D$  leading eigenvectors. The projection matrix onto the signal subspace of  $\widehat{\Gamma}_i$  is computed as follows

$$\mathbf{P}_{\text{sig}}(\widehat{\Gamma}_i) = \mathbf{V}(\widehat{\Gamma}_i) \left( \mathbf{V}^H(\widehat{\Gamma}_i) \mathbf{V}(\widehat{\Gamma}_i) \right)^{-1} \mathbf{V}^H(\widehat{\Gamma}_i). \quad (52)$$

Since each desired source is active during all the segments, its ATF,  $\mathbf{h}_j^d$ , is an eigenvector of  $\mathbf{V}(\widehat{\Gamma}_i)$ . Consequently, it is also an eigenvector of  $\mathbf{P}_{\text{sig}}(\widehat{\Gamma}_i)$  for all  $i$  with an eigenvalue 1, i.e.  $\mathbf{P}_{\text{sig}}(\widehat{\Gamma}_i)\mathbf{h}_j^d = \mathbf{h}_j^d$ . As a result, it is an eigenvector of  $\mathbf{P}_{\text{sig}} = \prod_{i=1}^{L_s} \mathbf{P}_{\text{sig}}(\widehat{\Gamma}_i)$  with eigenvalue 1 as well since  $\mathbf{P}_{\text{sig}}\mathbf{h}_j^d = \left( \prod_{i=1}^{L_s} \mathbf{P}_{\text{sig}}(\widehat{\Gamma}_i) \right) \mathbf{h}_j^d = \mathbf{h}_j^d$ . We consider the leading  $N_D$  eigenvectors of  $\mathbf{P}_{\text{sig}}$  and form the matrix  $\mathbf{P}_{\text{sig},N_D} \in \mathbb{C}^{M \times N_D}$ , whose columns are the eigenvectors. The intersection spectrum is defined as follows

$$\mathbf{P}_{\text{Intersect}}(\theta) = \mathbf{d}^H(\theta) \mathbf{P}_{\text{sig},N_D} \mathbf{P}_{\text{sig},N_D}^H \mathbf{d}(\theta). \quad (53)$$

We note that in addition to the dimension of the signal space of the matrix  $\mathbf{P}_{\text{sig}}$ , the intersection method requires the estimation of the dimension of the signal space of the correlation matrix for each segment (in order to compute  $\mathbf{V}(\widehat{\Gamma}_i)$ ).

## APPENDIX B ADDITIONAL EXPERIMENTAL RESULTS

We repeat the setting of the first experiment, only with an additional desired source, located at (2m, 3.5m, 2.5m). Fig. 10 is the same as Fig. 3, but the mean output SIR is taken over the two interferences and the two desired sources. Fig. 10(a) presents results for input SIR of  $-10\text{dB}$ , and Fig. 10(b) presents results for input SIR of  $-6\text{dB}$ . We see that the Riemannian approach is superior to the Euclidean one, resulting in higher SIRs. Also, the intersection method is more sensitive to the input SIR, resulting in higher output SIR for  $-6\text{dB}$ . Similar to the setting of one desired source, the Riemannian SbSp method is superior to the Riemannian DS method, as opposed to the Euclidean approach, for which they are on par. This is a consequence of the higher interference attenuation of the Riemannian approach, resulting in a better estimate of the signal space.

Next, we examine the performance of the Riemannian MVDR and the Euclidean MVDR beamformers, given by (48), for  $\widehat{\Gamma}_R$  and  $\widehat{\Gamma}_E$ , respectively. Fig. 11 presents the results. It is the same as Fig. 3, only for the MVDR beamformer. We see that the Riemannian approach is superior to the Euclidean one, however, with a slight advantage.

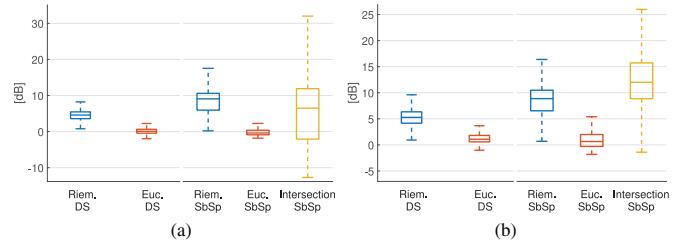


Fig. 10. Output SIR of the different beamformers for two desired sources and two interference sources. (a) The input SIR is  $-10\text{dB}$ . (b) The input SIR is  $-6\text{dB}$ .

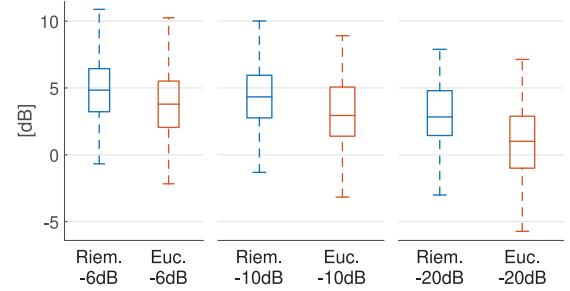


Fig. 11. The mean output SIR for the Riemannian and the Euclidean MVDR beamformers in the presence of two interference sources. The x-axis indicates the input SIR, and the y-axis indicates the output SIR.

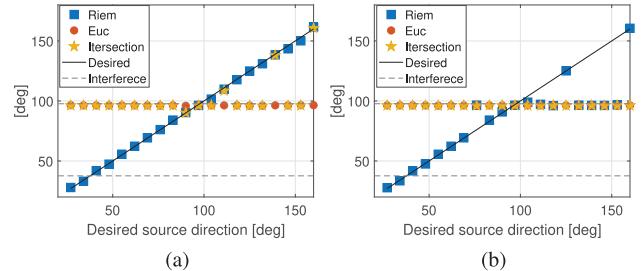


Fig. 12. Estimation of the DoA of the desired source for speech signals where (a) the input SIR is  $-6\text{dB}$ , and (b) the input SIR is  $-10\text{dB}$ .

We repeat the experiment with speech signals from the TIMIT dataset. For each source, the speaker and the time interval are chosen uniformly at random. Fig. 12 presents the results. We see that the proposed approach leads to improved results in comparison to the typically-used beamformers also for speech signals.

## APPENDIX C EXTENSION TO A STREAMING DATA SETTING

In a streaming data setting, we do not have access in advance to the entire signal, and the direction estimation is updated as more samples become available. Therefore, in this setting, we cannot compute the Riemannian mean using (17). To circumvent this, we turn to the estimator of the Riemannian mean proposed in [37], [46], which is updated after every received segment.

Each segment,  $i$ , is processed separately using  $L_w$  STFT windows, and an estimation of the correlation matrix,  $\widehat{\Gamma}_i$ , is computed using (16). Next, the estimation of the Riemannian

**Algorithm 3** Streaming DoA estimation in the presence of multiple interferences

**Input:** the current result of the STFT window of the received signal

**Output:** the estimated direction of the desired source  $\hat{\theta}$

- 1: Set  $\hat{\mathbf{R}}_0 = \mathbf{I}$
- 2: Repeat
  - 1) Accumulate  $L_w$  STFT windows (that form a segment)
  - 2) Compute its sample correlation matrix,  $\hat{\Gamma}_i$ , using (16)
  - 3) Compute  $\hat{\mathbf{R}}_i = \hat{\mathbf{R}}_{i-1}^{\frac{1}{2}} (\hat{\mathbf{R}}_{i-1}^{-\frac{1}{2}} \hat{\Gamma}_i \hat{\mathbf{R}}_{i-1}^{-\frac{1}{2}})^{\frac{1}{i}} \hat{\mathbf{R}}_{i-1}^{\frac{1}{2}}$
  - 4) Compute  $P_{DS}(\theta; \hat{\mathbf{R}}_i)$  using (18)
  - 5) Return  $\hat{\theta} = \text{argmax}_{\theta} P_{DS}(\theta; \hat{\mathbf{R}}_i)$

mean, denoted as  $\hat{\mathbf{R}}_i$ , which is the adaptive counterpart of  $\Gamma_R$ , is updated using the following update step

$$\hat{\mathbf{R}}_i = \hat{\mathbf{R}}_{i-1}^{\frac{1}{2}} (\hat{\mathbf{R}}_{i-1}^{-\frac{1}{2}} \hat{\Gamma}_i \hat{\mathbf{R}}_{i-1}^{-\frac{1}{2}})^{\frac{1}{i}} \hat{\mathbf{R}}_{i-1}^{\frac{1}{2}}. \quad (54)$$

The SRP of the DS beamformer is computed using (18) with  $\hat{\mathbf{R}}_i$  as an estimate of  $\hat{\Gamma}_R$ , and the direction to the desired source is set using (19). The algorithm is described in Algorithm 3.

We note that the current correlation matrix estimate has a full rank if  $L_w > M$ . This implies a latency of at least  $M$  times the duration of the STFT window. In case the STFT is performed with overlap this latency is reduced accordingly.

As for the Euclidean alternative, its estimator is updated in finer granularity at every STFT window,  $l$ . The estimator at the  $l$ th update is denoted by  $\mathbf{E}_l$  and is computed as follows

$$\mathbf{E}_l = \frac{n-1}{n} \mathbf{E}_{l-1} + \frac{1}{n} \mathbf{z}(l) \mathbf{z}^H(l). \quad (55)$$

Setting  $L_w = 1$  in step 2.1 in Algorithm 3, and substituting (55) in step 2.3 in Algorithm 3, results in the streaming version of the Euclidean counterpart.

#### APPENDIX D

#### ON THE PARTICULAR CHOICE OF THE RIEMANNIAN METRIC

In this work, we consider the Affine Invariant metric (also called Fisher Information metric) [47]. Another commonly-used metric in the space of HPD matrices is the Log-Euclidean metric [48], which could be viewed as a computationally efficient local approximation of the Affine Invariant metric. The induced Log-Euclidean distance is given by

$$d_{LE}^2(\Gamma_1, \Gamma_2) = \|\log(\Gamma_1) - \log(\Gamma_2)\|_F^2. \quad (56)$$

In the context of this work, the Affine Invariant metric is advantageous over the Log Euclidean because it better enhances the desired source subspace relative to the interference and noise subspace, as we show next. We remark that the derivation is similar to the derivation in Section V, where we demonstrate the advantage of the Affine Invariant metric over Euclidean geometry. First, we note that the following holds [49]:

$$\text{tr}(\Gamma_R) \leq \text{tr}(\Gamma_{LE}), \quad (57)$$

where  $\Gamma_{LE}$  denotes the Riemannian mean based on the Log-Euclidean metric. Second, the ATF to the desired source,  $\mathbf{h}_0$ , is a common eigenvector of all the population correlation matrices per segment. Consequently, according to Lemma 2 below, the mean correlation matrices based on both metrics have the same eigenvalue associated with the common eigenvector  $\mathbf{h}_0$ . More specifically, denoting  $\lambda_0(\Gamma) \equiv \frac{\mathbf{h}_0^H \Gamma \mathbf{h}_0}{\|\mathbf{h}_0\|^2}$ , we get  $\lambda_0(\Gamma_R) = \lambda_0(\Gamma_{LE})$ .

**Lemma 2:** Let  $\{\Gamma_j\}_j$  be the set of HPD matrices having a common eigenvalue  $\lambda_0$  associated with the common eigenvector  $\mathbf{h}_0$ . Then

$$\mathbf{h}_0^H \Gamma_R \mathbf{h}_0 = \mathbf{h}_0^H \Gamma_{LE} \mathbf{h}_0 = \mathbf{h}_0^H \Gamma_E \mathbf{h}_0 = \|\mathbf{h}_0\|^2 \lambda_0, \quad (58)$$

where  $\Gamma_R$ ,  $\Gamma_{LE}$ , and  $\Gamma_E$  are the Riemannian means based on the Affine Invariant and the Log-Euclidean metrics, and the Euclidean mean.

The rest of the eigenvectors of the correlation matrices span the interference and noise subspace. We recall that  $\lambda_i(\Gamma)$  is the  $i$ th eigenvalue, so we get

$$\frac{\lambda_0(\Gamma_R)}{\sum_{i=1}^{M-1} \lambda_i(\Gamma_R)} \geq \frac{\lambda_0(\Gamma_{LE})}{\sum_{i=1}^{M-1} \lambda_i(\Gamma_{LE})}, \quad (59)$$

which follows from Lemma 2 and (57).

We see that the Riemannian mean induced by the Affine Invariant metric captures better the desired signal subspace in comparison to the Riemannian mean induced by the Log Euclidean metric and the Euclidean mean (according to (24)), entailing an advantage in SbSp methods, for example.

#### ACKNOWLEDGMENT

The authors wish to thank the associate editor and the anonymous reviewers for their useful comments which helped to improve this manuscript.

#### REFERENCES

- [1] S. Gannot, M. Haardt, W. Kellermann, and P. Willett, “Introduction to the issue on acoustic source localization and tracking in dynamic real-life scenes,” *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 3–7, Mar. 2019.
- [2] V. Krishnameni, T. Kesavamurthy, and B. Aparna, “Beamforming for direction-of-arrival (DOA) estimation—a survey,” *Int. J. Comput. Appl.*, vol. 61, no. 11, pp. 4–11, 2013.
- [3] B. Widrow, P. Mantey, L. Griffiths, and B. Goode, “Adaptive antenna systems,” *Proc. IEEE*, vol. 55, no. 12, pp. 2143–2159, Dec. 1967.
- [4] R. T. Compton, “*Adaptive Antennas: Concepts and Performance*.” Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
- [5] F. Vook and R. Compton, “Bandwidth performance of linear adaptive arrays with tapped delay-line processing,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 901–908, Jul. 1992.
- [6] K. Yao, J. C. Chen, and R. E. Hudson, “Maximum-likelihood acoustic source localization: Experimental results,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3. Piscataway, NJ, USA: IEEE, 2002, pp. III-2949–III-2952.
- [7] F. Akbari, S. S. Moghaddam, and V. T. Vakili, “MUSIC and MVDR DOA estimation algorithms with higher resolution and accuracy,” in *Proc. 5th Int. Symp. Telecommun.* Piscataway, NJ, USA: IEEE, 2010, pp. 76–81.
- [8] D. Salvati, C. Dioli, and G. L. Foresti, “A weighted MVDR beamformer based on SVM learning for sound source localization,” *Pattern Recognit. Lett.*, vol. 84, pp. 15–21, Dec. 2016.
- [9] D. W. Rieken and D. R. Fuhrmann, “Generalizing MUSIC and MVDR for multiple noncoherent arrays,” *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2396–2406, Sep. 2004.

- [10] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [11] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [12] J. Xu, G. Liao, S. Zhu, and L. Huang, "Response vector constrained robust LCMV beamforming based on semidefinite programming," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5720–5732, Nov. 2015.
- [13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
- [14] F. Yan, M. Jin, and X. Qiao, "Low-complexity DOA estimation based on compressed MUSIC and its performance analysis," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1915–1930, Apr. 2013.
- [15] X. Zhang, L. Xu, L. Xu, and D. Xu, "Direction of departure (DOD) and direction of arrival (DOA) estimation in MIMO radar with reduced-dimension MUSIC," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1161–1163, Dec. 2010.
- [16] P. Vallet, X. Mestre, and P. Loubaton, "Performance analysis of an improved MUSIC DoA estimator," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6407–6422, Dec. 2015.
- [17] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, New York, NY, USA: Springer, 2001, pp. 157–180.
- [18] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.
- [19] D. Malioutov, M. Cetin, and A. S. 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.
- [20] N. Hu, Z. Ye, X. Xu, and M. Bao, "DOA estimation for sparse array via sparse signal reconstruction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 760–773, Apr. 2013.
- [21] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 26–29, Jan. 2014.
- [22] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 49, no. 2, pp. 760–773, Apr. 2013.
- [23] N. Hu, B. Sun, J. Wang, J. Dai, and C. Chang, "Source localization for sparse array using nonnegative sparse Bayesian learning," *Signal Process.*, vol. 127, pp. 37–43, 2016.
- [24] J. Dai and H. C. So, "Real-valued sparse Bayesian learning for DOA estimation with arbitrary linear arrays," *IEEE Trans. Signal Process.*, vol. 69, pp. 4977–4990, 2021.
- [25] R. Bhatia and J. Holbrook, "Riemannian geometry and matrix geometric means," *Linear Algebra Appl.*, vol. 413, nos. 2–3, pp. 594–618, 2006.
- [26] F. Nielsen and R. Bhatia, *Matrix Information Geometry*. Berlin, Germany: Springer, 2013.
- [27] L. T. Skovgaard, "A Riemannian geometry of the multivariate normal model," *Scand. J. Statist.*, vol. 11, no. 4, pp. 211–223, Jan. 1984.
- [28] O. Katz, R. R. Lederman, and R. Talmon, "Spectral flow on the manifold of SPD matrices for multimodal data processing," 2020, *arXiv:2009.08062*.
- [29] E. A. Habets, J. Benesty, and P. A. Naylor, "A speech distortion and interference rejection constraint beamformer," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 854–867, Mar. 2012.
- [30] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.
- [31] M. Arnaudon, F. Barbaresco, and L. Yang, "Riemannian medians and means with applications to radar signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 4, pp. 595–604, Aug. 2013.
- [32] H. Chahroud, R. M. Dansereau, S. Rajan, and B. Balaji, "Target detection through Riemannian geometric approach with application to drone detection," *IEEE Access*, vol. 9, pp. 123950–123963, 2021.
- [33] H. Chahroud, R. Dansereau, S. Rajan, and B. Balaji, "Improved covariance matrix estimation using Riemannian geometry for beamforming applications," in *Proc. IEEE Int. Radar Conf. (RADAR)*. Piscataway, NJ, USA: IEEE, 2020, pp. 693–697.
- [34] F. Hiai and D. Petz, "Riemannian metrics on positive definite matrices related to means," *Linear Algebra Appl.*, vol. 430, nos. 11–12, pp. 3105–3130, 2009.
- [35] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, 2005.
- [36] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.
- [37] J. Ho, G. Cheng, H. Salehian, and B. Vemuri, "Recursive Karcher expectation estimators and geometric law of large numbers," in *Proc. Artif. Intell. Statist. PMLR*, 2013, pp. 325–332.
- [38] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 581–585, May 2014.
- [39] Y. Lim and M. Pálfia, "Matrix power means and the Karcher mean," *J. Functional Anal.*, vol. 262, no. 4, pp. 1498–1514, 2012.
- [40] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.
- [41] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 4, pp. 546–555, May 2009.
- [42] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [43] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. America*, vol. 65, no. 4, pp. 943–950, 1979.
- [44] E. A. Habets, "Room impulse response (RIR) generator," 2008. [Online]. Available: <https://github.com/ehabets/RIR-Generator>
- [45] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.
- [46] G. Cheng, J. Ho, H. Salehian, and B. C. Vemuri, "Recursive computation of the Fréchet mean on non-positively curved Riemannian manifolds with applications," in *Riemannian Computer in Computer Vision*. Springer, 2016, pp. 21–43. [Online]. Available: [https://doi.org/10.1007/978-3-319-22957-7\\_2](https://doi.org/10.1007/978-3-319-22957-7_2)
- [47] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vision*, vol. 66, no. 1, pp. 41–66, 2006.
- [48] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrices," *SIAM J. Matrix Analysis Appl.*, vol. 29, no. 1, pp. 328–347, 2007.
- [49] M. Congedo, B. Afsari, A. Barachant, and M. Moakher, "Approximate joint diagonalization and geometric mean of symmetric positive definite matrices," *PLoS One*, vol. 10, no. 4, 2015, Art. no. e0121423.