

RTF Estimation Using Riemannian Geometry for Speech Enhancement in the Presence of Interferences

Or Ronai*, Yuval Sitton*, Amitay Bar, and Ronen Talmon

Viterbi Faculty of Electrical and Computer Engineering

Technion - Israel Institute of Technology

Haifa, Israel

{or.ronai@campus, yuvalsitton@campus, amitayb@campus, ronem@ee}.technion.ac.il

Abstract—We address the problem of multichannel audio signal enhancement in reverberant environments with interfering sources. We propose an approach that leverages the Riemannian geometry of the spatial correlation matrices of the received signals to estimate the relative transfer function (RTF) of the desired source. Specifically, we compute the spatial correlation matrices in short-time segments, and subsequently, their Riemannian mean, which preserves shared spatial components while attenuating unshared ones. This enables an effective rejection of intermittent interference, leading to accurate RTF estimation. We experimentally show that when the proposed RTF estimation is incorporated into the Minimum Variance Distortionless Response (MVDR) beamformer, it enhances the desired signal, outperforming the MVDR beamformer that is based on standard (Euclidean) RTF estimation. These favorable experimental results are demonstrated in challenging acoustic environments including multiple strong interfering sources, noise, and reverberations.

Index Terms—Speech enhancement, interference rejection, Riemannian geometry, RTF estimation

I. INTRODUCTION

Multichannel audio signal enhancement using microphone arrays has been an active area of research for many years [1], [2], with various applications involving hearing aids, speech recognition, hands-free communication, and teleconferencing, to name but a few [3]–[6]. Still, despite significant advancements, estimating a desired signal from noisy and reverberant measurements remains a challenge.

One prominent approach for signal enhancement in reverberant environments is based on the relative transfer function (RTF) [7], which is first estimated from the noisy and reverberant measurements, and then, incorporated into a beamformer for signal enhancement. Over the years, numerous RTF estimation methods have been developed. For example, in [7]–[9], the RTFs are estimated based on the power spectral density (PSD) and the cross-PSD of the desired signal in time-frequency bins containing the desired signal. However, these methods depend on estimating the probability of desired speech presence, which is unknown and challenging to estimate in noisy conditions, and even more so, in the presence of interferences. In [10], the RTFs are estimated using the least-squares (LS) method of the cross-PSD proposed in [7] using time segments where only the speaker of interest is active, which are assumed to be known. In [11], multiple interfering sources are considered and the RTFs are estimated using generalized eigenvalue decomposition (GEVD) of the PSD matrices. However, their method relies on time segments with specific and known source activity. In [12], it is proposed to estimate the acoustic transfer function (ATF) using techniques developed in [13] based on Bézout’s theorem. Their method assumes that the different sources are not active simultaneously. As suggested by

*Equal contribution. The authors thank Nimrod Peleg, Yair Moshe, and the Signal and Image Processing Lab (SIPL) staff for their support and helpful suggestions. This work was supported by the European Union’s Horizon 2020 research and innovation programme (grant No. 802735-ERC-DIFFOP).

the non-exhaustive survey above, existing RTF estimation methods face challenges in multi-source scenarios because they require time segments where only a single desired source is active, which are often unavailable, or estimates of the desired source presence.

In this paper, we consider a reverberant and noisy environment with interfering sources. The desired source is assumed to be constantly active, whereas the interfering sources are assumed to be intermittently active. The activity times of the different sources, their positions, and their power are unknown. Our goal is to estimate the desired signal as received in a reference microphone, where the main challenge is the presence of the interfering sources. To this end, we propose to exploit the Riemannian geometry of the spatial sample correlation matrices, which are Hermitian Positive Definite (HPD) matrices. Specifically, we compute the Riemannian mean of spatial sample correlation matrices in short time segments and leverage an intriguing property of the Riemannian mean – that it preserves shared spatial components while attenuating unshared ones [14]–[16] – for estimating the RTF of the desired source. The RTF estimation is implemented using the EVD of the Riemannian mean, and in turn, it is incorporated into the design of a beamformer that enhances the desired source. We demonstrate the proposed approach using the Minimum Variance Distortionless Response (MVDR) beamformer [17]. Usually, the MVDR beamformer is computed based on the spatial sample correlation matrix, which in our context, is viewed as the Euclidean mean and functions as a baseline in our experiments. We empirically show that the proposed approach obtains better-desired signal enhancement in terms of signal-to-interference ratio (SIR) compared with its Euclidean counterpart.

We remark that Riemannian geometry has already been shown to be useful in array signal processing, e.g., in [18]–[23], and specifically, in attenuating interfering sources for direction of arrival (DoA) estimation [16]. In addition, we consider the MVDR beamformer as a prototypical beamformer for signal enhancement, however, the proposed approach could be applied to other beamformers as well.

II. PROBLEM FORMULATION

Consider an array of M microphones in a noisy and reverberant environment with a single desired source and N interfering sources. All the sources are static. The signal received at the m th microphone is given by

$$z_m(n) = s^d(n) * h_m^d(n) + \sum_{j=1}^N s_j^i(n) * h_{jm}^i(n) + v_m(n), \quad (1)$$

where $s^d(n)$ and $s_j^i(n)$ are the signals from the desired source and the j th interfering source, respectively, $h_m^d(n)$ is the acoustic impulse response (AIR) between the desired source and the m th microphone, $h_{jm}^i(n)$ is the AIR between the j th interfering source

and the m th microphone, $v_m(n)$ is the noise at the m th microphone, and $*$ denotes convolution. The desired source is constantly active during the observation interval, while the interfering sources exhibit intermittent activity with no predefined structure. The desired source, the interfering sources, and the noise are assumed to be uncorrelated. In addition, the noise is assumed to be spatially white.

We analyze the received signal using the Short-Time Fourier Transform (STFT) with a window size of K samples. The signal $z_m(n)$ in (1) in the STFT domain is given by

$$z_m(l, k) = s^d(l, k)h_m^d(l, k) + \sum_{j=1}^N s_j^i(l, k)h_{jm}^i(l, k) + v_m(l, k), \quad (2)$$

where l denotes the time frame, k denotes the frequency index, $h_m^d(l, k)$ is the ATF between the desired source and the m th microphone, and $h_{jm}^i(l, k)$ is the ATF between the i th interfering source and the m th microphone. Since we assume all the sources are static, their ATFs do not change over time, so in the following, we omit their STFT time frame index l . We define the RTF between the desired source and the m th microphone with respect to the first microphone (for arbitrary indexing) as [7]

$$r_m^d(k) \triangleq \frac{h_m^d(k)}{h_1^d(k)}, \quad (3)$$

and we stack the RTF of the desired source into a vector $\mathbf{r}^d(k) \in \mathbb{C}^{M \times 1}$ as follows

$$\mathbf{r}^d(k) = \begin{bmatrix} 1 & h_2^d(k) & \dots & h_M^d(k) \\ h_1^d(k) & & & h_1^d(k) \end{bmatrix}^\top. \quad (4)$$

The definition of $r_{jm}^i(k)$, and $\mathbf{r}_j^i(k)$ for the j th interfering source follows similarly. Henceforth, we refer to the first microphone as the reference microphone.

We stack the received signals at all microphones into a vector $\mathbf{z}(l, k) = [z_1(l, k), \dots, z_M(l, k)]^\top \in \mathbb{C}^{M \times 1}$, which, using the RTFs, is given by

$$\mathbf{z}(l, k) = x^d(l, k)\mathbf{r}^d(k) + \sum_{j=1}^N x_j^i(l, k)\mathbf{r}_j^i(k) + \mathbf{v}(l, k), \quad (5)$$

where $x^d(l, k) \triangleq s^d(l, k)h_1^d(k)$, and $x_j^i(l, k) \triangleq s_j^i(l, k)h_{j1}^i(k)$ are the desired source and the j th interfering source received at the first microphone, respectively, and $\mathbf{v}(l, k)$ is defined as the column stack vector of $v_m(l, k)$. Similarly, we define $\mathbf{h}^d(k)$ and $\mathbf{h}_j^i(k)$ as the column stack vectors of $h_m^d(k)$ and $h_{jm}^i(k)$, respectively.

Our goal is to enhance the desired signal at the reference microphone, $x^d(l, k)$, based on the received signals at the microphone array $\mathbf{z}(l, k)$, for all l and k . The primary challenge lies in the presence of the interfering sources, which are unknown, positioned at unknown locations, and could be stronger than the desired source.

III. PROPOSED APPROACH

A key component in many beamformers is the RTF to the desired source. We propose to consider the Riemannian geometry of the HPD manifold for estimating the RTF of the desired source and show that it is particularly useful in the presence of interfering sources. Specifically, we propose to use the Riemannian mean of the spatial correlation matrices, considering the Affine-Invariant (AI) metric [24], [25]. The AI metric was shown to preserve the desired source subspace better and to attenuate the interference and noise subspace compared to other metrics [16, Appendix D].

Following [16], we divide the received signal, $\mathbf{z}(l, k)$, into L disjoint segments, each containing W consecutive STFT windows.

For each segment $i \in \{1, \dots, L\}$, we estimate the sample correlation matrix of the k th frequency, denoted by $\hat{\Gamma}_i(k) \in \mathbb{C}^{M \times M}$, as follows

$$\hat{\Gamma}_i(k) = \frac{1}{W} \sum_{l=(i-1) \cdot W+1}^{i \cdot W} \mathbf{z}(l, k)\mathbf{z}^H(l, k). \quad (6)$$

Note that the number of STFT windows W must be at least the number of the microphones M , so that $\hat{\Gamma}_i(k)$ has a full rank.

As each correlation matrix $\hat{\Gamma}_i(k)$ is a point on the HPD manifold, we compute the Riemannian mean of the set of L correlation matrices $\{\hat{\Gamma}_i(k)\}_i$ by

$$\hat{\Gamma}_R(k) = \arg \min_{\Gamma \in \mathcal{M}} \sum_i d_R^2(\Gamma, \hat{\Gamma}_i(k)), \quad (7)$$

where \mathcal{M} is the Riemannian manifold constituted by the space of HPD matrices, and $d_R^2(\Gamma_1, \Gamma_2)$ is the distance between two matrices Γ_1 and Γ_2 , induced by the AI metric. For the AI metric, there is no closed-form expression for the Riemannian mean for more than two matrices. Therefore, we compute the Riemannian mean using an iterative algorithm [26, Algorithm 1]. For brevity, the frequency index k is omitted from $\hat{\Gamma}_R$.

Next, we estimate the ATF of the desired source up to a complex scalar [11], [27] by computing the dominant eigenvector associated with the largest eigenvalue of $\hat{\Gamma}_R$ as follows

$$\hat{\mathbf{h}}^d = \arg \max_{\alpha \in \mathbb{C}^M, \alpha \neq 0} \frac{\alpha^H \hat{\Gamma}_R \alpha}{\alpha^H \alpha}. \quad (8)$$

Then, the RTF of the desired source, denoted by $\mathbf{r}^d(\hat{\Gamma}_R)$, is computed using (4). We note that RTF estimation is a non-blind estimation problem, in contrast to estimating the ATF, which is a blind estimation problem since the input, i.e., the source signal, is unknown. Furthermore, computing the RTF resolves the complex scalar ambiguity in the ATF estimation in (8). We also note that common algorithms for RTF estimation, e.g., [7]–[9] that rely on the ratio between the cross-PSD and the PSD of the signals are not effective in the presence of interfering sources, particularly if the desired signal is speech [28]. Consequently, we present a different approach using the EVD of the spatial correlation matrix. Although this approach seems to have disadvantages, such as the need to solve a harder problem of ATF estimation first and its susceptibility to interfering sources, we posit (and demonstrate empirically in the sequel) that using the Riemannian mean instead of the Euclidean mean offers significant benefits.

Next, we compute the MVDR coefficients [17] using the RTF estimation of the desired source, $\mathbf{r}^d(\hat{\Gamma}_R)$, by

$$\mathbf{w}_{\text{mvdr}}(k) = \frac{\hat{\Gamma}^{-1}(k)\mathbf{r}^d(\hat{\Gamma}_R)}{\mathbf{r}^{dH}(\hat{\Gamma}_R)\hat{\Gamma}^{-1}(k)\mathbf{r}^d(\hat{\Gamma}_R)}, \quad (9)$$

where $\hat{\Gamma}(k)$ is the sample correlation matrix over the entire interval. Note that $\hat{\Gamma}(k)$ can be recast as the Euclidean mean of $\{\hat{\Gamma}_i(k)\}_i$:

$$\hat{\Gamma}(k) = \frac{1}{L} \sum_{i=1}^L \hat{\Gamma}_i(k), \quad (10)$$

and thus, replacing $\hat{\Gamma}_R$ with $\hat{\Gamma}$ in (8) and (9) gives rise to the standard (Euclidean) RTF estimate $\mathbf{r}^d(\hat{\Gamma})$ and MVDR coefficients.

Finally, the estimated desired signal is given by

$$\hat{x}^d(l, k) = \mathbf{w}_{\text{mvdr}}^H(k)\mathbf{z}(l, k). \quad (11)$$

The signal in the time domain, $\hat{x}^d(n)$, is computed by the Inverse Short-Fourier Transform (ISTFT). The proposed approach is summarized in Algorithm 1. We conclude with two remarks. First, we

Algorithm 1 Desired signal enhancement

Input: the received signal in the STFT domain $\{z(l, k)\}_{l, k}$ **Output:** the enhanced desired signal $\hat{x}^d(n)$

- 1: **for** each frequency k **do**
 - 2: Divide $\{z(l, k)\}_l$ into L consecutive segments
 - 3: Compute $\{\hat{\Gamma}_i(k)\}_{i=1}^L$ using (6)
 - 4: Compute $\hat{\Gamma}_R$ of the set $\{\hat{\Gamma}_i(k)\}_{i=1}^L$ using [26, Algorithm 1]
 - 5: Estimate $r^d(\hat{\Gamma}_R)$ by (8) and (4)
 - 6: Compute $w_{\text{mvdr}}(k)$ using (9) and (10)
 - 7: Compute $\hat{x}^d(l, k)$ using (11)
 - 8: **end for**
 - 9: **return** the ISTFT of $\hat{x}^d(l, k)$, i.e. $\hat{x}^d(n)$
-

consider the MVDR beamformer for its popularity, however, the proposed approach could be applied to other beamformers as well. Second, the proposed approach could be implemented in an online manner similar to the online algorithm in [16].

A. Theoretical Support

For analysis, we consider the *population* correlation matrices, i.e., the expectation of (6). For each frequency, their Riemannian and Euclidean means are given by [16]

$$\Gamma = \sigma_d^2 \mathbf{h}^d \mathbf{h}^{dH} + \sum_{j=1}^N \mu_j^2 \mathbf{h}_j^i \mathbf{h}_j^{iH} + \sigma_v^2 \mathbf{I}, \quad (12)$$

where σ_d^2 and σ_v^2 are the power of the desired source and the noise, respectively, and $\mathbf{I} \in \mathbb{R}^{M \times M}$ is the identity matrix. The parameter μ_j depends on the used geometry. For the Riemannian mean, it is given by

$$\mu_j^2 = \frac{(\sigma_{i_j}^2 \|\mathbf{h}_j^i\|^2 + \sigma_v^2)^{\tau_j} (\sigma_v^2)^{1-\tau_j} - \sigma_v^2}{\|\mathbf{h}_j^i\|^2}, \quad (13)$$

and for the Euclidean mean, it is given by

$$\mu_j^2 = \sigma_{i_j}^2 \tau_j, \quad (14)$$

where $\sigma_{i_j}^2$ is the power of the j th interfering source, and τ_j is the relative number of segments during which the j th interfering source is active. We see that the coefficients of the desired source subspace σ_d^2 and the noise subspace σ_v^2 in (12) are equal for both the Euclidean and the Riemannian means. However, the coefficients of the interfering source subspace μ_j depend on the considered geometry. While for the Euclidean mean they depend only on the interference power and the duration of the activity, for the Riemannian mean they also depend on the noise power and the corresponding ATF. In the context of DoA estimation, it was shown that (14) is greater than (13) [16], indicating that the Riemannian mean leads to a much greater interference rejection, i.e., greater attenuation of the interfering sources subspace, than the Euclidean mean. Building on this result, the dominant eigenvector of the Riemannian mean is assumed to be related to the ATF of the desired source since the other spatial components associated with the interfering sources are attenuated.

IV. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed Riemannian approach and compare it to its Euclidean counterpart (The code is available [here](#)). We consider a reverberant enclosure ($8\text{m} \times 6\text{m} \times 3.5\text{m}$), with a reverberation time of $\beta = 150\text{ms}$, and a uniform linear microphone array of $M = 16$ microphones. The leftmost microphone is located at

(3.7m, 1m, 2m), and the distance between the microphones is 4.26cm along the x -axis. The AIR between a source and a microphone in the array is generated using a Room Impulse Response (RIR) generator [29] that implements the image method [30]. The TIMIT corpus [31] is used for the speech signals. The sampling frequency is 16KHz, and the length of the AIRs is 2048 samples. The STFT window size is $K = 1024$ samples with a 50% overlap using Hanning window. The performance is evaluated based on 100 Monte Carlo iterations, where the location of the desired source, the desired and interfering signals, and the additive noise are randomized in each iteration. The evaluation metric is the ΔSIR , defined as $\Delta\text{SIR} = \text{oSIR} - \text{iSIR}$, where oSIR is the output SIR (the SIR of \hat{x}^d), and iSIR is the input SIR (the SIR of z_1). In the following, we refer to SNR and SIR, indicating the SNR and SIR of z_1 , respectively.

In the first experiment, we consider a single desired source and two interfering sources. All the sources are positioned on an arc of radius 2.7m from the center of the array on the XY plane at a height of 1.5m. The desired source is positioned uniformly at random within a sector of $[65^\circ, 115^\circ]$. The interfering sources are located at 31.5° and 141.1° . The signal duration is 4.096s. We consider two segments, where the desired source is constantly active, while each interfering source is disjointly active in one of the segments. We report the results on this particular setting for convenience, but different configurations were tested, all producing similar results.

Figure 1a presents the ΔSIR as a function of the SIR at a fixed SNR of 20dB and Fig. 1b presents the ΔSIR as a function of the SNR with a SIR of -6dB . The proposed approach and its Euclidean counterpart appear in blue and red, respectively. We observe from Figs. 1a and 1b that the proposed approach results in improved SIR, outperforming its Euclidean counterpart by a large margin, reaching up to 20dB. The improved SIR indicates better signal enhancement obtained by greater interference rejection. Additionally, we see from Fig. 1b that the Euclidean approach is not sensitive to changes in the SNR. The reason is that the main challenge in this setting is the interfering sources rather than the noise. In contrast, the proposed approach is sensitive to the SNR. The higher the SNR, the larger the improvement achieved by the proposed approach. This is in accordance with the results in [16], where the direction of arrival estimation problem was considered, and it was shown both theoretically and experimentally that the Riemannian approach outperforms its Euclidean counterpart demonstrating higher sensitivity to the SNR. We also examine the performance for different reverberation times with SNR of 20dB and SIR of -6dB . Figure 1c presents the results. We see that the shorter the reverberation time, the higher the ΔSIR obtained by the Riemannian approach. We found that increased reverberation time increases the correlation between the RTF of the desired source and the RTF of an interfering source. This could imply that the estimated RTF used by the MVDR beamformer also contains a component that points toward an interfering source. This component becomes more dominant as the reverberation time increases. We do not present this result due to space limitations. To demonstrate that the proposed approach can be applied to other beamformers, we consider the Kronecker MVDR (KMVDR) method [32], [33], and incorporate the proposed Riemannian approach for estimating the RTF. The estimated RTF is in turn used in the beamformer proposed in [33, Eqs. 17 and 23]. We repeat the experiment above and present the results in Fig. 1d. We see similar trends, where the Riemannian approach demonstrates larger interference rejection.

To further quantitatively evaluate the proposed approach, we consider the perceptual evaluation of speech quality (PESQ) [34], and the short-time objective intelligibility (STOI) [35]. The results appear in

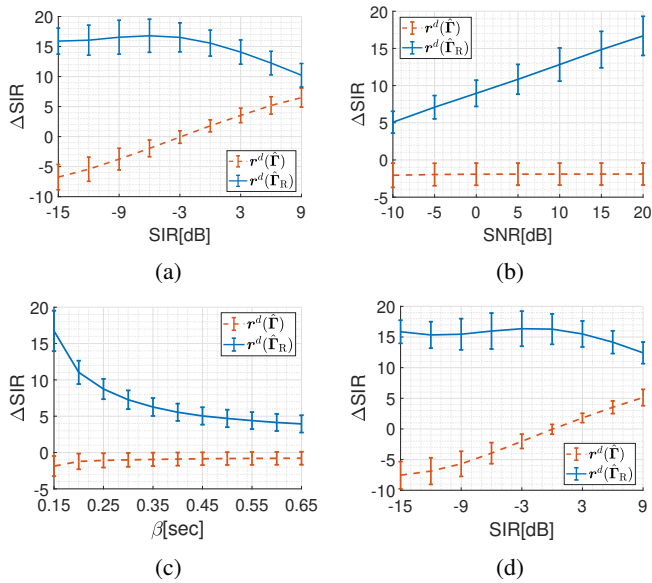


Fig. 1: Δ SIR obtained by the proposed approach (solid blue line) and its Euclidean counterpart (dashed red line) with the MVDR beamformer (a, b and c) and the KMVDR beamformer (d). (a) Δ SIR vs. SIR with SNR of 20dB and $\beta = 150$ ms. (b) Δ SIR vs. SNR with SIR of -6 dB and $\beta = 150$ ms. (c) Δ SIR vs. β with SIR of -6 dB and SNR of 20dB. (d) Same as (a) but with KMVDR.

TABLE I: The PESQ and STOI measures for SNR of 20dB, SIR of -6 dB, and $\beta = 150$ ms.

	Euclidean	Riemannian
PESQ	1.16 ± 0.37	2.57 ± 0.27
STOI	0.51 ± 0.09	0.90 ± 0.05

Table I. We see that the proposed approach demonstrates significantly higher PESQ and STOI.

Figure 2 depicts an example of the sonograms for SNR of 20dB and SIR of -6 dB of z_1 (Fig. 2a), x^d (Fig. 2b), \hat{x}^d obtained using the standard MVDR beamformer (Euclidean) (Fig. 2c), and \hat{x}^d obtained using the proposed Riemannian approach (Fig. 2d). We see that the proposed approach results in an enhanced signal that is closer to the desired signal. In contrast, the Euclidean approach results in a mixture of the desired and interfering signals. For example, between 2.9s and 3.1s, we see the effect of an interfering source by comparing Figs. 2a and 2b. The proposed Riemannian approach demonstrates superior attenuation of this interfering source as shown in Fig. 2d, compared to the Euclidean approach, as seen in Fig. 2c. We see from Fig. 2c that the Euclidean approach can lead to truncations at certain frequencies, for example, at 1.5KHz in the first half of the sonogram, and 2KHz and 3KHz in the second half of the sonogram. At these frequencies, the estimated RTF corresponds to an interfering source rather than the desired source. Since each interfering source is only partially active, at times when it is not active, the beamformer points at a nonactive source resulting in low power at the sonogram. Additionally, we see in Fig. 2d an effect of noise amplifications, for example at 0.5KHz, 3KHz, and 4.5KHz. This stems from the rejection of the interfering sources, which colors the noise and could amplify it.

Finally, we consider 5 interfering sources, arbitrarily located at 159.9° , 33.9° , 152.9° , 44.6° and 41.0° , each at a distance of 2.7m

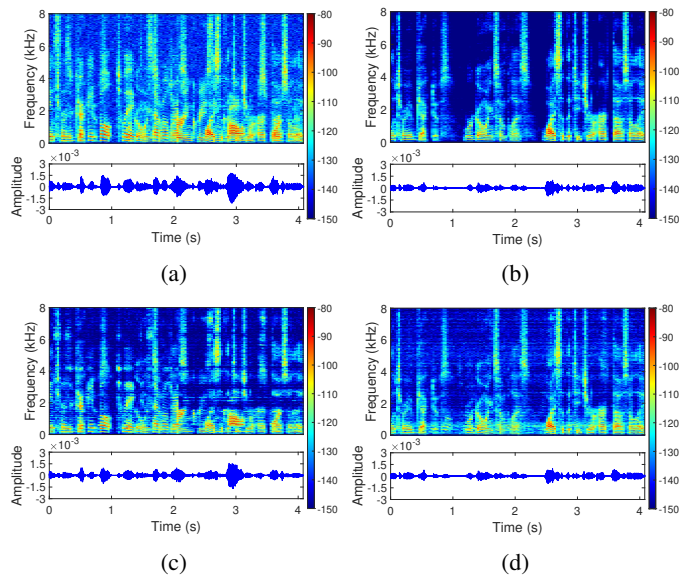


Fig. 2: Sonograms ([dB/Hz]) of the (a) signal received in the reference microphone, (b) desired signal received in the reference microphone, (c) MVDR output using the Euclidean approach, and (d) MVDR output using our approach.

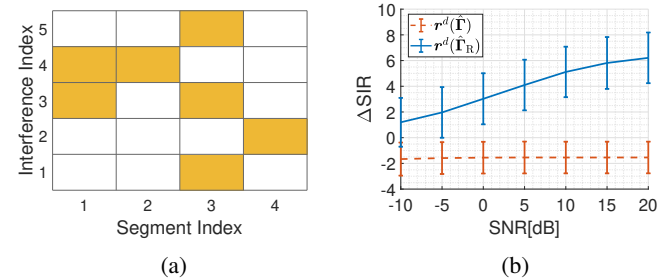


Fig. 3: (a) The activity pattern of the five interfering sources. (b) Same as Fig. 1b but for five interfering sources with SIR of -3 dB.

from the center of the array on the XY plane. The SIR is set to -3 dB. The signal duration is 8.192s, and it is divided into 4 segments. We note that the number of segments is smaller than the number of interfering sources. The randomly chosen activity pattern, fixed for all iterations, is shown in Fig. 3a. Figure 3b presents the results for different SNR values. We see that the proposed Riemannian approach leads to improved results in this setting as well.

REFERENCES

- [1] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2013.
- [2] P. Vary and R. Martin, *Digital Speech Transmission and Enhancement*, John Wiley & Sons, 2023.
- [3] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Trans. on Speech and Audio Process.*, vol. 11, no. 6, pp. 700–708, 2003.
- [4] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, "An overview of noise-robust automatic speech recognition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 4, pp. 745–777, 2014.
- [5] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*, John Wiley & Sons, 2018.

- [6] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 4, pp. 692–730, 2017.
- [7] 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, 2001.
- [8] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. on Speech and Audio Process.*, vol. 12, no. 5, pp. 451–459, 2004.
- [9] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 4, pp. 546–555, 2009.
- [10] O. Schwartz, S. Gannot, and E. A. Habets, "Multispeaker lmv beamformer and postfilter for source separation and noise reduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 5, pp. 940–951, 2017.
- [11] 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, 2009.
- [12] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a mimo acoustic signal processing perspective," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 3, pp. 1053–1065, 2007.
- [13] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech and Audio Process.*, vol. 13, no. 5, pp. 882–895, 2005.
- [14] O. Katz, R. R. Lederman, and R. Talmon, "Multimodal manifold learning using kernel interpolation along geodesic paths," *Information Fusion*, vol. 114, pp. 102637, 2025.
- [15] T. Shnitzer, H.-T. Wu, and R. Talmon, "Spatiotemporal analysis using riemannian composition of diffusion operators," *Applied and Computational Harmonic Analysis*, vol. 68, pp. 101583, 2024.
- [16] A. Bar and R. Talmon, "On interference-rejection using Riemannian geometry for direction of arrival estimation," *IEEE Trans. Signal Process.*, 2023.
- [17] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [18] L. Yang, M. Arnaudon, and F. Barbaresco, "Riemannian median, geometry of covariance matrices and radar target detection," in *IEEE Eur. Radar Conf.*, 2010, pp. 415–418.
- [19] B. Balaji and F. Barbaresco, "Application of Riemannian mean of covariance matrices to space-time adaptive processing," in *IEEE European Radar Conference*, 2012, pp. 50–53.
- [20] M. Arnaudon, F. Barbaresco, and L. Yang, "Riemannian medians and means with applications to radar signal processing," *IEEE J. Sel. Top. Signal Process.*, vol. 7, no. 4, pp. 595–604, 2013.
- [21] M. Coutino, R. Pribic, and G. Leus, "Direction of arrival estimation based on information geometry," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2016, pp. 3066–3070.
- [22] J. S. Picard, A. Bar, and R. Talmon, "Riemannian covariance fitting for direction-of-arrival estimation," *arXiv preprint arXiv:2404.03401*, 2024.
- [23] J. S. Picard, A. Bar, and R. Talmon, "Direct position determination by covariance-fitting on the Riemannian manifold of hermitian positive definite matrices," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2024, pp. 8521–8525.
- [24] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, pp. 127–154, 2006.
- [25] F. Hiai and D. Petz, "Riemannian metrics on positive definite matrices related to means," *Linear Algebra Appl.*, vol. 430, no. 11–12, pp. 3105–3130, 2009.
- [26] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomput.*, vol. 112, pp. 172–178, 2013.
- [27] J. Schmalenstroer, J. Heymann, L. Drude, C. Boeddecker, and R. Haeb-Umbach, "Multi-stage coherence drift based sampling rate synchronization for acoustic beamforming," in *IEEE Int. Workshop Multimedia Signal Process.*, 2017, pp. 1–6.
- [28] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," *Springer handbook of speech processing*, pp. 945–978, 2008.
- [29] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep.*, vol. 2, no. 2.4, pp. 1, 2006.
- [30] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [31] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," *Linguistic Data Consortium*, 1993.
- [32] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a kronecker product decomposition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, 2018.
- [33] X. Wang, G. Huang, I. Cohen, J. Benesty, and J. Chen, "Kronecker product adaptive beamforming for microphone arrays," in *IEEE APSIPA ASC*, 2021, pp. 49–54.
- [34] I.-T. Recommendation, "Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Rec. ITU-T P. 862*, 2001.
- [35] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.