Acoustic Source Localization Based on Geometric Projection in Reverberant and Noisy Environments

Tao Long, Jingdong Chen, *Senior Member, IEEE*, Gongping Huang, *Student Member, IEEE*, Jacob Benesty, and Israel Cohen, *Fellow, IEEE*

***Abstract*—Acoustic source localization (ASL) is a fundamental yet still challenging signal processing problem in sound acquisition, speech communication, and human–machine interfaces. Many ASL algorithms have been developed, such as the steered response power (SRP), the SRP-phase transform, the minimum variance distortionless response, the multiple signal classification (MUSIC), the householder transform-based methods, to name but a few. Most of those algorithms require hundreds or even thousands of snapshots to produce one reliable estimate, which make them difficult to track moving sources. Moreover, not much efforts have been reported in the literature to show the intrinsic relationships among those methods. This paper deals with the ASL problem with its focal point placed on how to achieve ASL with a short frame of acoustic signal (corresponding to a single snapshot in the frequency domain). It reformulates the ASL problem from the perspective of geometric projection. Four types of power functions are proposed, leading to several different algorithms for ASL. By analyzing those power functions, we show the equivalence between the popularly used conventional algorithms and our methods, which provides some new insights into the conventional algorithms. The relationships among different algorithms are discussed, which make it easy to comprehend the pros and cons of each of those methods. Experiments in real acoustic environments corroborate the theoretical analysis, which in turn justifies the contribution of this paper.**

***Index Terms*—Acoustic source localization, projection, steered response power, phase transform, householder transform, minimum variance distortionless response (MVDR), multiple signal classification (MUSIC).**

I. Introduction

Acoustic source localization (ASL) is a problem of estimating the position of radiating acoustic sources using temporal and spatial information provided by an array of microphones. It can either serve as an independent processor or as a first step for subsequent processing such as speech enhancement, beamforming, source separation in a large range of practical systems including smartspeakers, smart home systems, teleconferencing, camera surveillance, etc. [1]–[13]. Many methods have been developed over the last few decades, which can be divided into two main categories depending on how the source position is determined, i.e., two-step and single-step approaches. In the two-step approaches, the time-difference-of-arrival (TDOA) information among different pairs of microphone is first estimated. The position of a source of interest is then determined by virtue of triangulation or geometrical intersections [14]–[17]. In this class of techniques, the localization performance depends highly on the accuracy and robustness of TDOA estimation, which is challenging in the presence of strong reverberation and noise [18]. In comparison, the single-step approaches obtain directly the source location by searching the extremum value of a cost function defined from the microphone array outputs in either a two-dimensional (2D) or a three-dimensional (3D) grid.

In the single-step approaches, the most critical issue is the definition or selection of a cost function, which can produce an extremum (generally maximum) value at the grid coordinates corresponding to the source position. Many efforts have been devoted to this issue in the literature and a number of cost functions have been defined. The most popular one is the steered response power (SRP), which is essentially a delay-and-sum beamformer [19]. This method, however, suffers from a few drawbacks, the principal of which are its sensitivity to reverberation and low spatial resolution. An improved version of SRP based on the so-called phase transform (PHAT) is then developed, which is equivalent to forming the SRP cost function using the prewhitened array signals [20]–[24]. Both SRP and SRP-PHAT are fixed beamformers, which do not consider the characteristics or statistics of the application noise field and, therefore, are in general suboptimal in terms of localization performance. One way to improve the ASL performance is to take into consideration the noise statistics in constructing the cost function, leading to the minimum variance distortionless response (MVDR) based method [25], [26], which has higher resolution and better performance than the traditional SRP or

Manuscript received June 16, 2018; revised October 29, 2018; accepted November 29, 2018. Date of publication December 6, 2018; date of current version April 11, 2019. This work was supported in part by the NSFC fund for Young Scientists under Grant 61601370, in part by the NSFC Distinguished Young Scientists Fund under Grant 61425005, in part by the ISF-NSFC joint research program under Grants 2514/17 and 61761146001, and in part by Top International University Visiting Program for Outsanding Young scholars of Northwestern Polytechnical University. The Guest Editor coordinating the review of this manuscript and approving it for publication was Prof. Walter Kellermann. *(Corresponding author: Jingdong Chen.)*

T. Long and G. Huang are with the Center of Intelligent Acoustics and Immersive Communications and the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: longtao@nwpu.edu.cn; gongpinghuang@gmail.com).

J. Chen is with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jingdongchen@ieee.org).

J. Benesty is with the INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

I. Cohen is with the Andrew and Erna Viterby Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/JSTSP.2018.2885410

SRP-PHAT as long as the noise statistics are given or estimated accurately. Also, the multiple signal classification (MUSIC) method is another high resolution method which is based on the eigen subspace decomposition [27]–[29]. Finally, the Householder transform based technique was developed recently and it was shown to be more accurate and robust than the conventional methods if the algorithmic parameters are properly set [30], [31].

Although much effort has been devoted to it, ASL in practical environments remains a challenging problem. Particularly, the existing single-step ASL approaches generally require many (hundreds or even thousands of) signal samples to produce one reliable estimate, which make them not effective in dealing with moving sources. Also, not much work has been done to study the major difference and relationship among different algorithms, preventing us from comprehending the pros and cons of each algorithm. This paper is devoted to the ASL problem, with its focal point being on ASL with a short frame of signals, which corresponds to a single-snapshot source localization in the frequency domain. We present a single-step ASL theoretical framework from the perspective of geometric projection, based on which four types of narrowband power functions and three broadband fusion functions are proposed, which lead to several different ASL algorithms. The relationship and difference among different algorithms will be discussed.

The rest of this paper is organized as follows. In Section II, we present the signal model and the ASL problem formulation. Section III introduces the general concept of geometric projection and then discusses how to project the received signal vector onto a hypothesized steering vector for ASL using geometric projection. Section IV presents a frequency-domain single-snapshot ASL method based on the geometric projection, where we introduce four narrowband power functions and three different fusion methods for broadband signals. Section V deduces several ASL algorithms based on the four narrowband power functions, including the conventional SRP, SRP-PHAT, MVDR, MUSIC, and the Householder transform based techniques. We also study the relationships among different algorithms. In Section VI, we present some experiments to validate the theoretical analysis. Finally, some conclusions are drawn in Section VII.

II. Signal Model and Problem Formulation

Consider an array consisting of M omnidirectional microphones, placed in a certain geometry in a three dimensional space, as shown in Fig. 1. Let $\mathbf{r}_\mathrm{s} \in \mathbb{R}^3$ and $\mathbf{r}_m \in \mathbb{R}^3$ denote, respectively, the acoustic source location and the mth microphone location.

Let us first consider that the environment is free of reverberation and let us choose the first microphone as the reference. The output signal of the mth microphone at the time index t can then be expressed as [2]

$$\begin{aligned} y_m(t) &= x_m(t) + v_m(t) \\ &= x_1\left[t - \tau_{m1}(\mathbf{r}_\mathrm{s})\right] + v_m(t),\ m = 1, 2, \ldots, M, \end{aligned} \tag{1}$$

where $x_m(t)$ and $v_m(t)$ are, respectively, the (zero-mean) signal of interest and (zero-mean) additive noise at the mth microphone,

$$\tau_{m1}(\mathbf{r}_\mathrm{s}) \triangleq \frac{\|\mathbf{r}_\mathrm{s} - \mathbf{r}_m\| - \|\mathbf{r}_\mathrm{s} - \mathbf{r}_1\|}{c} \tag{2}$$

is the relative time delay between microphone m and 1, with $\|\cdot\|$ denoting the Euclidean norm, and c being the speed of sound in the air. Note that for clarity of exposition, we have neglected the propagation attenuation and multipath effects in the signal model in (1). But we will consider those effects in experiments in Section VI. If the reader is interested in the signal models that consider delay, attenuation, and multipath effects, please refer to [32].

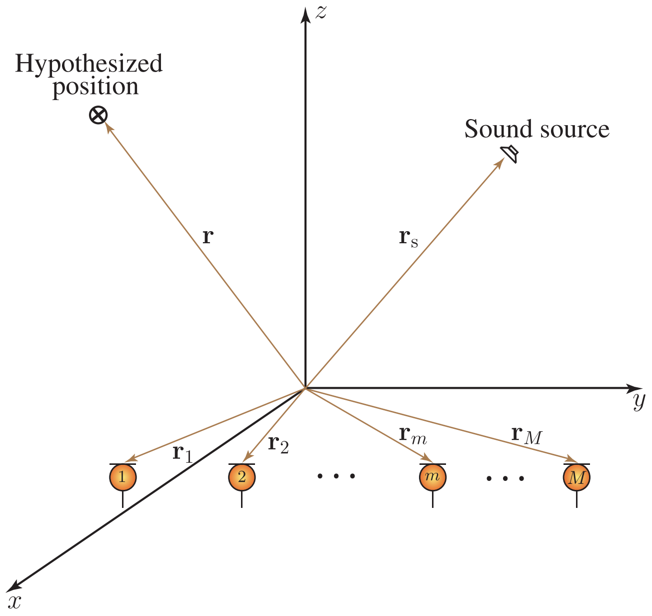

Fig. 1. Illustration of the ASL problem with microphone arrays, where $\mathbf{r}_m$, $\mathbf{r}_\mathrm{s}$, and $\mathbf{r}$ denote, respectively, the mth microphone location, the acoustic source location, and a hypothesized location.

In the frequency domain, the signal model given in (1) can be rewritten as

$$\begin{aligned} Y_m(f) &= X_m(f) + V_m(f) \\ &= e^{-\jmath 2\pi f \tau_{m1}(\mathbf{r}_s)} X_1(f) + V_m(f),\ m = 1, 2, \ldots, M, \end{aligned} \tag{3}$$

where f denotes the frequency, $Y_m(f)$, $X_m(f)$, and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $x_m(t)$, and $v_m(t)$, respectively, and $\jmath$ is the imaginary unit with $\jmath^2 = -1$.

Stacking the M frequency-domain microphone signals in a vector form, we obtain [32]

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f, \mathbf{r}_\mathrm{s}) X_1(f) + \mathbf{v}(f), \end{aligned} \tag{4}$$

where

$$\mathbf{y}(f) \triangleq \left[Y_1(f)\ Y_2(f)\ \cdots\ Y_M(f)\right]^T,$$

$$\mathbf{x}(f) \triangleq \left[X_1(f)\ X_2(f)\ \cdots\ X_M(f)\right]^T,$$

$$\mathbf{v}(f) \triangleq \left[V_1(f)\ V_2(f)\ \cdots\ V_M(f)\right]^T,$$

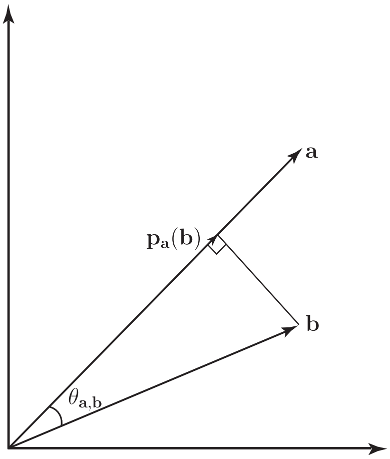

Fig. 2. Illustration of the projection of the vector $\mathbf{b}$ onto the vector $\mathbf{a}$ where $\mathbf{p_a}(\mathbf{b})$ is the projection vector and $\theta_{\mathbf{a},\mathbf{b}}$ is the angle between $\mathbf{b}$ and $\mathbf{a}$.

the superscript ^{T} is the transpose operator, and

$$\mathbf{d}\,(f,\mathbf{r}_\mathrm{s}) \triangleq \left[1 \; e^{-\jmath 2\pi f \tau_{21}(\mathbf{r}_\mathrm{s})} \; \cdots \; e^{-\jmath 2\pi f \tau_{M1}(\mathbf{r}_\mathrm{s})}\right]^T \tag{5}$$

is the signal propagation vector due to the source at $\mathbf{r}_\mathrm{s}$.

If we define a power function $\mathcal{P}(\mathbf{r})$ for any hypothesized position $\mathbf{r}$, the ASL problem is generally formulated as one of searching the maximum value of $\mathcal{P}(\mathbf{r})$ over a specified spatial grid, i.e.,

$$\widehat{\mathbf{r}}_\mathrm{s} = \arg\max_{\mathbf{r}} \mathcal{P}(\mathbf{r}). \tag{6}$$

III. Geometric Projection in High-Dimensional Space

In this section, we first describe the general concept of geometric projection in a high dimensional space, then discuss the geometric projection of a received signal vector onto a hypothesized steering vector for the ASL problem, which forms the basis for the presented ASL methods.

A. Notation and Definitions

Let $\mathbf{a} = [a_1 \; a_2 \; \cdots \; a_M]^T$ and $\mathbf{b} = [b_1 \; b_2 \; \cdots \; b_M]^T$ be two vectors in the M dimensional space $\mathbb{C}^M$, as shown in Fig. 2. The inner product of $\mathbf{b}$ and $\mathbf{a}$ is

$$\langle \mathbf{b}, \mathbf{a} \rangle \triangleq \sum_{m=1}^{M} b_m^* a_m = \mathbf{b}^H \mathbf{a}, \tag{7}$$

where superscript * and ^{H} denote complex conjugation and conjugate-transpose operator, respectively. The two vectors are orthogonal if their inner product is zero. The norm (length) of a vector is defined as

$$\|\mathbf{a}\| = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle} = \sqrt{\mathbf{a}^H \mathbf{a}}. \tag{8}$$

The projection of $\mathbf{b}$ onto $\mathbf{a}$, which is denoted by $\mathbf{p_a}(\mathbf{b})$, is defined as

$$\mathbf{p_a}(\mathbf{b}) \triangleq \frac{< \mathbf{b}, \mathbf{a} >}{\|\mathbf{a}\|^2}\mathbf{a} = \frac{\mathbf{b}^H \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}. \tag{9}$$

The angle between $\mathbf{b}$ and $\mathbf{a}$, i.e., $\theta_{\mathbf{a},\mathbf{b}}$, is defined as

$$\cos^2 \theta_{\mathbf{a},\mathbf{b}} \triangleq \frac{|< \mathbf{b}, \mathbf{a} >|^2}{\|\mathbf{b}\|^2 \, \|\mathbf{a}\|^2} = \frac{\left|\mathbf{b}^H \mathbf{a}\right|^2}{\|\mathbf{b}\|^2 \, \|\mathbf{a}\|^2} \tag{10}$$

and, obviously,

$$\begin{aligned} \|\mathbf{p_a}(\mathbf{b})\| &\leqslant \|\mathbf{b}\|, \\ \cos^2 \theta_{\mathbf{a},\mathbf{b}} &\leqslant 1, \end{aligned}$$

where equalities hold if and only if $\mathbf{b}$ and $\mathbf{a}$ are collinear.

B. Geometric Projection of the Received Signal Vector Onto a Hypothesized Steering Vector

Let $\mathbf{r}$ be a hypothesized position, we can define the steering vector corresponding to the hypothesized position as

$$\mathbf{d}\,(f,\mathbf{r}) \triangleq \left[1 \; e^{-\jmath 2\pi f \tau_{21}(\mathbf{r})} \; \cdots \; e^{-\jmath 2\pi f \tau_{M1}(\mathbf{r})}\right]^T. \tag{11}$$

To simplify the notation, in the rest of the paper, $\mathbf{y}\,(f)$, $\mathbf{x}\,(f)$, $\mathbf{v}\,(f)$, $\mathbf{d}\,(f,\mathbf{r}_\mathrm{s})$, and $\mathbf{d}\,(f,\mathbf{r})$ are written as $\mathbf{y}$, $\mathbf{x}$, $\mathbf{v}$, $\mathbf{d}_\mathrm{s}$, and $\mathbf{d}$, respectively.

The projection of the received signal vector $\mathbf{y}$ onto the hypothesized steering vector $\mathbf{d}$ is

$$\mathbf{p_d}(\mathbf{y}) = \frac{\mathbf{y}^H \mathbf{d}}{\|\mathbf{d}\|^2}\mathbf{d}. \tag{12}$$

Since $\|\mathbf{d}\|^2 = M$, the square norm of $\mathbf{p_d}(\mathbf{y})$ is

$$\begin{aligned} \|\mathbf{p_d}(\mathbf{y})\|^2 &= \frac{1}{\|\mathbf{d}\|^4}(\mathbf{d}\mathbf{y}^H \mathbf{d})^H (\mathbf{d}\mathbf{y}^H \mathbf{d}) \\ &= \frac{1}{\|\mathbf{d}\|^4}\mathbf{d}^H \mathbf{y} \mathbf{d}^H \mathbf{d} \mathbf{y}^H \mathbf{d} \\ &= \frac{1}{M}\mathbf{d}^H \mathbf{y}\mathbf{y}^H \mathbf{d}. \end{aligned} \tag{13}$$

From (9), (10) and (13), we also have

$$\cos^2 \theta_{\mathbf{d},\mathbf{y}} = \frac{\mathbf{d}^H \mathbf{y}\mathbf{y}^H \mathbf{d}}{M\,\|\mathbf{y}\|^2} = \frac{\|\mathbf{p_d}(\mathbf{y})\|^2}{\|\mathbf{y}\|^2}. \tag{14}$$

IV. Frequency-Domain Single Snapshot ASL Based on Geometric Projection

In this section, we first present the principle of ASL based on geometric projection. We then define four narrowband power functions and give some fusion methods for broadband ASL.

For the ASL problem, we make the following two assumptions.

- The clean signal $\mathbf{x}$ and the noise signal $\mathbf{v}$ are assumed to be independent.
- The noise signal $\mathbf{v}$ is spatially white with zero mean and covariance matrix $\sigma^2 \mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix.

Based on the aforementioned assumptions, we have

$$E\left[\|\mathbf{p_d}(\mathbf{y})\|^2\right] = E\left[\|\mathbf{p_d}(\mathbf{x})\|^2\right] + \sigma^2 \mathbf{I}_M, \tag{15}$$

where $E[\cdot]$ denotes mathematical expectation.

From (15), it can be found that the norm of the projection vector $\mathbf{p}_{\mathbf{d}}(\mathbf{y})$ increases as $\mathbf{d}$ approaches $\mathbf{d}_{\mathrm{s}}$ and $E\left[\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2\right]$ reaches its maximum when $\mathbf{r} = \mathbf{r}_{\mathrm{s}}$. As a result, the ASL problem can be written as

$$\widehat{\mathbf{r}}_{\mathrm{s}} = \arg\max_{\mathbf{r}} E\left[\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2\right]. \tag{16}$$

A. Narrowband Power Functions Based on the Geometric Projection

According to (6), a critical issue of ASL is to define a proper form of the power function $\mathcal{P}(\mathbf{r})$. Based on the geometric projection given in (16), we can now define the following narrowband power functions.

1) Power Function-I. $\mathcal{P}_1(\mathbf{r}, f)$: The first narrowband power function from $\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2$ is defined as

$$\mathcal{P}_1(\mathbf{r}, f) = M\,\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2 = \mathbf{d}^H \mathbf{y}\mathbf{y}^H \mathbf{d}. \tag{17}$$

2) Power Function-II. $\mathcal{P}_2(\mathbf{r}, f)$: Let us consider a normalized form of $\mathbf{y}$ as

$$\bar{\mathbf{y}} \triangleq \left[\frac{Y_1}{|Y_1|} \quad \frac{Y_2}{|Y_2|} \quad \cdots \quad \frac{Y_M}{|Y_M|}\right]^T = \mathbf{\Phi}\mathbf{y}, \tag{18}$$

where $\mathbf{\Phi} = \mathrm{diag}\left(\frac{1}{|Y_1|}, \quad \frac{1}{|Y_2|} \quad, \ldots, \quad \frac{1}{|Y_M|}\right)$ is an $M \times M$ diagonal matrix. Then, the second narrowband power function from $\|\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})\|^2$ can be defined as

$$\mathcal{P}_2(\mathbf{r}, f) = M\,\|\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})\|^2 = \mathbf{d}^H \mathbf{\Phi}\mathbf{y}\mathbf{y}^H \mathbf{\Phi}^H \mathbf{d}, \tag{19}$$

where $\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})$ is the projection of the normalized signal vector $\bar{\mathbf{y}}$ onto the hypothesized steering vector $\mathbf{d}$.

3) Power Function-III. $\mathcal{P}_3(\mathbf{r}, f)$: The third narrowband power function is defined as

$$\mathcal{P}_3(\mathbf{r}, f) = \|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^{\beta} = M^{-\frac{\beta}{2}}\left[\mathcal{P}_1(\mathbf{r}, f)\right]^{\frac{\beta}{2}}, \tag{20}$$

where β is a positive real number.

4) Power Function-IV. $\mathcal{P}_4(\mathbf{r}, f)$: Based on (14), it can also be checked that $\cos^2\theta_{\mathbf{d},\mathbf{y}}$ increases as $\mathbf{d}$ approaches $\mathbf{y}$. Consequently, we can define the fourth narrowband power function from $\cos^2\theta_{\mathbf{d},\mathbf{y}}$ as

$$\mathcal{P}_4(\mathbf{r}, f) = M\cos^2\theta_{\mathbf{d},\mathbf{y}} = \frac{\mathbf{d}^H\mathbf{y}\mathbf{y}^H\mathbf{d}}{\|\mathbf{y}\|^2} = \frac{\mathcal{P}_1(\mathbf{r}, f)}{\|\mathbf{y}\|^2}. \tag{21}$$

For the special case $|Y_1| = |Y_2| = \cdots = |Y_M| = \frac{1}{\sqrt{M}}\|\mathbf{y}\|$, we have

$$\begin{aligned}\mathcal{P}_2(\mathbf{r}, f) &= M\left\|\mathbf{p}_{\mathbf{d}}\left(\frac{\mathbf{y}}{|Y_1|}\right)\right\|^2 \\ &= \mathbf{d}^H\left(\frac{\mathbf{y}}{|Y_1|}\right)\left(\frac{\mathbf{y}}{|Y_1|}\right)^H\mathbf{d} \\ &= \frac{\mathbf{d}^H\mathbf{y}\mathbf{y}^H\mathbf{d}}{|Y_1|^2} = M\frac{\mathbf{d}^H\mathbf{y}\mathbf{y}^H\mathbf{d}}{\|\mathbf{y}\|^2} \\ &= M\mathcal{P}_4(\mathbf{r}, f).\end{aligned} \tag{22}$$

B. Fusion Methods for Broadband Sources

In the broadband situation, we need to fuse the narrowband power functions in IV-A across different frequency bands. We consider the following three types of fusion methods.

1) Arithmetic Fusion: The arithmetic fusion method simply sums the narrowband power functions across all the frequency bins in the frequency range of interest, i.e.,

$$\mathcal{P}_{\mathrm{A}}(\mathbf{r}) \triangleq \sum_{f=f_1}^{f_2}\mathcal{P}(\mathbf{r}, f), \tag{23}$$

where f_1 and f_2 are, respectively, the lower and upper cutoff frequencies of the frequency band of interest. From (15) and (17), one can expect that the ASL performance based on function-I using arithmetic fusion method is robust to white noise.

2) Geometric Fusion: The geometric fusion can be written as [26]

$$\mathcal{P}_{\mathrm{G}}(\mathbf{r}) \triangleq e^{\sum_{f=f_1}^{f_2}\ln\mathcal{P}(\mathbf{r},f)} = \prod_{f=f_1}^{f_2}\mathcal{P}(\mathbf{r}, f). \tag{24}$$

Let $\mathcal{P}_{n-\mathrm{G}}(\mathbf{r}), n = 1, 2, 3, 4$ denote the geometric fusion using the narrowband power functions I, II, III, IV, respectively. Then we can obtain

$$\begin{aligned}\mathcal{P}_{3-\mathrm{G}}(\mathbf{r}) &= \prod_{f=f_1}^{f_2} M^{-\frac{\beta}{2}}\left[\mathcal{P}_1(\mathbf{r}, f)\right]^{\frac{\beta}{2}} \\ &= M^{-\frac{\beta F}{2}}\left[\prod_{f=f_1}^{f_2}\mathcal{P}_1(\mathbf{r}, f)\right]^{\frac{\beta}{2}} \\ &= M^{-\frac{\beta F}{2}}\left[\mathcal{P}_{1-\mathrm{G}}(\mathbf{r})\right]^{\frac{\beta}{2}}\end{aligned} \tag{25}$$

where F is the total number of frequency bands, and

$$\begin{aligned}\mathcal{P}_{4-\mathrm{G}}(\mathbf{r}) &= \prod_{f=f_1}^{f_2}\frac{\mathcal{P}_1(\mathbf{r}, f)}{\|\mathbf{y}\|^2} = \left(\prod_{f=f_1}^{f_2}\frac{1}{\|\mathbf{y}\|^2}\right)\prod_{f=f_1}^{f_2}\mathcal{P}_1(\mathbf{r}, f) \\ &= \left(\prod_{f=f_1}^{f_2}\frac{1}{\|\mathbf{y}\|^2}\right)\mathcal{P}_{1-\mathrm{G}}(\mathbf{r}).\end{aligned} \tag{26}$$

From (25) and (26), one can see that the terms $M^{-\frac{\beta F}{2}}$ and $\prod_{f=f_1}^{f_2}\frac{1}{\|\mathbf{y}\|^2}$ are independent of $\mathbf{r}$. So, the geometric fusion functions defined from power functions I, III and IV, i.e., $\mathcal{P}_{1-\mathrm{G}}(\mathbf{r})$, $\mathcal{P}_{3-\mathrm{G}}(\mathbf{r})$ and $\mathcal{P}_{4-\mathrm{G}}(\mathbf{r})$ are equivalent up to a scale.

3) Normalized Fusion: The normalized fusion across different frequencies is

$$\mathcal{P}_{\mathrm{N}}(\mathbf{r}) \triangleq \sum_{f=f_1}^{f_2}\frac{\mathcal{P}(\mathbf{r}, f)}{\max_{\mathbf{r}}\mathcal{P}(\mathbf{r}, f)}. \tag{27}$$

Let $\mathcal{P}_{n-\mathrm{N}}(\mathbf{r}), n = 1, 2, 3, 4$ denote the normalized fusion using the narrowband power functions I, II, III, IV, respectively.

Then the normalized fusion $\mathcal{P}_{1-\mathrm{N}}$ is

$$\begin{aligned}\mathcal{P}_{1-\mathrm{N}}(\mathbf{r}, f) &= \sum_{f=f_1}^{f_2} \frac{\mathcal{P}_1(\mathbf{r}, f)}{\max_{\mathbf{r}} \mathcal{P}_1(\mathbf{r}, f)} \\ &= \sum_{f=f_1}^{f_2} \frac{\frac{\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2}{\|\mathbf{y}\|^2}}{\max_{\mathbf{r}} \frac{\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2}{\|\mathbf{y}\|^2}} \\ &= \sum_{f=f_1}^{f_2} \frac{\cos^2 \theta_{\mathbf{d},\mathbf{y}}}{\max_{\mathbf{r}} \cos^2 \theta_{\mathbf{d},\mathbf{y}}} \\ &= \sum_{f=f_1}^{f_2} \frac{\mathcal{P}_4(\mathbf{r}, f)}{\max_{\mathbf{r}} \mathcal{P}_4(\mathbf{r}, f)}. \qquad (28)\end{aligned}$$

The normalized fusion $\mathcal{P}_{4-\mathrm{N}}$ is

$$\mathcal{P}_{4-\mathrm{N}}(\mathbf{r}, f) = \sum_{f=f_1}^{f_2} \frac{\mathcal{P}_4(\mathbf{r}, f)}{\max_{\mathbf{r}} \mathcal{P}_4(\mathbf{r}, f)}. \qquad (29)$$

From (28) and (29), we see that the normalized power functions $\mathcal{P}_{1-\mathrm{N}}(\mathbf{r}, f)$ and $\mathcal{P}_{4-\mathrm{N}}(\mathbf{r}, f)$ are equivalent, both of which are based on $\mathcal{P}_4(\mathbf{r}, f)$.

C. Estimated Source Position

At last, based on the fusion power functions defined in IV-B, ASL can be achieved by searching the maximum of $\mathcal{P}_{\mathrm{Fusion}}$ (including $\mathcal{P}_{\mathrm{A}}$, $\mathcal{P}_{\mathrm{G}}$, and $\mathcal{P}_{\mathrm{N}}$) over a predefined spatial grid, i.e.,

$$\widehat{\mathbf{r}}_{\mathrm{s}} = \arg \max_{\mathbf{r}} \mathcal{P}_{\mathrm{Fusion}}(\mathbf{r}). \qquad (30)$$

V. ASL Algorithms Based on the Four Types of Power Functions from the Perspective of Geometric Projection

We presented four types of power functions from the perspective of geometric projection in Section IV. In this part, we will deduce different ASL algorithms based on those power functions, including: SRP, SRP-PHAT, Householder transform based methods, pseudo MUSIC, and pseudo MVDR. For each method, we will first demonstrate the power function, then explain its geometric meaning from the perspective of geometric projection. Note that we focus on the single snapshot (in the frequency domain) and the single source case in this section.

A. SRP

Power Function: The SRP can be computed by summing the cross-correlation function for all possible pairs of the set of microphones. The power function of the SRP method is given by

$$\mathcal{P}_{\mathrm{SRP}}(\mathbf{r}, f) = \sum_{m=1}^{M} \sum_{l=1}^{M} Y_m Y_l^* e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})}, \qquad (31)$$

where $\tau_{m\,l}(\mathbf{r}) = \tau_{m\,1}(\mathbf{r}) - \tau_{l\,1}(\mathbf{r})$.

Geometric Meaning: It can be shown that the SRP method is equivalent to the power function-I defined in (17), which is based on the projection of the received signal vector $\mathbf{y}$ onto the hypothesized steering vector $\mathbf{d}$. We have

$$\mathcal{P}_{\mathrm{SRP}}(\mathbf{r}, f) = \mathcal{P}_1(\mathbf{r}, f) = M\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2 = \mathbf{d}^H \mathbf{y}\mathbf{y}^H \mathbf{d}. \qquad (32)$$

Proof: Using the definitions of $\mathbf{d}$ and $\mathbf{y}$, we get the following result:

$$\begin{aligned}\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2 &= \frac{1}{M}\mathbf{d}^H \mathbf{y}\mathbf{y}^H \mathbf{d} \\ &= \frac{1}{M}\sum_{m=1}^{M} \sum_{l=1}^{M} Y_m Y_l^* e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})}. \qquad (33)\end{aligned}$$

It follows immediately that

$$\begin{aligned}\mathcal{P}_1(\mathbf{r}, f) &= M\|\mathbf{p}_{\mathbf{d}}(\mathbf{y})\|^2 \\ &= \sum_{m=1}^{M} \sum_{l=1}^{M} Y_m Y_l^* e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})} \\ &= \mathcal{P}_{\mathrm{SRP}}(\mathbf{r}, f), \qquad (34)\end{aligned}$$

which proves the equivalence between SRP and the power function-I.

B. SRP-PHAT

Power Function: The power function of the SRP-PHAT method is given by

$$\mathcal{P}_{\mathrm{SRP-PHAT}}(\mathbf{r}, f) = \sum_{m=1}^{M} \sum_{l=1}^{M} \frac{Y_m Y_l^*}{|Y_m Y_l^*|} e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})}. \qquad (35)$$

Geometric Meaning: It can be shown that the SRP-PHAT method is equivalent to the power function-II defined in (19), which is based on the projection of the normalized received signal $\bar{\mathbf{y}}$ onto the hypothesized steering vector $\mathbf{d}$. We have

$$\mathcal{P}_{\mathrm{SRP-PHAT}}(\mathbf{r}, f) = \mathcal{P}_2(\mathbf{r}, f) = M\|\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})\|^2. \qquad (36)$$

Proof: With the definitions of $\mathbf{d}$ and $\bar{\mathbf{y}}$, we have

$$\begin{aligned}\|\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})\|^2 &= \frac{1}{M}\mathbf{d}^H \bar{\mathbf{y}}\bar{\mathbf{y}}^H \mathbf{d} \\ &= \frac{1}{M}\mathbf{d}^H \mathbf{\Phi}\mathbf{y}\mathbf{y}^H \mathbf{\Phi}^H \mathbf{d} \\ &= \frac{1}{M}\sum_{m=1}^{M} \sum_{l=1}^{M} \frac{Y_m Y_l^*}{|Y_m Y_l^*|} e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})}. \qquad (37)\end{aligned}$$

Then, the power function-II is

$$\begin{aligned}\mathcal{P}_2(\mathbf{r}, f) &= M\|\mathbf{p}_{\mathbf{d}}(\bar{\mathbf{y}})\|^2 \\ &= \sum_{m=1}^{M} \sum_{l=1}^{M} \frac{Y_m Y_l^*}{|Y_m Y_l^*|} e^{\jmath 2\pi f \tau_{m\,l}(\mathbf{r})} \\ &= \mathcal{P}_{\mathrm{SRP-PHAT}}(\mathbf{r}, f), \qquad (38)\end{aligned}$$

which proves the equivalence between the SRP-PHAT and the power function-II.

C. Householder Transformation Based Method

Power Function: The Householder transformation can transform any given vector into a new vector, which is proportional to a unit vector. Mathematically, we define the Householder transformation matrix $\mathbf{T}(f, \mathbf{r})$ (written as $\mathbf{T}$) associated with the hypothesized steering vector $\mathbf{d}$ as [30]

$$\mathbf{T} \triangleq \mathbf{I}_M - \frac{2}{\mathbf{b}^H\mathbf{b}}\mathbf{b}\mathbf{b}^H, \tag{39}$$

where

$$\mathbf{b} \triangleq \mathbf{d} + \sqrt{M}\,\mathbf{i}_1 \tag{40}$$

and $\mathbf{i}_1 \triangleq \begin{bmatrix} 1 \ 0 \cdots 0 \end{bmatrix}^T$ is the first column of $\mathbf{I}_M$. It can be checked that $\mathbf{T}$ is Hermitian and unitary, i.e., $\mathbf{T}^H = \mathbf{T}$ and $\mathbf{T}\mathbf{T}^H = \mathbf{I}_M$. It is easy to verify that $\frac{-1}{\sqrt{M}}\mathbf{T}\mathbf{d} = \mathbf{i}_1$,

which means that the Householder transformation projects the vector $\mathbf{d}$ into a unit vector. Left-multiplying both sides of (4) by $-\mathbf{T}/\sqrt{M}$, we get

$$\frac{-1}{\sqrt{M}}\mathbf{T}\mathbf{y} = \mathbf{y}' = \begin{bmatrix} Y_1' \\ \mathbf{y}_2' \end{bmatrix} = \begin{bmatrix} X_1' \\ \mathbf{x}_2' \end{bmatrix} + \begin{bmatrix} V_1' \\ \mathbf{v}_2' \end{bmatrix}. \tag{41}$$

When $\mathbf{r} = \mathbf{r}_\mathrm{s}$, (41) becomes

$$\frac{-1}{\sqrt{M}}\mathbf{T}\mathbf{y} = \begin{bmatrix} Y_1' \\ \mathbf{y}_2{}' \end{bmatrix} = \begin{bmatrix} X_1 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} V_1' \\ \mathbf{v}_2{}' \end{bmatrix}. \tag{42}$$

We see how the Householder transformation yields a clear noise reference signal. Indeed, $Y_1{}' = X_1 + V_1{}'$ is the sum of the desired signal and noise, while the $(M-1)$-dimensional vector $\mathbf{y}_2{}' = \mathbf{v}_2{}'$ contains noise only if there is no reverberation.

From (41) and (42), the power function based on Householder transformation for any hypothesized location $\mathbf{r}$ can be defined as [31]

$$\mathcal{P}_\mathrm{HT}(\mathbf{r}, f) = |Y_1'|^\beta, \tag{43}$$

with β being a positive real number.

Geometric Meaning: It can be shown that the Householder transformation method is equivalent to the power function-III, which is also based on the projection of the received signal $\mathbf{y}$ onto hypothesized steering vector $\mathbf{d}$, i.e.,

$$\mathcal{P}_\mathrm{HT}(\mathbf{r}, f) = M^{-\frac{\beta}{2}}\,\|\mathbf{p}_\mathbf{d}(\mathbf{y})\|^\beta = M^{-\frac{\beta}{2}}\mathcal{P}_3(\mathbf{r}, f). \tag{44}$$

If $\beta = 2$, the Householder transformation method is equivalent to the SRP method up to a scale, i.e.,

$$\mathcal{P}_\mathrm{HT}(\mathbf{r}, f) = M^{-2}\mathcal{P}_\mathrm{SRP}(\mathbf{r}, f). \tag{45}$$

Proof: Figure 3 illustrates the geometric meaning based on the Householder transformation matrix $\mathbf{T}$ (a mirror reflection transform by a hyperplane $\mathbf{u}$), which reflects the hypothesized steering vector $\mathbf{d}$ onto the coordinate axis. As can be seen in Fig. 3, $\mathbf{d}$ is transformed to $\mathbf{d}''$, which lies in the $\mathbf{y}_1$ direction ($\mathbf{y}_1$ is an $M \times 1$ dimensional vector and $\mathbf{y}_1 = \begin{bmatrix} Y_1 \ 0 \cdots 0 \end{bmatrix}^T$), and $\mathbf{y}$ is transformed to $\mathbf{y}'' = \mathbf{T}\mathbf{y} \triangleq \begin{bmatrix} Y_1'' \ Y_2'' \cdots Y_M'' \end{bmatrix}^T$ without changing its norm. Based on the unitarity of the Householder transformation and the geometric relationship shown in Fig. 3, we can get

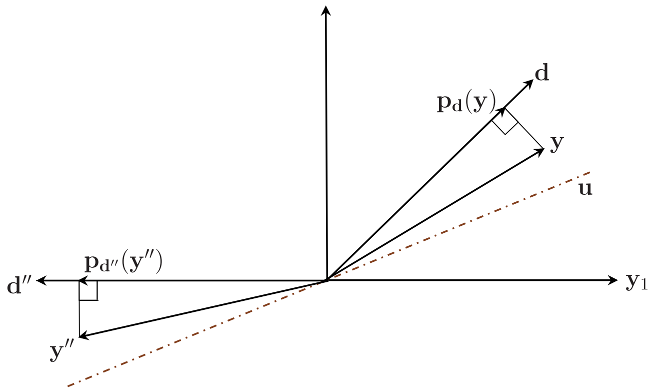
d
$\mathbf{p}_\mathbf{d}(\mathbf{y})$
y
u
$\mathbf{p}_{\mathbf{d}''}(\mathbf{y}'')$
$\mathbf{d}''$
$\mathbf{y}_1$
$\mathbf{y}''$

Fig. 3. Illustration of the Householder transformation from a projection perspective, where $\mathbf{T}$ is the Householder transformation matrix, $\mathbf{u}$ is the hyperplane associated with the vector $\mathbf{d}$. Based on the Householder transformation, $\mathbf{d}$ is transformed to $\mathbf{d}''$, and $\mathbf{y}$ is transformed to $\mathbf{y}''$.

$$|Y_1''| = \|\mathbf{p}_{\mathbf{d}''}(\mathbf{y}'')\| = \|\mathbf{p}_\mathbf{d}(\mathbf{y})\|, \tag{46}$$

where $\mathbf{p}_{\mathbf{d}''}(\mathbf{y}'')$ is the projection of $\mathbf{y}''$ onto $\mathbf{d}''$, and $\mathbf{p}_\mathbf{d}(\mathbf{y})$ is the projection of $\mathbf{y}$ onto $\mathbf{d}$.

From (43) and (46), we get

$$\begin{aligned} \mathcal{P}_\mathrm{HT}(\mathbf{r}, f) &= |Y_1'|^\beta = M^{-\frac{\beta}{2}}\,|Y_1''|^\beta \\ &= M^{-\frac{\beta}{2}}\,\|\mathbf{p}_{\mathbf{d}''}(\mathbf{y}'')\|^\beta \\ &= M^{-\frac{\beta}{2}}\,\|\mathbf{p}_\mathbf{d}(\mathbf{y})\|^\beta \\ &= M^{-\beta}\left[M\,\|\mathbf{p}_\mathbf{d}(\mathbf{y})\|^2\right]^{\frac{\beta}{2}}. \end{aligned} \tag{47}$$

From (17), (20) and (47), we obtain the relationship between $\mathcal{P}_\mathrm{HT}(\mathbf{r}, f)$ and power function-I and III:

$$\mathcal{P}_\mathrm{HT}(\mathbf{r}, f) = M^{-\beta}\left[\mathcal{P}_1(\mathbf{r}, f)\right]^{\frac{\beta}{2}} = M^{-\frac{\beta}{2}}\mathcal{P}_3(\mathbf{r}, f). \tag{48}$$

For $\beta = 2$, we have

$$\mathcal{P}_\mathrm{HT}(\mathbf{r}, f) = M^{-2}\mathcal{P}_1(\mathbf{r}, f) = M^{-2}\mathcal{P}_\mathrm{SRP}(\mathbf{r}, f), \tag{49}$$

which gives the relationship between the Householder transformation based method and the SRP method.

D. Pseudo MUSIC

Power Function: The MUSIC method, which is based on the eigen subspace decomposition, is popularly used to estimate DOA. Here, we can also define a power function for single snapshot ASL, similar to the MUSIC approach, leading to the pseudo MUSIC (PMUSIC).

Let us first apply the singular value decomposition to $\mathbf{y}\mathbf{y}^H$, i.e.,

$$\mathbf{y}\mathbf{y}^H = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H, \tag{50}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}\left(\sigma_1,\ 0,\ \ldots,\ 0\right)$ is an $M \times M$ diagonal matrix and $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \ \mathbf{u}_2 \cdots \mathbf{u}_M \end{bmatrix}$ is the $M \times M$ matrix consisting of the corresponding eigenvectors which are orthogonal with each other, with $\|\mathbf{u}_m\| = 1, m = 1, 2, \ldots, M$.

Then, we can define the power functions of PMUSIC as

$$\mathcal{P}_{\text{PMUSIC}-\text{S}}(\mathbf{r}, f) = \mathbf{d}^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{d} \tag{51}$$

or

$$\mathcal{P}_{\text{PMUSIC}-\text{N}}(\mathbf{r}, f) = \frac{1}{\mathbf{d}^H \mathbf{U}_{\text{N}} \mathbf{U}_{\text{N}}^H \mathbf{d}}, \tag{52}$$

where $\mathbf{U}_{\text{N}} = \left[\mathbf{u}_2 \ \mathbf{u}_3 \ \cdots \ \mathbf{u}_M \right]$.

Geometric Meaning: $\mathcal{P}_{\text{PMUSIC}-\text{S}}$ is based on the projection of the eigenvector $\mathbf{u}_1$ onto the hypothesized steering vector $\mathbf{d}$, while $\mathcal{P}_{\text{PMUSIC}-\text{N}}$ is based on the sum of the projections of the eigenvectors $\mathbf{u}_m$ onto the hypothesized steering vector $\mathbf{d}$. Both $\mathcal{P}_{\text{PMUSIC}-\text{S}}$ and $\mathcal{P}_{\text{PMUSIC}-\text{N}}$ are equivalent to the power function-IV defined in (21), which is based on the angle $\theta_{\mathbf{d},\mathbf{y}}$ between $\mathbf{y}$ and $\mathbf{d}$. Specifically,

$$\begin{aligned} \mathcal{P}_{\text{PMUSIC}-\text{S}}(\mathbf{r}, f) &= M \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_1) \right\|^2 = \mathcal{P}_4(\mathbf{r}, f), \\ \mathcal{P}_{\text{PMUSIC}-\text{N}}(\mathbf{r}, f) &= \frac{1}{M \sum_{i=2}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_i) \right\|^2} \\ &= \frac{1}{M - \mathcal{P}_4(\mathbf{r}, f)}. \end{aligned} \tag{53}$$

Proof: Let $\theta_{\mathbf{d},\mathbf{u}_m}$ be the angle between the eigenvector $\mathbf{u}_m$ and the hypothesized steering vector $\mathbf{d}$, and $\mathbf{p}_{\mathbf{d}}(\mathbf{u}_m)$ be the projection of $\mathbf{u}_m$ onto $\mathbf{d}$. Since $\mathbf{u}_m$, $m = 1, 2, \ldots, M$, are orthogonal with each other, it can be verified from (13) and (14) that

$$\sum_{m=1}^{M} \cos^2 \theta_{\mathbf{d},\mathbf{u}_m} = 1 \tag{54}$$

and

$$\sum_{m=1}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_m) \right\|^2 = \sum_{m=1}^{M} \left\| \mathbf{u}_m \right\|^2 \cos^2 \theta_{\mathbf{d},\mathbf{u}_m} = 1. \tag{55}$$

The first column of $\mathbf{U}$, denoted by $\mathbf{u}_1$, is the eigenvector corresponding to the maximum eigenvalue, and it can be verified that $\mathbf{u}_1$ is in the same direction as $\mathbf{y}$ and

$$\mathbf{u}_1 \mathbf{u}_1^H = \frac{\mathbf{y}\mathbf{y}^H}{\left\| \mathbf{y} \right\|^2}. \tag{56}$$

Let $\mathbf{p}_{\mathbf{d}}(\mathbf{u}_1)$ be the projection of the eigenvector $\mathbf{u}_1$ onto the hypothesized steering vector $\mathbf{d}$. We have

$$\begin{aligned} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_1) \right\|^2 &= \frac{1}{M} \mathbf{d}^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{d} = \frac{1}{M} \mathbf{d}^H \frac{\mathbf{y}\mathbf{y}^H}{\left\| \mathbf{y} \right\|^2} \mathbf{d} \\ &= \frac{\left\| \mathbf{p}_{\mathbf{d}}(\mathbf{y}) \right\|^2}{\left\| \mathbf{y} \right\|^2} = \cos^2 \theta_{\mathbf{d},\mathbf{y}}, \end{aligned} \tag{57}$$

where $\theta_{\mathbf{d},\mathbf{y}}$ is the angle between $\mathbf{y}$ and $\mathbf{d}$. So the PMUSIC power function from the eigenvector $\mathbf{u}_1$ is

$$\begin{aligned} \mathcal{P}_{\text{PMUSIC}-\text{S}}(\mathbf{r}, f) &= \mathbf{d}^H \mathbf{u}_1 \mathbf{u}_1^H \mathbf{d} = M \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_1) \right\|^2 \\ &= M \cos^2 \theta_{\mathbf{d},\mathbf{y}} = \mathcal{P}_4(\mathbf{r}, f). \end{aligned} \tag{58}$$

The projection of an eigenvector $\mathbf{u}_m$ onto the hypothesized steering vector $\mathbf{d}$ is

$$\mathbf{p}_{\mathbf{d}}(\mathbf{u}_m) = \frac{\mathbf{u}_m^H \mathbf{d}}{\left\| \mathbf{d} \right\|^2} \mathbf{d}. \tag{59}$$

Then, we have

$$\left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_m) \right\|^2 = \frac{1}{M} \mathbf{d}^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{d} \tag{60}$$

and

$$\sum_{i=2}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_i) \right\|^2 = \frac{1}{M} \mathbf{d}^H \mathbf{U}_{\text{N}} \mathbf{U}_{\text{N}}^H \mathbf{d}. \tag{61}$$

From (55) and (59), we also get

$$\sum_{m=1}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_m) \right\|^2 = \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_1) \right\|^2 + \sum_{i=2}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_i) \right\|^2 = 1. \tag{62}$$

So the PMUSIC power function from the eigenvectors $\mathbf{u}_m$ is

$$\begin{aligned} \mathcal{P}_{\text{PMUSIC}-\text{N}}(\mathbf{r}, f) &= \frac{1}{\mathbf{d}^H \mathbf{U}_{\text{N}} \mathbf{U}_{\text{N}}^H \mathbf{d}} \\ &= \frac{1}{M \sum_{i=2}^{M} \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_i) \right\|^2} \\ &= \frac{1}{M - M \left\| \mathbf{p}_{\mathbf{d}}(\mathbf{u}_1) \right\|^2} \\ &= \frac{1}{M - \mathcal{P}_4(\mathbf{r}, f)}. \end{aligned} \tag{63}$$

E. Pseudo MVDR

Power Function: The power function of the MVDR localization method is defined as the SRP with the MVDR beamformer:

$$\mathcal{P}_{\text{MVDR}-\text{M}}(\mathbf{r}, f) \triangleq \frac{1}{\mathbf{d}^H \mathbf{R}_{\mathbf{y}}^{-1} \mathbf{d}}, \tag{64}$$

where $\mathbf{R}_{\mathbf{y}} \triangleq \text{E}\left(\mathbf{y}\mathbf{y}^H \right)$ is the correlation matrix of $\mathbf{y}$. In real applications, diagonal loading may be needed if $\mathbf{R}_{\mathbf{y}}$ is ill-conditioned. Then, the power function of MVDR becomes [33]

$$\mathcal{P}_{\text{MVDR}-\text{M}}(\mathbf{r}, f) \triangleq \frac{1}{\mathbf{d}^H \left(\mathbf{R}_{\mathbf{y}} + \delta \mathbf{I}_M \right)^{-1} \mathbf{d}}, \tag{65}$$

where δ is the diagonal loading parameter whose value is relatively small as compared to the power of $\mathbf{y}$.

For the single snapshot case, we define the pseudo MVDR (PMVDR) power function as

$$\mathcal{P}_{\text{PMVDR}}(\mathbf{r}, f) \triangleq \frac{1}{\mathbf{d}^H (\mathbf{y}\mathbf{y}^H + \delta \mathbf{I}_M)^{-1} \mathbf{d}}. \tag{66}$$

Geometric Meaning: PMVDR is from the power function-IV defined in (21), which is based on the angle $\theta_{\mathbf{d},\mathbf{y}}$ between $\mathbf{y}$ and $\mathbf{d}$. It can be verified that

$$\mathcal{P}_{\text{PMVDR}}(\mathbf{r}, f) \approx \frac{\delta}{M - \mathcal{P}_4(\mathbf{r}, f)} = \delta \mathcal{P}_{\text{PMUSIC}-\text{N}}(\mathbf{r}, f). \tag{67}$$

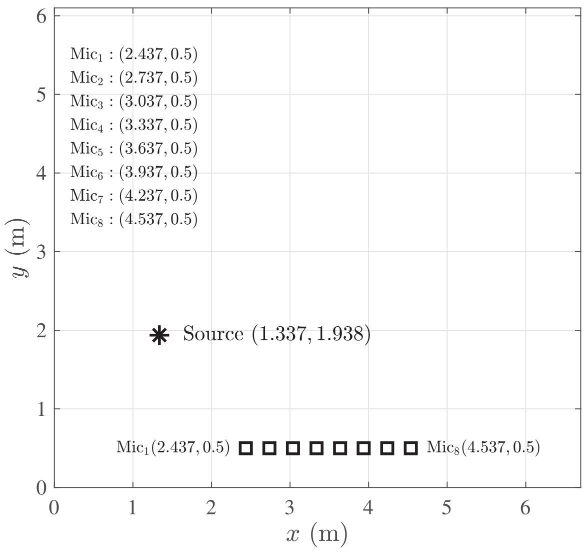

Fig. 4. Layout of the experimental setup in the varechoic chamber (unit is meter).

Proof: Using the Woodbury's identity, we get

$$\begin{aligned}&\left(\mathbf{y}\mathbf{y}^H + \delta\mathbf{I}_M\right)^{-1}\\&= \frac{1}{\delta}\mathbf{I}_M - \frac{1}{\delta}\mathbf{I}_M\mathbf{y}\left(1 + \mathbf{y}^H\frac{1}{\delta}\mathbf{I}_M\mathbf{y}\right)^{-1}\mathbf{y}^H\frac{1}{\delta}\mathbf{I}_M\\&= \frac{1}{\delta}\mathbf{I}_M - \frac{1}{\delta^2}\mathbf{y}\left(1 + \frac{1}{\delta}\left\|\mathbf{y}\right\|^2\right)^{-1}\mathbf{y}^H\\&\approx \frac{1}{\delta}\mathbf{I}_M - \frac{1}{\delta}\frac{\mathbf{y}\mathbf{y}^H}{\left\|\mathbf{y}\right\|^2},\end{aligned} \tag{68}$$

where the approximation is established because $\frac{1}{\delta}\left\|\mathbf{y}\right\|^2 \gg 1$.

Substituting (68) into (66), the power function of PMVDR is of the following form:

$$\begin{aligned}\mathcal{P}_{\mathrm{PMVDR}}(\mathbf{r}, f) &= \frac{1}{\mathbf{d}^H\left(\mathbf{y}\mathbf{y}^H + \delta\mathbf{I}_M\right)^{-1}\mathbf{d}}\\&\approx \frac{\delta}{\mathbf{d}^H\mathbf{d} - \frac{\mathbf{d}^H\mathbf{y}\mathbf{y}^H\mathbf{d}}{\left\|\mathbf{y}\right\|^2}}\\&= \frac{\delta}{M - \mathcal{P}_4(\mathbf{r}, f)}\\&= \delta\mathcal{P}_{\mathrm{PMUSIC-N}}(\mathbf{r}, f).\end{aligned} \tag{69}$$

VI. Experiments

A. Experimental Setup

The experiments are conducted with real impulse responses measured in the varechoic chamber at Bell Labs. The Bell Labs chamber is a rectangular room, which measures 6.7 m long by 6.1 m wide by 2.9 m high. The layout of the multichannel experimental setup is illustrated in Fig. 4, where a uniform linear array of 8 omnidirectional microphones is mounted 1.4 m ($z = 1.400$) above the floor and located, respectively, at (x, 0.500, 1.400), where $x = 2.437, 2.737, 3.037, 3.337, 3.637, 3.937, 4.237, 4.537$. For a detailed description of the varechoic chamber and how the reverberation time, T_{60}, is controlled, see [34], [35]. The acoustic channel impulse responses from the source to the eight microphones were measured at 48 kHz; but we downsample them to 16 kHz. To simulate a sound source, we place a loudspeaker at (1.337, 1.938, 1.600), playing back a clean speech signal. The clean speech signal is recorded in a quiet office room. The overall length of the signal is approximately 30-s long and sampled at 16 kHz. During the experiments, the microphone outputs are generated by convolving the source signal with the corresponding measured impulse responses and noise is then added to the convolved signals to control the SNR level.

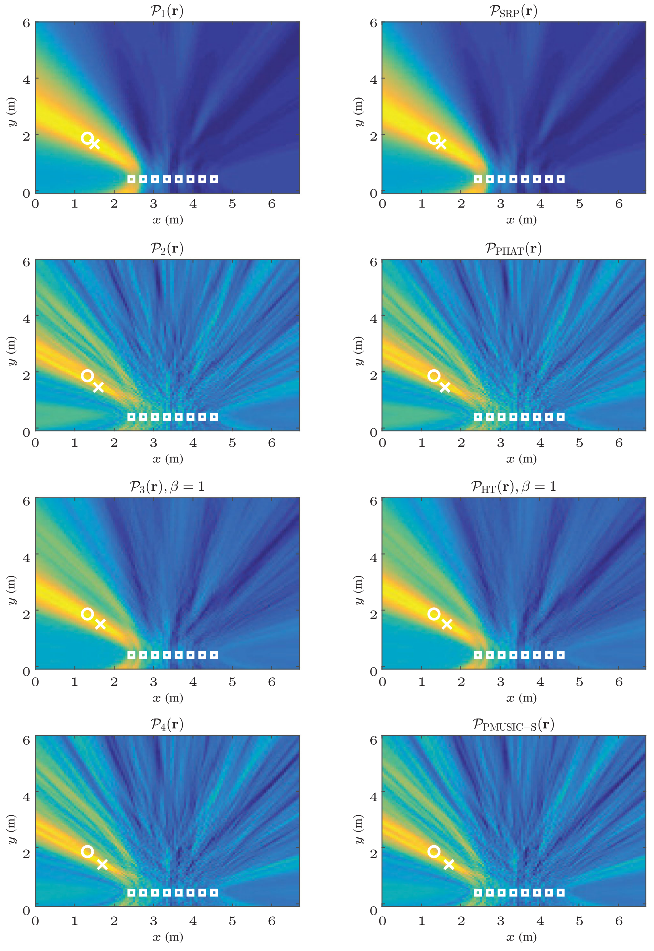

Fig. 5. Color map of the presented four narrowband power functions and conventional ASL algorithms in a 2D searching grid, where the white circle and the white cross represent, respectively, the true source and estimated positions. The true source position is $\mathbf{r}_s = (1.337, 1.938)$ m, the reverberation condition is $T_{60} = 380$ ms, and the background noise is white Gaussian noise with an input SNR of 10 dB.

In our experiment, ASL is carried out in the STFT domain. The array signals are partitioned into non-overlapping time frames of size 64 ms (1024 samples) and each frame is then transformed into the STFT domain using a 1024-point FFT. Two-dimensional ASL is performed every frame based on a single-snapshot method in the frequency domain. We compute

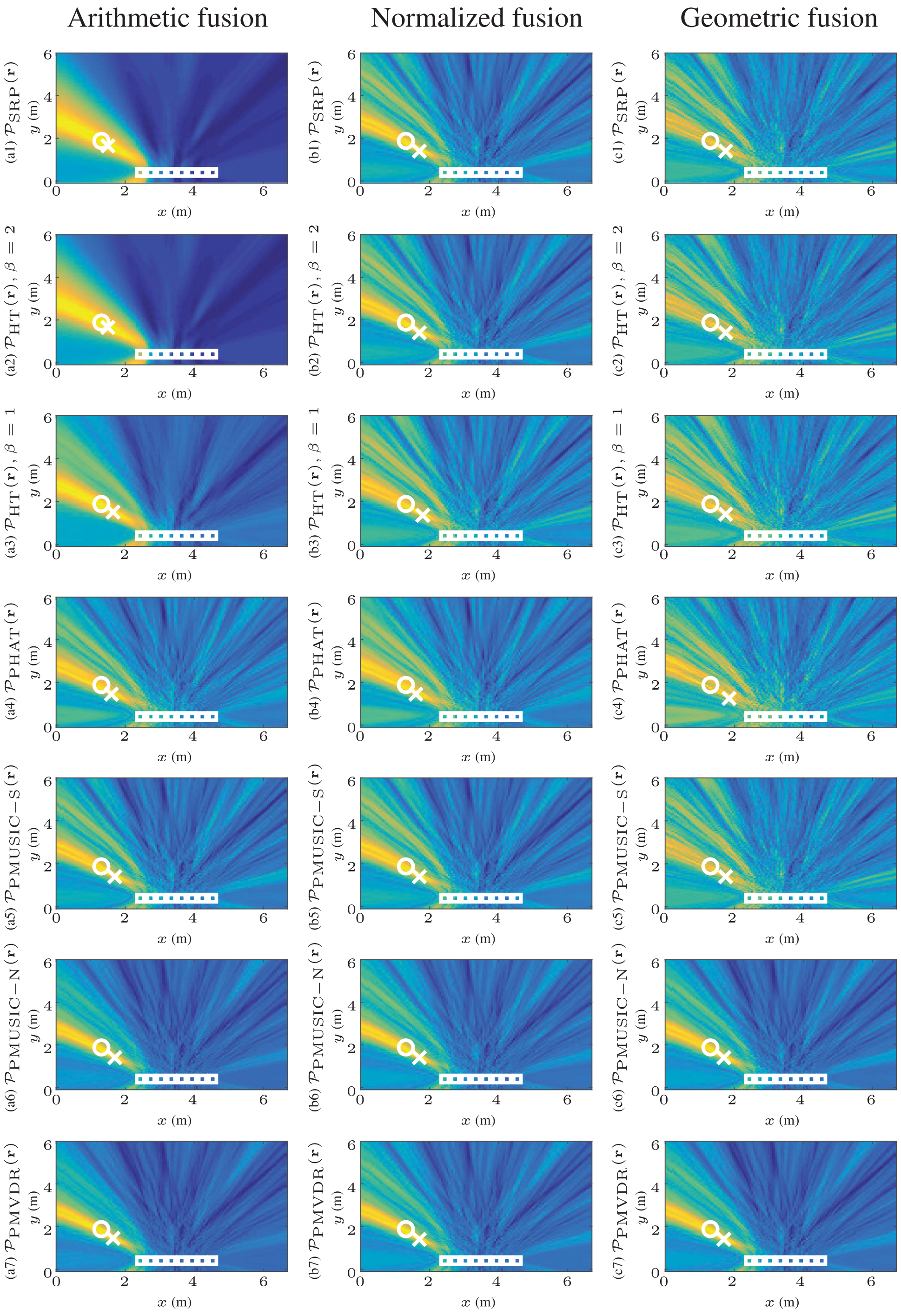

Fig. 6. Color map of the cost function plotted using the power function $\mathcal{P}(\mathbf{r})$ with different algorithms and different broadband fusion methods in the 2D searching grid for ASL, where the white circle and the white cross represent, respectively, the source and estimated locations. For each column: (a) Arithmetic fusion, (b) Normalized fusion, (c) Geometric fusion. For each row: (1) SRP, (2) SRP-PHAT, (3) HT ($\beta = 2$), (4) HT ($\beta = 1$), (5) PMUSIC-S, (6) PMUSIC-N, and (7) PMVDR ($\delta = 0.001$). The true source location is $\mathbf{r}_s = (1.337, 1.938)$ m, $T_{60} = 380$ ms, and the background noise is white Gaussian noise with an input SNR of 10 dB.

the power function $\mathcal{P}(\mathbf{r})$ using different algorithms and then search the maximum, thereby determining the source position.

B. Experimental Results

The first experiment verifies the equivalence of the proposed four kinds of power functions and the conventional ASL algorithms. We set $f_1 = 0$ Hz and $f_2 = 8000$ Hz as the lower and upper cutoff frequencies. Reverberation is configured so that the reverberation time T_{60} is 380 ms. The microphone signals are obtained by convolving the source signal with the corresponding measured impulse responses, and white Gaussian noise is then added with an input SNR of 10 dB. The arithmetic fusion method is used for all the conditions in this experiment.

We computed the power function for every position in the 2-D grid. The results are plotted in Fig. 5 as a color map. It can be observed that

- the power function-I based method is equivalent to the SRP algorithm,
- the power function-II based method is equivalent to the SRP-PHAT algorithm,
- the power function-III based method and the Householder transform algorithm are equivalent, and
- the power function-IV based method is equivalent to the PMUSIC-S algorithm.

The second experiment compares the ASL performances of seven algorithms with three broadband fusions. The experimental condition is the same as in the previous experiment. The results are plotted in Fig. 6, with three fusion methods: (a) arithmetic fusion, (b) normalized fusion, and (c) geometric fusion; and 7 power functions: (1) SRP, (2) SRP-PHAT, (3) HT ($\beta = 2$), (4) HT ($\beta = 1$), (5) PMUSIC-S, (6) PMUSIC-N, and (7) PMVDR ($\delta = 0.001$). Comparing different algorithms from the first column, we observe that

- the result of (a1) is the same as (a2), which corroborates that the SRP method is equivalent to the HT method with $\beta = 2$,
- the result of (a6) is the same as (a7), which corroborates that the PMUSIC-N method is equivalent to the PMVDR method, and
- the result of (a2) is different from that of (a3), which implies that the performance of the HT method depends on the value of the parameter β.

Comparing different broadband fusion methods, one can observe that

- the results of the geometric fusion with SRP, HT, and PMUSIC-S [as shown in (c1), (c2), (c3), and (c5)] are the same; this is obtained because the geometric power function-III and function-IV are equivalent to the geometric power function-I, and
- the results of the normalized fusion for different algorithms (as shown in the second column) are almost the same because all the normalized power functions are based on $\cos^2 \theta_{\mathbf{d},\mathbf{y}}$, where $\theta_{\mathbf{d},\mathbf{y}}$ is the angle between $\mathbf{y}$ and $\mathbf{d}$.

The third experiment evaluates the ASL performances of the four kinds of power functions in different reverberation and SNR conditions. Three different reverberation conditions are considered, i.e., $T_{60} = 240, 380, 580$ ms. Spatially white noise is considered with the input SNR changing from 0 dB to 20 dB with an increment of 2 dB. Arithmetic fusion method for broadband is used for all the conditions in this experiment. A total of 200 frames are used to compute the statistical performance results. The ASL performance measure used is the root mean square (RMS) error defined in a two-dimensional space, i.e.,

$$\mathrm{RMS} \triangleq \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left\{ \left[\widehat{x}_{\mathrm{s}}(n) - x_{\mathrm{s}}\right]^2 + \left[\widehat{y}_{\mathrm{s}}(n) - y_{\mathrm{s}}\right]^2 \right\}}, \tag{70}$$

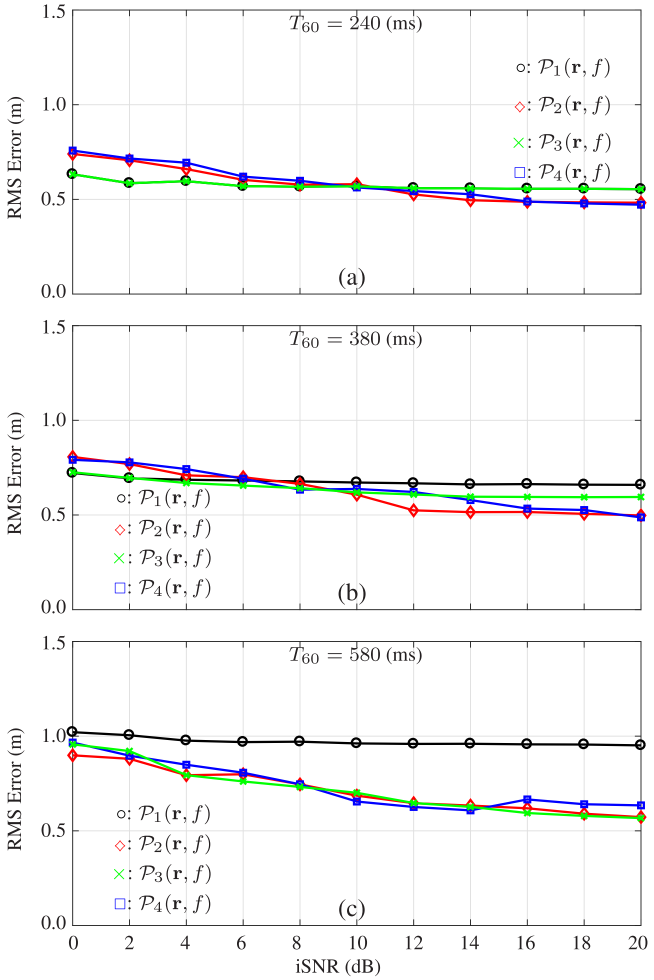

Fig. 7. RMS error with the four power functions versus the input SNR in white Gaussian noise case: (a) $T_{60} = 240$ ms, (b) $T_{60} = 380$ ms, and (c) $T_{60} = 580$ ms.

where N is the total number of analysis frames, $[\widehat{x}_{\mathrm{s}}(n), \widehat{y}_{\mathrm{s}}(n)]$ is the estimated source location for the n-th frame, and $(x_{\mathrm{s}}, y_{\mathrm{s}})$ is the true location.

We also define the percentage of the nonanomalous estimates as

$$\frac{N_{\mathrm{non}}}{N} \times 100\%, \tag{71}$$

where N_{non} is the number of nonanomalous ASL frames with the nonanomalous estimates being the ones that satisfy

$$\sqrt{(\widehat{x}_{\mathrm{s}} - x_{\mathrm{s}})^2 + (\widehat{y}_{\mathrm{s}} - y_{\mathrm{s}})^2} \leq 0.5 \text{ m}. \tag{72}$$

The RMS error and the percentage of the nonanomalous estimates as a function of the input SNR are plotted in Figs. 7 and 8, respectively. We make the following observations.

- The RMS error of the position estimates and the percentage of nonanomalous estimates of the power function-I based method do not change much with the input SNR. These results corroborate the analysis in Section IV-B that the ASL performance based on function-I using the arithmetic fusion method is robust to spatially white noise.

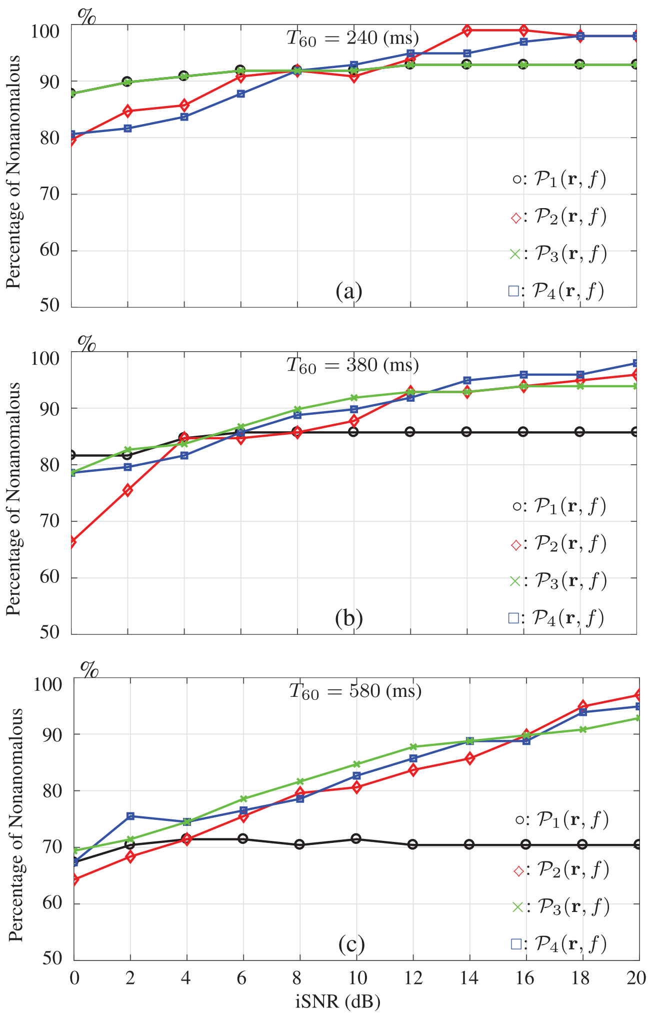

Fig. 8. Percentage of nonanomalous estimates with the four power functions versus the input SNR in white Gaussian noise case: (a) $T_{60} = 240$ ms, (b) $T_{60} = 380$ ms, and (c) $T_{60} = 580$ ms.

- The RMS error of the position estimates with the power function-II, III, and IV decreases with the increase of the input SNR, while their percentage of nonanomalous increases as the input SNR increases. This is reasonable. The higher the input SNR, the better is the TDOA performance.
- For the power function-III based methods, we can achieve better performance of ASL by adjusting the value of the parameter β to better fit the environment. (In this experiment, we set $\beta = 2.0$ for $T_{60} = 240$ ms, $\beta = 1.6$ for $T_{60} = 380$ ms, and $\beta = 0.8$ for $T_{60} = 580$ ms.)

VII. Conclusion

This paper deals with the problem of acoustic source localization. We presented a source localization framework from the perspective of geometric projection and four kinds of narrowband power functions and three fusion methods for broadband sources. We showed how most popularly used conventional algorithms could be cast into the presented framework based on the projection of the observation signal vector onto a hypothesized steering vector. Some new insights were presented as how different conventional algorithms are related to each other. Experiments in real acoustic environments corroborate the theoretical analysis.

References

[1] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Singapore: Wiley-IEEE, 2018.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Processing*. Berlin, Germany: Springer-Verlag, 2008.

[3] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "The ray space transform: A new framework for wave field processing," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5696–5706, Nov. 2016.

[4] F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources in the ray space for distributed microphone sensors," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Nov. 2017, pp. 170–174.

[5] M. Compagnoni *et al.*, "A geometrical–statistical approach to outlier removal for TDOA measurements," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3960–3975, Aug. 2017.

[6] Y. Dorfan, A. Plinge, G. Hazan, and S. Gannot, "Distributed expectation-maximization algorithm for speaker localization in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 3, pp. 682–695, Mar. 2018.

[7] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "Semi-supervised source localization on multiple manifolds with distributed microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 7, pp. 1477–1491, Jul. 2017.

[8] H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoust. Soc. Amer.*, vol. 120, pp. 2724–2736, Nov. 2006.

[9] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, May 2017.

[10] H. He, J. Chen, J. Benesty, Y. Zhou, and T. Yang, "Robust multichannel TDOA estimation for speaker localization using the impulsive characteristics of speech spectrum," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 2017, pp. 6130–6134.

[11] I. Merks, G. Enzner, and T. Zhang, "Sound source localization with binaural hearing aids using adaptive blind channel identification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2013, pp. 438–442.

[12] A. Marti, M. Cobos, and J. J. Lopez, "Real time speaker localization and detection system for camera steering in multiparticipant video conferencing environments," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2011, pp. 2592–2595.

[13] Y. Huang, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Jun. 2000, vol. 2, pp. II909–II912.

[14] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Trans. Speech, Audio Process.*, vol. 9, no. 8, pp. 943–956, Nov. 2001.

[15] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[16] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *J. Acoust. Soc. Amer.*, vol. 107, pp. 384–391, Jan. 2000.

[17] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Trans. Speech, Audio Process.*, vol. 5, no. 1, pp. 45–50, Jan. 1997.

[18] H. Do and H. F. Silverman, "A fast microphone array SRP-PHAT source location implementation using coarse-to-fine region contraction (CFRC)," in *Proc. IEEE Workshop Appl. Signal Process. Audio and Acoust.*, Oct. 2007, pp. 295–298.

[19] M. S. Bartlett, "Smoothing periodograms from time series with continuous spectra," *Nature*, vol. 161, no. 4096, pp. 686–687, 1948.

[20] J. H. DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*. Brown Univ. Providence, 2000.

[21] H. F. Silverman, Y. Yu, J. M. Sachar, and W. R. Patterson, "Performance of real-time source-location estimators for a large-aperture microphone array," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 593–606, Jul. 2005.

[22] H. Do, H. F. Silverman, and Y. Yu, "A real-time srp-phat source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, vol. 1, pp. I–121.

[23] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 5, pp. 499–508, Aug. 2004.

[24] K. D. Donohue, J. Hannemann, and H. G. Dietz, "Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments," *Signal Process.*, vol. 87, pp. 1677–1691, Jul. 2007.

[25] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[26] 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, Mar. 2014.

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

[28] M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigen structure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 4, pp. 817–827, Aug. 1984.

[29] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 823–831, Aug. 1985.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations. Third Edition*. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.

[31] G. Huang, J. Chen, and J. Benesty, "Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the householder transformation," *J. Acoust. Soc. Amer.*, vol. 138, pp. 3053–3060, Nov. 2015.

[32] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 170–170, 2006.

[33] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[34] W. C. Ward, G. Elko, R. Kubli, and W. McDougald, "The new varechoic chamber at at&t bell labs," in *Proc. Wallance Clement Sabine Centennial Symp.*, 1994, pp. 343–346.

[35] A. Härmä, "Acoustic measurement data from the varechoic chamber," *Technical Memorandum, Agere Systems*, 2001.

Jingdong Chen (M'99–SM'09) received the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences, Beijing, China, in 1998.

From 1998 to 1999, he was with the ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan, where he conducted research on speech synthesis, speech analysis, as well as objective measurements for evaluating speech synthesis. He then joined the Griffith University, Brisbane, Australia, where he engaged in research on robust speech recognition, and signal processing. From 2000 to 2001, he worked at ATR Spoken Language Translation Research Laboratories on robust speech recognition and speech enhancement. From 2001 to 2009, he was a Member of Technical Staff with Bell Laboratories, Murray Hill, NJ, USA, working on acoustic signal processing for telecommunications. He subsequently joined WeVoice, Inc., NJ, USA, as the Chief Scientist. He is currently a Professor with the Northwestern Polytechnical University, Xi'an, China. He authored and co-authored the books *Fundamentals of Signal Enhancement and Array Signal Processing* (Wiley, 2018), *Fundamentals of Differential Beamforming* (Springer, 2016), *A Conceptual Framework for Noise Reduction* (Springer, 2015), *Design of Circular Differential Microphone Arrays* (Springer, 2015), *Study and Design of Differential Microphone Arrays* (Springer, 2013), *Speech Enhancement in the STFT Domain* (Springer, 2011), *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study* (Springer, 2011), *Speech Enhancement in the Karhunen-Loève Expansion Domain* (Morgan&Claypool, 2011), *Noise Reduction in Speech Processing* (Springer, 2009), *Microphone Array Signal Processing* (Springer, 2008), and *Acoustic MIMO Signal Processing* (Springer, 2006). He is also a coeditor/coauthor of the book *Speech Enhancement* (Springer, 2005). His research interests include acoustic signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, microphone array signal processing, signal separation, and speech communication.

Dr. Chen was an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING from 2008 to 2014. He is currently a Technical Committee (TC) member of the IEEE Signal Processing Society (SPS) TC on Audio and Acoustic Signal Processing. He was the General Co-Chair of IWAENC 2016, the Technical Program Chair of the IEEE TENCON 2013, a Technical Program Co-Chair of the IEEE WASPAA 2009, IEEE ChinaSIP 2014, IEEE ICSPCC 2014, and IEEE ICSPCC 2015, and helped organize many other conferences. He was a recipient of the 2008 Best Paper Award from the IEEE Signal Processing Society (with Benesty, Huang, and Doclo), the best paper award from the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2011 (with Benesty), the Bell Labs Role Model Teamwork Award twice, respectively, in 2009 and 2007, the NASA Tech Brief Award twice, respectively, in 2010 and 2009, the Japan Trust International Research Grant from the Japan Key Technology Center in 1998, and the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998. He is also the co-author of a paper for which C. Pan received the IEEE R10 (Asia-Pacific Region) Distinguished Student Paper Award (First Prize) in 2016.

Tao Long received the Ph.D. degree in biomedical engineering from the Xi'an Jiaotong University, Xi'an, China, in 2014.

He is currently a Postdoctoral Researcher with the Northwestern Polytechnical University, Xi'an, China, and also a Visiting Researcher with the Israel Institute of Technology (Technion), Haifa, Israel. His research interests include acoustic source localization, noise reduction, speech enhancement, microphone array signal processing, and speech signal processing.

Gongping Huang (S'13) received the Bachelor's degree in electronics and information engineering from the Northwestern Polytechnical University (NPU), Xi'an, China, in 2012. He is currently working toward the Ph.D. degree at the Information and Communication Engineering, NPU.

He is a visiting Ph.D. student at INRS-EMT, University of Quebec, Quebec, Canada. His research interests include microphone array signal processing, noise reduction, speech enhancement, and audio and speech signal processing.

Jacob Benesty received the master degree in microwaves from Pierre & Marie Curie University, Paris, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, Orsay, France, in April 1991. During his Ph.D. (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms with the Centre National d'Etudes des Telecomunications (CNET), Paris, France. From January 1994 to July 1995, he worked with the Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff with Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, QB, Canada, as a Professor. He is also a Visiting Professor with the Technion—Israel Institute of Technology, Haifa, Israel, an Adjunct Professor with Aalborg University, Aalborg, Denmark, and a Guest Professor with Northwestern Polytechnical University, Xi'an, China. He is the inventor of many important technologies. In particular, he was the Lead Researcher with the Bell Labs, who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he conceived and designed the world-first PC-based multiparty hands-free full-duplex stereo conferencing system over IP networks. He has co-authored and co-edited/co-authored numerous books in the area of acoustic signal processing. His research interests include signal processing, acoustic signal processing, and multimedia communications.

Prof. Benesty is the Editor of the book series *Springer Topics in Signal Processing*. He was the General Chair and Technical Chair of many international conferences and a member of several IEEE technical committees. Four of his journal papers were awarded by the IEEE Signal processing Society and he was a recipient of the Gheorghe Cartianu Award from the Romanian Academy, in 2010.

Israel Cohen (M'01–SM'03–F'15) received the B.Sc. (*summa cum laude*), M.Sc., and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1990, 1993, and 1998, respectively. He is currently a Professor with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel. He is also a Visiting Professor with Northwestern Polytechnical University, Xi'an, China From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Ministry of Defense, Haifa, Israel. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT, USA. In 2001, he joined the Electrical Engineering Department, Technion—Israel Institute of Technology. His research interests are array processing, statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering. He is a Co-Editor of the Multichannel Speech Processing Section of the *Springer Handbook of Speech Processing* (Springer, 2008), and a co-author of *Fundamentals of Signal Enhancement and Array Signal Processing* (Wiley-IEEE Press, 2018).

Dr. Cohen was the recipient of the Norman Seiden Prize for Academic Excellence (2017), the SPS Signal Processing Letters Best Paper Award (2014), the Alexander Goldberg Prize for Excellence in Research (2010), and the Muriel and David Jacknow Award for Excellence in Teaching (2009). He is an Associate Member of the IEEE Audio and Acoustic Signal Processing Technical Committee, and a Distinguished Lecturer of the IEEE Signal Processing Society. He was an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, and a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee and the IEEE Speech and Language Processing Technical Committee.