Applied Acoustics 135 (2018) 85–90

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Direction of arrival estimation of multiple acoustic sources using a maximum likelihood method in the spherical harmonic domain

Yuxiang Hu[a], Jing Lu[a,*], Xiaojun Qiu[b]

[a] Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China
[b] Centre for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Multiple source DOA estimation
Maximum likelihood
Spherical harmonic domain
Alternating projection

ABSTRACT

Direction of arrival estimation (DOA) of multiple acoustic sources has been used for a wide range of applications, including room geometry inference, source separation and speech enhancement. The beamformer-based and subspace-based methods are most commonly used for spherical microphone arrays; however, the former suffers from spatial resolution limitations, while the later suffers from performance degradation in noisy environment. This letter proposes a multiple source DOA estimation approach based on the maximum likelihood method in the spherical harmonic domain and implements an efficient sequential iterative search of maxima on the cost function in the spherical harmonic domain. The proposed method avoids the division of the spherical Bessel function, which makes it suitable for both rigid-sphere and open-sphere configurations. Simulation results show that the proposed method has a significant superiority over the commonly used frequency smoothing multiple signal classification method. Experiments in a normal listening room and a reverberation room validate the effectiveness of the proposed method.

1. Introduction

The rotationally symmetric spatial directivity makes the spherical microphone array an appealing structure in many audio applications, among which the acoustic source localization, or the direction of arrival (DOA) estimation, plays an important role in speech enhancement [1], room impulse response analysis [2], and room geometry inference [3].

Various DOA estimation methods have been proposed, which can be generally classified as beamformer-based [2–5] and subspace-based [6–8]. The beamformer-based methods, such as those based on plane-wave decomposition (PWD) [4] and the minimum variance distortionless response (MVDR) beamformer [3], have the benefit of straightforward implementation, but suffer from low spatial resolution. The subspace-based methods, such as the multiple signal classification (MUSIC) [7], provide a high spatial resolution; however, they suffer from severe performance degradation when the signal-to-noise ratio (SNR) is low [6,9]. In order to improve the robustness of the DOA estimation of coherent sources, wideband expansion based on focusing matrices or frequency smoothing (FS) techniques has to be employed [8].

We proposed a maximum likelihood DOA estimation method in the spherical harmonic domain (SHMLE) recently, which is an attractive alternative DOA estimation method with advantages of high spatial resolution, strong robustness and straightforward wideband implementation [10]. The proposed SHMLE method only considered one source situation, while two or more sources often need to be localized in many practical applications. In this letter, the SHMLE method is extended to estimate the DOA of multiple sources. Generally speaking, the DOAs can be determined by searching maxima on the maximum likelihood (ML) cost function. However, the commonly used grid search method is only effective in finding the global maximum, which restricts its applicability in one source situation. To achieve effective DOA estimation of multiple sources, an efficient sequential iterative search method is introduced in the spherical harmonic (SH) domain. Experiments using a 32-element spherical microphone array validate the feasibility and superiority of the proposed method.

2. Method

2.1. Signal model in the spherical harmonic domain

The standard spherical coordinate system is utilized with r, θ and ϕ representing the radius, the elevation angle and the azimuth, respectively. The sound field is assumed to be composed of plane waves from L sources with $\Psi_l = (\theta_l, \phi_l)$ ($l = 1, 2, \ldots, L$) being the DOA of the l-th plane wave and $s_l(k)$ being its amplitude, where k denotes the wave

* Corresponding author.
E-mail address: lujing@nju.edu.cn (J. Lu).

https://doi.org/10.1016/j.apacoust.2018.02.005
Received 27 September 2017; Received in revised form 17 January 2018; Accepted 7 February 2018

number. The Q element spherical microphone array is distributed uniformly on a sphere with a radius of a centred at the origin of the coordinate system, and $\Omega_q = (\theta_q, \phi_q)$ is the angle position of the q-th microphone [11].

The sound pressure of the q-th microphone for the incident waves can be expressed as [12]

$$p(k,\Omega_q) = \sum_{l=1}^{L} s_l(k)e^{i\mathbf{k}_l^T\mathbf{r}_q} \approx \sum_{l=1}^{L} s_l(k)\sum_{n=0}^{N}\sum_{m=-n}^{n} b_n(k)Y_{n,m}^*(\Psi_l)Y_{n,m}(\Omega_q), \tag{1}$$

where $\mathbf{k}_l = -k(\cos\phi_l\sin\theta_l,\ \sin\phi_l\sin\theta_l,\ \cos\theta_l)^T$ and $\mathbf{r}_q = a(\cos\phi_q\sin\theta_q,\ \sin\phi_q\sin\theta_q,\ \cos\theta_q)^T$ denote the wave vector of the plane wave and position of the q-th microphone in the Cartesian coordinate. $Y_{n,m}$ is the spherical harmonic of order n and degree m, N is the highest order number for the plane wave decomposition and satisfies $(N+1)^2 < Q$. The superscript $(*)$ denotes complex conjugation. $b_n(k)$ is a function of array configuration [11]. Eq. (1) can be expressed in matrix form as

$$p(k,\Omega_q) \approx \mathbf{y}^T(\Omega_q)\mathbf{B}(k)\mathbf{Y}^H(\mathbf{\Psi})\mathbf{s}(k), \tag{2}$$

with

$$\mathbf{y}(\Omega_q) = [Y_{0,0}(\Omega_q),\ Y_{1,-1}(\Omega_q),\ Y_{1,0}(\Omega_q),\ Y_{1,1}(\Omega_q),\ ...,\ Y_{N,N}(\Omega_q)]^T, \tag{3}$$

$$\mathbf{y}(\Psi_l) = [Y_{0,0}(\Psi_l),\ Y_{1,-1}(\Psi_l),\ Y_{1,0}(\Psi_l),\ Y_{1,1}(\Psi_l),\ ...,\ Y_{N,N}(\Psi_l)]^T, \tag{4}$$

$$\mathbf{Y}(\mathbf{\Psi}) = [\mathbf{y}(\Psi_1), \mathbf{y}(\Psi_2), ..., \mathbf{y}(\Psi_L)\ ]^T, \tag{5}$$

$$\mathbf{B}(k) = diag\{b_0(k), b_1(k), b_1(k)\ , b_1(k), ..., b_N(k)\}, \tag{6}$$

$$\mathbf{s}(k) = [s_1(k), s_2(k), ..., s_L(k)]^T, \tag{7}$$

$$\mathbf{\Psi} = [\Psi_1, \Psi_2, ..., \Psi_L], \tag{8}$$

where the superscript $(^T)$ denotes the transpose.

In the presence of additive noise, the sound pressure at all Q microphones can be expressed as

$$\mathbf{p}(k,\mathbf{\Omega}) \approx \mathbf{Y}(\mathbf{\Omega})\mathbf{B}(k)\mathbf{Y}^H(\mathbf{\Psi})\mathbf{s}(k) + \boldsymbol{\nu}(k), \tag{9}$$

where

$$\mathbf{Y}(\mathbf{\Omega}) = [\mathbf{y}(\Omega_1), \mathbf{y}(\Omega_2), ..., \mathbf{y}(\Omega_Q)]^T, \tag{10}$$

$\mathbf{p}(k,\mathbf{\Omega}) = [p(k,\Omega_1),\ p(k,\Omega_2),\ ...,\ p(k,\Omega_Q)]^T$ is the vector of the sound pressure of Q microphones, and $\boldsymbol{\nu}(k) = [\nu_1(k),\ \nu_2(k),\ ...,\ \nu_Q(k)]^T$ is the additive sensor noise added to the system. The noise is assumed to be complex Gaussian, to be uncorrelated with the signal, to have zero mean, and for simplicity, to be spatially white with a covariance matrix $\mathbf{R}_{\boldsymbol{\nu}}(k) = \sigma_\nu^2\mathbf{I}_Q$, where σ_ν^2 is the unknown noise variance and $\mathbf{I}_Q$ is the identity matrix of order $Q \times Q$.

For the uniformly spatial sampling configuration used in this letter, the following orthogonal relation holds (note that $(N+1)^2 \le Q$) [11]

$$\frac{4\pi}{Q}\mathbf{Y}^H(\mathbf{\Omega})\mathbf{Y}(\mathbf{\Omega}) = \mathbf{I}_{(N+1)^2}. \tag{11}$$

The SH transform can be carried out by multiplying both sides of Eq. (9) from the left by $\frac{4\pi}{Q}\mathbf{Y}^H(\mathbf{\Omega})$, which yields

$$\mathbf{p_{nm}}(k) \approx \mathbf{B}(k)\mathbf{Y}^H(\mathbf{\Psi})\mathbf{s}(k) + \boldsymbol{\nu}_{\mathbf{nm}}(k), \tag{12}$$

where $\mathbf{p_{nm}}(k)$ is a vector containing $(N+1)^2$ SH domain coefficients, i.e.,

$$\mathbf{p_{nm}}(k) = [p_{0,0}(k),\ p_{1,-1}(k),\ p_{1,0}(k),\ p_{1,1}(k),\ ...,\ p_{N,N}(k)]^T. \tag{13}$$

The second term on the right side of Eq. (12) is the noise expressed in the SH domain, i.e. $\boldsymbol{\nu}_{\mathbf{nm}}(k) = \frac{4\pi}{Q}\mathbf{Y}^H(\Omega)\boldsymbol{\nu}(k)$, with the mean

$$E[\boldsymbol{\nu}_{\mathbf{nm}}(k)] = \frac{4\pi}{Q}\mathbf{Y}^H(\mathbf{\Omega})E[\boldsymbol{\nu}(k)] = \mathbf{0}, \tag{14}$$

and the covariance matrix

$$\mathbf{R_{nm}}(k) = E\left[\frac{4\pi}{Q}\mathbf{Y}^H(\mathbf{\Omega})\boldsymbol{\nu}(k)\boldsymbol{\nu}^H(k)\mathbf{Y}(\mathbf{\Omega})\frac{4\pi}{Q}\right] = \frac{4\pi}{Q}\cdot\sigma_\nu^2\mathbf{I}_{(N+1)^2}, \tag{15}$$

where $E(\cdot)$ denotes the statistical expectation. Apparently, the noise model in the SH domain is also zero-mean complex Gaussian.

2.2. Sound source DOA estimation in the spherical harmonic domain

Define $\mathbf{\Theta} = [\mathbf{\Psi}^T, \mathbf{S}^T, \sigma_n^2]^T$ as the vector of all the unknown parameters, where $\mathbf{S} = [\mathbf{s}(k_{\min})^T, ..., \mathbf{s}(k_{\max})^T\ ]^T$ contains the amplitudes of the source signals with k_{min} and k_{max} representing the minimum and maximum wave numbers and satisfying $ka \le N$. Throughout this paper, $\mathbf{\Psi}$, s and σ_ν^2 are assumed to be deterministic and unknown, while the observed data $\mathbf{p_{nm}}$ is considered random [13]. The likelihood function of $\mathbf{p_{nm}}$ given $\mathbf{\Theta}$ in the SH domain can be expressed as [9,13]

$$f(\mathbf{p_{nm}};\mathbf{\Theta}) = \frac{\exp\left\{-\sum_{k=k_{min}}^{k_{max}}[\mathbf{p_{nm}}(k)-\mathbf{V_{nm}}(k,\mathbf{\Psi})\mathbf{s}(k)]^H\mathbf{R_{nm}^{-1}}[\mathbf{p_{nm}}(k)-\mathbf{V_{nm}}(k,\mathbf{\Psi})\mathbf{s}(k)]\right\}}{(\pi^{(N+1)^2}|\mathbf{R_{nm}}|)^{k_{max}-k_{min}}}, \tag{16}$$

where $\mathbf{V_{nm}}(k,\Psi) = \mathbf{B}(k)\mathbf{Y}^H(\mathbf{\Psi})$ and $|\cdot|$ denotes the matrix determinant. The solution to Eq. (16) is given by [10]

$$\hat{\mathbf{\Psi}} = \operatorname{argmin}_{\mathbf{\Psi}} \sum_{k=k_{min}}^{k_{max}} \|\mathbf{p_{nm}}(k)-\mathbf{V_{nm}}(k,\mathbf{\Psi})\mathbf{V_{nm}}(k,\mathbf{\Psi})^\dagger\mathbf{p_{nm}}(k)\|^2, \tag{17}$$

where $(\cdot)^\dagger$denotes pseudo-inverse operation.

Define the cost function as

$$J(\mathbf{\Psi}) = -10\log_{10}\left(\sum_{k=k_{min}}^{k_{max}} \|\mathbf{p_{nm}}(k)-\mathbf{V_{nm}}(k,\mathbf{\Psi})\mathbf{V_{nm}}(k,\mathbf{\Psi})^\dagger\mathbf{p_{nm}}(k)\|^2\right), \tag{18}$$

then the wideband estimator can be described as

$$\hat{\mathbf{\Psi}} = \underset{\mathbf{\Psi}}{\operatorname{argmax}}\, J(\mathbf{\Psi}). \tag{19}$$

The SHMLE has the remarkable benefit of easy wideband implementation as described in Eqs. (17–19). This is superior over the other methods in the spherical harmonic domain, which usually require a quite cumbersome frequency smoothing (FS) technique to realize wideband DOA [8]. Compared with the maximum likelihood method in Ref. [14], the division of $b_n(k)$ is avoided, which makes the method proposed in this letter suitable for both rigid-sphere and open-sphere arrays. Note that for open-sphere arrays, $b_n(k)$ is close to 0 at frequencies corresponding to the zeros of the spherical Bessel functions.

2.3. DOA estimation of multiple sources

For one source situation, Eq. (19) can be solved using the grid search method. For P grid points and L sources situation, the computational load of Eq. (19) is $O(P^L)$, which is computationally prohibitive. Moreover, effective discrimination of the multiple maxima in the cost function is very difficult even if repetitive traversal is feasible. To alleviate these problems, a nonlinear optimization algorithm is applied in the SH domain with implementation of the alternating projection method [15]. The alternating projection method avoids the multidimensional search by estimating the location of one source sequentially while fixing the estimates of other source locations from the previous iteration.

For nonlinear optimization methods, the initial locations of the sound sources is critical to reach the global maximum. In this letter, the simplified grid search method is adopted to find initial locations, and the procedure of the method is described as follows.

(1) Estimate the location of the first source s_1 on a single source grid with