Robust Stochastic Maximum Likelihood Algorithm for DOA Estimation of Acoustic Sources in the Spherical Harmonic Domain

Hossein Lolaee
School of Electrical and Computer Eng.
College of Eng., University of Tehran
Tehran, Iran
hossein_lolaee@ut.ac.ir

Mohammad Ali Akhaee
School of Electrical and Computer Eng.
College of Eng., University of Tehran
Tehran, Iran
akhaee@ut.ac.ir

***Abstract*—The direction of arrival (DOA) estimation of the sound sources has been a popular signal processing research topic due to its widespread applications. Using spherical microphone array, DOA estimation can be applied in the spherical harmonic (SH) domain without any spatial ambiguity. However, the environment reverberation and noise can degrade the estimation performance. In this paper, we propose a novel iterative stochastic maximum likelihood (ML) algorithm for DOA estimation of multiple sound sources in the presence of spatially nonuniform noise in the SH domain. The main idea of the proposed algorithm is considering the general model of the received signal in the SH domain. We reduce the complexity of the ML estimation by breaking it down to two separate problems: noise parameters and DOA estimation problems. Simulation results indicate that the proposed algorithm improves the robustness of estimation, i.e, the root mean square error, by at least 7 dB compared to the recent methods in the reverberant and noisy environments.**

***Index Terms*—Direction of Arrival Estimation, Spherical Microphone Array, Spherical Harmonics**

I. Introduction

The direction of arrival (DOA) estimation of sound sources has been a popular signal processing research topic due to its widespread applications, including speech enhancement, dereverbration, and robot audition. The spherical microphone arrays have attracted more attention recently. The spatial symmetry of spherical arrays helps us to capture the 3-D information of sound sources without spatial ambiguity. The main advantage of analysis in the spherical harmonic (SH) domain is the decoupling of frequency-dependent and angular-dependent components [1].

Traditional DOA estimation methods can be divided into three categories: time-delay [2], beamforming [3], and subspace based methods [4], [5]; one of the famous and popular methods of the third category is the multiple signal classification (MUSIC) [4] and estimation of signal parameters via rotational invariance techniques (ESPRIT) [5]. Sound source reverberation causes correlated and coherent acoustic signals which degrades the performance of the traditional methods specially spectral based ones. Although in [6], [7], MUSIC and ESPRIT are applied in the SH domain, they lost accuracy in high reverberation.

A DOA estimation method is proposed based on independent component analysis (ICA) by using directional sparsity of sound sources in the series of [8]–[10]. In [8], the unmixing matrix was extracted by applying the ICA model to the SH domain signals and then DOA estimated by comparing its columns with the dictionary of possible plane-wave source directions steering vectors. Since this method suffers from low resolution, by combining ICA and sparse recovery methods its performance can be improved [9]. In [10], authors improve the convergence of sparse recovery solver by exploiting spatial location of the sound sources as a prime information. In all of these methods, DOA estimation strongly degrades in reverberant and noisy environment because the authors do not consider the noise in the their methods.

In this paper, first the received signal model is investigated in the SH domain. Then considering the general model of the received signal in the SH domain, an iterative stochastic maximum likelihood (ML) of DOA estimation is proposed for multiple sources in the presence of spatially nonuniform noise. The proposed ML estimator require an exhaustive search in the joint DOA and noise parameters space. In order to reduce the complexity, we break down the ML estimation to two separate problems. In the first problem, we estimate noise parameters while fixing the DOAs and in the second problem we obtain the DOAs from the estimated noise parameters. Simulation results indicate that the proposed algorithm shows at least 7 dB and 10 dB improvement in robustness in terms of root mean square error (RMSE) compared to the best results of MUSIC and ICA algorithms in the reverberant and noisy environments, respectively.

II. Signal Model

In this section, a model for the received signal in the SH domain is presented using the approach provided in [11]. Consider a spherical array of I identical omnidirectional microphones, where the i'th microphone located at Cartesian coordinates of $\boldsymbol{r}_i = [r\sin\theta_i\cos\phi_i, r\sin\theta_i\sin\phi_i, r\cos\theta_i]^T$, where (r, θ_i, ϕ_i) denote the corresponding spherical coordinates. The notations describing the spherical geometry are

ISBN 978-90-827970-1-5 © EURASIP 2018

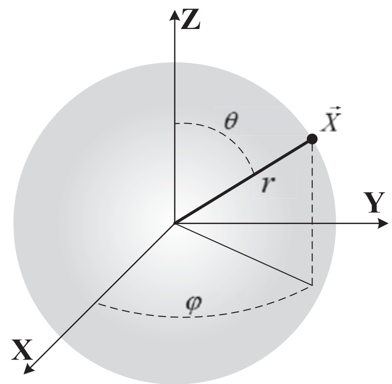

Fig. 1. The notations describing the spherical geometry

illustrated in Fig. 1. Assume that there exist L plain-wave source signals where l'th source impinging from the angular direction $\mathbf{\Psi}_l \triangleq (\theta_l', \phi_l')$ with wave-number k. The received signal at the i'th microphone from the l'th source at time t is $s_l\left(t - \tau_i(\mathbf{\Psi}_l)\right)$, where $\tau_i(\mathbf{\Psi}_l)$ is the propagation delay of the l'th source between the reference and i'th microphone. Considering narrow-band signal assumption of sound source, we have $s_l(t - \tau_i(\mathbf{\Psi}_l)) = e^{-j\boldsymbol{k}_l^T \mathbf{r}_i} s_l(t)$, where $\boldsymbol{k}_l = -[k \sin\theta_l' \cos\phi_l',\ k\sin\theta_l' \sin\phi_l', k\cos\theta_l']^T$ is the wave-vector corresponding to the l'th plane-wave. The received signal at i'th microphone at time t is [12]:

$$x_i(t) = \sum_{l=1}^{L} e^{-j\boldsymbol{k}_l^T \boldsymbol{r}_i} s_l(t) + n_i(t), \quad 1 \leq i \leq I, \tag{1}$$

where $n_i(t)$ is the additive white Gaussian noise with zero mean and variance σ^2 of the i'th microphone. On the other hand, by solving the wave equation in the spherical coordinates [13] and applying a proper truncation order N [14], the sound field can be approximated inside a sphere of radius $\hat{r}$ centered at the origin, as follows [8]:

$$e^{-j\boldsymbol{k}_l^T \boldsymbol{r}_i} = \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n(kr) Y_n^m(\mathbf{\Psi}_l) Y_n^m(\mathbf{\Phi}_i), \quad \| \boldsymbol{r_i} \| \leq \hat{r}, \tag{2}$$

where $b_n(kr) \triangleq 4\pi j^n j_n(kr)$ is the mode strength of order n for open sphere, $j = \sqrt{-1}$, j_n is the spherical Bessel function, $\mathbf{\Phi}_i \triangleq (\theta_i, \phi_i)$, $\hat{r} = \frac{2N}{ek_l}$, e is the Eulers number and $Y_n^m(\cdot)$ is the real-valued spherical harmonic of order n and degree m defined as:

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}}\, P_l^{|m|}(\cos\theta) \times \begin{cases} (-1)^m \sqrt{2} \cos(m\phi) & \text{for } m > 0 \\ 1 & \text{for } m = 0 \\ (-1)^m \sqrt{2} \sin(|m|\phi) & \text{for } m < 0 \end{cases}. \tag{3}$$

where P_n^m is the associated Legendre function of order n and degree m. Rewriting (2) in the matrix form, we have

$$\mathbf{A}(\mathbf{\Psi}) = \mathbf{Y}(\mathbf{\Phi}) \mathbf{B}(kr) \mathbf{Y}^T(\mathbf{\Psi}), \tag{4}$$

where $\mathbf{\Phi} \triangleq \{\mathbf{\Phi}_i, i = 1, \ldots, I\}$ and $\mathbf{Y}(\mathbf{\Psi})$ is the source spherical harmonics matrix of size $L \times (N+1)^2$ and defined as

$$\mathbf{Y}(\mathbf{\Psi}) \triangleq [\boldsymbol{y}(\mathbf{\Psi}_1)^T, \boldsymbol{y}(\mathbf{\Psi}_2)^T, \ldots, \boldsymbol{y}(\mathbf{\Psi}_L)^T]^T \tag{5}$$

where

$$\boldsymbol{y}(\mathbf{\Psi}_l) = [Y_0^0(\mathbf{\Psi}_l), Y_1^{-1}(\mathbf{\Psi}_l), Y_1^0(\mathbf{\Psi}_l), Y_1^1(\mathbf{\Psi}_l), \ldots, Y_N^N(\mathbf{\Psi}_l)],$$

the array spherical harmonics matrix, $\mathbf{Y}(\mathbf{\Phi})$, is the size of $I \times (N+1)^2$ and defined similar to (5) and mode strength matrix, $\mathbf{B}(kr)$, is the size of $(N+1)^2 \times (N+1)^2$ and defined as follows:

$$\mathbf{B}(kr) \triangleq \operatorname{diag}\left\{b_0(kr), b_1(kr), b_1(kr), b_1(kr), \ldots, b_N(kr)\right\}.$$

The spherical harmonics decomposition of the received signal can be obtained [15]:

$$\begin{aligned} x_{n,m}(t) &= \int_{\Omega \in S^2} x(t) Y_n^m(\Omega)\, d\Omega \\ &\cong \sum_{i=1}^{I} \alpha_i x_i(t) Y_n^m(\mathbf{\Phi}_i), \end{aligned} \tag{6}$$

where $\Omega = (\theta, \phi)$ and $x_{n,m}(t)$ are the coefficients of the spherical harmonics decomposition and α_i is the real-valued weighting parameter corresponding to the i'th microphone and obtained according to the spatial sampling scheme [16]. Equation (6) can be represented in a matrix form as:

$$\boldsymbol{x}_{\mathbf{nm}}(t) = \mathbf{Y}(\mathbf{\Phi})^T \mathbf{\Sigma} \boldsymbol{x}(t), \tag{7}$$

where $\mathbf{\Sigma} \triangleq \operatorname{diag}\left\{\alpha_1, \alpha_2, \ldots, \alpha_I\right\}$ and $\boldsymbol{x}_{\mathbf{nm}}(t)$ is defined as

$$\boldsymbol{x}_{\mathbf{nm}}(t) \triangleq [x_{0,0}(t), x_{1,-1}(t), x_{1,0}(t), x_{1,1}(t), \ldots, x_{N,N}(t)]^T. \tag{8}$$

Considering (6), the spherical harmonics are orthonormal as represented in [16]

$$\mathbf{Y}(\mathbf{\Phi})^T \mathbf{\Sigma} \mathbf{Y}(\mathbf{\Phi}) = \mathbf{I}, \tag{9}$$

where $\mathbf{I}$ is $(N+1)^2 \times (N+1)^2$ identity matrix. After some mathematical manipulation, the received signal model in the SH domain can be obtained:

$$\boldsymbol{b}(t) = \mathbf{Y}(\mathbf{\Psi})^T \boldsymbol{s}(t) + \boldsymbol{z}(t), \quad t = 1, \ldots, N_s \tag{10}$$

where $\boldsymbol{b}(t)$ is named higher-order ambisonic (HOA) signal vector of the order-N, $\boldsymbol{z}(t)$ is the noise vector in the SH domain and N_s is the number of snapshots. $\boldsymbol{b}(t)$ and $\boldsymbol{z}(t)$ are obtained as follows:

$$\boldsymbol{z}(t) = \mathbf{\Gamma} \boldsymbol{n}(t) \text{ and } \boldsymbol{b}(t) = \mathbf{\Gamma} \boldsymbol{x}(t) \tag{11}$$

where $\mathbf{\Gamma} \triangleq \mathbf{B}^{-1}(kr) \mathbf{Y}^H(\mathbf{\Phi}) \mathbf{\Sigma}$. According to (11), the HOA signal is a linear instantaneous mixture of the sources' signals. Transforming the received signals is performed by applying the time domain encoding filter, $\mathbf{\Gamma}$, to the received signal in the SH domain. It must be noticed that the transformation filter, $\mathbf{\Gamma}$, is known for the given array.

ISBN 978-90-827970-1-5 © EURASIP 2018

III. Robust Iterative Stochastic ML Estimation of DOA

In this section, a new ML DOA estimation of multiple sound sources in the SH domain is proposed by considering unknown and stochastic sources. It must be noted that the additive noise in (10) is spatially nonuniform white noise.

Suppose that the additive noise vector and sources' signals to be zero mean Gaussian with the covariance matrix $\mathbf{R}_n = \mathbf{Q} = \text{diag}\{q_1, q_2, \ldots, q_P\}$ and $\mathbf{R}_s = \sigma_s^2 \mathbf{I}_{\text{L}}$, respectively, where $\mathbf{I}_{\text{L}}$ is the $L \times L$ identity matrix. Also the noise is assumed to be independent of the sound sources. The set of unknown parameters are defined as $\boldsymbol{\Omega} \triangleq \{\boldsymbol{\Psi}, \mathbf{Q}\}$. Thus, the log-likelihood function of the HOA signal will be

$$L(\boldsymbol{b}; \boldsymbol{\Omega}) = -\frac{N_s}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{t=1}^{N_s} \boldsymbol{b}(t)^T \boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1} \boldsymbol{b}(t), \quad (12)$$

where $\boldsymbol{\Sigma}(\boldsymbol{\Omega}) = \sigma_s^2 \mathbf{Y}(\boldsymbol{\Psi}) \mathbf{Y}(\boldsymbol{\Psi})^T + \mathbf{Q}$ is the HOA signal covariance matrix. Thus, the ML estimator of $\boldsymbol{\Omega}$ can be written as:

$$\hat{\boldsymbol{\Omega}} = \arg\max_{\boldsymbol{\Omega}} L(\boldsymbol{b}; \boldsymbol{\Omega}). \quad (13)$$

Minimizing the objective function in (13), requires an exhaustive search in $(P + 2L)$-dimension space. This search procedure is time consuming. Thus, it is not practical. According to [17], to simplify this optimization problem, the ML estimation of $\boldsymbol{\Omega}$ is broken down to two separate problems for estimation of $\boldsymbol{\Psi}$ and $\mathbf{Q}$. After parameters separation, the optimization problem is solved iteratively as following. First, we fix $\boldsymbol{\Psi}$ in (12) with an initial value and derive the ML estimation of $\mathbf{Q}$. Then, the estimation of $\boldsymbol{\Psi}$ is obtained considering the estimated $\hat{\mathbf{Q}}$. After that, we estimate $\mathbf{Q}$ with the estimated $\hat{\boldsymbol{\Psi}}$ from the last iteration. We follow this procedure until the objective function, $\Delta L = L(\boldsymbol{b}; [\hat{\boldsymbol{\Omega}}]^i) - L(\boldsymbol{b}; [\hat{\boldsymbol{\Omega}}]^{i-1})$ be lower than a certain threshold, T_{thr}, where $[\hat{\boldsymbol{\Omega}}]^i$ and $[\hat{\boldsymbol{\Omega}}]^{i-1}$ denote estimated parameters in the iteration i and $i-1$.

The ML estimation of the noise vector elements with fixed $\boldsymbol{\Psi}$ can be described as:

$$\hat{q}_p = \arg\min_{q_p} N_s \ln(|\boldsymbol{\Sigma}(\mathbf{Q})|) + \sum_{t=1}^{N_s} \boldsymbol{b}(t)^T \boldsymbol{\Sigma}(\mathbf{Q})^{-1} \boldsymbol{b}(t). \quad (14)$$

Differentiating the above objective function with respect to q_p and doing some algebraic simplifications (See Appendix for derivation procedure), $\hat{q}_p$ is obtained as:

$$\hat{q}_p = \frac{N_s \left(1 - \boldsymbol{e}_p^T \boldsymbol{\Pi} \boldsymbol{e}_p\right)}{\sum_{t=1}^{N_s} \left(\boldsymbol{b}^T(t) \boldsymbol{\Sigma}_p^{-1}\right)^2}, \quad 1 \le p \le P, \quad (15)$$

where $\boldsymbol{e}_p$ and $\boldsymbol{\Sigma}_p^{-1}$ is the p'th column of $\mathbf{I}_{\text{P}}$ and $\boldsymbol{\Sigma}(\boldsymbol{\Omega})^{-1}$, respectively and

$$\boldsymbol{\Pi} \triangleq \tilde{\mathbf{Y}} \left(\frac{1}{\sigma_s^2} \mathbf{I}_{\text{L}} + \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}}\right)^{-1} \tilde{\mathbf{Y}}^T. \quad (16)$$

where $\tilde{\mathbf{Y}} \triangleq \mathbf{Q}^{-1/2} \mathbf{Y}(\boldsymbol{\Psi})$. After the estimation of $\mathbf{Q}$, the ML estimation of $\boldsymbol{\Psi}$ can be obtained as

$$\hat{\boldsymbol{\Psi}} = \arg\min_{\boldsymbol{\Psi}} N_s \ln(|\boldsymbol{\Sigma}(\boldsymbol{\Psi})|) + \sum_{t=1}^{N_s} \boldsymbol{b}(t)^T \boldsymbol{\Sigma}(\boldsymbol{\Psi})^{-1} \boldsymbol{b}(t). \quad (17)$$

To simplify the objective function in (17), the matrix determinant lemma gives the following equality

$$\det(\boldsymbol{\Sigma}(\boldsymbol{\Psi})) = \det(\mathbf{Q}) \det(\mathbf{I}_{\text{L}} + \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}}). \quad (18)$$

Using (18), the optimization problem in (17) can be restated as:

$$\begin{aligned} \hat{\boldsymbol{\Psi}} &= \arg\min_{\boldsymbol{\Psi}} N_s \ln\left(\det(\mathbf{Q}) \det(\mathbf{I}_{\text{L}} + \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}})\right) \\ &\quad + \sum_{t=1}^{N_s} \tilde{\boldsymbol{b}}^T(t) \boldsymbol{\Pi} \tilde{\boldsymbol{b}}(t) \\ &= \arg\min_{\boldsymbol{\Psi}} N_s \ln\left(\det(\mathbf{I}_{\text{L}} + \tilde{\mathbf{Y}}^T \tilde{\mathbf{Y}})\right) + \text{tr}\left\{\tilde{\mathbf{S}}_b \boldsymbol{\Pi}\right\}, \quad (19) \end{aligned}$$

where $\tilde{\boldsymbol{b}}(t) \triangleq \mathbf{Q}^{-1/2} \boldsymbol{b}(t)$ and $\tilde{\mathbf{S}}_b = \sum_{t=1}^{N_s} \tilde{\boldsymbol{b}}(t) \tilde{\boldsymbol{b}}^T(t)$. The optimization problem in (19) can be solved by Nelder-Mead direct search method [18]. Algorithm 1 summarizes the iterative stochastic ML (ISML) estimator considering spatially nonuniform noise.

Algorithm 1 ISML algorithm

Input: $\boldsymbol{b}(t), 1 \le t \le N_s$, the t-th vector of observation.
Output: $[\hat{\boldsymbol{\Psi}}]$, the vector of the estimated sources DOA.

1: **Initialization:** Initialize $[\hat{\boldsymbol{\Psi}}]^0$ randomly, $\left[\hat{\mathbf{Q}}\right]^0 = \text{diag}\{1, 1, \ldots, 1\} \in R^{P \times P}$ and $i = 1$.
2: **while** $\Delta L > T_{thr}$ **do**
3: Obtain $\left[\hat{\mathbf{Q}}\right]^i$ using $\left[\hat{\mathbf{Q}}\right]^{i-1}$ and (15).
4: Obtain $[\hat{\boldsymbol{\Psi}}]^i$ using $\left[\hat{\mathbf{Q}}\right]^i$, $[\hat{\boldsymbol{\Psi}}]^{i-1}$ and (19).
5: $[\hat{\boldsymbol{\Omega}}]^i \leftarrow \left\{[\hat{\boldsymbol{\Psi}}]^i, \left[\hat{\mathbf{Q}}\right]^i\right\}$
6: Compute $L(\boldsymbol{b}; [\hat{\boldsymbol{\Omega}}]^i)$ and then ΔL.
7: **end while**
8: **return** $[\hat{\boldsymbol{\Psi}}]$

IV. Simulation

In this section, the proposed ISML algorithm is evaluated and compared with the traditional standard MUSIC algorithm [4] and the recently proposed ICA based method [8] through various scenarios. In the conducted simulations, the SMA is an open array of radius 15 cm consisting 12 omnidirectional microphones. The microphone array is located in the coordinates (1.5 m, 6 m, 8 m) of a room of size 3 m × 8 m × 10 m. Three sound sources are located at 2 m distance of the array. Angular locations of these sources are reported in Table I. The sources play the speech signals with duration about one second which are sampled at 16 KHz. The signal to reverberation ratio (SRR) is almost equal to -3.5 dB and the room reverberation time (RT60) is approximately 350 ms. The room's impulse response between the sources and the array is obtained using MCRoomSim, a multichannel room acoustics simulator [19]. Microphone signals and additive white Gaussian noise are filtered with the HOA encoding filters which result in 2nd order HOA signals and SH domain noise, respectively. The

TABLE I
SOUND SOURCES DIRECTION

Source number	1	2	3
$(\phi^\circ, \theta^\circ)$	(40,50)	(70,-50)	(110,10)

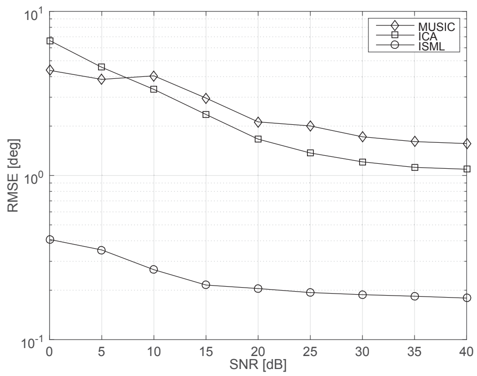

Fig. 2. Comparing RMSE of estimating ϕ (degree) in the ISML algorithm versus ICA and MUSIC methods.

length of the HOA encoding filters is 512 and designed such that its output SNR is maximized. Then, the HOA signals are filtered by bandpass filters with the pass-band of 500 to 3500 Hz. The optimization in (19) is performed by Nelder-Mead direct search method [18].

In Figs. 2 and 3, the average RMSE of the estimating ϕ and θ for ISML, ICA and MUSIC versus SNR are presented. The average of 100 different realizations are used to obtain each simulated point. As it can be seen, the proposed algorithm significantly improves RMSE compared to MUSIC and ICA. Performance of the ICA method is highly dropped in low SNR values due to not considering the environmental noise. In higher SNR values, the ICA assumption becomes closer to the reality, resulting in the ICA outperforms the MUSIC. The proposed algorithm exhibits a better performance because of regarding the spatially nonuniform noise model and reverberation. The signal is assumed to be independent and non-Gaussian for the ICA algorithm. But due to reverberation, both

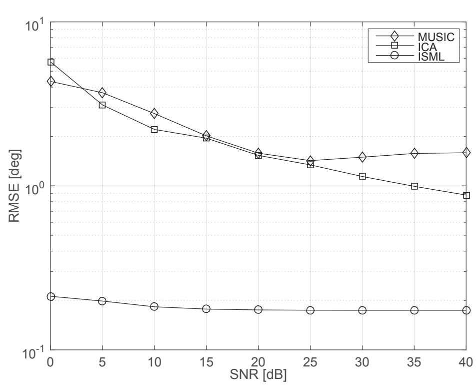

Fig. 3. Comparing RMSE of estimating θ (degree) in the ISML algorithm versus ICA and MUSIC methods.

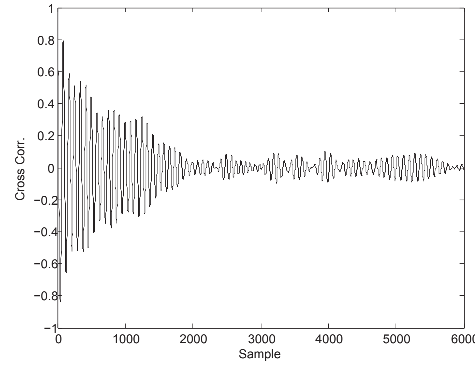

Fig. 4. Cross correlation between the first and the second HOA signals

assumptions are not realistic for DOA estimation in the SH domain. In order to show that the distribution of the HOA signal is Gaussian, Kolmogorov-Smirnov hypothesis test is used. The test result, with the 5% significance level, confirms that the HOA signals come from a Gaussian distribution. Also, the cross-correlation coefficient between the first and the second normalized HOA signals are calculated as $\rho = \mathrm{E}\{b_1 b_2\}$ which is equal to 0.834. For better visualization, the cross-correlation between the first and second HOA signals is plotted in Fig. 4. Considering the signal model matches to the HOA domain, the ISML can achieve lower RMSE of estimating ϕ and θ. Referring to this results, we can say that the proposed algorithm shows at least 10 dB improvement in robustness compared to the best results of MUSIC and ICA methods in the noisy environments.

To examine the robustness of the ISML in the reverberant environments, the average RMSE of the estimating ϕ and θ for 100 different realizations in 40 dB SNR versus different RT60 are reported in Table II and III. In the lower RT60s, ICA and ISML almost have the same performance. Because the ICA method does not consider the correlation in the HOA signals (referring to Fig. 4), its RMSE grows by increasing the RT60. Also, the MUSIC algorithm shows acceptable performance in the lower RT60 and degrades as RT60 increases. According to Tables II and III, the proposed algorithm shows at least 7 dB improvement in robustness compared to the best results of MUSIC and ICA methods in the reverberant environments as it was in the noisy environment.

V. Conclusion

In this paper, considering the general model of the received signal in the SH domain, we proposed an iterative stochastic ML estimation of DOA of multiple sources in the presence of spatially nonuniform noise. In order to reduce the complexity of the ML estimation, we break it down to two separate problems: noise parameters and DOA estimation problems. In the first problem, noise parameters is estimated while fixing the DOAs and in the second problem the DOAs from the estimated noise parameters is obtained. Simulation results demonstrated that the proposed algorithm shows at least 7dB and 10dB better robustness in terms of RMSE compared to

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE II
RMSE ESTIMATING ϕ (DEGREE) IN 40 dB SNR VERSUS DIFFERENT RT60S

RT60 [sec]	ISML	ICA	MUSIC
0.110	0.201	0.220	1.100
0.241	0.284	0.722	1.421
0.301	0.311	1.092	1.565
0.411	0.347	2.698	1.582
0.650	0.422	2.806	1.696
0.799	0.476	2.555	2.640

TABLE III
RMSE ESTIMATING θ (DEGREE) IN 40 dB SNR VERSUS DIFFERENT RT60S

RT60 [sec]	ISML	ICA	MUSIC
0.110	0.187	0.181	0.908
0.241	0.257	0.502	1.077
0.301	0.297	0.876	1.577
0.411	0.354	1.476	2.386
0.650	0.419	1.887	2.515
0.799	0.469	2.213	2.850

the MUSIC and ICA methods in the reverberant and noisy environments, respectively. Analyzing the convergence of the proposed algorithm and deriving the Cramer-Rao Bound for the received signal model in the SH domain is a part of our future work.

VI. APPENDIX

A. Derivation of (15)

Differentiating the objective function in (14) with respect to q_p and using the equality $\frac{\partial \boldsymbol{\Sigma}}{\partial q_p} = \boldsymbol{e}_p \boldsymbol{e}_p^T$ we have:

$$N_s \mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial q_p}\right\} - \sum_{t=1}^{N_s} \boldsymbol{b}(t)^T \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial q_p}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}(t) = N_s \mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\boldsymbol{e}_p^T\right\} - \sum_{t=1}^{N_s} \boldsymbol{b}(t)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\boldsymbol{e}_p^T\boldsymbol{\Sigma}^{-1}\boldsymbol{b}(t) = N_s \boldsymbol{e}_p^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p - \sum_{t=1}^{N_s}\left(\boldsymbol{b}(t)^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p\right)^2 = 0, \quad (20)$$

where $\boldsymbol{e}_p^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p$ are denoting the (p,p)-th element and p-th column of matrix $\boldsymbol{\Sigma}^{-1}$, respectively. Using the Woodburry matrix identity, inverse of $\boldsymbol{\Sigma}$ can be found as:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{Y}^T\left(\mathbf{I}_{\mathrm{L}} + \mathbf{Y}\mathbf{Q}^{-1}\mathbf{Y}^T\right)^{-1}\mathbf{Y}\mathbf{Q}^{-1}. \quad (21)$$

So $\boldsymbol{e}_p^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_p$ can be simplified as follows:

$$\left[\boldsymbol{\Sigma}^{-1}\right]_{pp} = \frac{1}{q_p}\left(1 - \boldsymbol{e}_p^T\boldsymbol{\Pi}\boldsymbol{e}_p\right), \quad (22)$$

with:

$$\boldsymbol{\Pi} \triangleq \tilde{\mathbf{Y}}^T\left(\mathbf{I}_{\mathrm{L}} + \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^T\right)^{-1}\tilde{\mathbf{Y}}. \quad (23)$$

Therefore, by substituting equation (22) into the last term of (20), q_p estimation will be:

$$\hat{q}_p = \frac{N_s\left(1 - \boldsymbol{e}_p^T\boldsymbol{\Pi}\boldsymbol{e}_p\right)}{\sum_{t=1}^{N_s}\left(\boldsymbol{b}(t)^T\boldsymbol{\Sigma}_p^{-1}\right)^2}, \quad 1 \leq p \leq P. \quad (24)$$

REFERENCES

[1] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008.

[2] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, no. 8, pp. 1950–1960, Aug. 2012.

[3] T. Tung, K. Yao, D. Chen, R. Hudson, and C. Reed, "Source localization and spatial filtering using wideband MUSIC and maximum power beamforming for multimedia applications," in *IEEE Workshop on Signal Processing Systems Design and Implementation*, 1999.

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

[5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[6] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011.

[7] L. Kumar, G. Bi, and R. M. Hegde, "The spherical harmonics root-music," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016.

[8] N. Epain and C. T. Jin, "Independent component analysis using spherical microphone arrays," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 91–102, Jan. 2012.

[9] T. Noohi, N. Epain, and C. T. Jin, "Direction of arrival estimation for spherical microphone arrays by combination of independent component analysis and sparse recovery," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013.

[10] ——, "Super-resolution acoustic imaging using sparse recovery with spatial priming," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015.

[11] L. Kumar and R. M. Hegde, "Stochastic cramer-rao bound analysis for DOA estimation in spherical harmonics domain," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1030–1034, Aug. 2015.

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

[13] E. G. Williams and J. A. Mann, "Fourier acoustics: Sound radiation and nearfield acoustical holography," *The Journal of the Acoustical Society of America*, vol. 108, no. 4, pp. 1373–1373, Oct. 2000.

[14] H. Sun, S. Yan, and U. P. Svensson, "Optimal higher order ambisonics encoding with predefined constraints," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 742–754, Mar. 2012.

[15] J. Driscoll and D. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, no. 2, pp. 202–250, Jun. 1994.

[16] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[17] C. Chen, F. Lorenzelli, R. Hudson, and K. Yao, "Maximum likelihood DOA estimation of multiple wideband sources in the presence of nonuniform sensor noise," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 835079, 2008.

[18] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, Jan. 1998.

[19] A. Wabnitz, N. Epain, C. Jin, and A. Van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, 2010, pp. 1–6.

ISBN 978-90-827970-1-5 © EURASIP 2018