Nested sampling for detection and localization of sound sources using a spherical microphone array

Ning Xiang 1 and Tomislav Jasa 2

1 Graduate Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy New York, 12180 USA; nxiang@rpi.edu
2 Thalgorithm Research, San Jose California, 95134 USA; tomislav.jasa@gmail.com
† This paper is an extended version of our paper published in MaxEnt 2023.

Abstract: Since its inception in 2004, nested sampling has been used in acoustics applications. This work applies nested sampling within a Bayesian framework to detection and localization of sound sources using a spherical microphone array. Beyond an existing work, this source localization task relies on spherical harmonics to establish parametric models that distinguish the background sound environment from the presence of sound sources. Upon a positive detection, the parametric models are also involved to estimate an unknown number of potentially multiple sound sources. For the purpose of source detection, a no-source scenario needs to be considered in addition to the presence of at least one sound source. Specifically, the spherical microphone array senses the sound environment. The acoustic data are analyzed via spherical Fourier transforms using Bayesian model comparison of two different models accounting for absence and presence of sound sources for the source detection. Upon a positive detection, potentially multiple source models are involved to analyze direction of arrivals (DoAs) using Bayesian model selection and parameter estimation for the sound source enumeration and localization. These are two levels (enumeration and localization) of inferential estimations being necessary to correctly localize potentially multiple sound sources. This paper discusses an efficient implementation of the nested sampling algorithm applied to the sound source detection and localization within the Bayesian framework.

Keywords: nested sampling, Bayesian model comparison, Bayesian model selection, parameter estimation, sound source detection, sound source localization.

0. Introduction

Nested Sampling (NS) was introduced by Skilling [1] as a numerical method for efficient Bayesian calculations. Soon afterward this method was applied to acoustics problems [2] where Jasa & Xiang explored using the Lebesgue integral as the mathematical foundation of the NS algorithm. Since then, that effort has resulted in a series of publications in acoustic applications [3–5]. This paper showcases that the NS algorithm has recently been applied in sound source detection and localization within a Bayesian framework. To detect and localize sound sources, the sound environment is sensed by a spherical microphone array whose signals are processed using a spherical Fourier transform. Spherical harmonics are exploited to process the acoustic data and to formulate the signal models. This paper emphasizes that source detection represents a model comparison problem, source enumeration represents a model selection problem, and source localization represents a parameter estimation problem all of which can be efficiently accomplished within Bayesian framework using the NS algorithm.

This paper presents a further development from the previous work [6] in that a background model for a no-source scenario needs to be established. The source detection problem is critically based on the model comparison between the no-source and one-source models. Special attention is given to the spherical harmonics when dealing with the background model and is separately dealt with in Sec. 1.2. In addition to these model
improvements, higher order (fourth) spherical components have been achieved due to further development of a 32-channel microphone array as illustrated in Fig. 1.

Figure 1. Spherical microphone array of radius $a = 3.5$ cm. Altogether 32 microphones are nearly uniformly flush-mounted over the rigid spherical surface.

1. Spherical Microphone Data and Models

This Section briefly introduces the data processing and the prediction models used for the sound source detection, enumeration, and direction of arrival estimations.

1.1. Microphone Array Data

When $Q$ microphones are arranged flush on a rigid sphere of radius $a$ nearly equidistantly, the sound pressure signals $P_{\text{mic}}$ are processed by

$$
D(\Theta) = \frac{4\pi}{Q} \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{Y_{mn}(\Theta)}{B(ka)} \sum_{q=1}^{Q} P_{\text{mic}}(\Theta_{q}) Y_{mn}^{*}(\Theta_{q}), \tag{1}
$$

with the third sum over $q = 1, \ldots, Q$ being spherical Fourier transform of $Q$ microphone signals $P_{\text{mic}}(\Theta_{q})$ at angular positions $\Theta_{q}$ and symbol $\ast$ stands for complex-conjugate. Function $B(ka)$ is a modal strength of the rigid sphere of radius $a$,

$$
B(ka) = jk \left[ j_n(ka) - \frac{j'_n(ka)}{h'_n(ka)} h_n(ka) \right], \tag{2}
$$

where $j = \sqrt{-1}$, $k = \omega/c$ is the propagation coefficient of sound waves. Functions $j_n(\cdot)$ and $h_n(\cdot)$ are spherical Bessel and Hankel functions, and $j'_n(\cdot)$ and $h'_n(\cdot)$, their derivatives, respectively. $\Theta = \{\theta, \phi\}$, collectively represents elevation and azimuth angles, while $\Theta_{q}$ specifies $Q$ microphone locations flush mounted on the spherical surface of the rigid sphere of radius $a$. Figure 1 shows a spherical microphone array of $Q = 32$ channels developed for this research. The spherical array is built upon a rigid sphere of radius $a = 3.5$ cm. In the following we denote $D = \{D(\Theta)\}$ as a two-dimensional matrix (vectors), representing the experimental data in the context of Bayesian inferential inversion.

Figure 2. Beamforming superposition of two sound sources using a spherical order $N = 4$. 
1.2. Prediction Models

In Eq. (1), $Y^m_n(\Theta)$ is so-called spherical harmonics of order $n$ and degree $m$, it is orthonormal and complete in a sense,

$$g(\Theta_s, \Theta) = 2\sqrt{\pi} \sum_{n=0}^{N} \sum_{m=-n}^{n} Y^m_n(\Theta_s) \ast Y^m_n(\Theta) \rightarrow \delta(\cos \theta - \cos \theta_s) \delta(\phi - \phi_s),$$  (3)

when $N \to \infty$. $\Theta_s$ is source angle in form of direction of arrival (DoA). $\delta(\cdot)$ represents Kronecker delta function. Using Eq. (3) we establish a predictive model of spherical beam-forming as

$$M_S(\Theta_S, \Theta) = \sum_{s=1}^{S} \frac{A_s g^2(\Theta_s, \Theta)}{\max_s g^2(\Theta_s, \Theta)}$$, (4)

where $A_s$ counts for different source energy strengths of individual sound sources. Note that $\Theta_S = \{A_1, A_2, \ldots, A_S, \Theta_1, \Theta_2, \ldots, \Theta_S\}$ collectively denotes both strength vector $A_S$ and angular directions (vectors) of $S$ sound sources, each angular vector contains one pair of elevation and azimuth angles $\Theta_s = \{\theta_s, \phi_s\}$. Variable $\Theta$ represents the angular range for possible sound sources to be localized. For $S \geq 1$, the kernel function $g(\Theta_s, \Theta)$ in Eq. (3) is processed for the upper order $(N + 1)^2 \leq Q$. That means, the integer-valued order $N$ of the spherical harmonics is limited by the number of microphone channels instead of infinity. Figure 2 illustrates a superposition of two simultaneous sources of equal amplitudes predicted by the model kernel in Eq. (3) for $N = 4$ before squaring operation to build source energy. The finite upper order $N$ is responsible to the width of lobes rather than a niddle-form ones. Figure 3 (a) illustrates the spherical microphone data for presence of two simultaneous sound sources, processed using Eqs. (1-2). While Figure 3 (b) illustrates the predicted map of the two simultaneous sound sources using the model in Eq. (4). The angular range is evaluated over $\Theta = \{0 \leq \theta \leq 180^\circ; 0 \leq \phi \leq 360^\circ\}$.

**Figure 3.** Comparison between the experimental data (a) processed according to Eq. (1) with the prediction model (b) in Eq. (4) of two sound sources using a spherical order $N = 4$.

When processing the microphone array data, there is no prior knowledge about the incoming sound field either with presence or absence of sound sources. It doesn’t make sense to pursue direction of arrival analysis, if no sound sources are present in the incoming microphone signals. For the model-based Bayesian detection we need to establish a background model. Special attention has to be given to the spherical harmonics
processing in this case. Specifically, \( M_0 \) represents the no-source model for \( S = 0 \). In this case, the kernel function \( g(\Theta_0, \Theta) \) in Eq. (3) is only calculated for \( N = 0 \), namely zero-order of the spherical harmonics.

\[
M_0(\Theta) = \frac{A_0 g^2(\Theta)}{\text{max}[g^2(\Theta)]},
\]

(5)

where the direction of ‘no-source’ \( \Theta_0 \) is irrelevant over the angular range \( \Theta \). For notation purpose, we collectively denote \( M_S = \{M_0, M_1, \ldots, M_S\} \) as being the prediction models for the directional of arrival analysis, while denote \( M_D = \{M_0, M_1\} \) for the sound source detection, a small subset of \( M_S \).

2. Bayesian Calculations

Given the data \( D \) as formulated in Sec. 1.1 and the prediction models \( M_S(\Theta_S) \) in Sec. 1.2, this work relies on Bayes theorem

\[
p(D|\Theta_S,M_S) \times P = L(\Theta_S) \times \Pi(\Theta_S),
\]

(6)

with \( \Pi(\Theta_S) = p(\Theta_S|M_S) \) being the prior probability and \( L(\Theta_S) = p(D|\Theta_S,M_S) \) being the likelihood function. The prior and the likelihood are both prior probabilistic in nature and need to be assigned a priori. This work applies the principle of the maximum entropy (MaxEnt), which leads to a uniform prior and a Student-T distribution for the likelihood (see Ref. [6] for details). The evidence \( Z \) in Eq. (6) plays a central role for the source detection and source enumeration problems and is determined by

\[
Z = \int_{\Theta_S} L(\Theta_S) \Pi(\Theta_S) \, d\Theta_S = \int_0^1 L(\mu) \, d\mu,
\]

(7)

where

\[
\mu(L_e) = \int_{L(\Theta_S) > L_e} \Pi(\Theta_S) \, d\Theta_S
\]

(8)

is the prior mass with \( L(\mu(L_e)) = L_e \), and \( 0 \leq L_e \leq L_{\text{max}} \) as derived by Skilling [7]. The NS algorithm generates a monotonically increasing partition of the likelihood range \([0, L_{\text{max}}]\)

\[
0 < L_0 < L_1 < \cdots < L_{T-1} < L_T < L_{\text{max}},
\]

(9)

via constrained sampling such that \( L_t \) with \( t \in \{0, 1, \ldots, T\} \) is sampled from the domain \( (\Theta_S : L(\Theta_S) > L_{t-1}) \). Observe that as \( L_t \) increases from 0 to \( L_{\text{max}}, \mu_t \) decreases from 1 to 0 where \( \mu_t = \mu(L_e) \) and the partition of Eq. (9) generates the monotonically decreasing sequence

\[
1 > \mu_0 > \mu_1 > \cdots > \mu_{T-1} > \mu_T > 0.
\]

(10)

Using the sequences in Eqs. (9-10), the one-dimensional integration on the far-right hand side of Eq. (7) is well approximated by

\[
Z \leftarrow \sum_{t=0}^T L_t \Delta \mu_t,
\]

(11)

with

\[
\Delta \mu_t = \mu_t - \mu_{t+1}, \quad \text{or} \quad \Delta \mu_t = \mu_{t-1} - \mu_t.
\]

(12)

Skilling [7] pointed out that the constrained prior mass is a statistical quantity and follows a shrinkage of

\[
\mu_t \approx e^{-t/P},
\]

(13)

after \( t \) iterations with \( P \) being an integer number for initializing random samples. A detailed proof of this result was given using order statistics in Appendix B of Jasa & Xiang [3].
NS was shown by Jasa & Xiang [3] to be a numerical implementation of Lebesgue integration where Eq. (11) represents the sum of weighted integrands of simple functions which are generated by partitioning the range rather than the domain of the function. An early account of this connection of the NS algorithm to Lebesgue integration can also be found in Jasa & Xiang [2], published in the Proceedings of the 25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2005).

3. Sound Source Detection, Enumeration, and Localization

The above described calculation is implemented for $P = 500$ [in Eq. (13)], the initial population with uniformly distributed prior ranges of all pending parameters, including sound source strength $A_s$, the elevation $\theta_S$ and azimuth angles $\phi_S$. For potentially simultaneous sound sources up to four ($S \leq 4$), the parameter space is of dimension up to $3 \times S$. The NS is applied to estimating Bayes factors for model order from 1 to 4 via evidences. It is used to rank a potential model accounting for an unknown number of simultaneous sound sources, so-called sound source enumeration process. This process has also been described previously in Xiang & Landschoot [6], followed by a DoA analysis based on the selected model $M_S$, that is carried out by Bayesian parameter estimation. When examining this effort critically, the authors recognize that the source enumeration, even it represents a higher level of inference via Bayesian model selection, it would still be incomplete if the machine sensory modality, such as in this application of a spherical microphone array, would not be notified that absence of sound sources oftentimes represents predominant portions of sound environment in practical scenarios. It will only make sense to pursue sound source enumeration and DoA analysis if any sound source is ever detected.

The sound source detection is carried out in the scope of this current work using Bayesian model comparison. The prediction models $M_D = \{M_0, M_1\}$ solely involve two models, $M_0$ in Eq. (5) and $M_1$ in Eq. (4) for $S = 1$. Note that Equation (5) is separately described because the $g^2(\Theta_0, \Theta)$ needs special attention, in which the spherical order is set to $N = 0$, while for $M_S$ for $S \geq 1$, the spherical order $N = 4$ due to the 32-channel spherical microphone array used for this work.

Figure 4. Sound source detection based on Bayesian model comparison. Bayesian evidence is estimated using both ‘no-source’ model $M_0$ and one-source model $M_1$. The evidence is expressed in unit [decibans], in honor of Thomas Bayes [8].

For the source detection, Bayesian evidence is estimated using the NS based on the ‘no-source’ model $M_0$ against ‘one-source’ model $M_1$. Specifically, for $M_0$-based sampling, there is still one pending parameter $A_0$ to sample. Figure 4 (Left) shows an experimental investigation when the microphone array data contain no sources, but noisy background signals. The evidence estimation using the NS demonstrates insignificant differences.
with that of $M_1$, indicating the source detection is negative. Figure 4 (Right) shows if the microphone array data contain sound sources, yet unknown number, the evidence estimation clearly shows significant differences in comparison with those of 'no-sources'. The source detection is positive.

Figure 5. Sound source enumeration based on Bayes factor estimation. The Bayes factors are expressed in unit [decibans], in honor of Thomas Bayes [8]. A two-source model is preferred by the Bayesian model selection. The evidence estimated using nested sampling also provides the posterior as a byproduct.

Upon a positive detection of sound sources, a further process involves Bayesian model selection. A set of sound source models from $M_0$ to $M_4$ is involved for estimating Bayes factors

$$B_{i,i-1} = 10 \lg \left( \frac{Z_i}{Z_{i-1}} \right), \quad [\text{decibans}]$$

for $i = 1, 2, \ldots 4$. Figure 5 illustrates one set of Bayes factor estimations. In this work, the evidence and Bayes factors are calculated in unit [decibans] denoting 10 times logarithm based 10 [$10 \lg$, in honor of Thomas Bayes [8]. In this case, the source enumeration using Bayesian model selection suggests that two-sources are contained in the data. At this stage the interest on specific DoA parameters is pushed into background. Upon selection of a two-source model using the NS, the exploration samples during the iterative NS process, also provide posterior samples for the model $M_2$ as byproduct; they are readily available once the evidence $Z_2$ for two sound sources is sufficiently explored. The posterior samples provide parameter estimates of two sound sources in terms of source strength $A_2$, and angular parameters $\Theta_2$. The data processed using Eq.(1) and the model prediction according to the posterior estimates are compared in Fig. 3.

4. Concluding Remarks

From its start of introduction into Bayesian calculations, Skilling’s nested sampling [1] had an immediate impact on room-acoustic research [2], where an early account of Lebesgue integral view on the nested sampling was first exposed in the MaxEnt Community in 2005. A thorough handling of its mathematical foundation of Lebesgue integral is given at a later point [3]; ‘Interpreting nested sampling as a statistical approximation of a Lebesgue integral opens the possibility of a large body of existing research to be applied in the analysis and possible extension of the algorithm.’ Over the past 15 years a stream of applications using NS in acoustics science and engineering has been emerging. Among others, this paper reports on an acoustic application of nested sampling using a spherical microphone array within Bayesian framework.

Funding: This research received no external funding.
Acknowledgments: The authors are grateful to Drs. John Skilling, Paul Goggans, and Kevin Knuth for their stimulating discussions. Mrs. Stephen Weikel, Christopher Landschoot, and Thomas Metzger have contributed to parts of this work in scope of their MS degree thesis projects.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DoA Direction of arrival
NS Nested sampling
MaxEnt Principle of maximum entropy

References