Nested sampling applied in Bayesian room-acoustics decay analysis[a)]

Tomislav Jasa
Thalgorithm Research, Toronto, Ontario, L4X 1B1, Canada

Ning Xiang[b)]
Graduate Program in Architectural Acoustics, School of Architecture, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 15 December 2011; revised 31 August 2012; accepted 10 September 2012)

Room-acoustic energy decays often exhibit single-rate or multiple-rate characteristics in a wide variety of rooms/halls. Both the energy decay order and decay parameter estimation are of practical significance in architectural acoustics applications, representing two different levels of Bayesian probabilistic inference. This paper discusses a model-based sound energy decay analysis within a Bayesian framework utilizing the nested sampling algorithm. The nested sampling algorithm is specifically developed to evaluate the Bayesian evidence required for determining the energy decay order with decay parameter estimates as a secondary result. Taking the energy decay analysis in architectural acoustics as an example, this paper demonstrates that two different levels of inference, decay model-selection and decay parameter estimation, can be cohesively accomplished by the nested sampling algorithm. © *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4754550]

PACS number(s): 43.60.Uv, 43.55.Br, 43.55.Mc, 43.60.Pt [ZHM] Pages: 3251–3262

I. INTRODUCTION

Bayesian methods have been utilized in a wide range of acoustics applications[1–16] with increasing attention being given to the task of performing Bayesian model selection. A few of the most recent applications of Bayesian model selection can be found in such acoustics applications as room-acoustics energy decay analysis[1] and geo-acoustics inversion,[2,3] where the model selection problems are tackled separately from that of parameter estimation. The major challenge in Bayesian model selection is the efficient calculation of the *Bayesian evidence* used to rank competing models. Xiang and Goggans[1] utilized marginalization of the acoustic model along with an assumption on the form of the posterior distribution which works well for many single-slope and double-slope decays; however, once the decay model is of third or higher order, or the second-slope decay is significantly low in level, these assumption can cease to be valid.[1,4] Battle *et al.*[3] accomplished model selection for geo-acoustics inversion problems using an importance sampling algorithm, the success of which depends critically on proper choice of the importance sampling distribution. In recent work by Xiang *et al.*[4] and Dettmer and Dosso,[5] Bayesian model selection applied to room-acoustic decay order estimation, room acoustics energy decay analysis, and geo-acoustic inversion problems was solved using the Bayesian information criterion (BIC). The BIC is based on the assumption that the posterior probability distribution is well approximated by a multivariate Gaussian probability distribution. Analysis of experimental multiple-rate energy decay data has indicated that the asymptotic approximation assumed by the BIC is critically sensitive to the maximum posterior estimation. Dettmer and Dosso[2] have recently applied the annealed importance sampling algorithm to model selection in the context of geo-acoustical inversion, which indicates a requirement for more elaborate model selection algorithms in the acoustics community. This paper applies the nested sampling algorithm proposed by Skilling[17,18] to Bayesian room-acoustics energy decay analysis. The paper presents the nested sampling algorithm as a numerical implementation of the Lebesgue integral as originally proposed by Jasa and Xiang[12] in order to provide acousticians an alternative understanding of the nested sampling algorithm's theoretical foundation.

This paper is organized as follows: Section II presents a brief introduction to sound energy decay analysis and demonstrate how this specific architectural acoustics application requires a model-based data analysis. Section III outlines the two levels of inference that are required for decay model selection and decay parameter estimation, and cohesively formulates both decay model selection and decay parameter estimation using Bayesian probability theory. Sections IV and V derive the nested sampling algorithm in detail, in the context of Bayesian model selection and Bayesian parameter estimation. Section VI discusses experimental results using experimentally measured data in the form of acoustical room impulse responses and subsequent Schroeder decay functions. Section VII discusses possible extensions to the nested sampling algorithm in the context of Bayesian model selection and Bayesian parameter estimation.

II. MODEL-BASED BAYESIAN INFERENCE

Sound energy decays often exhibit multiple-rate characteristics in a wide variety of enclosures which results in alternative parametric models (with each model corresponding to

[a)]Aspects of this work have been presented at the 19th ICA, Madrid, Spain and the 159th ASA Meeting in Baltimore, MD.
[b)]Author to whom correspondence should be addressed. Electronic mail: xiangn@rpi.edu

a given number of decay slopes) competing with each other in order to explain the experimental data. This model-based analysis is used to formulate the two levels of inference within Bayesian framework and should prove beneficial to a broad audience of acoustic scientists.

A generalized linear parametric model with s exponential decay terms,

$$\mathbf{H}_s = \mathbf{G}_s\mathbf{A}_s, \quad 1 \le s \le M, \tag{1}$$

is expected to describe the data $\mathbf{D}$ in the form of a Schroeder decay function,[19] with M representing the number of competing models under consideration. Both data $\mathbf{D}$ and the model $\mathbf{H}_s$ are column vectors of K elements, representing the (normalized) sound energy decays. $\mathbf{G}_s$ is a $K \times (s+1)$ matrix, with the jth column of $\mathbf{G}_s = [G_{kj}^{(s)}]$ given by

$$G_{kj}^{(s)}(T_j^{(s)}, t_k) = \begin{cases} t_K - t_k & \text{for} \quad j = 0 \\ e^{-(13.8\cdot t_k)/T_j^{(s)}} & \text{for} \quad j = 1, 2, \ldots, \end{cases} \tag{2}$$

where $0 \le k < K$. Time constant t_K represents the upper limit of Schroeder's integration and $0 \le j \le s$. The variable t_k represents discrete times. Column vector $\mathbf{A}_s = [A_0^{(s)}, A_1^{(s)}, \ldots, A_s^{(s)}]^{Tr}$ contains $s+1$ linear coefficients (Tr denotes a matrix transpose), while column vector $\mathbf{T}_s = [T_1^{(s)}, T_2^{(s)}, \ldots, T_s^{(s)}]^{Tr}$ contains s decay time parameters to be determined. The validity of this model for estimating the energy decay parameters in acoustically coupled spaces has been experimentally verified in Refs. 20 and 21, especially when t_K is large. Each parametric model $\mathbf{H}_s$ contains parameters $\mathbf{A}_s$ and $\mathbf{T}_s$. They are summarized by parameter vector $\mathbf{w}_s = [\mathbf{A}_s, \mathbf{T}_s]^{Tr}$, and the parametric model is denoted as $\mathbf{H}_s(\mathbf{w}_s)$ with $1 \le s \le M$ for ease of the following discussion. The subscript s of $\mathbf{w}_s$ and the superscript (s) of parameters $A_j^{(s)}$ and $T_j^{(s)}$ is the decay order; namely, the decay model contains s different exponential decay terms with s different decay times.

III. TWO LEVELS OF BAYESIAN INFERENCE

Bayesian probabilistic inference encompasses both parameter estimation, which can be considered as the first level of inference, and model selection, which can be considered the second level of inference.[1,22,23] Focusing on model selection first is scientifically logical, as a poor model can give misleading results no matter how accurately the model parameters themselves have been estimated.

A. Model selection: The second level of inference

Selecting a model that fits the data best is a poor method for model selection as a more complex model can always fit the data either equally well or better, but may generalize poorly or lead to over-fitting of data.[22,23] In practice an investigator posses sufficient domain knowledge in order to focus on a finite number of parametric models, all of which describe the data reasonably well. This fact is incorporated into the background information I. Bayesian model selection applies Bayes' theorem to a competing set of models $\mathbf{H}_s$ given the data, $\mathbf{D}$, and the background information, I, by the relation

$$p(s|\mathbf{D}) = \frac{p(s)p(\mathbf{D}|s)}{p(\mathbf{D})}, \tag{3}$$

while pushing any interest in model parameter values into the background. While the background information I should be explicitly given on the right side of solidus $(\cdot|I)$ of each probabilistic quantity, it is omitted for simplifying notations throughout this paper. Model $\mathbf{H}_s$ is also specified by s throughout this paper for simplicity, but still bearing in mind that s in the following also represents the decay model in Eq. (1) via Eq. (2). Bayes' theorem in the form of Eq. (3) represents how one's prior knowledge about model $\mathbf{H}_s$, expressed by *prior probability* $p(s)$, is updated in presence of the data $\mathbf{D}$. Quantity $p(\mathbf{D}|s)$ represents the probability that the given model $\mathbf{H}_s$ generates the data, and is termed the *marginal likelihood* or *evidence* of the data in the context of the model selection. Probability $p(s|\mathbf{D})$ on the left-hand side of the equation is the *posterior probability* of the model $\mathbf{H}_s$ given the data, representing the updated knowledge about the model *after* the data become available. When comparing two competing models, the posterior ratio or so called Bayes factor[24] between two different models $\{\mathbf{H}_i, \mathbf{H}_j\}$,

$$\frac{p(i|\mathbf{D})}{p(j|\mathbf{D})} = \frac{p(i)}{p(j)}\frac{p(\mathbf{D}|i)}{p(\mathbf{D}|j)}, \tag{4}$$

can be determined as the quantity $p(\mathbf{D})$ cancels out. Note that the Bayes factor only ranks the models under consideration against each other; it cannot rule out the fact that there may be a much better model that has not been considered in the analysis. The model selection problem investigated in this paper will be one which can be incorporated in the background information I: Looking only within a specified set Φ of competing models $\{\mathbf{H}_1, \ldots, \mathbf{H}_M\}$, which model is most preferred by the data, and how strongly is it supported by the data relative to the competing alternatives in the set.[25] Upon the presence of the data, only one model among the competing model set will be selected. The first ratio on the right-hand side in Eq. (4), termed the *prior ratio*, represents the degree of one's initial belief on how much model $\mathbf{H}_i$ would have been preferred over model $\mathbf{H}_j$ before the data were acquired, while the second ratio on the right-hand side of Eq. (4), termed the *marginal likelihood ratio* or *evidence ratio*, measures how much the data prefer the model $\mathbf{H}_i$ over $\mathbf{H}_j$. Assigning equal prior probability among the models results in a logarithmic Bayes factor

$$\log\left[\frac{p(i|\mathbf{D})}{p(j|\mathbf{D})}\right] = \log(p(\mathbf{D}|i)) - \log(p(\mathbf{D}|j)). \tag{5}$$

Equation (5) indicates that the marginal likelihood or evidence $p(\mathbf{D}|s)$ plays a central role in Bayesian model selection. The second level of Bayesian inference intrinsically embodies Occam's razor in a quantitative way by penalizing over-parameterized models, assigning them larger probabilities only if the complexity of the data justifies the additional complication of the model. The result is a concise model that provides a good fit to the data.[26] A detailed discussion on the quantitative implementation of Occam's razor or the principle of parsimony is given in recent Refs. 22 and 27.

B. Parameter estimation: The first level of inference

Parameter estimation, as the first level of Bayesian inference, applies Bayes' theorem to the parameter vector $\mathbf{w}_s$ and the data $\mathbf{D}$ given a specific model $\mathbf{H}_s(\mathbf{w}_s)$

$$p(\mathbf{w}_s|\mathbf{D},s)=\frac{p(\mathbf{w}_s|s)p(\mathbf{D}|\mathbf{w}_s,s)}{p(\mathbf{D}|s)}. \tag{6}$$

Bayes' theorem used in this problem represents how one's prior knowledge about parameters $\mathbf{w}_s$ given the specific model $\mathbf{H}_s(\mathbf{w}_s)$ and the background information I, as expressed by prior probability $p(\mathbf{w}_s|s)$, are updated in the presence of data $\mathbf{D}$. The background information I specifies that one model $\mathbf{H}_s$ in Eq. (1) is given via model selection which approximates the data $\mathbf{D}$ such that the residual error vector $\mathbf{e}$ between the data $\mathbf{D}$ and the model $\mathbf{H}_s$,

$$\mathbf{e}=\mathbf{D}-\mathbf{H}_s, \tag{7}$$

is finitely bounded. Probability $p(\mathbf{D}|\mathbf{w}_s,s)$ expresses the likelihood that the parameter set $\mathbf{w}_s$ associated with the given model $\mathbf{H}_s$ generates the data $\mathbf{D}$. It is therefore termed the likelihood function and can also be written as

$$L(\mathbf{w}_s)=p(\mathbf{D}|\mathbf{w}_s,\,s). \tag{8}$$

The likelihood function $L(\mathbf{w}_s)$ for the sound energy decay analysis has been well discussed in previous publications,[14,21] it is the probability of the residual error $\mathbf{e}$. The background information I states that the only available information about the error $\mathbf{e}^2$ is that it corresponds to a finite but unknown bounded value which implies a finite variance σ^2. If the finite amount of error could not be warranted, the model is obviously wrong and has to be reestablished within the current problem at hand. Taking the finite variance of the error as the only information regarding the error statistics, application of the principle of maximum entropy results in a Gaussian probability density function for $\mathbf{e}$,[27]

$$L(\mathbf{w}_s,\sigma)=(\sqrt{2\pi\sigma})^{-K}\exp\left(\frac{-E}{\sigma^2}\right), \tag{9}$$

with

$$E=\frac{\mathbf{e}^{Tr}\mathbf{e}}{2}, \tag{10}$$

and the unspecified variance σ^2. The maximum entropy principle also leads to logical independence of e_k in the residual error vector $\mathbf{e}$, which has been used to multiply the probabilities for each data point according to the product rule. This assignment is not the same as if the residual error is taken to be Gaussian white. The maximum-entropy assignment of a Gaussian likelihood function follows from the finite variance of the error and the fact that no further information about the residual errors is available.[27]

The notation $L(\mathbf{w}_s,\sigma)$ emphasizes that the finite error variance σ^2 is unspecified when applying the maximum entropy principle. In the parameter estimation problem, the variance σ^2 is a nuisance parameter, being necessary for the probability assignment, but of no interest for the relevant parameter values. The integral over all possible values of σ by assigning a Jeffreys prior[25] is another important tool within the Bayesian framework, termed marginalization. The marginalization results in[14]

$$L(\mathbf{w}_s)=\Gamma\left(\frac{K}{2}\right)\frac{(2\pi E)^{-K/2}}{2}, \tag{11}$$

where $\Gamma(\cdot)$ is the gamma function.

Probability $p(\mathbf{w}_s|\mathbf{D},s)$ on the left-hand side of Eq. (6) is the *posterior probability* of the parameters $\mathbf{w}_s$, representing the updated knowledge about the parameters once the data become available. The integral over the entire parameter space results in

$$\int_{\mathbf{w}_s}p(\mathbf{w}_s|\mathbf{D},s)\,d\mathbf{w}_s=1. \tag{12}$$

An integration over the entire parameter space on both sides of Eq. (6), along with Eq. (8) yields

$$p(\mathbf{D}|s)=Z_s=\int_{\mathbf{w}_s}L(\mathbf{w}_s)p(\mathbf{w}_s)\,d\mathbf{w}_s, \tag{13}$$

where $p(\mathbf{w}_s)=p(\mathbf{w}_s|s)$ for simplicity. Quantity $p(\mathbf{D}|s)$ in the denominator on the right-hand side of Eq. (6), now defined by Eq. (13), is exactly the same as the marginal likelihood in Eq. (3) and plays a central role in model selection. Equation (6) can then be expressed as

$$\overbrace{p(\mathbf{w}_s|\mathbf{D},s)}^{\text{posterior}}\times\underbrace{Z_s}_{\text{evidence}}=\overbrace{p(\mathbf{w}_s)}^{\text{prior}}\times\underbrace{L(\mathbf{w}_s)}_{\text{likelihood}}. \tag{14}$$

The quantity $Z_s=p(\mathbf{D}|s)$ is known as the *evidence* for model $\mathbf{H}_s$ in the context of Bayesian model selection.[22] The logarithm of the evidence $\log(Z_s)$ is often presented in the context of Bayes factors as discussed in Sec. III A. Equation (14) states the logical relations among the quantities of the Bayesian inference: Prior probability $p(\mathbf{w}_s)$ and the likelihood function $L(\mathbf{w}_s)$ are the inputs, while the posterior probability distribution $p(\mathbf{w}_s|\mathbf{D},s)$ and the evidence Z_s are the outputs of the Bayesian inference.[17]

IV. SIMPLE FUNCTIONS AND NESTED SAMPLING

Nested sampling, proposed by Skilling,[17] provides an algorithm for estimating the evidence Z_s in addition to determining the decay parameter estimates, thus solving the two levels of Bayesian inference.

Evaluating the evidence Z_s by numerical integration can be problematic due to difficulty in determining an efficient partition of the (typically multi-dimensional) model parameter space. A uniformly fine partition of the parameter space will be required unless the location, shape, and size of the posterior probability distribution is known *a priori*. Monte Carlo integration methods such as importance sampling[28] remove the need to define a uniform partition of the parameter space and instead rely on defining sampling distributions which, in effect, partition the parameter space in an efficient way. The difficulty with these methods is in defining appropriate sampling distributions which can be especially challenging for high dimensional posterior probability distributions. Annealing methods such as thermodynamic

integration[29] and annealed importance sampling[30] use the form of the posterior probability distribution itself to define sampling distributions. The difficulty with annealing methods is defining an efficient temperature schedule. Nested sampling avoids annealing and instead focuses on partitioning of the range of the likelihood function. Partitioning the range as opposed to the domain of a function in order to determine the integral is well established in the mathematical literature through the theory of Lebesgue integration.[31] Interpreting nested sampling as a statistical approximation of a Lebesgue integral as done by Jasa and Xiang[12] opens the possibility of a large body of existing research to be applied in the analysis and possible extension of the algorithm.

A. Simple functions and Lebesgue integration

A starting point of the Lebesgue integral is to define the integral for *simple functions* which take on only a finite number of values over the domain and approximate the integral for the distribution by a limit of such integrals.[31] An ordered partition of the range of the likelihood function $L(\mathbf{w}_s)$,

$$0 = L_0 < \cdots < L_n, \tag{15}$$

defines the sequence of nested sets

$$\varnothing \subset \Omega_n \subset \cdots \subset \Omega_0 = \Omega, \tag{16}$$

where

$$\Omega_i = \{\mathbf{w}_s | L(\mathbf{w}_s) > L_i\}, \tag{17}$$

$$\Omega = \{\mathbf{w}_s | L(\mathbf{w}_s) > 0\} = \mathbf{w}_s, \tag{18}$$

$$\varnothing = \text{the null set}. \tag{19}$$

Using the nested sets, $\Omega_n \subset \cdots \subset \Omega_0 = \Omega$, construct a sequence of disjoint sets

$$A_0, \ldots, A_n, \tag{20}$$

where

$$A_i = \begin{cases} \Omega_i \cap \bar{\Omega}_{i+1} & \text{for } 0 \le i \le n-1 \\ \Omega_n & \text{for } i = n, \end{cases} \tag{21}$$

with $\bar{\Omega}_{i+1}$ denoting the complement of Ω_{i+1} [see Fig. 1(b) where $L(\mathbf{w}_s)$ is defined for a one-dimensional parameter space $\mathbf{w}_s$]. The simple function $S^n(\mathbf{w}_s)$ is then defined by

$$S^n(\mathbf{w}_s) = \sum_{i=0}^{n} L_i 1_{A_i}(\mathbf{w}_s), \tag{22}$$

where $1_X(\mathbf{w}_s)$ is the indicator function for any set $\varnothing \subset X \subset \Omega$

$$1_X(\mathbf{w}_s) = \begin{cases} 1 & \text{if } \mathbf{w}_s \in X \\ 0 & \text{otherwise}, \end{cases} \tag{23}$$

an example of which is shown in Fig. 1(a) where $L(\mathbf{w}_s)$ is defined for a one-dimensional parameter space $\mathbf{w}_s$.

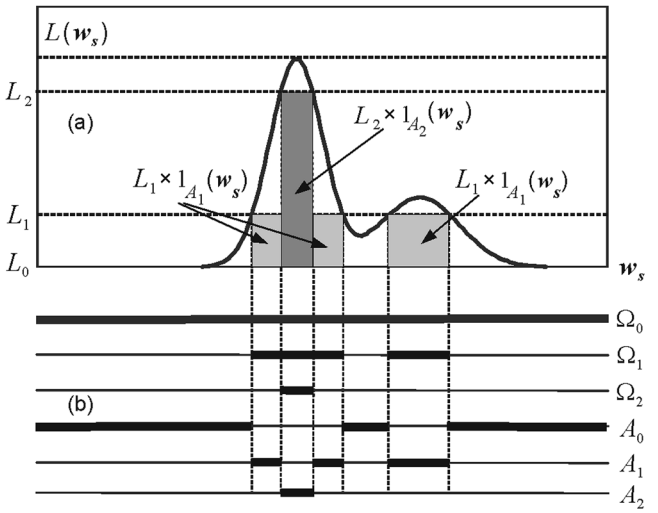

FIG. 1. Construction of the simple function $S^n(\mathbf{w}_s) = \Sigma_{i=1}^n L_i 1_{A_i}(\mathbf{w}_s)$ along with sets $\Omega_0, \Omega_1, \Omega_2$, and A_0, A_1, A_2 for the likelihood function $L(\mathbf{w}_s)$ given the ordered partition $L_0 < L_1 < L_2$.

Given the prior probability distribution $p(\mathbf{w}_s)$, the measure $\mu(\Omega_i)$, or probability $\mathbb{P}[\Omega_i]$ of a set Ω_i is given by

$$\mu(\Omega_i) = \mathbb{P}[\Omega_i] = \int_{\Omega_i} p(\mathbf{w}_s) d\mathbf{w}_s, \tag{24}$$

with

$$\mu(A_i) = \begin{cases} \mu(\Omega_i) - \mu(\Omega_{i+1}) & \text{if } 0 \le i \le n-1 \\ \mu(\Omega_i) & i = n, \end{cases} \tag{25}$$

and

$$\mu(\Omega) = 1, \tag{26}$$

$$\mu(\varnothing) = 0 \tag{27}$$

by the properties of measures.[31] The integral of $S^n(\mathbf{w}_s)$ is defined by the finite sum

$$\int_{\mathbf{w}_s} S^n(\mathbf{w}_s)\, d\mu = \sum_{i=0}^{n} L_i \mu(A_i). \tag{28}$$

The final step of the simple function approximation is to quantify the relationship between

$$\sum_{i=0}^{n} L_i \mu(A_i) \tag{29}$$

and the Bayesian evidence

$$Z_s = \int_{\mathbf{w}_s} L(\mathbf{w}_s) p(\mathbf{w}_s)\, d\mathbf{w}_s. \tag{30}$$

If the size n of the ordered partition $0 = L_0 < \cdots < L_n$ increases such that $S^n(\mathbf{w}_s)$ converges to $L(\mathbf{w}_s)$ in a pointwise, monotonically increasing manner, the integral of the simple function will converge to the evidence Z_s in a monotonically increasing manner as is shown in Appendix A

(along with definitions for monotonic and point-wise monotonic convergence) resulting in the approximation

$$Z_s \approx \sum_{i=0}^{n} L_i \mu(A_i). \tag{31}$$

As the range of $L(\mathbf{w}_s)$ is one-dimensional regardless of the dimensionality of the parameter space $\mathbf{w}_s$, the computational cost in Eq. (31) scales linearly in n assuming that $\mu(A_i)$ is known, which allows for fine partitions of the range of $L(\mathbf{w}_s)$ to be computationally tractable. This benefit is tempered by the realization that exact determination of the measure $\mu(A_i)$ requires the evaluation of an integral over a potentially high dimensional parameter space $\mathbf{w}_s$. The utility of using the simple function approximation to efficiently estimate the evidence Z_s will depend on the algorithm chosen to define $0 = L_0 < \cdots < L_n$ along with estimates of $\mu(A_i)$. Nested sampling, presented in Sec. IV B, is one such algorithm.

B. Nested sampling

The nested sampling algorithm is an iterative process where the input at each iteration i is given by

(i) The prior probability distribution $p(\mathbf{w}_s)$,
(ii) likelihood L_i, which defines the set Ω_i as given by Eq. (17).

For each iteration:

(i) Generate $r(i)$ independent random samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ from the constrained prior probability distribution

$$p_{\Omega_i}^C(\mathbf{w}_s) = p(\mathbf{w}_s) \times 1_{\Omega_i}(\mathbf{w}_s). \tag{32}$$

(ii) Create a list of samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{(r(i))}$, from $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ which are ordered by likelihood such that $L_{(1)} < \cdots < L_{(r(i))}$, where $L_{(i)} = L(\mathbf{w}_s^{(i)})$ and $\mathbf{v}_s^i = \mathbf{w}_s^{(1)}$.

The output of each iteration i consists of

(i) The likelihood $L_{i+1} = L_{(1)}$ which defines the set $\Omega_{i+1} = \Omega_{(1)}$ as given by Eq. (17),
(ii) sample $\mathbf{v}_s^i = \mathbf{w}_s^{(1)}$, with $\mathbf{w}_s^{(1)}$ being the sample with the lowest likelihood value within iteration i.

An example of a nested sampling iteration is shown in Fig. 2.

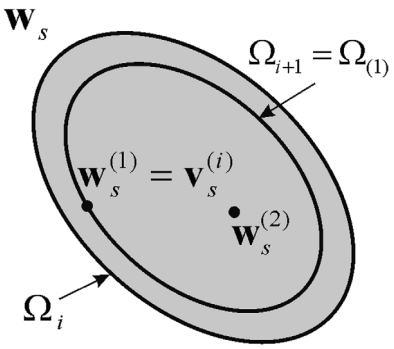

FIG. 2. Given the set Ω_i with associated likelihood L_i, $r(i) = 2$ samples $\mathbf{w}_s^{(1)}, \mathbf{w}_s^{(2)}$ are generated from the constrained prior probability distribution $p_{\Omega_i}^C(\mathbf{w}_s)$ which are sorted by likelihood such that $L_i < L_{(1)} < L_{(2)}$, where $L_{(j)} = L(\mathbf{w}_s^{(j)})$. The nested sampling algorithm assigns $L_{i+1} = L_{(1)}$, $\Omega_{i+1} = \Omega_{(1)}$.

Starting with $L_0 = 0$, the nested sampling algorithm when iterated n times, will produce an ordered partition $0 = L_0 < \cdots < L_n$ of $L(\mathbf{w}_s)$. An important property of the nested sampling algorithm is the ability to estimate the measures $\mu(\Omega_i)$: as $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ are randomly generated at every iteration, the likelihood values L_i, the sets Ω_i and measures $\mu(\Omega_i)$ produced by the Nested Sampling algorithm are random or stochastic in nature, i.e., $\mu(\Omega_n), \ldots, \mu(\Omega_0)$ are random variables. The joint probability density function for $\mu(\Omega_n), \ldots, \mu(\Omega_0)$ is given by

$$f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\mu_n, \ldots, \mu_0) = \left[\prod_{j=1}^{n} f_{\mu(\Omega_j)|\mu(\Omega_{j-1})}(\mu_j)\right] f_{\mu(\Omega_0)}(\mu_0), \tag{33}$$

with

$$f_{\mu(\Omega_j)|\mu(\Omega_{j-1})}(\mu_j) = \begin{cases} \dfrac{r(j)}{\mu(\Omega_{j-1})}\left[\dfrac{\mu_j}{\mu(\Omega_{j-1})}\right]^{r(j)-1} & \text{if } 0 \le \mu_j \le \mu(\Omega_{j-1}) \\ 0 & \text{otherwise,} \end{cases} \tag{34}$$

as shown in Appendix B, and

$$f_{\mu(\Omega_0)}(\mu_0) = \delta(\mu_0 - 1), \tag{35}$$

which signifies that $\mu(\Omega_0) = \mu(\Omega) = 1$.

C. Expected values and nested sampling

Using the probability density function

$$f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\mu_n, \ldots, \mu_0), \tag{36}$$

one can calculate the expected value

$$\langle g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] \rangle = \int_{\bar{\mu}} g(\bar{\mu}) f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\bar{\mu})\, d\bar{\mu}, \tag{37}$$

where $g(\bullet)$ is a function of the measures $\mu(\Omega_n), \ldots, \mu(\Omega_0)$ and $\bar{\mu} = \mu_n, \ldots, \mu_0$.

The expectation of Eq. (37) can be quantified by Monte Carlo estimates

$$\bar{g} = \frac{1}{m}\sum_{j=1}^{m} g(\bar{\mu}^j), \tag{38}$$

and

$$\mathrm{var}(\bar{g}) = \frac{1}{m-1}\sum_{j=1}^{m} (g(\bar{\mu}^j) - \bar{g})^2, \tag{39}$$

where sample $\bar{\mu}^j$ is drawn from $f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\mu_n, \ldots, \mu_0)$ as follows: Set $\mu_0 = 1$.

For $1 \leq k \leq n$ do

(i) Generate $r(k)$ random numbers $u_1, \ldots, u_{r(k)}$ from the uniform distribution $U[0, \mu_{k-1}]$.
(ii) Create an ordered list $u_{(1)}, \ldots, u_{(r(k))}$ from $u_1, \ldots, u_{r(k)}$.
(iii) Set $\mu_k = u_{(r(k))}$.

end
Set $\bar{\mu}^j = (\mu_n, \ldots, \mu_0)$.

The result in a stochastic estimate of the expectation given by

$$\langle g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] \rangle \approx \bar{g} \pm \sqrt{\mathrm{var}(\bar{g})}. \tag{40}$$

An important special case of the expectation $\langle g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] \rangle$ is the evaluation of the evidence Z_s. In this case, one chooses

$$g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] = \sum_{i=0}^{n-1} L_i \times [\mu(\Omega_i) - \mu(\Omega_{i+1})] + L_n \times \mu(\Omega_n) \tag{41}$$

$$= \sum_{i=0}^{n} L_i \times \mu(A_i) \tag{42}$$

Note that approximations of Z_s using $g[\mu(\Omega_n), \ldots, \mu(\Omega_0)]$ [defined by Eq. (41)], and quantified by Monte Carlo estimates [as in Eqs. (38) and (39)] are stochastic in nature, i.e.,

$$\overbrace{Z_s}^{\text{stochastic}} \approx \sum_{i=0}^{n} \underbrace{L_i}_{\text{deterministic}} \times \overbrace{\mu(A_i)}^{\text{stochastic}}, \tag{43}$$

examples of which will be given in Sec. VI. Other functions $g(\bullet)$ can be defined for Bayesian parameter estimation, examples of which will be given in Sec. V.

D. Determining the nested sampling parameters

The nested sampling algorithm requires the specification of the number of nested sampling iterations n, as well as the number of samples $r(i)$ drawn at every iteration i. Examining how changes in n and $r(i)$ influence the resulting ordered partition $0 = L_0 < \cdots < L_n$ and distribution $f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\mu_n, \ldots, \mu_0)$ is important in determining the utility of the nested sampling algorithm in order to approximate the evidence Z_s.

Increasing n alone does not result in a point-wise monotonic convergence of $S^n(\mathbf{w}_s)$ to $L(\mathbf{w}_s)$. Thus $n \to \infty$ alone is not a sufficient condition for the nested sampling algorithm to estimate the evidence Z_s exactly [even assuming that $\mu(A_i)$ can be determined without error]. Letting $r(i) \to \infty$ results in an exhaustive sampling of $p^C_{\Omega_i}(\mathbf{w}_s)$ at each iteration i, which in turn generates a dense ordered partition $0 = L_0 < \cdots < L_n$. Thus the combined conditions $r(i) \to \infty$, $n \to \infty$, $n > \max[r(i)]$ will result in the nested sampling algorithm defining $S^n(\mathbf{w}_s)$ which has point-wise monotonic convergence to $L(\mathbf{w}_s)$.

Consider the approximate value for the measure

$$\log[\mu(\Omega_n)] \approx -\sum_{i=0}^{n-1} \frac{1}{r(i)} \pm \sqrt{\sum_{i=0}^{n-1} \left[\frac{1}{r(i)}\right]^2}, \tag{44}$$

after n iterations of the nested sampling algorithm which is derived in Appendix D. The relation

$$\lim_{r(i)\to\infty} \pm \sqrt{\sum_{i=0}^{n-1} \left[\frac{1}{r(i)}\right]^2} = 0, \tag{45}$$

implies that $r(i) \to \infty$ results in an exact estimate for $\log[\mu(\Omega_n)]$, and as such $r(i) \to \infty$, $n \to \infty$, $n > \max[r(i)]$ are sufficient conditions for the Nested Sampling algorithm to produce an exact estimate of the evidence Z_s. A drawback of increasing $r(i)$ is that

$$\lim_{r(i)\to\infty} \log[\mu(\Omega_i)] \approx \lim_{r(i)\to\infty} -\sum_{i=0}^{n-1} \frac{1}{r(i)} = 0, \tag{46}$$

which implies that

$$\mu(\Omega_n) \approx 1 = \mu(\Omega_0). \tag{47}$$

Thus large values for $r(i)$ result in a slow reduction in the volume or measure of the parameter space at each iteration, with a trade off of increasing accuracy in assigning the measures $\mu(\Omega_n)$. As discussed in Sec. IV, typical likelihood functions are concentrated in small volumes or measures of the parameter space. An optimal choice for $r(i)$ will result in a fast reduction in the measure of the parameter space at each iteration i which quickly isolates the dominant region of the likelihood function in addition to constructing a sufficiently dense ordered partition $0 = L_0 < \cdots < L_n$, which will estimate the evidence Z_s to a required degree of accuracy. The ability to use small values for $r(i)$ in practice results in nested sampling being a viable algorithm useful for estimating the evidence Z_s as will be shown in Sec. VI.

Given Eq. (44) and assuming that $r(i)$ have been determined for $0 \leq i \leq n$, a robust stopping criteria would be to choose a value of n for a fixed $\log(\delta)$ such that

$$-\sum_{i=0}^{n-1} \frac{1}{r(i)} < \log(\delta), \tag{48}$$

where $0 < \delta < 1$ would represent an upper bound on the remaining measure or volume of the parameter space which to explore. One would then perform m additional iterations until

$$L_{n+m} \times \mu(\Omega_{n+m}) < \epsilon, \tag{49}$$

for a fixed value of ϵ. This combined stopping criteria will be less susceptible to errors due to plateaus in the likelihood function as a predefined measure or volume of the parameter space is always explored, although as with any numerical algorithm choosing an optimal stopping criteria will likely be problem-specific.

E. Implementation details

Efficiently generating samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ from $p^C_{\Omega_i}(\mathbf{w}_s)$ is key to implementing the nested sampling

algorithm. The presence of high dimensional parameter spaces $\mathbf{w}_s$ in many acoustics applications suggests that a Markov Chain Monte Carlo (MCMC) approach to generate the samples would be the best candidate. Choosing an appropriate proposal distribution is crucial to efficient sampling with MCMC. As the volume $\log(\mu(\Omega_i))$ is reduced by an approximate factor of $-1/r(i)$ with each iteration i of the nested sampling algorithm, using a fixed MCMC proposal distribution for all iterations will likely prove inefficient. A heuristic approach which scales the log-volume of MCMC proposal distribution by a factor of $-1/r(i)$ at each iteration i will potentially improve MCMC sampling efficiency. The slice sampling MCMC algorithm[14,32] which has the ability to automatically tune a proposal distribution in order to correct for cases where scaling may lead to inefficient sampling, seems particularly suited for nested sampling.[12] Any MCMC approach used will be problematic in that the samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ generated will be *dependent* (although the dependence between samples can be minimized by choosing samples which are uncorrelated). As the proofs in the Appendixes rely on *independent* samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$ using a MCMC approach violates the assumptions at the heart of the nested sampling algorithm. While MCMC sampling (and slice sampling, in particular) cannot generate the required independent samples $\mathbf{w}_s^1, \ldots, \mathbf{w}_s^{r(i)}$, it is likely to be the only feasible approach in higher dimensional parameter spaces $\mathbf{w}_s$. As such, the samples generated from a MCMC approach are used as if they were *independent*.

V. PARAMETER ESTIMATION BY NESTED SAMPLING

Parameter estimates can be determined through the moments of the posterior distribution, which can be defined by the expected values

$$\langle h(\mathbf{w}_s)\rangle = \frac{\int_{\mathbf{w}_s} h(\mathbf{w}_s)L(\mathbf{w}_s)p(\mathbf{w}_s)\,d\mathbf{w}_s}{\int_{\mathbf{w}_s} L(\mathbf{w}_s)p(\mathbf{w}_s)\,d\mathbf{w}_s}. \tag{50}$$

Given samples $\{\mathbf{v}_s^0, \ldots, \mathbf{v}_s^n\}$ generated by the nested sampling algorithm, let $h_i = h(\mathbf{v}_s^i)$. Define the function

$$g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] = \frac{\sum_{i=0}^{n} h_i L_i \mu(A_i)}{\sum_{i=0}^{n} L_i \mu(A_i)} \tag{51}$$

$$= \frac{\sum_{i=0}^{n-1} h_i L_i[\mu(\Omega_i) - \mu(\Omega_{i+1})] + h_n L_n(\Omega_n)}{\sum_{i=0}^{n-1} L_i[\mu(\Omega_i) - \mu(\Omega_{i+1})] + L_n\mu(\Omega_n)}. \tag{52}$$

As in the case of the simple function approximation to the evidence Z_s one has

$$g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] \approx \langle h(\mathbf{w}_s)\rangle \tag{53}$$

(by a simple change in the proof given in Appendix A), from which it follows that $g[\mu(\Omega_n), \ldots, \mu(\Omega_0)]$ and the nested samples $\{\mathbf{v}_s^0, \ldots, \mathbf{v}_s^n\}$ can provide an estimate of $\langle h(\mathbf{w}_s)\rangle$, and solve the Bayesian parameter estimation problem. Important examples of $g[\mu(\Omega_n), \ldots, \mu(\Omega_0)]$ include defining estimates of the mean values of the kth parameter $\hat{w}_s^k$ using

$$h_i = h(\mathbf{v}_s^i) = \nu_s^{ik}, \tag{54}$$

and defining estimates of the jth row, kth column of the covariance matrix $\langle \hat{C}\rangle = \langle \hat{C}_{jk}\rangle$ using

$$h_i = h(\mathbf{v}_s^i) = (\mathbf{v}_s^{rj} - \hat{w}_s^j)(\mathbf{v}_s^{r\,k} - \hat{w}_s^k). \tag{55}$$

As discussed in Sec. IV C, parameter estimates will be stochastic in nature and can be qualified by Monte Carlo approximations given by Eqs. (38) and (39). Given a covariance matrix, the individual variance σ_j^2 and the standard deviations σ_j of the jth decay parameter can be determined by calculating eigenvalues of the co-variance matrix $\langle \hat{C}\rangle$ as discussed in Ref. 20. The inter-relationship between decay parameters is measured by the cross-correlation coefficient

$$\mathrm{ccc}(j,k) = \hat{C}_{jk}/\sqrt{\hat{C}_{jj}\hat{C}_{kk}}. \tag{56}$$

VI. EXPERIMENTAL RESULTS

This section presents an experimental example which is carried out in a one-eighth scale model of two coupled acoustic spaces in order to illustrate the application of the nested sampling algorithm in acoustical measurements. When given in the original scale, the primary room which contains a dodecahedral sound source measures 7.2 m $\times$ 6.32 m $\times$ 4.88 m, while the secondary room measures 7.2 m $\times$ 9.76 m $\times$ 7.6 m. An aperture (opening area) with size of 7.2 m $\times$ 0.6 m couples the primary room with the secondary room. The interior surfaces of two rooms are treated to feature diffuse reflections. The measurement of the room impulse response in the primary room is accomplished using maximum-length sequences where the measured data is bandpass filtered at 1 kHz (octave band).

Prior to any acoustic measurements, little prior information is available to determine whether the energy decay function $\mathbf{D}$ in this frequency band supports a single-slope model or a multiple-slope model. The prior information is such that the energy decays in closed/separated rooms (single-space cases) should be well characterized by a single (natural) reverberation time for each room, respectively, since most interior surfaces are featured with highly diffusively reflecting materials to create diffuse sound fields. In each of the separated rooms, the sound energy decays are most likely expected to exhibit single-rate decay characteristics. When the separated rooms are coupled to each other by the aperture, one would not expect that the number of decay slopes would be greater than three. At this stage, there is no prior information as to what decay rates are present. This specific data analysis task exemplifies a typical application of both the first level and the second level of Bayesian inference:

Which of the expected three models, the single-slope, the double-slope, or the triple-slope decay model do the data support? At this inference level, values of the decay rates are not of interest at all until the specific model is identified. To highlight the competing nature of the expected decay models, Fig. 3 illustrates the normalized Schroeder decay curves via an experimentally measured room impulse response using the scale-model coupled-rooms. Both the second order ($s=2$) and the third order ($s=3$) decay models seem to describe the experimental data well, where the third decay order model seems to improve the curve-fitting slightly.

The nested sampling algorithm was applied to the experimental data using the first ($s=1$), second ($s=2$), and third ($s=3$) order decay models. The prior distribution used for each model is given in Table I with the algorithm parameters $r(i)=8$ and 293 iterations for all three models. The results of the nested sampling algorithm applied to the second order decay ($s=2$) model are used for illustrative purposes.

From Eq. (48) it was determined that 293 iterations of the nested sampling algorithm would reduce the volume of the unexplored parameter space of the posterior probability density function by a factor of 1×10^{-15} which was deemed adequate based on prior experience with energy decay distributions. Figure 4(a) shows the log likelihood $\log(L_i)$ for each iteration i of the nested sampling algorithm. The $\log(L_i)$ values in Fig. 4(a) climb from ca. -1240 neper at the $i=1$st iteration of the algorithm up to ca. 1500 neper at the $i=125$th iteration, at which point $\log(L_i)$ remains relatively constant, as is shown in Fig. 4(b). Two hundred samples $(\bar{\mu}^1, \ldots, \bar{\mu}^{200})$ were generated from the probability density function

$$f_{\mu(\Omega_n),\ldots,\mu(\Omega_0)}(\mu_n, \ldots, \mu_0) \tag{57}$$

following the procedure described in Sec. IV C. Each sample $\bar{\mu}^t$ for $1 \le t \le 200$ represents a sequence of measures

$$[\mu(\Omega_0), \ldots, \mu(\Omega_{293})], \tag{58}$$

given 293 iterations of the nested sampling algorithm (as an example, samples $\bar{\mu}_{10}$ and $\bar{\mu}_{25}$ are shown as a plot of $\log[\mu(\Omega_i)]$ for each iteration i in Fig. 5). The function

$$g[\mu(\Omega_n), \ldots, \mu(\Omega_0)] = \sum_{i=0}^{292} L_i \times [\mu(\Omega_i) - \mu(\Omega_{i+1})] + L_{293} \times \mu(\Omega_{293}) \tag{59}$$

was calculated for each sample $\bar{\mu}^t$ using the L_i values. The collected set of $g[\mu(\Omega_n), \ldots, \mu(\Omega_0)]$ values for the samples $\bar{\mu}^t$ are plotted in a histogram, shown in Fig. 6(c), and represent a Monte Carlo approximation of the pair distribution function for the nested sampling estimate of $\log(Z_s)$. The estimate of $\log(Z_s)$ was quantified by Monte Carlo approximations given

TABLE I. Uniform prior parameters for three different decay models.

Model		A_0	A_1	T_1 (s)	A_2	T_2 (s)	A_3	T_3 (s)
1st order	Min	$1E-7$	0.01	0.1	—	—	—	—
	Max	$1E-5$	0.1	1.0	—	—	—	—
2nd order	Min	$1E-7$	0.01	0.1	0.001	0.4	—	—
	Max	$1E-5$	0.1	1.0	0.05	2.5	—	—
3rd order	Min	$1E-7$	0.01	0.1	0.001	0.4	0.0001	0.6
	Max	$1E-5$	0.1	1.0	0.05	2.5	0.005	3.5

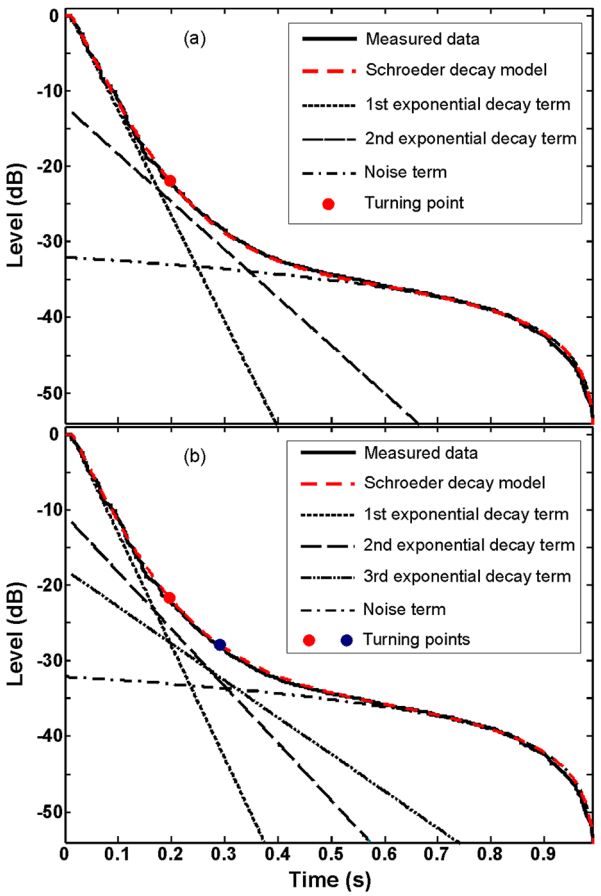

FIG. 3. (Color online) Experimentally measured Schroeder decay curves compared with the double-slope and triple-slope decay model curves, along with decomposed exponential decay-slopes, the noise term, and turning points. (a) Comparison with the double-slope decay model. (b) Comparison with the triple-slope decay model.

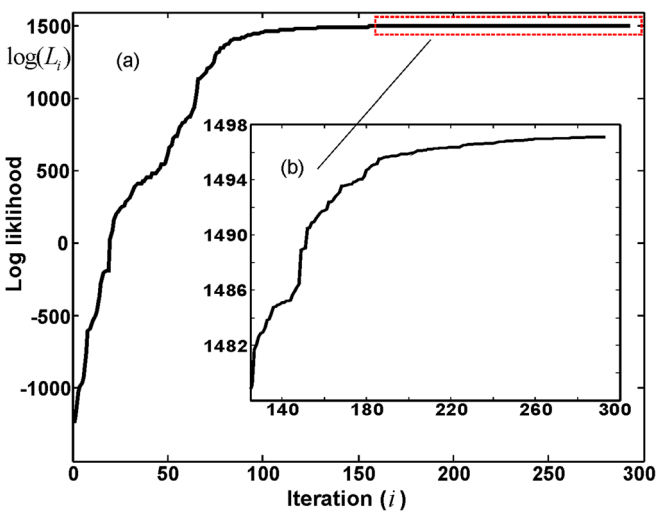

FIG. 4. (Color online) Log likelihood $\log(L_i)$ values increasing with iteration i using a double slope ($s=2$) decay model defined using experimentally measured data as described in Sec. VI. (a) Entire course of $\log(L_i)$ for iterations $1 \le i \le 293$. (b) Magnified segment of $\log(L_i)$ between iterations $125 \le i \le 293$.

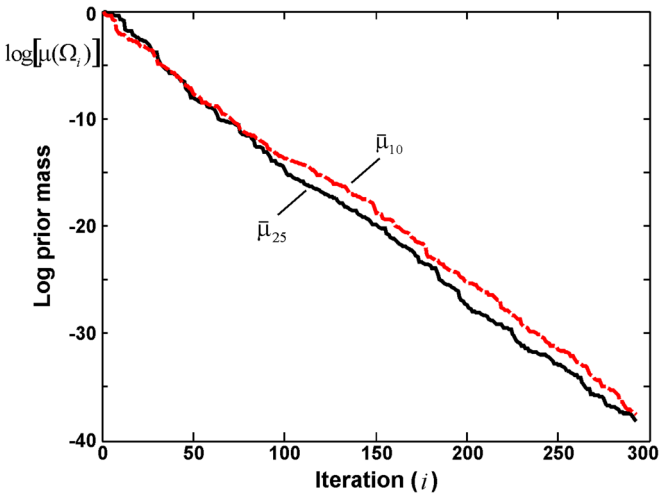

FIG. 5. (Color online) $\log[\mu(\Omega_i)]$ values for samples $\bar{\mu}_{10}$ and $\bar{\mu}_{25}$ decreasing with iteration i calculated for a double slope ($s=2$) decay model defined using experimentally measured data as described in Sec. VI. Each sample $\bar{\mu}_{10}$ and $\bar{\mu}_{25}$ represents a sequence of measures $[\mu(\Omega_0),\ldots,\mu(\Omega_{293})]$ for the 293 iterations of the nested sampling algorithm.

by Eqs. (38) and (39) with a mean value of 1475.5 neper and variance 2.55 neper, also shown in Fig. 6(c).

The procedure described in the previous paragraph was applied to the first order ($s=1$) and third order ($s=3$) decay models using the parameters given in Tables I. The resulting histograms representing estimates of $\log(Z_s)$ of each model are superimposed in Fig. 6(a) and shown individually in Figs. 6(b)–6(d). Figure 6(a) shows a sufficient separation in the histograms to assume that the nested sampling algorithm was successful in discriminating between the three models. The mean value of $\log(Z_s)$ estimate for the second order ($s=2$) decay model is approximately 25 neper higher than that of the single order ($s=1$) model. When increasing the decay order to three ($s=3$), the mean value of the $\log(Z_s)$ estimate declines significantly to a value of approximately

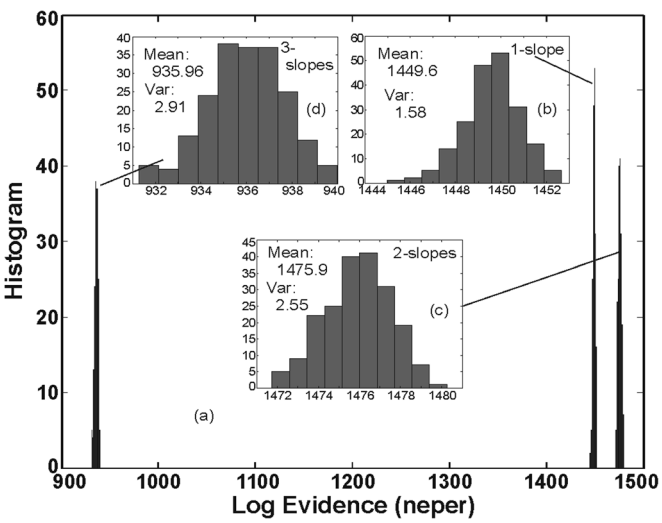

FIG. 6. Superimposed histograms of estimates for $\log(Z_s)$ each generated from 200 samples $\bar{\mu}^t$ applied to the single-slope ($s=1$), double-slope ($s=2$), and triple-slope ($s=3$) decay model evaluated from an acoustically measured Schroeder decay function in a coupled-volume system described in Sec. VI.

TABLE II. Mean values for decay parameter estimates and error bars for the double slope ($s=2$) model evaluated from an acoustically measured Schroeder decay function in a coupled-volume system described in Sec. VI.

	Double-slope parameters
A_0(dB)	−62.01
A_1(dB)	−5.15 (±1.9E - 4)[a]
T_1(s)	0.369 (±2.1E - 4)
A_2(dB)	−14.91 (±1.1E - 4)[a]
T_2(s)	0.947 (±2.9E - 3)

[a] dA_1,dA_2: listed linearly.

925 neper. Occam's razor, implicitly encapsulated in the Bayesian evidence has successfully penalized the over-parameterized third order ($s=3$) decay model. After ranking the three decay models, the second-slope model survives from the competing alternatives. The decay parameters and covariance matrix given the second order ($s=2$) model were then estimated using Eqs. (54) and (55). The mean values of the parameter and covariance matrix estimates were determined by Monte Carlo approximation given by Eq. (38). The error bars for parameter estimates were calculated from the estimated covariance matrix. The parameter estimates and error bars for the second order ($s=2$) model are shown in Table II.

VII. SUMMARY AND FUTURE WORK

This paper has demonstrated that the nested sampling algorithm can successfully discriminate the number of energy decays present in acoustically coupled spaces within the Bayesian framework. Estimating the number of energy decays is an example of two levels of Bayesian inference often encountered in the architectural acoustics practice: decay order estimation, being a model selection problem corresponding to the higher (2nd) level of inference, and decay parameter estimation, being parameter estimation problem corresponding to the 1st level of inference. The brief formulation of the two levels of Bayesian inference following a top-down approach (from the higher level to the lower level) is presented in Sec. III which discusses the importance of the Bayesian evidence Z_s.

This paper presents the basics of the Lebesgue integral in Sec. IV A through the simple function approximation and then demonstrates how the simple function approximation can define a numerical algorithm which can be used in order to evaluate the Bayesian evidence Z_s. Separating the concept of the simple function approximation from the nested sampling algorithm allows for an alternative view toward understanding the theoretical basis of the nested sampling algorithm which may allow acousticians to apply or extend the algorithm to other problem domains.

A topic of future research is to explore alternative methods to nested sampling which have been developed in order to estimate measures or volumes of sets (see Dyer *et al.*,[33] for example) which could be used with the simple function approximation described in Sec. IV A.

The experimental example demonstrates the function of Occam's razor within the Bayesian framework by penalizing the over-parameterized (three-slope) model while choosing the

second order model over the simpler first order model. While evaluating the Bayesian evidence Z_s, the nested samples are all stored in memory, and a straightforward calculation allows for these samples to be used for decay parameter estimation.

An important open problem of the nested sampling algorithm is how to choose values of $r(i)$. The experimental example in Sec. VI used a fixed value for $r(i)$ for all iterations which is likely to be a non-optimal approach. The authors believe that choosing $r(i)$ optimally will not be a simple task. Similar to many numerical algorithms, the selection of $r(i)$ involves a tradeoff between computational efficiency and accuracy of the algorithm as discussed in Sec. IV E. In order to reduce the amount of user tuning required, an adaptive approach to choosing $r(i)$ is one possible area of future research.

ACKNOWLEDGMENTS

The authors wish to thank Dr. John Skilling for his constructive comments on an early stage of this work, Dr. Paul Goggans, Jonathan Botts, and Cameron Fackler for their stimulating discussions, and Jonathan Botts and Cameron Fackler for proofreading.

APPENDIX A: SIMPLE FUNCTION AND LEBESGUE INTEGRATION

For completeness this appendix outlines the concept of the simple function and Lebesgue integration, interested readers are refereed to the text by Dudley[31] for detailed explanations of the notation and terminology used.

Definition: Given a sequence $a_n : \mathbb{N} \to \mathbb{R}$ and $b \in \mathbb{R}$. If $\lim_{n\to\infty} a_n = b$ and $a_n \leq b$ for all n, then a_n converges to b in a monotonic manner denoted by $a_n \uparrow b$.

Definition: Given a simple function $S^n : \mathbb{R}^m \to \mathbb{R}$ and $f : \mathbb{R}^m \to \mathbb{R}$. If $\lim_{n\to\infty} S^n(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x}$, and $S^n(\mathbf{x}) \leq f(\mathbf{x})$ for all $\mathbf{x}$ and n, then $S^n(\mathbf{x})$ converges to $f(\mathbf{x})$ in a point-wise monotonic manner denoted by $S^n(\mathbf{x}) \uparrow f(\mathbf{x})$.

Theorem 1: *Given* $(\mathbb{R}^n, \mathcal{B}, \mu)$, *where* $\mathcal{B}$ *are Borel sets, and* μ *a measure. Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a measurable function such that* $f(\mathbf{x}) \geq 0$. *For any sequence of measurable functions* $f_n : \mathbb{R}^n \to \mathbb{R}$ *such that* $f(\mathbf{x}) \geq 0$ *and* $f_n(\mathbf{x}) \uparrow f(\mathbf{x})$ *one has* $\int f_n(\mathbf{x})\, d\mu \uparrow \int f_n(\mathbf{x})\, d\mu$.

Proof: See Ref. 31. ■

Theorem 2: *Given* $(\mathbb{R}^n, \mathcal{B})$, $d\mathbf{x}$, *where* $\mathcal{B}$ *are Borrel sets, and* $d\mathbf{x}$ *a Lebesgue measure. Let* $p : \mathbb{R}^n \to \mathbb{R}$ *be a measurable function such that* $p(\mathbf{x}) \geq 0$ *and define* $\mu(X) = \int 1_X(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}$ *for* $X \subset \mathcal{B}$. *For any measurable function* $L : \mathbb{R}^n \to \mathbb{R}$ *such that* $L(\mathbf{x}) \geq 0$, *and a sequence of simple functions* $S^n(\mathbf{x})$ *such that* $0 \leq S^n(\mathbf{x}) \uparrow L(\mathbf{x})$ *one has* $\int S^n(\mathbf{x})\, d\mu \uparrow \int L(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}$.

Proof:

$$\int S^n(\mathbf{x})\, d\mu = \sum_{i=0}^{n} L_i \mu(A_i) = \sum_{i=0}^{n} L_i \int 1_{A_i}(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \tag{A1}$$

$$= \int \Big(\sum_{i=0}^{n} L_i 1_{A_i}(\mathbf{x}) \Big) p(\mathbf{x})\, d\mathbf{x} \tag{A2}$$

$$= \int S^n(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}. \tag{A3}$$

Now

$$S^n(\mathbf{x}) \uparrow L(\mathbf{x}) \Rightarrow S^n(\mathbf{x}) p(\mathbf{x}) \uparrow L(\mathbf{x}) p(\mathbf{x}). \tag{A4}$$

Thus, by Theorem 1

$$\int S^n(\mathbf{x})\, d\mu = \int S^n(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \uparrow \int L(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}, \tag{A5}$$

which implies

$$\sum_{i=0}^{n} L_i \mu(A_i) \uparrow \int L(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x}. \tag{A6}$$

■

APPENDIX B: NESTED SAMPLING AND MEASURES OF SETS

Consider the set $\Omega_i = \{\mathbf{w}_s | L(\mathbf{w}_s) > L_i\}$ as shown in 7(a) with measure $\mu_i = \mu(\Omega_i)$. Define a function μ: $L(\mathbf{w}_s) \to [0,1]$ with $\mu(L_i) = \mu(\Omega_i)$ as shown in Fig. 7(b). A random sample $\mathbf{w}_s^1$ generated from $p_{\Omega_i}^C(\mathbf{w}_s)$ defines the random variables $L_1 = L(\mathbf{w}_s^1)$ and $\mu_1 = \mu(L_1)$ as shown in Figs. 7(c) and 7(d). The cumulative density function for the measure μ_1 is given by

$$F_{\mu_1|\mu_i}(a) = \mathbb{P}[\mu_1 \leq a | \mu_1 \leq \mu_i] \tag{B1}$$

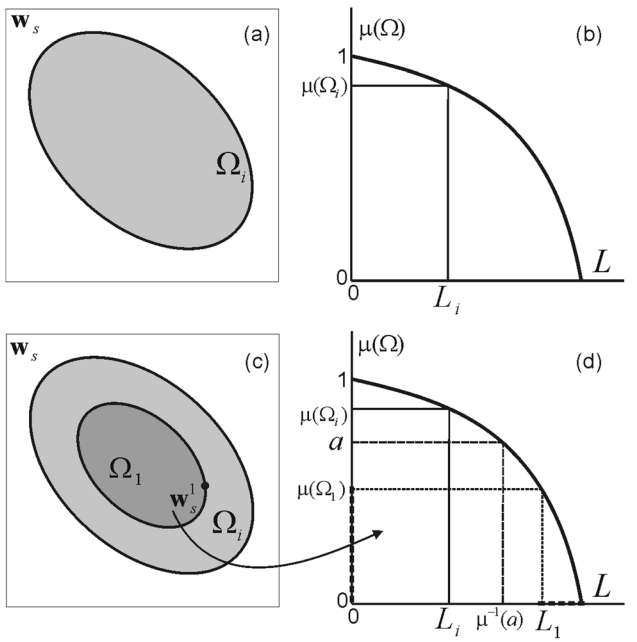

FIG. 7. Given the set Ω_i, the associated likelihood L_i and measure $\mu(\Omega_i)$ are shown in (a). Generate a sample $\mathbf{w}_s^1$ from the constrained prior probability distribution $p_{\Omega_i}^C(\mathbf{w}_s)$ as shown in (b). From sample $\mathbf{w}_s^1$ with the associated likelihood value L_1, one has $L_1 \geq L_i, \Omega_1 \subset \Omega_i$, and $\mu(\Omega_1) < \mu(\Omega_i)$ as shown in (b). The shaded region on the horizontal axis defines values for which $\mu(\Omega_1) \leq a$ along with a corresponding shaded region on the horizontal axis which defines values for which $L_1 \geq \mu^{-1}(a)$.

$$= \begin{cases} 0 & \text{if} \quad a < 0 \\ \dfrac{\mathbb{P}[\mu_1 \le a]}{\mathbb{P}[\mu_1 \le \mu_i]} & \text{if} \quad 0 \le a \le \mu_i \\ 1 & \text{if} \quad a > \mu_i, \end{cases} \tag{B2}$$

where $\mathbb{P}[A]$ denotes the probability of the event A. As $\mu(\cdot)$ is a monotonically decreasing function, the inverse function $\mu^{-1}(\cdot)$ is well defined which results in

$$\frac{\mathbb{P}[\mu_1 \le a]}{\mathbb{P}[\mu_1 \le \mu_i]} = \frac{\mathbb{P}[L_1 > \mu^{-1}(a)]}{\mathbb{P}[L_1 > L_i]}, \tag{B3}$$

as shown in Fig. 7(d). Given

$$\mathbb{P}[L_1 > \mu^{-1}(a)] = \mu^{-1}(\mu(a)) = a, \tag{B4}$$

$$\mathbb{P}[L_1 > L_i] = \mu_i, \tag{B5}$$

the cumulative density function and probability density functions are given by

$$F_{\mu_1|\mu_i}(a) = \begin{cases} 0 & \text{if} \quad a < 0 \\ \dfrac{a}{\mu_i} & \text{if} \quad 0 \le a \le \mu_i \\ 1 & \text{if} \quad a > \mu_i, \end{cases} \tag{B6}$$

$$F_{\mu_1|\mu_i}(a) = \begin{cases} \dfrac{1}{\mu_i} & \text{if} \quad 0 \le a \le \mu_i \\ 0 & \text{otherwise}, \end{cases} \tag{B7}$$

which implies that μ_1 is (by definition) a uniformly distributed random variable $U[0, \mu_i]$. Thus generating a sample $\mathbf{w}_s^1$ from $p^C_{\Omega_i}(\mathbf{w}_s)$ is equivalent to generating a sample μ_1 from the uniform distribution $U[0, \mu_i]$.

Let $\mathbf{w}_s^1,\ldots,\mathbf{w}_s^{r(i)}$ be a set of independent samples randomly generated $p^C_{\Omega_i}(\mathbf{w}_s)$, which define independent random variables $\mu_1=\mu(\Omega_1), \ldots, \mu_{r(i)} = \mu(\Omega_{r(i)})$ generated from the uniform distribution $U[0, \mu_i]$. Define an ordered list $\mu_{(r(i))} < \cdots < \mu_{(1)}$ from the set $\mu_1, \ldots, \mu_{r(i)}$. The random variable $\mu_{(1)}$ is an order statistic[34] with probability density function

$$f_{\mu_{(1)}|\mu_i}(a) = r(i) f_{\mu_1|\mu_i}(a)[F_{\mu_1|\mu_i}(a)]^{r(i)-1}. \tag{B8}$$

Substituting Eqs. (B6) and (B7) into Eq. (B8) results in

$$f_{\mu_{(1)}|\mu_i}(a) = \begin{cases} \dfrac{r(i)}{\mu_i}\left[\dfrac{a}{\mu_i}\right]^{r(i)-1} & \text{if} \quad 0 \le a \le \mu_i \\ 0 & \text{otherwise}, \end{cases} \tag{B9}$$

with

$$\langle \log(\mu_{(1)}|\mu_i)\rangle = \log(\mu_i) - \frac{1}{r(i)}, \tag{B10}$$

$$\mathrm{var}\{\log(\mu_{(1)})|\mu_i\} = \log(\mu_i) + \left[\frac{1}{r(i)}\right]^2. \tag{B11}$$

APPENDIX C: VOLUME REDUCTION FACTOR

The reduction in the volume of the parameter space Ω after n iterations of nested sampling is given by a factor of

$$\mu(\Omega_n) = \prod_{i=0}^{n-1} \frac{\mu(\Omega_{i+1})}{\mu(\Omega_i)}, \tag{C1}$$

or

$$\log[\mu(\Omega_n)] = \sum_{i=0}^{n-1} \log[\mu(\Omega_{i+1})] - \log[\mu(\Omega_i)]. \tag{C2}$$

From Eqs. (B10) and (B11) in Appendix B,

$$\langle \log[\mu(\Omega_n)]\rangle = -\sum_{i=0}^{n-1} \frac{1}{r(i)} \tag{C3}$$

and

$$\mathrm{var}\{\log[\mu(\Omega_n)]\} = \sum_{i=0}^{n-1} \left[\frac{1}{r(i)}\right]^2. \tag{C4}$$

Using Eqs. (C3) and (C4), the probability density function of $\log[\mu(\Omega_n)]$ can be approximated (using a form of the Lindeberg central limit theory) by the normal density

$$N\left\{-\sum_{i=0}^{n-1}\frac{1}{r(i)}, \sum_{i=0}^{n-1}\left[\frac{1}{r(i)}\right]^2\right\}. \tag{C5}$$

which results in

$$\log[\mu(\Omega_n)] \approx -\sum_{i=0}^{n-1}\frac{1}{r(i)} \pm \sqrt{\sum_{i=0}^{n-1}\left[\frac{1}{r(i)}\right]^2}. \tag{C6}$$

[1] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," J. Acoust. Soc. Am. **113**, 2685–2697 (2003).

[2] J. Dettmer and S. E. Dosso, "Bayesian evidence computation for model selection in non-linear geoacoustic inference problems," J. Acoust. Soc. Am. **128**, 3406–3415 (2010).

[3] D. Battle, P. Gerstoft, W. S. Hodgkiss, and W. A. Kuperman, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).

[4] N. Xiang, P. Goggans, T. Jasa, and P. Robinson, "Characterization of sound energy decays in multiple coupled-volume systems," J. Acoust. Soc. Am. **129**, 741–752 (2011).

[5] J. Dettmer, Ch. W. Holland, and S. E. Dosso, "Analyzing lateral seabed variability with Bayesian inference of seabed reflection data," J. Acoust. Soc. Am. **126**, 56–69 (2009).

[6] J. J. Remus and L. M. Collins, "Comparison of adaptive psychometric procedures motivated by the Theory of Optimal Experiments: Simulated and experimental results," J. Acoust. Soc. Am. **123**, 315–326 (2008).

[7] S. E. Dosso and M. J. Wilmut, "Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion," J. Acoust. Soc. Am. **124**, 82–89 (2008).

[8] C. Yardim, P. Gerstoft, and W. S. Hodgkiss, "Tracking of geoacoustic parameters using Kalman and particle filters," J. Acoust. Soc. Am. **125**, 764–760 (2009).

[9] S. E. Dosso, P. L. Nielsen, and Ch. H. Harrison, "Bayesian inversion of reverberation and propagation data for geoacoustic and scattering parameters," J. Acoust. Soc. Am. **125**, 2867–2880 (2009).

[10]Y.-M. Jiang and N. R. Chapman, "The impact of ocean sound speed variability on the uncertainty of geoacoustic parameter estimatesm," J. Acoust. Soc. Am. **125**, 2881–2895 (2009).
[11]G. Kim, Y. Lu, Y. Hu, and Ph. C. Loizoua, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am. **126**, 1486–1494 (2009).
[12]T. Jasa and N. Xiang, "Using nested sampling in the analysis of multi-rate sound energy decay in acoustically coupled rooms," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, editied by K. H. Knuth, A. E. Abbas, R. D. Morris, and J. P. Castle (AIP, Melville, NY, 2005), Vol. 803, pp. 189–196.
[13]N. Xiang, and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1424 (2001).
[14]T. Jasa and N. Xiang, "Efficient estimation of decay parameters in acoustically coupled spaces using slice sampling," J. Acoust. Soc. Am. **126**, 1269–1279 (2009).
[15]C. C. Anderson, A. Q. Bauer, M. R. Holland, M. Pakula, P. Laugier, G. L. Bretthorst, and J. G. Millera, "Inverse problems in cancellous bone: Estimation of the ultrasonic properties of fast and slow waves using Bayesian probability theory," J. Acoust. Soc. Am. **128** 2940–2948 (2010).
[16]H.-J. Pu, X.-J. Qiu, and J.-Q. Wang, "Different sound decay patterns and energy feedback in coupled volumes," J. Acoust. Soc. Am. **129**, 1972–1980 (2011).
[17]J. Skilling, "Nested sampling," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by R. Fisher, R. Preuss, and U. von Toussant (AIP, Melville, NY, 2004), Vol. 735, pp. 395–405.
[18]J. Skilling, "Nested sampling for Bayesian computations," in *Proceedings of the 8th World Meeting on Bayesian Statistics*, Alicante, Spain (June 2006).
[19]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[20]N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayeisan decay time estimation," J. Acoust. Soc. Am. **117**, 3707–3715 (2005).
[21]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**, 3744–3749 (2006).
[22]D. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, 2002), Chap. 28.
[23]W. H. Jefferys and J. O. Berger, "Ockhams razor and Bayesian analysis," Am. Sci. **80**, 64–72 (1992).
[24]R. E. Kass and A. E. Raftery, "Bayes factors," J. Am. Stat. Assoc. **90**, 773–795 (1995).
[25]G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation* (Springer, New York, 1988), Chap. 5.
[26]M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," J. Am. Stat. Assoc. **96**, 746–774 (2001).
[27]P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*, (Cambridge University Press, Cambridge, 2005), Chap. 8.
[28]C. C. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, New York, 1999), Chap. 3.3.
[29]J. J. K. O Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer, New York, 1996), Chap. 4.9.
[30]R. M. Neal, "Annealed importance sampling," Stat. Comput. **11**, 125–139 (2001).
[31]R. M. Dudley, *Real Analysis and Probability* (Cambridge University Press, Cambridge, 2002), Chaps. 3 and 4, pp. 85–146.
[32]R. M. Neal, "Slice sampling," Ann. Stat. **31**, 706–767 (2003).
[33]M. Dyer, Alan Frieze, and Ravi Kannan, "A random polynomial-time algorithm for approximating the volume of convex bodies," J. ACM **38**, 1–17 (1991).
[34]C. Rose and M. D. Smith, "Order statistics," in *Mathematical Statistics with Mathematica* (Springer, New York, 2002), Chap. 9.4.