공학석사 학위논문

부족한 통계적 정보 하에서 베이지안 보정을 위한 새로운 초기값 탐색 알고리즘: 탐욕스러운 확률론적 분할 탐색 방법

A New Initial Point Search Algorithm for Bayesian Calibration with Insufficient Statistical Information: Greedy Stochastic Section Search

2022 년 2 월

서울대학교 대학원

기계공학부

이 현 찬

부족한 통계적 정보 하에서 베이지안 보정을 위한 새로운 초기값 탐색 알고리즘: 탐욕스러운 확률론적 분할 탐색 방법

A New Initial Point Search Algorithm for Bayesian Calibration with Insufficient Statistical Information: Greedy Stochastic Section Search

지도 교수 윤 병 동

이 논문을 공학석사 학위논문으로 제출함

2021 년 10 월

서울대학교 대학원

기계공학부

이 현 찬

이현찬의 공학석사 학위논문을 인준함

2021 년 12 월

위 원 장 김 윤 영 (인)

부위원장 윤 병 동 (인)

위 원 김 도 년 (인)

Abstract

A New Initial Point Search Algorithm for Bayesian Calibration with Insufficient Statistical Information: Greedy Stochastic Section Search

Hyeonchan Lee
Department of Mechanical Engineering
The Graduate School
Seoul National University

Computer-Aided Engineering (CAE) models are essential in the product design, operation, and maintenance processes. However, CAE models have not only aleatory (irreducible) uncertainty but also epistemic (reducible) uncertainty caused by excessive assumptions or a lack of knowledge in modeling actual physical phenomena. These uncertainties degrade the predictive performance and eventually reduce the usability of the CAE model. However, model calibration can improve the validity of models. On the other hand, with the improvement of internet-of-things (IoT), wireless sensors, and computing power, the technologies of integrating the physical system and CAE model are attracting attention, which is called a digital twin. Because CAE models are updated and calibrated based on data acquired from physical system online, the importance of model calibration techniques are growing. Bayesian calibration, a kind of statistical model calibration, utilizes not only observed data but also prior knowledge in a statistical manner using Bayes' theorem.

However, there is a problem of complex integration when performing complex model calibration that deals with high-dimensional multi-modal distributions. Markov Chain Monte Carlo (MCMC) solves this problem with the Monte Carlo integration method. In the MCMC method, the initial point significantly affects the burn-in period that wastes the sample in the early sampling stage and the calibration performance. However, a proper initial point cannot be obtained if the prior knowledge is insufficient. In this thesis, we developed a cost-effective, stochastic algorithm, the Greedy Stochastic Section Search (GSSS), that can systematically explore parametric space to the proper initial point even when the prior knowledge is insufficient. We verified the algorithm's performance through a numerical example with MoG6 in 6D distribution and calibration of engineering examples: the cantilever beam and the transmission line model.

Keywords: Bayesian calibration
Computer-Aided Engineering models
Digital twin
Markov Chain Monte Carlo
Initial point search algorithm

Student Number: 2020-28845

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Motivation

Numerous experiments on the engineering system are required for engineering decisions. Computer-Aided Engineering (CAE) models play an essential role in engineering decision-making, including designing products and maintenance, as they can be performed at a lower cost and faster than analysis based on experiments.[1, 2] However, there are uncertainties in many variables related to the CAE model. the predictive performance of the CAE model deteriorates due to various types of uncertainties. Uncertainties can be classified into three categories depending on the sources: physical uncertainty, modeling uncertainty, and statistical uncertainty. These uncertainties can be classified more specifically to aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty is reducible if more data is provided; on the other hand, epistemic uncertainty is irreducible though more data is provided.[1, 3, 4]

Model calibration is a process of finding a set of model parameters that obtain the best fit between the model's prediction and measurement while reducing the uncertainties.[1, 2, 5, 6] Deterministic model calibration reduces the difference between observation and the predicted model response value. On the other hand, the statistical model calibration method reduces the uncertainty of the calibration variables by using a statistical method. Because of the uncertainties associated with the CAE model, deterministic calibration methods can degrade model performance. For this reason, statistical model calibration is getting attention because it can take into account various uncertainties.[2]

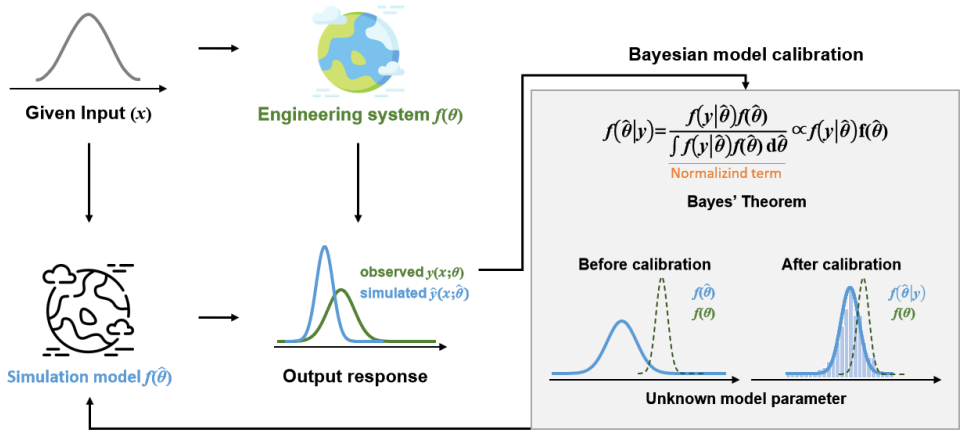


Figure 1-1 Bayesian model calibration process

Figure 1-1 describes the process of Bayesian model calibration. Bayesian calibration, one of the statistical calibration methods, is a method that has the advantage of using prior knowledge as a prior probability or distribution in Bayes' theorem and updating posterior distribution sequentially with repeated experiments.[1, 6, 7] However, calculating the normalizing term in the Bayes' theorem, complex integration must be calculated and limits the use of Bayesian model calibration.[8, 9]

Monte Carlo integration with the sampling method can approximate complex distribution. Markov Chain Monte Carlo (MCMC) method is the most efficient and effective sampling method.[9, 10] Since the MCMC algorithm acquires a new sample based on the current sample if the initial point is selected in the low-density region of the target distribution, many samples acquired in the early stage are wasted due to a long burn-in period. In the worst case, all samples acquired in a limited time may be meaningless.[11] Figure 1-2 shows the effect of an initial point. In this figure, (a) ~ (d) shows the result of sampling started from an appropriate initial point; on the other hand, the rest started from an improper initial point. Each figure shows the approximation results of acquiring 100, 500, 1000, and 5000 samples,

respectively, as the alphabetical order increases. In the case of the proper initial point, it can be seen that the distribution is approximated similarly with only 1000 samples as seen in (c), but in the case of an improper initial point, the approximate result with 1000 samples does not explain the target distribution at all in (g) As you can see. Even when 5000 samples are acquired in (h), it can be seen that inaccurate approximation results are obtained if the samples in the burn-in period are not removed. In this research, we propose an initial point search algorithm, the GSSS, to overcome problems that overcome the problem of insufficient prior knowledge and improve the usability of MCMC.

1.2 Dissertation Layout

This thesis is organized as follows. In chapter 2, preliminaries are explained, such as the probabilistic relationship between CAE model prediction and physics system response, the concept of Bayesian model calibration, the MCMC method, and the Golden Section Search (GSS) algorithm. In chapter 3, the proposed method, GSSS, is described in detail. This chapter consists of the main concepts of the greedy choice method and the stochastic search method. In chapter 4, three case studies are represented to verify the proposed method. These examples include one numerical example and two engineering calibration examples. Finally, chapter 5 is a conclusion. Summary and suggestions for future study are presented.

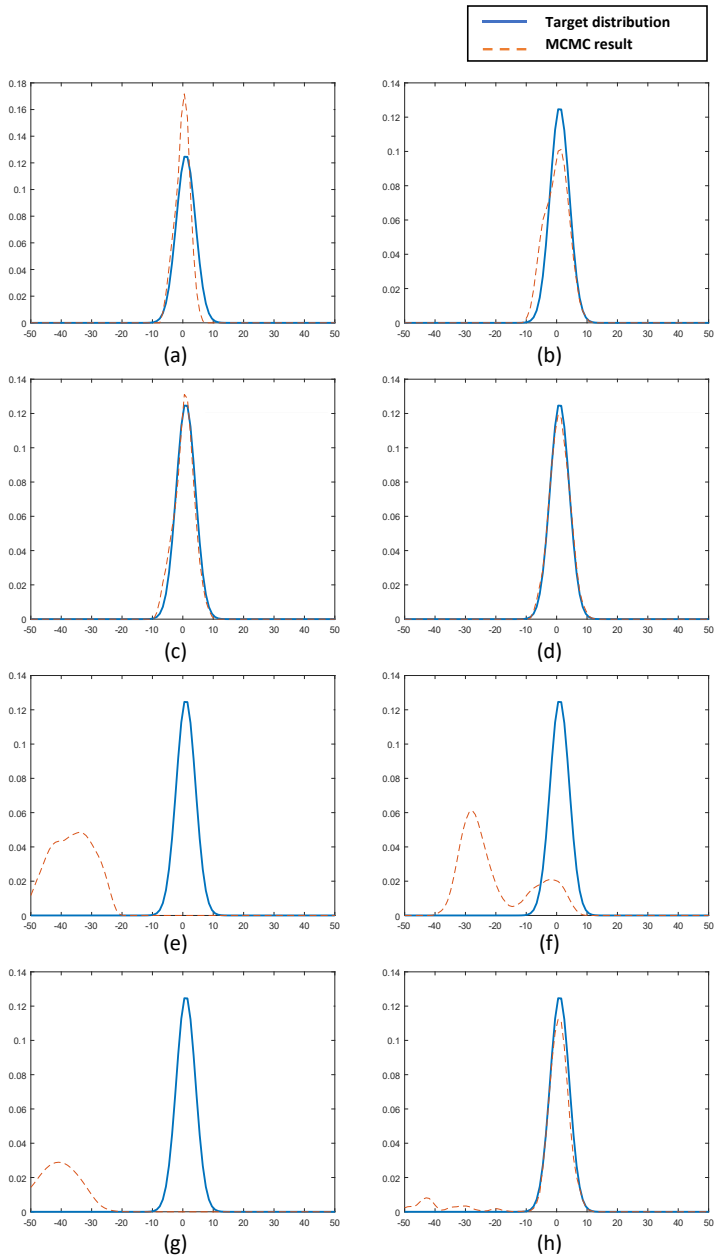


Figure 1-2 The effect of bad initialization in the MCMC method,

(100, 500, 1000, 5000 samples)

(a)~(d) : a proper initial point,

(e)~(h) : an improper initial point

Chapter 2. Research Background

This chapter introduces the preliminaries and research background. Chapter 2.1 describes the probabilistic relationship between the CAE model and physical system response, and Chapter 2.2 the concept of Bayesian calibration and practical difficulties related to integral. In Chapter 2.3, we explain the concept of MCMC and the initial point problem, which is research motivation. Finally, Chapter 2.4 explains the concept and limitations of the Golden Section Search (GSS) method.

2.1 Probabilistic Relation of System Response and CAE Model Prediction

Experimental observations include not only epistemic (reducible) but also aleatory (irreducible) uncertainty. Physical, environmental uncertainty and measurement error are examples of aleatory uncertainty. An observed response can be expressed as Eq. (2.1), considering uncertainty.[2, 12, 13]

$$\boldsymbol{y_{obs} = y_{true} + \epsilon} \qquad \textbf{(2.1)}$$

Y_{true} represents a true response of a system, y_{obs} observed response and ϵ aleatory uncertainty. y_{obs} is defined as a random variable because of epsilon.

CAE model requires model parameters (e.g., material property, dimensions) in addition to model input to predict a response. If modeling is conducted with sufficient knowledge, so modeling error is negligible, a true response can be replaced with a model prediction (y_{hat}).

$$\boldsymbol{y_{obs} = \hat{y}(X, \theta) + \epsilon} \qquad \textbf{(2.2)}$$

Eq.(2.2) shows a probabilistic relationship between model prediction ($\hat{y}$) and observed response (y_{obs}). A likelihood, $f(y_{obs}|\theta)$, can be defined and utilized in Bayesian model calibration.[14]

2.2 Bayesian Model Calibration

Statistical model calibration is a process that reduces uncertainty in calibration variables to improve the CAE model's predictive performance. Bayesian model calibration is a kind of statistical model calibration that uses Bayesian inference, an estimation method using Bayesian statistics. [1, 6, 13] In Bayesian statistics, parameters are assumed not a deterministic value but random variables, so the degree of belief can be expressed as a probability. Using Bayes' theorem, prior knowledge about calibration variables is transformed to the prior probability of calibration variables ($P(\theta)$) and updated with observed data (y). Updated distribution, called posterior distribution ($P(\theta|y)$), means knowledge about calibration variables after observing data.[15]

$$\boldsymbol{P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}} \qquad \textbf{(2.3)}$$

$P(y|\theta)$ in Bayes' theorem (2.3) is the likelihood. Likelihood means how likely data y is observed given θ and is function of calibration variable θ. Likelihood function quantify the similarity between observation and simulated result for a given parameter θ.

In the process of Bayesian calibration, observed data is included in likelihood, so it affects the posterior distribution. In a statistical model with θ as a parameter, the more similar the data generated to the observed data, the higher the likelihood, and the less similar, the lower the likelihood. [1, 5, 6, 13]

$P(y)$ is a marginal probability and plays a role of normalizing constant. Calculating normalizing terms in the Bayes theorem is intractable as the number of calibration variables increase because it requires high dimensional integration. Model calibration associated with multiple calibration variables investigates high dimensional multi-modal distributions.[9] This integral problem makes difficult to use the Bayesian calibration method.

Monte Carlo sampling and Monte Carlo integration are widely used to accomplish Bayesian calibration of complex systems. [16] The sampling method is also a solution to the integration problem in summarizing posterior results with posterior mean or maximum a posteriori (MAP). [1, 8, 9]

2.3 Markov Chain Monte Carlo (MCMC) Method

MCMC is the most efficient sampling method utilizing the Markov chain's property. The Metropolis-Hasting (MH) algorithm is the most basic MCMC algorithm and is a proper algorithm for understanding the MCMC method. MH algorithm iteratively draws a new sample (θ^*) from proposal distribution ($q(\theta^*|\theta)$) and decide to accept it or remain at the old sample (θ) with acceptance probability defined as

$$\boldsymbol{A(\theta, \theta^*) = \min\left\{1, \frac{P(\theta^*)q(\theta|\theta^*)}{P(\theta)q(\theta^*|\theta)}\right\}} \quad \textbf{(2.4)}$$

where $P(\theta)$ is the target distribution. MH algorithm is summarized as flow chart in **Algorithm 1**.

Algorithm 1 Metropolis-Hastings algorithm

Initialize θ^0
for I = 1,2,3 … **do**
Propose: $\theta^{candidate} \sim q(\theta^i|\theta^{i-1})$
Accept Probability:
$A\left(\theta^{i-1}, \theta^{\text{candidate}}\right) = \min\left\{1, \frac{P\left(\theta^{\text{candidate}}\right)q\left(\theta^{i-1}\middle|\theta^{\text{candidate}*}\right)}{P(\theta^{i-1})q\left(\theta^{\text{candidnate}}\middle|\theta^{i-1}\right)}\right\}$
$u \sim Uniform(0, 1)$
if $u < A(\theta^{i-1}, \theta^{candidate})$:
accept the proposal: $\theta^i \leftarrow x^{candidate}$
else:
reject the proposal: $\theta^i \leftarrow x^{i-1}$
end if
end for

If we can calculate the pdf value of any point of the target distribution, the MCMC algorithm can draw samples from the target distribution without expressing analytical form. Therefore, when performing Bayesian calibration using MCMC, samples can be directly obtained from the posterior distribution without being expressed in the analytical form of each of the prior, likelihood, and normalizing terms. Furthermore, because the ratio of P(θ) and $P(\theta^*)$ is required, not each of the values in calculating the acceptance probability, it is unnecessary to calculate the normalizing term in the posterior sampling process as follows:

$$\frac{\boldsymbol{P(\theta^*|y)}}{\boldsymbol{P(\theta|y)}} = \frac{\frac{\boldsymbol{P(y|\theta^*)P(\theta^*)}}{\boldsymbol{P(y)}}}{\frac{\boldsymbol{P(y|\theta)P(\theta)}}{\boldsymbol{P(y)}}} = \frac{\boldsymbol{P(y|\theta^*)P(\theta^*)}}{\boldsymbol{P(y|\theta)P(\theta)}} \qquad \textbf{(2.5)}$$

Therefore, if prior ($P(\theta)$) and likelihood ($P(y|\theta)$) can be computed at any values, we can apply MCMC algorithms. Researchers have been focused on proposal distribution and developed many MCMC algorithms like Metropolis-hasting, slice sampling, Hamiltonian MC. [9, 10, 16, 17] These algorithms have unique methods of proposing a new sample, that is, proposal distribution. Figure 2-1 represents some results of three MCMC algorithms: (a) Metropolis-hastings, (b) Slice sampling, and (c) Hamiltonian MC. However, determining an initial point of MCMC is previously unstudied because MCMC is an algorithm that converges regardless of the initial value with infinite time and computing resources. However, it is trivial that the assumption is invalid. Since all MCMC algorithms propose new samples based on the previous sample, the initial point affects sampling results performed within a limited time. Poor initial points cause a long burin-in period, an early stage of sampling, that wastes computational cost or fails to reach the probable region.[18] Therefore, the approximation result is unacceptable if the MCMC algorithm starts in an insignificant region independent of the target distribution. To minimize the effect of an initial point, multiple-initialize method is adopted. However, this strategy requires a substantial computational

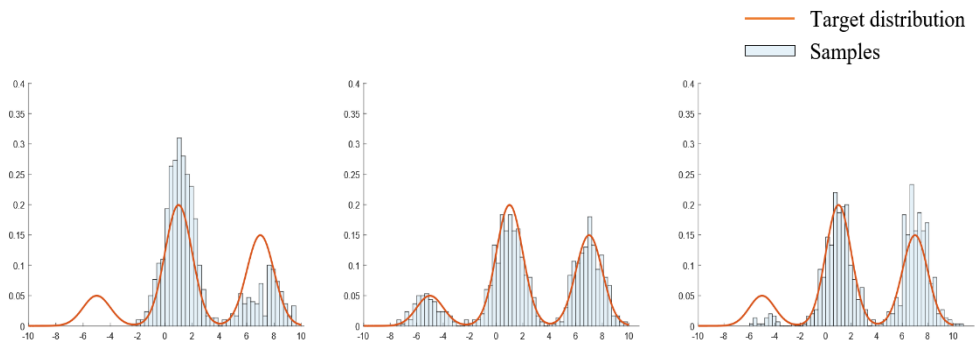


Figure 2-1 Sampling results of various MCMC algorithms.
(a) Metropolis-hastings algorithm, (b) slice sampling algorithm,
(c) Hamiltonian MC

cost.[7, 19]

Many studies have selected an initial point relying only on prior knowledge despite the importance of an initial point. Accordingly, it has limitations to use the MCMC method with insufficient prior knowledge. A more systematic algorithm is required for selecting an initial point of the MCMC algorithm.

2.4 Golden Section Search (GSS) Algorithm

The GSS is one of the most efficient optimization algorithms for one-dimensional uni-modal distribution.[20] Because it is a gradient-free algorithm, the GSS is proper to deal with the objective functions which are difficult to differentiate. The algorithm, proposed in 1953 by Kiefer, has been used in various fields and applications.[21-23] Figure 2-2 shows the flow chart of the GSS algorithm in one dimension, and Figure 2-3 is a schematic explanation.

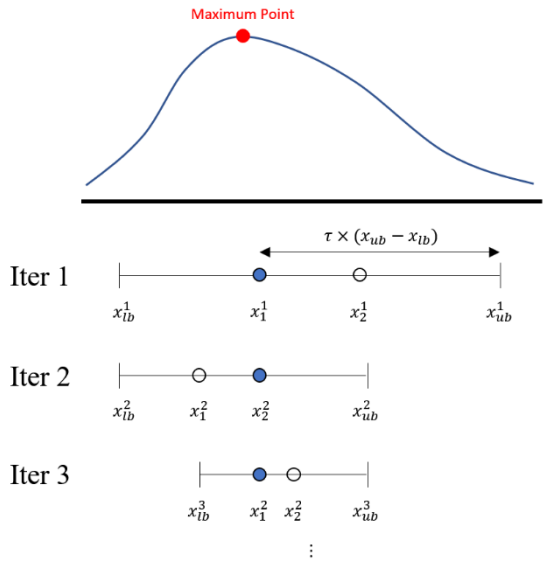


Figure 2-3 A schematic explanation of the GSS algorithm in one dimesion

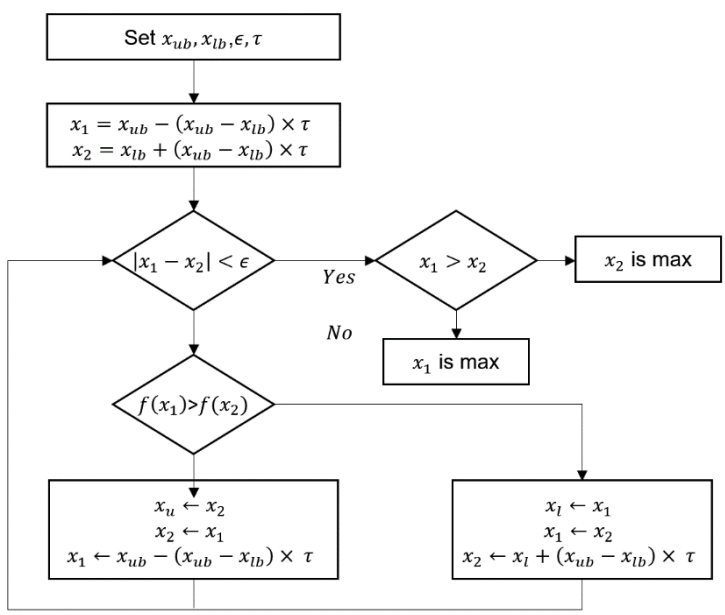


Figure 2-2 The flow chart of GSS algorithm in one dimension

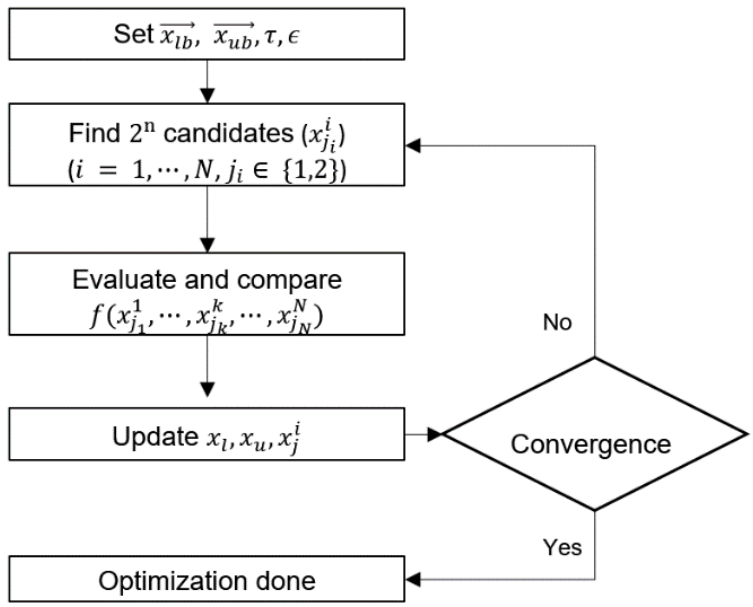


Figure 2-4 The flow chart of GSS algorithm in N dimension

[21] used the algorithm to search the initial point of hyperparameter optimization of the kriging model. This study showed that the GSS algorithm outperformed the genetic algorithm as the initial point search method. The authors reduced the dimensionality of the hyperparameter to a single variable using a radial basis function (RBF) so that they could apply the GSS algorithm without modifying it.

Some researchers use the GSS algorithm to optimize 2-D space and analyze the algorithm's characteristics. According to [24], the time complexity of the GSS in n-D is $O(2^N)$. Furthermore, the search space dimension is larger, the elimination region smaller. The flow chart of the GSS algorithm of the N dimension is shown in Figure 2-4. For example, every iteration eliminates 40% of search space by comparing two candidate points in 1-D and, on the other hand, 16% by comparing four candidate points in the 2-D case. Also, if the boundary condition is fixed, the result of algorithm implementation is determined. This deterministic

characteristic is a limitation to applying GSS to the initial search task because it is impossible to converge multiple modes.

Chapter 3. Greedy Stochastic Section Search (GSSS) algorithm

This chapter proposes the GSSS algorithm, a new initial point search algorithm for the MCMC method. The GSSS is an algorithm proposed to overcome the exponential increase of computational cost in a high dimension of GSS and the problem that cannot be applied to multi-modal distribution due to its deterministic nature. GSSS reduces the computational cost by applying a greedy algorithm to the boundary update process at every iteration of the GSS to search a subset of the search space. In this process, the necessary reference points were stochastically selected, and the split ratio was defined as a random variable so that the algorithm had stochastic properties.

The GSSS algorithm can be applied only with the constraints of the upper bound and the lower bound, and Bayesian calibration can be started in different modes of the high-dimensional multi-modal distribution through iterative search. The overall process of GSSS is summarized in Figure 3-2.

3.1 Greedy Choice Method

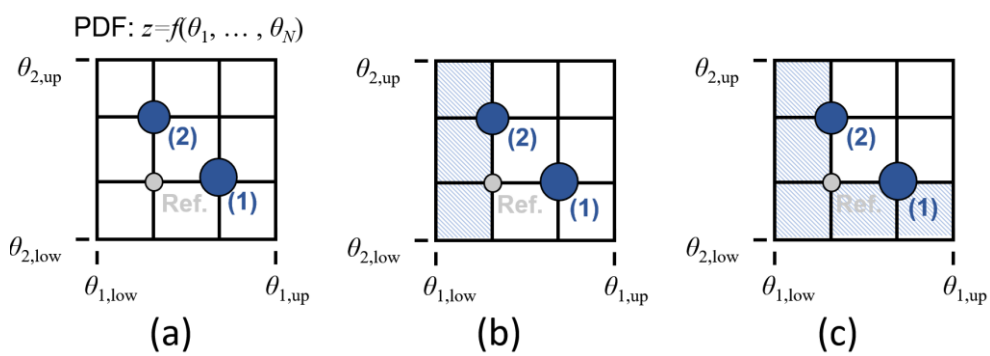


Figure 3-1 A iteration of GSSS algorithm

The greedy algorithm is a method that makes the best choice at each point without considering long-term performance.[25] A greedy algorithm is used to reduce computational costs in various fields. [25-27]

Here, the greedy concept is used to reduce the search space searched in the iteration of GSS. GSS searches all lattice points of the search space determined by the upper and lower boundary and split ratio by brute-force search and updates the boundary at once. As explained in Section 2.4, the lattice points increase exponentially as the dimension increases, making it difficult to use. In GSSS, the Greedy choice method does not search all lattice points and update them at once but determines one reference point among lattice points and updates the boundary for each dimension based on this.

In Figure 3-1, (a)~(c) shows the single iteration of the update process of GSSS. The gray-colored point is determined as a reference point, and the candidate points are to be compared accordingly. Iteratively updates each dimension based on the reference point. Figure 3-1(b) is the process of updating the dimension. When comparing the reference point and the value of point1, the shadowed blue area is eliminated because point 1 has a more significant value. Figure 3-1(c) updates the boundary of the dimension in the same way. If the algorithm is used in more than two-dimensional space, repeat this process for all dimensions. If the Greedy algorithm was not used in 2D, the values of all four lattice points had to be calculated, but the greedy algorithm was applied to update the three-point calculations. In 2D, 1 point is saved, but the cost saved increases exponentially as the dimension increases.

The most crucial role in the greedy choice method is the reference point. When the reference point is determined, selecting the reference point is important because the candidate point to be evaluated in the iteration is determined accordingly. It is dealt with in the following stochastic method.

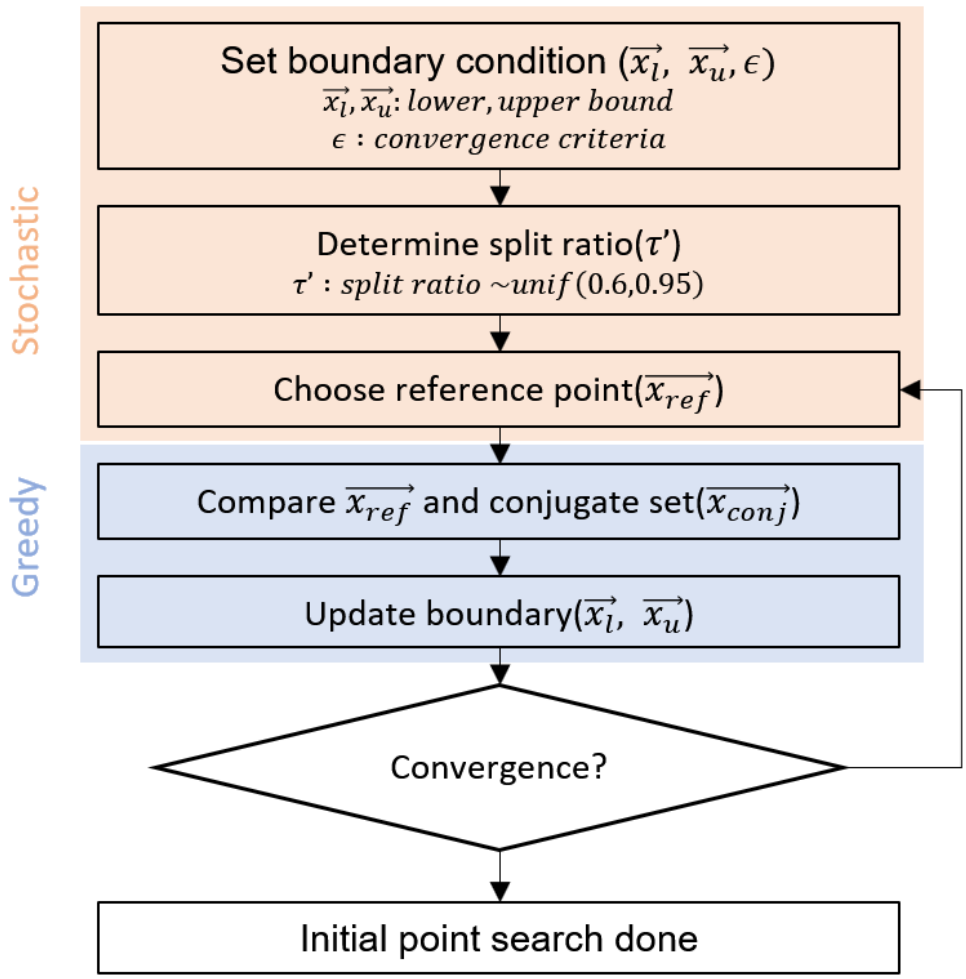


Figure 3-2 The summary of the proposed method, the GSSS algorithm.

3.2 Stochastic Search Method

We made the GSSS algorithm stochastic in two ways. The first is a method of defining the split ratio of the GSS algorithm as a random variable, and the other is a Density-based Reference Point Sampling method that selects a reference point in a probabilistic way based on the search density in the process of the greedy choice method.

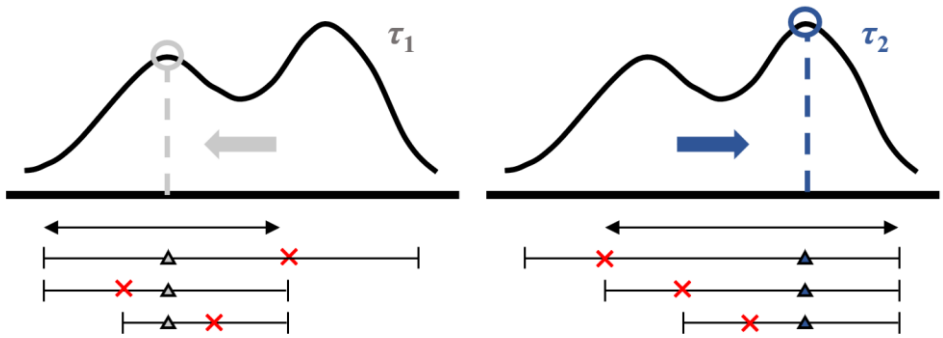


Figure 3-4 The effect of a split ratio

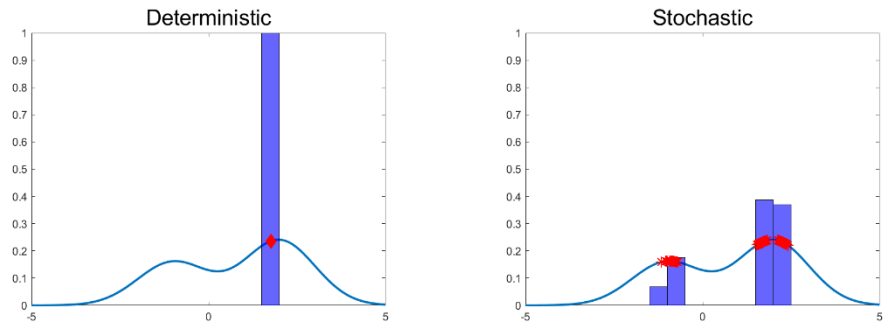


Figure 3-3 A numerical example to show effect of a split ratio

As shown in Figure 3-4, the mode in which the GSS algorithm converges varies according to the split ratio in the multi-modal distribution. Considering this, the split ratio, which was a deterministic value in GSS with a golden ratio (0.618), was defined as a random variable so that it could converge to different modes when the GSSS algorithm was repeatedly applied. Figure 3-3 is the result of 1000 repetitions by defining the GSS and split ratio as a uniform random variable with a range of 0.6 to 0.95. In the deterministic GSS algorithm of split ratio, it can be seen that all simulations converge to one mode. However, in the case of the random variable split ratio, the algorithm converges evenly in both modes.

Next, Density-based Reference Point Sampling is a method of selecting a reference point in the greedy choice process. (Figure 3-5) The theoretical basis of this method is the upper confidence bound (UCB), which is used to make decisions with partial information feedback in multi-armed bandit or Monte Carlo tree search methods. [28-30] Density-based uncertainty quantification is performed to apply the UCB method, which quantifies uncertainty in the discrete domain based on frequency, to the continuous domain. The method runs the algorithm, estimates the density of the search space using kernel density estimation (KDE) for the points searched up to the previous iteration and calculates the density of all lattice points. [27, 31, 32] In this case, high density means that there is much information about the region adjacent to the corresponding lattice point, which means that uncertainty is low. Conversely, a low density means less information and high uncertainty, so a strategy that determines a lattice point with high uncertainty as a reference point with high probability makes it possible to search the search space evenly. The uncertainty of each lattice point is quantified as U_i. This expression is an expression of replacing frequency with density in the term meaning uncertainty in UCB. At this time, instead of simply determining the point with

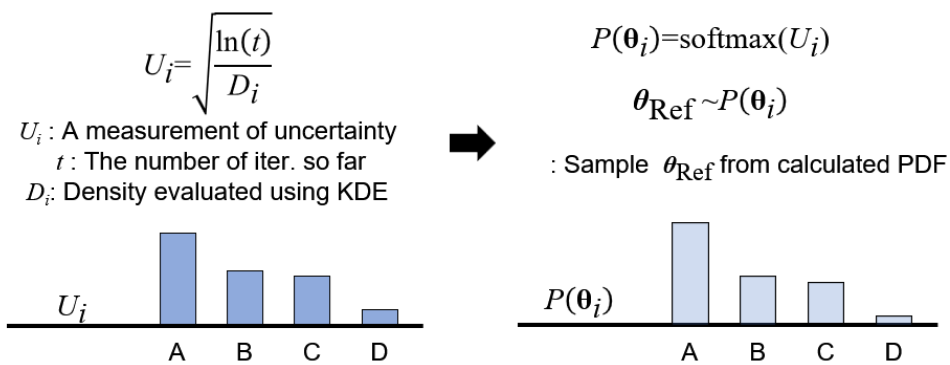


Figure 3-5 Process of probability calculation
in Density-based Reference Point Sampling

the highest uncertainty as to the reference point, the uncertainty is converted into a probability using the softmax function, and the reference point is determined probabilistically based on this. [33]

Chapter 4. Case Studies

In this section, to verify the performance of GSSS, apply it to one numerical example and two engineering examples to analyze the results. In the numerical example, the characteristics of the initial value search method are compared using the known multi-dimensional multi-modal distribution to help the understanding of the proposed method and to analyze the effects of the greedy and stochastic search methods, respectively. The CAE model is calibrated through two calibration case studies, and the performance according to the initial value search method is compared. The effectiveness of the proposed method can be compared with the error and inference time of the calibration result according to the initialization method. For all examples, initial values are selected and compared using sampling from non-informative prior (sampling method), GSS, and GSSS methods.

4.1 Numerical Example : Mixture of Gaussian (MoG) 6 in six dimension

A Mixture of Gaussian 6 (MoG6) distribution is a gaussian mixture distribution with six modes.[34, 35] This research aims to develop a cost-effective, stochastic algorithm for high-dimensional multi-modal distributions, so MoG6 in 6D distribution is a proper distribution. Information on each mode of this distribution is shown in a Table 4-1. Except x_1 and x_2, other element of modes are set to be 0; the value of $x_3 \sim x_6$ are zero.

This example applies only initial point search methods, excluding estimating distribution with the MCMC algorithm. For the non-informative distribution, a uniform distribution in which the boundary of each dimension is [-100, 100] was chosen. The boundary between GSS and GSSS is the same as the sampling method, and the convergence criterion (ϵ) is set to 0.01. Each method is executed 10,000 times.

The number of calculations and Kullback-Leibler divergence (KLD) with the ideal case is used as a metric to compare initial search methods. The former quantifies computational costs and the latter mode search performance. To explain more details about KLD, converging to the low-density region means failing to find proper initial points. The more evenly model converges to each mode, the better the performance.[36] So, the ideal case is where search results converge to each mode with a probability of 1/6 and never converge to the low-density region. The searched initial point is classified into 6 modes (from mode 1 to mode 6) and low-density region (search failure), transformed as frequency and compared with the ideal distribution. The lower the KLD, the smaller difference from the ideal distribution, so the smaller the KLD, the better method is.

Table 4-1 The coordinate of modes of MoG6 distribution

Mode number	x_1	x_2
1	-5	-8.66
2	5	-8.66
3	10	0
4	5	8.66
5	-5	8.66
6	-10	0

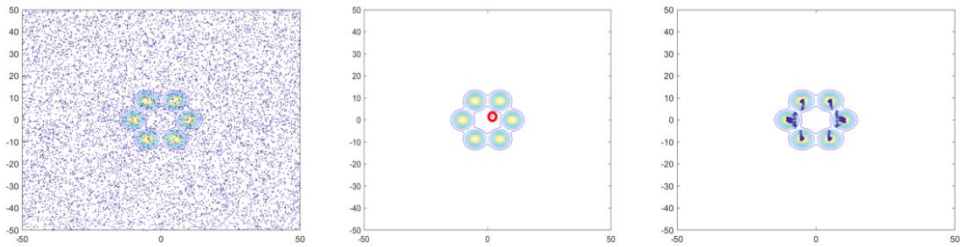


Figure 4-1 The qualitative result of the numerical example

Qualitative results are shown in Figure 4-1, which is visualized by marginalizing the search results from each method. It can be seen that the results of the sampling method are randomly selected within the boundary. In GSS and GSSS, it can be seen that all points converge near the mode, but in the case of GSS, it can be seen that all trials converge to the same point. On the other hand, it can be seen that the results of GSSS converge to all modes. In Figure 4-3, we represented the convergence ratio for each mode. In the marginal plot, the sampling method also seemed to converge on some modes, but all implementation never converged to any modes. GSS converges on all points in mode 3, and GSSS converges evenly on six modes. The results of quantification with KLD and elapsed real-time are summarized in Figure 4-2. The bootstrap method was used to estimate the variation of the result. KLD, a metric of initial point evaluation, is shown in Figure 4-2(b) in log-scale. Finally, we did a test for the time complexity analysis. GSS requires 604 calculations to find an initial point, but the GSSS 158.1 on average. This result suggests that GSSS is an algorithm that converges in various modes while being cost-effective.

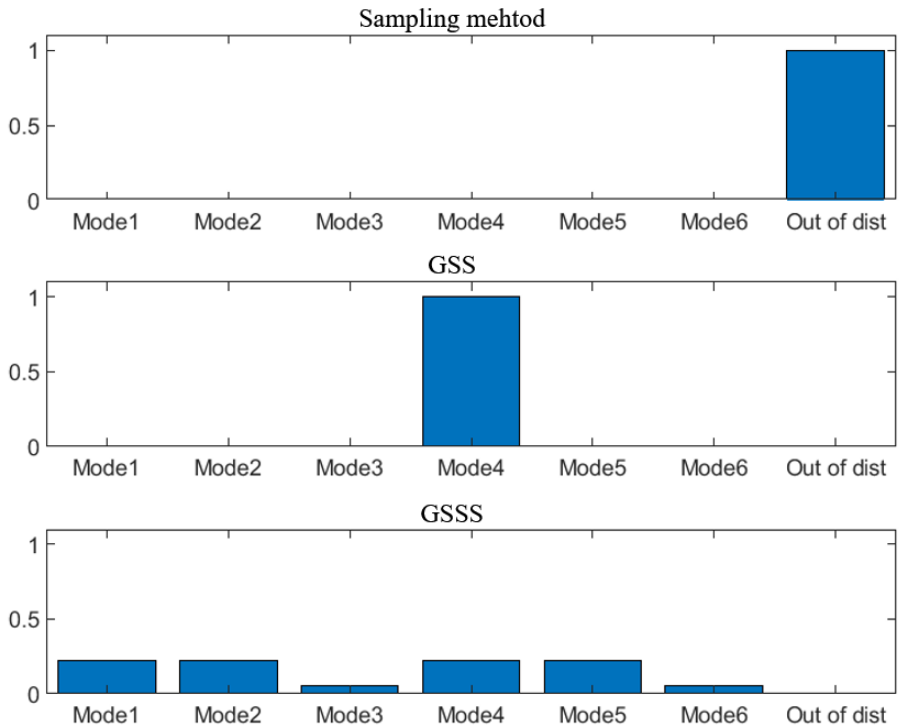
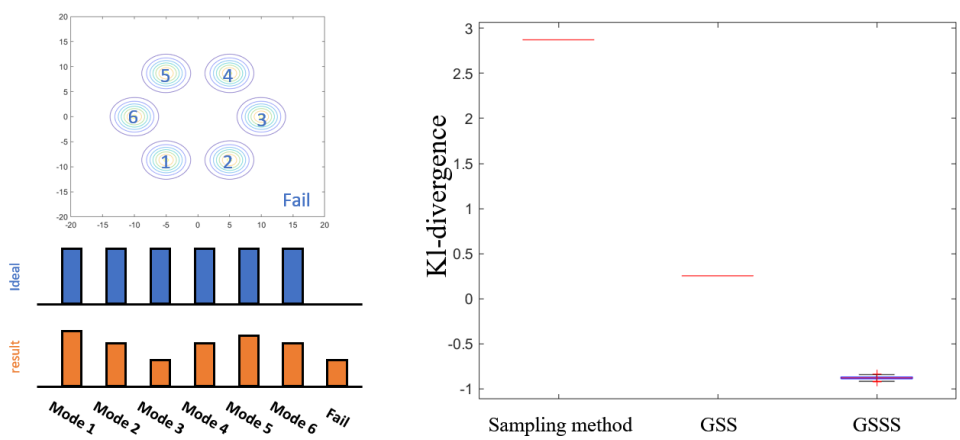


Figure 4-3 The frequency of the initialization method



(a) (b)

Figure 4-2 The quantitative result: Kullback-Leibler divergence (KLD)

(a) the schematics of the distribution, ideal result and drawn result

(b) KLD results by initialization methods

4.2 Engineering Example 1: Bayesian Calibration of Cantilever Beam

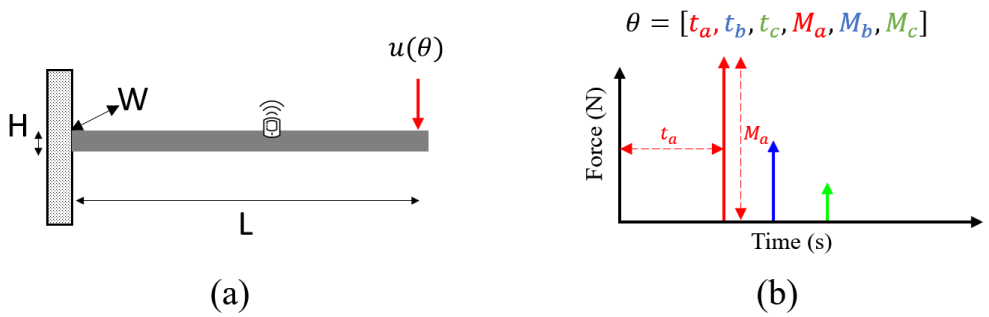


Figure 4-4 (a) Schematics of cantilever beam and (b) excitation visualization

Two Bayesian calibration case studies are provided to demonstrate the effectiveness of the proposed method. The first case study is to infer vibration excitation information which includes excitation timing and magnitude. In diagnosing a bearing fault or a mechanical fault of On-Load Tap Changer (OLTC), the vibration excitation time difference or magnitude can be used as a health index so that this case study can be expanded to the fault diagnosis method.[37, 38]

Figure 4-4(a) is a schematic of the FEM model, and Figure 4-4 (b) shows how the calibration variable is visualized in this example. The variables of the model include dimension, mechanical property, excitation time and magnitude. Among these, only excitation time and magnitude were defined as calibration inputs (θ) and other known model inputs (x). This case study is assumed to have three excitation events. The response is 5 seconds of vibration signal acquired with a sampling rate of 2,000 Hz.

Calibration vector whose elements are calibration variables $\theta = [t_a, t_b, t_c, M_a, M_b, M_c]$ determine response. As the vibration signal observed is the superposition of the responses of

excitations, what matters is not each element but the time-magnitude pair of the vector. For example, $\theta_1 = [t_1, t_2, t_3, M_1, M_2, M_3]$ and $\theta_2 = [t_2, t_3, t_1, M_2, M_3, M_1]$ are different vectors, but generate the same vibration signals. This means that the target distribution of Bayesian calibration is multi-modal. This is illustrated graphically in Figure 4-5. Figure 4-5(a) is a different vector, but when visualized, it explains that these vectors are semantically equivalent, and Figure 4-5(b) is a figure showing the same response generated from these vectors.

However, since the target probability distribution is unknown, it is necessary to rearrange the time-magnitude in chronological order through sorting to compare the calibration results. For example, $\theta = [t_2, t_3, t_1, M_2, M_3, M_1]$ $where$ $(t1 < t2 < t3)$ then $\theta^* = [t_1, t_2, t_3, M_1, M_2, M_3]$ Rearrange θ to θ^* and compare it with the ground truth vector.

We use the slice sampling algorithm to estimate probability distribution, which is a type of MCMC algorithm. Using sampling results, we acquire 60 chains of 1000 samples and estimate the calibration variable with Maximum a Posteriori (MAP). And in order to analyze the effect of the quality of the initial point, we compared and analyzed the initialization

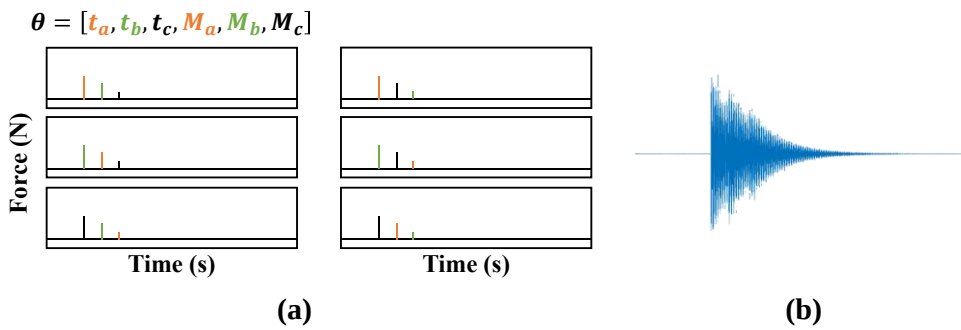


Figure 4-5 The multi-modal characteristics

(a) visualization of calibration vectors

(b) simulated response from calibration vectors (θ)

methods. The non-informative prior for the sampling method used a uniform distribution, and the boundary was set to time [0, 2.5] and magnitude [0, 10].

The Bayesian calibration process proceeds as Figure 4-6. The target data is generated with ground truth values as in Table 4-2. Continuous wavelet transform is performed on the target signal and the signal generated by model simulation. Features that include time, scale, and magnitude (d, d*) are extracted from a ridge of cwt. This distribution of d, d* is estimated using KDE as the likelihood function ($f(d|\theta)$) then, multiply this by the prior to calculate the posterior. Repeat Step2 and Step3 until enough samples are acquired.

Table 4-2 The ground truth vector of engineering example 1

	Time(s)	**Magnitude(N)**
1st excitation	1	4
2nd excitation	1.25	2
3rd excitation	1.5	1

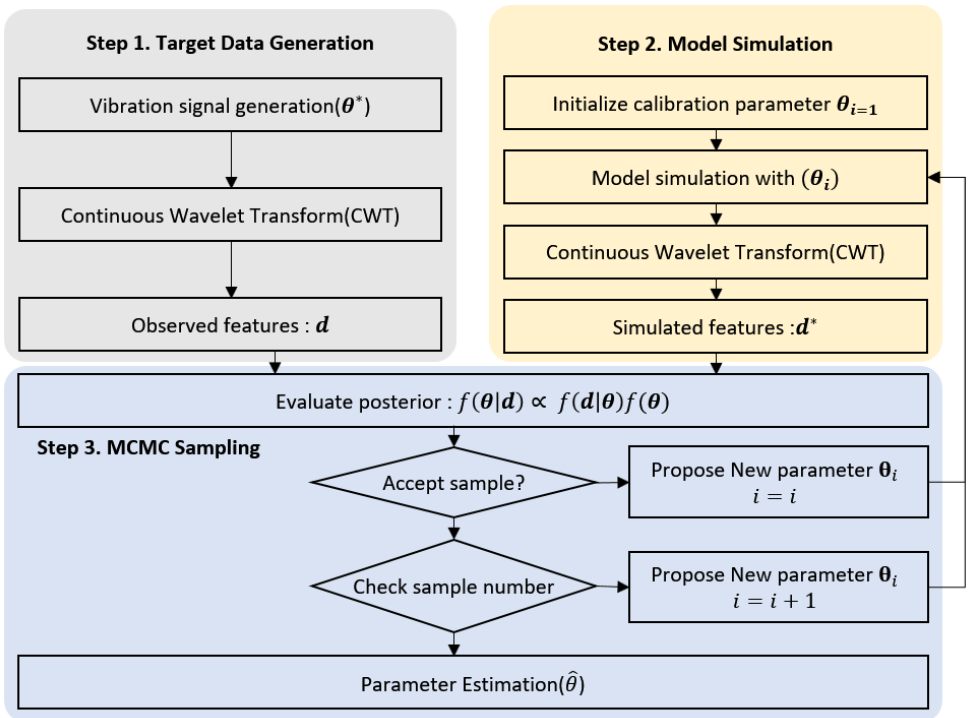


Figure 4-6 Procedures for a Bayesian model calibration of the engineering example 1

Figure 4-7, Figure 4-8, and Figure 4-9 show the qualitative result. The $x-axis$ means excitation time and the $y-axis$ excitation magnitude in these figures. Three excitations are plotted with different colors; red represents the first excitation, blue the second, and green the third. All excitations are not distinguished clearly in the sampling initialization method in Figure 4-7. In the GSS initialization case, the calibration result is not valid except for the first excitation. Especially, estimation of the last excitation is unacceptable. (see Figure 4-8) However, the three excitations are distinguished, and the estimation of the first excitation is remarkably accurate. (see Figure 4-9) The first excitation estimation result is the best for all initialization methods, followed by the second and third. It is because the excitation

magnitude becomes smaller as it goes back and therefore has little influence on the likelihood function.

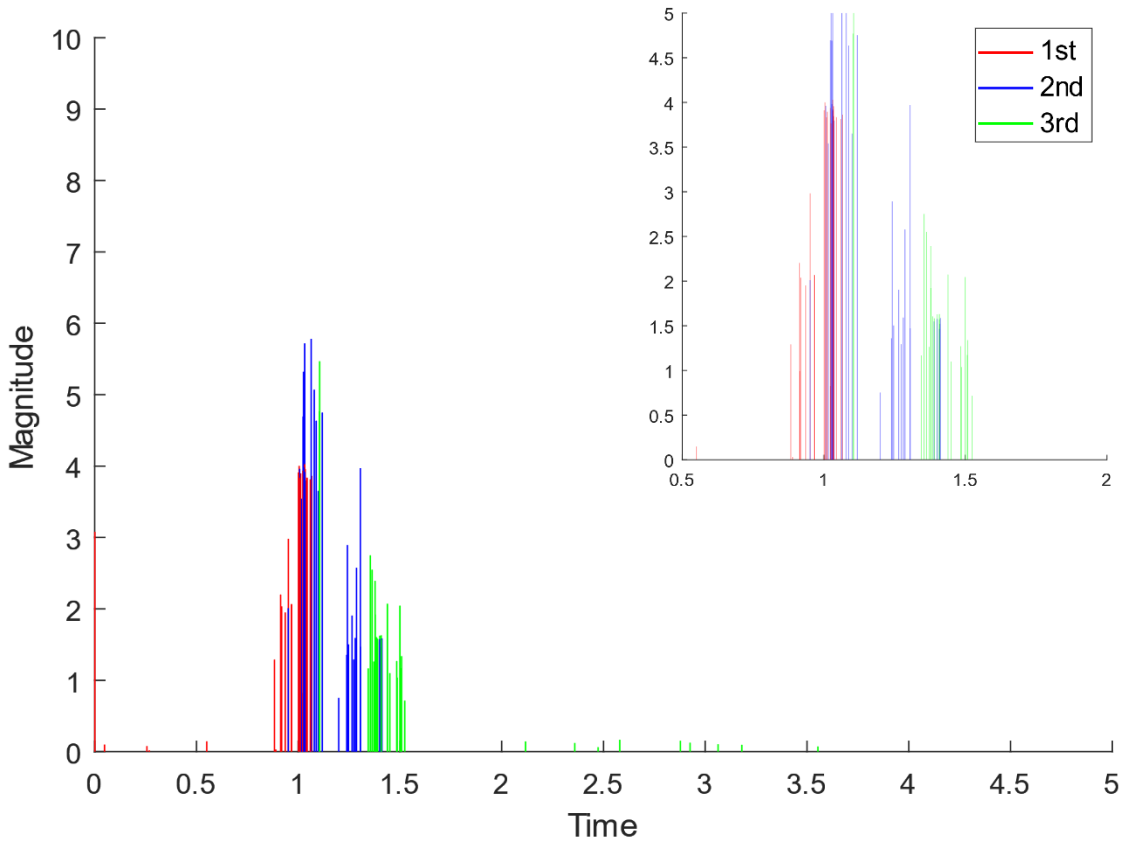


Figure 4-7 The result of the sampling method initialization

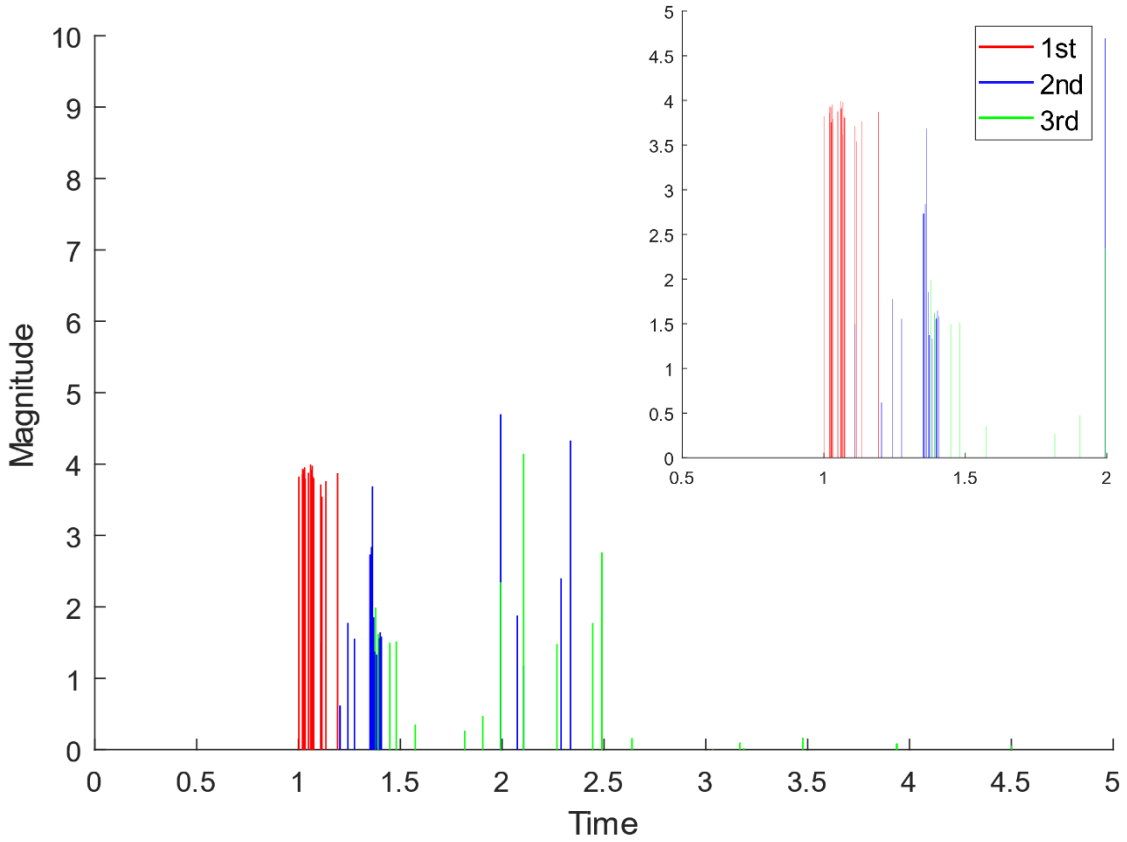


Figure 4-8 The result of the GSS method initialization

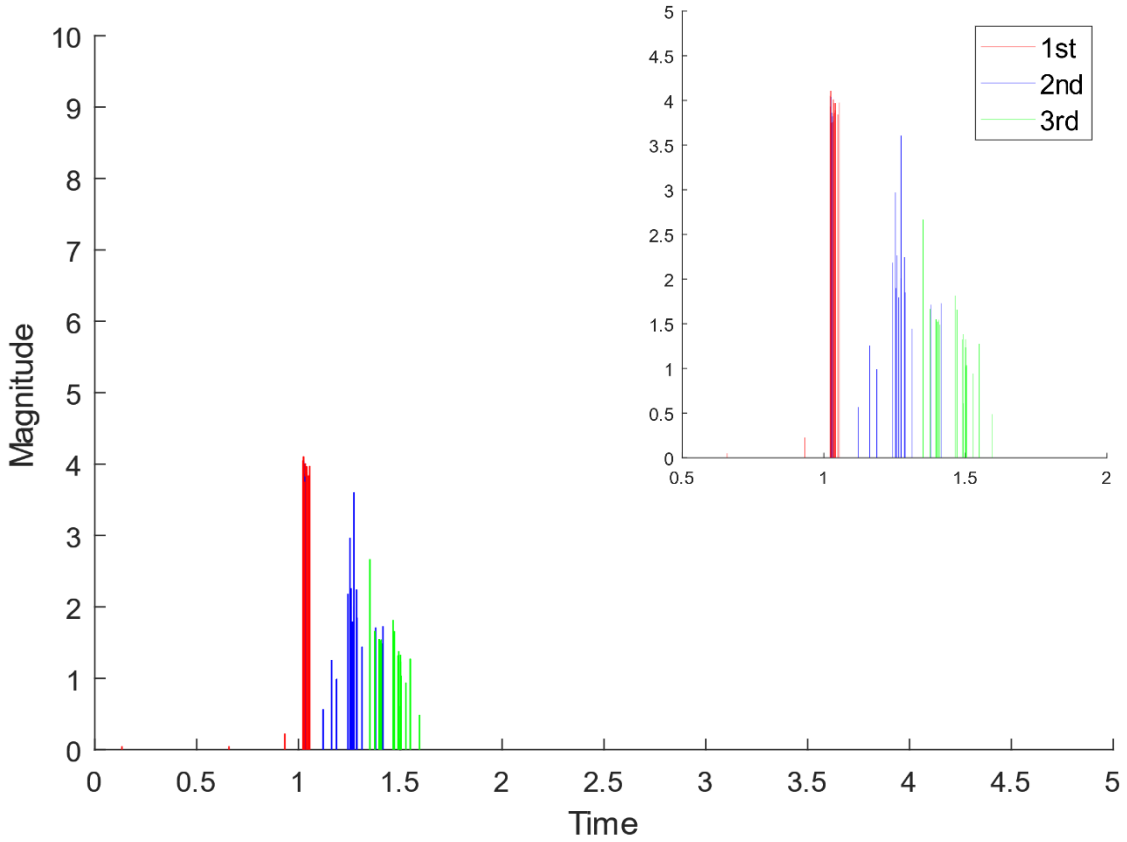
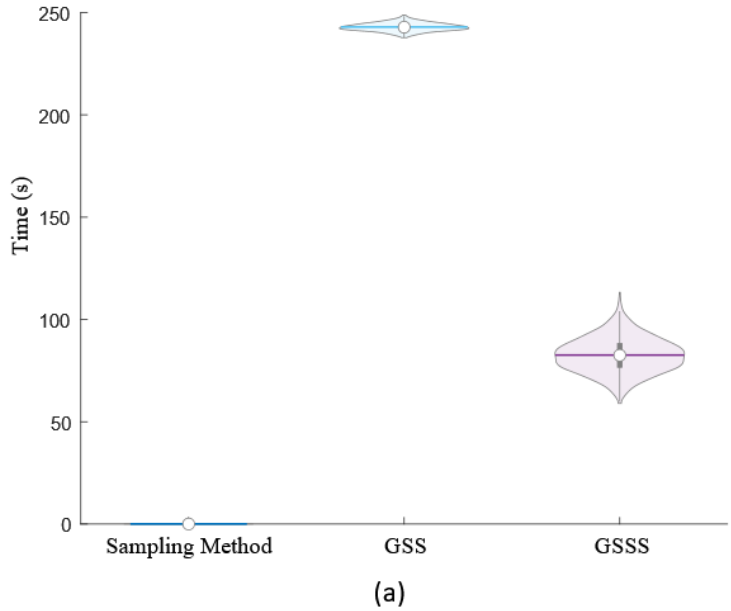


Figure 4-9 The result of the GSSS method initialization

Quantitatively, Figure 4-10(a) shows the implementation time of initialization methods and Figure 4-10(b) MSE of the Bayesian calibration result. The sampling initialization takes 0.051s, GSS initialization 242.92s, and GSSS initialization 82.29s on average. As the sampling initialization is drawing a random number from known distribution, it takes little time. GSSS initialization required about 34% of the time required for GSS initialization. However, the MSE result of sampling initialization was 3.635, which was inaccurate. The MSE of the GSSS initialization case was 1.1458, but GSS initialization obtained 0.6552, which is lower than that. To summarize the case study, better calibration results were obtained more efficiently when GSSS initialization was used.



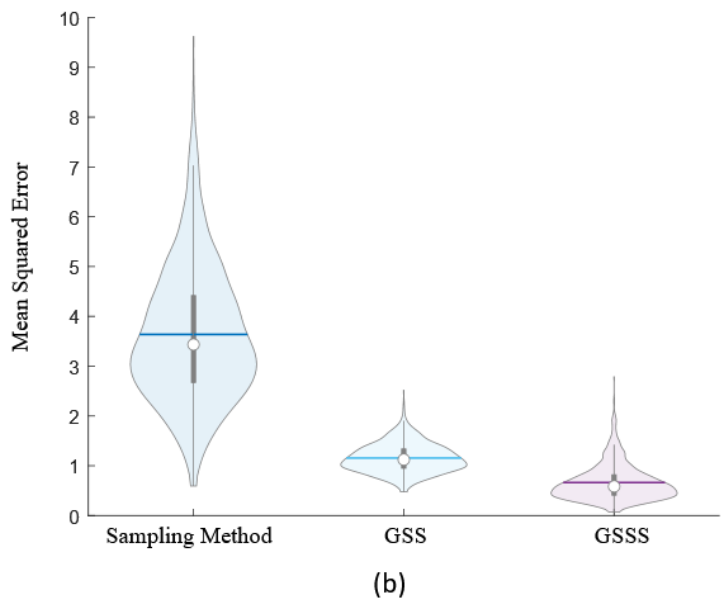


Figure 4-10 The quantitative result of engineering example 1

4.3 Engineering Example 2: Bayesian Calibration of The Transmission Line Model

The transmission line model is a model that can simulate the behavior of a human hemodynamic system.[39] It is a recursive model that can generate blood pressure waveform data generated at 55 locations in the human body given five normalized parameters: young's modulus (E), thickness (T), radius (R), peripheral resistance (PR), and length (L). Cardiovascular diseases such as Abdominal Aortic Aneurysm (AAA) or arterial stenoses can be efficiently detected from blood pressure.[40, 41] Cardiovascular disease may be diagnosed by estimating parameters from the acquired blood pressure waveform data. Figure 4-11 shows an example of a blood pressure waveform.

The calibration process is as follows. As in Engineering Example 1:

Bayesian Calibration of Cantilever Beam, target data is generated using a specific calibration

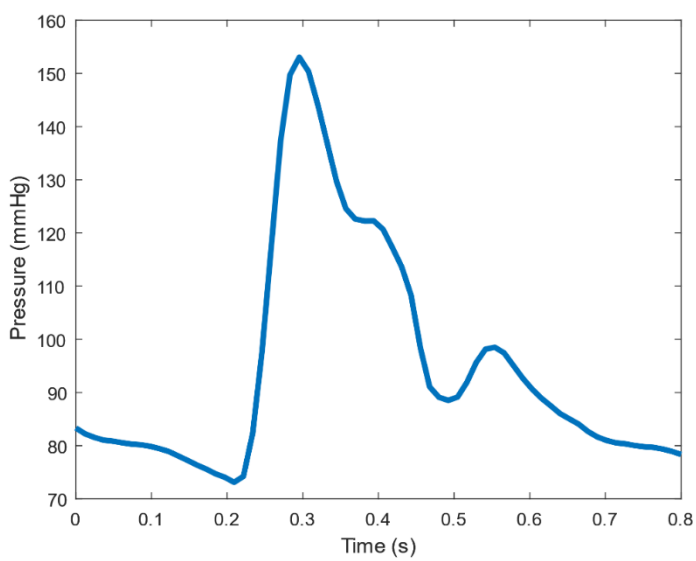


Figure 4-11 The simulated blood pressure from the transmission

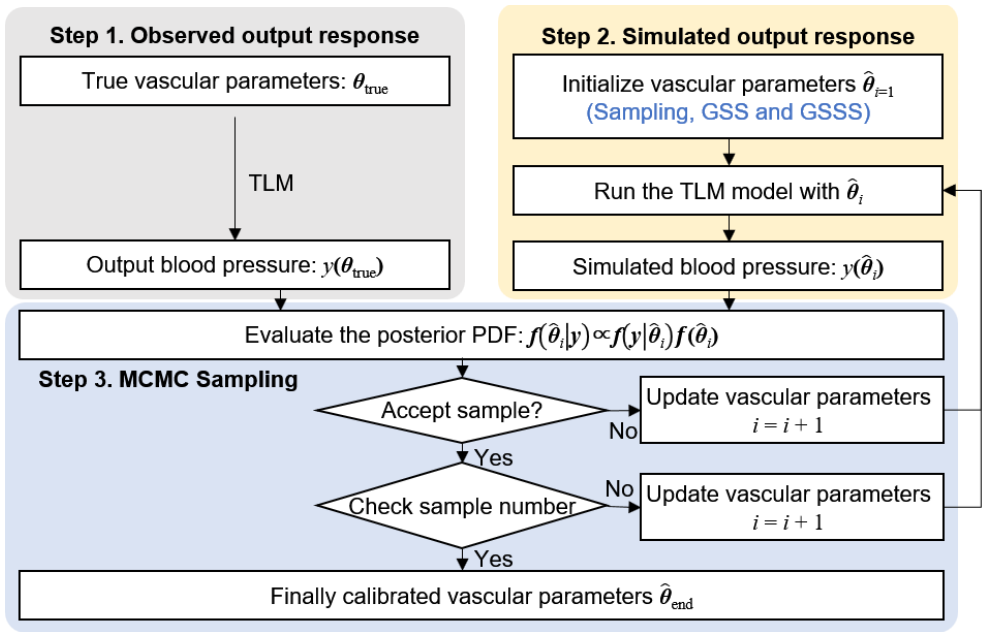


Figure 4-12 Procedures for a Bayesian model calibration of the engineering example 2

vector as the ground truth, and estimation is performed by applying the Bayesian calibration method. All elements of the ground truth vector are one. Unlike case study 1, in this case, raw data was used without signal pre-processing. Overall calibration process is shown in Figure 4-12. Assuming that each data point has the same Gaussian noise type measurement error, the likelihood was calculated in the same way as the research background as follows. The detailed formulation is as follow:

$$y_{obs} = y + \epsilon$$

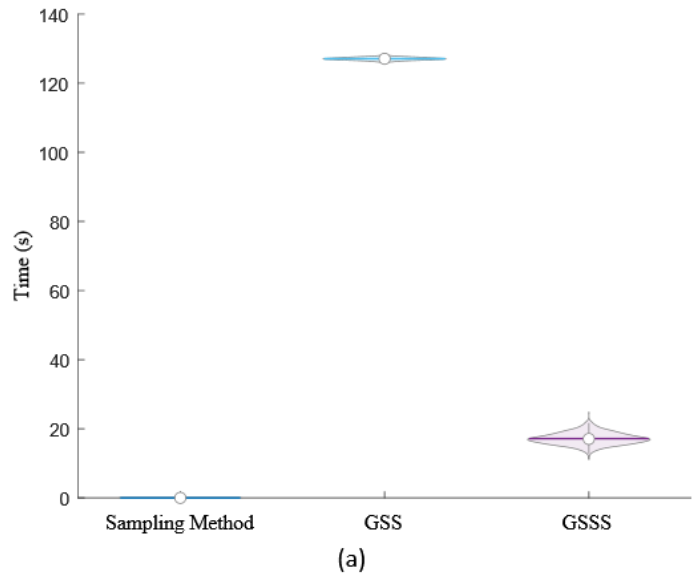
$$y_{obs} - y \sim N(0, \Sigma) \; where \; \Sigma = \sigma \cdot I_{\{N \times N\}}$$

$$L(\theta; y) \propto \prod_{i=1}^{N} \exp\left[\frac{(y_i - y_{obs,i})^2}{2\sigma^2}\right]$$

$$\log L(\theta; y) \propto \sum_{i=1}^{N} \frac{\left(y_i - y_{obs,i}\right)^2}{2\sigma^2}$$

The prior knowledge is modeled as uniform distribution whose minimum and maximum are 0.8 and 1.2, respectively, and the posterior distribution was calculated as the product of non-informative prior and likelihood. Also, in this example, the MSE of the Bayesian calibration result initialized with random sampling, GSS, and GSSS is compared with the computation time.

Quantitatively, Figure 4-13(a) shows the implementation time of initialization methods and Figure 4-13(b) MSE of the Bayesian calibration result. The sampling initialization take 0.1275s, GSS initialization 127.09s, and GSSS initialization 17s in average. However, similar to the Cantilever Beam example, the GSSS initialization showed the best calibration performance.



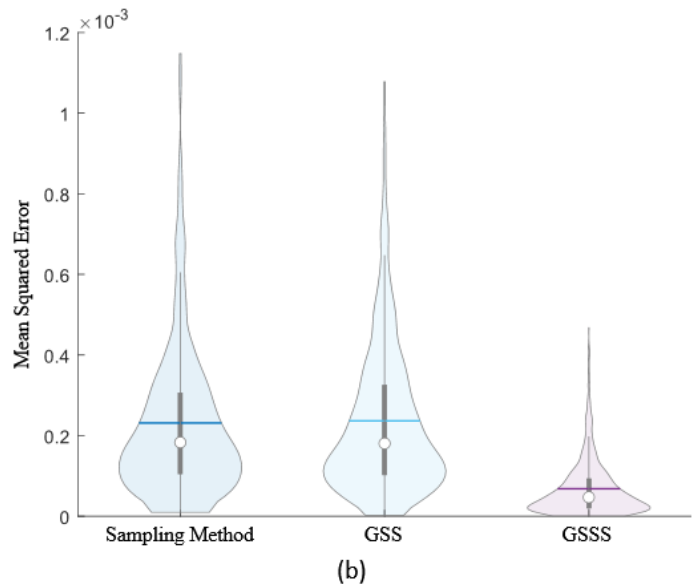


Figure 4-13 The quantitative result of engineering example 2

Chapter 5. Conclusions

5.1 Summary and Contributions

This thesis proposed the GSSS algorithm, an initial point search algorithm, to perform Bayesian model calibration using the MCMC method with insufficient statistical information.

Contribution 1: Analysis of the effect of the initial point of the MCMC method for Bayesian model calibration

Although it has not received attention in the current MCMC study, when applying the MCMC method practically, the initial value is greatly affected by resource limitations, numerical errors, the high dimensionality of distribution, and multi-modal characteristics. In a study related to Bayesian calibration using MCMC, the initial value was selected using prior knowledge, but it was difficult to apply in the absence of prior knowledge. In order to use the MCMC through this study, the necessity of systematically selecting initial values was suggested.

Contribution 2: Development of cost-effective initial point search algorithms applicable to the high-dimensional multi-modal distribution

This thesis proposed the GSSS algorithm, an initial point search algorithm, to perform Bayesian model calibration using the MCMC method with insufficient statistical information. The proposed method was developed by introducing the greedy and stochastic properties to the GSS algorithm. The greedy choice method reduces the computational cost in high dimensional, modifying the time complexity of $O(2^N)$ to $O(N)$. The stochastic search method allows the algorithm to search multiple modes while evenly searching the entire search space. In summary, The proposed method can search high-dimensional distribution efficiently, and furthermore, it is applicable to multi-modal distribution.

Contribution 3: Conducting a case study in the various field, applying the proposed method to examples in various fields

We verified the utility of the GSSS algorithm by conducting three case studies: one numerical example and two engineering examples. In a numerical example using MoG6 distribution, the multi-modal distribution search performance was quantified using KL-divergence as a metric, and the cost was compared. Then, the cantilever beam example and the transmission model example were performed, and the performance was quantified by MSE and compared. This process confirmed that the proposed method had a positive effect on the Bayesian model calibration.

5.2 Suggestions for Future Research

This thesis proposed the necessitation of a proper initial point of the MCMC method and a new initial point search algorithm. We demonstrated three case studies to validate the excellence of the proposed method.

Issue 1: Excessive simplification of the algorithm in higher dimensional space

In the case studies presented in this thesis, all distributions used a six-dimensional distribution. However, the algorithm's time complexity is very efficient as $O(N)$, performance degradation is concerned as the dimension increases. Therefore, it may be necessary to improve the algorithm to control the trade-off between computational cost and algorithm performance and analyze algorithm performance according to dimension.

Issue 2: Absence of evaluation process for the searched initial point in the implementation process

Although the proposed method can search the multi-modal distribution efficiently, all the initial values searched may not be proper. Suppose Bayesian model calibration can be selectively performed based on a metric that can evaluate the suitability of the selected initial value. In that case, the robustness of the proposed method can be improved. For this, it is necessary to devise a metric that can evaluate the suitability of the initial value.

Issue 3: Application to other fields using the MCMC algorithms

Although the developed method is an initial value study of MCMC for model calibration, the research motivation was to solve the challenge of using MCMC in the condition of insufficient prior knowledge. Therefore, it will be possible to expand the scope of application by solving problems in other fields that use MCMC.

References

1. Lee G, Kim W, Oh H, Youn BD, Kim NH. Review of statistical model calibration and validation-from the perspective of uncertainty structures. Struct Multidiscip O. 2019;60(4):1619-44.

2. Youn BD, Jung BC, Xi ZM, Kim SB, Lee WR. A hierarchical framework for statistical model calibration in engineering product development. Comput Method Appl M. 2011;200(13-16):1421-31.

3. Oberkampf WL, Helton JC, Joslyn CA, Wojtkiewicz SF, Ferson S. Challenge problems: uncertainty in system response given uncertain parameters. Reliab Eng Syst Safe. 2004;85(1-3):11-9.

4. Urbina A, Mahadevan S, Paez TL. Quantification of margins and uncertainties of complex systems in the presence of aleatoric and epistemic uncertainty. Reliab Eng Syst Safe. 2011;96(9):1114-25.

5. Jeong CH, Choi SH, Lee I. Bayesian inference of the flow resistivity of a sound absorber and the room's influence on the Sabine absorption coefficients (L) (a)). J Acoust Soc Am. 2017;141(3):1711-4.

6. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc B. 2001;63:425-50.

7. Liu JS, Liang FM, Wong WH. The multiple-try method and local optimization in metropolis sampling. J Am Stat Assoc. 2000;95(449):121-34.

8. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452-9.

9. Andrieu C, de Freitas N, Doucet A, Jordan MI. An introduction to MCMC for machine learning. Mach Learn. 2003;50(1-2):5-43.

10. van Ravenzwaaij D, Cassey P, Brown SD. A simple introduction to Markov Chain Monte-Carlo sampling. Psychon B Rev. 2018;25(1):143-54.

11. Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:170102434. 2017.

12. Mousaviraad SM, He W, Diez M, Stern F. Framework for Convergence and Validation of Stochastic Uncertainty Quantification and Relationship to Deterministic Verification and Validation. Int J Uncertain Quan. 2013;3(5):371-95.

13. Higdon D, Nakhleh C, Gattiker J, Williams B. A Bayesian calibration approach to the thermal problem. Comput Method Appl M. 2008;197(29-32):2431-41.

14. Hou D, Hassan IG, Wang L. Review on building energy model calibration by Bayesian inference. Renew Sust Energ Rev. 2021;143.

15. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis: CRC press; 2013.

16. Kuczera G, Kavetski D, Renard B, Thyer M. A limited-memory acceleration strategy for MCMC sampling in hierarchical Bayesian calibration of hydrological models. Water Resources Research. 2010;46(7).

17. Robert CP, Elvira V, Tawn N, Wu CY. Accelerating MCMC algorithms. Wires Comput Stat. 2018;10(5).

18. Gilks WR, Roberts GO. Strategies for improving MCMC. Markov chain Monte Carlo in practice. 1996;6:89-114.

19. Marti R, Resende MGC, Ribeiro CC. Multi-start methods for combinatorial optimization. Eur J Oper Res. 2013;226(1):1-8.

20. Chapra SC. Applied numerical methods with MATLAB for engineers and scientists: McGraw-Hill Higher Education; 2008.

21. Ollar J, Mortished C, Jones R, Sienz J, Toropov V. Gradient based hyper-parameter optimisation for well conditioned kriging metamodels. Struct Multidiscip O. 2017;55(6):2029-44.

22. Yazici I, Yaylaci EK, Yalcin F. Modified golden section search based MPPT algorithm for the WECS. Eng Sci Technol. 2021;24(5):1123-33.

23. Kaveh A, Dadras A, Malek NG. Robust design optimization of laminated plates under uncertain bounded buckling loads. Struct Multidiscip O. 2019;59(3):877-91.

24. Chang YC. N-Dimension Golden Section Search: Its Variants and Limitations. Int Conf Biomed. 2009:2219-24.

25. Cormen TH, Leiserson CE, Rivest RL, Stein C. Introduction to algorithms: MIT press; 2009.

26. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1-2):203-14.

27. Lee K, Cho H, Lee I. Variable selection using Gaussian process regression-based metrics for high-dimensional model approximation with limited data. Struct Multidiscip O. 2019;59(5):1439-54.

28. Auer P. Using confidence bounds for exploitation-exploration trade-offs. Journal of Machine Learning Research. 2002;3(Nov):397-422.

29. Russo D, Van Roy B. Learning to Optimize via Posterior Sampling. Math Oper Res. 2014;39(4):1221-43.

30. Sutton RS, Barto AG. Introduction to reinforcement learning: MIT press Cambridge; 1998.

31. Pokrajac D, Latecki LJ, Lazarevic A, Nikolic J. Variable width elliptic Gaussian kernels for probability density estimation. Telsiks 2007: 8th International

Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Vols 1 and 2. 2007:593-+.

32. Al-Momani N, Ebrahem MAH, Eidous O. Variable scale Kernel density estimation for simple linear degradation model. Electron J Appl Stat. 2021;14(2):359-72.

33. Blanchard P, Higham DJ, Higham NJ. Accurately computing the log-sum-exp and softmax functions. Ima J Numer Anal. 2021;41(4):2311-30.

34. Gournelos T, Kotinas V, Poulos S. Fitting a Gaussian mixture model to bivariate distributions of monthly river flows and suspended sediments. J Hydrol. 2020;590.

35. Song JM, Zhao SJ, Ermon S. A-NICE-MC: Adversarial Training for MCMC. Adv Neur In. 2017;30.

36. Son H, Choi H, Kim W, Youn BD, Lee G. A comparative study of statistical validation metrics with consideration of variance to address type II errors in statistical model validation. Struct Multidiscip O. 2022;65(2):1-19.

37. Kang P, Birtwhistle B. Condition assessment of power transformer on-load tap-changers (OLTC) using wavelet analysis. Pica 2001: 22nd Ieee Power Engineering Society International Conference on Power Industry Computer Applications. 2001:330-5.

38. Chen AY, Kurfess TR. Signal processing techniques for rolling element bearing spall size estimation. Mech Syst Signal Pr. 2019;117:16-32.

39. He W, Xiao HG, Liu XH. Numerical Simulation of Human Systemic Arterial Hemodynamics Based on a Transmission Line Model and Recursive Algorithm. J Mech Med Biol. 2012;12(1).

40. Xiao HG, Avolio A, Zhao MG. Modeling and Hemodynamic Simulation of

Human Arterial Stenosis Via Transmission Line Model. J Mech Med Biol. 2016;16(5).

41. Elhanaf A, Guaily A, Elsaid A. Numerical simulation of blood flow in abdominal aortic aneurysms: Effects of blood shear-thinning and viscoelastic properties. Math Comput Simulat. 2019;160:55-71.

Abstract (Korean)

부족한 통계적 정보 하에서 베이지안 보정을 위한 새로운 초기값 탐색 알고리즘: 탐욕스러운 확률론적 분할 탐색 방법

서울대학교 공과대학
기계공학부 대학원
이 현 찬

FEM 모델 등 Computer-Aided Engineering (CAE) 모델은 공학적 제품의 설계 및 운영 뿐 아니라 유지보수 과정에서도 중요한 역할을 수행한다. 그러나 CAE 모델은 내재적 불확실성 뿐만 아니라 모델링 과정에서 과도한 가정이나 지식의 부족으로 발생하는 인식론적 불확실성이 존재한다. 이러한 불확실성은 CAE 모델의 예측 성능을 저하하고 결국 유용성을 떨어뜨린다. 모델 보정은 모델의 적합성을 향상키는 과정을 말한다. 한편, 사물 인터넷 기술, 무선 센서와 계산 성능의 발전으로 물리적 시스템과 CAE 모델의 융합을 의미하는 디지털 트윈 기술이 주목을 받고 있다. 디지털 트윈의 개념하에서 물리적 시스템으로부터 취득된 데이터에 기반해 모델을 주기적으로 갱신할 필요가 있기 때문에, 모델 보정기술의 중요성은 점차 증가하고 있다.

베이지안 모델 보정은 통계적 모델 보정의 한 갈래로 취득된 데이터에 더해 사전지식을 확률 형태로 활용할 수 있다는 장점이 있다. 하지만 이 과정에서 복잡한 적분 계산을 해야 하는 문제점이 있어 다차원의 다봉 분포에 적용하기 어려움이 있다. 마코프 체인 몬테 카를로 방법을 활용한 몬테 카를로 적분으로 이러한 문제를 해결할 수 있다. 마코프 체인 몬테 카를로 방법을 사용할 때, 초깃값은 계산 비용의 낭비를 야기하는 번 인(burn-in)에 결정적인 영향을 주지만, 지금까지 초기값 선정은 엔지니어의 사전지식에 의존했다. 따라서 사전 지식이 부족한 경우 적합한 초기값 선정을 할 수 없어 결국 베이지안 모델 보정을 수행할 수 없었다. 이러한 문제를 해결하기 위해 본 학위 논문에서 사전 지식이 부족한 상황에서 베이지안 모델 보정에 사용되는 마코프 체인 몬테 카를로 방법을 위한 효율적인 확률론적 초기값 탐색 알고리즘을 제안한다. 제안된 알고리즘은 하나의 가우시안 혼합 모델을 사용한 수치 예시와 캔틸레버 빔의 유한 요소 모델을 사용하는 모델 보정 예시, 사람의 혈류 역학을 모사하는 트랜스미션 모델 (Transmission model)의 보정 예시를 통해 개발된 알고리즘의 유효성을 검증했다.

주제어: 베이지안 모델 보정 (Bayesian model calibration)
CAE 모델 (computer-Aided Engineering models)
디지털 트윈 (digital twin)
마코프 체인 몬테 카를로 (markov chain monte carlo)
초기값 탐색 알고리즘 (initial point search algorithm)

학 번: 2020-28845