

EFFICIENT MODEL CHOICE AND PARAMETER ESTIMATION BY USING NESTED SAMPLING APPLIED IN EDDY-CURRENT TESTING

C. Cai¹, T. Rodel² and M. Lambert¹

¹ L2S UMR8506, CNRS-SUPELEC-Univ Paris-Sud, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

² SATIE, ENS-Cachan, 61 avenue du Président Wilson, 94230 Cachan, France

ABSTRACT

In many applications, such as Eddy-Current Testing (ECT), we are often interested in the joint model choice and parameter estimation. Nested Sampling (NS) is one of the possible methods. The key step that reflects the efficiency of the NS algorithm is how to get samples with hard constraint on the likelihood value. This contribution is based on the classical idea where the new sample is drawn within a hyper-ellipsoid, the latter being located from Gaussian approximation. This sampling strategy can automatically guarantee the hard constraint on the likelihood. Meanwhile, it shows the best sampling efficiency for models which have Gaussian-like likelihood distributions. We apply this method in ECT. The simulation results show that this method has high model choice ability and good parameter estimation accuracy, and low computational cost meanwhile.

Index Terms— Nested Sampling, metamodeling, model choice, Bayesian, parameter estimation.

1. INTRODUCTION

In Non-Destructive Testing (NDT), our main research area, there are many demands for joint model choice and parameter estimation. Take the flaw detection and characterization problem [1, 2, 3, 4] for example, it will be very useful if the method can tell automatically what kind of flaw, a hole or a crack, that we are dealing with and what are the corresponding dimensions if it is a crack. The former is a model choice problem while the later is a parameter estimation problem. Nested Sampling (NS) is one of those methods which can jointly solve both of them. In [5], the author proposes several alternative methods, most of them based on Markov Chain Monte Carlo (MCMC) sampling method. Compared to NS, MCMC-based methods are much more expensive in computational cost.

In Bayesian inference, the model choice can be made by comparing the marginal likelihoods, also called evidences, of

different model assumptions. However, approximating the marginal likelihood is a very difficult task which is at least a power factor more expensive than the parameter estimation problem in computational cost. NS is first proposed in [6] in which the authors proposed to transform the multi-dimensional likelihood marginalization problem into a single-dimension integration problem. This reduces significantly the computational cost. Parameter estimation can then be carried out easily by averaging the posterior samples which are by-products of evidence approximation in NS. The most difficult step in NS is how to efficiently generate random samples from the given posterior law with the constraint that the likelihoods of the new samples are larger than a given threshold. Aiming at this, the authors proposed an ellipsoidal NS method in [7] and extended later to multi-nested sampling in [8, 9] for multimodal likelihood distributions. Most of the problems of our concern have unimodal likelihoods, so we are going to follow the same idea as in [7] to propose an efficient NS method for joint model choice and parameter estimation for Eddy Current Testing (ECT) applications.

The first main contribution of this work is to introduce metamodeling method [10] into NS for improving the computational efficiency. Secondly, a complementary discussion will be given on the location of hyper-ellipsoidal contours in NS. Suggestions on varying the active samples in order to improve the convergence rate of the algorithm will be proposed in particular.

2. PROBLEM STATEMENT

2.1. Bayesian inference for model choice

For simplicity, the following discussion is based on choice between two models. It can be directly extended to higher number of models.

For a model of concern \mathcal{M}_i where the corresponding unknown parameters are denoted by $\mathbf{x}_i \in \mathbb{R}^{N_i}$, the forward model with a general Gaussian noise model can be described by

$$\mathbf{y} = f_{\mathcal{M}_i}(\mathbf{x}_i) + \epsilon, \text{ for } i = 1, 2. \quad (1)$$

$\mathbf{y} \in \mathbb{C}^M$ are the observations and $f_{\mathcal{M}_i}(\mathbf{x}_i)$ is the function describing the physical phenomenon from \mathbf{x}_i to \mathbf{y} regarding

This work is part of the research project ByPASS. Thanks to the French National Agency of Research (ANR) for funding it and CEA-LIST for providing metamodel databases. Useful comments of D. Lesselier are acknowledged.

to the model \mathcal{M}_i . In Eq. (1), N_i is the number of unknown parameters in model \mathcal{M}_i ; M is the total dimension of observations; $\epsilon \sim \mathcal{N}(0, \Sigma)$ is the complex Gaussian noise where Σ is the covariance matrix.

Bayes law indicates that the posterior distribution $p(\mathbf{y}|\mathbf{x}_i, \mathcal{M}_i)$ can be rewritten as

$$p(\mathbf{y}|\mathbf{x}_i, \mathcal{M}_i) = \frac{p(\mathbf{y}|\mathbf{x}_i, \mathcal{M}_i) p(\mathbf{x}_i|\mathcal{M}_i)}{p(\mathbf{y}|\mathcal{M}_i)}, \quad (2)$$

where $p(\mathbf{y}|\mathbf{x}_i|\mathcal{M}_i)$ is the Gaussian likelihood function. $p(\mathbf{x}_i|\mathcal{M}_i)$ is the prior distribution of parameters in \mathcal{M}_i . With no further information, it can be assumed to be uniformly distributed. $p(\mathbf{y}|\mathcal{M}_i)$ is the evidence of model \mathcal{M}_i and the marginal probability of \mathbf{y} within the model of concern also. In Bayesian inference, model evidence is often used to select the correct forward model. However, it is very difficult to obtain in practice. One way, most likely the only way, is to integrate the product of the likelihood and the prior distribution over the entire parameter space which gives the evidence for model \mathcal{M}_i as

$$Z_i = p(\mathbf{y}|\mathcal{M}_i) = \int_{\mathbf{x}_i} p(\mathbf{y}|\mathbf{x}_i, \mathcal{M}_i) p(\mathbf{x}_i|\mathcal{M}_i) d\mathbf{x}_i. \quad (3)$$

This is the reason for which model evidence is also called marginal likelihood. The model which has the maximal evidence can then be selected as the correct one.

2.2. Principle of nested sampling

The evidence given in Eq. (3) is a multi-dimensional integration problem. It is still a very difficult problem to approximate it. Many methods [11, 12, 13, 14, 6] have been proposed for this objective, NS being one of them. It transforms the multi-dimension intergration into a single-dimension integration by introducing the prior mass X defined as

$$X_i(\lambda) = \int_{L_{\mathcal{M}_i}(\mathbf{x}_i) > \lambda} p(\mathbf{x}_i|\mathcal{M}_i) d\mathbf{x}_i, \quad (4)$$

where $L_{\mathcal{M}_i}(\mathbf{x}_i) = p(\mathbf{y}|\mathbf{x}_i, \mathcal{M}_i)$ is the likelihood as a function of \mathbf{x}_i regarding to \mathcal{M}_i . The prior mass can be seen as the hyper-volume of the prior distribution for parameters whose corresponding likelihoods are larger than a given value λ . So, it is a scalar function decreasing from 1 to 0 when the likelihood increases from 0 to its maximum. Given the prior mass defined as in Eq. (4), the evidence in Eq. (3) can then be rewritten as

$$Z_i = \int_0^1 L(X_i|\mathcal{M}_i) dX_i \approx \sum_{k=1}^K L(X_i^k|\mathcal{M}_i) \Delta X^k, \quad (5)$$

where $L(X_i|\mathcal{M}_i)$ is the same likelihood function as $L_{\mathcal{M}_i}(\mathbf{x}_i)$ but as a function of the prior mass X_i . It becomes an one-dimensional integration problem which can be approximated

by summing up sub-areas surrounded by the likelihoods and the prior mass values. One of the main challenges of NS method is to get proper samples which can result in a good approximation for Eq. (5) and meanwhile to have high computational efficiency.

A general NS algorithm scheme is presented in Algo. 1 where J is the number of active samples at each iteration, k is the iteration index. For the above sampling procedure,

Algorithm 1 a general nested sampling method.

- 1: Initializing J active samples $\{\mathbf{x}_i^j\}_{j=1, \dots, J}$ from the prior distribution $p(\mathbf{x}_i|\mathcal{M}_i)$, $\hat{Z}_i = 0$, $k = 1$.
 - 2: Locating the sample among all active samples which gives the minimum likelihood $L_{\min} = \min \{L_{\mathcal{M}_i}(\mathbf{x}_i^j), j = 1, \dots, J\}$.
 - 3: Generating a new independent sample from the rest prior volume where $L(\mathbf{x}_i|\mathcal{M}_i) > L_{\min}$ and using it to replace the one located in step 2.
 - 4: Up-dating the model evidence $\hat{Z}_i = \hat{Z}_i + L_{\min} \Delta X^k$ where the prior mass step $\Delta X^k = 0.5 (X_i^{k+1} - X_i^{k-1})$, $k = k + 1$.
 - 5: Repeating steps 2-4 until $\frac{X_i^k L_{\max}}{\hat{Z}_i} < r_\sigma$ where $L_{\max} = \max \{L_{\mathcal{M}_i}(\mathbf{x}_i^j), j = 1, \dots, J\}$ and r_σ is a threshold indicating the tolerance on maximal updating ratio for \hat{Z}_i .
 - 6: Increasing \hat{Z}_i by $\frac{1}{J} X_i^K \sum_{j=1}^J L_{\mathcal{M}_i}(\mathbf{x}_i^j)$, where K is the total number of iterations after terminating the loop from 2 to 5.
-

the prior mass can be approximated by $X_i^k \approx \exp(-\frac{k}{J})$ (see [6]). It is sorted out in descending order automatically. The trapezoidal increasing step ΔX^k used in step 4 is for better approximating the integration compared to the classical one $\Delta X^k = X_i^k - X_i^{k-1}$.

Samples from the posterior distribution are by-products of the evidence approximation. The Posterior Mean (PM) estimation and the corresponding uncertainty for the parameters can then be approximated by the average and the variance of all samples. In order to get correct uncertainty approximation, over-sampling near the likelihood maximum should be avoided. For this reason, a proper value for r_σ should be used in step 5.

2.3. Main difficulties and contributions

In NS, the first difficulty that arises is step 3 in Algo. 1: how to generate new independent sample subject to the constraint that the corresponding likelihood is greater than a given threshold? To deal with it, Skilling et al. [6] proposed to use an algorithm based on MCMC method which has the disadvantage of producing correlated samples. As pointed out in [5], the evidence approximation accuracy from correlated samples is still an open question. In [7], the authors suggest to use a so-called ellipsoidal nested sampling algorithm in which the hard constraint on the likelihood value is approximated by a hyper-ellipsoid. An enlarging factor is then employed to make sure that the parameter space is completely traversed. In addition, a rejection procedure is performed to ensure that

the new sample satisfies the constraint on the likelihood. Following the same approach, our contribution is completed with a detailed discussion on localization of the hyper-ellipsoidal contour and an analysis of the convergence speed with the aim of improving the computational efficiency.

The second difficulty is the high computational cost due to the likelihood evaluation for each sample. As seen from the NS algorithm scheme (Algo. 1), without considering the rejected samples, $(J + K)$ samples from the posterior law are available by the end. Each sample requires a likelihood evaluation, each needing a forward projection from \mathbf{x}_i to \mathbf{y} as described by the function $f_{\mathcal{M}_i}(\mathbf{x}_i)$ in Eq. (1). This requires the computation of at least $(J + K)$ forward models. Unfortunately, in many applications, the forward model evaluation can be computationally expensive, such as in ECT as discussed in § 3 and § 4. To overcome this difficulty, we propose to use a metamodeling method [10]. A database is pre-trained by using a simulation software CIVA where an accurate modeling method [15] is employed for the forward model. Then, at each forward model evaluation in NS, only a Kriging interpolation is needed which drastically reduces the computational burden. As shown in § 4, the evidence approximation for a three-parameter model can be performed within minutes instead of hours with the *exact* model. It should be mentioned that this metamodel-based NS method is efficient only for low-dimensional models ($N_i < 20$). Since most of the problems that we are dealing with are non-linear, for large-dimensional ones, both the database pre-training and the kriging interpolation can become computationally unfeasible.

3. METHOD

3.1. Hyper-ellipsoidal contour location

A new sample generation in ellipsoidal NS includes the following steps.

- S1. Locating the parameters \mathbf{x}_c , \mathbf{R} and L_c which define the hyper-ellipsoidal contour

$$\mathcal{C}_i: (\mathbf{x}_i - \mathbf{x}_c)^T \mathbf{R} (\mathbf{x}_i - \mathbf{x}_c) = L_c.$$

- S2. Generating a new sample $\mathbf{x}_i^{\text{new}}$ inside \mathcal{C}_i .

- S3. Rejecting the sample if $L_{\mathcal{M}_i}(\mathbf{x}_i^{\text{new}}) < L_{\min}$ and repeating S2.

For a given evidence approximation problem, the computational efficiency and the approximation accuracy depend on the choice of the hyper-ellipsoid in step 1. There are two criteria to verify whether the located \mathcal{C}_i is a good one. First, \mathcal{C}_i needs to be *large* enough to cover all regions that $L_{\mathcal{M}_i}(\mathbf{x}_i) < L_{\min}$. This is to make sure that all regions are traversed during the iterations so that the final approximated evidence has high accuracy. Second, \mathcal{C}_i needs to be *small* enough so that the rejection ratio can be low enough for what concerns the computational cost.

As addressed in [7], \mathbf{x}_c can be set at the center of active samples, \mathbf{R} can be estimated by inverting the covariance matrix of active samples. As for L_c , it is said in [7] that *create an ellipsoid that just touches the maximum coordinate values of the existing points*. If it means that \mathcal{C}_i should pass through the farthest active sample from the center, then it will be inappropriate since one cannot make sure to have all active samples inside \mathcal{C}_i . We suggest to use L_c located as follows

$$L_c = \max \left\{ (\mathbf{x}_i^j - \mathbf{x}_c)^T \mathbf{R} (\mathbf{x}_i^j - \mathbf{x}_c), j = 1, \dots, J \right\}. \quad (6)$$

This method is equivalent to approximate the likelihood by a Gaussian distribution which leads to a computationally efficient algorithm for Gaussian-like likelihood distributions. To get independent samples uniformly distributed in a hyper-ellipsoid, it can be done by generating random samples inside of a n -sphere first [16, 17], then performing dilatation and rotation according to \mathbf{R} calculated by Eq. 6. We are going to use the Gaussian projection method discussed in [17] and proven later in [18] to be one of the most efficient.

3.2. Discussion on convergence rate

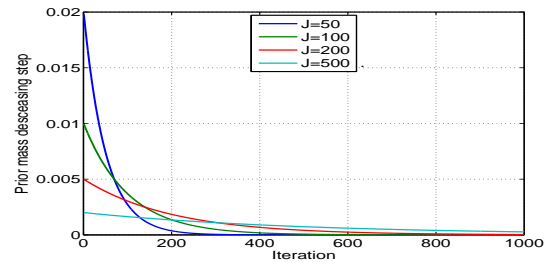


Fig. 1: convergence rate in terms of prior mass: $r_c(X_i^k)$ v.s. iterations for $J = 50, 100, 200$ and 500 .

The decreasing step per iteration for the prior mass is

$$r_c(X_i^k) = X_i^k - X_i^{k+1} \approx C(J) \exp \left\{ -\frac{k}{J} \right\}, \quad (7)$$

where J is the number of active samples. The first term $C(J) = 1 - \exp \left\{ -\frac{1}{J} \right\}$ depends on the number of active samples, the second depends also on the iteration index k . In brief, the convergence rate decreases rapidly along the iterations. As shown in Fig. 1(a) for several common used values for J , the convergence rate in terms of the prior mass descends so rapidly that $r_c(X_i^k)$ becomes less than 0.1% of the initial prior mass $X_i^{k=0} = 1$ when $k > 200$. In situations with sharp likelihood distributions, such as the 3D flaw characterization discussed in § 4, the consequence of this rapid descending rate makes the prior mass step very small before burning-up.

Considering the significant influence of J on the convergence rate, the convergence rate can be improved by varying the active numbers during the iterations for the aim of

reducing the first term in Eq. (7) to compensate the fast descent of the second term. However, attention must be paid on the choice of J , because the evidence approximation error is dominated stochastically by a term $O(\frac{1}{\sqrt{J}})$ as shown in [5].

4. SIMULATIONS

4.1. Example in Eddy-Current Testing

To test our metamodel-based NS method, an ECT example as shown in Fig. 2 is going to be used. The goal is to retrieve the width w and the length l of a single crack from the measurement of the variation of impedance of a coil moving at a height of l_0 above the plate. The configurations for the two models under discussion are listed in Tab. 1.

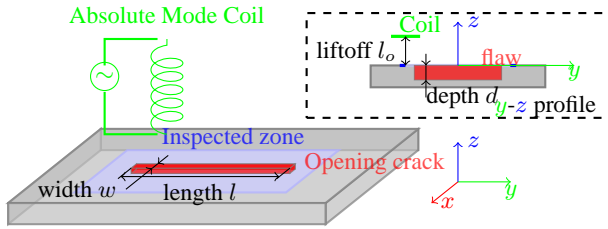


Fig. 2: test example in Eddy-Current Testing: an opening crack present in a plate of Inconel 600 with thickness of 1.55 mm. The $y - z$ profile is shown at the top right

Table 1: model configurations for the simulation test.

Model	unknown parameter	fixed parameter
\mathcal{M}_1	d, l and w	$l_0 = 0.303$ mm
\mathcal{M}_2	d, l, w and l_0	—

4.2. Model choice and flaw characterization results

In order to test our model choice, we first simulate the data \mathbf{y} by using the model \mathcal{M}_1 where Gaussian noises are added as well based on Eq. (1). Then, we estimate the model evidences for \mathcal{M}_1 and \mathcal{M}_2 by using the ellipsoidal NS method discussed in § 3. Unknown parameters and the corresponding estimation uncertainties are estimated from NS samples at the end. The results are given in Tab. 2 where the better results are marked in bold. We see that the model evidence for \mathcal{M}_1 is larger than that for \mathcal{M}_2 , which means that \mathcal{M}_1 is the *correct* model, even though \mathbf{y} is simulated using the same $l_0 = 0.303$ mm as in \mathcal{M}_2 . In terms of computational time, \mathcal{M}_2 costs five times less than \mathcal{M}_1 . This is because the computational cost for metamodel kriging interpolation grows exponentially as a function of the dimension of unknown parameters. This is also why this metamodel-based NS method is only suitable for small dimensional problems.

Table 2: nested sampling results of model evidence and parameter estimation by using simulated data from model \mathcal{M}_1 with Gaussian noise SNR = 20 dB.

Model		\mathcal{M}_1	\mathcal{M}_2
	<i>true value (mm)</i>		
Estimated parameters	$d = 0.85$	0.85 \pm 0.06	0.83 \pm 0.06
	$l = 18$	18.08 \pm 0.44	17.83 \pm 0.46
$\hat{\mathbf{x}} \pm \sigma_e$ (mm)	$w = 0.15$	0.15 \pm 0.02	0.17 \pm 0.02
	$l_o = 0.303$	0.31 \pm 0.06	–
Estimated evidence	\hat{Z}	1.4 $\times 10^{-985}$	8.1 $\times 10^{-1100}$
Time t (min.)		40	5

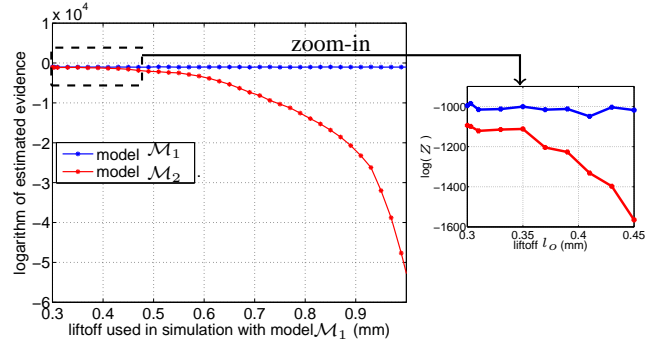


Fig. 3: estimated logarithm model evidences v.s. varying liftoff l_0 from 0.3 mm to 1.0 mm used in data simulation with model \mathcal{M}_1 .

By varying the liftoff used in the data simulation and performing the same model choice, we obtain the curve shown in Fig. 3. We see that the evidence difference becomes larger and larger when l_0 goes far away from $l_0 = 0.303$ mm. The model choice becomes also much easier.

5. CONCLUSION

In this work, we have discussed the metamodel-based NS method for joint model choice and parameter estimation. A detailed description on generation of random samples in ellipsoidal NS with constraint on the likelihood value by using Gaussian approximation is given. To analyze the performance of this metamodel-based NS method, simulation tests are performed based on an ECT example. The results show that the method has high selection ability for model choice, since even a small difference caused by metamodel interpolations can be distinguished, and it has high estimation accuracy for unknown parameters. In terms of the computational cost, the latter increases exponentially as a function of the unknown parameter dimension. This shows that this algorithm is only efficient for solving low-dimension problems. However, our analysis shows that the computational efficiency can be improved by varying the active samples in NS. This will be one of our future works.

6. REFERENCES

- [1] M. V. Kreutzbruck, K. Allweins, T. Ruhl, M. Muck, C. Heiden, H.-J. Krause, and R. Hohmann, "Defect detection and classification using a SQUID based multiple frequency eddy current NDE system," *IEEE Transactions on Applied Superconductivity*, vol. 11, no. 1, pp. 1032–1037, Mar. 2001.
- [2] D. Premel and A. Baussard, "Eddy-current evaluation of three-dimensional flaws in flat conductive materials using a Bayesian approach," *Inverse Problems*, vol. 18, no. 6, pp. 1873–1889, 2002.
- [3] J. I. De La Rosa, G. A. Fleury, S. E. Osuna, and M. E. Davoust, "Markov chain Monte Carlo posterior density approximation for a groove-dimensioning purpose," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 1, pp. 112–22, 2006.
- [4] T. Khan and P. Ramuhalli, "A recursive Bayesian estimation method for solving electromagnetic nondestructive evaluation inverse problems," *IEEE Transactions on Magnetism*, vol. 44, no. 7, pp. 1845–55, July 2008.
- [5] N. Chopin and C. P. Robert, "Properties of nested sampling," *Biometrika*, vol. 97, no. 3, pp. 741–755, 2010.
- [6] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Analysis*, vol. 1, no. 4, pp. 833–680, 2006.
- [7] P. Mukherjee, D. Parkinson, and A. R. Liddle, "A nested sampling algorithm for cosmological model selection," *The Astrophysical Journal*, vol. 638, pp. L51–L54, 2006.
- [8] F. Feroz, M. P. Hobson, and M. Bridges, "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 1601–1614, 2009.
- [9] F. Feroz, M.P. Hobson, E. Cameron, and A. N. Pettitt, "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics," *Instrumentation and Methods for Astrophysics*, vol. preprinted on arXiv, pp. 1–28, 2014.
- [10] S. Bilicz, M. Lambert, and Sz. Gyimthy, "Kriging-based generation of optimal databases as forward and inverse surrogate models," *Inverse Problems*, vol. 26, no. 7, pp. 074012–27, 2010.
- [11] A. Gelman and J. Hill, "Bayesian model choice: asymptotics and exact calculations," *Journal of the Royal Statistical Society*, vol. B 56, pp. 501–14, 1994.
- [12] P. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–32, 1995.
- [13] X. Meng and W. Wong, "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration," *Statistica Sinica*, vol. 6, pp. 831–60, 1996.
- [14] X. Meng and S. Schilling, "Warp bridge sampling," *Journal of Computational and Graphical Statistics*, vol. 11, pp. 552–66, 2002.
- [15] R. Miorelli, C. Reboud, T. Theodoulidis, N. Poulakis, and D. Lesselier, "Efficient modeling of ECT signals for realistic cracks in layered half-space," *IEEE Transactions on Magnetism*, vol. 49, no. 6, pp. 2886–92, June 2013.
- [16] M. E. Muller, "A note on a method for generating points uniformly on n -dimensional spheres," *Communications of the ACM*, vol. 2, no. 4, pp. 19–20, 1959.
- [17] D. E. Knuth, *The Art of Computer Programming*, vol. 2. Seminumerical Algorithms, chapter 3, pp. 1–136, Addison-Wesley publishing company, second edition, 1981.
- [18] J. Poland, "Three different algorithms for generating uniformly distributed random points on the n -sphere," 2000, <http://www-alg.ist.hokudai.ac.jp/~jan/randsphere.pdf> (last consultation the 6th of October 2014).