Université de Liège

Faculté des Sciences appliquées
Département d'aérospatiale et mécanique

Uncertainty quantification and stochastic modelling

Referring professor: Arnst Maarten

Bayesian inference

Alessandro Pase - ID: S2304388
Eugenio Squinzi - ID: S2303701

Academic year 2023/2024

Contents

List of Figures

List of Tables

1. Introduction

The report aims to present the results obtained following an analysis based on the application of a Markov chain Monte-Carlo method to explore the solution to a probabilistic inverse problem. First the posterior probability density function will be calculated which, in the Bayesian formulation of the inverse problem, describes the state of knowledge we have about the data as it follows from prior information, then a Markov chain Monte Carlo method will be implemented to sample from the Bayesian posterior. The sampling method implemented is the Metropolis-Hasting method. The effect of the σ value will then be studied by varying it from the initial case and evaluating how this change impacts the results. The convergence properties of the chain, taking into account the mean, will be studied using tools available in the literature. In conclusion, an adaptive method will then be implemented to ensure convergence in a real case in which the posterior function is not easily integrated.

1.1. Physical model

The model from which the data comes is a seismic surveyor with six different stations.

$$\begin{aligned} (x_1, y_1) &= (3,\ 15)\ \text{km}, & (x_2, y_2) &= (3,\ 16)\ \text{km}, \\ (x_3, y_3) &= (4,\ 15)\ \text{km}, & (x_4, y_4) &= (4,\ 16)\ \text{km}, \\ (x_5, y_5) &= (5,\ 15)\ \text{km}, & (x_6, y_6) &= (5,\ 16)\ \text{km}. \end{aligned}$$

Table 1: Position of the six seismic stations

The experiment starts with the activation of the seismic source and the arrival times of the seismic wave at the station are recorded and are:

$$\begin{aligned} t_1^{obs} &= 3.12\ \text{s}, & t_2^{obs} &= 3.26\ \text{s}, \\ t_3^{obs} &= 2.98\ \text{s}, & t_4^{obs} &= 3.12\ \text{s}, \\ t_5^{obs} &= 2.84\ \text{s}, & t_6^{obs} &= 2.98\ \text{s}. \end{aligned}$$

Table 2: Observed arrival times of the seismic waves

It is assumed that the potential impact of the experimental uncertainties on these data can be modelled in terms of independent additive Gaussian random variables with a mean value of 0 and a standard deviation of $\sigma = 0.10$ s. The seismic waves travel at a velocity of $v = 5$ km/s. The coordinates of the seismic stations and the velocity of the seismic waves are assumed to be perfectly known.
It is assumed that all seismic waves propagate isotropically in all directions and, therefore, for a given $(x,\ y)$, the arrival of the seismic waves at the seismic stations are

$$t_j = \frac{1}{v}\sqrt{(x_j - x)^2 + (y_j - y)^2}, \quad j = 1, ..., 6$$

2. Posterior probability density function

Bayesian inference is a powerful framework for making statistical inferences, which leverages the principles of probability theory to update our beliefs about a particular hypothesis or model as we acquire new data. At the heart of Bayesian inference are two fundamental concepts: the likelihood and the posterior probability density function. These components play a crucial role in quantifying uncertainty, making predictions and informing decision-making processes. The likelihood is a fundamental concept in Bayesian inference, representing the probability of observing the data given a particular hypothesis or model. Formally, if we denote the likelihood as $l(\boldsymbol{\theta})$, it quantifies how well the parameters $\boldsymbol{\theta}$ of our model explain the observed data. Mathematically, it is expressed as:

$$l(\boldsymbol{\theta}) = \pi(\mathbf{x^{obs}}|\boldsymbol{\theta})$$

In this equation, $\pi(\mathbf{x}^{\mathbf{obs}}|\boldsymbol{\theta})$ represents the probability of observing the data, assuming that the model parameters are set to $\boldsymbol{\theta}$, which in our case is the epicenter of the seismic event.
The posterior probability density function is another crucial element in Bayesian inference. It represents our updated beliefs about the model parameters after taking into account the observed data. The posterior PDF is denoted as $\pi(\boldsymbol{\theta}|\mathbf{x}^{\mathbf{obs}})$ and is calculated using Bayes' theorem:

$$\pi(\boldsymbol{\theta}|\mathbf{x}^{\mathbf{obs}}) = l(\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta})$$

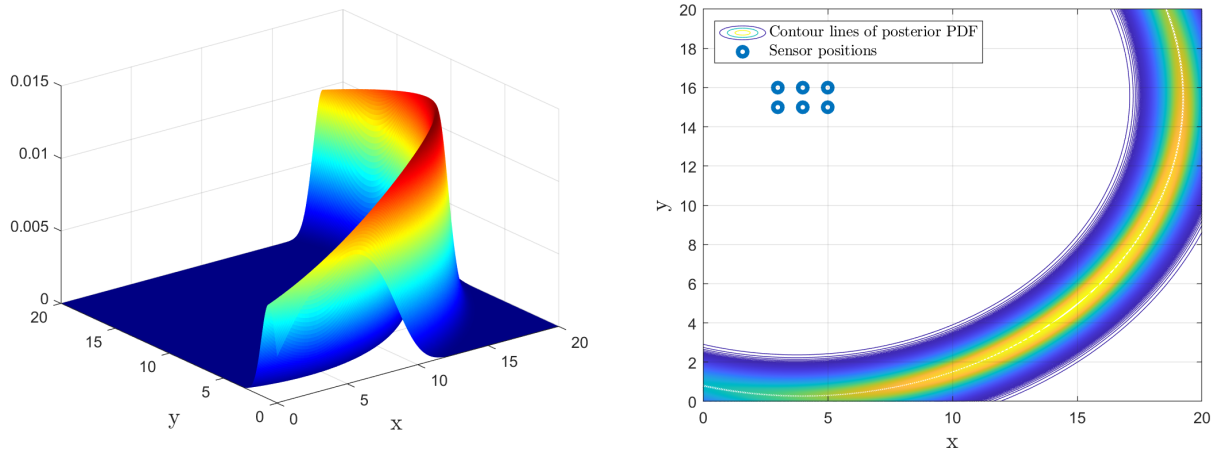
In this equation, $l(\boldsymbol{\theta})$ is the likelihood previously defined and $\pi(\boldsymbol{\theta})$ is the prior probability distribution, which represents our initial knowledge or beliefs about the parameters of a model before observing the data. The posterior PDF encapsulates our updated knowledge about the model parameters, taking into account both our prior beliefs (prior distribution) and the information contained in the data.

2.1. Posterior probability density function calculation and analysis

Considering the problem presented in Section 1.1, we have no information on the epicentre, so we can consider a uniform prior probability density function: $\pi(x, y) \propto c$. This constant will then be chosen so that the posterior probability density function integrated over the entire domain is 1.
Because the experimental uncertainties are assumed to be representable in terms of independent additive Gaussian random variables with a mean value of 0 and a standard deviation of σ, the likelihood function in the Bayesian formulation of the inverse problem is given by

$$l(x, y) = \prod_{j=1}^{6} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(t_j - t_j^{\text{obs}})^2}{2\sigma} \right)$$

Thus, the posterior probability density function can be obtained using the Bayesian formulation of the inverse problem.



(a) Surface plot of the posterior PDF

(b) Contour plot of the posterior PDF

Figure 1: Graphical representation of the posterior probability density function

The posterior probability density function identifies the probability that the epicentre is located at a specific point. The posterior probability density function is consistent with what could be expected from the data in that the rightmost points in the reference system receive the signal sooner and, given an x, the southernmost station receives the signal sooner than the one to the north. As can be seen from Fig. 1, one area can be identified as highly probable than the others, which is in agreement with the previous observation, but it cannot be excluded that the epicentre may not also be located far from this area, since there is uncertainty on the datas. The function also presents a circular distribution. This phenomenon is coherent with the model as seismic waves propagate isotropically.

3. Markov chain Monte Carlo method

3.1. Introduction to the Metropolis-Hasting sampling method

The Bayesian posterior is, in general, a multivariable probability density function, and if the problem size is greater than 2, it becomes complex to use it for calculation, visualisation and exploration. This is why the Bayesian posterior sampling technique is often used [3].
To sample directly from a known probability function, except in very simple cases, a direct method is not available, but Markov chain Monte Carlo (MCMC) methods are often used, which aim to construct a Markov chain in such a way that the probable probability distribution to be analysed is an invariant probability distribution. In MCMC methods the goal is to construct a transition kernel $K(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta})$ for which $\pi(\boldsymbol{\theta}|\mathbf{x^{obs}})$ is an invariant density, the target Bayesian posterior [3]:

$$\pi(\hat{\boldsymbol{\theta}}|\mathbf{x^{obs}}) = \int_{\mathbb{R}^m} K(\hat{\boldsymbol{\theta}}|\mathbf{x^{obs}}) d\boldsymbol{\theta}$$

The algorithm used for sampling is the random-walk Metropolis-Hasting method which involves the use of a proposal density function from which a sample is sampled and accepted or not in the chain with a probability $\alpha(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{k-1})$ defined as [3]:

$$\alpha(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{k-1}) = \min\left\{1, \frac{\pi(\hat{\boldsymbol{\theta}}|\mathbf{x^{obs}}) q(\boldsymbol{\theta}^{k-1}|\hat{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta}^{k-1}|\mathbf{x^{obs}}) q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{k-1})}\right\}$$

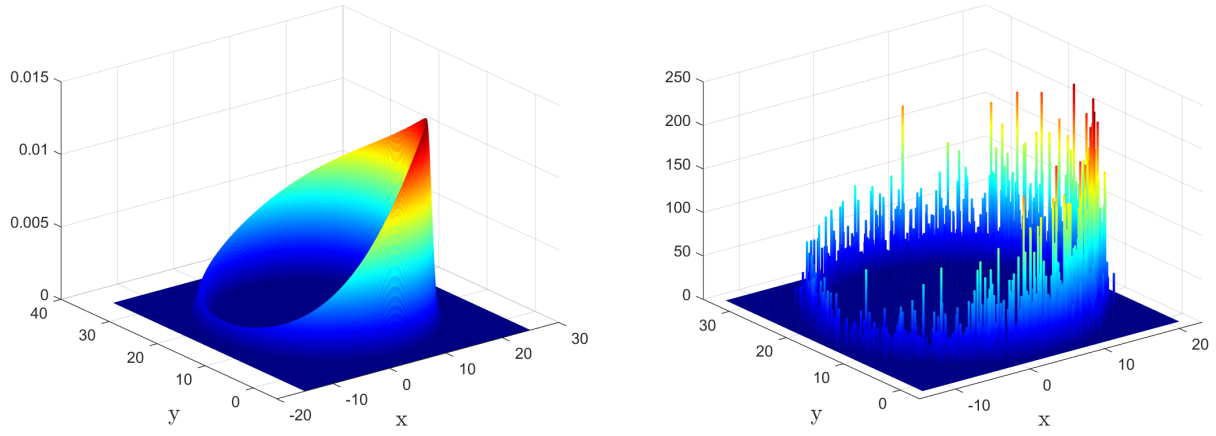
The random-walk Metropolis-Hasting method, which is the method implemented in the presented project, uses a multivariate Gaussian probability density function as a proposal density function [3]:

$$q(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}^{k-1}) = \exp\left(-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{k-1})^T [C]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{k-1})\right)$$

As a choice of matrix $[C]$ one can use the approximation $[C] = \frac{2.4^2}{d}[\Sigma]$ with d the problem size and $[\Sigma]$ the covariance matrix estimate of the target probability distribution [2].
The choice of the first sample, on the other hand, is random so as to eliminate any bias on the method, as long as it is at a point where the posterior PDF is non-zero.

3.2. Presentation of the results

The results of the first analysis are presented, using the model provided in the problem statement without variations. A diagonal formulation is used for the matrix $[\Sigma]$, using the variance values calculated numerically by integrating the posterior PDF.
As can be seen from Fig. 6, the distribution follows the pattern of the posterior PDF well.



(a) Surface plot of the posterior PDF

(b) 3D histogram of the MCMC samples - iterations: 1e5

Figure 2: Comparison of sample distribution and true posterior PDF

Another indicator of the goodness of the MCMC is the trend of the mean. As can be seen in Fig. 3, the mean as the iterations change converges to the value of the mean calculated numerically from the posterior probability density function. The *burn-in* technique was used to analyse the convergence: it consists in discarding the first part of the samples of the chain to reduce the effect of the random starting value and acheive a better chain convergence, as it will be further discussed in Section 5.

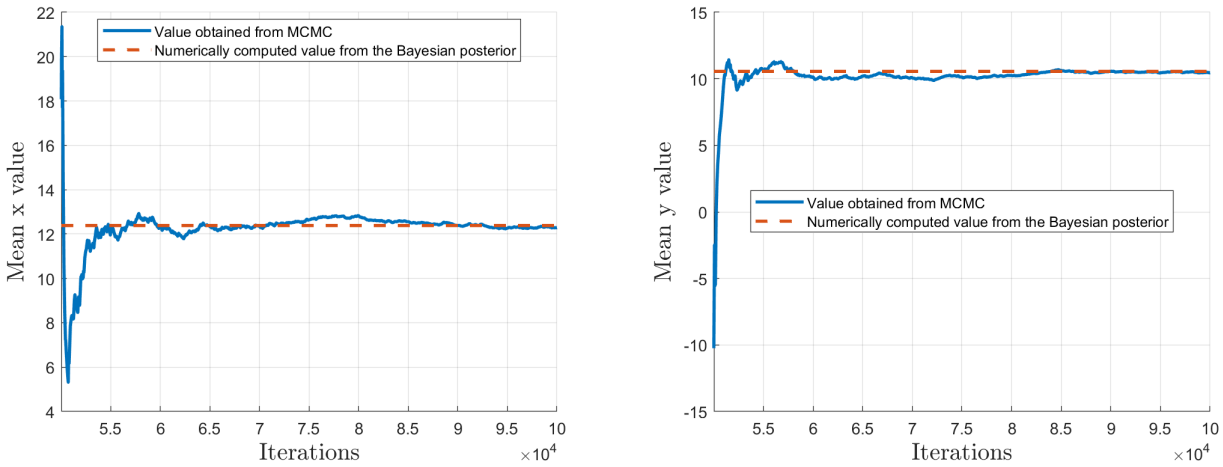


Figure 3: Convergence of the mean values with *burn-in* technique

Although there is good convergence of the mean, the sampling method choices made are not perfect in the case under analysis. The covariance matrix has very high values and therefore the Gaussian jumping distribution is very *open*: this causes the jump between one sample and the next to be very large. The large distance between the two samples results in a low probability of the new sample being accepted and consequently the acceptance rate is very low. In the case under analysis, an acceptance rate of 0.05 is obtained, a long way from the optimum rate of 0.23 (if the size is greater than 1) [2].
A further indication that the scale of the jumping distribution is too large can be obtained by viewing the chain on top of a contour plot of the posterior PDF. As can be seen from Fig. 4, the distribution of points jumps from one side of the contour to the other, confirming what was analysed above.

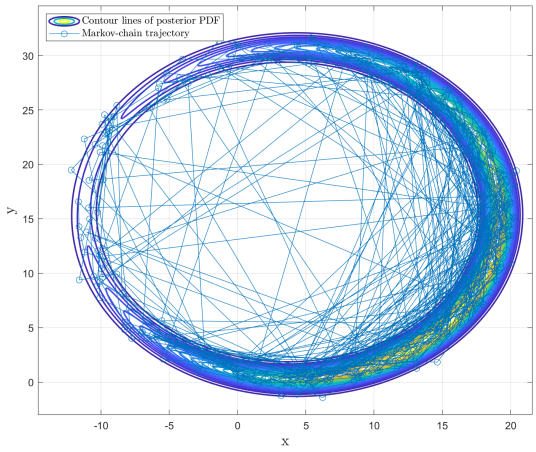
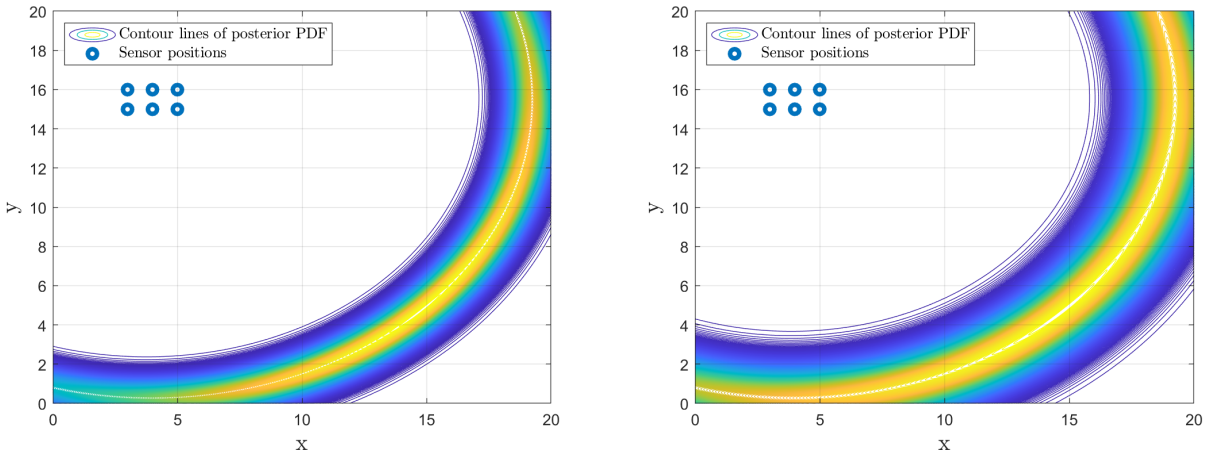


Figure 4: Trajectory of the chain on top of a contour plot of the Bayesian posterior probability density function - Number of samples: 1e4, for visualization reasons

One solution to this problem is to scale the $[C]$ matrix so as to tighten the jumping distribution and reduce the distance between samples. The scaling must be done carefully, however: if you reduce it too much, each sample will be close to the previous one, which results in a strong difficulty to explore the function well, thus affecting the convergence of the chain. The choice of this scaling coefficient related to convergence will be analysed in detail in Section 5.

4. Case for a higher incidence of experimental uncertainty

In the present Section of the report, the analysis performed with $\sigma = 0.10$ are repeated considering $\sigma = 0.25$, so modifying the incidence of experimental uncertainty. The first step is to re-evaluate the probability posterior and compare it with the one obtained in Chapter 2.

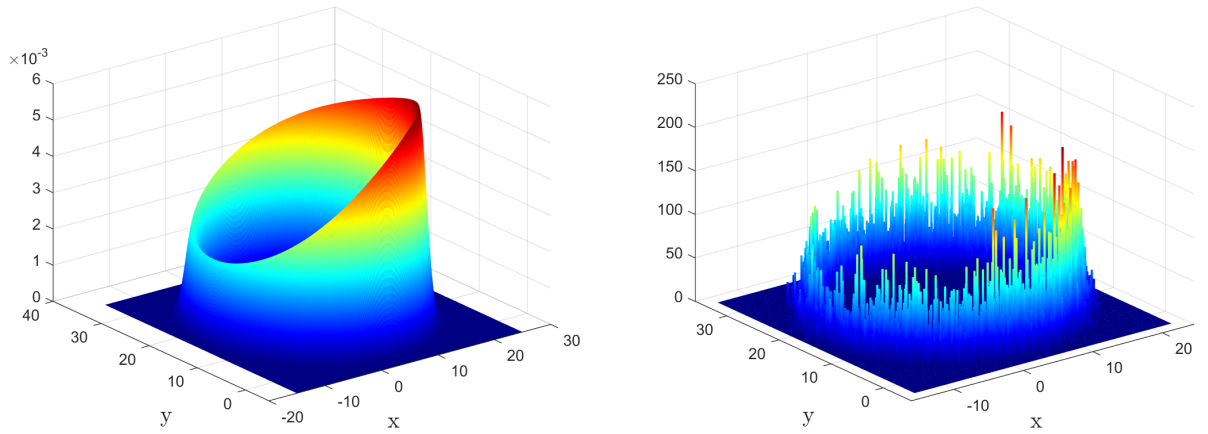


(a) Contour plot with $\sigma = 0.1$

(b) Contour plot with $\sigma = 0.25$

Figure 5: Comparison between the contour plots with two different values of σ

A greater standard deviation results in a probability distribution that maintains the same overall shape as the previous scenario, but is considerably broader. This means that the range of values where the likelihood of the epicenter occurring is significant becomes more wide.
The analysis, then, proceeds with the application of the random-walk Metropolis-Hastings sampling MCMC to the new posterior probability density function. Since the shape of the distribution is the same as in the previous case, the same type of MCMC can be deployed to explore the parametric field: so, once again a normal initial proposal distribution is adopted; however, in this case an update of the components of the covariance matrix is needed: in order to have an equivalent analysis to the one of Section 3, the diagonal components must be updated to the variances in x and y of the new posterior. Results are studied in the same way as the ones obtained for $\sigma = 0.10$ in Section 3.



(a) Surface plot of the posterior PDF - $\sigma = 0.25$

(b) 3D histogram with $\sigma = 0.25$ - iterations: 2e5

Figure 6: Comparison of sample distribution and true posterior PDF

Several noteworthy observations can be made regarding the results. Firstly, the acceptance rate of the Markov chain increases from 0.05 for $\sigma = 0.10$ to 0.08 for $\sigma = 0.25$. This is a consequence of the Bayesian posterior being more extensive, which, in turn, boosts the probability

of accepting samples. Secondly, when using $n = 100000$ as was the case with the previous value of σ, the representation of the target posterior becomes less accurate, and convergence of the means in both x and y is not achieved. This deficiency arises because the portion of the parametric space where the posterior probability is substantial is significantly broader, necessitating a higher number of samples to explore it with sufficient precision. Consequently, the number of samples in the chain was increased to $n = 200000$, resulting in a notably precise representation of the target posterior probability and successful convergence of the means within a small tolerance.

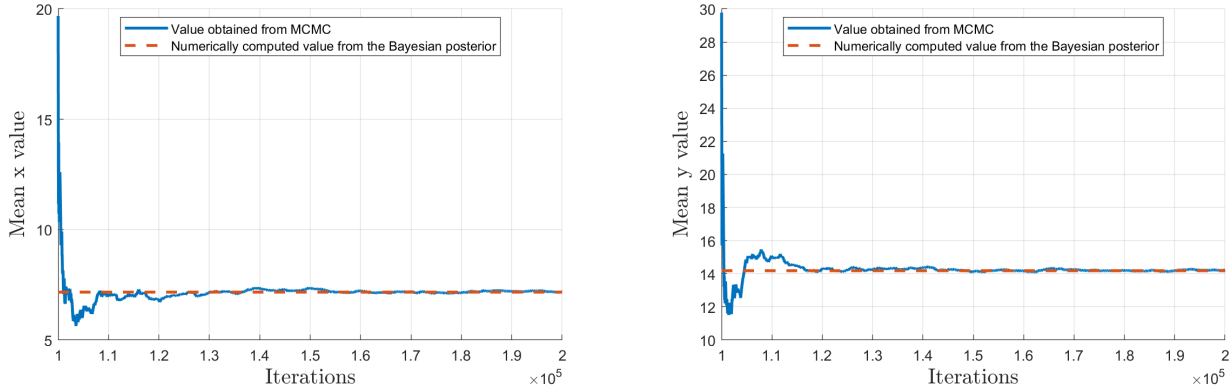


Figure 7: Convergence of the mean value with *burn-in* technique - $\sigma = 0.25$

5. Study of MCMC convergence and regularity

5.1. Introduction and error bound analysis

As previously said, the idea at the basis of the deployment of a Monte Carlo Markov chain is to explore the parametric space and obtain a set of random samples able to provide a representative picture of a target probability distribution. However, the possibility to use the samples to extract statistical quantities with enough precision to characterize the phenomenon depends on the so-called *mixing properties* of the chain itself: in particular, the present section of the report focuses specifically on the study of MCMC convergence, regularity and the evaluation of their effects on the representation drawn from the samples.

Firstly, convergence is the ability of the MCMC algorithm to reach a stationary distribution whose samples accurately represent the target distribution; on the other hand, regularity means that the MCMC is able to reach every region of the parametric field with positive probability under the target distribution. So, it is clear that a sufficient regularity of the chain is a prerequisite for the convergence of the chain itself: if the way the chain explores the domain is incorrect, the representation of the target probability distribution will be biased and results will not converge.

The topic of convergence is widely discussed by the authors Casella and Robert [4]: indeed, they focus on the use of the samples of the MCMC to predict the statistical properties of the target distribution. In more depth, they provide insights to the behaviour of error bounds of mean and standard deviation with respect to the number of samples in the chain based on the extension of the Central Limit Theorem, henceforth referred as CLT: CLT states that, as the number of samples increases, the distribution tends to the normal distribution; so, under this assumption, the error bound of the mean is given by $\sigma/\sqrt{n}$ which means that the precision of the estimate of the mean grows if a larger sample size is considered. To get an example of this behavior in our case study, a set of 10 different chains were created and their mean was calculated and plotted with respect to the number of iterations; as previously done in Section 3, the *burn-in* technique is applied. It can be easily observed that the bound predicted by the CLT provides a good estimation of the behaviour of the mean.

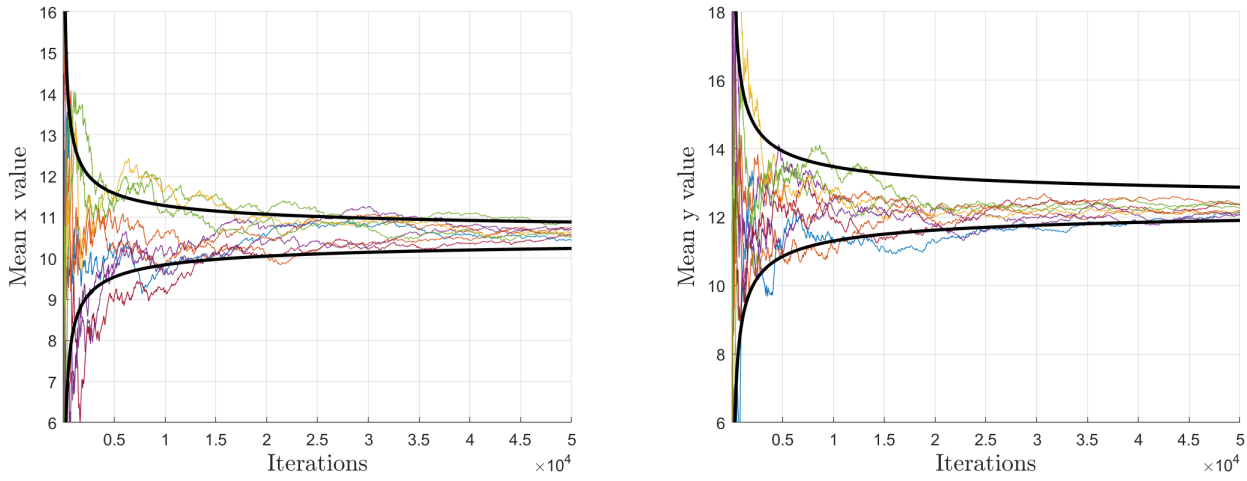


Figure 8: Mean error bounds for a set of different Markov chains

5.2. Convergence diagnostic

The importance of convergence assessment to ensure enough result reliability is strongly emphasized by Casella and Robert: such assessment can be performed following different procedures, some of which are hereby presented. The first possible approach is to inspect the behaviour of the statistical quantities of the sample pool for increasing number of samples in the chain: if convergence happens, the quantities should settle to a good approximation of the ones of the target probability distribution. A second possibility is the Gelman-Rubin convergence diagnostic [1]: the base concept is to run $m > 1$ different chains of n samples and study how samples are correlated inside each chain and between chains. An important observation to be made is that Gelman and Rubin suggest to pre-process the samples obtained from the chain by applying the *burn-in* technique: following Gelman and Rubin's procedure, the first half of the samples is discarded. The *within-chain* variance W and the *between-chain* variance B (variance of the means of the different chains) are evaluated and, then, the empirical variance of all chains combined is obtained as:

$$\hat{\sigma}^2 = \frac{(n-1)W}{n} + \frac{B}{n}$$

Convergence check is based on the assumption that the distribution to be studied is Gaussian, while a t-student distribution can be used to obtain a Bayesian credible interval; its mean $\hat{\mu}$ is obtained as the mean of all the samples from each chain, while its variance and degrees of freedom can be computed as:

$$\hat{V} = \hat{\sigma}^2 + \frac{B}{nm}$$

$$d = \frac{2\hat{V}^2}{Var(\hat{V})}$$

Lastly, the $\hat{R}$ parameter is obtained as:

$$\hat{R} = \sqrt{\frac{(d+3)\hat{V}}{(d+1)W}}$$

If all chains have converged, this parameter should be close to 1: as a rule of thumb, convergence is ensure with $\hat{R} \leq 1.1$. This type of analysis was implemented and convergence was studied: in particular, a parametric analysis was performed considering 4 values of the scaling factor (1, 0.1, 0.01, 0.001) and 4 sample sizes (100000, 50000, 10000, 1000). For each case, 10 chains were simulated and the maximum $\hat{R}$ parameter is reported in Tab. 3.

	100000	**50000**	**10000**	**1000**
1	1.0010	1.0016	1.0068	1.5639
0.1	1.0019	1.0128	1.0733	1.6728
0.01	1.0042	1.0164	1.5069	1.7902
0.001	1.2225	1.2489	1.6919	1.7767

Table 3: Convergence indices from Gelman-Rubin diagnostic

The first observation to be made is that 1 and 0.1 factors provide good convergence even if few samples are considered: this allows to obtain reliable results while also reducing the computational cost of the MCMC and the post-processing of the samples. On the other hand, the results obtained for 0.01 and 0.001 show that convergence is guaranteed only for an higher number of samples.
Indeed, if strong oscillations and no convergence are observed in the behaviour of the statistical parameters estimated from the samples of a Metropolis-Hastings MCMC, the most likely reason comes from issues in chain regularity: in other words, in the way the chain explores the parametric space. This once again highlights the need to perform a convergence check whenever working with MCMC. Many ways to deal with this kind of problem are available: in first instance, the way the chain explores the domain can be changed by considering a larger sample size and experimenting with the starting sample. However, if the aim of the analysis is to obtain a reliable tool for an efficient and effective sampling regardless of the initial guess, the way to proceed is to act on the initial proposal distribution: once the distribution type is chosen, its parameters must be tuned in order to obtain sufficient chain regularity and convergence, as was pointed out in Section 3; an example of this procedure can be observed in Table 3 regarding the choice of the scaling factor of the covariance matrix. Furthermore, another possibility is to develop adaptive algorithms that are able to change such parameters for each chain iteration to guarantee the satisfaction of the mixing properties of the chain, regularity and convergence; the development of an adaptive algorithm is the topic of the next chapter.

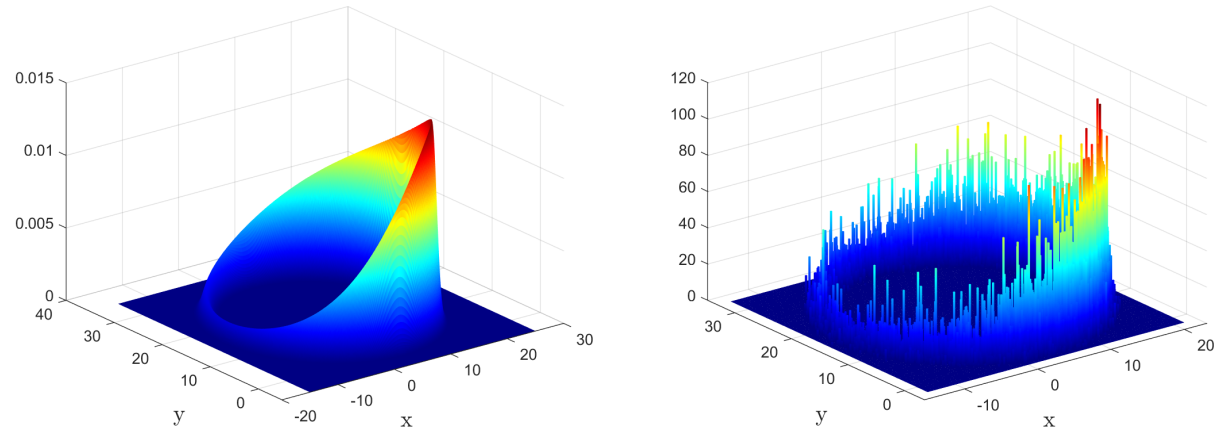
6. Adaptive Metropolis algorithms

6.1. Adaptive method for the scaling value choice

As previously discussed in the preceding sections, it is imperative to select an appropriate scaling coefficient to avoid excessively large distance between consecutive samples while simultaneously preserving the convergence of the method.
To mitigate the need for a trial-and-error approach in determining the scaling coefficient, an adaptive method, as proposed by Gelman et al. [2], can be employed:

1. Start with an initial estimate of the scaling factor at $\frac{2.4^2}{d}$, where d represents the dimensionality of the problem.
2. After a certain number of iterations, adjust the Metropolis jumping rule by incrementing or decrementing the scaling factor if the acceptance rate is excessively high or low, respectively. The objective is to converge the acceptance rate to the optimal value of 0.23, especially when multiple parameters are updated simultaneously.

After implementing the algorithm, a simulation was conducted using the adaptive method. As desired, an acceptance rate of 0.23 was achieved, and the results are visible in the following figures.
As can be observed from Fig. 11, the histogram is significantly more accurate compared to the one obtained in Fig. 6. This is because a larger number of points are accepted, resulting in a denser representation that effectively captures all the features of the Bayesian posterior PDF.
In this case as well, there is a notable convergence of the mean as the samples increase (applying the *burn-in*). Therefore, the reduction of the scaling has not compromised the convergence.



(a) Surface plot of the posterior PDF

(b) 3D histogram using the adaptive scaling - iterations: 1e5

Figure 9: Comparison of sample distribution with adaptive scaling factor and true posterior PDF

The most noticeable difference is observed when analyzing the chain's trajectory on the contour of the Bayesian posterior PDF. By using an adaptive method for scaling selection, a chain is obtained that correctly explores the entire function without jumping from side to side. This confirms an optimal acceptance rate value of 0.23, which is achieved using this method.

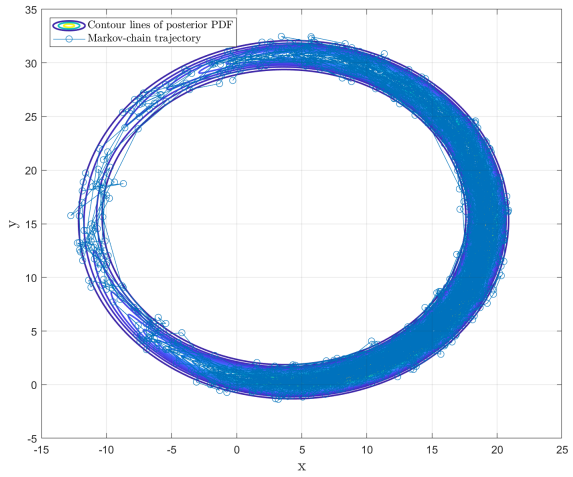


Figure 10: Trajectory of the chain on top of a contour plot of the Bayesian posterior probability density function with adaptive scaling factor - Number of samples: 1e4, for visualization reasons

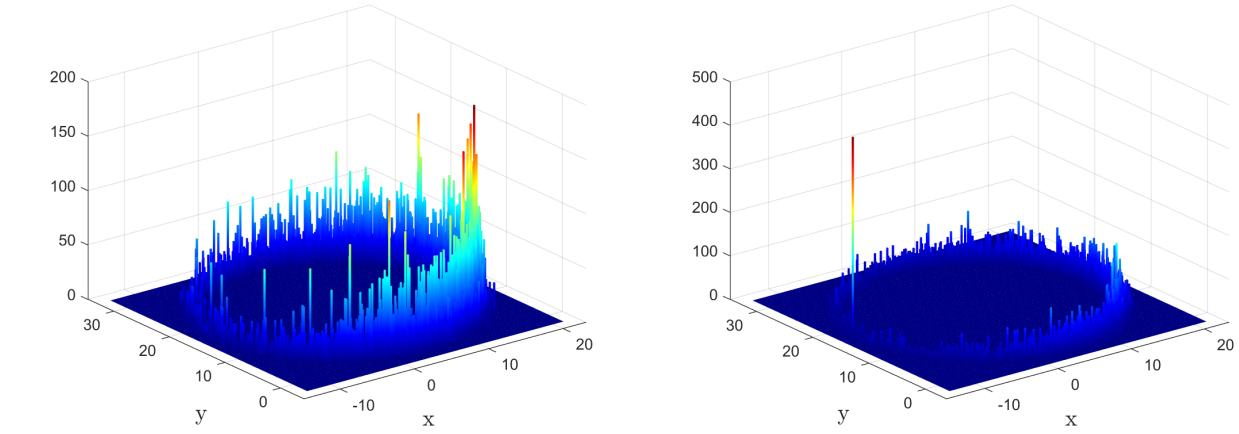
For the implementation of the method, it is possible to refer to the MATLAB code enclosed. One precaution taken was to impose an upper limit on the scaling value at 0.5 to prevent uncontrolled growth of the scaling factor.

6.2. Adaptive method for the covariance matrix

Another consideration that can be made is related to the value of the covariance matrix $[\Sigma]$. In the previous formulations, the values of σ_{xx}^2 and σ_{yy}^2, calculated by numerical integration of the posterior, are considered known. In real cases, however, the Bayesian posterior PDF is not easily integrated. Adaptive methods have also been developed to solve this problem. In this project, the adaptive method proposed by Smith [5] is presented. During a non-adaptive phase lasting for a duration denoted by k_0, we calculate a sequence of chain values: $q^0, q^1, q^2, \ldots, q^{k-1}$, utilizing the initial covariance matrix denoted as $[\Sigma_0]$. Once the adaptation process begins, the updated covariance matrix for the chain at the k-th step is determined as follows:

$$[\Sigma_k] = \frac{2.4^2}{d}\text{cov}(q^0, q^1, q^2, \ldots, q^{k-1}) + \varepsilon \mathbb{I}_p$$

The choice of the value for k_0 is made to strike a balance between achieving effective mixing and ensuring an initial stage of the chain progression with sufficient point diversity to prevent the covariance matrix from becoming singular. The inclusion of the term $\varepsilon\mathbb{I}_p$ guarantees that $[\Sigma_k]$ remains positive definite, although often ε can be set to zero.
As an exercise, we assume minimal prior knowledge of the posterior, treating it as a black-box function. In light of this, we have chosen a guess for the covariance matrix, denoted as $\Sigma_0 = \mathbb{I}$, this means that the jumping distribution is a perfectly symmetric multivariate gaussian distribution. The value of k_0 is chosen as $k_0 = 100$, as recommended in the text by Smith [5], which corresponds to 0.1% of the total iterations. With these choices, the chain exhibits excellent performance, comparable to that achieved using the correct values of σ_{ii}^2. The results are not presented to avoid redundancy.
To test the algorithm, one can attempt to push these values to their limits. For instance, using a value of $k_0 = 1000$, which corresponds to 1% of the total iterations, and a matrix $[\Sigma_0] = \text{diag}(\sigma_{yy}^2, \sigma_{xx}^2)$, representing a Gaussian distribution rotated by 90° relative to the correct distribution. In this extreme case, the resulting chain poorly describes the behavior of the posterior PDF. With this particularly challenging choice, we require more iterations to mitigate the effect of the unfavorable initial guess. As evident from the figure, there is a completely erroneous peak due to the fact that, since the jumping distribution is far from the ideal one, the algorithm struggles to generate a subsequent accepted sample. By reducing the confidence interval to $k_0 = 100$, returning to the initial value of 0.1%, even with an evidently incorrect guess for the matrix, the algorithm manages to adapt.



(a) 3D histogram of the MCMC - $k_0 = 100$

(b) 3D histogram of the MCMC - $k_0 = 1000$

Figure 11: Comparison between MCMC samples with two different trust intervals

7. Conclusions

In this report, an experimental problem affected by uncertainty was analysed using Bayesian inference. As analysed, the convergence and regularity of the Markov-chains is a fundamental requirement to allow the samples to correctly explore the Bayesian posterior and requires an in-depth study of the proposal distribution in order to guarantee these properties. The implementation of the random-walk Metropolis-Hasting method in the context of Monte Carlo Markov-chains was analysed, evaluating the strong advantages of adaptive methods that iteratively modify the proposal distribution in order to improve convergence and correct exploration of the posterior probability. As a final analysis, a more real-world case was simulated in which the posterior is not easily tractable and thus the values of its statistical moments cannot be calculated. This analysis made it possible to study the impact of the choices underlying the iterative method on the final MCMC.

References

[1] A. Gelman and D. B. Rubin. *Inference from Iterative Simulations Using Multiple Sequences*. Statistical Science, 1992, Vol. 7 No. 4.

[2] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian data analysis*. Chapman & Hall/CRC, 2014.

[3] A. Maarten. *Bayesian inference - Notes from the course: Uncertainty quantification and stochastic modeling*. University of Liège - Accademic year: 2023/24.

[4] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[5] R. C. Smith. *Uncertinty Quantification - Theory, Implementation and Applications*. SIAM - Philadelphia.