

Bayesian inference for the identification of model parameters in atmospheric entry problems

Master's thesis submitted in partial fulfillment of the requirements
for the degree of Master in Engineering Physics,

ANTHONY BOSCO

Advisor:
Prof. Maarten Arnst

Jury members:
Prof. Christophe Geuzaine
Prof. Vincent Terrapon
Ir. Joffrey Coheur

ACKNOWLEDGEMENTS

First, I'd like to thank my advisor, Prof. Maarten Arnst, as well as Joffrey Coheur, for all the help they provided me, answering the many questions I had, and their countless suggestions to improve the state of this work. In extenso, I would also like to thank the other members of the COSTMO group, Kevin and Kim, for accepting to share their office with Armin and I during the second quadrimester!

Then, I would also like to thank Prof. Christophe Geuzaine and Prof. Vincent Terrapon for accepting to be a part of the jury evaluating this work, and I hope they'll find it interesting.

I'd like to thank the many friends I've made during these few years in the Faculty of Applied Sciences. I really enjoyed all these years, whether it is the parties or the group work. I wish you all the best in your futur careers!

Last but not least, I'd like to say thank you to my parents, for giving me the opportunity to carry out such interesting studies and always supporting me during these five years.

Anthony Bosco

CONTENTS

1	Introduction	1
1.1	Motivations	1
1.1.1	Atmospheric entry flows	1
1.1.2	Thermal protection system	2
1.2	Objective of the thesis	2
1.3	Outline	3
2	Litterature review and theoretical background	4
2.1	Introduction	4
2.2	Modelling the pyrolysis process	5
2.3	Parameter estimation	6
2.3.1	Notions of inverse problems	6
2.3.2	Bayesian inference for parameter estimation	7
2.3.3	Parameter identification applied to the pyrolysis problem	9
3	Numerical solution and sensitivity analysis of the pyrolysis model	11
3.1	Numerical solution of the kinetic equations	11
3.1.1	Numerical solution of a dynamical system	11
3.1.2	Application to the pyrolysis problem	12
3.2	Exact solution to a simplified problem	12
3.3	Sensitivity analysis	15

3.3.1	Motivation	15
3.3.2	Sensitivity analysis	17
3.4	Extension to multiple reactions	19
3.5	Linearised gas production	21
4	Adjoint method for gradient computations	23
4.1	Motivation for the adjoint method	23
4.2	Adjoint method for the computation of the gradient	24
4.2.1	Adjoint method for a non-linear system of equations	24
4.2.2	Discrete adjoint method for a non-linear system of dynamical equations	25
4.2.3	Application to a least-square cost function	26
4.2.4	Illustration on the pyrolysis problem	27
5	Numerical methods for Bayesian inference	32
5.1	Linearised inverse problem	32
5.2	Hamiltonian Monte Carlo	33
5.2.1	Hamiltonian dynamics	34
5.2.2	Time integration of Hamilton's equations	35
5.2.3	Motivation for the Hamiltonian Monte Carlo method	37
5.2.4	Choice for the matrix $[M]$	40
5.2.5	Hamiltonian Monte Carlo algorithm	40
5.2.6	Application to the simplified pyrolysis problem	42
5.3	Markov Chain Monte Carlo based on Itô stochastic differential equations	49
5.3.1	Introduction	49
5.3.2	Numerical solution of the Stochastic differential equation	50
5.3.3	Application to a quadratic potential	51
5.3.4	Application to the simplified pyrolysis problem	56
5.3.5	Comparison with the Hamiltonian Monte Carlo method	59
5.4	Influence of the measurement noises	60
5.5	Extension to multiple reactions	65
6	Conclusion and future works	71

CONTENTS

6.1 Conclusions	71
6.2 Future work	72

LIST OF FIGURES

1.1	Artistic rendition of Dragon entering the Earth’s atmosphere – Courtesy of SpaceX.	1
2.1	Scatter plot of a Markov chain generated by an MCMC algorithm for A and E .	9
3.1	Comparison between the exact solution and the numerical solution (with $\Delta t = 0.1$ s) for the gas production with a constant heating rate $\beta = 2$ [K/S], $E = 4000000$ [J/mol], $A = 100000$ [1/s] and $m = 5$.	14
3.2	Convergence of the implicit Euler solver upon refining the time-step.	14
3.3	Effect of the variation of A on the gas production, with $E = 100000$ [J/mol], $m = 5$.	15
3.4	Effect of the variation of E on the gas production, with $A = 4000000$ [1/s], $m = 5$.	16
3.5	Effect of the variation of m on the gas production, with $A = 4000000$ [1/s], $E = 100000$ [J/mol].	16
3.6	Sensitivity of the gas production with respect to A , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.	18
3.7	Sensitivity of the gas production with respect to E , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.	19
3.8	Sensitivity with respect to m , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.	19
3.9	Gas production associated with the first and second reaction.	20
3.10	Linear combination of the gas production associated with each reactions.	21
4.1	Synthetic data generated using the forward problem with the parameters from Table 4.1.	27

4.2	Adjoint solution λ as a function of time (with $\Delta t = 0.1$ s), where the markers are points at which the cost function is defined (with $\Delta t_r = 10$ s).	28
4.3	Convergence of the gradient of the cost function as the time step is refined.	29
4.4	Comparison of the CPU time for the computation of the gradient with three different methods for a growing number of reactions.	30
4.5	Comparison of the CPU time for the computation of the gradient with the forward sensitivity and adjoint methods for a growing number of reactions.	30
5.1	Example of the phase space representation of a trajectory generated by Hamiltonian dynamics.	35
5.2	Comparison between the exact solution (dashed line) and numerical solution (solid line) for the explicit Euler method and Störmer-Verlet method.	37
5.3	Principle of the Hamiltonian Monte Carlo algorithm illustrated in the phase space. Adapted from [8].	41
5.4	Trajectory of the parameters generated by simulating the Hamiltonian dynamics for the simplified pyrolysis problem, with $\Delta t = 0.1$, $\tau = 25$, when $[M] = [\hat{\Sigma}_q]^{-1}$	43
5.5	Trajectory of the parameters generated by simulating the Hamiltonian dynamics for the simplified pyrolysis problem, with $\Delta t = 0.1$, $\tau = 25$, when the matrix $[M'] = \text{diag}([M])$	44
5.6	Markov chain generated by the Hamiltonian Monte Carlo algorithm. The blue circles represents samples drawn when $[M] = [\hat{\Sigma}_q]^{-1}$, while the red dots are samples generated by the diagonal approximation.	45
5.7	First 250 samples generated by the Hamiltonian Monte Carlo method for the two choices of the matrix $[M]$, compared to the contour of the linearised problem's bivariate posterior density for the parameters A and m	46
5.8	Scatter plot of the Markov chains in the original variables \mathbf{q}	46
5.9	Scatter plot of the Markov chains in the transformed variables \mathbf{Q} when the matrix $[M]$ is an estimation of the covariance matrix of the posterior.	47
5.10	Scatter plot of the Markov chains in the transformed variables \mathbf{Q} when the matrix $[M]$ is a diagonal matrix.	47
5.11	Posterior mean of A and m . The blue line is obtained when the matrix $[M]$ is the inverse of the covariance matrix of the posterior density, while the red line is obtained with a diagonal approximation.	48
5.12	Convergence of the acceptance rate with the time step Δt	49
5.13	Solution of the Itô-SDE for an harmonic oscillator.	52
5.14	Solution of the Itô-SDE for an harmonic oscillator in the $p - q$ plane.	53

5.15 Kernel density estimations of the marginal pdf for q and p for an harmonic oscillator, compared to a normal density with zero mean and unit standard deviation.	53
5.16 Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 0.01$ and $\Delta t = 0.1$ using the Stormer-Verlet scheme.	54
5.17 Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 0.01$ and $\Delta t = 0.1$ - zoom on the first iterations.	54
5.18 Kernel density estimation based on the solution of the Ito SDE with $\Delta t = 0.1$ and $f_0 = 0.01$	55
5.19 Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 1000$ and $\Delta t = 0.1$	55
5.20 Kernel density estimation based on the solution of the Ito SDE with $\Delta t = 0.1$ and $f_0 = 1000$	56
5.21 Markov chain generated by the Itô-SDE method. The blue line represents the Markov chain when the matrix $[M]$ is the inverse of the variance of posterior, while the red line is the Markov chain generated with a diagonal approximation.	57
5.22 500 first steps of the trajectory generated by the Itô-SDE method for the two choices of the matrix $[M]$, compared to the contour of the linearised problem's bivariate posterior density for the parameters A and m	58
5.23 Convergence of the posterior mean of A	58
5.24 Itô-SDE method: Posterior mean of A obtained with $\tau = 4000$ and $\Delta t = 0.1$	59
5.25 Synthetic data with Gaussian noise of standard deviation $\sigma = 5 \times 10^{-4}$ used for the Bayesian inference. The parameters used to generate the data are the same as in Figure 4.1.	60
5.26 Kernel density estimate of the linearised posterior density for A	61
5.27 Kernel density estimate in the new variables of the posterior obtained with linearisation of the inverse problem.	62
5.28 Markov chain generated by the three sampling algorithms.	63
5.29 Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the Itô-SDE method.	64
5.30 Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the HMC method.	64
5.31 Propagation of the samples obtained with the HMC method to the production of one gas.	65
5.32 Propagation of the samples obtained with the Ito-SDE method to the production of one gas.	65
5.33 Synthetic data used for the production of three gases.	66

LIST OF FIGURES

5.34	Markov chain generated for the parameter $F_{2,1}$ using the Itô-SDE method.	67
5.35	Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the Itô-SDE method. The blue dots are the kinetic parameters of the first reaction, and the red dots are those of the second reaction.	68
5.36	Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the HMC method. The blue dots are the kinetic parameters of the first reaction, and the red dots are those of the second reaction.	68
5.37	Propagation of the samples obtained with the Ito-SDE method to the production of the three gases.	69
5.38	Propagation of the samples obtained with the HMC method to the production of the three gases.	70

CHAPTER 1

INTRODUCTION

1.1 Motivations

1.1.1 Atmospheric entry flows

Vehicles entering the atmosphere typically do so at high velocities. For example, probably one of the most famous space mission, Apollo 11, entered the Earth's atmosphere after returning from the Moon with a velocity greater than 10 km/s. Due to these high velocities, atmospheric entry flows operate in the hypersonic regime, resulting in a shock layer, which can be seen in Figure 1.1.

The shock layer is characterised by high temperatures inside the shock layer, reaching up to 10000 K at the edge of the shock layer. At these temperatures, the gas molecules surrounding the flow are starting to react (dissociation and ionisation). The high temperatures in this reacting shock layer results in high thermal exchange between the flow and the vehicle.



Figure 1.1: Artistic rendition of Dragon entering the Earth's atmosphere – Courtesy of SpaceX.

1.1.2 Thermal protection system

In order for the vehicles to withstand the high heat exchanges, they are equipped with a heat shield, or thermal protection system (TPS). Designing the heat shield appears to be one of most important part in the success of the atmospheric entry process. Depending on the nature of the mission, different materials and different thicknesses for the heat shield can be used. Indeed one may think about different entry velocities, or differences in the composition of the atmosphere that may also influence the heat exchange, for instance when entering Mars' atmosphere.

In this work, ablative thermal protection systems will be considered. For instance, these materials are usually needed when entry speeds are higher than 8 km/s and where high heating rate are encountered. High heating rate results in the ablation (loss of material) in the outer surface of the heat shield, which may be caused by sublimation, chemical reaction or mechanical ablation (spallation). An important phenomenon that appears in the inner region of the heat shield is pyrolysis, which refers to the thermal degradation of the material due to its exposition to elevated temperatures. The pyrolysis process releases gases, often simply called pyrolysis gases. These gases can then percolate through the porous material, and eventually reach the surface of the heat shield, interacting with the boundary layer. It is then observed that the pyrolysis gases creates a blockage effect of the heat fluxes, reducing the heating on the heat shield.

A typical ablative material used in heat shields is the phenolic-impregnated carbon ablator (PICA) developed at NASA Ames research center in the 1990s and used, for example, on the Mars Science Laboratory. It consists of a carbon preform matrix, which is impregnated with a pyrolyzing phenolic resin. An improvement of PICA was also recently developed by SpaceX and is called PICA-x. This material possess the advantage of having a low density, which is crucial when the total mass of the entry vehicle is considered. Moreover, it possesses a lower thermal conductivity than other ablative materials, making it a suitable material for thermal protection systems.

Nowadays, the design of thermal protection system can be performed using numerical methods. These numerical methods can be, for instance, based on a unified flow-material model, where the model for the flow and for the response the thermal protection system are solved using the same computational domain, where a modified version of the Navier Stokes equations are solved (*e.g.* Volume Averaged Navier Stokes) [26].

Those equations require the development of model for the chemical reactions occurring in the material at elevated temperatures. For instance, the degradation of the material, *i.e.* the mass loss, caused by pyrolysis process and the production of gases must be quantified.

1.2 Objective of the thesis

The model used for the production of pyrolysis gases, based on the work from Lachaud [20], is highly empirical and requires the determination of several model parameters, assumed to be constant with the temperature. These parameters are identified based on experiments performed on small samples of the material composing the thermal protection system, where the change of mass of the sample can be measured by thermogravimetric analysis and the

production of pyrolysis gas is measured by mass spectroscopy [7].

Parameters can be identified by deterministic optimisation. However, this approach is usually not fitted to take into account the uncertainty in the measurements. Probabilistic methods, where the parameters are considered to be random variables with a given probability density function allows the quantification of the measurement uncertainties.

The interest in the quantification of uncertainties in the parameters lies in the fact that they can be propagated to obtain the production of pyrolysis gases used in the numerical models for the thermal protection system. The different outcomes of the numerical simulations can then be used as a way to assess the reliability of the thermal protection system, reducing the need for high security margin in the conception of TPS.

The present work will focus on the parameters identification related to the pyrolysis process. As of now, probabilistic methods are already used [11], which lead to the creation of an Bayesian inference toolbox using for instance Metropolis-Hastings algorithm. The goal of this work is to extend the general framework already developed for the inference of parameters in the pyrolysis process by using more recent and robust inference methods. These methods use, for instance, information on the geometry of the parameters' probability density function using gradients and Hessian matrices, allowing a robust and efficient exploration of the parameter space.

1.3 Outline

- In the second chapter, the theoretical background behind the pyrolysis process will be considered. In particular, a general model for the chemical kinetic aspect of the pyrolysis gases production is presented. In a second part, the methodology for the estimation of parameters with Bayesian inference is explained.
- In the third and fourth chapter, a numerical solver for the time integration of the equations appearing in the pyrolysis model in chapter 2 is developed. Moreover, methods for the computation of the sensitivities of the gas production are investigated and used to obtain a linearised version of the pyrolysis model. Finally, an adjoint method for the computation of the gradient of cost functions appearing in the Bayesian inference formulation is developed, and its performance is compared to other typical methods.
- The fifth chapter is concerned with the development of robust sampling methods for Bayesian inference using the gradient of the posterior distribution. In particular, the linearised pyrolysis is used to obtain an exact solution of the inverse problem. Then, the Hamiltonian Monte Carlo algorithm is introduced, and a method based on an Itô stochastic differential equation is investigated. These algorithms are then applied to treat different test cases based on the pyrolysis model established in chapter 2.

CHAPTER 2

LITTERATURE REVIEW AND THEORETICAL BACKGROUND

In this chapter, the model for the pyrolysis phenomenon used in this work is described in detail. Furthermore, the framework for parameter estimation using this model is laid out. In particular, the use of a probabilistic method is considered.

2.1 Introduction

Pyrolysis is a process which consists in the thermal degradation of a material due to elevated temperatures. It can be observed in many different applications: from the thermal protection system of an atmospheric entry vehicle to the burning of match. [20]

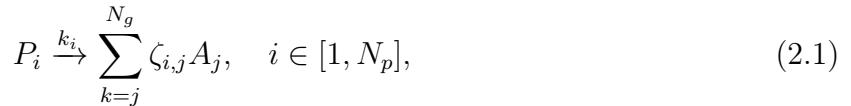
In the present work, we are interested in the former case. In these situations, pyrolysis is observed when entry vehicles reaches the atmosphere of a planet and undergo considerable heating. Pyrolyzing material such as PICA are then used as materials in thermal protection. In the aerospace application, one of the current concern is to be able to reproduce numerically in an accurate manner the physico-chemical process that a entry vehicle undergoes when entering in the atmosphere. Indeed Pyrolysis is an important part of the thermal protection system. Indeed, as it is heated, the pyrolyzing material releases gases. Those gases can percolate through the porous material and interact with the boundary layer. These interactions helps in the protection of the entry vehicle.

Numerical simulations thus require the knowledge of an accurate pyrolysis model. However, the pyrolysis is usually difficult to model in a certain way. The equations involved in the model are relatively simple, but building a model require the knowledge on the chemical reactions that happens, which are usually unknown. Experiments on the materials used as thermal protection system aim at a better characterisation of the gases that are released by the pyrolysis process, but the reaction scheme remains arbitrary.

2.2 Modelling the pyrolysis process

The pyrolysis process has been widely studied, notably in the fire engineering community. Although the production of pyrolysis gases depends on the material considered, a general model for the physico-chemical process can be developed. In the case of a phenolic resin, experimental results from [16] suggests that the pyrolysis process occurs in two main chemical reactions, for which the individual kinetics can be studied. A more general physico-chemical model able to represent an arbitrary number of chemical reactions for the pyrolysis process can then be deduced.

Following, for instance, the work of Lachaud [20], let us assume that the material can be separated into several fictitious solid phases. These solid phases can all react and produce pyrolysis gases, for instance, at different temperature levels. A general notation for the decomposition of a fictitious solid phases P_i into gaseous species A_j with stoichiometric coefficients $\zeta_{i,j}$ can be introduced as follows:



where N_g is the number of pyrolysis gas produced by the solid phase, N_p is the number of fictitious solid phases and k_i is the reaction rate.

The kinetics of the decomposition of a solid phase P_i can then be expressed using the empirical relation

$$\frac{d\xi_i}{dt} = f(\xi_i)k_i(T), \quad (2.2)$$

where the dimensionless quantity $\xi_i \in [0, 1]$ is the advancement of reaction. The function $f(\xi_i)$ is assumed to be a non-linear function of the advancement of the form

$$f(\xi_i) = (1 - \xi_i)^{m_i}, \quad (2.3)$$

where the exponent m_i is called the reaction order. The reaction rate $k_i(T)$ is assumed to be following an Arrhenius law of the form

$$k_i(T) = A_i \exp\left(-\frac{E_i}{\mathcal{R}T}\right), \quad (2.4)$$

where A_i is the Arrhenius pre-exponential factor, E_i is the activation energy and \mathcal{R} is the perfect gas constant. The pre-exponential factor has units [s^{-1}] and the activation energy is in [J/mol]. Finally, the model for the kinetics of the decomposition of a solid phase P_i is given by

$$\frac{d\xi_i}{dt} = (1 - \xi_i)^{m_i} A_i \exp\left(-\frac{E_i}{\mathcal{R}T}\right). \quad (2.5)$$

Once the kinetics of the decomposition of each solid phases are determined, the production of the gaseous specie A_j can be computed. Lachaud [20] suggested that the gas production π_j of a pyrolysis gas j is given by a linear combination of the contribution of each reaction, *i.e.*

$$\pi_j = \sum_{i=1}^{N_p} \zeta_{ij} \varepsilon_{\text{mv}} \rho_{\text{mv}} F_i \frac{d\xi_i}{dt}, \quad (2.6)$$

where $\varepsilon_{\text{mv}}\rho_{\text{mv}}$ is the initial average virgin-matrix density and F_i is the fraction of mass that is decomposed with the reaction i . Adopting the notation from [11], this model can also be written in a vector form

$$\boldsymbol{\pi} = \varepsilon_{\text{mv}}\rho_{\text{mv}}[F]\dot{\boldsymbol{\xi}}, \quad (2.7)$$

where $[F]$ is a matrix of dimensions $N_g \times N_p$.

Finally, the total production of pyrolysis gas is obtained by summing the gas production of each gaseous species,

$$\Pi = \sum_j^{N_g} \pi_j. \quad (2.8)$$

It should be noted that the pyrolysis model mentioned here assumes that the fictitious solid phases are known. In practice, the outcome of experiments only allows the quantification of the gas production to be known. The fictitious solid phases and their decomposition have to be deduced from the experimental data first. This implies that, in general, there exists several admissible reaction schemes that can reproduce similar experimental observations.

2.3 Parameter estimation

Many parameters define the pyrolysis model described previously. In order to ensure meaningful results from the model, these parameters need to be identified based on experimental data available in the literature.

2.3.1 Notions of inverse problems

Parameter estimation can be seen as a form of inverse problem, *i.e.*, considering a forward problem given by the model

$$\mathbf{y}(\boldsymbol{\theta}) = \mathbf{d}, \quad (2.9)$$

which, given the model parameters $\boldsymbol{\theta}$ predicts the outcome \mathbf{d} of the physical process of interest. Inverse problems can thus be formulated as follows: given experimental observations \mathbf{d}^{obs} , find the model parameters $\boldsymbol{\theta}$ such that Eq.(2.9) is satisfied.

However, due to the experimental nature of the observations, it should be noted that the observations are not exact, which is caused by the presence of measurement errors (or noise) $\boldsymbol{\varepsilon}$, such that, the inverse problem requires to find model parameters $\boldsymbol{\theta}$ that satisfy

$$\mathbf{y}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} = \mathbf{d}^{\text{obs}}, \quad (2.10)$$

is satisfied, where the noise $\boldsymbol{\varepsilon}$ is assumed to be additive.

The solution of these inverse problems can be obtained by either a deterministic or probabilistic approach. In the former, the solution can be cast into an optimisation problem, consisting for instance in minimizing the least-square error, *i.e.*

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{d}^{\text{obs}} - \mathbf{y}(\boldsymbol{\theta})\|^2. \quad (2.11)$$

It also appears that inverse problems may be ill-posed, such that Eq.(2.11) is unstable. Regularization methods to overcome this issue can be considered. For instance, a regularizing term can be added to the optimisation problem [9], such that the solution of the inverse problem can be written

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{d}^{\text{obs}} - \mathbf{y}(\boldsymbol{\theta})\|^2 + \alpha R(\boldsymbol{\theta}), \quad (2.12)$$

where α is a regularization parameter. This regularization term should be chosen in accordance with the noise level, noting that it should not be too small otherwise the unstable least-square method is recovered, nor should it be too large otherwise the least-square terms becomes less important and the influence of the observed data in the optimisation process becomes negligible.

2.3.2 Bayesian inference for parameter estimation

A second approach that can be considered when solving an inverse problem consists in adopting a probabilistic point of view [28]. An interesting aspect of this point of view is that it allows the interpretation of the parameters $\boldsymbol{\theta}$ as a distribution $\pi(\boldsymbol{\theta}|\mathbf{d})$, which naturally takes into account the uncertainties in the parameters resulting from the uncertainties in the experimental measurements.

The distribution $\pi(\boldsymbol{\theta}|\mathbf{d})$, called posterior distribution, is obtained by writing Bayes' theorem, *i.e.*

$$\pi(\boldsymbol{\theta}|\mathbf{d}) = \frac{\pi_0(\boldsymbol{\theta}) \pi(\mathbf{d}|\boldsymbol{\theta})}{\int_{\Theta} \pi_0(\boldsymbol{\theta}) \pi(\mathbf{d}|\boldsymbol{\theta}) d\boldsymbol{\theta}}, \quad (2.13)$$

where $\pi_0(\boldsymbol{\theta})$ is called the prior, which contains a priori beliefs on the parameters. For instance, if it is known that some parameters must be positive, a possibility is to model its prior as an indicator function that is equal to 0 when $\boldsymbol{\theta}$ is negative and 1 when it is positive. These priors are however not continuous, which can be problematic if the gradient of the posterior density is needed. When no prior information is known, uninformative priors are used or other physical intuition on the possible values. It can also be modelled as a uninformative prior, or as Gaussian distribution.

The second term, $\pi(\mathbf{d}|\boldsymbol{\theta})$ is called the likelihood and measures the probability of observing data \mathbf{d} given the parameters $\boldsymbol{\theta}$ through the model $\mathbf{y}(\boldsymbol{\theta})$. If it is assumed that the noise $\boldsymbol{\varepsilon}$ and the parameters $\boldsymbol{\theta}$ are independent random variables, and that the measurement errors $\boldsymbol{\varepsilon}$ are a Gaussian noises with covariance matrix $[\Sigma]$, the likelihood may be written as

$$\pi(\mathbf{d}|\boldsymbol{\theta}) = k \exp \left(-\frac{1}{2} (\mathbf{d} - \mathbf{y}(\boldsymbol{\theta}))^T [\Sigma]^{-1} (\mathbf{d} - \mathbf{y}(\boldsymbol{\theta})) \right), \quad (2.14)$$

where k is a normalization constant.

An equivalence between the probabilistic and deterministic approach can be made. Indeed, considering the maximum likelihood estimator

$$\boldsymbol{\theta}^{\text{ML}} = \arg \max \pi(\boldsymbol{\theta}|\mathbf{d}) \quad (2.15)$$

$$= \arg \min \frac{1}{2} (\mathbf{d} - \mathbf{y}(\boldsymbol{\theta}))^T [\Sigma]^{-1} (\mathbf{d} - \mathbf{y}(\boldsymbol{\theta})), \quad (2.16)$$

the least-square minimization problem is recovered. Similarly, the maximum a posteriori estimator can be interpreted as the regularized optimisation problem in Eq.(2.12).

Solving the inverse problem consists in characterizing the posterior density to extract statistical estimators and quantify the uncertainty. Monte Carlo methods are thus required to obtain estimations of statistical descriptors of the posterior distribution defined, for a function $g(\boldsymbol{\theta})$ that is square-integrable with respect to the posterior density, as

$$\bar{y} = \int_{\Theta} g(\boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{d}) d\boldsymbol{\theta}. \quad (2.17)$$

In a few words, Monte Carlo methods [4] works by drawing ν independent and identically distributed samples $\{\boldsymbol{\theta}^{(\ell)}, 1 \leq \ell \leq \nu\}$. These independent samples are then used to approximate the integral, provided that the integral exists, by computing

$$\bar{y}^{\nu} = \frac{1}{\nu} \sum_{\ell=1}^{\nu} g(\boldsymbol{\theta}^{\ell}). \quad (2.18)$$

By the law of large number, the estimator \bar{y}^{ν} converges almost surely to \bar{y} as $\nu \rightarrow +\infty$.

However, as the number of parameters $\boldsymbol{\theta}$ grows, the Monte Carlo method cannot generally be used. Indeed the method used to draw samples are usually not adapted for sampling from a high dimensional distribution, unless the posterior distribution is Gaussian. This is the case if the relation model-observable in Eq.(2.9) is linear.

In practice, MCMC methods generate a Markov Chain that possesses the posterior distribution of interest as a limiting distribution. It should be noted that the elements of the sequence $\{\boldsymbol{\theta}^{(\ell)}, 1 \leq \ell \leq \nu\}$ thus generated are in general not independent, since they follow the Markov property, *i.e.* the next randomly generated sample depends on the current sample, but not on the previous ones. However, it can be shown that for an ergodic Markov chain, a law of large numbers holds, and the estimation of statistical descriptor in Eq.(2.18) remains valid.

One of the most popular algorithm for generating Markov Chains that samples from a distribution is the Metropolis-Hastings algorithm, which can be summarized as follows: starting from $\boldsymbol{\theta}^{(\ell)}$, a candidate $\boldsymbol{\theta}^*$ is generated from a proposal distribution $P(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(\ell)})$. The proposal distribution is assumed to be easy to sample from, while remaining close to the posterior distribution. Then, the proposal is either accepted or rejected based on a simple criterion: if

$$\xi \sim \mathcal{U}(0, 1) \leq \alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(\ell)}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^* | \mathbf{d}) P(\boldsymbol{\theta}^* | \boldsymbol{\theta}^{(\ell)})}{\pi(\boldsymbol{\theta}^{(\ell)} | \mathbf{d}) P(\boldsymbol{\theta}^{(\ell)} | \boldsymbol{\theta}^*)} \right\},$$

then the proposal is accepted and $\boldsymbol{\theta}^{(\ell+1)} = \boldsymbol{\theta}^*$, else it is rejected and $\boldsymbol{\theta}^{(\ell+1)} = \boldsymbol{\theta}^{(\ell)}$.

If the proposal distribution is chosen to be symmetric, which is, for example, the case for a Gaussian probability density function, the condition simplifies into

$$\alpha(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(\ell)}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^* | \mathbf{d})}{\pi(\boldsymbol{\theta}^{(\ell)} | \mathbf{d})} \right\},$$

where it can be seen that the proposal $\boldsymbol{\theta}^*$ is always accepted if $\pi(\boldsymbol{\theta}^* | \mathbf{d}) \geq \pi(\boldsymbol{\theta}^{(\ell)} | \mathbf{d})$.

The Metropolis-Hastings algorithm can require manual tuning of the proposal distribution. For example, in the case of the Gaussian distribution, the covariance matrix must be chosen such that it allows the Markov chain to efficiently explore the posterior distribution. Moreover, it should also be tuned so that the acceptation rate of the samples is high enough for the algorithm to be efficient in terms of computational cost.

2.3.3 Parameter identification applied to the pyrolysis problem

Recent experiments for the pyrolysis of phenolic impregnated carbon ablators used in thermal protection systems have been carried out by Bessire and Minton [7] in 2017, with the aim to improve the understanding of the decomposition mechanisms in the material. To do so, samples of PICA were heated at a constant heating rate and the experiment was reproduced for different heating rates. The evolution of the sample's mass was obtained by thermogravimetric analysis, and the pyrolysis gases production were measured by mass spectroscopy. The authors were able to observe the production of fourteen pyrolysis gases.

Based on those experiments, Coheur [11] identified a reaction scheme consisting of the decomposition of five fictive solid phase. The parameters appearing in this reaction scheme can then be identified with the Bayesian inference methodology. Current works in the quantification of uncertainty are based on the use of the Metropolis-Hastings algorithm and some of its variants, where the covariance matrix of the proposal can be adapted based on information on the previous samples.

When applied to real data where the number of parameters can be important, these algorithms results in low acceptance rate. In particular, when dimension of the parameter space increases, it becomes harder to generate suitable proposals. Moreover, it appears that some parameters are highly correlated, which makes the tuning of the covariance matrix harder. For instance, the parameters A and E in Eq.(2.5) are highly correlated. (Figure 2.1).

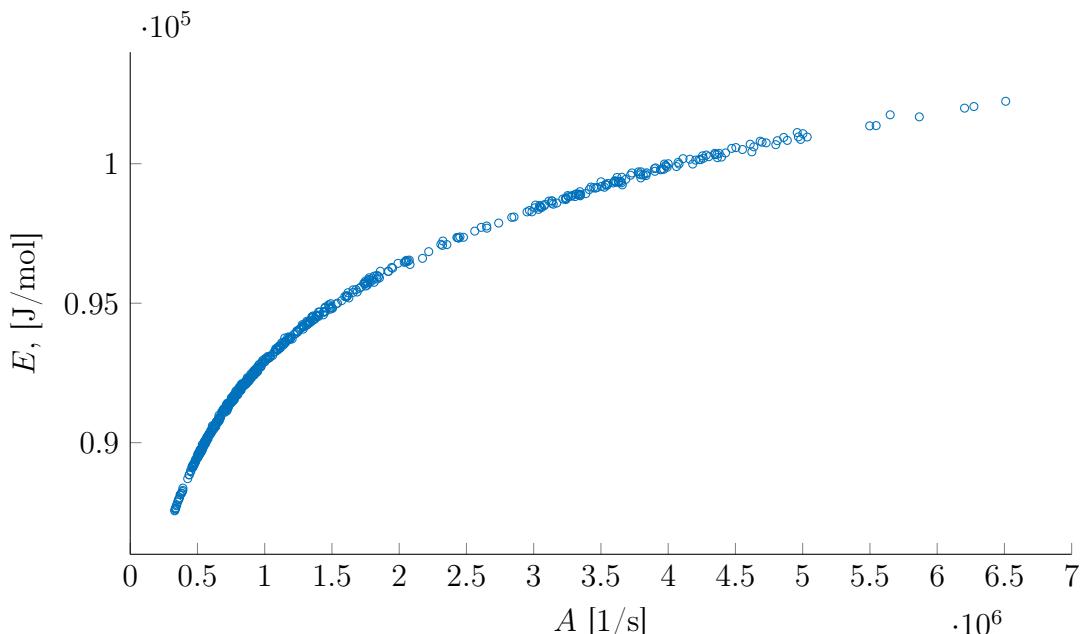


Figure 2.1: Scatter plot of a Markov chain generated by an MCMC algorithm for A and E .

For instance, when parameters are correlated, their joint probability density function is characterised by the narrow shape of Figure 2.1, which means that if the proposal distribution generates a new value θ^* that is located further away from the current value $\theta^{(\ell)}$, it is highly probable that the proposal will be rejected, whereas if the proposal distribution generates new values that are close to the current value, it results in a Markov chain that does not efficiently explore the posterior distribution. This is in this type of situation where algorithm using gradient information on the posterior density can significantly improve the results.

The goal of this present work is thus to develop methods for sampling from a posterior distribution that are more efficient and robust than the Metropolis-Hastings algorithm. Methods that requires less tuning are interesting, in particular, in the case of pyrolysis, when the exact model is not known and can vary given the observations.

In particular, methods based on information about the geometry of the posterior distribution trough the gradient, such as the Hamiltonian Monte Carlo method, are promising, since they tend to perform well in situation with correlated variables and higher dimensions of the posterior.

CHAPTER 3

NUMERICAL SOLUTION AND SENSITIVITY ANALYSIS OF THE PYROLYSIS MODEL

In this chapter, a numerical solver based on the implicit Euler method for the system of non-linear ordinary differential equations characterising the advancement of the pyrolysis reactions is developed. Furthermore, the effect of the different parameters appearing in those equations are assessed by the means discrete sensitivity analysis based on the numerical solver used, and this method is used to derive a linearised version of the pyrolysis model.

3.1 Numerical solution of the kinetic equations

3.1.1 Numerical solution of a dynamical system

Let us consider a general system of dynamical equations of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t, \boldsymbol{\theta}), \quad (3.1)$$

with initial condition $\mathbf{x}(0) = \mathbf{x}_0$. In Eq.(3.1), $\boldsymbol{\theta}$ represents the different parameters defining the dynamics of the problem. For instance, in the case of the advancement equations for the pyrolysis process in Eq.(2.5), it represents kinetic parameters $\{A_i, E_i, m_i\}, i = 1, \dots, N_R$.

The dynamical system of equations can then be discretised using an implicit single step time integration method, resulting in a non-linear algebraic system of equations of the form

$$\mathbf{F}^{k+1}(\mathbf{x}^{k+1}, \mathbf{x}^k, t^{k+1}, \boldsymbol{\theta}) = \mathbf{0}, \quad (3.2)$$

where \mathbf{F}^{k+1} is called the residual. This system of equations is then solved using the Newton-Raphson algorithm, by evaluating, until a convergence criterion is met, the expression

$$\mathbf{x}_{(\ell)}^{k+1} = \mathbf{x}_{(\ell-1)}^{k+1} - \left(\frac{\partial \mathbf{F}^{k+1}}{\partial \mathbf{x}^{k+1}} \right)^{-1}_{(\ell-1)} \mathbf{F}^{k+1} \left(\mathbf{x}_{(\ell-1)}^{k+1}, \mathbf{x}^k, t^{k+1}, \boldsymbol{\theta} \right), \quad (3.3)$$

where the starting point of the algorithm can be chosen as $\mathbf{x}_{(0)}^{k+1} = \mathbf{x}^k$. In the case where an implicit Euler method is used with a time increment Δt , the discretisation of Eq.(3.1) reads

$$\mathbf{F}^{k+1}(\mathbf{x}^{k+1}, \mathbf{x}^k, t^{k+1}, \boldsymbol{\theta}) = \mathbf{x}^{k+1} - \mathbf{x}^k - \Delta t \mathbf{f}(\mathbf{x}^{k+1}, t, \boldsymbol{\theta}), \quad (3.4)$$

and the Jacobian matrix in the Newton-Raphson update is given by

$$\frac{\partial \mathbf{F}^{k+1}}{\partial \mathbf{x}^{k+1}} = [I] - \Delta t \frac{\partial \mathbf{f}}{\partial \mathbf{x}^{k+1}}(\mathbf{x}^{k+1}, t, \boldsymbol{\theta}). \quad (3.5)$$

Moreover, it can be noticed that, if the dynamical equations are not coupled, which is typically the case in the system of equations describing the kinetics of the pyrolysis reactions, the Jacobian matrix is a diagonal matrix.

3.1.2 Application to the pyrolysis problem

Let us consider the system of equations describing the kinetics of the pyrolysis reactions in Eq.(2.5). Posing $\xi_i = 1 - \eta_i$, it may be rewritten as

$$\dot{\eta}_i = -\eta_i^{m_i} A_i \exp\left(-\frac{E_i}{RT}\right), \quad i = 1, \dots, N_R, \quad (3.6)$$

with initial condition $\eta_i(0) = 1$. An implicit Euler discretisation with time-step Δt results in the system of equations

$$F_i^{k+1}(\eta_i^{k+1}, \eta_i^k, t^{k+1}, \boldsymbol{\theta}) = \eta_i^{k+1} - \eta_i^k + \Delta t (\eta_i^{k+1})^{m_i} A_i \exp\left(-\frac{E_i}{RT^{k+1}}\right) = 0, \quad (3.7)$$

and the Jacobian matrix is given by

$$\frac{\partial F_i^{k+1}}{\partial \eta_j^{k+1}} = \left[1 + \Delta t m_i (\eta_i^{k+1})^{m_i-1} A_i \exp\left(-\frac{E_i}{RT^{k+1}}\right) \right] \delta_{ij}, \quad (3.8)$$

where δ_{ij} is the Kronecker delta.

3.2 Exact solution to a simplified problem

In order to assess the precision of the aforementioned numerical method, we now consider a toy problem consisting of a sample of initial mass $m_0 = 1$ with a single solid phase S producing a gaseous specie G , *i.e.*



Furthermore, let us assume that the sample is heated at a constant heating rate β , defined as

$$\beta = \frac{dT}{dt}, \quad (3.10)$$

such that the evolution of the temperature is given by

$$T(t) = \beta t + T_0, \quad (3.11)$$

where T_0 is the initial temperature of the sample. In this situation the production of the gas G can be written as

$$\pi_G = -\dot{\eta} \quad (3.12)$$

$$= \eta^m A \exp\left(-\frac{E}{RT(t)}\right). \quad (3.13)$$

In this situation, an exact solution can be derived, which will then be used to assess the convergence of the numerical solver.

Using the chain rule, the time derivative in Eq.(3.6) can be rewritten as

$$\frac{d\eta}{dt} = \beta \frac{d\eta}{dT}, \quad (3.14)$$

and Eq.(3.6) can thus be expressed as

$$\beta \frac{d\eta}{dT} = -\eta^m A \exp\left(-\frac{E}{RT}\right). \quad (3.15)$$

Eq.(3.15) can be integrated by writing

$$\beta \int_1^\eta \frac{d\tilde{\eta}}{\tilde{\eta}^m} = -A \int_{T_0}^T \exp\left(-\frac{E}{R\tilde{T}}\right) d\tilde{T}, \quad (3.16)$$

such that the left hand side reads

$$\beta \int_1^\eta \frac{d\tilde{\eta}}{\tilde{\eta}^m} = \begin{cases} \beta \frac{\eta^{1-m} - 1}{1-m}, & m \neq 1, \\ \beta \ln(\eta), & m = 1. \end{cases} \quad (3.17)$$

Introducing the change of variables $\zeta = \frac{E}{RT}$ and integrating by parts, the right hand side becomes

$$-A \int_{T_0}^T \exp\left(-\frac{E}{R\tilde{T}} d\tilde{T}\right) = -\frac{AE}{R} \left\{ \left[\frac{\exp(-\tilde{\zeta})}{\tilde{\zeta}} \right]_{\zeta_0}^\zeta + \int_{\zeta_0}^\zeta \frac{\exp(-\tilde{\zeta})}{\tilde{\zeta}} d\tilde{\zeta} \right\} \quad (3.18)$$

$$= -\frac{AE}{R} \left\{ \left[\frac{\exp(-\tilde{\zeta})}{\tilde{\zeta}} \right]_{\zeta_0}^\zeta - E_1(\zeta) + E_1(\zeta_0) \right\}, \quad (3.19)$$

where $E_1(x)$ is the exponential integral function [1] defined as

$$E_1(x) = \int_x^\infty \frac{\exp(-x)}{x} dx. \quad (3.20)$$

Finally, the solution of Eq.(3.15) can be expressed as

$$\eta(T) = \begin{cases} \left[1 - \frac{(1-m)}{\beta} \left(AT \exp\left(-\frac{E}{RT}\right) - \frac{AE}{R} E_1\left(\frac{E}{RT}\right) + C_0 \right) \right]^{\frac{1}{1-m}}, & m \neq 1, \\ \exp\left[-\frac{1}{\beta} \left(AT \exp\left(-\frac{E}{RT}\right) - \frac{AE}{R} E_1\left(\frac{E}{RT}\right) + C_0 \right)\right], & m = 1, \end{cases} \quad (3.21)$$

where $C_0 = -AT \exp\left(-\frac{E}{RT_0}\right) + \frac{AE}{R} E_1\left(\frac{E}{RT_0}\right)$.

Figure 3.1 shows the production of gas computed with the analytical solution for $\eta(T)$, compared to the numerical solution obtained with an implicit Euler integration scheme, with $\Delta t = 0.1$ s. Good agreement between both solutions can be observed.

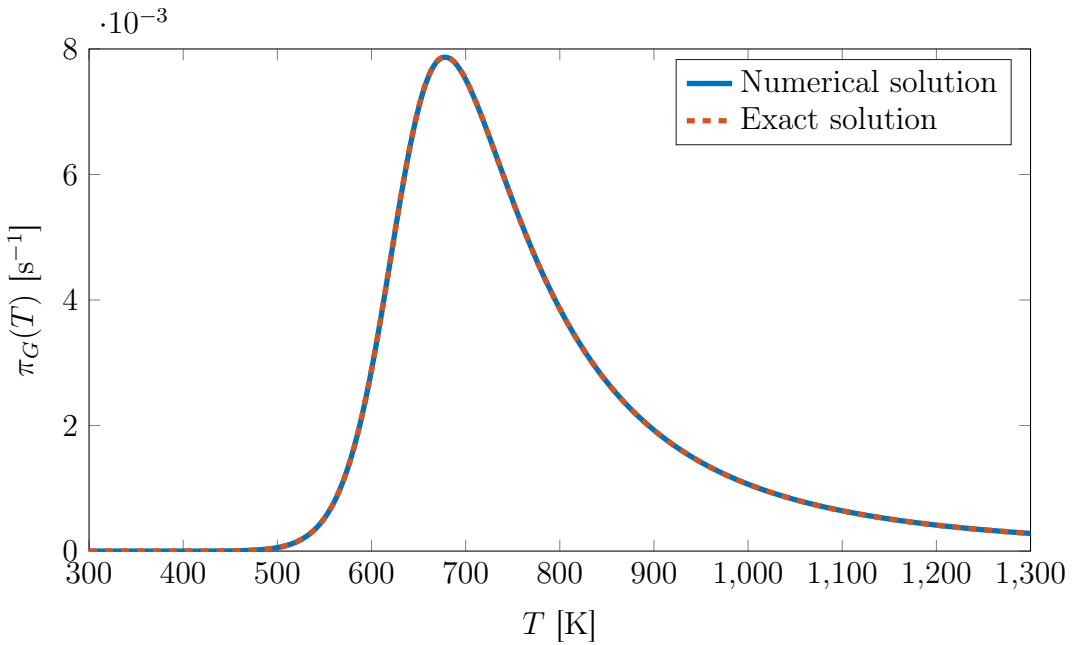


Figure 3.1: Comparison between the exact solution and the numerical solution (with $\Delta t = 0.1$ s) for the gas production with a constant heating rate $\beta = 2$ [K/S], $E = 4000000$ [J/mol], $A = 100000$ [1/s] and $m = 5$.

Furthermore, one can define the error for a given time-step Δt between the exact solution and the numerical solution as

$$\varepsilon_{\Delta t} = \max_k |\pi(t_k) - \pi^k|, \quad (3.22)$$

which should be a decreasing function of the time-step. In particular, in the case of an implicit Euler discretisation, it is expected that the graph of the error with respect to the time-step, in logarithmic axis, is a line of slope equals to one, since the implicit Euler method is of order one. Figure 3.2 shows that the numerical time integration is indeed of order one.

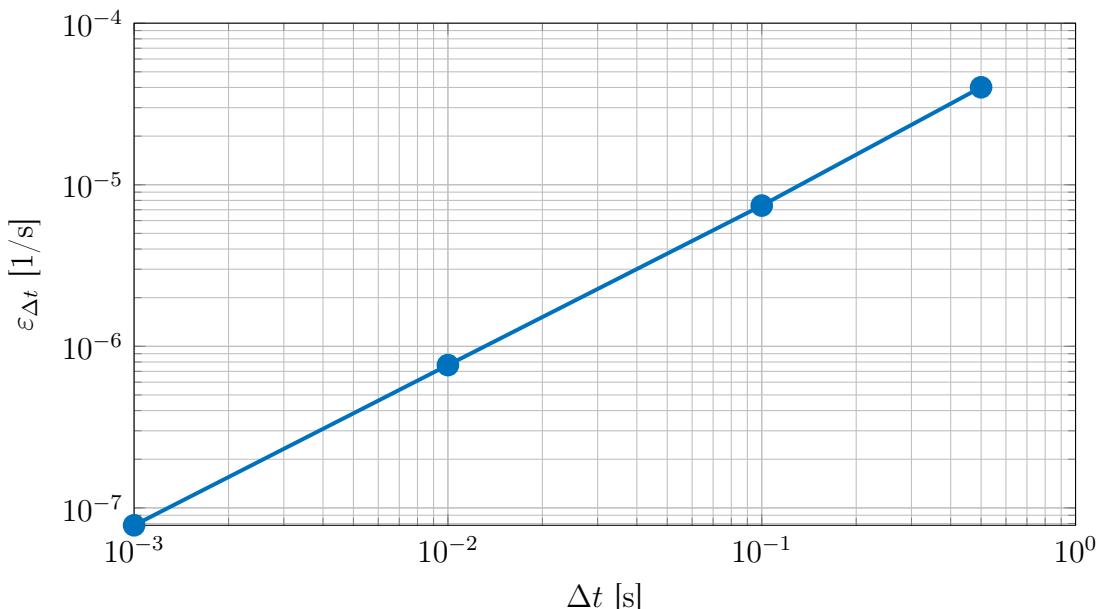


Figure 3.2: Convergence of the implicit Euler solver upon refining the time-step.

3.3 Sensitivity analysis

3.3.1 Motivation

In order to perform parameter estimation, it is important to first gain some insight on the effect of the parameters, for example, on the pyrolysis gas production. Let us again consider the constant heating rate problem solved previously for several values of each parameters and compare it to the solution obtained in the previous section.

In figure 3.3, the exponential pre-factor in the Arrhenius equation is considered. It can be seen that increasing A results in a shifting to the left and an increase in the reaction peak. Moreover, it appears to be shifting slightly despite large variation of the parameter. In the context of parameter identification, this could imply that this parameter will have larger uncertainties than the others.

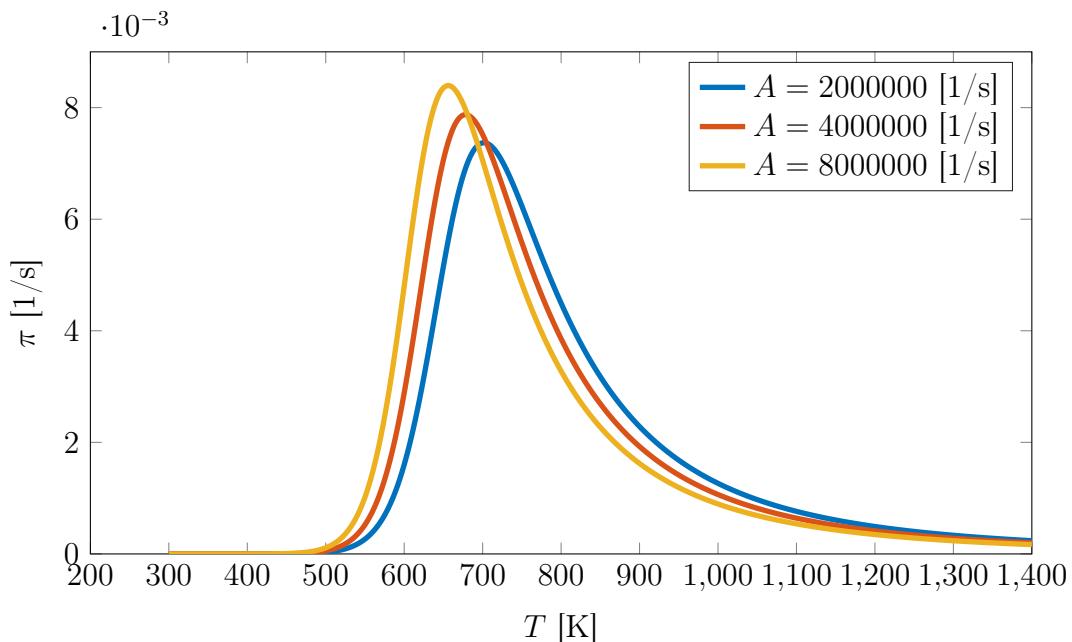


Figure 3.3: Effect of the variation of A on the gas production, with $E = 100000$ [J/mol], $m = 5$.

Figure 3.4 shows the effect of a variation of the activation energy on the gas production. Since it represents the energy required for the reaction to take place, it seems natural that it is shifted to the right when it increase, as more thermal energy will be needed for a chemical reaction to occur.

It is interesting to notice, as it was mentioned in the previous chapter, that A and E are correlated. This can be seen for instance since varying their values results in the opposite effect. Moreover, the variation of A required to produce a similar variation to that of E is high. In other words, to obtain similar gas production curves, a small variation of E must be associated with a large variation of A , which justify the exponential correlation between the parameters shown in the previous chapter.

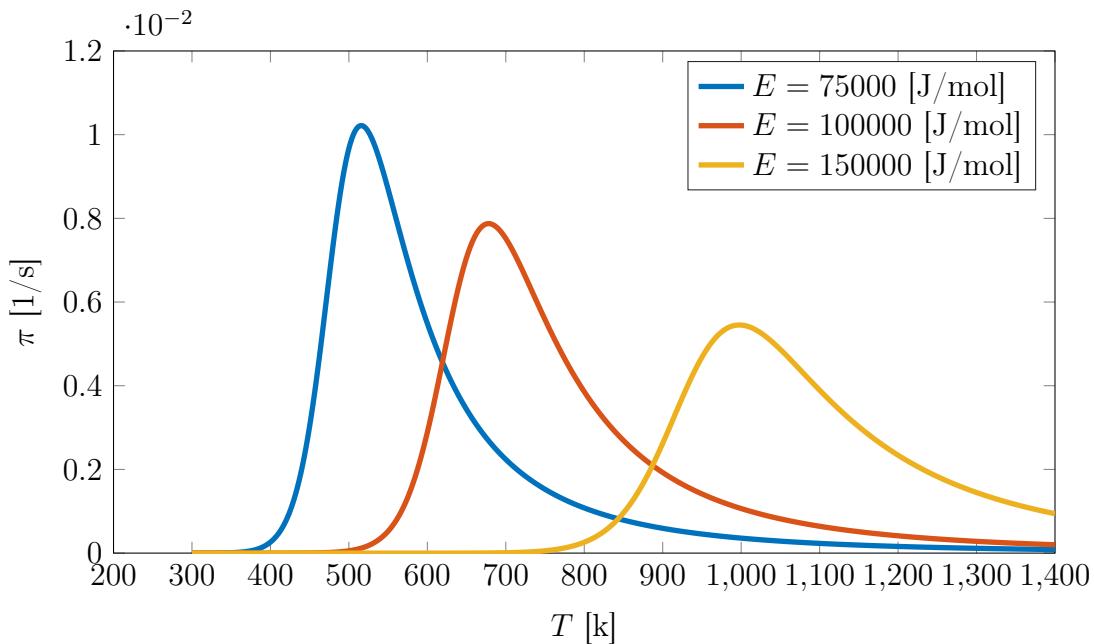


Figure 3.4: Effect of the variation of E on the gas production, with $A = 4000000$ [1/s], $m = 5$.

Finally, Figure 3.5 shows the variation of the gas production when modifying the parameter m . In this case, it can be seen that higher values of m leave the temperature at which the reaction peak occurs unchanged, but greatly impact its magnitude which suggests that, for smaller values of m , the rate of change of $\eta(t)$ at the peak is significantly greater. Moreover, since the rate of change is greater, η also reaches values close to zero faster, at which point the rate of change of η also goes to zero.

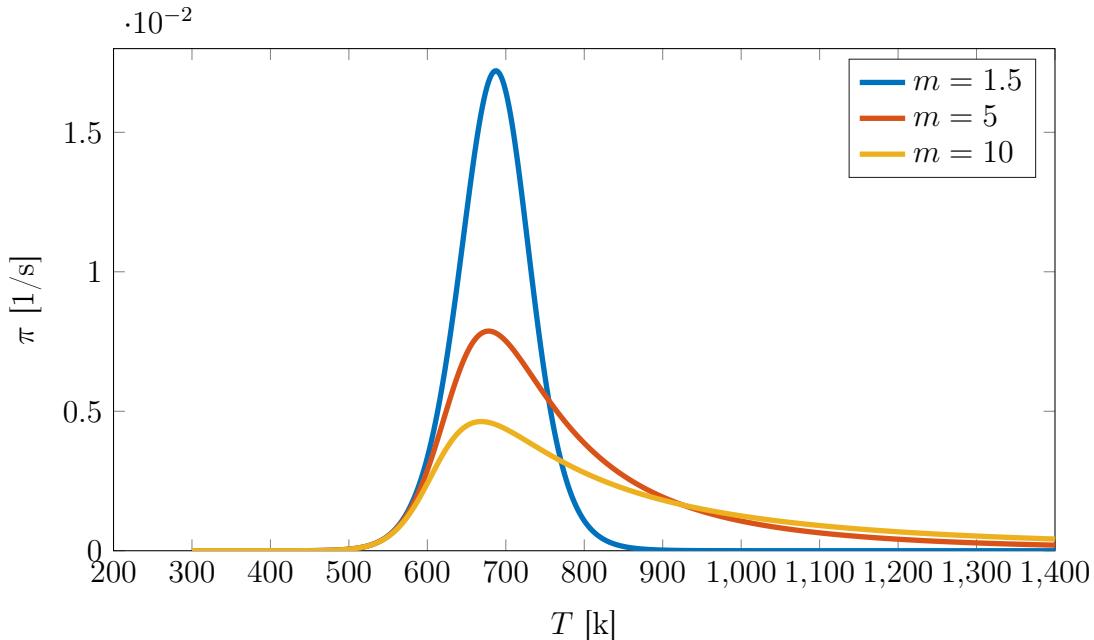


Figure 3.5: Effect of the variation of m on the gas production, with $A = 4000000$ [1/s], $E = 100000$ [J/mol].

3.3.2 Sensitivity analysis

The influence of the parameters on the gas production can also be assessed by the means of a first order sensitivity analysis, which consists in analysing the effect of an infinitesimal change in model parameters $\boldsymbol{\theta}$ on the output of the system $\mathbf{x}(t, \boldsymbol{\theta})$. The sensitivity matrix $[S]$ is thus defined as

$$S_{ij}(t, \boldsymbol{\theta}) = \frac{\partial x_i}{\partial \theta_j}, \quad (3.23)$$

where $[S]$ is a matrix of dimensions $(N_x \times N_\theta)$, N_x is the number of dynamical equations and N_θ is the number of parameters.

A first numerical approach to obtain the coefficients of this matrix consists in using a forward finite differences approximation of the derivatives in Eq.(3.23), *i.e.*

$$\frac{\partial x_i}{\partial \theta_j} = \frac{x_i(t, \boldsymbol{\theta} + \varepsilon \mathbf{e}_j) - x_i(t, \boldsymbol{\theta})}{\varepsilon} + \mathcal{O}(\varepsilon), \quad (3.24)$$

which requires solving $N_\theta + 1$ a system of N_x non-linear equations.

However, when the solution of the system is obtained through numerical simulation, it should be noted that the derivative computed with finite difference contains both an error due to the numerical solution, as well as an error arising from the finite difference approximation itself. More accurate methods such a centred finite differences also require a greater amount of evaluation of the model output $\mathbf{x}(t, \boldsymbol{\theta})$, which can be an issue if an efficient computation of the sensitivity matrix is sought.

When the sensitivity analysis of a dynamical system of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t, \boldsymbol{\theta}) \quad (3.25)$$

is considered, a more efficient method for the computation of $[S]$ can be developed. Differentiating Eq.(3.25) with respect to the parameters $\boldsymbol{\theta}$, an evolution equation for the sensitivity matrix can be obtained, *i.e.*

$$\dot{S}_{ij} = \frac{\partial f_i}{\partial x_k} S_{kj} + \frac{\partial f_i}{\partial \theta_j}, \quad (3.26)$$

with $S_{ij}(0, \boldsymbol{\theta}) = 0$ since it is assumed that the initial condition \mathbf{x}_0 does not depend on the parameters. This evolution equation can then be solved using similar numerical methods as the dynamical system. Using a similar approach, the discretised dynamical system

$$\mathbf{F}^{k+1}(\mathbf{x}^{k+1}, \mathbf{x}^k, t^{k+1}, \boldsymbol{\theta}) = \mathbf{0}, \quad (3.27)$$

can directly be differentiated with respect to $\boldsymbol{\theta}$, *i.e.*

$$\frac{\partial F_i^{k+1}}{\partial x_\ell^{k+1}} S_{\ell j}^{k+1} + \frac{\partial F_i^{k+1}}{\partial x_\ell^k} S_{\ell j}^k = -\frac{\partial F_i^{k+1}}{\partial \theta_j}, \quad (3.28)$$

such that the evolution of sensitivity matrix is obtained by solving a linear, explicit system of equations for the coefficients S_{ij} , which can be solved in parallel with the system of equation for \mathbf{x} , since the Jacobian matrix has already been computed and given the fact that, if the same implicit Euler scheme is considered, one has

$$\frac{\partial F_i^{k+1}}{\partial x_\ell^k} = \delta_{i\ell}. \quad (3.29)$$

Moreover, it can be seen that both Eq.(3.26) and Eq.(3.28) produce the same results if the same time discretisation is used.

In terms of computational cost, the finite difference method requires the resolution of a system of N_x non-linear differential equations $N_\theta+1$ times ($2N_\theta+1$ for centred finite difference), while the discrete forward sensitivity method require solving the system of equations once, as well as solving $N_x \times N_\theta$ linear equations for the elements of the sensitivity matrix.

If the quantity of interest is not the sensitivity of the states \mathbf{x} , but the sensitivity of a function $\mathbf{y}(\mathbf{x}, t, \boldsymbol{\theta})$, which is, for instance, the case when studying the production of pyrolysis gases, the sensitivity matrix $[\tilde{S}]$ associated to the function \mathbf{y} can be expressed using the chain rule, *i.e.*

$$\tilde{S}_{ij} = \frac{\partial y_i}{\partial x_\ell} S_{\ell j} + \frac{\partial y_i}{\partial \theta_j}. \quad (3.30)$$

Applying this methodology to the problem in which a sample is heated at a constant heating rate β and releases a single pyrolysis gas G , the sensitivity matrix reduces to the gradient of π_G with respect to $\boldsymbol{\theta} = (A, E, m)^T$.

Figure 3.6 shows the sensitivity of the gas production for the parameter A . An interesting aspect is that it is similar in shape, but reversed, to the graph in Figure 3.7, which further confirms that the parameters have opposite effects, and hints at a possible correlation between A and E . The order of magnitude of the derivatives suggests that a small variation of the parameter A also has less influence on the gas production, compared to E and m , which further confirms what was observed previously by manually varying the parameters. In particular, Figure 3.8 shows the sensitivity of the gas production to the parameter m . It is the component with the higher magnitude, and it was indeed seen in Figure 3.5. Finally, Figures 3.6 to 3.8 also compares the sensitivity of the gas production computed with the forward sensitivity method and finite difference (where $\varepsilon = 0.01\theta_j$), and it can be seen that the two methods are in agreement.

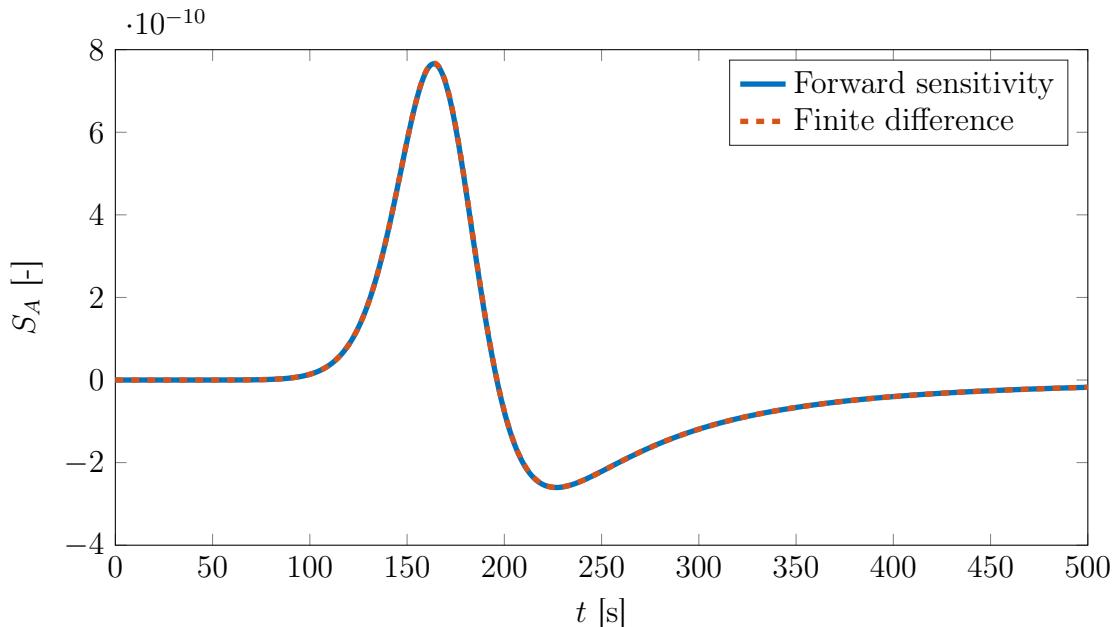


Figure 3.6: Sensitivity of the gas production with respect to A , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.

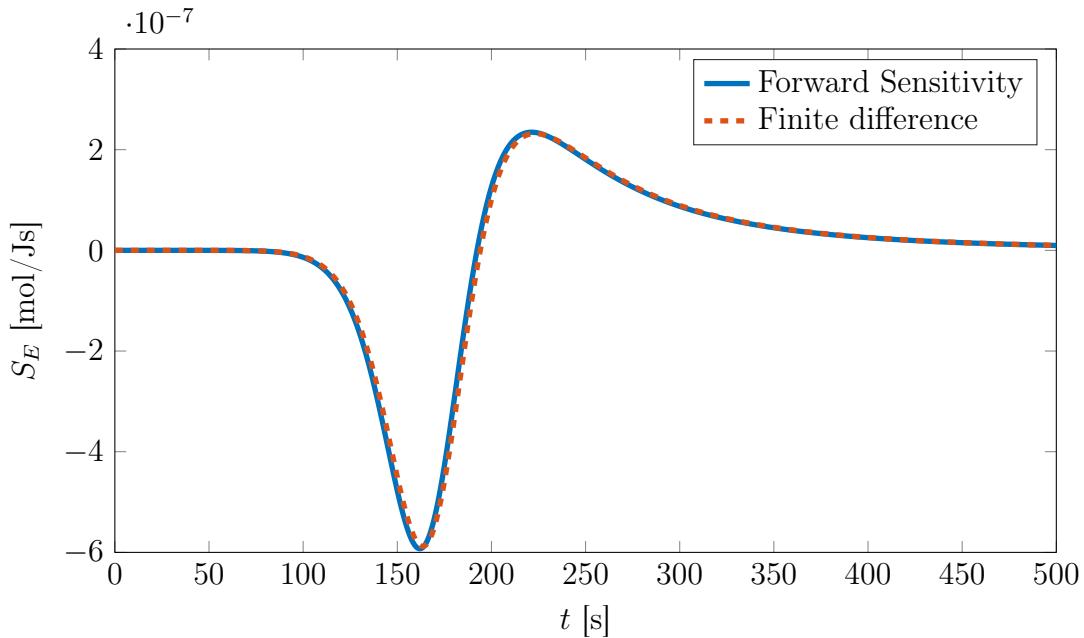


Figure 3.7: Sensitivity of the gas production with respect to E , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.

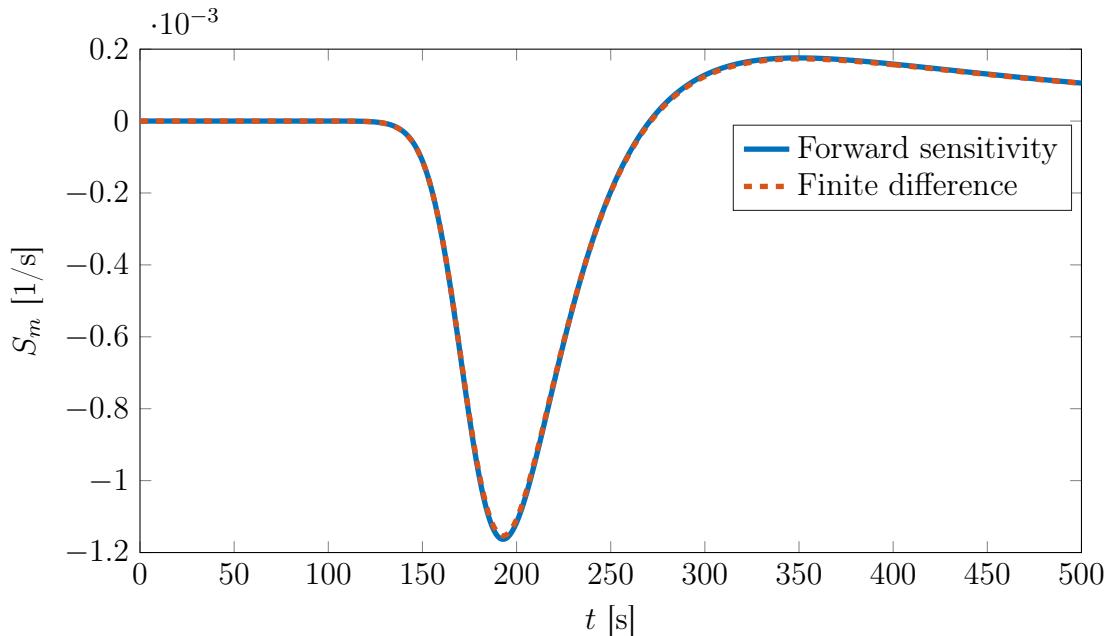


Figure 3.8: Sensitivity with respect to m , with $\Delta t = 0.1$ s. The finite difference solution is obtained with $\varepsilon = 0.01\theta_j$.

3.4 Extension to multiple reactions

Up to now, the numerical methods develop to solve the dynamical equations and obtain the gas production were only applied to the production of a single gas stemming from a unique reaction. However, the numerical solver, as well as the sensitivity analysis can readily be

extended to models where several reactions and gases are considered. Here, we will illustrate these more complete models on the production of a gas, resulting from two reactions, *i.e.*



and the gas production π_A , which now contains two reaction peaks, becomes a linear combination of the production associated with each reactions, *i.e.*

$$\pi_A = F_1 \dot{\eta}_1 + F_2 \dot{\eta}_2. \quad (3.33)$$

	A [1/s]	E [J/mol]	m [-]	F [-]
Reaction 1	4000000	100000	5	0.48
Reaction 2	2000000	75000	10	0.45

Table 3.1: Values of the parameters used in the two reaction pyrolysis model.

The two individual production curves associated to η_1 and η_2 obtained using the data from Table 3.1 are shown in Figure 3.9, and the linear combination of the two is shown in Figure 3.10, where the two reactions peak can readily be identified.

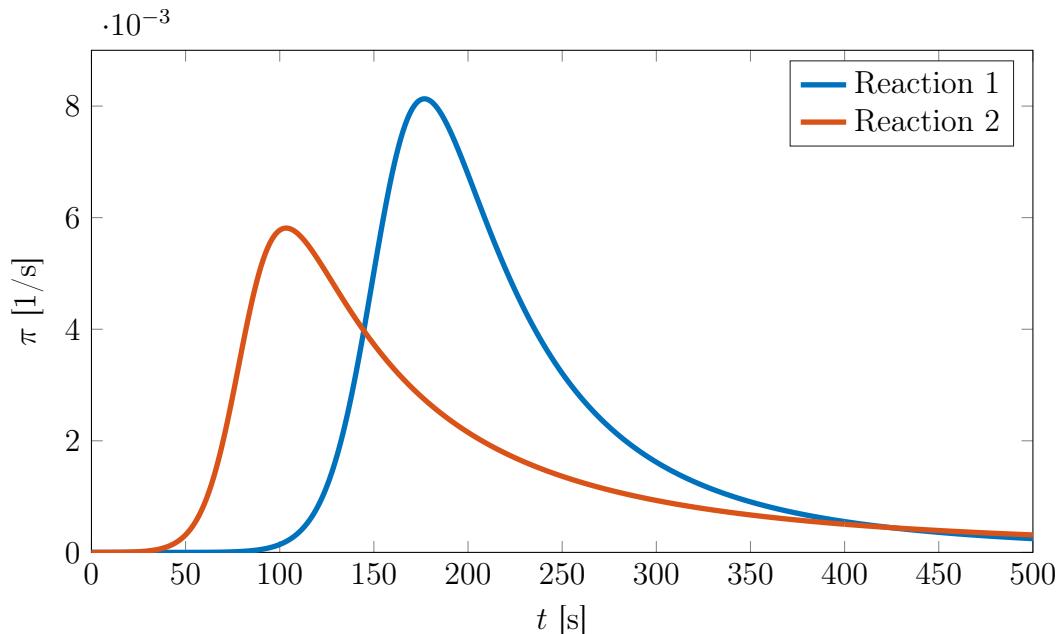


Figure 3.9: Gas production associated with the first and second reaction.

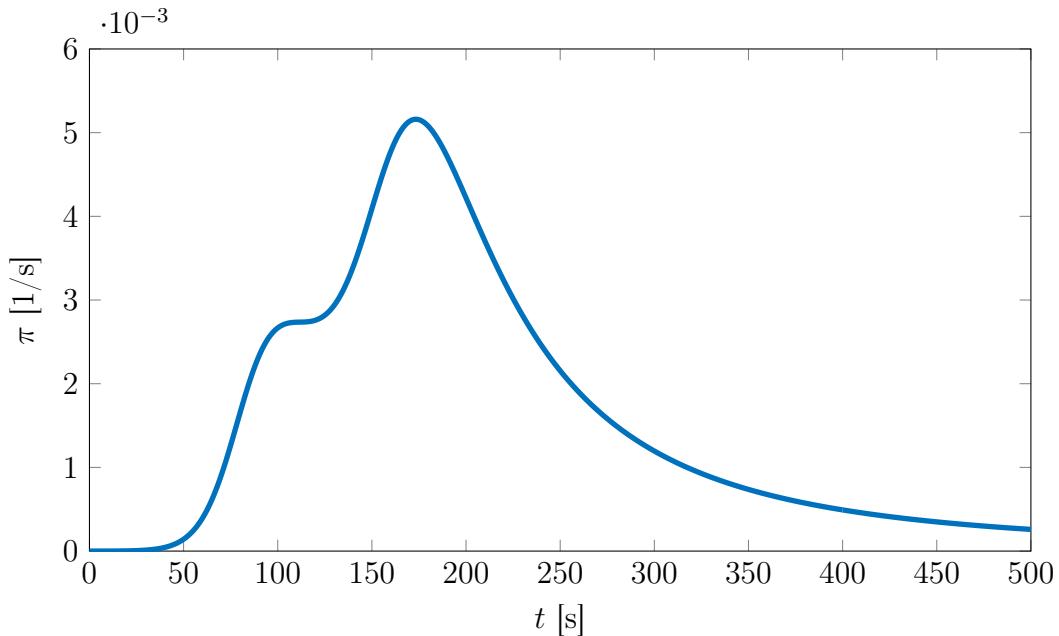


Figure 3.10: Linear combination of the gas production associated with each reactions.

Furthermore, while the development of the sensitivity analysis method remains the same, some simplifications arise. Indeed, in a general case involving multiple gases and reactions, since the mass fractions $F_{i,j}$ do not appear explicitly in the dynamical equations, the computation of the sensitivity matrix $[S_{i,j}]$ obtained by the solution of the evolution equations in Eq.(3.28) can be carried out by ignoring the the mass fractions.

The contribution of these parameters to the sensitivity matrix associated to the gas production $[\tilde{S}_{i,j}]$ in Eq.(3.30) only appears in the second term. Moreover, since the production of a gas i is obtained by a linear combination of individual gas production, the derivative of π_j with respect to $F_{k,\ell}$ can be readily obtained without further computations. Indeed, one has

$$\frac{\partial \pi_i}{\partial F_{k,\ell}} = \varepsilon_{mv} \sum_j \frac{\partial F_{i,j} \dot{\eta}_j}{\partial F_{k,\ell}} \quad (3.34)$$

$$= \varepsilon_{mv} \dot{\eta}_\ell \delta_{ij,k\ell}. \quad (3.35)$$

In particular, in the case of two reactions and one gas considered previously, one simply has

$$\frac{\partial \pi_A}{\partial F_1} = \dot{\eta}_1, \quad \frac{\partial \pi_A}{\partial F_2} = \dot{\eta}_2, \quad (3.36)$$

and the evolution of the sensitivities with respect to F_i are simply the individual gas production curves already shown in Figure 3.9.

3.5 Linearised gas production

Finally, to conclude this chapter, it should also be mentioned that the sensitivity matrix can be used to linearise the gas production model by the use of a first order Taylor expansion

around a reference values of the parameters $\boldsymbol{\theta}_0$. One has

$$\pi_i(t, \boldsymbol{\theta}) \approx \pi_i(t, \boldsymbol{\theta}_0) + [\tilde{S}_{ij}(t, \boldsymbol{\theta}_0)](\theta_j - \theta_{0,j}). \quad (3.37)$$

In the context of Bayesian inference, the reference parameters vector $\boldsymbol{\theta}_0$ can, for instance, be chosen as the maximum likelihood or maximum a posteriori estimators, provided that those estimators are available. Moreover, an advantage that arise when considering linear models is that it has been shown [28] that the solution of a linear inverse problem can be readily computed. Indeed, as it will be shown in Chapter 5, the posterior distribution of a linear inverse problem is a Gaussian distribution, for which, as mentioned in Chapter 2, samples can be readily obtained without having to rely on more advanced sampling methods.

CHAPTER 4

ADJOINT METHOD FOR GRADIENT COMPUTATIONS

In the previous chapter, a methodology for the sensitivity analysis of a model defined by a system of dynamical equations was developed. In particular, the forward sensitivity method using the time discretisation used to solve the system of equations was considered. In this chapter, a discrete adjoint method used to efficiently compute the gradient of cost functions appearing, for instance, in optimisation problems and Bayesian inference will be developed. In particular, it will be applied to a least-square type of cost function, which appears often as the log-likelihood in the Bayesian formulation of inverse problems, and it will be illustrated in the case of a simplified pyrolysis model. Finally, its performance, compared to the finite difference and forward sensitivity methods, will be assessed.

4.1 Motivation for the adjoint method

Let us consider a cost function $J(\boldsymbol{\theta})$ of the type

$$J(\boldsymbol{\theta}) = \sum_{r=0}^{N_r} j(\mathbf{x}(t^r, \boldsymbol{\theta}), \boldsymbol{\theta}, t^r), \quad (4.1)$$

where $\mathbf{x}(t, \boldsymbol{\theta})$ is given by Eq.(3.1). First, it should be noted that, if the equations for $\mathbf{x}(t, \boldsymbol{\theta})$ are solved using the numerical methods presented in the previous chapter, it is required that the time discretisation Δt is fine enough so that the numerical solution is known at $t = t^r$.

This type of cost function is a generic form of functions appearing in optimisation and Bayesian inference, when, for example, measurements are available at a discrete set of time instants. It is usually necessary to compute the gradient of this cost function with respect to the parameters $\boldsymbol{\theta}$ when considering an optimisation method, but this derivative is also required when using more advanced sampling algorithm for Bayesian inference, such as Hamiltonian Monte Carlo. For such sampling algorithms, which are often used for a large number of iterations, efficient methods for the computation of the gradient are mandatory.

Similarly to the previous chapter, several approaches can be explored. First, the N_θ deriva-

tives can be computed by forward finite differences approximation, *i.e.*

$$\frac{\partial J}{\partial \theta_j} \approx \frac{J(\boldsymbol{\theta} + \varepsilon \mathbf{e}_j) - J(\boldsymbol{\theta})}{\varepsilon} \quad (4.2)$$

although the same weaknesses in terms of precision and computational cost can readily be identified.

The sensitivity matrix computed with the forward sensitivity method can also be used. Indeed, differentiating Eq.(4.1) with respect to the parameters, one has

$$\frac{\partial J}{\partial \theta_j} = \sum_{r=0}^N \left(\frac{\partial j}{\partial x_\ell} S_{\ell j} + \frac{\partial j}{\partial \theta_j} \right). \quad (4.3)$$

While solving the equations for the elements of the matrix $[S]$ is more efficient than solving the forward problem for finite differences, it still requires a non-negligible amount of computations, in particular when the number of parameters increases.

4.2 Adjoint method for the computation of the gradient

Another method commonly used in optimisation for computing the gradient of a cost function constrained by evolution equations such as Eq.(3.1) is the adjoint method, where the gradient is computed by the means of an adjoint solution $\boldsymbol{\lambda}(t, \boldsymbol{\theta})$.

4.2.1 Adjoint method for a non-linear system of equations

A first insight on the adjoint method can be obtained by first considering a simpler problem than the non-linear dynamical equations from Eq.(2.5). A typical example is that of the computation of the gradient of a function $j(\mathbf{x}(t, \boldsymbol{\theta}), \boldsymbol{\theta}, t)$ where \mathbf{x} satisfies the constraint

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) = 0, \quad (4.4)$$

which can be, for instance, a non-linear system of equations resulting from a finite element discretisation of a partial differential equation.

The adjoint equations can be derived from the sensitivity analysis. The derivative of $j(\mathbf{x}(t, \boldsymbol{\theta}), \boldsymbol{\theta}, t)$ is

$$\frac{dj}{d\theta_j} = \frac{\partial j}{\partial x_\ell} S_{\ell j} + \frac{\partial j}{\partial \theta_j}. \quad (4.5)$$

Taking the derivative of the constraint in Eq.(4.4) with respect to $\boldsymbol{\theta}$, one has

$$\frac{df_i}{d\theta_j} = \frac{\partial f_i}{\partial x_\ell} S_{\ell j} + \frac{\partial f_i}{\partial \theta_j} = 0, \quad (4.6)$$

such that $S_{\ell j} = - \left(\frac{\partial f_i}{\partial x_\ell} \right)^{-1} \frac{\partial f_i}{\partial \theta_j}$. Introducing this relation into Eq.(4.5) results in

$$\frac{dj}{d\theta_j} = \frac{\partial j}{\partial x_\ell} \left(\frac{\partial f_i}{\partial x_\ell} \right)^{-1} \frac{\partial f_i}{\partial \theta_j} + \frac{\partial j}{\partial \theta_j}. \quad (4.7)$$

The adjoint $\boldsymbol{\lambda}$ is then introduced as the solution of a linear system of equations

$$\frac{\partial f_i}{\partial x_\ell} \lambda_i = \frac{\partial j}{\partial x_\ell} \quad (4.8)$$

and the gradient is given by

$$\frac{dj}{d\theta_j} = \lambda_i \frac{\partial f_i}{\partial \theta_j} + \frac{\partial j}{\partial \theta_j}. \quad (4.9)$$

The advantage of this formulation is that it allows us to ignore the sensitivity matrix and replace it with a linear system of equations involving derivatives that are simpler to compute. A less intuitive yet equivalent method for obtaining the adjoint equations relies on the use of the Lagrange method in optimisation, where it can be shown that the adjoint solution $\boldsymbol{\lambda}$ is, in fact, a vector containing the Lagrange multipliers.

4.2.2 Discrete adjoint method for a non-linear system of dynamical equations

The methodology used to obtain the adjoint solution presented in [9] can be extended to the general form of cost function in Eq.(4.1). First, let us form the Lagrangian

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{r=0}^{N_r} j(\mathbf{x}(t^r, \boldsymbol{\theta}), \boldsymbol{\theta}, t^r) + \sum_{k=0}^{N-1} \lambda_i^{k+1} F_i^{k+1}, \quad (4.10)$$

where the last term is null by definition. It is important to notice that the summation is not carried out on the same indices since it is not required that the cost function is defined at each time step, but it is still assumed that

$$\{t^r\}_{r=0}^{N_r} \subset \{t^k\}_{k=0}^N, \quad N_r \Delta t_r = N \Delta t,$$

which, in practice, means that the time discretisation of the dynamical system of equations must be adapted to the data. It is not mandatory to assume that the final times of the sums match, but it will simplify the final formulation of the adjoint equations, and in the context of Bayesian inference, it is not absurd to assume that the numerical solution can be stopped at the last time an experimental data is available.

The adjoint equation can then be found by looking at the first variation of this Lagrangian. One has

$$\delta \mathcal{L} = \sum_{r=0}^{N_r} \left(\frac{\partial j}{\partial x_j^r} \delta x_j^r + \frac{\partial j}{\partial \theta_j} \delta \theta_j \right) + \sum_{k=0}^{N-1} \lambda_i^{k+1} \left(\frac{\partial F_i^{k+1}}{\partial x_j^k} \delta x_j^k + \frac{F_i^{k+1}}{\partial x_j^{k+1}} \delta x_j^{k+1} + \frac{F_i^{k+1}}{\partial \theta_j} \delta \theta_j \right). \quad (4.11)$$

Noticing that

$$\sum_{k=0}^{N-1} \lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial x_i^{k+1}} = \sum_{k=1}^N \lambda_i^k \frac{\partial F_i^k}{\partial x_i^k},$$

and that $\delta x_j^0 = 0$, the variation of the Lagrangian may be rewritten as

$$\begin{aligned}\delta\mathcal{L} = & \sum_{r=1}^{N_r-1} \left(\frac{\partial j}{\partial x_j^r} + \lambda_i^{r+1} \frac{\partial F_i^{r+1}}{\partial x_j^r} + \lambda_i^r \frac{\partial F_i^r}{\partial x_j^r} \right) \delta x_j^r + \sum_{k=1, t^k \neq t^r}^{N-1} \left(\lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial x_j^k} + \lambda_i^k \frac{\partial F_i^k}{\partial x_j^k} \right) \delta x_j^k \\ & + \left(\frac{\partial j}{\partial x_j^N} + \lambda_i^N \frac{\partial F_i^N}{\partial x_j^N} \right) \delta x_j^N + \left(\sum_{r=0}^N \frac{\partial j}{\partial \theta_j} + \sum_{k=0}^{N-1} \lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial \theta_j} \right) \delta \theta_j.\end{aligned}\quad (4.12)$$

Eq.(4.12) suggests that the gradient of the cost function $J(\boldsymbol{\theta})$ can be obtained by enforcing that the three first terms are equal to zero. One obtains the following system of adjoint equations

$$\lambda_i^k \frac{\partial F_i^k}{\partial x_j^k} + \lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial x_j^k} = 0 \quad 1 \leq k \leq N-1, t^k \neq t^r, \quad (4.13)$$

$$\lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial x_j^k} + \lambda_i^k \frac{\partial F_i^k}{\partial x_j^k} + \frac{\partial j}{\partial x_j^k} = 0 \quad 1 \leq k \leq N-1, t^k = t^r, \quad (4.14)$$

with the initial condition

$$\lambda_i^N \frac{\partial F_i^N}{\partial x_j^N} + \frac{\partial j}{\partial x_j^N} = 0. \quad (4.15)$$

The gradient of the cost function is then given by the last term in Eq.(4.12), *i.e.*

$$\frac{\partial J}{\partial \theta_j} = \sum_{r=0}^{N_r} \frac{\partial j}{\partial \theta_j} + \sum_{k=0}^{N-1} \lambda_i^{k+1} \frac{\partial F_i^{k+1}}{\partial \theta_j}. \quad (4.16)$$

The system of equation for the adjoint solution $\boldsymbol{\lambda}$ is a backward linear system of equation. This implies that the complete forward system has to be solved prior to solving the adjoint system, in contrast with the evolution equations in the forward sensitivity method which were solved in parallel with the forward system. However, instead of adding $N_x \times N_\theta$ equations to the problem, the adjoint method only adds N_x equations, which is interesting in terms of computational cost, especially when the number of parameters is high.

4.2.3 Application to a least-square cost function

Let us now assume that the cost function has the form

$$J(\boldsymbol{\theta}) = \sum_{r=0}^{N_r} \frac{1}{2} (\tilde{y}_i^r - y_i^r) C_{ik} (\tilde{y}_k^r - y_k^r), \quad (4.17)$$

where $\tilde{\mathbf{y}}$ is, for example, data available from experiments and $[C]$ is a symmetric positive definite matrix. Let us apply the methodology proposed above to compute the gradient.

The derivatives of the residual F_i^{k+1} appearing in the adjoint equations are easy to compute, and, defining

$$j(y^r, \boldsymbol{\theta}) = \frac{1}{2} (\tilde{y}_i^r - y_i^r) C_{ik} (\tilde{y}_k^r - y_k^r), \quad (4.18)$$

	A [1/s]	E [J/mol]	m [-]
Synthetic data	4000000	100000	5
Evaluation of the gradient	4500005	150000	5.5

Table 4.1: Parameters used to generate synthetic data and to evaluate the gradient.

the derivative of the j with respect to x_j^r may be written as

$$\frac{\partial j}{\partial x_j^r} = \frac{1}{2} \frac{\partial [(\tilde{y}_i^r - y_i^r) C_{ik} (\tilde{y}_k^r - y_k^r)]}{\partial x_j^r} \quad (4.19)$$

$$= -\frac{1}{2} (\tilde{y}_i^r - y_i^r) C_{ik} \frac{\partial y_k^r}{\partial x_j^r} - \frac{1}{2} \frac{\partial y_i^r}{\partial x_j^r} C_{ik} (\tilde{y}_k^r - y_k^r) \quad (4.20)$$

$$= -(\tilde{y}_i^r - y_i^r) C_{ik} \frac{\partial y_k^r}{\partial x_j^r} \quad (4.21)$$

Similarly, the derivative of j with respect to θ_j is simply given by

$$\frac{\partial j}{\partial \theta_j} = -(\tilde{y}_i^r - y_i^r) C_{ik} \frac{\partial y_k^r}{\partial \theta_j}. \quad (4.22)$$

4.2.4 Illustration on the pyrolysis problem

As an illustration of the method, let us consider once more the single reaction problem from the previous chapter. To do so, synthetic data for the gas production π were produced using the forward model for the parameters in Table 4.1, with a constant time-step $\Delta t_r = 10$ s between the measurement times , as shown in Figure 4.1.

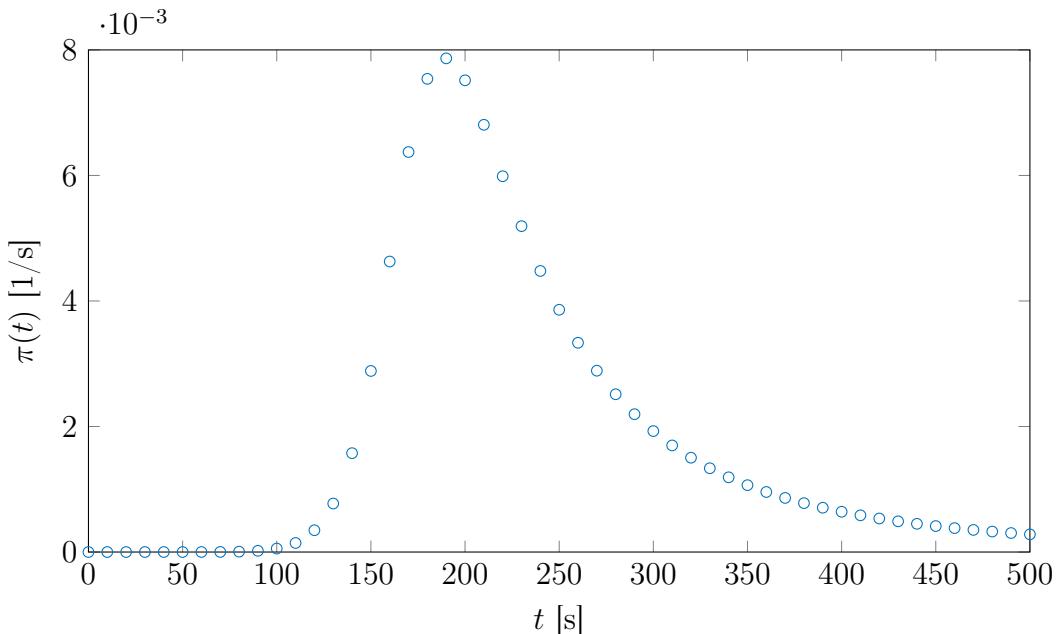


Figure 4.1: Synthetic data generated using the forward problem with the parameters from Table 4.1.

The least square cost function appearing in the inverse problem formulation is then written as

$$J(\boldsymbol{\theta}) = \frac{1}{2} \sum_r^{N_r} (\boldsymbol{\pi}^r - \boldsymbol{\pi}(t^r, \boldsymbol{\theta}))^T [C] (\boldsymbol{\pi}^r - \boldsymbol{\pi}(t^r, \boldsymbol{\theta})). \quad (4.23)$$

where the matrix $[C]$ is chosen to be a diagonal matrix with positive elements and the cost function is evaluated with a set of parameters different from the ones used to generate the data (Table 4.1). Once the problem has been formally defined, the adjoint equations can be solved. The solution $\lambda(t, \boldsymbol{\theta})$ is shown in Figure 4.2, when the gradient is evaluated using the parameters from Table 4.1. The markers show the solution at points where the cost function is defined. At those points, the gradient of the cost function appears in the adjoint equations, which results in the discontinuous peaks observed in the graph of $\lambda(t, \boldsymbol{\theta})$.

Once the adjoint solution is known, the gradient of the cost function can be readily computed using Eq(GradAdjoint). Table 4.2 reports the value of each component of the gradient computed by the adjoint method, as well as the finite differences approximation (with $\varepsilon_j = 0.01\theta_j$) and the gradient obtained with the forward sensitivity method.

	Adjoint method	Finite differences	Forward sensitivity
$\partial_A J$	-0.0113	-0.0112	-0.0113
$\partial_E J$	16.6474	16.5582	16.6474
$\partial_m J$	-209256.3	-209052.9	-209256.3

Table 4.2: Gradient of the cost function obtained with the adjoint method, the forward sensitivity method and the finite difference method.

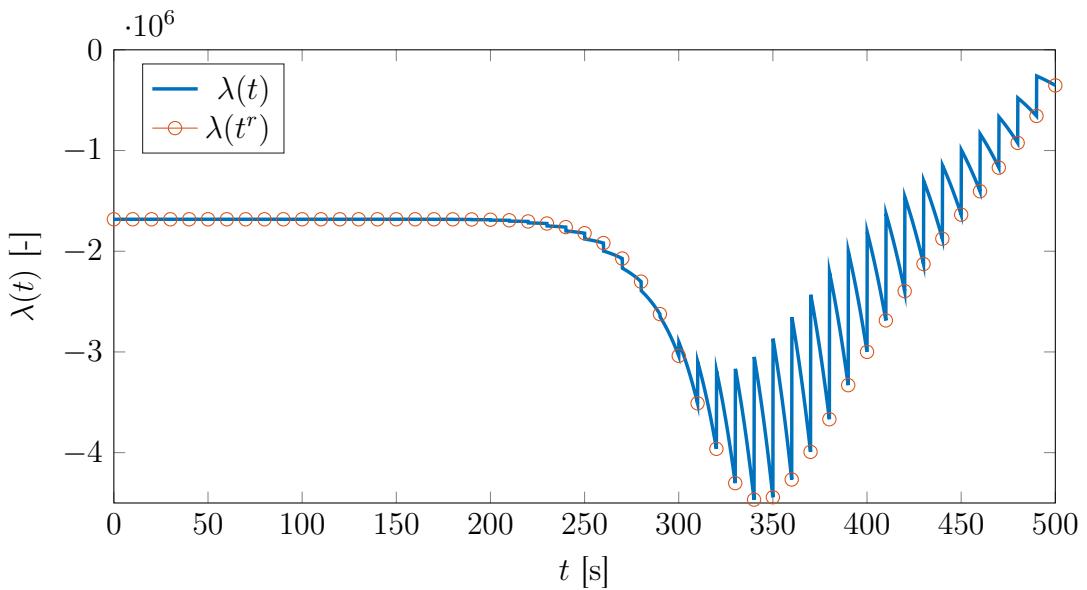


Figure 4.2: Adjoint solution λ as a function of time (with $\Delta t = 0.1$ s), where the markers are points at which the cost function is defined (with $\Delta t_r = 10$ s).

It is interesting to notice that the adjoint method and the forward sensitivity approach give the same value for each components of the gradient. This is indeed expected since these methods compute the exact gradient of the discretised system, whereas the finite difference

method is an approximation of the derivatives, entailed with an error depending on the value of ε in Eq.(4.2)

Because the accuracy of the adjoint method is constrained by the numerical solver used, it is also interesting to consider the convergence of the gradient as the time step is refined. This is shown in Figure 4.3, where the value of the gradient with respect to A and E for several Δt is shown. It can indeed be observed that these gradients converge to a single value (which can be identified as their exact value) as the time step is refined, and that, as for the forward problem, a choice must be made between the precision and the computational cost.

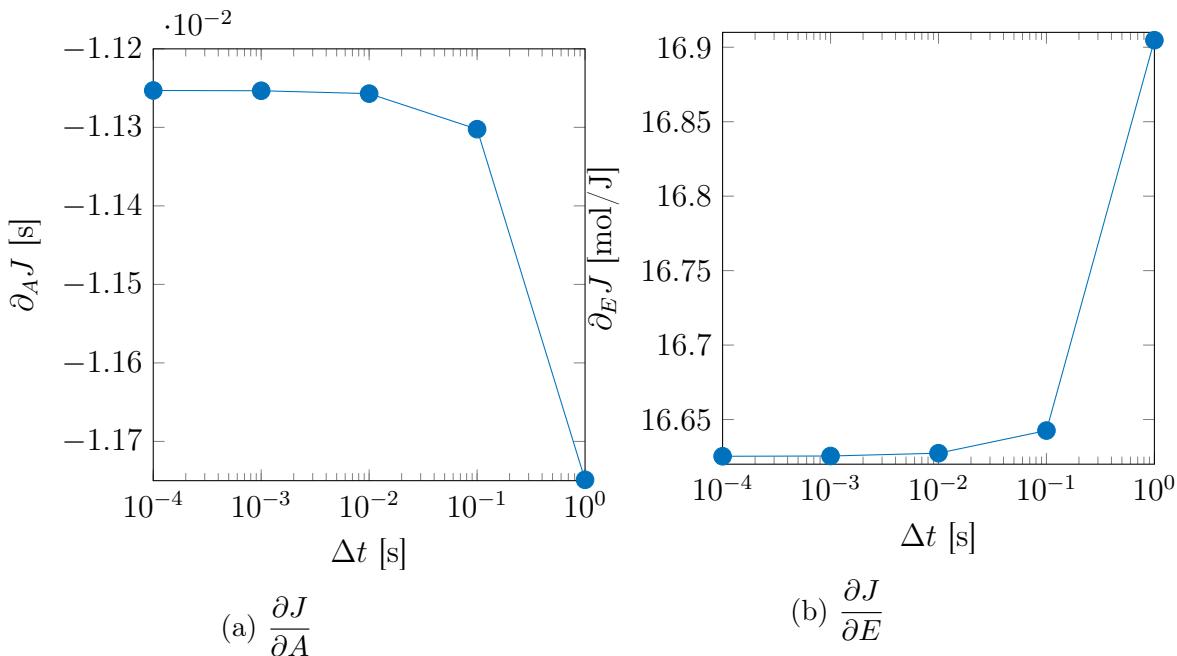


Figure 4.3: Convergence of the gradient of the cost function as the time step is refined.

It should be noted that the simplified problem is not fitted to assess the performance of the adjoint method compared to finite differences or forward sensitivity since only one reaction is considered. In order to set up a comparison, let us consider models that are more expensive in terms of computational cost. To do so, let us consider an increasing number of reactions producing only one gas.

This choice is motivated by the fact that the adjoint method, as well as the forward sensitivity method, are only expensive for computing the gradient of the cost function with respect to the parameters appearing in the dynamical equations, while the gradient of the cost function with respect to the mass fractions $F_{i,j}$ is trivial to compute and does not add much to the comparison. Hence, this setup allows a meaningful comparison while remaining rather simple. In particular, it is expected that adding more gases will result in a non negligible increase of the computational cost only if the gradient of the cost function is computed for each parameters using the finite difference method.

Figure 4.4 shows the CPU time for a number of reaction increasing from one to five using the three methods for computing the gradient described previously (with $\Delta t = 0.1$ s in the numerical model). As expected, as the number of reactions increases, the CPU time grows as well. In particular, the CPU time associated with the finite difference method is considerably larger than with the two other methods, as well as suffering from the accuracy

issues mentioned previously.

Furthermore, Figure 4.5 shows the the same results with only for the forward sensitivity and adjoint methods. As expected, the CPU time is higher for the former, and seems to increase fast despite the number of reactions remaining small, while the CPU time associated with the adjoint method remains almost constant.

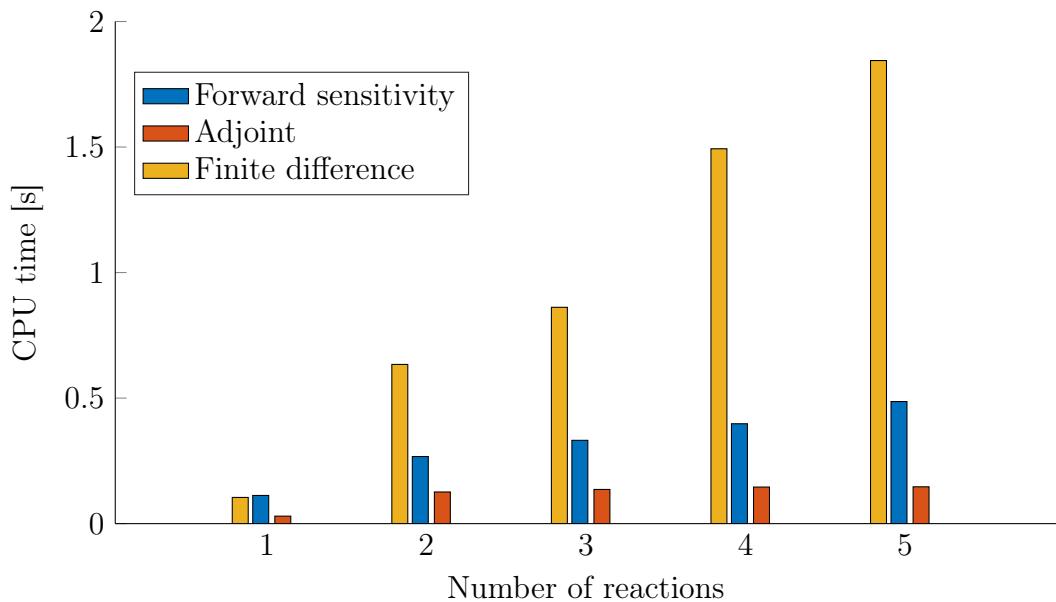


Figure 4.4: Comparison of the CPU time for the computation of the gradient with three different methods for a growing number of reactions.

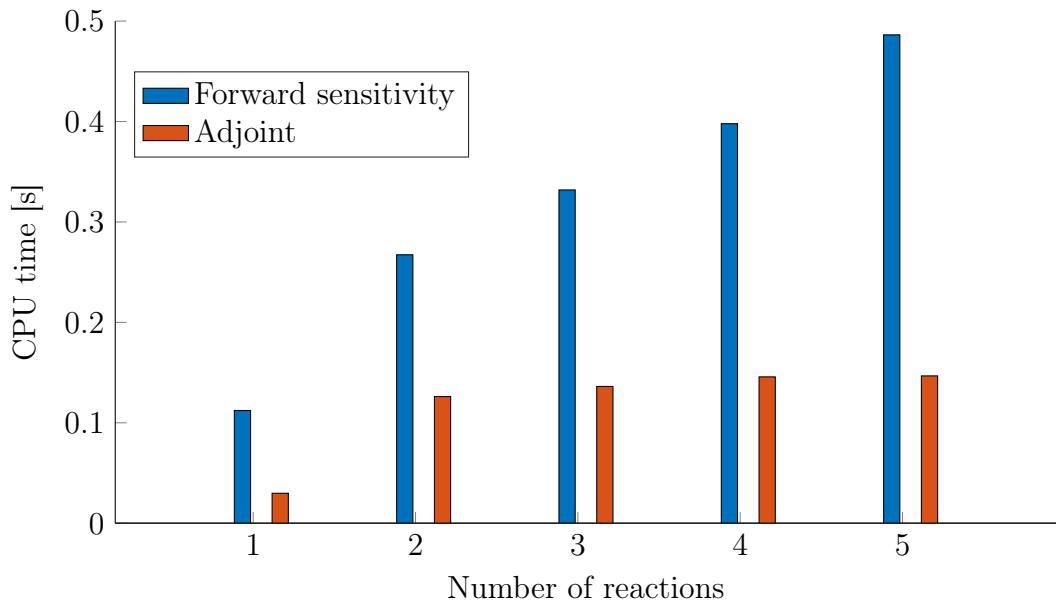


Figure 4.5: Comparison of the CPU time for the computation of the gradient with the forward sensitivity and adjoint methods for a growing number of reactions.

In conclusion, the discrete adjoint method developed in this chapter allows a fast and accurate computation of the gradient of a least-square cost function. It is particularly

important since methods for Bayesian inference that use the gradient of a least square cost function require an efficient computation to be used with a relatively high number of iterations, while the additional precision compared to the finite difference method should also impact the quality of the results obtained.

CHAPTER 5

NUMERICAL METHODS FOR BAYESIAN INFERENCE

In the second chapter, the Bayesian formulation for the solution of inverse problems was described by the means of a probability density function. In particular, the parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$ of a forward model $\mathbf{y}(\boldsymbol{\theta})$ that maps $\boldsymbol{\theta}$ to observations $\mathbf{d} \in D \subset \mathbb{R}^m$ were assumed to be random variables. The solution of the inverse problem, which consists in finding $\boldsymbol{\theta}$ knowing some observations \mathbf{d}^{obs} with measurement errors $\boldsymbol{\varepsilon}$,

$$\mathbf{y}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon},$$

is interpreted as a joint probability density function $\pi(\boldsymbol{\theta} | \mathbf{d}^{\text{obs}})$ using Bayes' formula.

The goal of this chapter is to introduce numerical methods used to sample from the posterior density. In particular, Markov Chain Monte Carlo methods using information on the geometry of the posterior distribution through its gradient are explored. These methods are then applied to test cases involving the pyrolysis model developed in the previous chapters.

5.1 Linearised inverse problem

Before tackling more involved problems, let us first consider a linear inverse problem. Indeed, when the relation between the parameters and the observed data is linear, the posterior density $\pi(\boldsymbol{\theta} | \mathbf{d})$ takes a simple form that can directly be analysed without employing Markov Chain Monte Carlo methods. Indeed, Tarantola [28] showed that if the forward model is linear, *i.e.* of the form

$$\mathbf{y}(\boldsymbol{\theta}) = [G]\boldsymbol{\theta}, \quad (5.1)$$

where the matrix $[G]$ does not depend on the parameters, then for a well chosen prior density, the posterior density can be written as a Gaussian distribution. Because methods for sampling independent samples from a Gaussian distribution are efficient, even in high dimensions, having access to a Gaussian approximation of the solution can prove itself useful.

For instance, if a Gaussian prior is used, the posterior density can be written as

$$\pi(\boldsymbol{\theta} | \mathbf{d}) = k \exp \left(-\frac{1}{2} \|(\mathbf{d} - [G]\boldsymbol{\theta})\|_{\Sigma}^2 - \frac{1}{2} \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|_P^2 \right), \quad (5.2)$$

where $\|\mathbf{x}\|_A^2 = \mathbf{x}^T [A] \mathbf{x}$, $[\Sigma]$ is the covariance matrix of the likelihood, $[P]$ is the covariance matrix of the prior density, $\tilde{\boldsymbol{\theta}}$ is a reference value of the parameters and k is a normalisation constant. Developing Eq.(5.2), Tarantola showed that the posterior density can be rewritten in the form

$$\pi(\boldsymbol{\theta} | \mathbf{d}) = k \exp \left(-\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T [\hat{\Sigma}]^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right), \quad (5.3)$$

which is a Gaussian probability density function with a covariance matrix

$$[\hat{\Sigma}] = ([G]^T [\Sigma]^{-1} [G] + [P]^{-1})^{-1}. \quad (5.4)$$

and of mean

$$\hat{\boldsymbol{\theta}} = [\hat{\Sigma}] \left([G]^T [\Sigma]^{-1} \mathbf{d} + [P]^{-1} \tilde{\boldsymbol{\theta}} \right). \quad (5.5)$$

This methodology can then be applied to the linearised gas production given by Eq.(3.37), where the linearisation is performed, for instance, around the maximum likelihood estimator, which can be obtained by deterministic optimisation. To do so, let us further assume that an uninformative prior is used, such that the terms related to the Gaussian prior in Eq.(5.2) vanish. Then, replacing the linear model by the linearised gas production, one obtains the following posterior density:

$$\pi(\boldsymbol{\theta} | \mathbf{d}) = k \exp \left(-\frac{1}{2} \sum_{r=0}^{N_r} \|(\mathbf{d}^r - \pi(t^r, \tilde{\boldsymbol{\theta}}) - [G(t^r, \tilde{\boldsymbol{\theta}})](\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})\|_\Sigma^2 \right). \quad (5.6)$$

The solution can then be expressed in the form of Eq.(5.3) with covariance matrix

$$[\hat{\Sigma}] = \left(\sum_{r=0}^{N_r} [G(t^r, \tilde{\boldsymbol{\theta}})]^T [\Sigma]^{-1} [G(t^r, \tilde{\boldsymbol{\theta}})] \right)^{-1}. \quad (5.7)$$

and of mean

$$\hat{\boldsymbol{\theta}} = [\hat{\Sigma}] \sum_{r=0}^{N_r} [G(t^r, \tilde{\boldsymbol{\theta}})]^T [\Sigma]^{-1} (\mathbf{d}^r - \mathbf{y}(t^r, \tilde{\boldsymbol{\theta}}) + [G(t^r, \tilde{\boldsymbol{\theta}})] \tilde{\boldsymbol{\theta}}). \quad (5.8)$$

Having access to an estimation of the covariance matrix $[\hat{\Sigma}]$ can greatly improve the efficiency of Markov Chain Monte Carlo methods, as it will be seen in the next sections.

Moreover, it also allows to run preliminary diagnostics on the posterior density before using any sampling methods. For example, as it will be shown in the following sections in the case of the simplified pyrolysis model, an uninformative prior was usually used when comparing the methods on data sets with small measurement noises. However, when the uncertainty on the data is higher, the linear approximation resulted in a distribution where some parameters were taking negative values, which suggested that using an informative prior was wrong.

5.2 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo, also called hybrid Monte Carlo, was first developed by Duane [12] for applications in lattice field theory, and was later extended to applications in statistical

inference [22]. In a few words, it is based on the generation of samples by simulating the dynamics of an Hamiltonian system. The strength of this method is that, through Hamilton's equation, it uses information on the geometry of the distribution, which can help overcoming the typical issues related to higher dimension and correlation between the parameters arising in the Metropolis-Hastings sampling algorithm.

5.2.1 Hamiltonian dynamics

To formally define the Hamiltonian Monte Carlo, some notions from Hamiltonian mechanics are necessary. Let us consider a system evolving in a d -dimensional space localised by its coordinates q_i . The equations describing its motion in the Hamiltonian formalism can be expressed as follows:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad (5.9)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad (5.10)$$

with initial conditions $q_i(0) = q_i^0$ and $p_i(0) = p_i^0$. The Hamiltonian $H(\mathbf{q}, \mathbf{p})$ is introduced as a measure of the total energy of the system. It often happens that Hamiltonians are separable and can be split into the contribution of two terms, namely the potential energy $U(\mathbf{q})$ and the kinetic energy $T(\mathbf{p})$, such that $H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + T(\mathbf{p})$. It should be noted that it is here assumed that both the potential energy and kinetic energy only depends on the position and the momentum respectively. Hamilton's equation can then be written as

$$\frac{dq_i}{dt} = \frac{\partial T}{\partial p_i}, \quad (5.11)$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i}. \quad (5.12)$$

The Hamiltonian dynamics possesses some important properties that are worth mentioning [21]:

1. The Hamiltonian dynamics is reversible. Let us define the Hamiltonian flow as

$$\phi_t : (\mathbf{q}_0, \mathbf{p}_0) \rightarrow (\mathbf{q}(t), \mathbf{p}(t)), \quad t \geq 0.$$

Then, it appears that reversed flow ϕ_{-t} is simply obtained by negating the time derivatives in Hamilton's equation.

2. The Hamiltonian is conserved, since

$$\frac{dH}{dt} = \left(\frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial t} + \frac{\partial H}{\partial q_i} \frac{\partial q_i}{\partial t} \right) \quad (5.13)$$

$$= 0. \quad (5.14)$$

3. The Hamiltonian flow ϕ_t is symplectic. Defining $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ and rewriting Hamilton's equations as

$$\frac{d\mathbf{x}}{dt} = [J] \nabla_{\mathbf{x}} H, \quad [J] = \begin{bmatrix} 0 & [\mathbb{I}] \\ -[\mathbb{I}] & 0 \end{bmatrix},$$

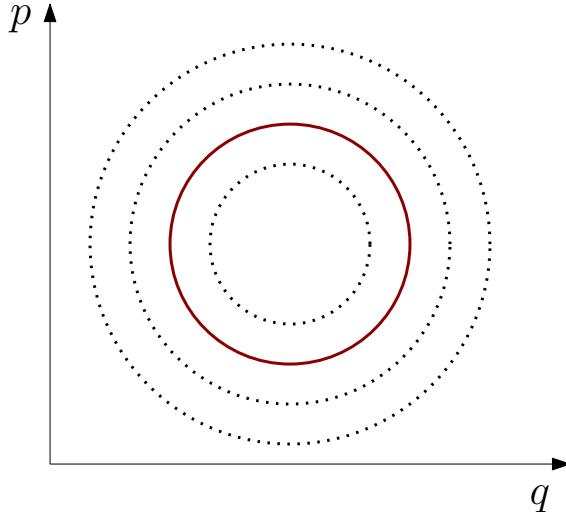


Figure 5.1: Example of the phase space representation of a trajectory generated by Hamiltonian dynamics.

where $[\mathbb{I}]$ is the $d \times d$ identity matrix. The flow ϕ_t is said to be symplectic if

$$(\nabla_{\mathbf{x}}\phi_t)^T [J]\nabla_{\mathbf{x}}\phi_t = [J].$$

Since $\det([J]) = 1$, one has $\det(\nabla_{\mathbf{z}}\phi_t)^2 = 1$ which implies volume conservation in the phase space, since

$$\begin{aligned} \int_{\phi_t(V)} d\mathbf{q}d\mathbf{p} &= \int_V \det(\nabla_{\mathbf{z}}\phi_t) d\mathbf{q}d\mathbf{p} \\ &= \int_V d\mathbf{q}d\mathbf{p}. \end{aligned}$$

A direct implication of these properties is that the trajectory of $(q_i(t), p_i(t))$ in the $2d$ -dimensional phase space follows an hypersurface of constant Hamiltonian,

$$\mathcal{S} = \{(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^{2d} : H((\mathbf{p}, \mathbf{q}) = E\},$$

where the constant E depends on the initial conditions of the problem. For example, this can be seen in Figure 5.1 in the case $d = 1$. The dotted lines are hypersurfaces of constant Hamiltonian and the solid red line is an example of a trajectory obtained by solving Hamilton's equations.

5.2.2 Time integration of Hamilton's equations

Except in simpler cases such as the harmonic oscillator, Hamilton's equation cannot be solved analytically and numerical method for the time integration must be developed. An additional difficulty that arises when solving the Hamiltonian dynamics lies in the fact that the numerical solver should respect the different properties stated previously.

Let us first consider the discretisation of Eq.(5.12) using the explicit Euler method, which

is given by

$$p_i^{k+1} = p_i^k - \Delta t \frac{\partial U}{\partial q_i}(\mathbf{q}^k), \quad (5.15)$$

$$q_i^{k+1} = q_i^k + \Delta t \frac{\partial T}{\partial p_i}(\mathbf{p}^k). \quad (5.16)$$

It is possible to show that this numerical scheme is not suitable for Hamiltonian dynamics. Defining the numerical flow $\Phi_{\Delta t}(\mathbf{p}^k, \mathbf{q}^k) = (\mathbf{p}^{k+1}, \mathbf{q}^{k+1})$, it follows that the determinant of the Jacobian matrix of the numerical flow is, in general, not equal to one, and the third property does not hold. It is generally observed, even in simple situations such as an harmonic oscillator, that the Hamiltonian is also not conserved with the Euler method, which can be observed in the phase space where the trajectory drifts to infinity.

Time integration of the Hamiltonian dynamics therefore requires a specific class of temporal integrators, often called symplectic integrators [18], which are specifically designed with the properties of the Hamiltonian in mind. For instance, Euler's method can be made symplectic by evaluating the gradient of the kinetic energy in Eq.(5.16) at \mathbf{p}^{k+1} . Among the symplectic integrators, the Störmer-Verlet scheme, which is of order two, is often used. It is given by

$$p_i^{k+1/2} = p_i^k - \frac{\Delta t}{2} \frac{\partial U}{\partial q_i}(\mathbf{q}^k), \quad (5.17)$$

$$q_i^{k+1} = q_i^k + \Delta t \frac{\partial T}{\partial p_i}(\mathbf{p}^{k+1/2}), \quad (5.18)$$

$$p_i^{k+1} = p_i^{k+1/2} - \frac{\Delta t}{2} \frac{\partial U}{\partial q_i}(\mathbf{q}^{k+1}). \quad (5.19)$$

As an illustration of the time integration of the Hamiltonian dynamics and problems that can arise if the time integration method is not chosen appropriately, let us consider, as in [22], a one dimensional problem where the kinetic energy and potential energy are given by

$$T = \frac{p^2}{2}, \quad U = \frac{q^2}{2},$$

with initial conditions $p_0 = 0$, $q_0 = 1$. The exact solution to this problem is simply $q(t) = \cos(t)$, $p(t) = -\sin(t)$, $0 \leq t \leq \tau$. Figure 5.2 shows a comparison between the exact solution and the numerical solution obtained with the two time integration methods mentioned above, where $\Delta t = 0.1$, $\tau = 10$. As expected, the numerical trajectory obtained with the explicit Euler method drifts from the exact trajectory and the Hamiltonian is not conserved, whereas the Störmer-Verlet is able to accurately reproduce the exact trajectory.

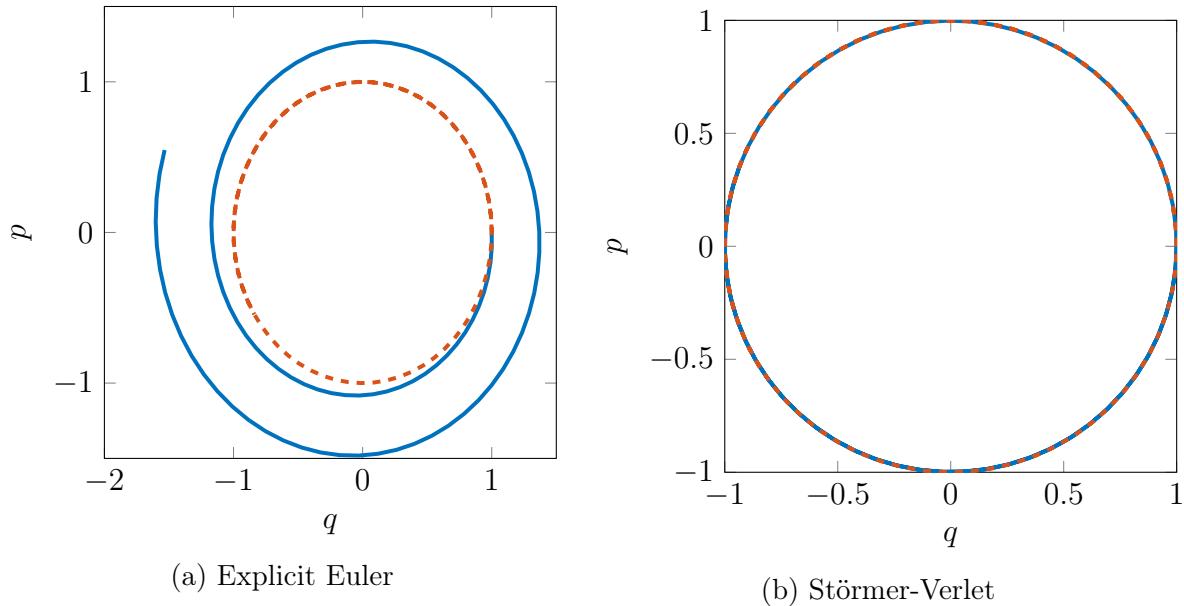


Figure 5.2: Comparison between the exact solution (dashed line) and numerical solution (solid line) for the explicit Euler method and Störmer-Verlet method.

5.2.3 Motivation for the Hamiltonian Monte Carlo method

Let us consider a Hamiltonian system composed of N particles evolving in a 3 dimensional space such that $d = 3N$, which can for instance represent the motion of a gas composed of a large number of molecules N . In practice, the evolution of the system is given through the solution of Hamilton's equation. Once the equations have been solved, we can characterise observable values of the system $f(\mathbf{q}, \mathbf{p})$ by computing its time average on the trajectory, *i.e.*

$$\bar{f} = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(\mathbf{q}, \mathbf{p}) d\tau. \quad (5.20)$$

However, since the system is composed of many particles, it is often impossible to directly solve these equations. In statistical physics, the focus is not on solving these equations for each particle, but rather studying an ensemble of many identical replicas of the system which each represents a possible state of the system, where one state of the system represents a point in the phase space (\mathbf{q}, \mathbf{p}) . Introducing the probability density $\rho(\mathbf{q}, \mathbf{p})$ as the probability of the system being in a volume $d\mathbf{q}d\mathbf{p}$, *i.e* the probability of being in one sub-system of the ensemble of systems, this allows the computation of the observable quantities of the system under the ergodicity assumption, *i.e.* that the time averages are equal to the ensemble averages. Here, we will place ourselves in the case of the so-called canonical ensemble, where the system considered has a fixed temperature but can exchange energy with its surrounding environment. At equilibrium, the energy of the system fluctuates but remains constant in average [21].

The conservation of volume in the phase space (\mathbf{q}, \mathbf{p}) under an Hamiltonian flow ϕ_t is generally referred as Liouville's theorem. Another interesting formulation of this theorem consists in defining the probability density function $\rho(\mathbf{q}, \mathbf{p}, t)$ which characterize the probability of finding the system in a volume $d\mathbf{q}d\mathbf{p}$ around the state (\mathbf{q}, \mathbf{p}) . The volume preservation in the phase space implies that the time derivative of $\rho(\mathbf{q}, \mathbf{p})$ follows a continuity equation for

an incompressible flow. This results in the following formulation of Liouville's theorem:

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0, \quad (5.21)$$

where we introduced the Poisson brackets

$$\{x, y\} = \left(\frac{\partial x}{\partial q_i} \frac{\partial y}{\partial p_i} - \frac{\partial x}{\partial p_i} \frac{\partial y}{\partial q_i} \right). \quad (5.22)$$

This formulation of Liouville's theorem hence states that the density $\rho(\mathbf{q}, \mathbf{p}, t)$ is constant under the Hamiltonian flow ϕ_t for an observer following the trajectory on the phase plane (\mathbf{q}, \mathbf{p}) .

Considering a function $f(\mathbf{q}, \mathbf{p})$ that is square integrable with respect to $\rho(\mathbf{q}, \mathbf{p}, t)$, probabilistic averages of $f(\mathbf{q}, \mathbf{p})$ are defined as

$$\bar{f} = \int_{\mathbb{R}^{2d}} f(\mathbf{q}, \mathbf{p}) \rho(\mathbf{q}, \mathbf{p}, t) d\mathbf{q} d\mathbf{p}. \quad (5.23)$$

Since it is assumed that the function $f(\mathbf{q}, \mathbf{p})$ does not explicitly depends on time, a direct implication of Liouville's theorem is that the time derivative of \bar{f} is

$$\frac{d\bar{f}}{dt} = \overline{\{f, H\}}. \quad (5.24)$$

In particular, this implies that, for a system with a time independent Hamiltonian $H(\mathbf{q}, \mathbf{p})$, the average Hamiltonian of the system is a constant. Let us also assume that the density $\rho(\mathbf{q}, \mathbf{p}, t)$ is time independent, which is the case for systems at equilibrium, such that Liouville's theorem rewrites

$$\{\rho, H\} = 0, \quad (5.25)$$

which implies $\rho(\mathbf{q}, \mathbf{p}) = \rho(H(\mathbf{q}, \mathbf{p}))$.

These properties can be used to construct the density $\rho(\mathbf{q}, \mathbf{p})$ using, for instance, the maximum entropy principle [4]. Defining the Shannon entropy associated with $\rho(\mathbf{q}, \mathbf{p})$ as

$$s(\rho) = - \int_{\mathbb{R}^{2d}} \rho(\mathbf{q}, \mathbf{p}) \log(\rho(\mathbf{q}, \mathbf{p})) d\mathbf{q} d\mathbf{p}, \quad (5.26)$$

the maximum entropy principle consists in finding $\rho(\mathbf{q}, \mathbf{p})$ that maximise the entropy $s(\rho)$ such that a given set of constraints is satisfied. In the case of the canonical ensemble, the constraint is that the average of the Hamiltonian is constant. The optimisation problem then reads

$$\max_{\rho} s(\rho), \quad (5.27)$$

subject to:

$$\int_{\mathbb{R}^{2d}} \rho(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} - 1 = 0, \quad (5.28)$$

$$\int_{\mathbb{R}^{2d}} H(\mathbf{q}, \mathbf{p}) \rho(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} - \bar{H} = 0. \quad (5.29)$$

The Lagrangian \mathcal{L} associated to the optimisation problem reads

$$\mathcal{L} = -s(\rho) - \lambda_0 \left(\int_{\mathbb{R}^{2d}} \rho(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} - 1 \right) - \lambda_1 \left(\int_{\mathbb{R}^{2d}} H(\mathbf{q}, \mathbf{p}) \rho(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} - \bar{H} \right), \quad (5.30)$$

where λ_0 and λ_1 are the Lagrange multipliers, and, taking the variation of the Lagrangian, it can be shown [4] that the solution to the optimisation problems results in the Gibbs density

$$\rho(\mathbf{q}, \mathbf{p}) = k(\lambda_0) \exp(-\lambda_1 H(\mathbf{q}, \mathbf{p})), \quad (5.31)$$

where $k(\lambda_0)$, which is a normalisation constant, and λ_1 are determined such that the constraints are satisfied.

This introduction of the probability density function $\rho(\mathbf{q}, \mathbf{p})$ suggests that the Hamiltonian dynamics could be used to sample from the posterior distribution in Bayesian inference. Assuming that the parameters $\boldsymbol{\theta}$ in the Bayesian formulation corresponds to the coordinates \mathbf{q} in the Hamiltonian dynamics, and introducing the momentum \mathbf{p} as an auxiliary variable, the potential energy can be chosen as

$$U(\mathbf{q}) = -\log(\pi(\mathbf{q} | \mathbf{d}^{\text{obs}})) \quad (5.32)$$

$$= U_{\text{Prior}}(\mathbf{q}) + U_{\text{L}}(\mathbf{q}), \quad (5.33)$$

where $U_{\text{Prior}}(\mathbf{q})$ and $U_{\text{L}}(\mathbf{q})$ are the contributions to the potential from the prior and the likelihood respectively. The potential due to the likelihood takes the general form

$$U_{\text{L}}(\mathbf{q}) = \sum_{r=0}^{N_r} \frac{1}{2} (\tilde{y}_i^r - y_i^r) C_{ik} (\tilde{y}_k^r - y_k^r), \quad (5.34)$$

for which the gradient required to simulate the Hamiltonian dynamics can be computed with the adjoint method developed in the previous chapter. The choice of the log-posterior as the potential energy makes the trajectories dependent on the geometry of the posterior density through the computation of its gradient.

It should be mentioned that the potential $U(\mathbf{q})$ must be a continuous function from \mathbb{R}^d to \mathbb{R} for the derivative to be defined. Hence when the prior distribution for a parameter α is chosen to be uniform on an interval $[a, b]$, *i.e.*

$$\pi_0(\alpha) = \mathbb{I}_{[a,b]}(\alpha),$$

where \mathbb{I} is the indicator function, the derivative is not defined and a regularisation of the prior must be performed. Regularisation methods includes smoothing the prior by convolution with a Gaussian kernel [17] or by applying a change of variables [3].

The kinetic energy term is usually chosen to be quadratic, *i.e.*

$$T(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T [M]^{-1} \mathbf{p} + C, \quad (5.35)$$

where $[M]$ is a symmetric positive definite matrix that is assumed to be constant, and C is a constant resulting from the normalisation constant in the posterior density. Under these choices for the potential energy and the kinetic energy, Hamilton's equations can finally be rewritten as

$$\frac{d\mathbf{q}}{dt} = [M]^{-1} \mathbf{p}, \quad (5.36)$$

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial U_{\text{prior}}(\mathbf{q})}{\partial \mathbf{q}} - \frac{\partial U_{\text{L}}(\mathbf{q})}{\partial \mathbf{q}}. \quad (5.37)$$

5.2.4 Choice for the matrix $[M]$

The choice of the matrix $[M]$ in Hamilton's equation influences the performance of the method. Following the developments from [3], let us write Cholesky decomposition of $[M]^{-1} = [L][L]^T$, where L is a lower triangular matrix. Introducing the change of variables

$$\mathbf{Q} = [L]^{-1}\mathbf{q}, \quad \mathbf{P} = [L]^T\mathbf{p},$$

such that Hamilton's equation can be expressed as

$$\frac{d\mathbf{Q}}{dt} = \mathbf{P}, \tag{5.38}$$

$$\frac{d\mathbf{P}}{dt} = -\frac{\partial U_{\text{prior}}([L]\mathbf{Q})}{\partial \mathbf{Q}} - \frac{\partial U_L([L]\mathbf{Q})}{\partial \mathbf{Q}}. \tag{5.39}$$

The matrix $[M]$ can thus be interpreted as a change of scales and correlations, such that the Hamiltonian dynamics in the original variables is equivalent to the dynamics in the transformed variables [3]. This suggests that, if $[M]$ is chosen carefully, the original inference problem can be made similar to an inference problem for the transformed variables (\mathbf{Q}, \mathbf{P}) with a covariance matrix $[\Sigma_Q] \approx [\mathbb{I}]$. It can be shown that it is the case when the matrix $[M]$ is chosen to be the covariance matrix of \mathbf{q} .

In practice, it often happens that the covariance matrix of \mathbf{q} is not directly accessible, but can be estimated. For instance, when looking at Gaussian posterior distributions, the covariance matrix is given by the inverse of the Hessian matrix of the potential $U(\mathbf{q})$. Hence, even for arbitrary posterior, an estimate of the covariance matrix can be approached by the Hessian matrix, evaluated, for example, at the maximum likelihood estimator.

Another choice for the estimation of the covariance matrix is to consider the linearised inverse problem in Eq.(5.1). Under this assumption, the posterior distribution was found to be a Gaussian distribution with covariance matrix given by Eq.(5.7), which can be used as an estimate for the covariance matrix of \mathbf{q} .

5.2.5 Hamiltonian Monte Carlo algorithm

Hamiltonian Monte Carlo seeks to explore the posterior distribution by simulating the Hamiltonian dynamics during a time period τ , such that, at iteration (ℓ) the algorithm propose a new set of positions and momenta

$$(\mathbf{q}^{(\ell+1)}, \mathbf{p}^{(\ell+1)}) = \phi_\tau \left(\mathbf{q}^{(\ell)}, \mathbf{p}_0^{(\ell+1)} \right),$$

where the initial momentum is chosen as $\mathbf{p}_0^{(\ell+1)} \sim \mathcal{N}(\mathbf{0}, [M])$.

Choosing the initial momentum at iteration (ℓ) to be a random variable allows the exploration of the posterior by looking at trajectories defined on different energy level at each iterations. Figure 5.3 shows the principle of the Hamiltonian Monte Carlo method when we look at the trajectories in the phase space. The red arrows are trajectories generated by the Hamiltonian dynamics, while the blue arrows show the generation of a new momentum variable at the start of a new iteration.

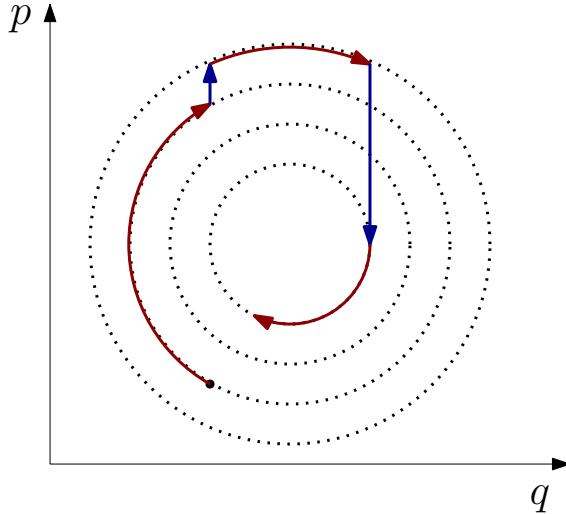


Figure 5.3: Principle of the Hamiltonian Monte Carlo algorithm illustrated in the phase space. Adapted from [8].

If the time integration of the Hamiltonian dynamics is not exact, which is the case of numerical methods are used, a accept-reject step as in the Metropolis-Hastings algorithm is used, and the transition is accepted if

$$\mathcal{U}(0, 1) \leq \min \left\{ 1, \frac{\exp(-H(\mathbf{q}^{(\ell+1)}, \mathbf{p}^{(\ell+1)}))}{\exp(-H(\mathbf{q}^{(\ell)}, \mathbf{p}_0^{(\ell+1)}))} \right\}. \quad (5.40)$$

A consequence of the accept-reject step is that if the Hamiltonian is not conserved, which is the case when the explicit Euler time integration method is used, it is expected that the algorithm rejects more proposals. Similarly, if a coarse time discretisation is used, the Hamiltonian at the end of the trajectory will also differ from the initial Hamiltonian, which will also result in lower acceptance rate. The final Hamiltonian Monte Carlo algorithm is summarised in Alg.(1).

According to Neal [22], the sequence generated by the Hamiltonian Monte Carlo algorithm $\{(\mathbf{q}^\ell, \mathbf{p}^\ell), \ell = 1, 2, \dots, N\}$ forms an ergodic Markov Chain, such that the law of large numbers holds. Hence, probabilistic averages can be evaluated, for a function $f(\mathbf{q}, \mathbf{p})$ square-integrable with respect to the Gibbs density as

$$\int_{\mathbb{R}^{2d}} f(\mathbf{q}, \mathbf{p}) k \exp(-U(\mathbf{q}) - T(\mathbf{p})) d\mathbf{q} d\mathbf{p} \approx \frac{1}{N} \sum_{\ell=1}^N f(\mathbf{q}^\ell, \mathbf{p}^\ell). \quad (5.41)$$

However, since \mathbf{p} is an auxiliary variable, probabilistic average for the inference problem are carried out for a function $f(\mathbf{q})$ square-integrable with respect to the posterior density. Because the kinetic energy was chosen to be a quadratic form $T(\mathbf{p}) = \mathbf{p}^T [M]^{-1} \mathbf{p}$, the integral of $k_p \exp(-T(\mathbf{p}))$ over \mathbb{R}^d is equal to one, and one has

$$\int_{\mathbb{R}^d} f(\mathbf{q}) k_q \exp(-U(\mathbf{q})) d\mathbf{q} \approx \frac{1}{N} \sum_{\ell=1}^N f(\mathbf{q}^\ell). \quad (5.42)$$

Algorithm 1 Hamiltonian Monte Carlo

```
1: Initialize:  $\mathbf{q}^{(1)} = \mathbf{q}_0$ ,  $[M]$ ,  $\varepsilon$ ,  $L$ ,  $N_{\text{iter}}$ , Integrator
2: for  $i = 1 : N_{\text{iter}}$  do
3:   Draw  $\mathbf{p}_0 \sim \mathcal{N}(\mathbf{0}, [M])$ 
4:    $[\mathbf{q}^*, \mathbf{p}^*] = \text{TimeIntegration}(\mathbf{q}^{(i)}, \mathbf{p}_0, \varepsilon, L, \text{Integrator})$ 
5:   Compute the acceptance probability

$$\alpha(H^*, H^0) = \min \{1, \exp(-H(\mathbf{q}^*, \mathbf{p}^*) + H(\mathbf{q}^{(i)}, \mathbf{p}_0))\}$$

6:   if  $\mathcal{U}(0, 1) \leq \alpha(H^*, H^0)$  then
7:      $\mathbf{q}^{(i+1)} = \mathbf{q}^*$ 
8:   else
9:      $\mathbf{q}^{(i+1)} = \mathbf{q}^{(i)}$ 
10:  end if
11: end for
```

5.2.6 Application to the simplified pyrolysis problem

To illustrate some of the properties of the Hamiltonian Monte Carlo method, let us apply the sampling algorithm to the simplified pyrolysis problem from the previous chapters. Synthetic data were generated using the forward problem in accordance with Figure 4.1, to which a Gaussian noise of standard deviation $\sigma = 10^{-5}$ was added.

The inference problem consists in estimating the kinetic parameters (A, E, m) . Because the noise added to the data is relatively small, the maximum likelihood estimator is considered to be the value of the parameters used to generate the data. Furthermore, we assume that the prior is uninformative, such that the posterior takes the form of Eq.(5.6). First, the linearised inverse problem is used to obtain an estimation of the covariance matrix of the posterior density, $[\hat{\Sigma}_q]$, using Eq.(5.7). Then, the matrix $[M]$ is set to be the inverse of $[\Sigma_q]$, as established previously.

Figure 5.4 shows the trajectories generated by the Hamiltonian dynamics using the Störmer-Verlet integrator with $\Delta t = 0.1$ and $\tau = 25$ during a single Hamiltonian Monte Carlo iteration (ℓ).

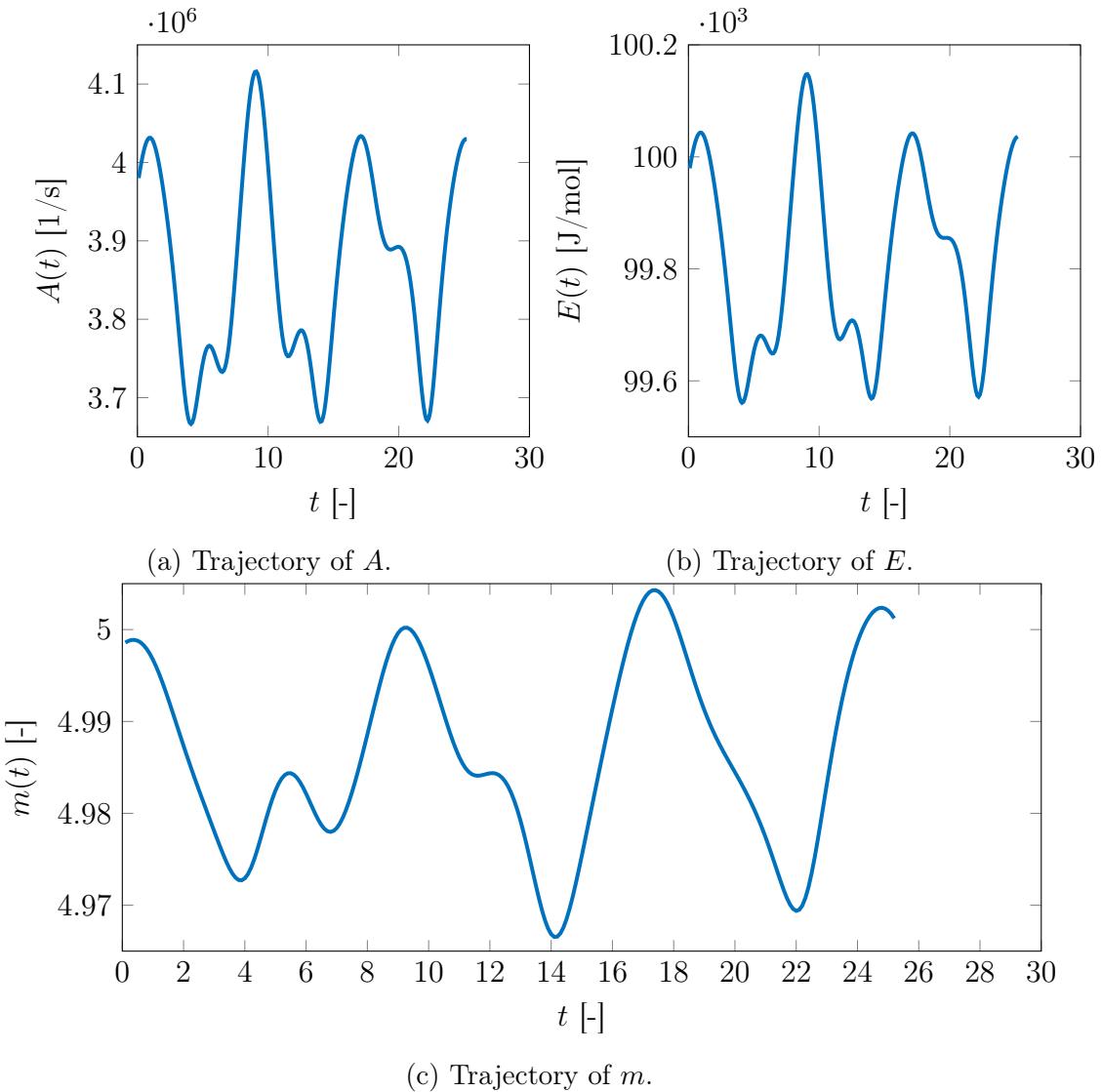


Figure 5.4: Trajectory of the parameters generated by simulating the Hamiltonian dynamics for the simplified pyrolysis problem, with $\Delta t = 0.1$, $\tau = 25$, when $[M] = [\hat{\Sigma}_q]^{-1}$.

Several observations can be drawn from analysing the trajectories. First, it is interesting to notice that, while the trajectories are not periodic, they still tend to oscillate over time. This implies that, for some final time τ , there is a possibility of generating a proposal that is equal or close to the initial values of the parameters at the iteration (ℓ) . For the algorithm to efficiently sample the posterior density, this type of behaviour should be avoided by carefully choosing the simulation time τ .

Furthermore, these trajectories explore the parameter space relatively well. As reported in [3, 22], when the matrix $[M] = [\hat{\Sigma}_q]^{-1}$, the trajectories generated by the Hamiltonian dynamics should explore the parameter space in an efficient manner for parameters with high uncertainties, as well as parameters with lower uncertainties. To illustrate this property, these trajectories can be compared to the ones obtained when the matrix $[M]$ is now chosen, for example, as $[M'] = \text{diag}([M])$, noting that if the posterior density was Gaussian, this could mean that the covariance matrix of the posterior is approximated by the diagonal elements of the Hessian matrix of the potential $U(\mathbf{q})$, thus neglecting any correlation between the parameters. The new trajectories are illustrated in Figure 5.5.

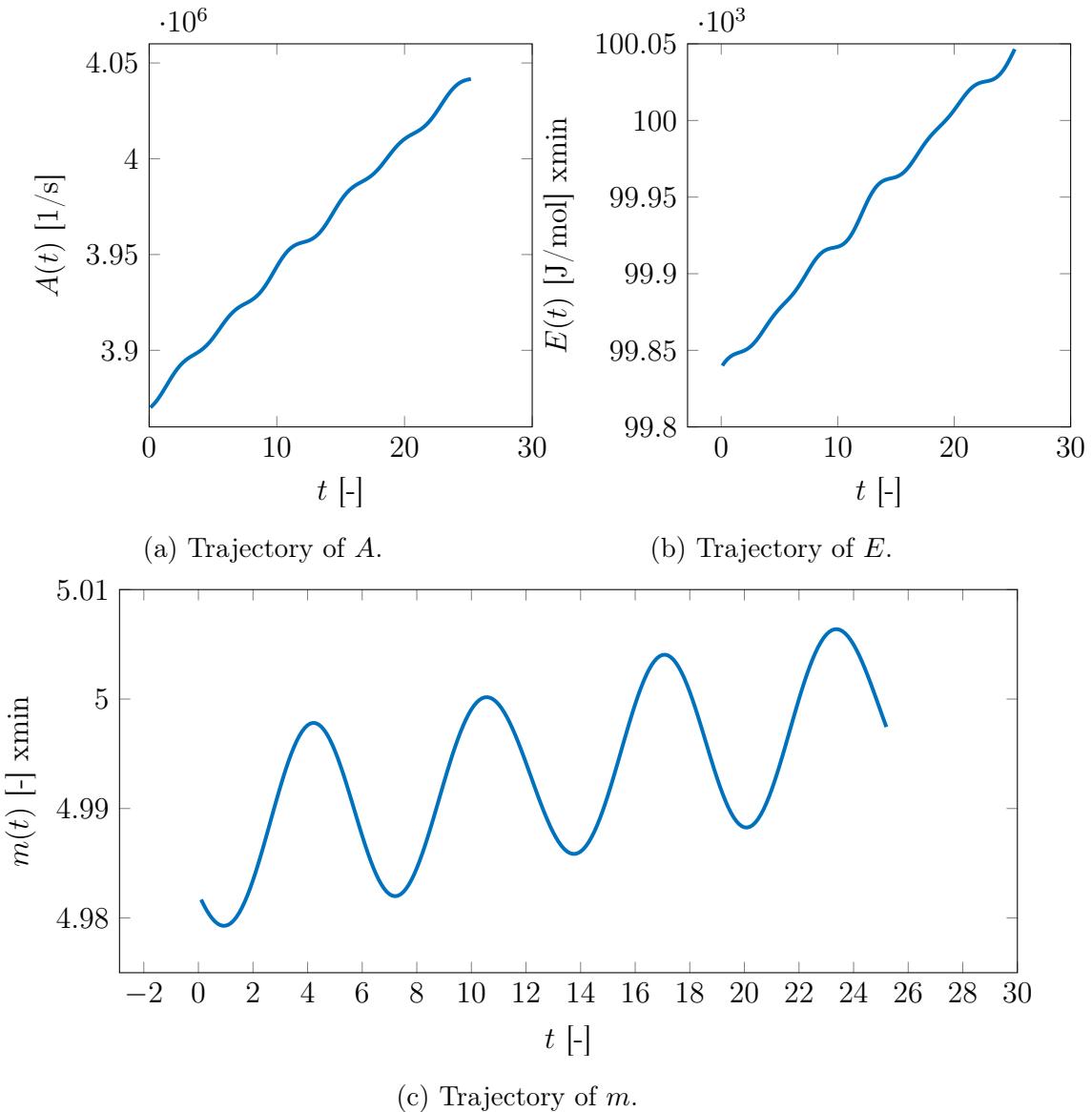


Figure 5.5: Trajectory of the parameters generated by simulating the Hamiltonian dynamics for the simplified pyrolysis problem, with $\Delta t = 0.1$, $\tau = 25$, when the matrix $[M'] = \text{diag}([M])$.

It can be observed that the trajectories, in particular for the parameters A and E , explore the parameter space less efficiently. This makes sense since these two parameters possess the highest uncertainty, as assessed from $[\hat{\Sigma}_q]$. Furthermore, an interesting property of the matrix $[M]$ can be deduced. The faster exploration of the parameter space in the first case means that smaller values of τ can be used to generate candidates located further away from the parameters sampled at the previous iteration, hence needing less evaluation of the gradient, which is beneficial in terms of computational cost.

Let us now consider that $\tau = 4$, and apply the Hamiltonian Monte Carlo algorithm for 1000 iterations. The resulting Markov chains for the three parameters are shown in Figure 5.6 for the two definitions of the mass matrix $[M]$.

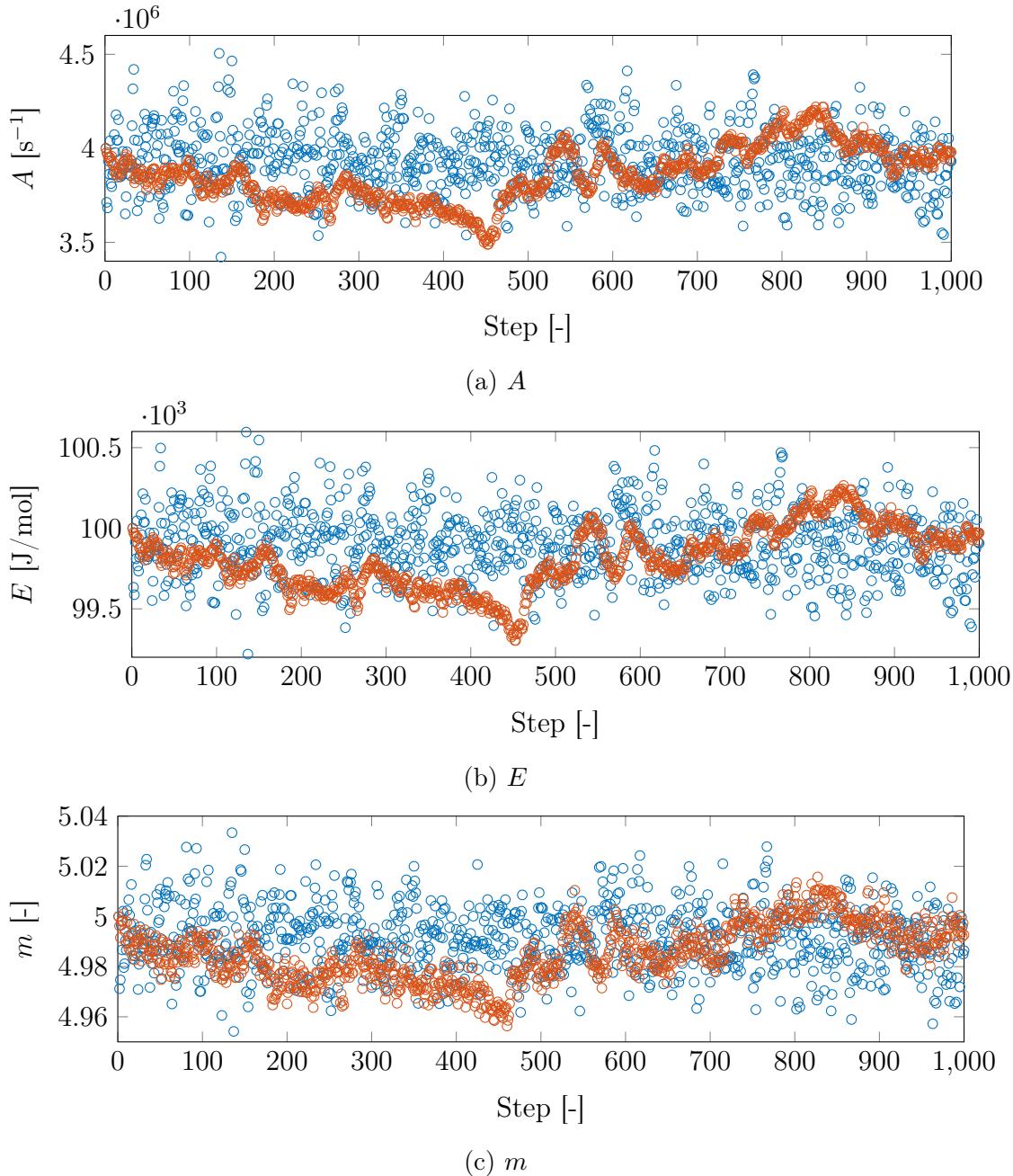


Figure 5.6: Markov chain generated by the Hamiltonian Monte Carlo algorithm. The blue circles represents samples drawn when $[M] = [\hat{\Sigma}_q]^{-1}$, while the red dots are samples generated by the diagonal approximation.

When the estimation of the covariance matrix is taken to be the covariance matrix of the linearised inverse problem, the final time $\tau = 4$ is usually sufficient. However, as it can be seen in Figure 5.5, the value of the parameters A and E barely change in the case where the diagonal matrix is used. This can be seen in the Markov chains from Figure 5.6. Indeed, the Markov chain associated with the first choice of the mass matrix seems to explore the parameter space relatively well, whereas it is clearly not the case for the second choice of $[M]$. Moreover, the Markov chain in the latter case, in particular for the first two parameters, possess the shape of a Markov chain associated with a random walk.

In Figure 5.7, we compare the first 250 samples obtained using the aforementioned matrices

with plots of the posterior density for the linearised inverse problem. When $[M] = [\hat{\Sigma}_q]^{-1}$, the samples are exploring the whole posterior density, while in the second case, the samples only covers a small part of the posterior density.

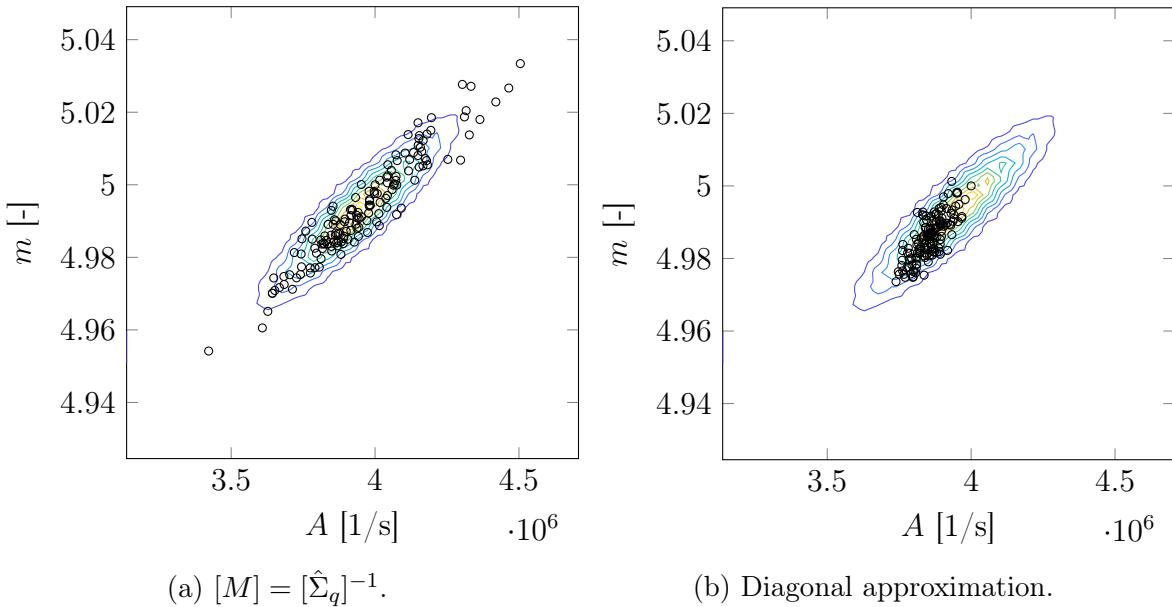


Figure 5.7: First 250 samples generated by the Hamiltonian Monte Carlo method for the two choices of the matrix $[M]$, compared to the contour of the linearised problem's bivariate posterior density for the parameters A and m .

Scatter plots of the Markov chains can be seen in Figure 5.8. These suggest that the parameters A and E are strongly correlated, and that these parameters are also correlated with m . This is, for instance, similar to what was found in [11] using the Metropolis-Hastings algorithm.

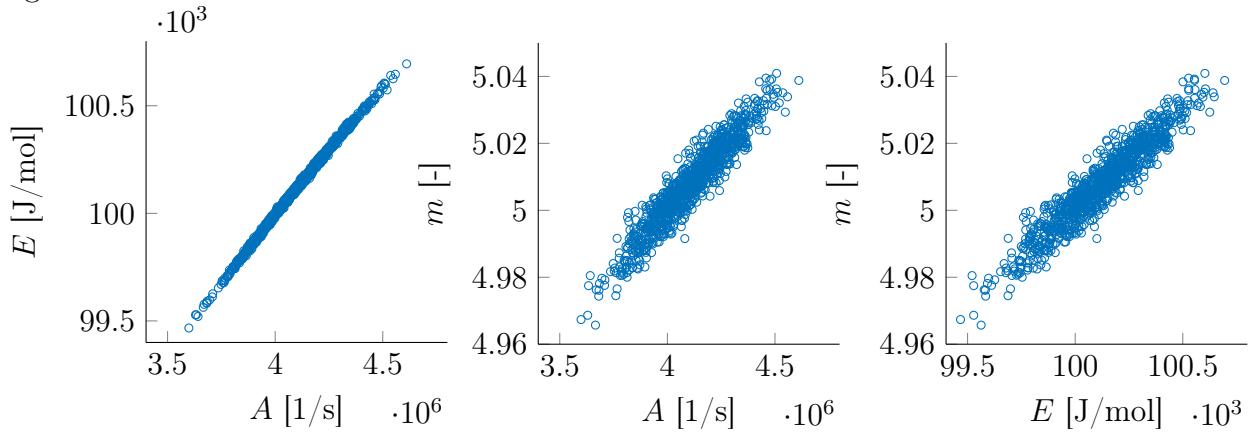


Figure 5.8: Scatter plot of the Markov chains in the original variables \mathbf{q} .

Since the matrix $[M]^{-1} = [L][L]^T$ in Hamilton's equations can be interpreted as a result of a change of variables $\mathbf{Q} = [L]^{-1}\mathbf{q}$, it is also interesting to look at these scatter plots in the transformed variables. Figure 5.9 shows the different scatter plots of the Markov Chains in transformed variables when $[M]$ is an estimate of the covariance matrix.

The effect of the change of correlations is clearly noticeable as the scatter plots' shapes are reminiscent of distributions with uncorrelated variables. In this case, the effect of the transformation of variables can be seen as an alternative to the re-parametrisation of the pyrolysis model that was introduced in [11] in order to improve the efficiency of the Metropolis-Hastings algorithm.

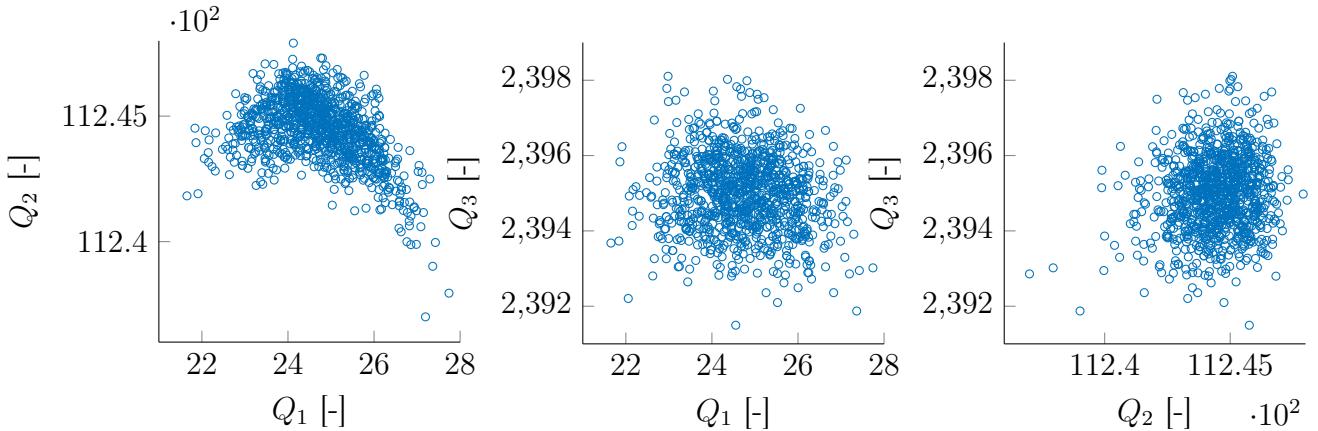


Figure 5.9: Scatter plot of the Markov chains in the transformed variables \mathbf{Q} when the matrix $[M]$ is an estimation of the covariance matrix of the posterior.

Figure 5.10 shows the transformed variables when the matrix is chosen to be a diagonal matrix. The resulting Markov Chains are simply those of Figure 5.8 in which the parameters have been normalised.

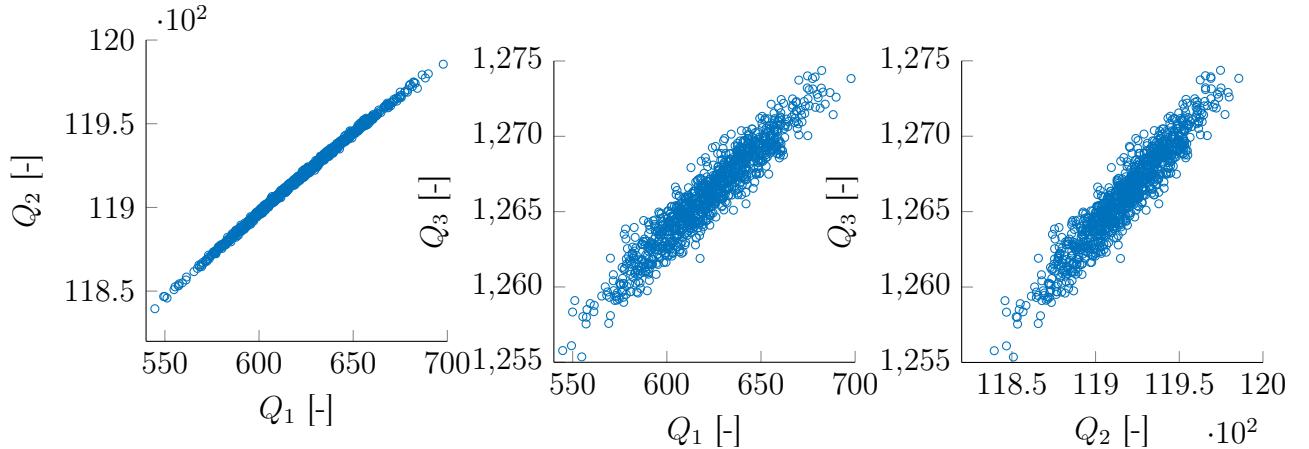


Figure 5.10: Scatter plot of the Markov chains in the transformed variables \mathbf{Q} when the matrix $[M]$ is a diagonal matrix.

Furthermore, Figure 5.6 suggests that the Markov chain produced by diagonal matrix has not converged. This is, for example, shown in Figure 5.11, where the convergence of the posterior means $\bar{\mathbf{A}}$ and $\bar{\mathbf{m}}$ obtained with the two matrices is assessed.

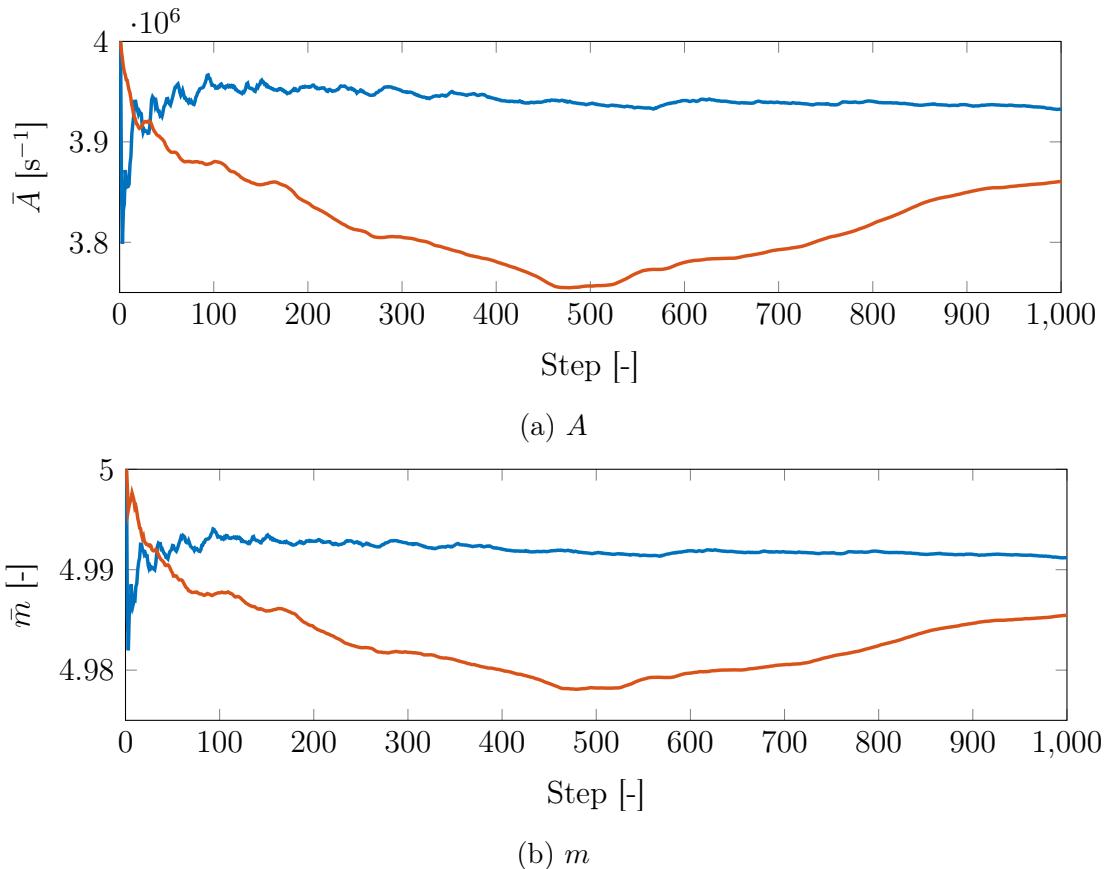


Figure 5.11: Posterior mean of A and m . The blue line is obtained when the matrix $[M]$ is the inverse of the covariance matrix of the posterior density, while the red line is obtained with a diagonal approximation.

The posterior mean of both parameters converges rather quickly for the Markov chain obtained with the estimation of the covariance matrix of the posterior density, while it has yet to converge after 1000 iterations when we assume a diagonal matrix.

It was assumed earlier that the method used for the time integration also had an influence on the performance of the algorithm. First, using the explicit Euler integration method with the final time $\tau = 4$ and a time step $\Delta t = 0.1$ produces an acceptance rate close to 0 percent for 1000 iterations. This suggests that this trajectory is too long, such that, in the phase plane, it has already drifted away from the hypersurface of constant Hamiltonian. Reducing the trajectory length by fixing $\tau = 0.5$ improves the acceptance rate, which is now close to fifty percent. However, by taking smaller trajectories, the elements of the Markov chain are now highly correlated. Hence, in practice, the explicit Euler integrator should not be used. However, the symplectic Euler method, which is given by

$$p_i^{k+1} = p_i^k - \Delta t \frac{\partial U}{\partial q_i}(\mathbf{q}^k), \quad (5.43)$$

$$q_i^{k+1} = q_i^k + \Delta t \frac{\partial T}{\partial p_i}(\mathbf{p}^{k+1}), \quad (5.44)$$

allows to obtain, with $\tau = 4$, a time step $\Delta t = 0.1$ and the same number of iterations, an acceptance rate close to 95%. Over several runs of the algorithm with different values of the time step, it was observed, as suggested by Neal [22], that the acceptance rate increases

upon reducing the time step (Figure 5.12). This follows from the fact that the Euler method is of order one, while the Störmer-Verlet integrator is of order two. The lack of data for the Symplectic Euler method for $\Delta t = 0.5$ results from the fact that the solution diverged. Moreover, the improvement in the acceptance rate for smaller time step is characterised by an increased computational cost. As a rule of thumb, for $\tau = 0.4$ and 1000 iterations, the time step $\Delta t = 0.01$ requires in the order of 400000 evaluations of the model and its gradient.

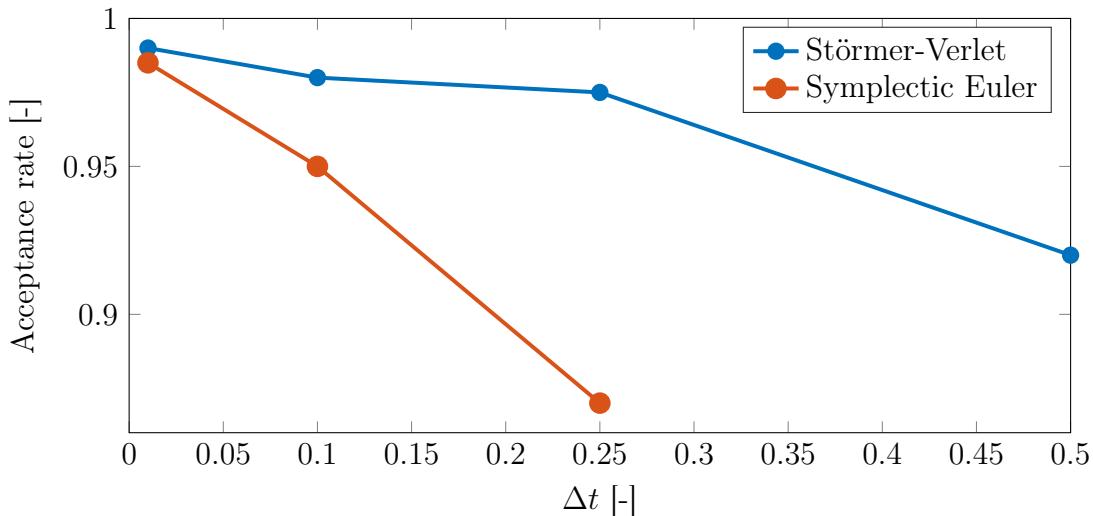


Figure 5.12: Convergence of the acceptance rate with the time step Δt .

In conclusion, despite being applied to a simplified model, this analysis allowed us to investigate the effect of some of the degrees of freedom in the Hamiltonian Monte Carlo, namely the matrix $[M]$, the time step Δt , the final time τ and the time integration method. In particular, the introduction of the matrix $[M]$ as an estimation of the covariance matrix of the posterior in Hamilton's equation greatly improve the performance of the method, and the use of the Störmer-Verlet scheme should be preferred to the other integration methods.

5.3 Markov Chain Monte Carlo based on Itô stochastic differential equations

5.3.1 Introduction

A method slightly different from the Hamiltonian Monte Carlo consists in generating a Markov Chain by simulating the trajectory of an Itô stochastic differential equation [3], given by

$$d\mathbf{q} = \mathbf{p} dt, \quad (5.45)$$

$$d\mathbf{p} = -\frac{\partial U}{\partial \mathbf{q}} dt - \frac{1}{2} f_0 \mathbf{p} dt + \sqrt{f_0} d\mathbf{W}, \quad (5.46)$$

with initial conditions $\mathbf{q}(0) = \mathbf{q}_0$, $\mathbf{p}(0) = \mathbf{p}_0$. Note that this system of equations is in fact a dissipative Hamiltonian system with a stochastic forcing term. The constant f_0 is a positive parameter and $\{\mathbf{W}(t), t \geq 0\}$ is a normalised Wiener process.

Under some regularity conditions exposed in [3] *i.e.*,

1. $U(\mathbf{q})$ is infinitely differentiable,
2. $\inf_{\|\mathbf{q}\| > r} U(\mathbf{q}) \rightarrow +\infty$ as $r \rightarrow +\infty$,
3. a valid pdf can be associated to $U(\mathbf{q})$, *i.e.*, $\int_{\mathbb{R}^d} \exp(-U(\mathbf{q})) d\mathbf{q} < +\infty$.

the solution to this system of stochastic differential equations is a stochastic process $\{(\mathbf{q}(t), \mathbf{p}(t)), t \geq 0\}$ whose long term solution is characterised by the Gibbs distribution

$$\rho(\mathbf{q}, \mathbf{p}) = k \exp \left(-U(\mathbf{q}) - \frac{\|\mathbf{p}\|^2}{2} \right), \quad (5.47)$$

where k is a normalisation constant. This results follows from a similar argument as the one made in the case of the canonical ensemble. Indeed, under these conditions, it can be shown [3] that the evolution of the probability density $\rho(\mathbf{q}, \mathbf{p})$ follows the stationary Fokker-Plank equation [27] to which the Gibbs density appears as a solution.

In particular, the solution is ergodic to the Gibbs distribution, which means that probabilistic averages with respect to $P(\mathbf{q}, \mathbf{p})$ can be estimated as time averages over trajectories, *i.e.*, for a square-integrable function $f(\mathbf{q}, \mathbf{p})$ with respect to the Gibbs density,

$$\int_{\mathbb{R}^{2d}} f(\mathbf{q}, \mathbf{p}) \rho(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} = \lim_{\tau \rightarrow +\infty} \int_0^\tau f(\mathbf{q}(t), \mathbf{p}(t)) dt. \quad (5.48)$$

As a result, similarly to the Hamiltonian Monte Carlo method, the potential in the stochastic differential equation is set to $U(\mathbf{q}) = -\log(\pi(\mathbf{q} | \mathbf{d}^{\text{obs}}))$, such that the random variables $\mathbf{q}(t)$ are samples drawn from the posterior density.

Furthermore, as in the Hamiltonian Monte Carlo method, Arnst [3] suggested the introduction of a symmetric positive definite matrix $[M]$, whose role is similar to the mass matrix in the previous method in the sense that it provides a change of scales and correlations aimed to improve the efficiency of the method. The system of stochastic differential equations thus becomes

$$d\mathbf{q} = [M]^{-1} \mathbf{p} dt, \quad (5.49)$$

$$d\mathbf{p} = -\frac{\partial U}{\partial \mathbf{q}} dt - \frac{1}{2} f_0 \mathbf{p} dt + \sqrt{f_0} [L]^{-T} d\mathbf{W}, \quad (5.50)$$

where $[L]$ is a lower triangular matrix such that $[M]^{-1} = [L][L]^T$. The choice of the matrix $[M]$ is thus the same as the one used in the Hamiltonian Monte Carlo method. Under this transformation of the stochastic differential equations, the Gibbs density becomes

$$\rho(\mathbf{q}, \mathbf{p}) = k \exp \left(-U(\mathbf{q}) - \frac{1}{2} \mathbf{p}^T [M]^{-1} \mathbf{p} \right). \quad (5.51)$$

5.3.2 Numerical solution of the Stochastic differential equation

It is generally not possible to find an exact solution to the system of equations in Eqs.(5.49-5.50) and numerical methods for the time integration must be used. Following the same

approach as in the previous section, two time integration methods will be considered. First, the explicit Euler method is given by

$$\mathbf{q}^{k+1} = \mathbf{q}^k + \Delta t[M]^{-1}\mathbf{p}^k \quad (5.52)$$

$$\mathbf{p}^{k+1} = (1 - \frac{\Delta t f_0}{2})\mathbf{p}^k - \Delta t \frac{\partial U(\mathbf{q}^k)}{\mathbf{q}} + \Delta t \sqrt{f_0}[L]^{-T} \Delta \mathbf{W}^{k+1}, \quad (5.53)$$

where the sequence $\{\Delta \mathbf{W}^{k+1}, k = 0, 1, \dots, \nu - 1\}$ are mutually independent multivariate Gaussian random variables with mean zero and covariance matrix $\Delta t[\mathbb{I}]$, where $[\mathbb{I}]$ is the $d \times d$ identity matrix. Referring to [17], the Störmer-Verlet method can also be applied to the stochastic differential equations, *i.e.*

$$\mathbf{q}^{k+1/2} = \mathbf{q}^k + \frac{\Delta t}{2}[M]^{-1}\mathbf{p}^k, \quad (5.54)$$

$$\mathbf{p}^{k+1} = \frac{1 - \beta}{1 + \beta}\mathbf{p}^k - \frac{\Delta t}{1 + \beta} \frac{\partial U(\mathbf{q}^{k+1/2})}{\partial \mathbf{q}} + \frac{\sqrt{f_0}[L]^{-T}}{1 + \beta} \Delta \mathbf{W}^{k+1}, \quad (5.55)$$

$$\mathbf{q}^{k+1} = \mathbf{q}^{k+1/2} + \frac{\Delta t}{2}[M]^{-1}\mathbf{p}^{k+1}, \quad (5.56)$$

where $\beta = f_0 \Delta t / 4$.

5.3.3 Application to a quadratic potential

As an illustration, let us apply these numerical methods to solve the SDEs (with $f_0 = 1$) in the case of a one dimensional harmonic oscillator considered in the previous section when solving Hamilton's equation, *i.e* with the potential $U(q) = \frac{1}{2}q^2$ and kinetic energy $E(p) = \frac{1}{2}p^2$, with initial conditions $q(0) = p(0) = 1$ and the Euler and Stormer-Verlet schemes are used with $\Delta t = 0.1$ and $\tau = 100000$. In accordance with Eq.(5.47), the long time behaviour of the solution $(q(t), p(t))$ should be that of a bivariate Gaussian distribution with zero mean and identity covariance matrix.

Figures 5.13 to 5.15 shows the results of the numerical solution. In particular, Figure 5.13 shows the stochastic processes obtained for both numerical schemes, which confirms that the marginal pdfs associated to $q(t)$ and $p(t)$ are indeed both centred at $q = p = 0$ and are similar to samples obtained from a normal distribution. Moreover, Figure 5.14 is a representation of the samples in the $p - q$ plan, and the circular shape observed is indeed reminiscent of that of a bivariate normal distribution.

However, it appears that the explicit Euler scheme does not exactly results in a normal probability density function for both q and p . Indeed, as shown in Figure 5.15, which shows the result of a kernel density estimation based on the samples generated by solving the stochastic differential equation, the Euler scheme seems to result in a pdf that is not a normal pdf with unit covariance, whereas the results obtained with the Stormer-Verlet scheme fits the graph of a normal pdf. This behaviour stem from different reasons. First, the Euler scheme is of order one, while the Stormer-Verlet scheme is of order two, and accuracy errors could explain some of the disparities. Moreover, as mentioned previously, the explicit Euler method is not suited for the time intregration of Hamiltonian dynamics, while symplectic schemes such as the Stormer-Verlet schemes are designed specifically to tackle these problems.

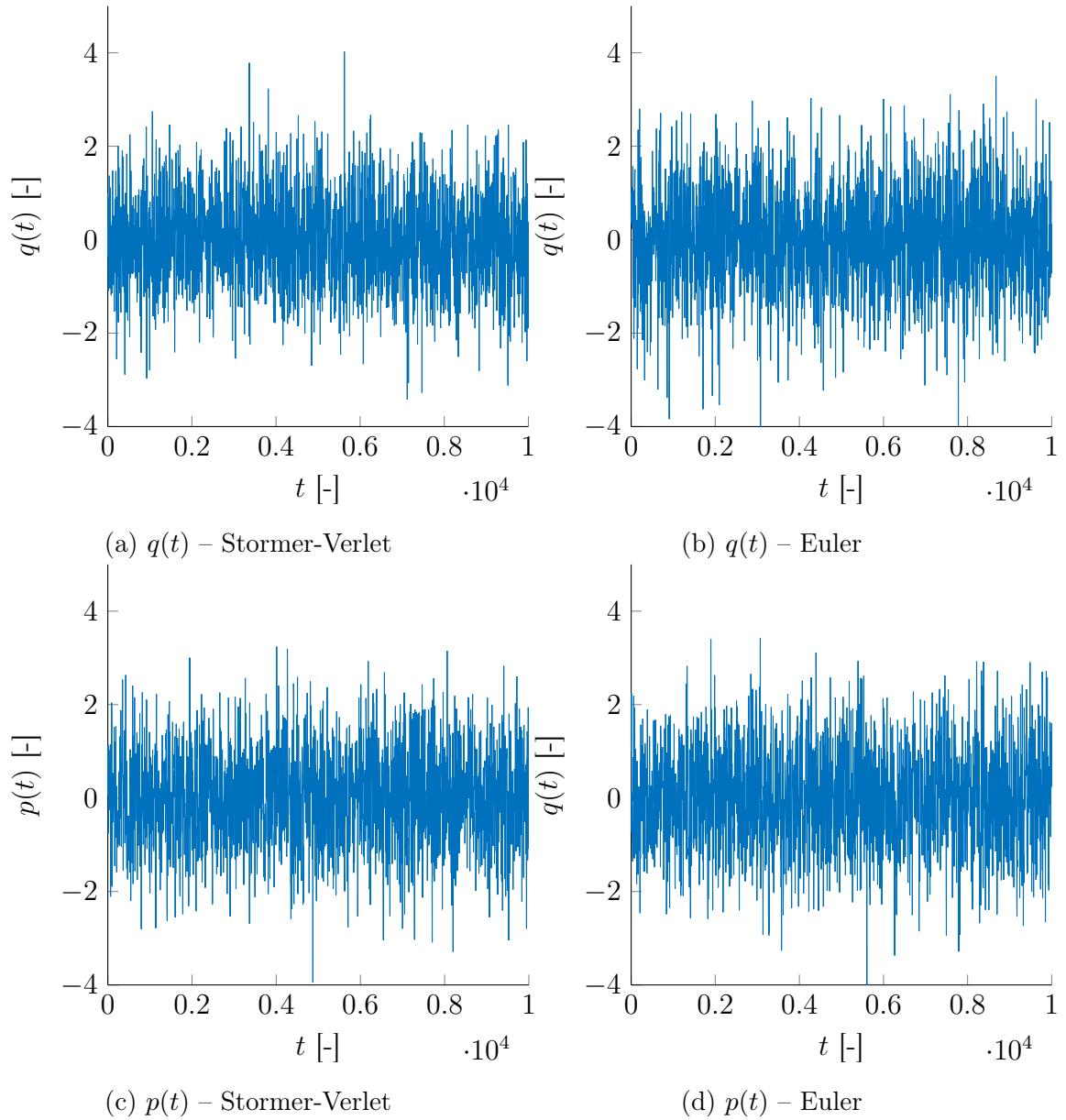


Figure 5.13: Solution of the Itô-SDE for an harmonic oscillator.

Furthermore, in the case of a quadratic potential, the system of equations in 5.50 can readily be written as

$$dq = pdt, \quad (5.57)$$

$$dp = -qdt - \frac{1}{2}f_0pd़t + \sqrt{f_0}dW, \quad (5.58)$$

with initial conditions $q(0) = q_0$, $p(0) = p_0$. This system, rewritten in matrix form, reads

$$dz = \begin{bmatrix} 0 & 1 \\ -1 & -\frac{1}{2}f_0 \end{bmatrix} z dt + \begin{pmatrix} 0 \\ \sqrt{f_0} \end{pmatrix} dW, \quad z = \begin{pmatrix} q \\ p \end{pmatrix}, \quad (5.59)$$

from which an exact solution can be obtained [3]. It is given by

$$z(t) = \exp \left(- \begin{bmatrix} 0 & -1 \\ 1 & \frac{1}{2}f_0 \end{bmatrix} t \right) z_0 + \int_0^t \exp \left(- \begin{bmatrix} 0 & -1 \\ 1 & \frac{1}{2}f_0 \end{bmatrix} (t-s) \right) \begin{pmatrix} 0 \\ \sqrt{f_0} \end{pmatrix} dW. \quad (5.60)$$

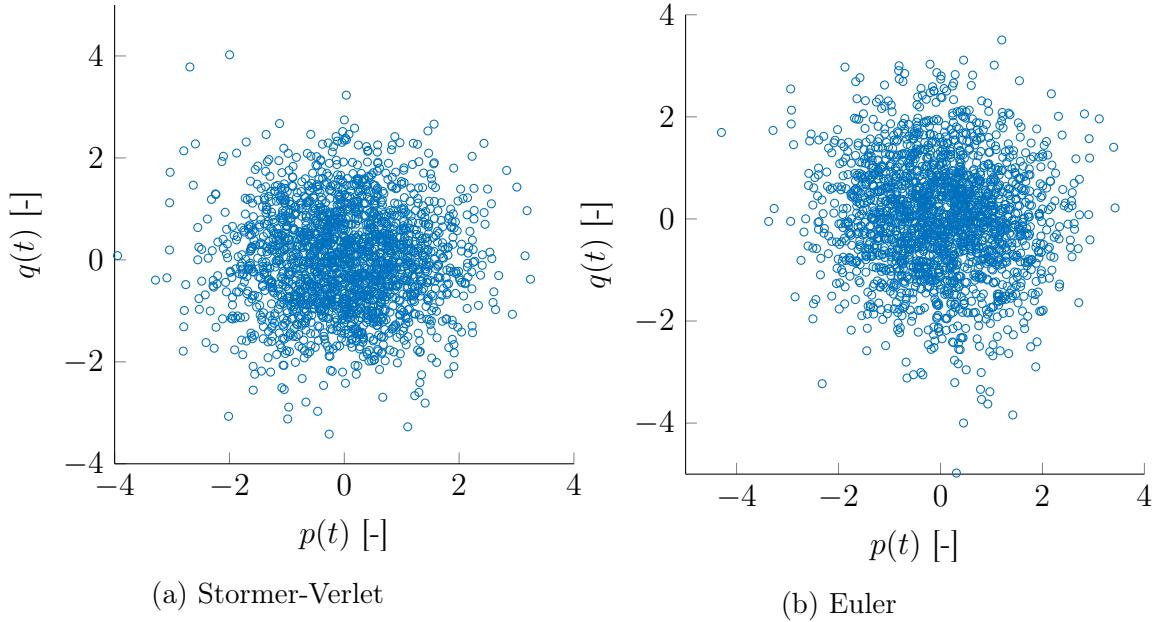


Figure 5.14: Solution of the Itô-SDE for an harmonic oscillator in the $p - q$ plane.

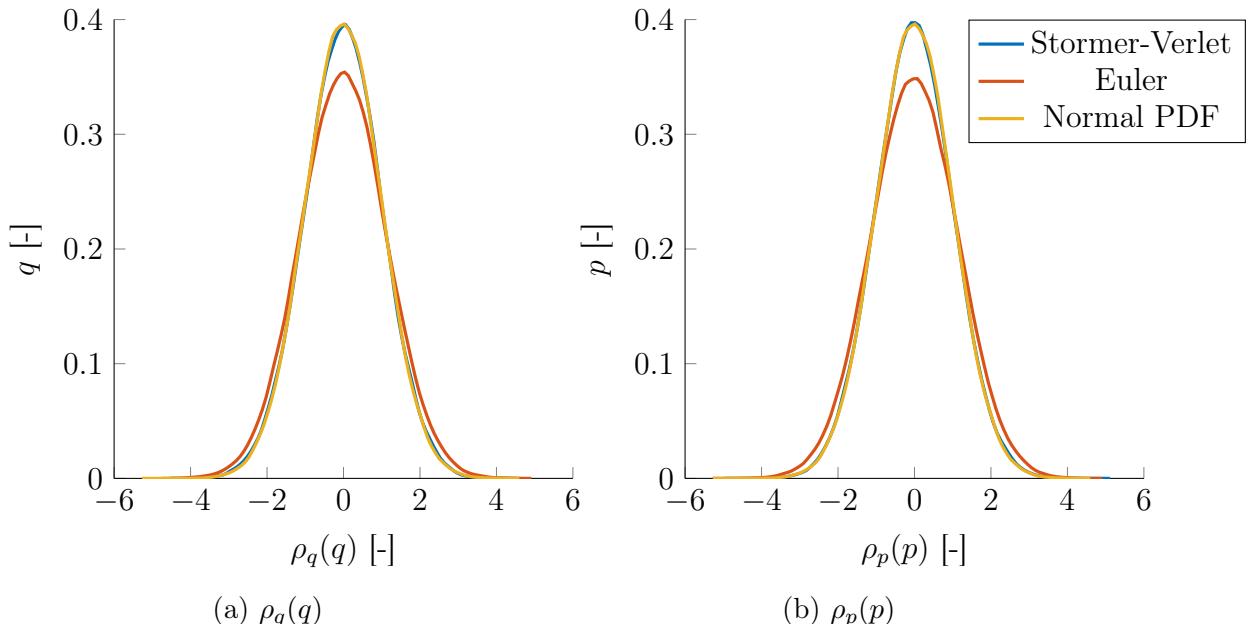


Figure 5.15: Kernel density estimations of the marginal pdf for q and p for an harmonic oscillator, compared to a normal density with zero mean and unit standard deviation.

The eigenvalues of the matrix appearing in the exponential allows the quantification of different regimes depending on the value of the parameter f_0 . Indeed, the eigenvalues are given by

$$\lambda_{1,2} = \frac{f_0/2 \pm \sqrt{(f_0/2)^2 - 4}}{2}.$$

When $f_0 > 4$, the eigenvalues are real and the system is said to be overdamped, while the case $f_0 < 4$ gives rise to complex conjugate eigenvalues, and the system is said to be underdamped. Finally, in the case $f_0 = 4$, the matrix possesses only one eigenvalue, and the system is said

to be critically damped. In particular, in the analysis of dynamical systems, the complex eigenvalues appearing in the underdamped case are usually associated with oscillations in the solution. It also appears that the critically damped case $f_0 = 4$ is, in the case of a quadratic potential, the fastest to reach the equilibrium solution of the system [3]. As an illustration of the influence of the parameter f_0 , let us consider two extreme cases: first, let us chose, for example, $f_0 = 0.01$ while every other parameters remains the same as the previous example. The numerical results obtained with both the Euler and Stormer-Verlet schemes are shown in Figures 5.16 and 5.17.

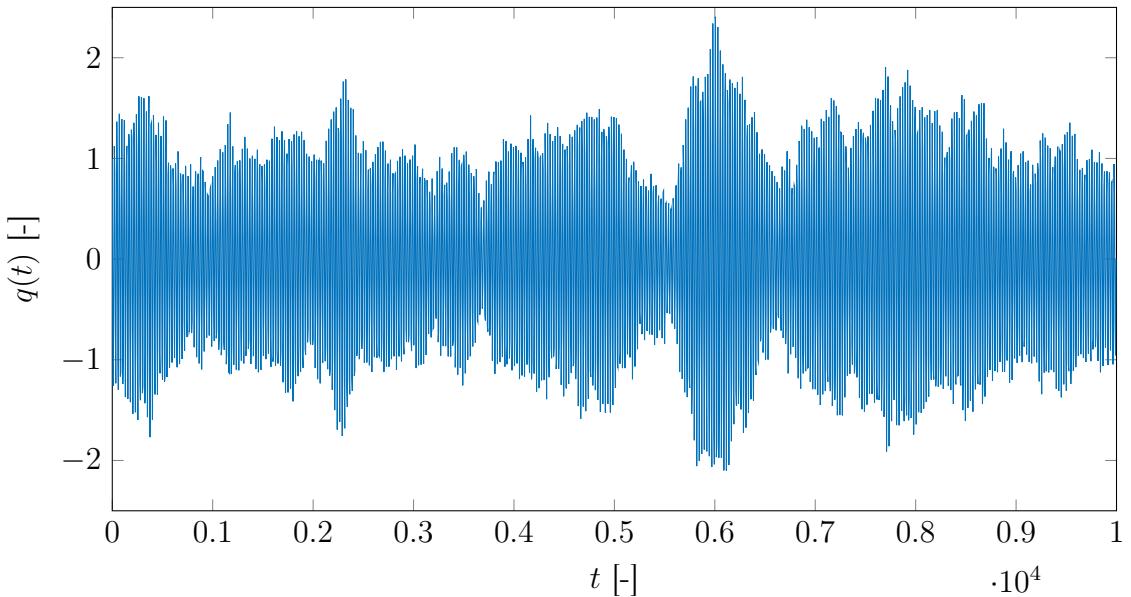


Figure 5.16: Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 0.01$ and $\Delta t = 0.1$ using the Stormer-Verlet scheme.

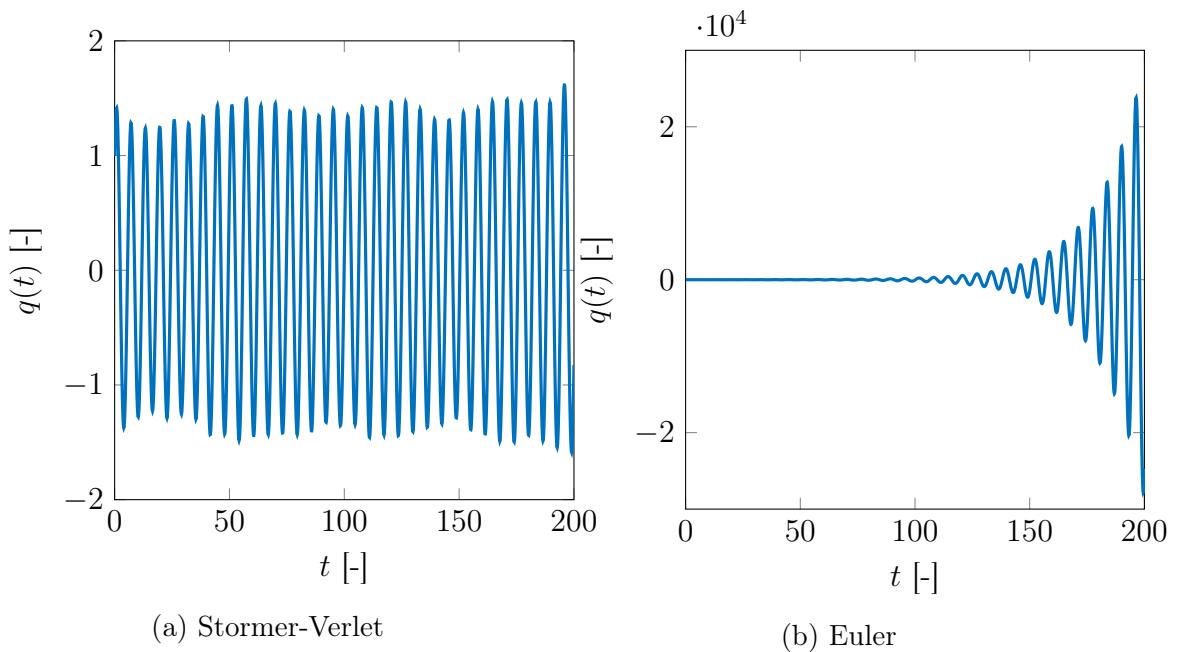


Figure 5.17: Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 0.01$ and $\Delta t = 0.1$ - zoom on the first iterations.

It can be seen that, when choosing small values of f_0 , the solution of the Ito SDE is characterised by the presence of oscillations around $q = 0$. Moreover, it is also interesting to notice that, for $\Delta t = 0.1$, the Euler scheme leads to an unstable solution depending on the value of the parameter f_0 . Figure 5.18 shows the kernel density estimation based on the samples generated by the Ito SDE. It is indeed a probability density function centred in zero, but it is not a normal distribution. Too low values of f_0 should hence be avoided in order to produce meaningful results.

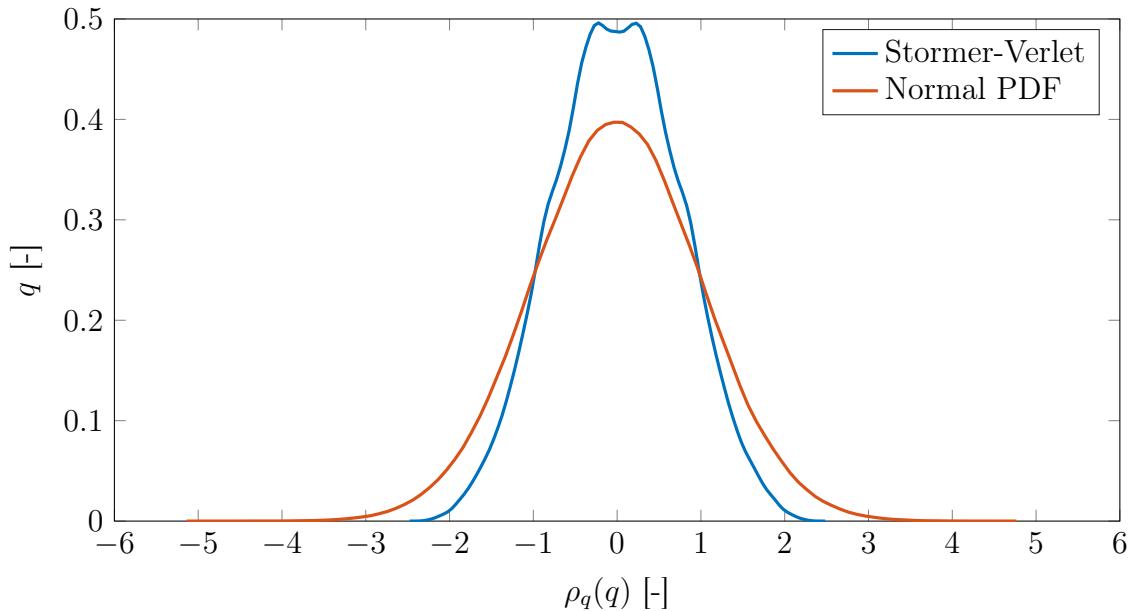


Figure 5.18: Kernel density estimation based on the solution of the Ito SDE with $\Delta t = 0.1$ and $f_0 = 0.01$.

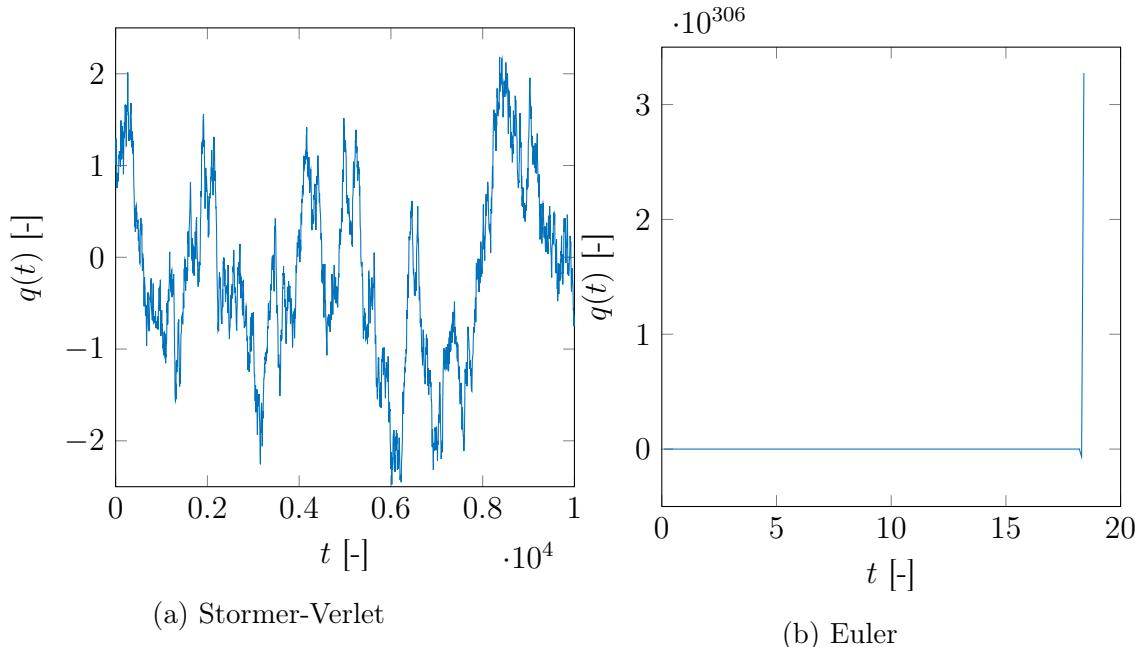


Figure 5.19: Numerical solution of the Ito SDE for an harmonic oscillator with $f_0 = 1000$ and $\Delta t = 0.1$.

Similarly, Figure 5.19 shows the solution of the Ito SDE when $f_0 = 1000$. In this configuration, no oscillation is observed, but the solution converges slowly to a normal pdf. Indeed, this can be observed when performing a kernel density estimation of the solution (Figure 5.20). The resulting pdf follows the curve of a normal distribution, but some irregularities are observed. The explicit Euler method is again unstable. In conclusion, values of f_0 that are too high should also be avoided. It should however be noted that this study on the parameter f_0 was performed in a simplified case and may differ for more complex potential. However, since no exact solution exists for such potentials, these results allow us to gain a first insight on the choice of the parameter f_0 .

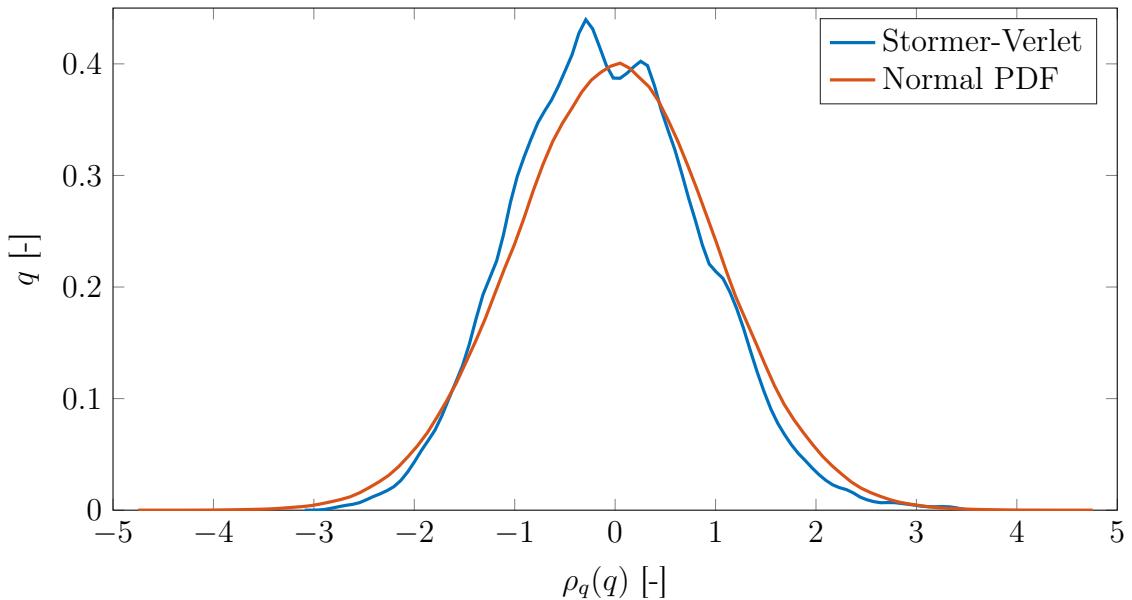


Figure 5.20: Kernel density estimation based on the solution of the Ito SDE with $\Delta t = 0.1$ and $f_0 = 1000$.

When the potential is chosen as the negative of the logarithm of the posterior distribution, the sequence $\{(\mathbf{q}^k, \mathbf{p}^k), k = 0, 1, \dots, \nu - 1\}$ generated by the simulation of the stochastic differential equations can be interpreted as a Markov Chain where the elements $\{\mathbf{q}^k, k = 0, 1, \dots, \nu - 1\}$ are samples drawn from the posterior distribution. Owing to the ergodicity property of the solution, probabilistic averages can be computed using this Markov chain. For a function $f(\mathbf{q}, \mathbf{p})$ that is square integrable with respect to the Gibbs density, the ergodicity property implies that

$$\int_{\mathbb{R}^{2d}} f(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p} \approx \frac{1}{\nu} \sum_{k=0}^{\nu-1} f(\mathbf{q}^k, \mathbf{p}^k), \quad (5.61)$$

and probabilistic averages for the posterior distribution are given by

$$\int_{\mathbb{R}^d} f(\mathbf{q}) k_q \exp(-U(\mathbf{q})) d\mathbf{q} \approx \frac{1}{\nu} \sum_{k=0}^{\nu-1} f(\mathbf{q}^k). \quad (5.62)$$

5.3.4 Application to the simplified pyrolysis problem

The Itô-SDE method can be applied to inference problem related to the simplified pyrolysis problem studied in the previous section. We first compare the effect of the matrix $[M]$ on the

Markov chain generated by the method. To do so, we set the free parameter $f_0 = 4$ as in [3], $\Delta t = 0.1$ and $\tau = 1000$.

Figure 5.21 shows the three different Markov chains obtained with the matrix $[M]$ set to the inverse of the covariance matrix of the posterior estimated by the linearisation of the inverse problem around the maximum likelihood estimator in the first case, and to a diagonal matrix $[M'] = \text{diag}([M])$ in the second case. The first choice makes the trajectories explore the parameter space in a efficient manner for each parameters, while in the second case, the values of m , which is the least uncertain parameter, are explored more efficiently than those with more uncertainty.

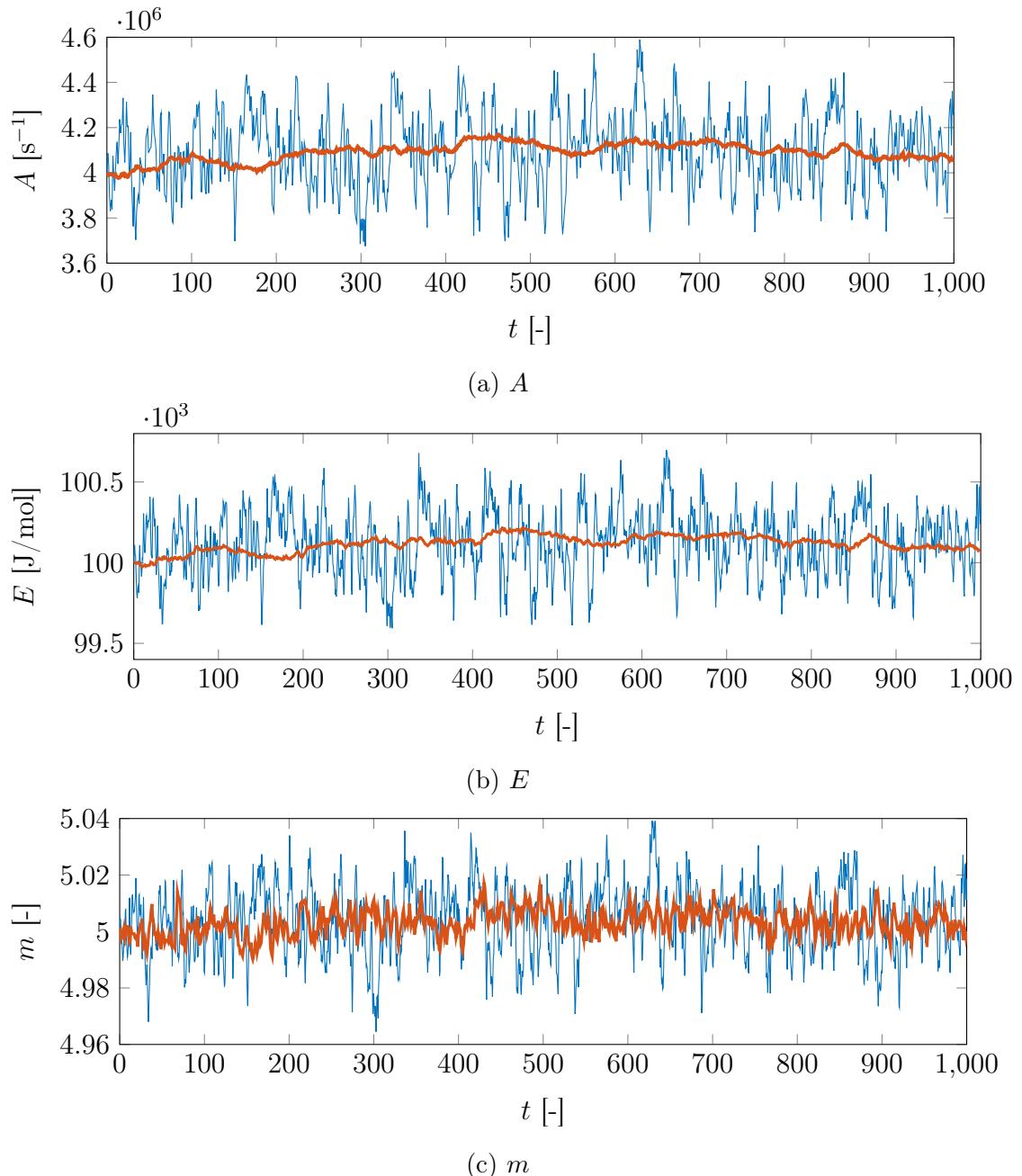


Figure 5.21: Markov chain generated by the Itô-SDE method. The blue line represents the Markov chain when the matrix $[M]$ is the inverse of the variance of posterior, while the red line is the Markov chain generated with a diagonal approximation.

Another way of visualising the effect of the matrix $[M]$ is by looking at a surface plot of a bivariate posterior density and assess how the trajectory evolve in time. In Figure 5.22, when the bivariate posterior density of the linearised inverse problem for A and m is considered, the trajectory simulated when $[M]$ is an estimation of the posterior's covariance matrix explores correctly the posterior density, whereas in the second case, most of the evolution of the trajectory consists in moving along the direction of m .

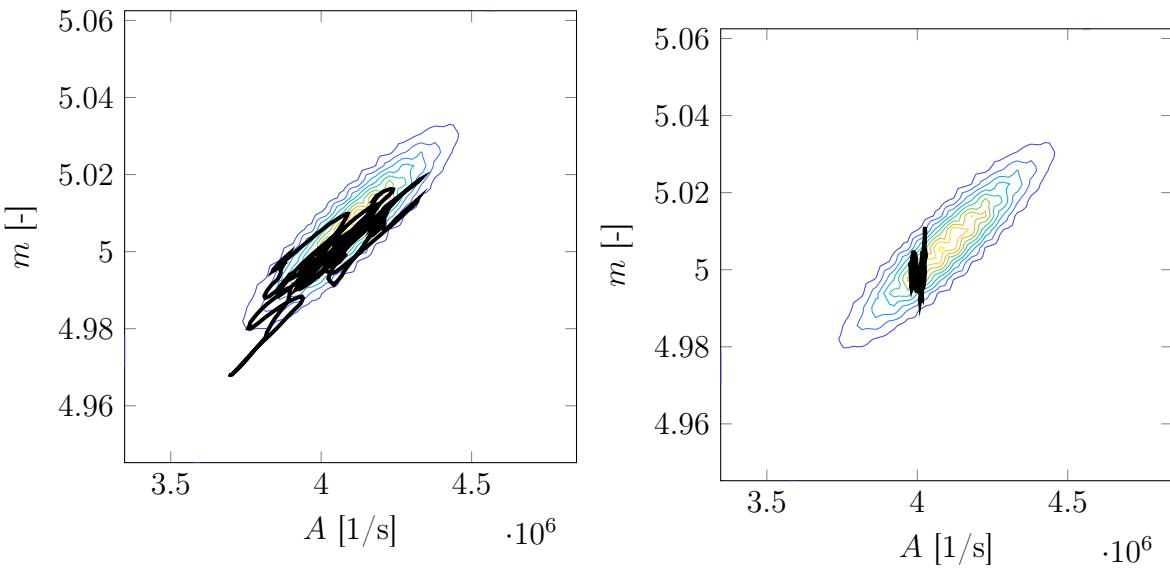


Figure 5.22: 500 first steps of the trajectory generated by the Itô-SDE method for the two choices of the matrix $[M]$, compared to the contour of the linearised problem's bivariate posterior density for the parameters A and m .

Finally, we consider in Figure 5.23 the convergence of the posterior mean of A for $\tau = 1000$ and $\Delta t = 0.1$ as a way to assess the efficiency of the method. A convergence to the true posterior mean seems to be start appearing as the time increases.

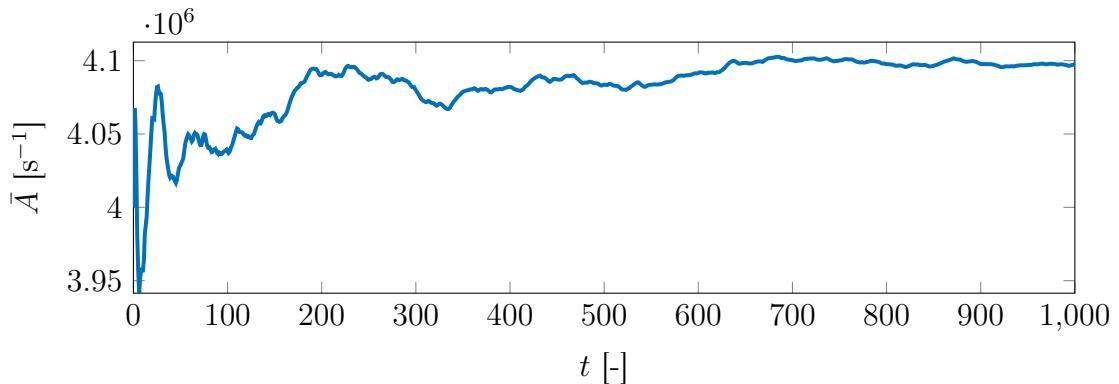


Figure 5.23: Convergence of the posterior mean of A .

5.3.5 Comparison with the Hamiltonian Monte Carlo method

The Hamiltonian Monte Carlo and Itô-SDE methods are based on a similar principle: samples are drawn from the posterior probability density function by simulating the dynamics of an Hamiltonian system characterized by a system of (stochastic) differential equations. However, there is a fundamental difference, based on the way these methods handle the solution from these equations. Hamiltonian Monte Carlo uses the final time step of a reversible trajectory to generate a Metropolis-Hastings proposal, whereas the Itô-SDE method directly uses the irreversible trajectory as samples from the posterior density.

If we consider, for example, the convergence of the mean for both methods in Figures 5.11a and 5.23, we may think, at first glance, that the Hamiltonian Monte Carlo method is more efficient. However, it should be noticed that 1000 iterations of the HMC algorithm, for $\tau = 4$ and $\Delta t = 0.1$ require the evaluation of 40000 time steps of Hamiltonian trajectories. If we now consider the Itô-SDE method where $\tau = 4000$ and only consider every 40 time steps in the solution, we obtain the graph in Figure 5.24 for the convergence of the posterior mean of A . For a similar computational time, it can be concluded that the Itô-SDE method and Hamiltonian Monte Carlo actually produce relatively similar results.

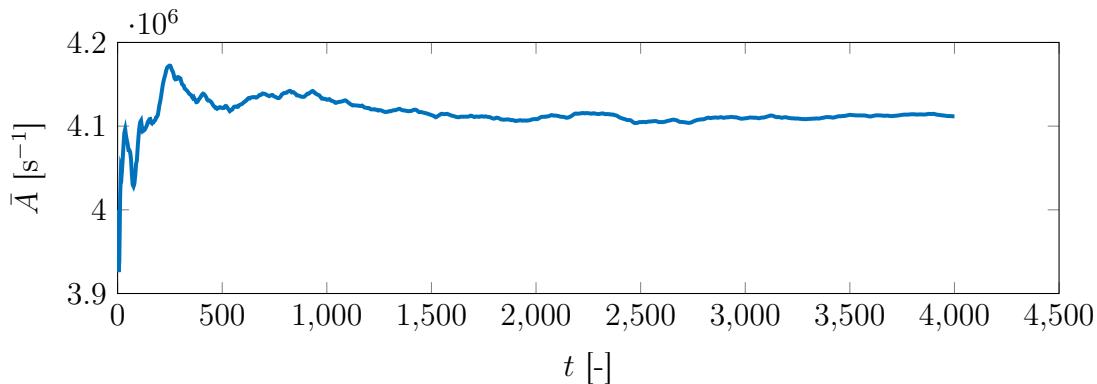


Figure 5.24: Itô-SDE method: Posterior mean of A obtained with $\tau = 4000$ and $\Delta t = 0.1$.

A major difference that was observed when applying the methods to the simplified pyrolysis problem is that, when the matrix $[M]$ is not chosen efficiently, the methods behave differently. Hamiltonian Monte Carlo produces a Markov chain resembling a random walk due to the inefficient transition resulting from the Hamiltonian dynamics, while the trajectory in the Itô-SDE method only evolves in the direction of the least uncertain parameters.

While the cause of these issues is the same, the different behaviours observed could potentially be explained by how the randomisation in the methods is handled. In HMC, it is due to the resampling of the momentum at each iteration, so that will still explore the posterior distribution, albeit slowly, whereas it is directly incorporated in the Itô-SDE as the stochastic forcing term.

5.4 Influence of the measurement noises

Up to now, the methods were introduced by solving an inverse problem for the simplified pyrolysis model using unrealistically low measurement noises. Using the simplified pyrolysis problem, new synthetic data were generated and a Gaussian noise of standard deviation $\sigma = 5 \times 10^{-4}$ was added, as seen in Figure 5.25, where perturbed data are compared with the gas production curve used to generate them. The goal is to assess how the algorithms compare when the noise in the data becomes non negligible. Indeed, in real pyrolysis experiments, the measurement noises on the gas production are usually important, as seen in [7].

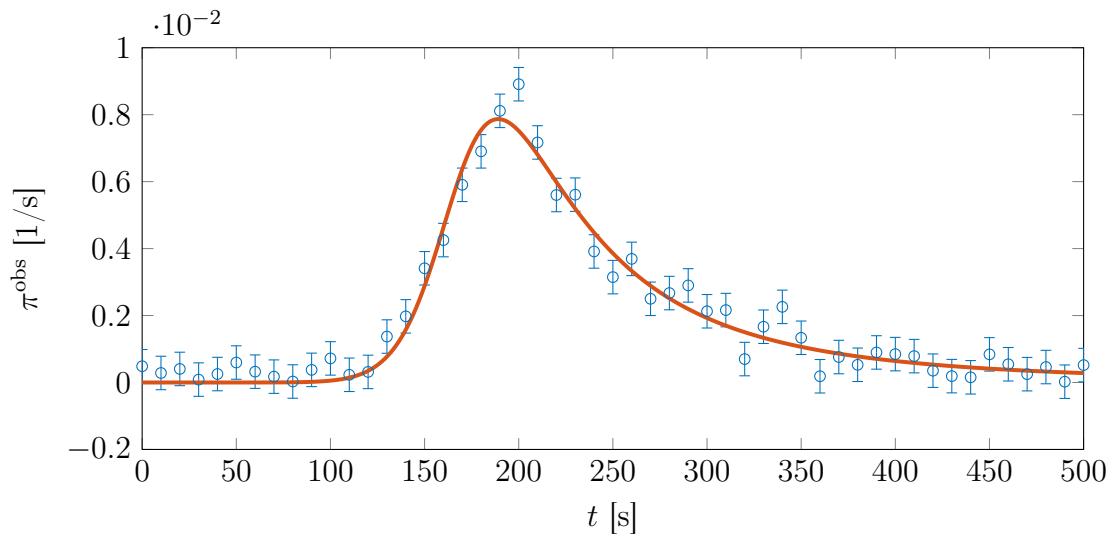


Figure 5.25: Synthetic data with Gaussian noise of standard deviation $\sigma = 5 \times 10^{-4}$ used for the Bayesian inference. The parameters used to generate the data are the same as in Figure 4.1.

A preliminary estimation of the posterior density with an uninformative prior was performed by using the linearised inverse problem around the parameters used to generate the data. The result obtained suggests that this type of prior density results in a posterior density that allows negative values of the parameters, which is not physically acceptable. For example, the kernel density estimation in Figure 5.26 shows that it is indeed the case for the parameter A . When the data becomes noisier, the uncertainty on the parameters increase. In the pyrolysis problem, we expect that the parameters become more correlated based on the sensitivity analysis carried out earlier in this work.

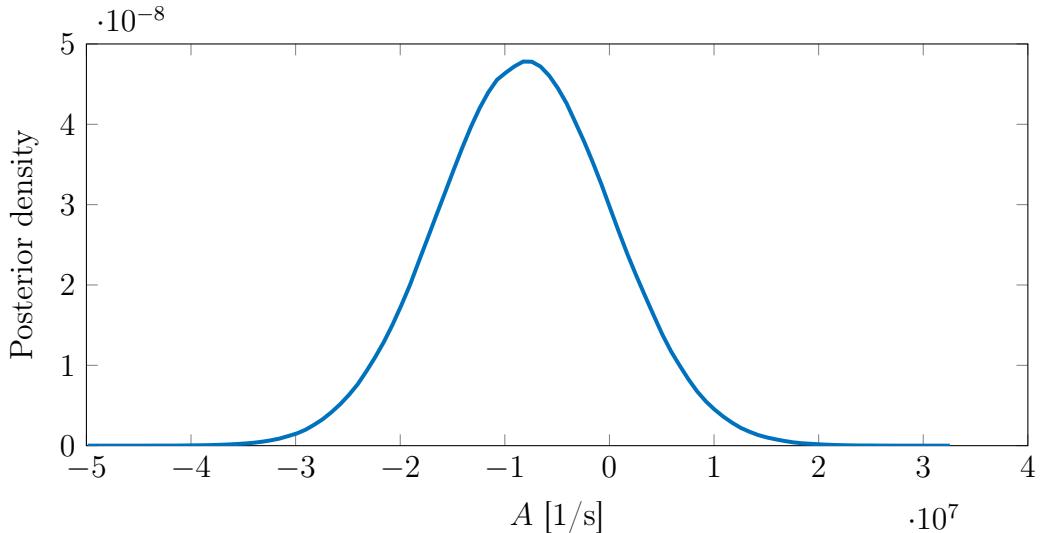


Figure 5.26: Kernel density estimate of the linearised posterior density for A .

We therefore need to specify the prior density to ensure positive values for the parameters. Following the developments in [3], the following prior density is considered:

$$\pi_0(\mathbf{q}) = k_p \frac{1}{A} \mathbb{I}_{\mathbb{R}_0^+}(A) \frac{1}{E} \mathbb{I}_{\mathbb{R}_0^+}(E) \frac{1}{m} \mathbb{I}_{\mathbb{R}_0^+}(m). \quad (5.63)$$

This type of prior density defines a potential $U(\mathbf{q})$ which is not continuous and therefore not differentiable. However, It can be made regular by introducing the change of variables $\tilde{\mathbf{q}} = \mathbf{f}(\mathbf{q})$ such that

$$\tilde{A} = \log(A), \quad \tilde{E} = \log(E), \quad \tilde{m} = \log(m). \quad (5.64)$$

Under this change of variables, the posterior density reads

$$\pi(\tilde{\mathbf{q}} | \mathbf{d}^{\text{obs}}) = k_p \frac{1}{A} \frac{1}{E} \frac{1}{m} \pi(\mathbf{d}^{\text{obs}} | \mathbf{f}^{-1}(\tilde{\mathbf{q}})) \det \left(\frac{\partial \mathbf{f}^{-1}(\tilde{\mathbf{q}})}{\partial \tilde{\mathbf{q}}} \right) \quad (5.65)$$

$$= k_p \pi(\mathbf{d}^{\text{obs}} | \mathbf{f}^{-1}(\tilde{\mathbf{q}})). \quad (5.66)$$

Furthermore, the gradients appearing in the dynamical equations for each methods, as well as the sensitivity matrix used to linearise the inverse problem must be adapted, *i.e.*

$$\frac{\partial U}{\partial \tilde{q}_j} = \frac{\partial U}{\partial q_i} \frac{\partial q_i}{\partial \tilde{q}_j}, \quad \frac{\partial \pi_i}{\partial \tilde{q}_j} = \frac{\partial \pi_i}{\partial q_\ell} \frac{\partial q_\ell}{\partial \tilde{q}_j}. \quad (5.67)$$

Once the changes of variable is performed, a linearisation of the inverse problem in the new variables is performed, so as to obtain a new estimation of the covariance matrix and assess the effect of the new prior on the posterior density.

In Figure 5.27, the effect of the prior introduced to enforce that the parameters must be positive on the posterior is shown. In this case, the parameter A becomes indeed non-negative, but also that its probability density becomes non-Gaussian even for a linearised problem.

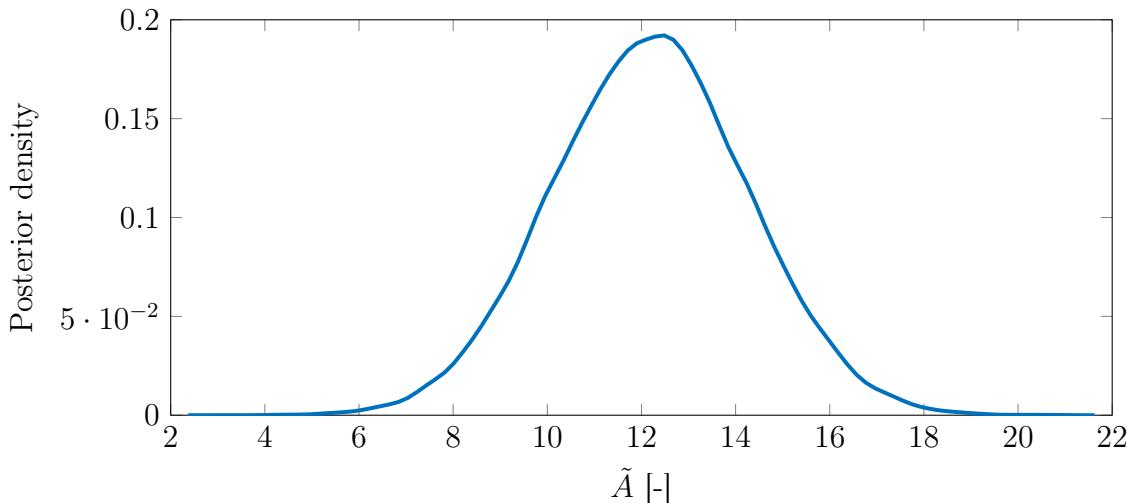


Figure 5.27: Kernel density estimate in the new variables of the posterior obtained with linearisation of the inverse problem.

The inference problem was performed with HMC and Itô-SDE, which were set up with the matrix $[M] = [\hat{\Sigma}]^{-1}$, the Störmer-Verlet integration scheme and a time step of $\Delta t = 0.1$. In the Itô-SDE method, the final time τ was chosen to be

$$\tau = 10000$$

, for the HMC method, $\tau = 1$ and the algorithm was run for 10000 iterations. Moreover, the random walk Metropolis-Hastings algorithm was also used, using a Gaussian proposal. In order not to manually tune the algorithm, the covariance matrix is set to $(2.4/\sqrt{3})^2[\hat{\Sigma}]$ where $[\hat{\Sigma}]$ is the covariance matrix obtained by linearisation of the inverse problem. This choice of proposal is motivated here as it does not require manual tuning and is the most efficient choice for sampling from a multivariate normal distribution with 3 parameters [3, 14]. The Metropolis-Hastings algorithm is then run for 100000 iterations. Note that these settings were chosen so that a similar number of model evaluation for each method is achieved.

Figure 5.28 shows the different Markov chains for \tilde{A} obtained with three algorithms for the arbitrary number of iterations. Each of these Markov Chains seems to follow a similar trajectory in the parameter space. However, we observe an acceptance rate close to 99% for the Hamiltonian Monte Carlo, and about 13.7% with Metropolis Hastings, which suggests that, despite the fact that the two first methods are more computationally expensive for a similar number of steps due to the additional numerical cost associated with the adjoint method, the samples they generate allows a better representation of the posterior density. A better acceptance rate could be obtained with Metropolis-Hastings, but may require manual tuning of the proposal.

Table 5.1 shows the evaluation of the posterior mean and standard deviation using the samples from the Markov Chains, and are compared to independent sample drawn from the posterior density of the linear problem. The results between each algorithms are relatively close. The differences with the linear approximation, in particular for \tilde{A} , suggests that the inverse problem cannot be treated as a linear problem. When the noise in the data was negligible, the uncertainty on the parameters was rather small. However, it can be seen here that the parameters can now range between more values. Moreover, the correlation between these parameter becomes greater, as it can be seen for instance, on scatter plots in Figures

5.29 and 5.30. These scatter plots are consistent with what is exposed in [11]. In particular, the exponential relation between A and E can be observed.

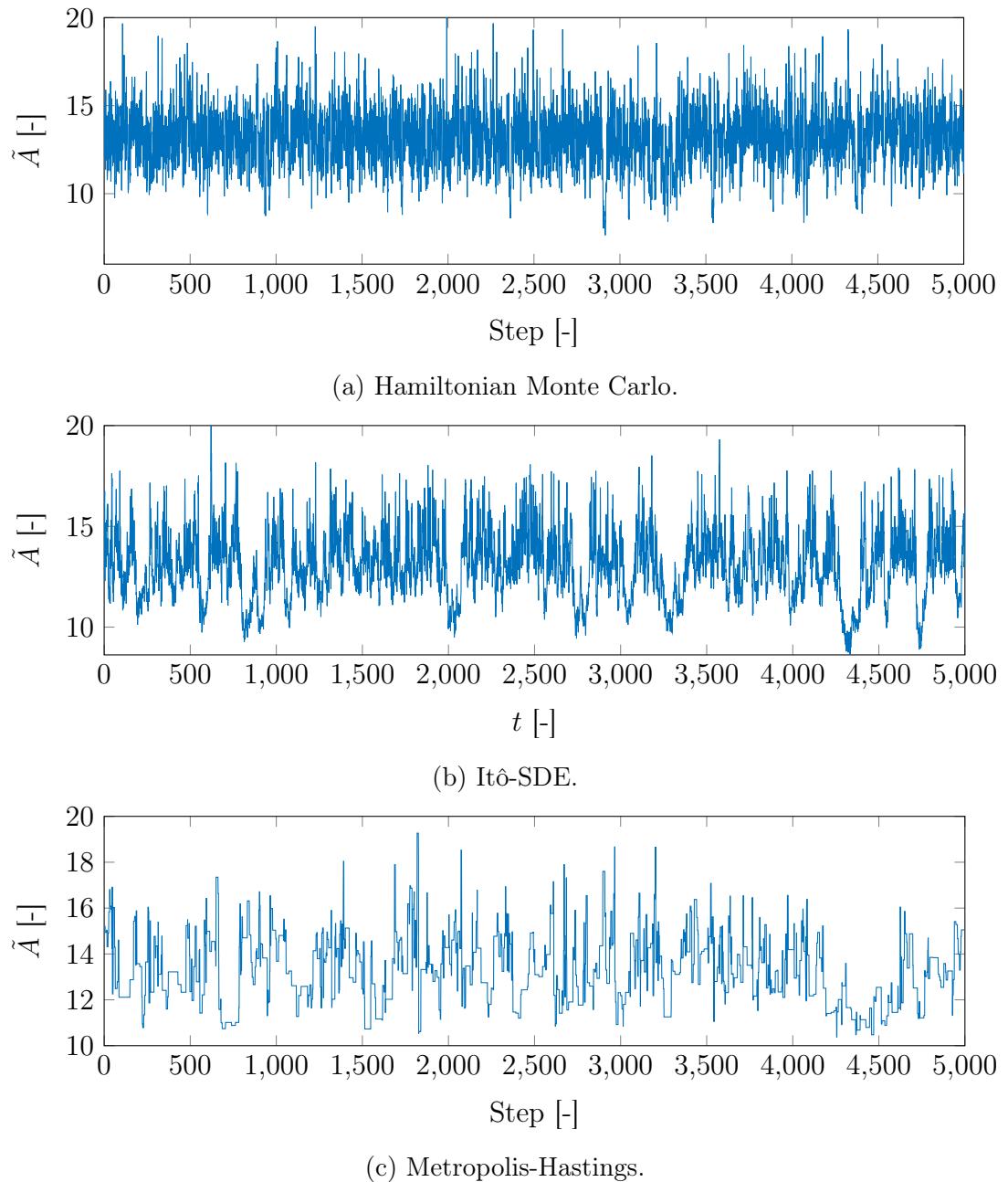


Figure 5.28: Markov chain generated by the three sampling algorithms.

	Hamiltonian Monte Carlo	Ito-SDE	Metropolis Hastings	Linear approx.
Mean \tilde{A}	13.1002	13.1015	13.2893	12.1754
Std \tilde{A}	1.7457	1.7118	1.4815	2.0792
Mean \tilde{E}	11.3996	11.4001	11.4118	11.3647
Std \tilde{E}	0.0977	0.0960	0.0807	0.1036
Mean \tilde{m}	1.4554	1.4548	1.4672	1.4132
Std \tilde{m}	0.1183	0.1165	0.1009	0.1255

Table 5.1: Posterior means obtained with the three methods, as well as from i.i.d samples drawn from the posterior of the linearised problem.

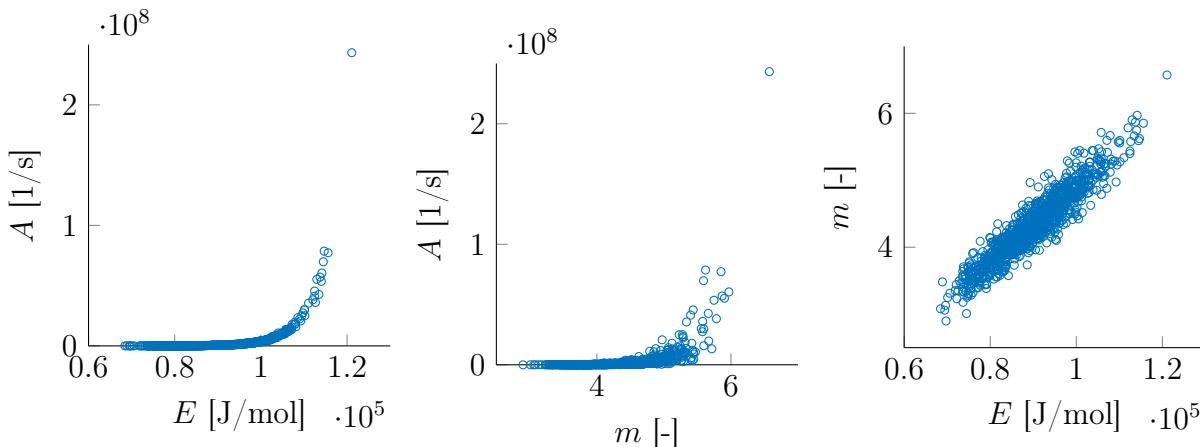


Figure 5.29: Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the Itô-SDE method.

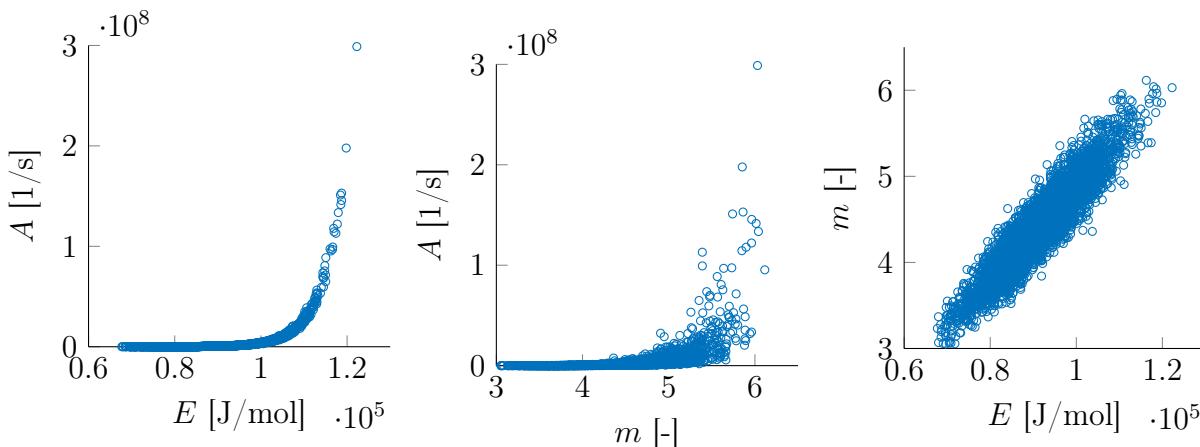


Figure 5.30: Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the HMC method.

Once samples from the posterior density are drawn, they can be propagated to the gas production. This is for instance done in Figures 5.31 and 5.32, where the shaded area is the represents all the possible outcomes obtained when propagating the samples, the solid blue line is the gas production curve used to generate the data. The possible gas production curves resulting from the propagation of the uncertainty in the parameters and the data shows relatively good agreement, and both the Ito-SDE and HMC methods produce similar results.

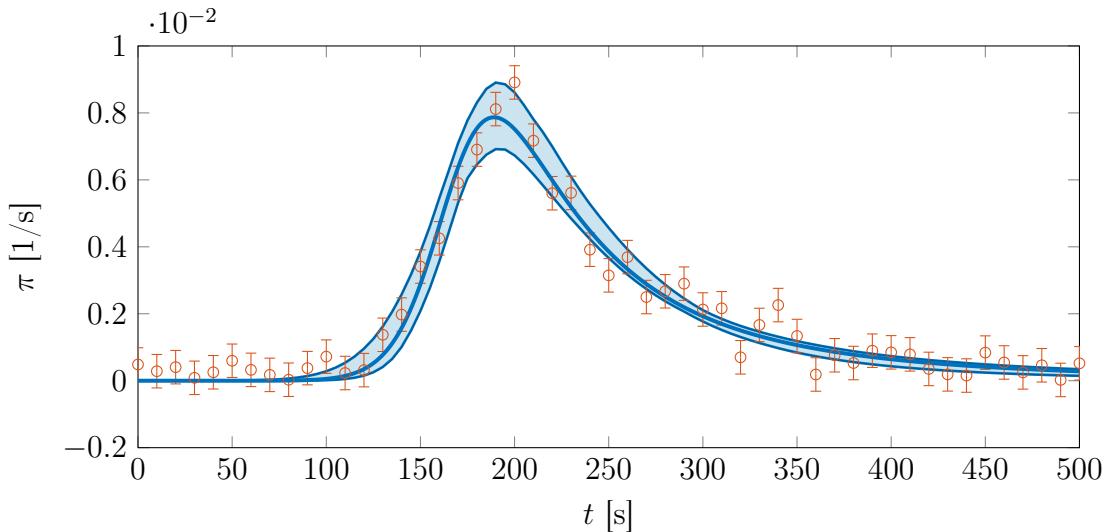


Figure 5.31: Propagation of the samples obtained with the HMC method to the production of one gas.

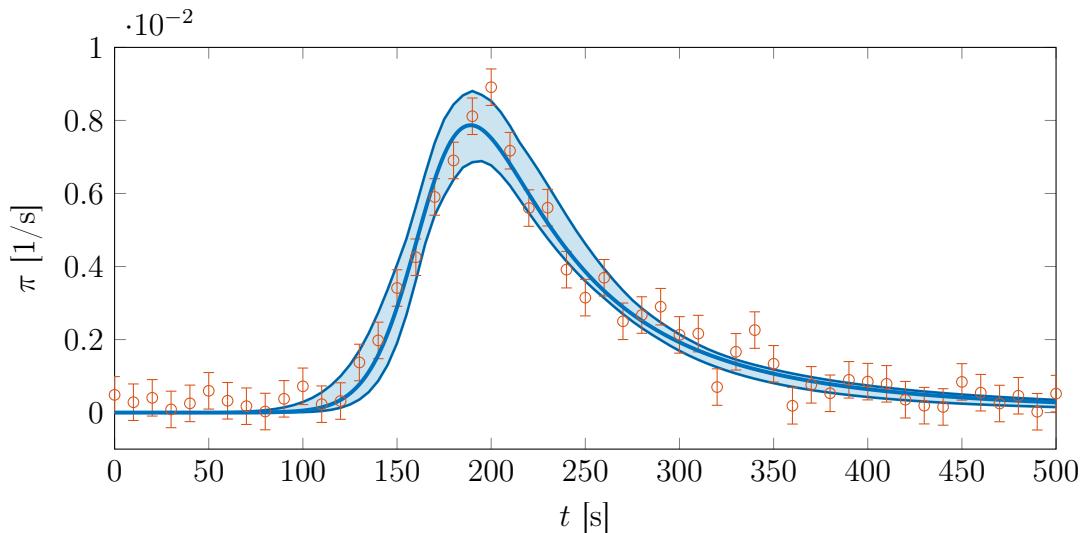
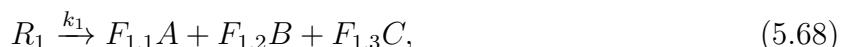


Figure 5.32: Propagation of the samples obtained with the Ito-SDE method to the production of one gas.

5.5 Extension to multiple reactions

Good results have been obtained when considering the simplified pyrolysis problem where a single fictitious solid phase produced only one gas. Let us now consider a slightly more involved problem, where two fictitious solid phases produce three gases. The reaction scheme considered is the following, using similar notations as in [11]:



and the production of each gases is hence given, in vector form, by

$$\boldsymbol{\pi} = \varepsilon_{\text{mv}} \rho_{\text{mv}} [F] \boldsymbol{\xi}, \quad (5.70)$$

where

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_A \\ \pi_B \\ \pi_C \end{pmatrix}, \quad \dot{\boldsymbol{\xi}} = \begin{pmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \end{pmatrix}, \quad [F] = \begin{bmatrix} F_{1,1} & F_{2,1} \\ F_{1,2} & 0 \\ F_{1,3} & F_{2,3} \end{bmatrix}.$$

This time, the inference problem consists in estimating the kinetic parameters, as well as the mass fractions $F_{i,j}$, which amounts to 11 parameters in total. Synthetic data are generated with the parameters from Tables 5.2 and 5.3, to which a Gaussian noise of standard deviation $(\sigma_A, \sigma_B, \sigma_C) = (2.5 \cdot 10^{-4}, 2 \cdot 10^{-4}, 3.5 \cdot 10^{-4})$ is added, and the constant $\varepsilon_{\text{mv}} \rho_{\text{mv}}$ was arbitrarily set to one, as shown in Figure 5.33.

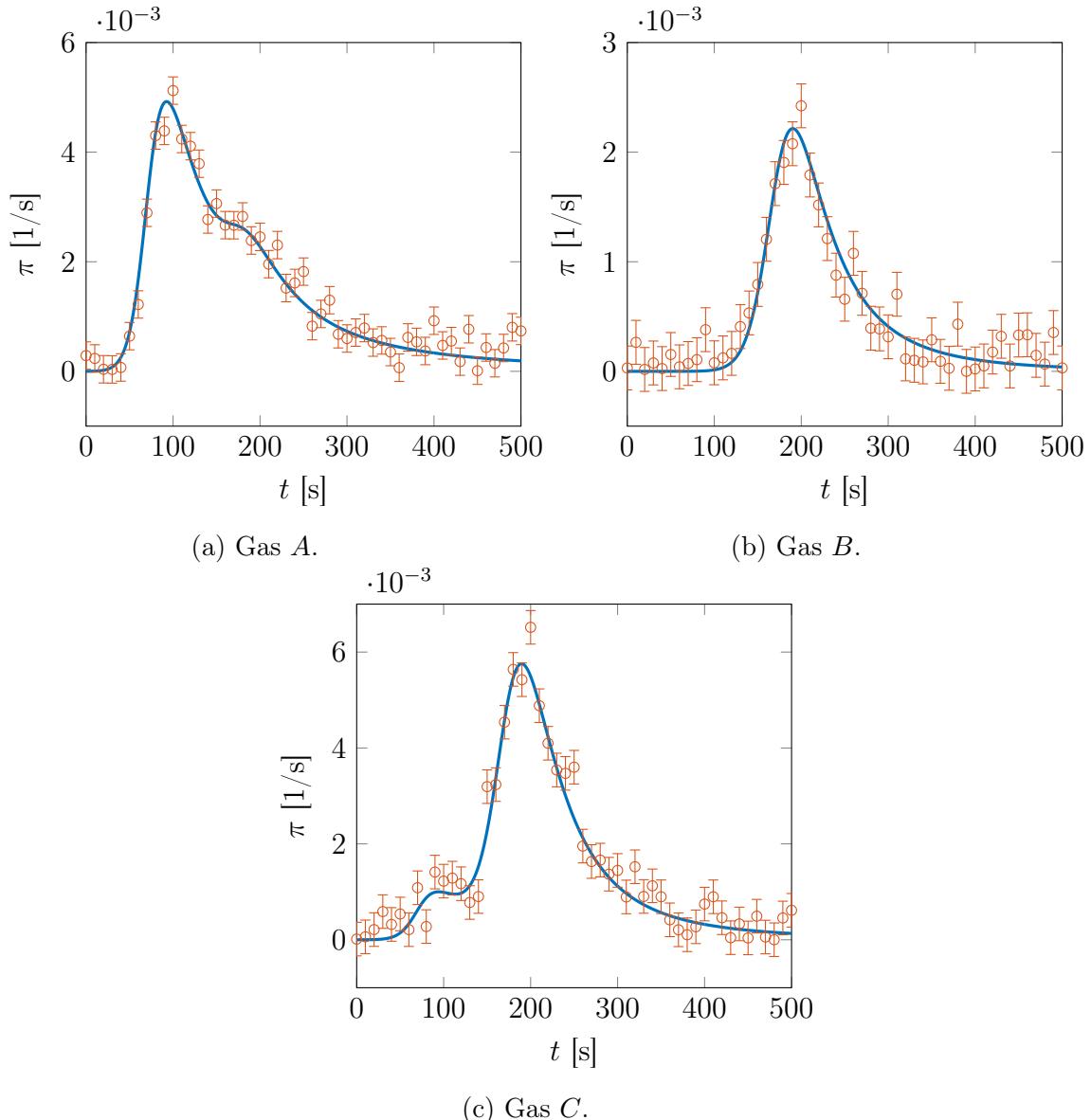


Figure 5.33: Synthetic data used for the production of three gases.

	Reaction 1	Reaction 2
A [1/s]	4000000	1000000
E [J/mol]	100000	66000
m [-]	4	8

Table 5.2: Kinetic parameters used in the two reactions model.

	Reaction 1	Reaction 2
Gas A	0.1	0.7
Gas B	0.24	0
Gas C	0.59	0.14

Table 5.3: Elements of the matrix $[F]$ used in the two reactions model.

As in the previous section, a first study of the problem with the linearised solution provides a first insight on the behaviour of the posterior distribution. Again, the kinetic parameters must be positive, while it is also assumed that the mass fractions $F_{i,j} \in]0, 1[$. Similarly, the posterior distribution from the linearised problem suggests that an uninformative prior allows negative values for the kinetic parameters. However, the parameters $F_{i,j}$ seem to remain between 0 and 1. This suggests that an uninformative prior for these parameters may be sufficient. Similarly as in the previous section, prior densities of the form

$$\pi_0(q_i^{\text{kin}}) = \frac{1}{q_i} \mathbb{I}_{\mathbb{R}_0^+}(q_i), \quad (5.71)$$

are considered. A preliminary run of the inference problem where only the prior for the kinetic parameter is specified shows that, despite the posterior density of the linearised problem providing samples where the parameters $F_{i,j}$ were always taking values in the interval where they are defined, samples generated with the numerical methods were taking values greater than one (Figure 5.34).

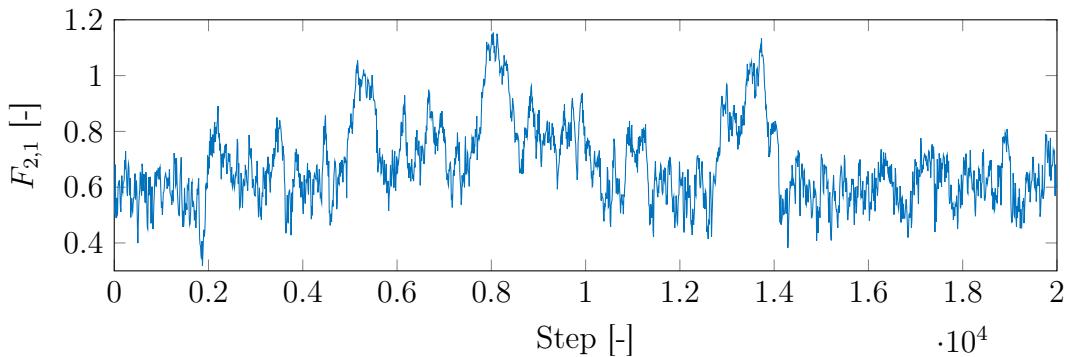


Figure 5.34: Markov chain generated for the parameter $F_{2,1}$ using the Itô-SDE method.

To ensure that the values of the parameters $F_{i,j}$ are indeed in the interval $]0, 1[$, prior densities can be constructed following the same methodology as in [3], *i.e.*

$$\pi_0(F_{i,j}) = \frac{1}{\cos^2((F_{i,j} - \frac{1}{2})\pi)} \mathbb{I}_{]0,1[}(F_{i,j}). \quad (5.72)$$

Under the change of variables $\tilde{\mathbf{q}} = \mathbf{f}(\mathbf{q})$ such that

$$\tilde{q}_i^{\text{kin}} = \log(q_i^{\text{kin}}), \quad \tilde{F}_{i,j} = \tan\left((F_{i,j} - \frac{1}{2})\pi\right), \quad (5.73)$$

the new parameters are now defined in \mathbb{R} . Using the Jacobian of this change of variables, the posterior density becomes

$$\pi(\tilde{\mathbf{q}} | \mathbf{d}^{\text{obs}}) = k_p \pi(\mathbf{d}^{\text{obs}} | \mathbf{f}^{-1}(\tilde{\mathbf{q}})), \quad (5.74)$$

and the derivatives involved in the linearised inverse problem and in the dynamics equations are transformed according to Eq.(5.67). Once the prior is specified, the parameter inference can be performed. The Itô-SDE and HMC methods are considered, with $\Delta t = 0.1$ and $[M] = [\hat{\Sigma}]$. The trajectory length is set to $\tau = 10000$ for the Itô-SDE method and $\tau = 1$ for the HMC algorithm (with $N_{\text{iter}}=10000$).

Because the number of parameters in the model has increased compared to the problem in which only one gas produced by one reaction was studied, it becomes more tedious to properly show individual Markov chains for each parameters. However, we can still establish some comparison with the previous case. For instance, Figure 5.35 shows the scatter plot of the kinetic parameters appearing in the two reactions. These scatter plots can readily be compared with those in Figure 5.29 and in [11]. Similar results are also observed with the Hamiltonian Monte Carlo method (Figure 5.36), with an acceptance rate of 98%. For instance, the acceptance rate obtained with the Metropolis Hastings algorithm was, in this case, less than 10%.

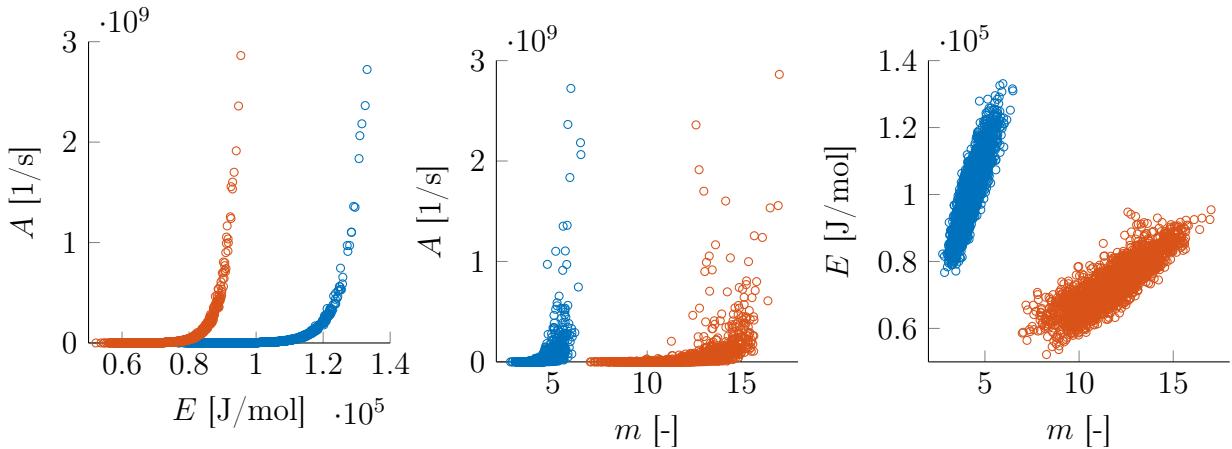


Figure 5.35: Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the Itô-SDE method. The blue dots are the kinetic parameters of the first reaction, and the red dots are those of the second reaction.

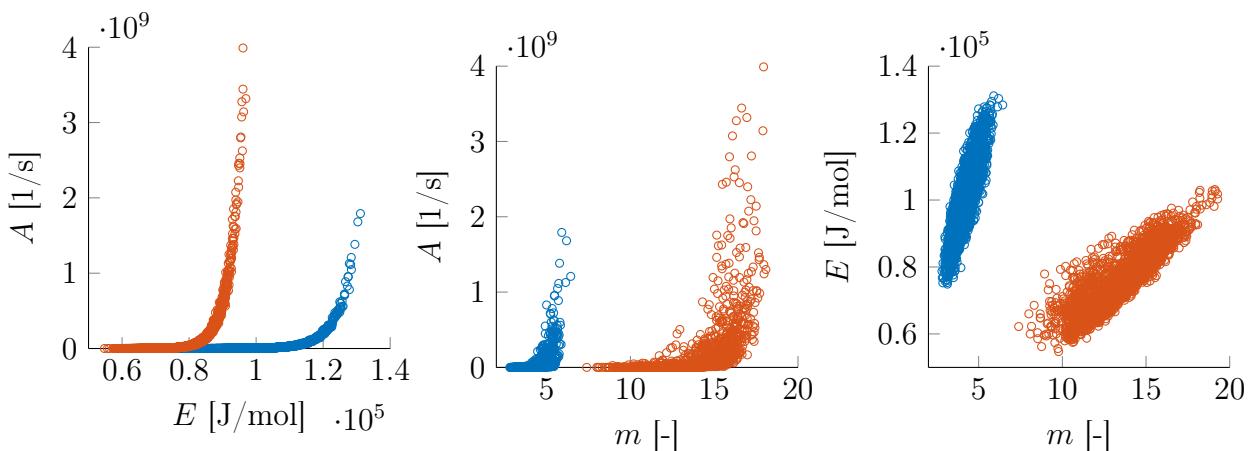


Figure 5.36: Scatter plots of the Markov chains in the original variables \mathbf{q} obtained with the HMC method. The blue dots are the kinetic parameters of the first reaction, and the red dots are those of the second reaction.

The samples obtained for the eleven parameters can then be propagated to the production of the three gases, which is shown, for both methods, in Figures 5.37 and 5.38. Similarly to the previous section, a good agreement between the gas production curves obtained with the samples, and the uncertainties in the data can be generally be observed, with similar results with each methods. However, the lower bound in the production peak of the first gas seems to be under-evaluated compared to the behaviour observed for the third gas. A possible explanation can be the difference in magnitude between the production peaks in the first and third gases.

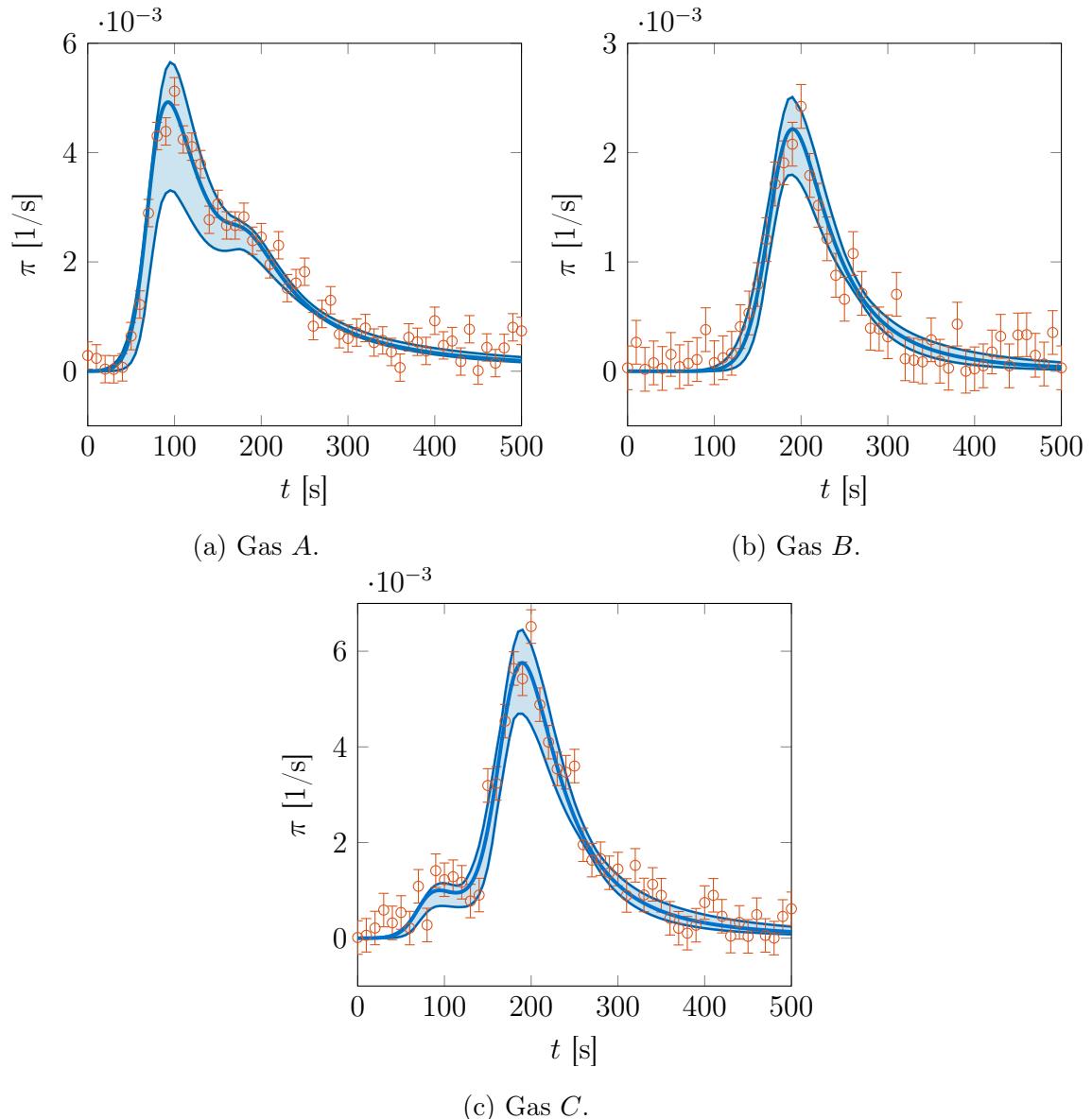


Figure 5.37: Propagation of the samples obtained with the Ito-SDE method to the production of the three gases.

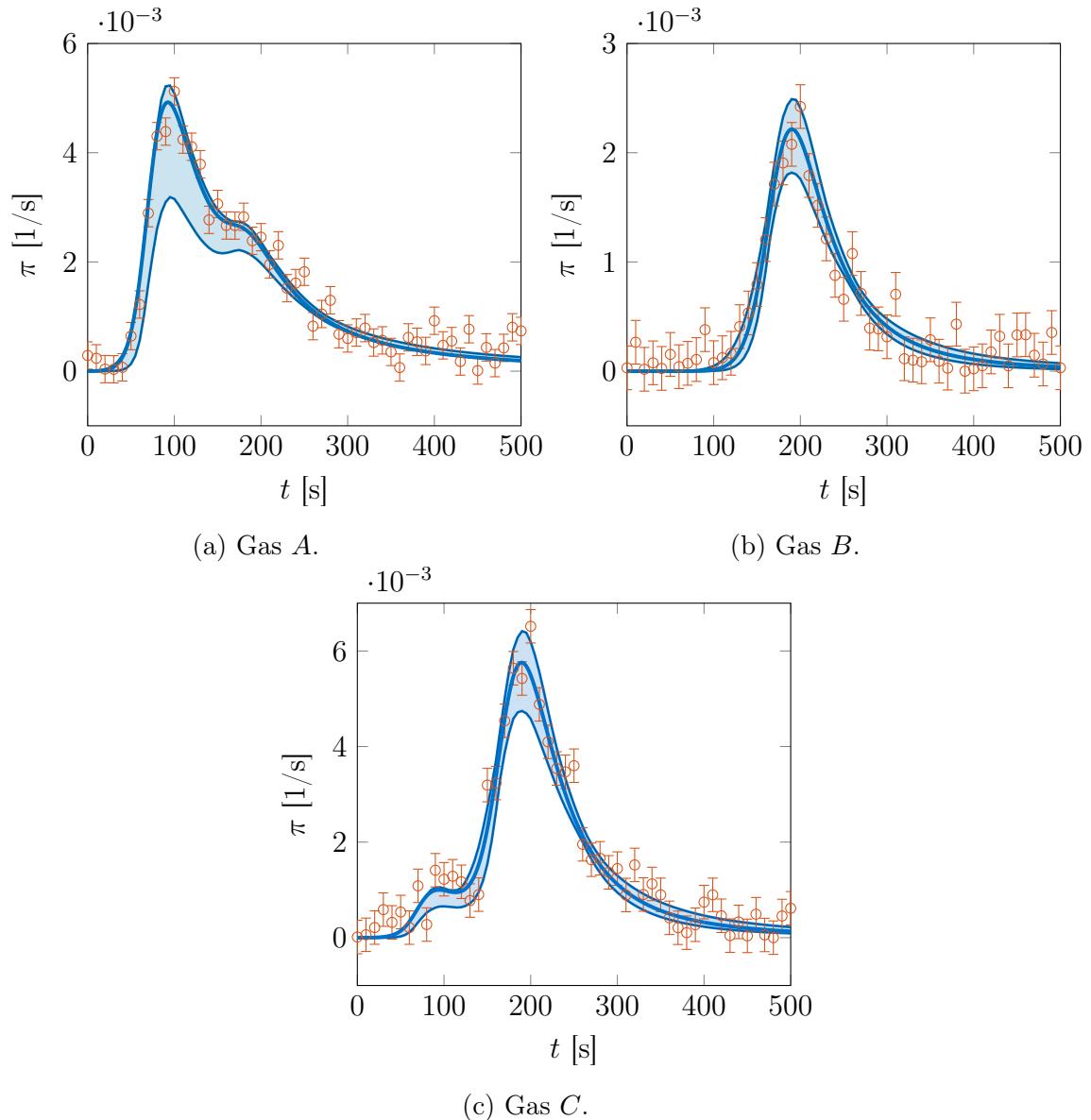


Figure 5.38: Propagation of the samples obtained with the HMC method to the production of the three gases.

In conclusion, both the Ito SDE and Hamiltonian Monte Carlo methods provide similar results when applied to problems involving the pyrolysis model. In particular, they were able to reproduce similar results as in [11] in several cases without requiring manual tuning. Moreover, while the Metropolis-Hastings method does not require the computation of gradient, its low acceptance rate results in a higher number of model evaluation in order to obtain a meaningful number of samples. The two methods developed in this work have an additional numerical cost resulting from the computation of the gradient, but generate much better samples. Coupled with an efficient gradient computation method, these methods seems hence more attractive than the Metropolis-Hastings algorithm.

CHAPTER 6

CONCLUSION AND FUTURE WORKS

6.1 Conclusions

- The first part of this work was related to the study of numerical properties of the dynamical system involved in the pyrolysis model. In particular, the implicit Euler scheme was developed and validated on a simplified pyrolysis model. Using this time integrator, methods for sensitivity analysis were developed and allowed the establishment of a linearised pyrolysis model.
- These numerical methods were then extended to able to compute the gradients appearing, for instance, in cost functions related to optimisation and Bayesian inference. A discrete forward sensitivity method was derived and was illustrated on the production of gaseous species in pyrolysis. Then, an adjoint method designed for discrete problems involving a least-square cost function in the likelihood of the Bayesian formulation of inverse problems was developed and validated on the pyrolysis model. The performance of these methods were assessed for an increasing number of parameters and the adjoint method was found to perform better than the forward sensitivity method, while the forward finite difference method was the least efficient of the three.
- The second part of this work was focused on the Bayesian setting for parameter identification. In particular, the methods for the computation of the gradient allowed establishing a linear approximation of the inverse problem, for which an exact solution as a Gaussian distribution can be found. The covariance matrix appearing in the solution was shown to be of great importance, as it improved the performance of the inference methods developed in this work. Moreover these gradients also allowed the development of Bayesian inference methods that relied on the gradient of a cost function similar to the one mentioned when developing the adjoint method. These methods use the gradient as an information on the geometry of the posterior density, allowing the Markov chain to explore the whole distribution of the posterior.
- In particular, two methods based on the simulation of trajectories from a dynamical system were used and compared on a simplified pyrolysis case consisting of a single gas and reaction, and showed promising results. These methods, namely the Hamiltonian Monte

Carlo algorithm and the Ito-SDE method indeed require less tuning, while generating better samples than the Metropolis-Hastings algorithm. Indeed, the Metropolis-Hastings algorithm resulted in low acceptance rate, leading to less meaningful samples. Despite having a higher numerical cost, the HMC and Ito-SDE methods are hence more attractive. Moreover, the influence of the tuning possibilities in these methods were discussed and their influence on the performance of the methods were assessed.

- The two methods were then applied to more involved problem consisting in multiple gases and reactions. The results obtained with each method were in accordance with those in [11]. In particular, the samples generated by these two methods were propagated to the production of gas, and the results were in good agreement with the experimental data and the associated measurement errors.

6.2 Future work

- Future works should include the transcription of the code developed during this work to the Bayesian inference library pyBit developed by Joffrey Coheur. To do so, some optimisation of the codes might be necessary for improved performance. Furthermore, these methods for computing gradients can and should be extended to other time stepping methods (*e.g.* to Runge-Kutta integration schemes [25]) as well as other model of pyrolysis so as to obtain a more robust toolbox. These will require to be able to compute the different matrices that depends on the residual of the numerical method.
- Furthermore, here, only simple cases were considered while real experiments on pyrolysis were left behind. The numerical methods presented in this work could, for instance, be used on the data from [7]. However, some preliminary tests revealed that the covariance matrix obtained with the reaction scheme established in [11] were ill-conditioned. Hence, future work could include adapting the reaction scheme to these methods. To adapt the reaction scheme, it may be interesting to explore, for instance, the eigenvalue spectrum of the covariance matrix, which may allow to construct better models.
- The numerical methods employed here should be developed with optimisation solvers to benefit fully from the properties that were investigated, regarding for instance the linearisation of the inverse problem. The optimal points found by an optimisation algorithm could be used as starting points for the Bayesian method, in order to quantify the uncertainty in the parameters. This optimisation solver could, for instance, be a quasi-Newton method which uses the gradients obtained by the adjoint method developed in Chapter 4. In turn, the quasi-Newton method provides an approximation of the Hessian matrix, which could be used as an alternative for the matrix $[M]$ in the inference methods.
- Moreover, many improvements of the inference methods developed in this work exist. In fact, the algorithms implemented are, in the most case, basic versions of these methods. For example, the extension from [15] where concepts of differential geometry are used to construct a more efficient method, notably by considering the matrix $[M(\mathbf{q})]$ as a local metric in a Riemannian manifold. These methods could improve some results, since the present implementation only consider $[M]$ to be a constant, which may be valid only in the vicinity of the parameters \mathbf{q} at which it

is evaluated and may not be as efficient when the algorithms sample point located further away. Furthermore, it should be interesting to study in more details the difference in properties such as convergence between methods such as Hamiltonian Monte Carlo, which is based on a reversible equation and the Ito-SDE method based on an irreversible hamiltonian system. Finally, it should also be noted that MCMC methods are not the only possibility to perform parameter estimations, and some recent methods based, for instance, on artificial intelligence [24] may be interesting to investigate.

BIBLIOGRAPHY

- [1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.
- [2] M. Arnst and C. Soize. Identification and sampling of bayesian posteriors of high-dimensional symmetric positive-definite matrices for data-driven updating of computational models. *Computer Methods in Applied Mechanics and Engineering*, 352:300 – 323, 2019.
- [3] M. Arnst, B. Abello Álvarez, J.-P. Ponthot, and R. Boman. Itô-sde mcmc method for bayesian characterization of errors associated with data limitations in stochastic expansion methods for uncertainty quantification. *Journal of Computational Physics*, 349:59 – 79, 2017.
- [4] Maarten Arnst. *Stochastic modelling*. Lecture notes, University of Liège, 2018.
- [5] Richard C. Aster, Brian Borchers, and Clifford H. Thurber. *Parameter Estimation and Inverse Problems*. Academic Press, Boston, second edition edition, 2013.
- [6] Charles H. Bennett. Mass tensor molecular dynamics. *Journal of Computational Physics*, 19(3):267 – 279, 1975.
- [7] Brody K. Bessire and Timothy K. Minton. Decomposition of phenolic impregnated carbon ablator (pica) as a function of temperature and heating rate. *ACS Applied Materials & Interfaces*, 9(25):21422–21437, 2017. PMID: 28544848.
- [8] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo, 2017. arxiv:1701.02434.
- [9] Marc Bonnet. *Problèmes inverses*. Lecture notes, Paris, 2004.
- [10] D. Calvetti and E. Somersalo. Inverse problems: From regularization to bayesian inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(3):e1427, 2018.
- [11] Joffrey Coheur. Bayesian parameter inference for pica devolatilization pyrolysis at high heating rates. *NASA Ames Research Center, Proceedings of the Summer Program 2018*, 2018.
- [12] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987.

- [13] Colin Fox, Geoff K. Nicholls, and Sze M. Tan. *Introduction to inverse problems*. Lecture notes, 2010.
- [14] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.
- [15] Mark Girolami, Ben Calderhead, and Siu A. Chin. Riemann manifold langevin and hamiltonian monte carlo methods. *J. of the Royal Statistical Society, Series B (Methodological)*, 2011.
- [16] Howard E. Goldstein. Pyrolysis kinetics of nylon 6–6, phenolic resin, and their composites. *Journal of Macromolecular Science*, 1969.
- [17] J. Guilleminot and C. Soize. Itô sde-based generator for a class of non-gaussian vector-valued random fields in uncertainty quantification. *SIAM Journal on Scientific Computing*, 36(6):A2763–A2786, 2014.
- [18] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*; 2nd ed. Springer, Dordrecht, 2006.
- [19] Victor Isakov. *Inverse Problems for Partial Differential Equations*. Springer, 2017.
- [20] J. Lachaud, J.B. Scoggins, T.E. Magin, M.G. Meyer, and N.N. Mansour. A generic local thermal equilibrium model for porous reactive materials submitted to high temperatures. *International Journal of Heat and Mass Transfer*, 108:1406 – 1417, 2017.
- [21] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. *Free Energy Computations*. IMPERIAL COLLEGE PRESS, 2010.
- [22] Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162, 2010.
- [23] Francisco Duarte Moura Neto and Antônio José da Silva Neto. *An Introduction to Inverse Problems with Applications*. Springer, 2013.
- [24] Tong Qin, Kailiang Wu, and Dongbin Xiu. Data driven governing equations approximation using deep neural networks. *Journal of Computational Physics*, 395:620 – 635, 2019.
- [25] Adrian Sandu. On the properties of runge-kutta discrete adjoints. In Vassil N. Alexandrov, Geert Dick van Albada, Peter M. A. Sloot, and Jack Dongarra, editors, *Computational Science – ICCS 2006*, pages 550–557, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [26] Pierre Schrooyen. *Numerical simulation of aerothermal flows through ablative Thermal Protection Systems*. Université catholique de Louvain, 2015.
- [27] C Soize. *The Fokker-Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions*. WORLD SCIENTIFIC, 1994.
- [28] Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.