Simulation of Random Variables, Hamiltonian Dynamics, and Sampling via Markov Chains

Author:

Alfonso Mateos Vicente

Tutor:

Noé Blassel

École des Ponts
ParisTech

Ingénierie Mathématique et Informatique
École des Ponts ParisTech
France
September 01, 2023

Contents

1 Simulation of Random Variables **4**

1.1 Inverse CDF method 4

1.2 Rejection Sampling method 6

1.2.1 Selection of the ”enveloping” distribution 7

1.2.2 Results of the Rejection Sampling method 7

1.3 Empirical validation of the Law of Large Numbers 8

1.4 Empirical Validation of the Central Limit Theorem 9

1.5 Variance reduction techniques 12

1.5.1 Preliminaries 13

1.5.2 Control variates 14

1.5.3 Importance sampling 15

1.5.4 Stratified sampling 16

1.5.5 Antithetic variates 17

1.5.6 Comparative Review 18

2 Hamiltonian dynamics **20**

2.1 Introduction to Hamiltonian dynamics 21

2.2 Symplectic schemes 22

2.2.1 Analytical solution 23

2.2.2 Euler method 25

2.2.3 Failure of standard methods 26

2.2.4 Constructing Symplectic Schemes 27

2.2.5 Symplectic Euler method 28

2.2.5.1 Linear Stability Analysis of the Sympletic Euler Scheme 30

2.2.6 Stormer-Verlet method 30

2.2.6.1 Linear Stability Analysis of the Verlet Scheme 32

2.2.7 Backward Error Analysis 33

2.3 Solar System Simulation 34

2.3.1 Modeling the Solar System: Gravitational Dynamics 34

2.3.2 Störmer-Verlet Scheme for Gravitational Dynamics 35

2.3.3 Implementation of the Solar System Simulation 37

2.3.3.1 Experiment 1: Expanding to the N-Body Problem 39

2.3.3.2 Experiment 2: Using Symplectic Euler Method 40

3 Markov Chains **42**

3.1 Introduction to Markov Chains 42

3.1.1 Stationary Distribution 42

3.1.2 Irreducibility and Uniqueness 43

3.2 Markov Chain Montecarlo Methods 43

3.2.1 Metropolis-Hastings Algorithm 43

3.2.1.1 Example: Sampling from a Gamma Distribution 44

3.2.2 Gibbs Sampling 44

3.2.2.1 Example: Sampling from a Bivariate Normal Distribution 45

3.2.3 Müller-Brown Potential 46

3.2.3.1 Metropolis-Hastings Algorithm for the Müller-Brown Potential . . . 47

3.2.3.2 Gibbs Sampling Algorithm for the Müller-Brown Potential 48

3.2.3.3 Comparison between Metropolis-Hastings and Gibbs Sampling . . 49
3.3 Simulation of a 2D molecular system 50
3.3.1 Theoretical Framework and Lennard-Jones Potential 50
3.3.2 Implementation using the Metropolis Hastings Algorithm 50

Abstract

This paper presents a comprehensive examination of simulation techniques in the realms of random variables, Hamiltonian dynamics, and Markov chain sampling. The initial section is devoted to the simulation of random variables, leveraging the Inverse Cumulative Distribution Function (CDF) method and Rejection Sampling method. Empirical validations of the Law of Large Numbers and the Central Limit Theorem are conducted to ascertain the efficacy of these methods. We also explore various variance reduction techniques, including control variates, importance sampling, stratified sampling, and antithetic variates, culminating in a detailed comparative analysis.

The subsequent section shifts focus to Hamiltonian dynamics, beginning with an introduction to the core principles, followed by an exploration of symplectic schemes. This includes analytical solutions, the Euler method, and a critical examination of the failure of standard methods. The development of Symplectic Schemes, particularly the Symplectic Euler and Stormer-Verlet methods, is presented with a comprehensive linear stability analysis. The application of these methods is illustrated through the simulation of the Solar System's gravitational dynamics, extending to experiments involving the N-Body Problem and the Symplectic Euler Method.

The paper culminates with a section on Sampling via Markov Chains. This section explores the application of Euler-Moruyama schemes to 2D potentials, delve into the intricacies of Langevin dynamics and splitting schemes, and present the final implementation in a full 2D molecular system. This aims to deepen the understanding of Markov chain sampling techniques and their practical applications in complex systems.

1 Simulation of Random Variables

This section explores the simulation of random variables, a cornerstone in statistical and dynamical system studies. In Section 1.1, we introduce the Inverse CDF method (1.1), providing a foundational overview of its mechanics and application in simulating random variables.

Progressing to Section 1.2 (1.2), the manuscript transitions to the Rejection Sampling method. This section is further divided to enhance clarity and depth: Section 1.2.1 (1.2.1) delves into the selection of the appropriate enveloping distribution, and Section 1.2.2 (1.2.2) presents the outcomes and insights from applying this method.

The narrative continues in Section 1.3 (1.3), where the empirical validation of the Law of Large Numbers is discussed, and in Section 1.4 (1.4), focusing on the Central Limit Theorem. These sections underscore the practical validation and implications of these pivotal statistical laws in simulation scenarios.

Concluding this section, a detailed examination of various variance reduction techniques is presented in Section 1.5 (1.5). Starting with fundamental concepts in Section 1.5.1 (1.5.1), it then explores control variates (Section 1.5.2, 1.5.2), importance sampling (Section 1.5.3, 1.5.3), stratified sampling (Section 1.5.4, 1.5.4), and antithetic variates (Section 1.5.5, 1.5.5), culminating in a comparative review in Section 1.5.6 (1.5.6). This part of the manuscript synthesizes and compares these techniques, highlighting their role in enhancing the accuracy and efficiency of simulation processes.

1.1 Inverse CDF method

Inverse CDF method is a basic method to generate pseudo-random numbers from any probability distribution given its cumulative distribution function. First of all, let's introduce the following statement extracted from [2]:

Theorem 1.1 (Inverse Transform Sampling)**.** For any random variable $X \in \mathbb{R}$, the random variable $F_X^{-1}(U)$ has the same distribution as X, where F_X^{-1} is the generalized inverse of the cumulative distribution function F_X of X and U is uniform on $[0, 1]$.

For continuous random variables, the inverse probability integral transform is indeed the inverse of the probability integral transform, which states that for a continuous random variable X with cumulative distribution function F_X, the random variable $U = F_X(X)$ is uniform on $[0, 1]$.

Knowing this, the idea is to generate a uniform random variable U and then apply the inverse of the cumulative distribution function F^{-1} to obtain a random variable with the desired distribution. The Algorithm 1 is as follows:

Algorithm 1: Inverse CDF mehtod

1. Generate a set of random numbers $U \sim \mathcal{U}(0, 1)$;
2. Find the inverse of the cumulative distribution function F^{-1} ;
3. Apply the inverse to the set of random numbers $X = F^{-1}(U)$;

Let's see this with some examples. Using the exponential distribution, we know its probability density function is $f(x) = \lambda e^{-\lambda x}$. Also, we alredy know its cumulative distribution function which is $F(x) = 1 - e^{-\lambda x}$. Then, the inverse of the cumulative distribution function is:

$$F^{-1}(x) = -\frac{1}{\lambda} \ln(1 - x) \tag{1}$$

So, we can generate a set of random numbers $U \sim \mathcal{U}(0, 1)$ and apply the inverse to obtain a set of random numbers $X = F^{-1}(U) \sim \mathcal{E}(\lambda)$. The Figure 1 shows the histogram of the generated random numbers and the probability density function of the exponential distribution.

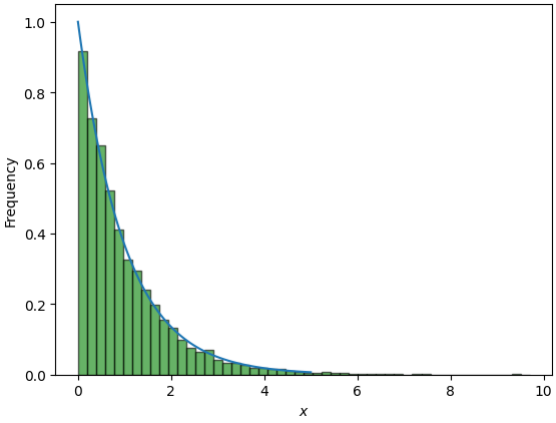

Figure 1: Histogram of Equation (1) with $\lambda = 1$ applied to $\mathcal{U}$.

With this picture, we can see that the histogram of the generated random numbers aligns remarkably well with the target distribution, however, we need to make more trials to be sure. We already know the theoretical mean and variance of the exponential distribution, so we can compare them with the mean of the samples while the number of trials increases.

We define the error as:

$$error = \frac{1}{n} \sum_{i=1}^{n} |x_i - \mu|$$

where x_i is the i-th random number generated and μ is the theoretical mean of the distribution.

Recall that the uniform and exponential means are:

$$\mu_{\mathcal{U}} = \frac{a + b}{2}$$

$$\mu_{\mathcal{E}} = \frac{1}{\lambda}$$

The Figure 2 and Figure 3 shows the error to the theoretical mean as the number of trials increases for the uniform and exponential distribution.

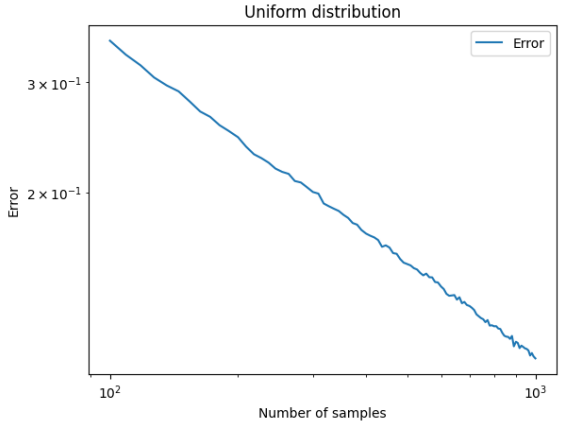

Figure 2: Error to the theoretical mean as the number of trials increases for the uniform distribution. Average of 10000 trials per point.

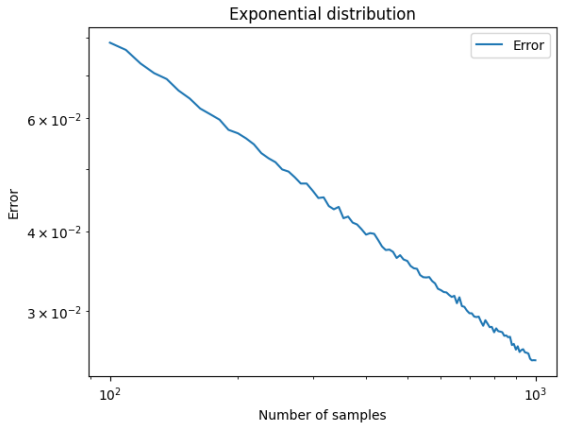

Figure 3: Error to the theoretical mean as the number of trials increases for the exponential distribution. Average of 10000 trials per point.

In drawing conclusions from the observed alignment of the histograms of generated random numbers with the target distribution, t here appears to be a noteworthy relation, as evidenced by Figures 2 and 3, since the error tends to zero as the number of trials increases.

1.2 Rejection Sampling method

The Rejection Sampling method is a method to generate pseudo-random numbers for any probability distribution given its probability density function. The idea is to generate a set of random numbers from a probability distribution that is easy to sample from and then reject the numbers that are not in the desired distribution. The Algorithm 2 is as follows:

Algorithm 2: Rejection Sampling method

1. Generate a set of random numbers $X \sim g(x)$;
2. Generate a set of random numbers $U \sim \mathcal{U}(0, 1)$;
3. If $U \leq \frac{f(X)}{Mg(X)}$ then accept X, otherwise reject X ;

Being $g(x)$ the probability density function with which we already know how to generate random numbers, denoted as the "enveloping" distribution; $f(x)$ the target probability density function; and M a factor we can choose manually and can be optimized. The formal explanation of why this algorithm works can be found in [12].

Let's see this with one example. We want to generate a set of random numbers from the following probability density function:

$$f(x) = 0.3e^{-0.2x^2} + 0.7e^{-0.2(x-5)^2} \tag{2}$$

1.2.1 Selection of the "enveloping" distribution

In this subsection we will explain the process of selecting an optimised "enveloping" distribution. In fact, we could use any distribution that envelops the desired one, but this will probably be very inefficient because there will be many samples that will not be under the objective function, so our goal is to obtain a distribution that envelops the desired one but with the smallest possible space between them. In this report we have chosen the normal distribution as the "enveloping" distribution because it is easy to sample by using the *Quantile function* as the inverse CDF and applying the previous sampling method, and it is a good candidate for enveloping the desired distribution. Knowing this, we have three parameters to optimise: the mean, μ; the variance, σ; and the scale parameter, M. Note that for this problem, we have to define the bounds as a, b.

First of all, let's start by defining the scale parameter M. In this moment, what we are looking for is a parameter that envelops the desired distribution, therefore it is a good idea to define M as follows:

$$M = \max_{x \in \mathbb{R}} \frac{f(x)}{g(x)} \tag{3}$$

In this way, we ensure that whatever the function is, we will envelope it. Knowing this parameter, we only have to choose μ and σ in the order of minimizing M. So we have a problem of optimization with two variables. The problem is the following:

$$\begin{aligned} \min_{\mu,\sigma} \quad & M(\mu,\sigma) \\ \text{s.t.} \quad & M(\mu,\sigma) = \max_{x \in \mathbb{R}} \frac{f(x)}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \\ & \sigma \geq 0, a \leq \mu \leq b \end{aligned} \tag{4}$$

We can solve this problem with multiple methods. In our case, we have chosen the Nelder-Mead's method (also called downhill simplex method) [11].

1.2.2 Results of the Rejection Sampling method

Once we have selected the "enveloping" distribution, we can apply the Rejection Sampling method. In our case, the enveloping function is the N($\mu \approx 3.644$,$\sigma \approx 3.041$) and the scalator parameter $M \approx 1.545$. The Figure 4 shows the histogram of the generated random numbers and the probability density function of the desired and the easy distribution.

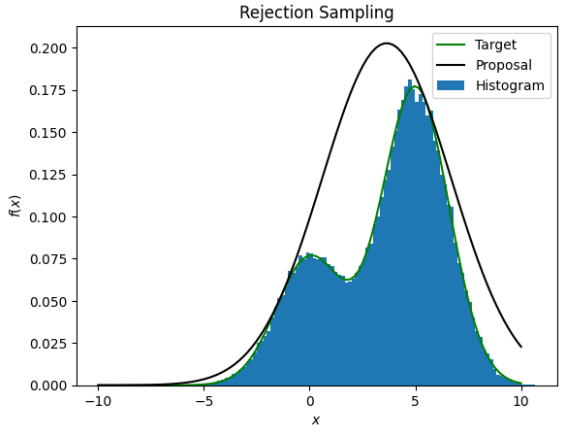

Figure 4: Histogram of the generated random numbers, Equation (2) and PDF of $\mathcal{N}(3.644, 3.041)$ scaled by $M \approx 1.545$.

First of all, note that the distribution chosen with the Equation 4 fits perfectly with the target distribution. Moreover, we can see that the histogram of the generated random numbers aligns remarkably well with the target distribution. In addition to the graph display itself, we can now also compare the mean of the samples with the theoretical mean as the number of trials increases. The Figure 5 shows the error to the theoretical mean. As disclaimer, we will use the following approximation for the mean: $\frac{1}{n}\sum_{i=1}^{n} x_i$ and the following approximation for the variance: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

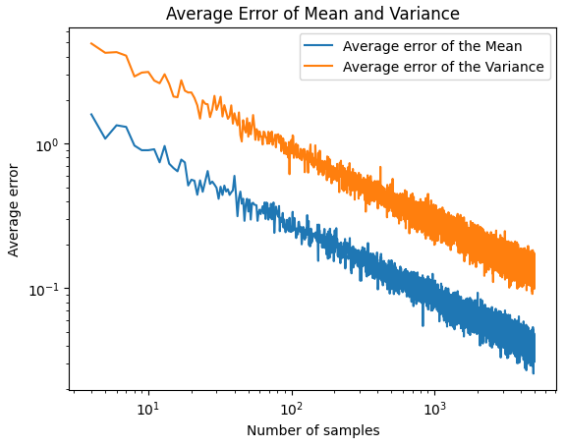

Figure 5: Error to the theoretical mean as the number of trials increases

In drawing conclusions we can see that the error tends to zero as the number of trials increases, so we can conclude that the Rejection Sampling method is a good method to generate pseudo-random numbers from any probability distribution given its probability density function.

1.3 Empirical validation of the Law of Large Numbers

This section explores the empirical validation of the Law of Large Numbers, a pivotal theorem in probability theory. The Law of Large Numbers (LLN) is a fundamental theorem in probability theory, asserting that the average of a sequence of independent and identically distributed random

variables with finite expected value converges to this expected value as the number of terms in the sequence increases. Formally, the theorem can be expressed as follows, extracted from [10]:

Theorem 1.2 (Law of Large Numbers)**.** Let $(X_j)_{j\geq 1}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(X_1^2) < +\infty$ then

$$\mathbb{E}(\bar{X}_n - \mathbb{E}(X_1))^2 = \frac{\mathrm{Var}(X_1)}{n} \xrightarrow[n\to\infty]{} 0 \tag{5}$$

Given our established ability to generate random numbers reflecting specific probability distributions, we can select a distribution, generate a set of numbers corresponding to it, and then contrast the calculated averages and means of the target distribution while escalating the number of trials. Now we can define the following error:

$$error = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{6}$$

where x_i is the i-th random number generated and μ is the theoretical mean of the distribution. Note that in the theorem was stated that the esperance of the difference, behaves as the variance divided by the number of trials, so we can plot the variance as the number of trials increases and calculate the slope of the regression line. If the theory is correct, the slope should be around -1 since the variance of the normal distribution is $\mathbb{V}(\bar{X}) = \frac{sigma^2}{n}$, therefore, the variance is inversely proportional to the number of variables. The Figure 6 shows the variance as the number of variables increases and the regression line.

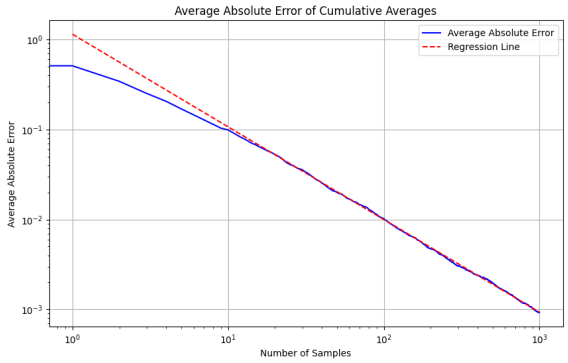

Figure 6: Logarithmic depiction of the absolute error relative to the mean. Average of 1000 trials per point. SLOPE!

The visual representation in Figure 6 substantiates that the error indeed diminishes as the number of trials augments, validating the assertion of the Law of Large Numbers that the experimental mean approaches the theoretical mean with an increasing number of observations. This is further corroborated by the linear regression, which yields a slope of -1.000, in alignment with the theoretical prediction.

1.4 Empirical Validation of the Central Limit Theorem

The Central Limit Theorem (CLT) posits a pivotal foundational theorem in probability theory, signifying that, irrespective of the shape of the original distribution, the distribution of sample means

will approximate a normal distribution as the sample size increases. To empirically validate this theorem, we can simulate a series of random numbers from any given probability distribution and systematically calculate their mean. By repeating this process, we can construct a histogram of the means and scrutinize whether it converges to a normal distribution, aligning with the theorem's prediction. Formally, the theorem can be expressed as follows: The Central Limit Theorem (CLT) posits a pivotal foundational theorem in probability theory, signifying that, irrespective of the shape of the original distribution, the distribution of sample means will approximate a normal distribution as the sample size burgeons. To empirically validate this theorem, we can simulate a series of random numbers from any given probability distribution and systematically calculate their mean. By perpetuating this process, we can construct a histogram of the means and scrutinize whether it converges to a normal distribution, aligning with the theorem's prediction. Formally, the theorem can be expressed as follows, extracted from [10]:

Theorem 1.3 (Central Limit Theorem)**.** Let $(X_j)_{j\geq 1}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(X_1^2) < +\infty$ and $\sqrt{\mathrm{Var}(X_1)} > 0$. We note $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ and $\sigma^2 = \mathrm{Var}(X_1)$. Then for $n \to \infty$,

$$\sqrt{n}\frac{\bar{X}_n - \mathbb{E}(X_1)}{\sigma} \xrightarrow[n\to\infty]{} Y \sim \mathcal{N}(0,1) \tag{7}$$

For illustrative purposes, consider the uniform distribution on $[0, 1]$, which yields a histogram as depicted in Figure 7.

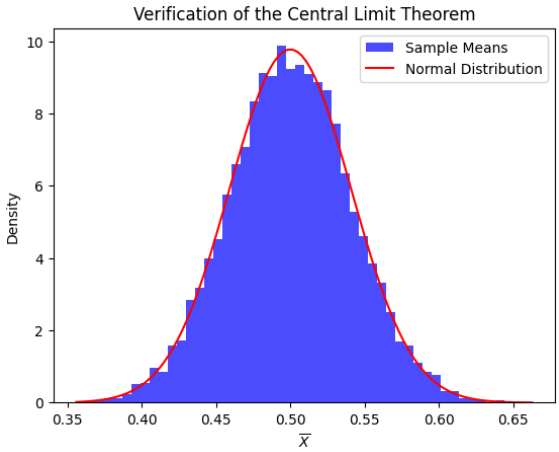

Figure 7: Histogram illustrating the convergence of the means of the uniform distribution with $n_{samples} = 10000$

In congruence with the methodology espoused in previous sections, we have juxtaposed the empirical mean of the samples with the theoretical mean, progressively augmenting the number of trials. Figure 8 delineates the deviation from the theoretical mean, and Figure 9 represents the deviation from the theoretical variance, both as functions of the number of trials. Therefore the formula of the deviation is:

$$error_{mean} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu|$$

where x_i is the i-th random number generated and μ is the theoretical mean of the distribution.

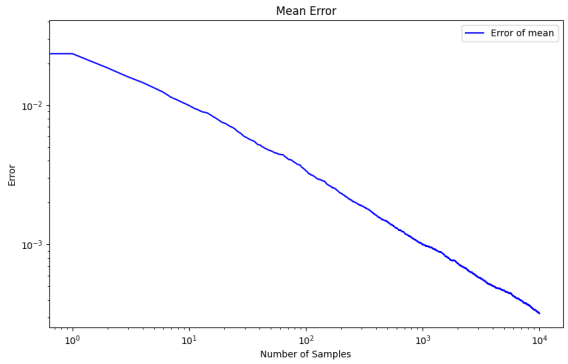

Figure 8: Error from the theoretical mean as a function of the number of trials. Average of 10000 trials per point.

And for the case of the variance:

$$error_{variance} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \sigma^2|$$

where x_i is the i-th random number generated and σ^2 is the theoretical variance of the distribution.

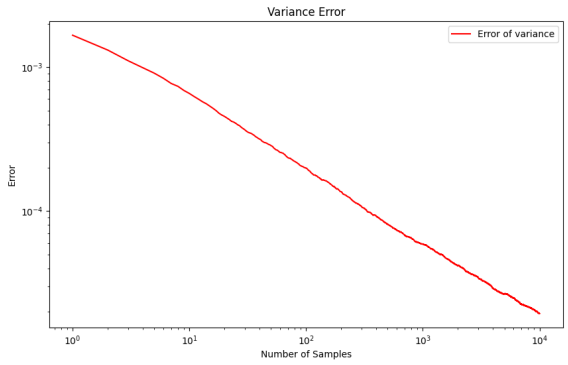

Figure 9: Error from the theoretical variance as a function of the number of trials. Average of 10000 trials per point.

A discernible insight gleaned from Figure 8 and Figure 9 is the palpable decrement in deviations from the theoretical values as the trials proliferate.

Finally, another way to validate the CLT is to plot the variance as the number of variables increases and calculate the slope of the regression line. If the theory is correct, the slope should be around -1 since the variance of the normal distribution is $\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}$, therefore the variance is inversely proportional to the number of variables. The Figure 10 shows the variance as the number of variables increases and the regression line.

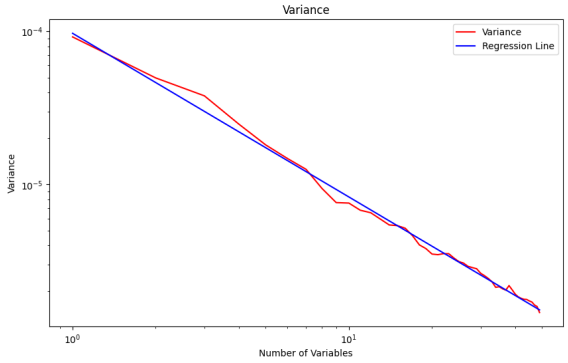

Figure 10: Deviation from the theoretical variance as a function of the number of trials. Average of 1000 trials per point.

The slope of the regression line is -1 in alignment with the theoretical prediction. This corroborates the validity of the Central Limit Theorem, as the variance of the means of the samples does indeed diminish as the number of trials increases.

1.5 Variance reduction techniques

In the theory of Monte Carlo methods, variance reduction techniques are a pivotal tool to increase the precision of the estimates of the expected value of a random variable. In this section, we will focus on three techniques: Control variates, Importance sampling and Antithetic variates. Also, we will introduce an example problem in which we will apply these techniques to compare them.

First of all, let introduce the variance reduction formally. Let X be a random variable drawn from a distribution of mean $\mathbb{E}(X) = \mu$ and finite variance $\mathrm{Var}(X) = \sigma^2$. The estimator of the expected value of X, i.e. the empirical mean of the sample, is defined as follows:

$$\hat{\mathbb{E}}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

So we want to reduce the variance of our estimation. Since the variance is defined as follows:

$$\mathrm{Var}(\hat{\mathbb{E}}(X)) = \frac{\mathrm{Var}(X)}{n} \tag{9}$$

Consequently, we are presented with two avenues for optimization: increasing the value of n, or diminishing the variance of X. Assuming that n is predetermined and unalterable, our focus would then shift to minimizing the variance of X.

Since we are going to introduce and compare the three techniques, first of all we need to introduce the problem we are going to solve. The problem is that we want to estimate the following integral:

$$I = \int_0^1 x^2 \, dx \tag{10}$$

It should be noted that the selection of the integral for this demonstration was intentional; a readily solvable integral was chosen for its ease of analytical computation, allowing for a straightforward comparison with theoretical values. Nonetheless, the methods illustrated herein are equally

applicable and potent for evaluating integrals that pose substantial challenges to analytical computation.

1.5.1 Preliminaries

First of all we have to compute which is the estimation we already can have without applying any variance reduction technique. We can compute the integral analytically. We know that the mean of the variable in a probability space is defined as follows:

$$\mathbb{E}(g(X)) = \int_a^b g(x)f(x)\,dx \tag{11}$$

Since $f(x) = \frac{1}{b-a}$ for the uniform distribution on $[a, b]$, we can apply Equation (11) to Equation (10) and we obtain:

$$I = \int_0^1 x^2\,dx = \mathbb{E}(g(X)) = \int_0^1 x^2 \frac{1}{1-0}\,dx = \frac{1}{3} \tag{12}$$

So, we only have to generate a set of random numbers $X \sim \mathcal{U}(0, 1)$, apply $g(x) = x^2$ and get the mean of the sample which is the estimate of the integral. The Figure 11 shows the error to the theoretical value as the number of trials increases. The error to the theoretical value is defined as:

$$error = \frac{1}{n}\sum_{i=1}^{n} |x_i - \mu|$$

where x_i is the i-th random number generated and μ is the theoretical mean of the distribution. Note that this formula will be used in all the following sections.

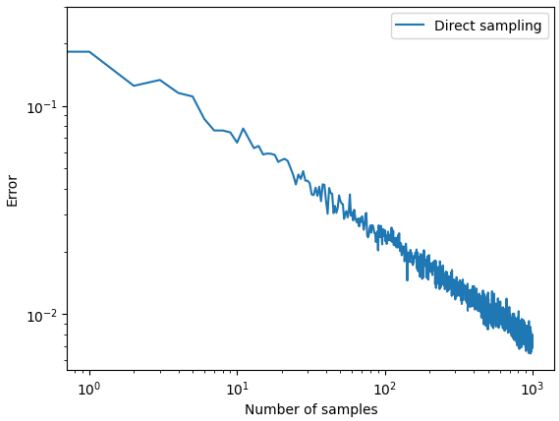

Figure 11: Error to the theoretical value as the number of trials increases using the direct method. Average of 100 trials per point.

Following the Equation (9), we can compute the variance of the estimation as follows:

$$\mathrm{Var}(X) = \mathrm{Var}(\hat{\mathbb{E}}(X))\,n \tag{13}$$

In this case, we have computed that $\mathrm{Var}(X) \approx 8.977 \cdot 10^{-2}$.

1.5.2 Control variates

Control variates is a variance reduction technique with the following idea: Let μ the parameter we want to estimate, and assume we have a statistic Y such that $\mathbb{E}(Y) = \tau$. Then, we can estimate μ by estimating $\mathbb{E}(Y)$ as $\hat{\mathbb{E}}(Y)$ and correcting the bias with the following formula:

$$\hat{\mu} = \mu + c(\tau - \hat{\mathbb{E}}(Y)) \tag{14}$$

Being c a constant which minimize the variance of the estimation. It is computed as follows:

$$c = -\frac{\mathrm{Cov}(\mu, \tau)}{\mathrm{Var}(\tau)} \tag{15}$$

Proof. Using the Equation (14) we can compute the variance of the estimation as follows:

$$\begin{gathered}\mathrm{Var}(\hat{\mu}) = \\ \mathbb{E}[(\mu + c(\tau - \hat{\mathbb{E}}(Y)))^2] - \mathbb{E}[\mu + c(\tau - \hat{\mathbb{E}}(Y))]^2 = \\ \mathbb{E}[\mu^2 + c^2(\tau - \hat{\mathbb{E}}(Y))^2 + 2c\sigma(\tau - \hat{\mathbb{E}}(Y))] - \mathbb{E}[\mu]^2 - c^2\mathbb{E}[\tau - \hat{\mathbb{E}}(Y))]^2 - 2c\mathbb{E}[\sigma(\tau - \hat{\mathbb{E}}(Y))] = \\ \mathrm{Var}(\mu) + c^2\mathrm{Var}(\tau) + 2cCov(\mu, \tau)\end{gathered}$$

Therefore, since we want to minimize the variance of the estimation, we can differentiate with respect to c and equate to zero:

$$\tfrac{\partial \mathrm{Var}(\hat{\mu})}{\partial c} = 2c\mathrm{Var}(\tau) + 2\mathrm{Cov}(\mu, \tau) = 0$$

And we obtain the following expression for c:

$$c = -\tfrac{\mathrm{Cov}(\mu,\tau)}{\mathrm{Var}(\tau)}$$

□

In our case, we can use the knowledge of the mean of the uniform distribution to correct the bias, so in each iteration we can compute the mean of the sample of random numbers of $\mathcal{U}(0, 1)$, and knowing that the mean of the uniform distirbution is 0.5 we can apply Equation (14) to obtain the estimate of the integral. The Figure 12 shows the error to the theoretical value as the number of trials increases, i.e. the error of the estimation of the integral:

$$error = \frac{1}{n}\sum_{i=1}^{n} |x_i - \frac{1}{3}|$$

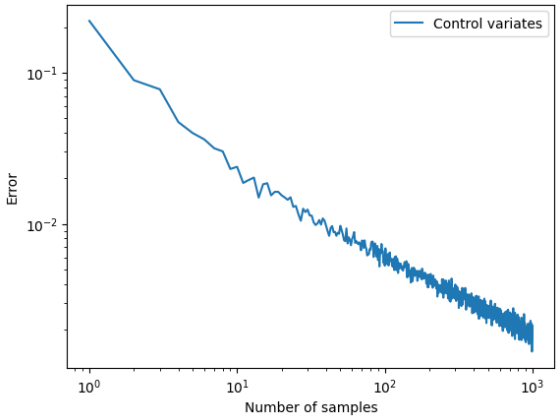

Figure 12: Error to the theoretical value as the number of trials increases using the control variates method. Average of 100 trials per point.

In this case, we have computed that $\mathrm{Var}(X) \approx 5.168 \cdot 10^{-3}$.

1.5.3 Importance sampling

Importance sampling is another variance reduction technique. Instead of using the knowledge of another estimator to reduce the bias of our sample, as done in the control variates method, in this case the idea is using the knowledge of the actual function we want to integrate, so instead of using a uniform distribution, we can use another distribution that is more similar to the function we want to integrate, in order to try more samples in the areas where the function is more important.

In our example, we know that the function is a parabola, so instead of using a sample which follows an uniform distribution, maybe we can use a distribution that fits better with the shape of the function. In this case, we have chosen $Beta(2.9, 1)$. In order to illustrate this, in the Figure 13 we can see the function we want to integrate, the PDF of $Beta(2.9, 1)$ and the PDF of $\mathcal{U}(0, 1)$.

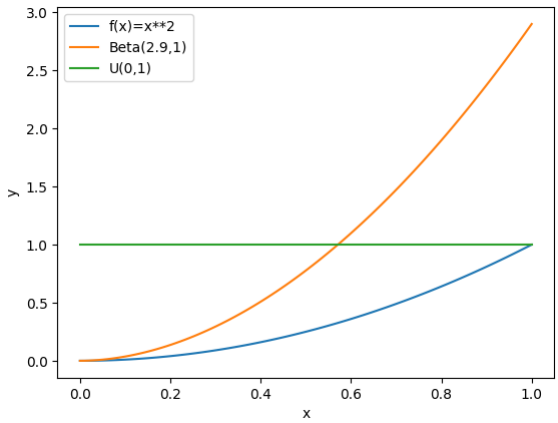

Figure 13: Function we want to integrate, PDF of $Beta(2.9, 1)$ and PDF of $\mathcal{U}(0, 1)$

As we can see, the PDF of $Beta(2.9, 1)$ fits better with the shape of the function we want to integrate, so we can expect that the error will decrease faster than the direct method and the control variates method.

So the idea is to generate a sample of random numbers $X \sim Beta(2.9, 1)$, and take $Y = \frac{f(X)}{g(X)}$ as the estimator of the integral, being $f(x) = x^2$ and $g(x)$ the PDF of $Beta(2.9, 1)$. The algorithm is as follows:

Algorithm 3: Importance Sampling method

1. Generate a set of random numbers $X \sim Beta(2.9, 1)$;
2. Compute $Y = \frac{f(X)}{g(X)}$;
3. Compute the mean of Y as the estimator of the integral ;

With the error:

$$error = |\bar{\mathbb{E}}(X) - \frac{1}{3}|$$

The Figure 14 shows the error to the theoretical value as the number of trials increases.

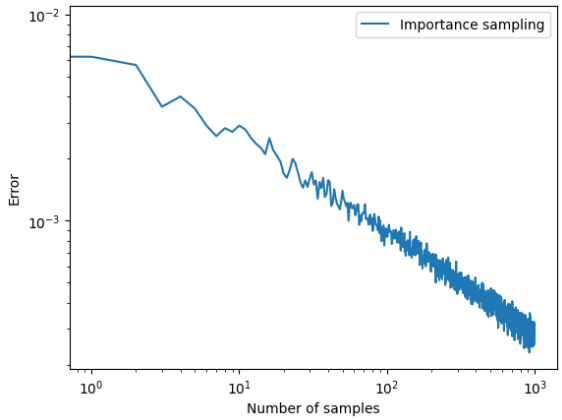

Figure 14: Error to the theoretical value as the number of trials increases using the importance sampling method. Average of 100 trials per point.

In this case, we have computed that $\mathrm{Var}(X) \approx 1.274 \cdot 10^{-4}$.

1.5.4 Stratified sampling

The idea is to divide the interval of the stratified sampling is pretty simple. We divide the interval in n subintervals, and we generate a sample of random numbers for each subinterval. Doing this, we are trying to reduce the variance of the mean of the distribution we are considering, so the sample is more representative of the distribution. The alrogithm is as follows:

Algorithm 4: Stratified Sampling method

1. Divide the interval in n subintervals ;
2. Generate a set of random numbers $X \sim \mathcal{U}(a_i, b_i)$ for each subinterval ;
3. Compute the mean of X as the estimator of the integral ;

With the error:

$$error = |\bar{\mathbb{E}}(X) - \frac{1}{3}|$$

The Figure 15 shows the error to the theoretical value as the number of trials increases.

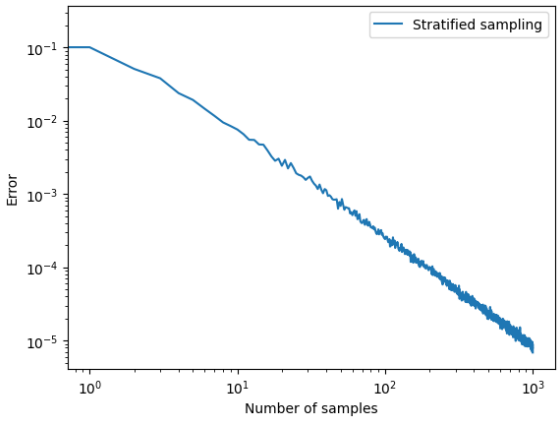

Figure 15: Error to the theoretical value as the number of trials increases using the stratified sampling method. Average of 100 trials per point.

In this case, we have computed that $\mathrm{Var}(X) \approx 1.108 \cdot 10^{-7}$.

1.5.5 Antithetic variates

The Antithetic variates method consists on taking for each random sample, its antithetic, i.e. the symmetric with respect to the mean of the distribution. The idea is that the variance of the mean of the distribution is reduced, since the mean of the antithetic is the same as the mean of the distribution. In our case, we have generated a sample $\mathcal{U} \sim \mathcal{U}(0, 0.5)$ and its antithetic $\mathcal{U}' = \{1 - x \ : x \in \mathcal{U}\}$. We can apply Equation (8) to $f(X)$ to obtain the estimate of the integral. The algorithm is as follows:

Algorithm 5: Antithetic Variates method

1. Generate a set of random numbers $X \sim \mathcal{U}(0, 0.5)$;
2. Compute the antithetic of X ;
3. Compute the mean of $X \bigcup X'$ as the estimator of the integral ;

With the error:

$$error = |\bar{\mathbb{E}}(X) - \frac{1}{3}|$$

The Figure 16 shows the error to the theoretical value as the number of trials increases.

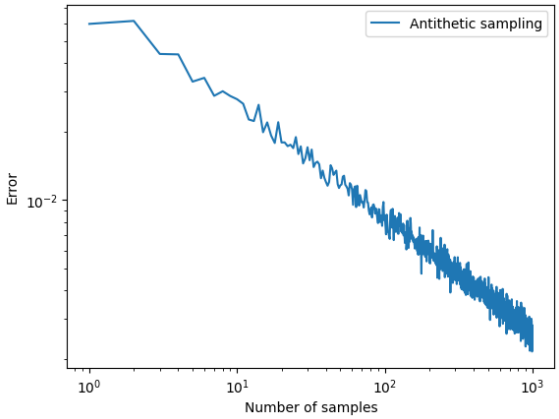

Figure 16: Error to the theoretical value as the number of trials increases using the antithetic variates method. Average of 100 trials per point.

In this case, we have computed that $\mathrm{Var}(X) \approx 1.178 \cdot 10^{-2}$.

1.5.6 Comparative Review

Having looked at the methods and confirmed the convergence of each approach described in this study, we now turn to a more focused comparison of these techniques. The goal is simple: find out which method works best for our specific problem. See Figure 17 for a visual representation of the error in relation to the theoretical value as we increase the number of trials for each method.

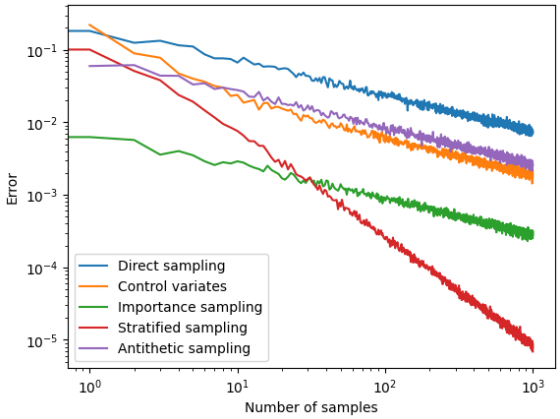

Figure 17: Error variation with increased trials for each method.

Note that the error was defined as follows:

$$error = |\bar{\mathbb{E}}(X) - \frac{1}{3}|$$

Looking at the data, it's clear that stratified sampling takes the lead with its quicker convergence compared to the other methods. However, claiming it as the undisputed champion would be premature, especially when we've explored just one problem. To firm up this initial finding, we need to dig deeper and explore a variety of problems.

But before moving on to look at more problems, let's look at the comparative table of variances for each method:

Method	Variance
Direct sampling	$8.977194 \cdot 10^{-2}$
Control variates sampling	$5.168373 \cdot 10^{-3}$
Importance sampling	$1.274508 \cdot 10^{-4}$
Stratified sampling	$1.107865 \cdot 10^{-7}$
Antithetic sampling	$1.178326 \cdot 10^{-2}$

Recall that the formula used to compute the variance is:

$$\mathrm{Var}(X) = \mathrm{Var}(\hat{\mathbb{E}}(X))\ n$$

With a sample size of $n = 1000$.

Analyzing the table, we discern significant disparities in variance between the different sampling methods, accentuating the prominence of stratified sampling, which registers the minimal variance,. This numerical inferiority in variance corroborates the preliminary observation about its superior convergence rate, offering more stable and reliable estimates. Stratified sampling outperformed other methods in our study primarily due to its targeted approach of dividing the population into homogeneous strata. This division ensures proportional representation of all subgroups, reducing sampling bias and variance within each stratum. Consequently, it yields more accurate and representative results. Furthermore, by adjusting the sample size within each stratum based on variability, stratified sampling enhances overall efficiency and precision. These methodological strengths contribute to its superior performance in terms of reliability and convergence rates, especially in diverse and complex datasets. Contrastingly, direct sampling exhibits the maximum variance, revealing its comparative inefficiency and instability in procuring estimates for this specific problem. The remaining methods, while overshadowed by the efficacy of stratified sampling, still exhibit markedly lower variances than direct sampling.

Next, we'll broaden our investigation to a more complex problem, applying the methods we've discussed to estimate the following integrals:

$$\text{Parabola} = \int_0^1 x^2\ dx \tag{16}$$

$$\text{Gaussian} = \int_0^1 e^{-x^2}\ dx \tag{17}$$

$$\text{Sine} = \int_0^1 \sin(x)\ dx \tag{18}$$

$$\text{Polynomial} = \int_0^1 x^3 - 2x^2 + x\ dx \tag{19}$$

$$\text{Exponential} = \int_0^1 e^x\ dx \tag{20}$$

Using these examples, we can now draw comparisons between the methods in a broader context. Figure 18 illustrates the error of each method applied to each function.

This error is defined as:

$$error = \mid method(f, a, b, iters) - \int_a^b f dx \mid$$

where $method$ is the method we are using, f is the function we are integrating, a and b are the limits of the integral and $iters$ is the number of iterations we are using, in this case $iters = 1000$.

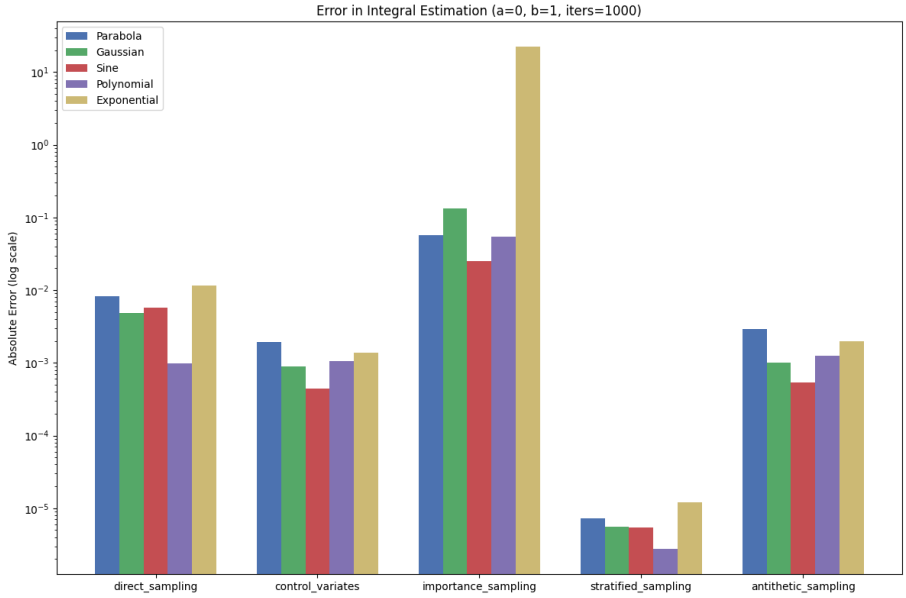

Figure 18: Method-wise error in each function.

Note that in the cace of importance sampling, we have used the gaussian distribution as the sampling distribution, therefore the error could be higher than the other methods, since the gaussian distribution might not be the best sampling distribution for all the functions. To conclude, we can say that the stratified sampling method is the best method for this problem, since it has the lowest error in all the functions.

2 Hamiltonian dynamics

This section delves into the realm of Hamiltonian dynamics, beginning with an in-depth exploration of symplectic schemes in Section 2.1 (2.2). This section starts with a discussion on the analytical solution in Section 2.1.1 (2.2.1), setting the groundwork for understanding symplectic integration methods.

In Section 2.1.2 (2.2.2), the focus shifts to the Euler method, detailing its application and significance in Hamiltonian systems. Following this, Section 2.1.3 (2.2.3) discusses the limitations and failures of standard methods in this context, highlighting the need for more sophisticated approaches.

Section 2.1.4 (2.2.4) addresses the construction of symplectic schemes, a crucial step for ac-

curate simulations in Hamiltonian dynamics. This is further expanded in Section 2.1.5 (2.2.5), which specifically examines the Symplectic Euler method. A detailed Linear Stability Analysis of the Symplectic Euler scheme is then presented in Section 2.1.5.1 (2.2.5.1).

The narrative continues with the Stormer-Verlet method in Section 2.1.6 (2.2.6), another important technique in symplectic integration, followed by a Linear Stability Analysis of the Verlet scheme in Section 2.1.6.1 (2.2.6.1). Section 2.1.7 (2.2.7) concludes this part with a discussion on Backward Error Analysis, providing insights into the error characteristics of symplectic integrators.

Section 2.2 (2.3) shifts the focus to the practical application of these concepts in simulating the solar system. Section 2.2.1 (2.3.1) begins with modeling gravitational dynamics, followed by the implementation of the Störmer-Verlet scheme in Section 2.2.2 (2.3.2). The implementation details of these models are further elaborated in Section 2.2.3 (2.3.3), with two experiments: Experiment 1 in Section 2.2.3.1 (2.3.3.1) expands to the N-Body problem, and Experiment 2 in Section 2.2.3.2 (2.3.3.2) utilizes the Symplectic Euler Method, providing practical insights into their application in complex dynamical systems.

2.1 Introduction to Hamiltonian dynamics

Hamiltonian dynamics, also called "Hamiltonian mechanics", is a reformulation of the clasical mechanics which describes the temporal evolution of a physical system in terms of pairs of variables: the generalized coordinates q_i and their momenta p_i.

$$p(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_n(t) \end{bmatrix} \quad q(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ q_n(t) \end{bmatrix}$$

In this section we consider the time evolution of a isolated system described at a microscopic level, i.e. a system of particles. The state of the system is described by the position of the particles q_i and their momenta p_i. We denote D the dimension of the positions and momenta variables. Therefore $D = 3N$ when the system is composed of N particles in a 3-dimensional physical space. Also, we assume that for the system there is a function $H(q, p, t)$ which describes the energy of the system.

The Hamiltonian dynamics is defined by the following equations:

$$\begin{cases} \frac{dq(t)}{dt} = \nabla_p H(q(t), p(t)) \\ \frac{dp(t)}{dt} = -\nabla_q H(q(t), p(t)) \end{cases} \tag{21}$$

With initial condition $p(0) = p^0,\ q(0) = q^0$ that should be provided. Now, introducing the following matrix:

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

And denoting $y = (q, p)$, we can rewrite the Hamiltonian dynamics equations as follows:

$$\frac{dy}{dt} = J\nabla H(y) = J\begin{bmatrix}\nabla_q H(y)\\ \nabla_p H(y)\end{bmatrix} \tag{22}$$

A Hamiltonian function is the sum of the kinetic energy and the potential energy of the system:

$$H(q,p) = K(p) + V(q)$$

A very common physical interpretation of the Hamiltonian function is the following:

$$H(q,p) = V(q) + \frac{1}{2}p^T M^{-1} p$$

Where M is the mass matrix of the system, note that we are supposing that the mass is stable over time. In this case, we can reformulate the Hamiltonian dynamics equations as follows:

$$\begin{cases}\frac{dq(t)}{dt} = M^{-1}p(t)\\ \frac{dp(t)}{dt} = -\nabla_q V(q(t))\end{cases} \tag{23}$$

Therefore, if we consider this equations in terms of positions we get:

$$M\frac{d^2q(t)}{dt^2} = -\nabla_q V(q(t))$$

Which is the Newton's second law of motion. So, we can see that the Hamiltonian dynamics is a generalization of the Newton's second law of motion.

One of the most important property of the Hamiltonian dynamics is the following:

Theorem 2.1 (Conservation of energy)**.** Let $H(q,p)$ be the Hamiltonian function of a system. Then, the energy of the system is conserved over time, i.e. $H(q(t),p(t)) = H(q^0,p^0)$ for all t.

Proof. Deriving the Hamiltonian function with respect to time we get:

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial H}{\partial q}\frac{dq}{dt} + \frac{\partial H}{\partial p}\frac{dp}{dt} =\\ &= \nabla_q H\frac{dq}{dt} + \nabla_p H\frac{dp}{dt} =\\ &= \nabla_q H J \nabla_p H - \nabla_p H J \nabla_q H = 0\end{aligned}$$

Since the hamiltonian H is defined as the total energy of the system, we can conclude that the energy of the system is conserved over time. □

2.2 Symplectic schemes

In this section we will introduce the symplectic schemes, which are a family of numerical methods to solve the Hamiltonian dynamics equations. The idea is to discretize the Hamiltonian dynamics equations (22) in order to obtain a numerical approximation of the solution. The symplectic schemes are a family of numerical methods which preserve the energy conservation property of the Hamiltonian dynamics equations, i.e. the energy of the system is conserved over time.

Definition 2.1. For an open set $U \subset \chi$, a mapping $g : U \to \mathbb{R}^{2D}$ of class C^1 is symplectic if $\nabla g(q,p)$ satisfies

$$(\nabla g)^T J \nabla g = J, \quad \forall (q,p) \in U$$

2.2.1 Analytical solution

Let us introduce a simple example problem to illustrate the symplectic schemes: the harmonic oscillator. The harmonic oscillator is a system composed of a particle of mass m attached to a spring with spring constant k. The position of the particle is denoted by $q(t)$. The potential energy of the system is defined as follows:

$$V(q) = \frac{1}{2}kq^2$$

Therefore, the Hamiltonian function of the system is defined as follows:

$$H(q,p) = \frac{1}{2m}p^2 + \frac{1}{2}kq^2 \tag{24}$$

Applying this Hamiltonian function to the Hamiltonian dynamics equations (22) we obtain the following equations:

$$\begin{cases} \frac{dq(t)}{dt} = \frac{1}{m}p(t) \\ \frac{dp(t)}{dt} = -kq(t) \end{cases}$$

Let's proceed step by step to solve the given system of differential equations for the harmonic oscillator. We are given the Hamiltonian:

$$H(q,p) = \frac{1}{2m}p^2 + \frac{1}{2}kq^2$$

And the Hamilton's equations:

$$\begin{cases} \frac{dq(t)}{dt} = \frac{1}{m}p(t) \\ \frac{dp(t)}{dt} = -kq(t) \end{cases}$$

We can combine the two first-order differential equations into a single second-order differential equation by substituting the expression for $\dot{x}(t)$ into the derivative $\dot{p}(t)$. Substituting $\dot{x}(t) = \frac{1}{m}p(t)$ into the derivative of the second equation gives us:

$$\frac{d^2q(t)}{dt^2} = -\frac{k}{m}q(t)$$

This is a second-order homogeneous linear differential equation.

To solve this second-order differential equation, we can use the characteristic equation method. Assume a solution of the form:

$$q(t) = e^{rt}$$

where r is a constant to be determined. Substituting this into the second-order equation gives:

$$r^2 e^{rt} + \frac{k}{m}e^{rt} = 0$$

Since e^{rt} is never zero, we can divide through by it to get the characteristic equation:

$$r^2 + \frac{k}{m} = 0$$

Solving for r gives us:

$$r = \pm i\sqrt{\frac{k}{m}}$$

where i is the imaginary unit.

Given that the roots are complex, the general solution of the equation is:

$$\begin{cases} q(t) = A\cos(\omega t) + B\sin(\omega t) \\ p(t) = -m\omega A\sin(\omega t) + m\omega B\cos(\omega t) \end{cases}$$

where A and B are arbitrary constants determined by initial conditions and $\omega = \sqrt{\frac{k}{m}}$ is the angular frequency of the oscillator. In order to determine the constants A and B, we need initial conditions. Specifically, we need the initial position $x(0)$ and initial velocity $\dot{x}(0)$. Let's say, for example:

$$q(0) = q_0 \quad \text{and} \quad \dot{q}(0) = v_0$$

Substituting these into the general solution and its derivative gives:

$$q_0 = A \quad \text{and} \quad v_0 = B\omega$$

Thus, if initial conditions are provided, A and B can be determined to give the particular solution for the system. For example, if we take $x_0 = 1$ and $v_0 = 0$, we get that $A = 1$ and $B = 0$. Therefore, the solution of the harmonic oscillator is:

$$\begin{cases} q(t) = \cos(\omega t) = \cos(\sqrt{\frac{k}{m}}t) \\ p(t) = -m\omega\sin(\omega t) = -\sqrt{km}\sin(\sqrt{\frac{k}{m}}t) \end{cases}$$

The Figure 19 shows the phase space of the harmonic oscillator with $k = 1$ and $m = 1$.

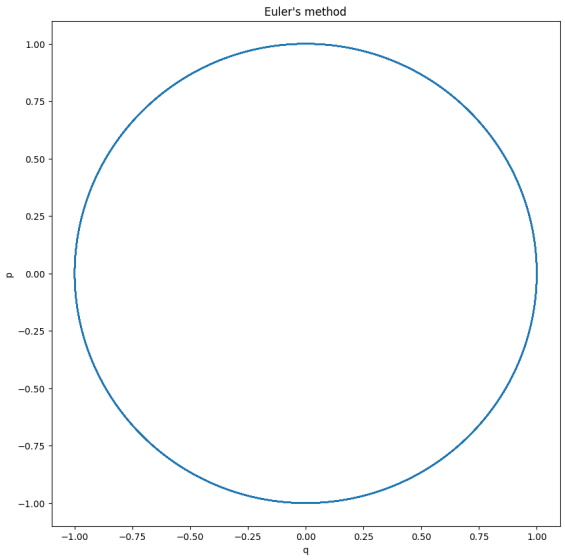

Figure 19: Phase space of the harmonic oscillator with $k = 1$ and $m = 1$ and initial conditions $q_0 = 1$ and $v_0 = 0$.

2.2.2 Euler method

The Euler method is a numerical method to solve ordinary differential equations. The idea is to discretize the differential equation in order to obtain a numerical approximation of the solution. The Euler method discretizes the differential equation as follows:

$$\begin{cases} p^{n+1} &= p^n - \Delta t \frac{\partial H}{\partial q}(q^n, p^n) = p^n - \Delta t \nabla_q V(q^n) \\ q^{n+1} &= q^n + \Delta t \frac{\partial H}{\partial p}(q^n, p^n) = q^n + \Delta t M^{-1} p^n \end{cases} \tag{25}$$

Being Δt the time step of the discretization and $p^n \approx p_{n\Delta t}$. Applying this sequence to the harmonic oscillator, which equations are described in (24), we obtain the following phase space:

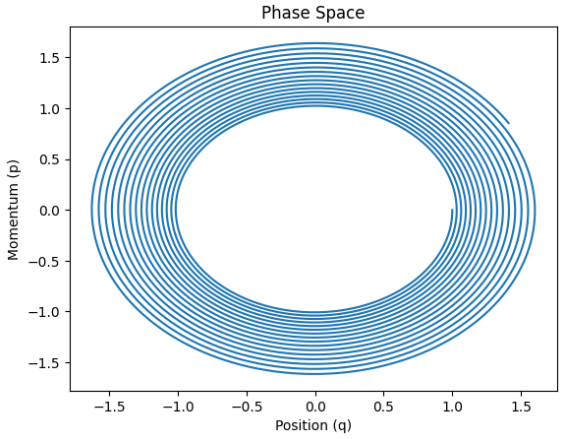

Figure 20: Phase space of the harmonic oscillator using the Euler method with $\Delta t = 0.01$.

As we can discern, this trajectory is not even near to the real trajectory of the harmonic oscillator in ideal conditions. The Figure 20 shows the phase space of the harmonic oscillator using the Euler

method with $\Delta t = 0.1$. We can see that the phase space is not a closed curve, which is the correct behaviour of the harmonic oscillator. Also, the energy is not conserved over time, which is another property of the harmonic oscillator. The Figure 21 shows the energy of the harmonic oscillator using the Euler method with $\Delta t = 0.1$.

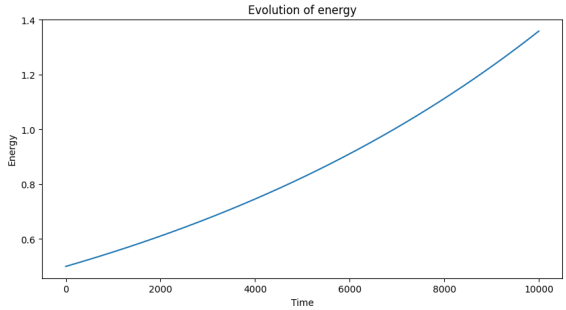

Figure 21: Energy of the harmonic oscillator using the Euler method with $\Delta t = 0.01$.

As we can see, the energy is not conserved over time, which is another property of the harmonic oscillator. The energy is increasing over time, which is not the correct behaviour of the harmonic oscillator. With this results we can conclude that the Euler method is not a good method to solve the Hamiltonian dynamics equations. This is a consequence of the fact that the Euler method does not preserve the symplectic structure of the Hamiltonian dynamics equations. Therefore, in the next section we will introduce a new method that preserves the symplectic structure of the Hamiltonian dynamics equations which is a slight modification of the Euler method.

2.2.3 Failure of standard methods

As we alredy know, the Hamiltonian dynamics is a standard ordinal differential equation (ODE), so it can be approximated by any standard integration scheme. However, as we saw using the Euler scheme, the energy increases over the time so it doesn't work. To provide a further view of this, in this section, we want to introduce the mathematical approach to see why this methods are failing. Considering again the problem of the harmonic oscillator, whose Hamiltonian is provided in 24. And taking the euler scheme given in 25 we can rewrite it to:

$$\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} = A^{n+1} \begin{pmatrix} p^0 \\ q^0 \end{pmatrix}$$

Being $A = \begin{pmatrix} 1 & -k\Delta t \\ \frac{1}{m}\Delta t & 1 \end{pmatrix}$. Now we diagonalize this matrix, so we obtain that its eigenvalues are:

$$\lambda_1 = 1 + i\Delta t\sqrt{\frac{k}{m}} = i + i\Delta t\omega, \quad \lambda_2 = 1 - i\Delta t\sqrt{\frac{k}{m}} = i - i\Delta t\omega$$

Since $\lambda_1\lambda_2 = 1 + (\Delta t\omega)^2 > 1$, we observe that the eigenvalues' magnitudes are greater than one. This indicates that the energy of the system, represented by the norm of the vector (p, q), will grow exponentially over time. To see this, consider that the energy at the n-th step is proportional to the square of the norm of the vector (p^n, q^n). Since each step of the Euler method multiplies the previous vector by A, and the eigenvalues of A have magnitudes greater than one, the norm

of this vector, and hence the energy, will increase exponentially with each step. This exponential growth is a clear indication of the method's failure to conserve energy, a fundamental characteristic of the harmonic oscillator in a closed system. Therefore, the Euler method, along with other similar standard integration schemes, is unsuitable for simulating systems where energy conservation is crucial, necessitating the development of more sophisticated methods that can maintain the energy constant over time.

2.2.4 Constructing Symplectic Schemes

To develop an integrator, we split the Hamiltonian $H(q,p)$ into simpler components:

$$H(q,p) = H_1(q,p) + H_2(q,p)$$

With:

$$H_1(q,p) = \frac{1}{2}p^T M^{-1} p \quad \text{(kinetic energy)}$$
$$H_2(q,p) = V(q) \quad \text{(potential energy)}$$

After the Hamiltonian is decomposed, the associated dynamics become:
For $H1$:

$$\dot{q} = M^{-1}p$$
$$\dot{p} = 0$$

For $H2$:

$$\dot{q} = 0$$
$$\dot{p} = -\nabla V(q)$$

Now, let's talk about flows. Think of a flow as a way to evolve a point in phase space over time based on our differential equations.

For the dynamics associated with H_1, we can integrate with respect to time, t:

$$\int \dot{q} dt = \int M^{-1} p dt$$

This gives:

$$q(t) = q(0) + tM^{-1}p$$

Similarly, for momentum, $p(t) = p(0)$ because $\dot{p} = 0$.

Thus, the flow for H_1, denoted as ϕ_1^t, evolves as:

$$\phi_1^t(q,p) = (q + tM^{-1}p, p)$$

For H_2, given $\dot{q} = 0$, $q(t) = q(0)$. But for momentum, we get:

$$\int \dot{p} dt = \int -\nabla V(q) dt$$

$$p(t) = p(0) - t\nabla V(q)$$

Thus, the flow for H_2, ϕ_2^t, is:

$$\phi_2^t(q,p) = (q, p - t\nabla V(q))$$

Combining these flows yields our symplectic schemes. The sequence of combining matters.
1. Kinetic (ϕ_1) followed by Potential (ϕ_2):

$$\begin{cases} q_{n+1} &= q_n + \Delta t M^{-1} p_n \\ p_{n+1} &= p_n - \Delta t \nabla V(q_{n+1}) \end{cases}$$

2. Potential (ϕ_2) followed by Kinetic (ϕ_1):

$$\begin{cases} p_{n+1} &= p_n - \Delta t \nabla V(q_n) \\ q_{n+1} &= q_n + \Delta t M^{-1} p_{n+1} \end{cases}$$

Theorem 2.2 (Symplecity of the Hamiltonian flow)**.** Let $H(q,p)$ be a $C^2(U)$ function, where U is an open set of $\mathbb{R}^{2D}$. Then, for any fixed $t \in \mathbb{R}$ such that the flow ϕ^t is defined, the mapping ϕ^t is symplectic.

Proof. Proof in [14, Chapter 2.1.2]. □

2.2.5 Symplectic Euler method

The symplectic Euler method is a slight modification of the Euler method that preserves the symplectic structure of Hamiltonian systems. This new method discretizes the Hamiltonian dynamics equations (22) as follows:

$$\begin{cases} p^{n+1} &= p^n - \Delta t \frac{\partial H}{\partial q}(q^n, p^n) = p^n - \Delta t \nabla_q V(q^n) \\ q^{n+1} &= q^n + \Delta t \frac{\partial H}{\partial p}(q^{n+1}, p^{n+1}) = q^n + \Delta t M^{-1} p^{n+1} \end{cases} \tag{26}$$

Being Δt the time step of the discretization. Applying this sequence to the harmonic oscillator, which equations are described in (24), we obtain the following phase space:

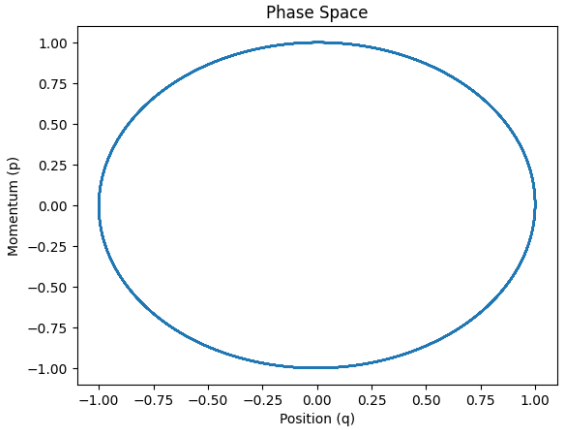

Figure 22: Phase space of the harmonic oscillator using the symplectic Euler method with $\Delta t = 0.01$.

The Figure 22 shows the phase space of the harmonic oscillator using the symplectic Euler method. We can see that the trajectory in the phase space is a closed curve, which is the correct behaviour of the harmonic oscillator.

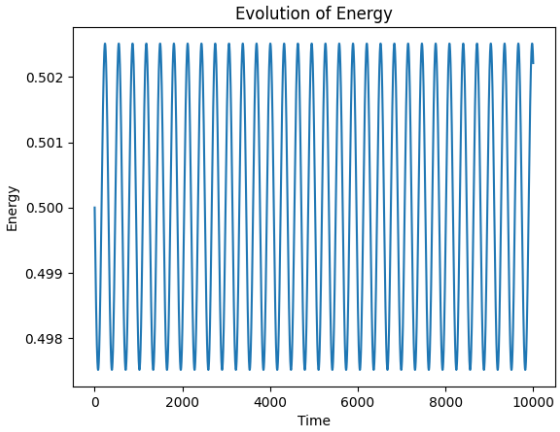

Figure 23: Energy of the harmonic oscillator using the symplectic Euler method with $\Delta t = 0.01$.

The Figure 23 shows the energy of the harmonic oscillator using the symplectic Euler method. We can see that the energy is approximately conserved over time, as we wanted for the Hamiltonian dynamics. Note that is not a straight line but the average of the energy is the same, so we can conclude that the energy is conserved over time.

Now, if we define the next map:

$$\Gamma_{\Delta t}^{Euler} = \max_{n \in \mathbb{N}} \{|H_{\Delta t}(p^n, q^n) - H_{\Delta t}(p^0, q^0)|\}$$

The Figure 24 shows the function $\Gamma_{\Delta t}^{Euler}$ over $\frac{1}{\Delta t}$.

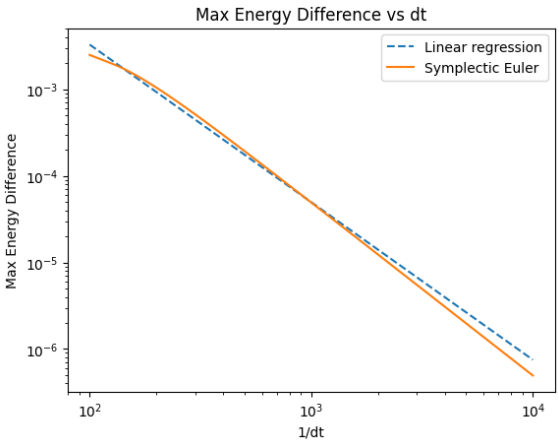

Figure 24: Function $\Gamma_{\Delta t}^{Euler}$ over $\frac{1}{\Delta t}$

To conclude the section, we can see in Figure 24 that the difference decreases as the time step decreases, which is the expected behaviour of any scheme that preserves the symplectic structure of the Hamiltonian dynamics equations. Therefore, we can conclude that the symplectic Euler method

is a good method to solve the Hamiltonian dynamics equations.

2.2.5.1 Linear Stability Analysis of the Sympletic Euler Scheme

Using the equation 26 and replacing all the termusing the harmonic oscillator problem, one can rewrite the scheme as:

$$\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} = A^{n+1} \begin{pmatrix} p^0 \\ q^0 \end{pmatrix}$$

With

$$A = \begin{pmatrix} 1 & -\Delta t k \\ \frac{\Delta t}{m} & 1 - \frac{\Delta t^2 k}{m} \end{pmatrix}$$

The eigenvalues are:

$$\lambda_1 = \frac{-\Delta t^2 k - \Delta t \sqrt{k(\Delta t^2 k - 4m)}}{2m} + 1$$
$$\lambda_2 = \frac{-\Delta t^2 k + \Delta t \sqrt{k(\Delta t^2 k - 4m)}}{2m} + 1$$

The fact that $\lambda_1 \lambda_2 = 1$ indicates that the energy of the system is conserved over time.

2.2.6 Stormer-Verlet method

The Störmer-Verlet method, often simply referred to as the Verlet method, is a numerical technique used to integrate ordinary differential equations of the form $\dot{y} = f(y)$. It is particularly popular in molecular dynamics simulations and other problems modeled by Hamiltonian systems. The method can be derived directly from the Taylor series expansion of the solution. The central idea behind the method is to use information from both the current and previous time steps to predict the value at the next time step.

The Störmer-Verlet method applied to the Hamiltonian equations is as follows (using dHdq and dHdp to denote the partial derivatives of H with respect to q and p, respectively):

$$\begin{cases} p^{n+\frac{1}{2}} = p^n - \frac{\Delta t}{2} \nabla_q V(q^n) \\ q^{n+1} = q^n + \Delta t M^{-1} p^{t+1/2} \\ p^{n+1} = p^{n+\frac{1}{2}} - \frac{\Delta t}{2} \nabla_q V(q^{n+1}) \end{cases} \tag{27}$$

The numerical flow of this scheme is noted as ϕ_t^{Verlet}. In matter of fact, note that:

$$\Phi_{\Delta t}^{Verlet} = \phi_{\Delta t/2}^2 \circ \phi_{\Delta t}^1 \circ \phi_{\Delta t}^2$$

Therefore, it is easy to prove that, by construction, is a sympletic scheme. So given the problem of the harmonic oscillator using the Störmer-Verlet method, the update equations become:

$$\begin{cases} p^{n+\frac{1}{2}} = p^n - \frac{\Delta t}{2} k q^n \\ q^{n+1} = q^n + \Delta t M^{-1} p^{t+1/2} \\ p^{n+1} = p^{n+\frac{1}{2}} - \frac{\Delta t}{2} k q^{n+1} \end{cases} \tag{28}$$

The Figure 25 shows the phase space of the harmonic oscillator using the Störmer-Verlet method.

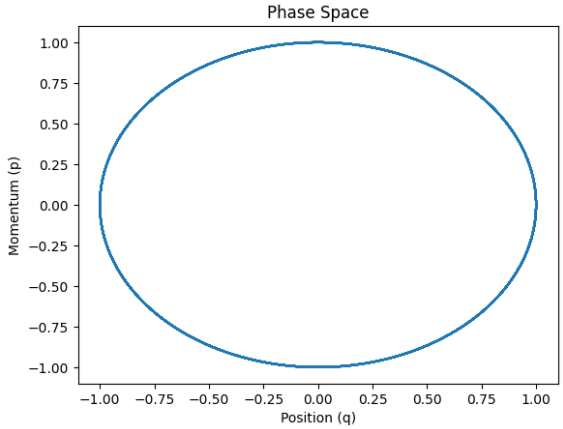

Figure 25: Phase space of the harmonic oscillator using the Störmer-Verlet method with $\Delta t = 0.01$.

The Figure 25 shows the phase space of the harmonic oscillator using the Störmer-Verlet method. We can see that the trajectory in the phase space is a closed curve, which is the correct behaviour of the harmonic oscillator.

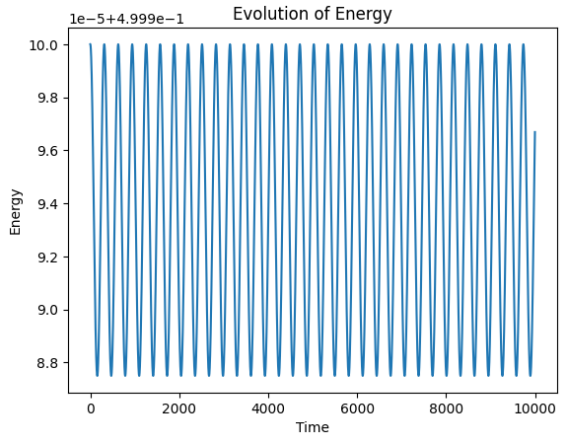

Figure 26: Energy of the harmonic oscillator using the Störmer-Verlet method with $\Delta t = 0.01$.

The Figure 26 shows the energy of the harmonic oscillator using the Störmer-Verlet method. We can see that the energy is approximately conserved over time, as we wanted for the Hamiltonian dynamics. Like the behaviour of the Euler Sympletic method, we can see that in this case, the energy continues to oscillate around the average value, however the scale of this oscillation is much smaller than in the Euler Sympletic method. To see this, we define the following function:

$$\Gamma_{\Delta t}^{Verlet} = \max_{h=1...N_{\Delta t}} \{|H(q^n, p^n) - H(q^0, p^0)|\}$$

So now we are prepared to overview the difference between both methods. The Figure 27 shows the function $\Gamma_{\Delta t}^{Euler}$ and $\Gamma_{\Delta t}^{Verlet}$ over $\frac{1}{\Delta t}$.

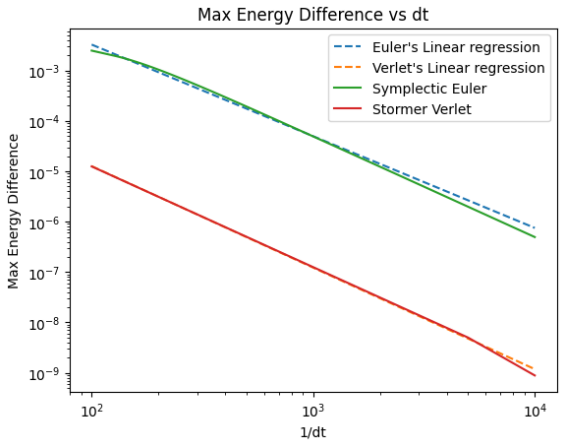

Figure 27: Function $\Gamma_{\Delta t}^{Euler}$ and $\Gamma_{\Delta t}^{Verlet}$ over $\frac{1}{\Delta t}$

As we can see, the Störmer-Verlet method is much better than the Euler Sympletic method. Note that the slope of the linear regression of the Störmer-Verlet method is approximately -2.0147 and the slope of the linear regression of the Euler Sympletic method is approximately -1.8196 Probably, this is a consequence of the fact that the Störmer-Verlet method is a second order method, while the Euler Sympletic method is a first order method. Therefore, we can conclude that the Störmer-Verlet method is a better method to solve the Hamiltonian dynamics equations than the Euler Sympletic method.

2.2.6.1 Linear Stability Analysis of the Verlet Scheme

Using the equation 28 and replacing all the terms by its definition, and considering that $\omega = \sqrt{\frac{k}{m}}$, we can rewrite the scheme as:

$$\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} = A \begin{pmatrix} p^n \\ q^n \end{pmatrix}$$

With

$$A = \begin{pmatrix} 1 - \frac{(\omega\Delta t)^2}{2} & \Delta t \\ -\omega^2 \Delta t (1 - \frac{(\omega\Delta t)^2}{4}) & 1 - \frac{(\omega\Delta t)^2}{2} \end{pmatrix}$$

Noting $\xi = \frac{(\omega\Delta t)^2}{2}$, the eigenvalues are the solutions of λ in:

$$(1 - \xi - \lambda)^2 + \xi(2 - \xi) = 0$$

Therefore, the eigenvalues are:

$$\begin{cases} \lambda_1 = 1 - \xi + i\sqrt{\xi(2-\xi)} & \lambda_2 = 1 - \xi - i\sqrt{\xi(2-\xi)} \;\; if\ \xi(2-\xi) \geq 0 \\ \lambda_1 = 1 - \xi + \sqrt{\xi(\xi-2)} & \lambda_2 = 1 - \xi - \sqrt{\xi(\xi-2)} \;\; if\ \xi(2-\xi) \leq 0 \end{cases}$$

In linear stability analysis, a scheme is said to be stable if the eigenvalues of the amplification matrix lie inside the unit circle in the complex plane. This means that their absolute values should be less than or equal to 1 for all possible values of ξ. This ensures that errors do not grow unboundedly as we march forward in time.

Given the eigenvalues derived above:

$$\lambda_1 = \begin{cases} 1 - \xi + i\sqrt{\xi(2-\xi)} \text{ if } \xi(2-\xi) \geq 0 \\ 1 - \xi + \sqrt{\xi(\xi-2)} \text{ if } \xi(2-\xi) \leq 0 \end{cases}$$

$$\lambda_2 = \begin{cases} 1 - \xi - i\sqrt{\xi(2-\xi)} \text{ if } \xi(2-\xi) \geq 0 \\ 1 - \xi - \sqrt{\xi(\xi-2)} \text{ if } \xi(2-\xi) \leq 0 \end{cases}$$

The eigenvalues are complex conjugates. Their magnitudes are:

$$|\lambda_1| = |\lambda_2| = \sqrt{(1-\xi)^2 + \xi(2-\xi)}$$

For stability, we require $|\lambda_1| \leq 1$. This translates to the condition:

$$(1-\xi)^2 + \xi(2-\xi) \leq 1$$

This inequality leads to a tautology so it is always stable. This is consistent with the observed property of the Störmer-Verlet scheme which conserves the energy over time. This, in combination with its second-order accuracy, makes it preferable over the Euler sympletic method for problems modeled by Hamiltonian systems.

2.2.7 Backward Error Analysis

In this section, we will introduce the backward error analysis. The backward error analysis is a technique to analyze the error of a numerical method. We can define the idea of this technique as: For a given problem $P(x)$, an exact input x and an approximate solution $\hat{x}$, we can define the backward error as the smallest perturbation Δx such that $P(x + \Delta x) = \hat{x}$. Therefore the formal definitions, extracted with little modifications from [7], is:

Definition 2.2 (Backward error)**.** Let $\hat{x}$ be an approximate solution to the equation $P(d) = x$. Then the backward error of $\hat{x}$ is defined by:

$$\eta = \min\{\epsilon : P(x + \Delta x) = \hat{x}, ||\Delta x|| \leq \epsilon\}$$

The following Figure 28 extracted from [8] illustrated the idea of the backward error analysis.

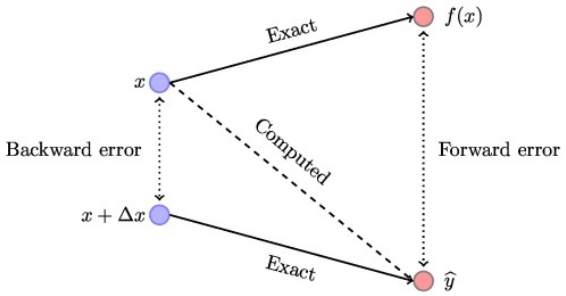

Figure 28: Backward error analysis.

2.3 Solar System Simulation

In this section, we'll harness the power of Hamiltonian dynamics to craft a simulation of the solar system. We'll adopt the Stormer-Verlet scheme, renowned for its energy-conserving properties, as our primary numerical method.

First of all, let's do a brief introduction to the problem we want to solve. We can see the solar system as a set of N bodies, each of them with a mass m_i a position q_i and a velocity v_i. The position and velocity are vectors in $\mathbb{R}^2$. The gravitational force between two bodies is given by:

2.3.1 Modeling the Solar System: Gravitational Dynamics

The motion of celestial bodies in the solar system, primarily planets, moons, and the Sun, can be described by Newton's law of universal gravitation. For N bodies, the gravitational force exerted on the i^{th} body due to the j^{th} body is given by:

$$\mathbf{F}_{ij} = -\frac{Gm_i m_j(\mathbf{q}_i - \mathbf{q}_j)}{|\mathbf{q}_j - \mathbf{q}_i|^3}$$

where G is the gravitational constant, approximately 6.67430×10^{-11} Nm^2/kg^2, $\mathbf{q}_i$ and $\mathbf{q}_j$ are the position vectors of the i^{th} and j^{th} bodies respectively and m_i and m_j are the masses of the i^{th} and j^{th} bodies respectively. Therefore, the total gravitational force acting on the i^{th} body due to all other bodies is:

$$\mathbf{F}_i = \sum_{j=1, j\neq i}^{N} \mathbf{F}_{ij}$$

To use Hamiltonian mechanics and the Störmer-Verlet scheme, we can represent the system in terms of its Hamiltonian, which is the sum of its kinetic and potential energies. For the i^{th} body:

1. Kinetic Energy, T_i:

$$T_i = \frac{1}{2} m_i \mathbf{v}_i \cdot \mathbf{v}_i$$

where $\mathbf{v}_i$ is the velocity of the i^{th} body.

2. Potential Energy, U_{ij}, due to the interaction between the i^{th} and j^{th} bodies:

$$U_{ij} = -\frac{Gm_i m_j}{|\mathbf{q}_j - \mathbf{q}_i|}$$

The total Hamiltonian H for the system is the sum of the kinetic and potential energies for all body pairs:

$$H = \sum_{i=1}^{N} T_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} U_{ij}$$

This Hamiltonian can be used to derive the equations of motion using Hamilton's equations, which can then be solved using the Störmer-Verlet scheme.

In the next section, we'll dive deeper into the implementation details, but this sets up the mathematical foundation for our solar system simulation based on gravitational interactions.

2.3.2 Störmer-Verlet Scheme for Gravitational Dynamics

Now that we have the Hamiltonian for the system, we can derive use it inside the Störmer-Verlet scheme to derive the equations of motion. The Störmer-Verlet scheme for the i^{th} body is:

$$\begin{cases} \mathbf{p}_i^{n+\frac{1}{2}} = \mathbf{p}_i^n - \frac{\Delta t}{2} \nabla_{\mathbf{q}_i} U(\mathbf{q}_i^n) \\ \mathbf{q}_i^{n+1} = \mathbf{q}_i^n + \Delta t M^{-1} \mathbf{p}_i^{n+1/2} \\ \mathbf{p}_i^{n+1} = \mathbf{p}_i^{n+\frac{1}{2}} - \frac{\Delta t}{2} \nabla_{\mathbf{q}_i} U(\mathbf{q}_i^{n+1}) \end{cases}$$

where $U(\mathbf{q}_i)$ is the potential energy of the i^{th} body due to all other bodies. The total potential energy of the system is the sum of the potential energies of all body pairs:

$$U = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} U_{ij}$$

where U_{ij} is the potential energy of the i^{th} body due to the j^{th} body. Therefore, the gradient of the potential energy of the i^{th} body is:

$$\nabla_{\mathbf{q}_i} U(\mathbf{q}_i) = \nabla_{\mathbf{q}_i} U(\mathbf{q}_i) = \sum_{j=1, j \neq i}^{N} \nabla_{\mathbf{q}_i} U_{ij}(\mathbf{q}_i, \mathbf{q}_j)$$

Given that:

$$U_{ij} = -\frac{G m_i m_j}{|\mathbf{q}_j - \mathbf{q}_i|}$$

We can compute the gradient of U_{ij} with respect to $\mathbf{q}_i$. Firstly, we can express the distance $|\mathbf{q}_j - \mathbf{q}_i|$ as r_{ij}:

$$r_{ij} = |\mathbf{q}_j - \mathbf{q}_i|$$

Thus, the potential energy becomes:

$$U_{ij} = -\frac{G m_i m_j}{r_{ij}}$$

To find the gradient of U_{ij} with respect to $\mathbf{q}_i$, we need to differentiate with respect to $\mathbf{q}_i$:

$$\nabla_{\mathbf{q}_i} U_{ij} = \frac{\partial U_{ij}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{q}_i}$$

First, differentiate U_{ij} with respect to r_{ij}:

$$\frac{\partial U_{ij}}{\partial r_{ij}} = G m_i m_j \frac{1}{{r_{ij}}^2}$$

Next, differentiate r_{ij} with respect to $\mathbf{q}_i$:

$$r_{ij} = \sqrt{(\mathbf{q}_j - \mathbf{q}_i) \cdot (\mathbf{q}_j - \mathbf{q}_i)}$$

$$\frac{\partial r_{ij}}{\partial \mathbf{q}_i} = \frac{\mathbf{q}_j - \mathbf{q}_i}{r_{ij}}$$

Putting it all together:

$$\nabla_{\mathbf{q}_i} U_{ij} = G m_i m_j \frac{1}{{r_{ij}}^2} \frac{\mathbf{q}_i - \mathbf{q}_j}{r_{ij}} = G m_i m_j \frac{\mathbf{q}_i - \mathbf{q}_j}{{r_{ij}}^3}$$

Thus, the gradient of the potential energy U_{ij} with respect to $\mathbf{q}_i$ is:

$$\nabla_{\mathbf{q}_i} U_{ij} = -G m_i m_j \frac{\mathbf{q}_i - \mathbf{q}_j}{|\mathbf{q}_j - \mathbf{q}_i|^3}$$

Now, plugging this back into the equation for the gradient of the total potential energy with respect to $\mathbf{q}_i$:

$$\nabla_{\mathbf{q}_i} U(\mathbf{q}_i) = \sum_{j=1, j \neq i}^{N} -G m_i m_j \frac{\mathbf{q}_i - \mathbf{q}_j}{|\mathbf{q}_j - \mathbf{q}_i|^3}$$

This is the gradient of the potential energy with respect to the position of the i^{th} body, taking into account all the interactions with other bodies in the system. Before executing the Störmer-Verlet scheme, let's show the potential energy of the system. The Figure 29 shows the potential energy of the system.

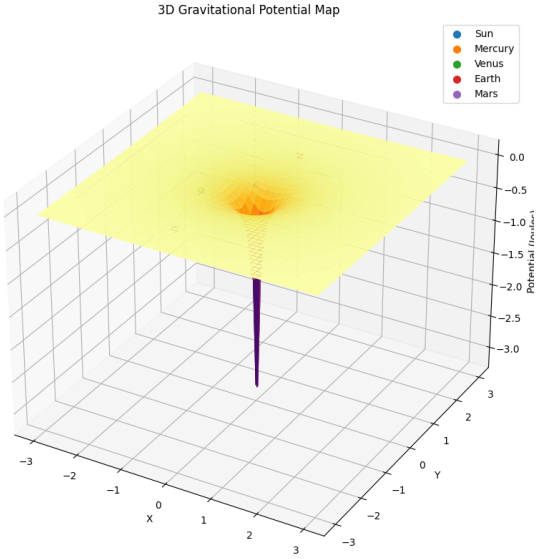

Figure 29: Potential energy of the part of the solar system.

As we could have expected, the Sun has, by far, the most negative potential energy, since it is the body with the highest mass. Also we can observe that the potential energy is negative, which is the expected behaviour of the gravitational potential energy.

2.3.3 Implementation of the Solar System Simulation

First of all, we need to define the parameters of the simulation. The Table 1 shows the parameters of the simulation.

Celestial Body	Mass (Relative to Sun)	Position (AU)	Mean velocity (AU/Year)
Sun	1.0000	0.000	0.00
Mercury	1.6505e-7	0.390	9.99
Venus	2.4335e-6	0.720	7.38
Earth	2.9860e-6	1.000	6.28
Mars	3.2085e-7	1.520	5.08
Jupiter	9.4950e-4	5.187	2.76

Table 1: Data of Celestial Bodies in the Solar System

Therefore, the units that we have used are:

- Mass: $M_{\odot}$ (Solar Mass)
- Distance: AU (Astronomical Unit)
- Time: Earth year

This is because we want to do a simulation of the solar system inside a machine that uses discrete numbers, therefore if we use the SI units, we could get some troubles with the precision of the

machine, while if we use the units that we have used, we get numbers that are not too big or too small, so we can get a good precision. Also, using this numbers we get that the gravitational constant is:

$$G = 4\pi^2 \, \frac{\text{AU}^3}{\text{year}^2 \, M_\odot}$$

Note that we have used the real masses of the planets and the Sun, and the real distance between them. Therefore, we can see that the simulation is very similar to the real solar system. All of these parameters can be found on the NASA website [1]. Now we are prepared to implement the simulation.

For illustrative purposes, we'll simulate the orbits of the Sun, Mercury, Venus, Earth and Mars. The Figure 30 shows the orbits of the planets of the solar system during 365 days and with a time step of 1 day.

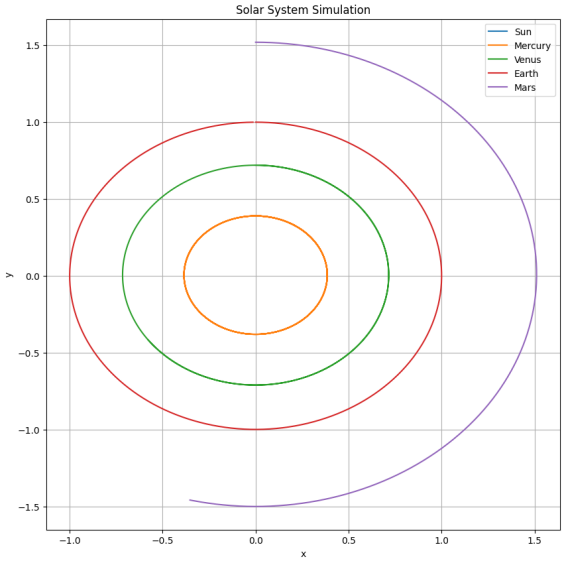

Figure 30: Simulation of the orbits of part of the planets of the solar system using the Störmer-Verlet scheme for $\Delta t = 1$ day and $T = 365$ days.

We can see that the orbits are closed curves, which is the expected behaviour of the planets of the solar system. Also, we can see that the orbits are not perfect circles, which is also the expected behaviour of the planets of the solar system. Note that for illustrative purposes, since it is known that the Earth has an orbit of 365.25 days, we have simulated the orbits for 365 days. If we had plotted also other planets like Jupiter, the plot would have been very messy, so we have decided to plot only the inner planets.

Now let's see the energy of the system. The Figure 31 shows the energy of the system for the same simulation.

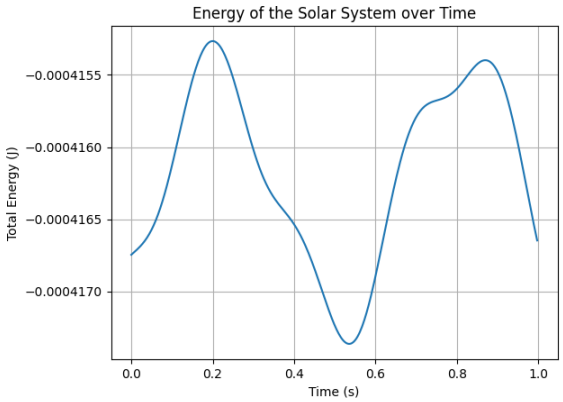

Figure 31: Energy of the simulation of part of the planets of the solar system using the Störmer-Verlet scheme for $\Delta t = 1$ day and $T = 365$ days.

We can see that the energy is approximately conserved over time, as we wanted for the Hamiltonian dynamics. Note that is not a straight line but the average of the energy is the same and we can see a slight of oscillations that may be repeated in the future if the simulation is done for more time. So, let's do the simulation for 20 years. The Figure 32 shows the energy of the system for the same simulation.

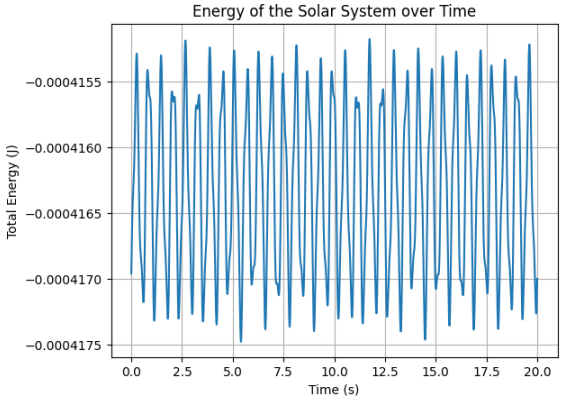

Figure 32: Energy of the simulation of part of the planets of the solar system using the Störmer-Verlet scheme for $\Delta t = 1$ day and $T = 20$ years.

As predicted, we can observe that the energy is approximately preserved.

Now that we have the system created and we have seen that the energy is preserved, we can do some experiments.

2.3.3.1 Experiment 1: Expanding to the N-Body Problem

In our initial simulation, as detailed in the referenced table, we included data for Jupiter but did not incorporate it into the actual simulation. This was primarily due to Jupiter's significant distance from the inner planets, which could potentially complicate the visual clarity of our model. However, given Jupiter's substantial mass – the largest among the planets – its inclusion is crucial for a more comprehensive understanding of the solar system's dynamics. This necessitates an expansion of our

model to an N-body problem, allowing for the gravitational influence of not just the Sun but also the other planets to be accounted for.

To explore this, we have extended our simulation to include Jupiter, aiming to observe its impact on the orbital paths of the other planets. The updated Figure 33 illustrates the orbits of the solar system's planets over a period of 20 years with a daily time step, now incorporating Jupiter. This addition provides a more realistic and complex representation of the gravitational interplay within our solar system.

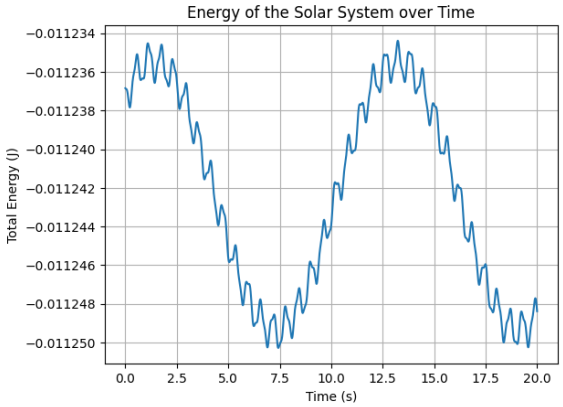

Figure 33: Simulation of the orbits of the planets of the solar system using the Störmer-Verlet scheme for $\Delta t = 1$ day and $T = 20$ years, now including Jupiter.

2.3.3.2 Experiment 2: Using Symplectic Euler Method

We alredy know that the Störmer-Verlet method performs better than the Euler Symplectic method, but let's see how the solar system behaves using the Euler Sympletic method. The Figure 34 shows the orbits of the planets of the solar system using the Euler Symplectic method for $\Delta t = 1$ day and $T = 20$ years.

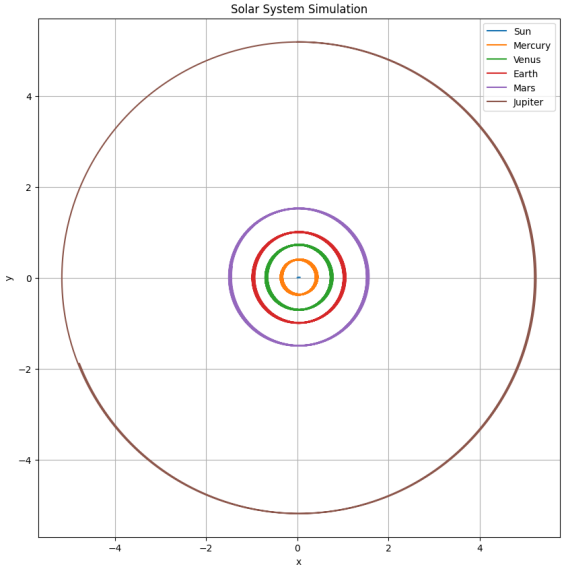

Figure 34: Simulation of the orbits of the planets of the solar system using the Euler Sympletic method for $\Delta t = 1$ day and $T = 20$ years.

Also we can plot the energy of the system. The Figure 35 shows the energy of the system for the same simulation.

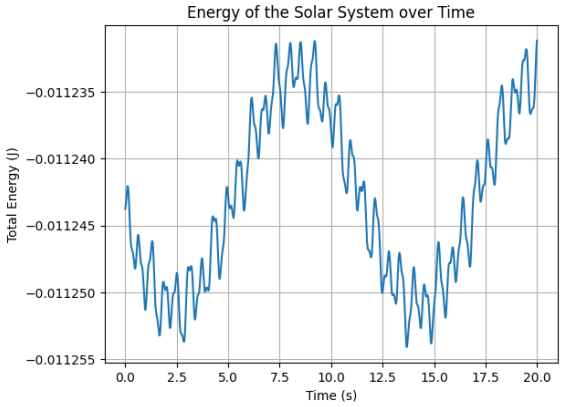

Figure 35: Energy of the simulation of the orbits of the planets of the solar system using the Euler Sympletic method for $\Delta t = 1$ day and $T = 20$ years.

Comparing it to the results obtained using the Störmer-Verlet method, again we can see that the first one is better in the following table:

Method	$\max H(x) - \min H(x)$	$\max \lvert H(X) - \mathbb{E}(H(X)) \rvert$
Symplectic Euler	2.43e-5	1.26e-5
Störmer-Verlet	1.59e-5	9.04e-6

Table 2: Data of Celestial Bodies in the Solar System

Therefore, in this problem we can conclude that the Störmer-Verlet method is better in terms of

conservation of the Energy than the Euler Sympletic method.

3 Markov Chains

This sections aims to show how Markov Chains are extremely useful in the field of Hamiltonian Dynamics. For this we will introduce the concept of Markov Chains and its properties, then we will explain the Markov Chain Montecarlo Methods (MCMC) and finally we will show how to use MCMC to sample from a distribution.

3.1 Introduction to Markov Chains

Let us start by introducing the concept of *Markov Chain*. A Markov Chain is a stochastic process that satisfies the Markov property. The Markov property states that the conditional probability distribution for the state at the next step depends only on the current state and not on the previous states. The Markov property can be expressed mathematically as:

$$\mathbb{P}(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x | X_n = x_n)$$

for any states $x, x_1, x_2, \ldots, x_n$.

A Markov Chain is usually characterised by the matrix of transitions, which is the matrix of the probabilities of transitioning between states of a Markov Chain. The Markov Chain is a square matrix $K(x, y)$ such that $K(x, y) \geq 0$ and $\sum_y K(x, y) = 1$, where the rows and columns are indexed by the states of the Markov Process. The entry in the i^{th} row and j^{th} column of the Markov Chain is the probability of transitioning from state i to state j.

3.1.1 Stationary Distribution

The final objective of the Chapter 3 is to introduce the Markov Chain Montecarlo Methods (MCMC). As first step to introduce MCMC, we need to introduce the concept of stationary distribution.

Definition 3.1 (Stationary Distribution)**.** Let $(X_n)_{n\geq 0}$ be a Markov Chain in E with transition matrix P. A probability measure pi on E is called a *stationary distribution* for $(X_n)_{n\geq 0}$ if it satisfies

$$\pi P = \pi$$

The denomination of stationary distribution comes from the following result.

Proposition 3.1. Let π be a stationary distribution for $(X_n)_{n\geq 0}$. For any $n \geq 0$,

$$\mathbb{P}_\pi(X_n = y) = \pi(y) \quad for\ all\ x \in E$$

in other words, π is a stationary distribution if and only if the distribution of X_n is π for all $n \geq 0$.

The proof can be found in [6].

Theorem 3.1 (Fundamental Theorem of Markov Chains)**.** Let χ be a finite set and $K(x, y)$ a Markov Chain indexed by χ. If there is $n_0 \in \mathbb{N}$ such that $K^n(x, y) \geq 0$ for all $n > n_0$, then K has a unique stationary distribution π such that

$$K^n(x,y) \xrightarrow{n\to\infty} \pi(y) \quad for\ each\ x,y \in \chi$$

This theorem has been extracted from [13].

3.1.2 Irreducibility and Uniqueness

One may ask if the stationary distribution is unique. The answer is no, it is not unique. However, there exists a condition that guarantees the uniqueness of the stationary distribution. This condition is called irreducibility.

Definition 3.2. A Markov Chain is said to be *irreducible* if for any two states x and y, there exists $n \geq 0$ such that $K^n(x,y) > 0$.

Proposition 3.2. Let K be a Markov Chain with a finite state space χ. If K is irreducible, then it has a unique stationary distribution π.

3.2 Markov Chain Montecarlo Methods

In this section, we will introduce the Markov Chain Montecarlo Methods (MCMC). The MCMC are a class of algorithms for sampling from a probability distribution. The idea of MCMC is to construct a Markov Chain that has the desired distribution as its stationary distribution. Therefore, the Markov Chain will converge to the desired distribution as the number of steps increases.

3.2.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is a MCMC algorithm for sampling from a probability distribution. The algorithm is defined as follows:

Algorithm 6: Metropolis-Hastings

1. Choose an initial state x_0
2. For $n = 0, 1, 2, \ldots$
   a. Sample y from $q(x_n, \cdot)$
   b. Compute the acceptance probability $\alpha(x_n, y)$
   c. Sample u from $\mathcal{U}(0,1)$
   d. If $u < \alpha(x_n, y)$, set $x_{n+1} = y$
   e. Else, set $x_{n+1} = x_n$
3. Return $(x_n)_{n\geq 0}$

where $q(x,\cdot)$ is a probability distribution on χ and $\alpha(x,y)$ is the acceptance probability defined as:

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right)$$

where π is the desired distribution. The proof of the correctness of the algorithm can be found in [3].

3.2.1.1 Example: Sampling from a Gamma Distribution

Let's see an example of the Metropolis-Hastings algorithm. We will use the Metropolis-Hastings algorithm to sample from a Gamma distribution. The Gamma distribution is a two-parameter family of continuous probability distributions. The probability density function of the Gamma distribution is:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $\alpha, \beta > 0$ and $\Gamma(\alpha)$ is the Gamma function. The Gamma distribution is often used to model the time until an event occurs. For example, the time until a machine part fails, the time until a queue empties, the time until a customer arrives, etc. The Gamma distribution is also used to model the size of insurance claims and rainfall. Executing the algorithm with the following parameters: $\alpha = 2$ and $\beta = 1$ and $q(x, \cdot) = \mathcal{N}(x, \delta t)$ with $\delta t = 1$, we get the following results:

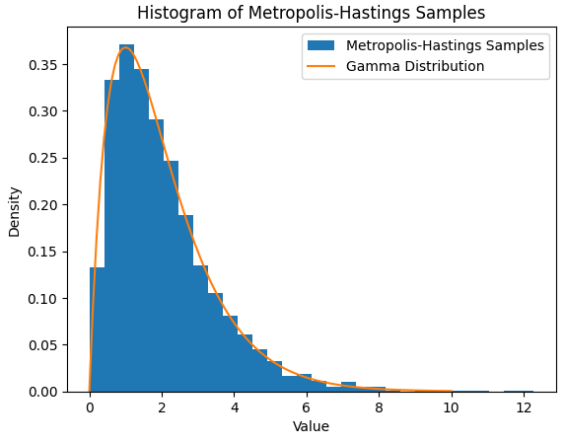

Figure 36: Sampling from a Gamma distribution using the Metropolis-Hastings algorithm with $\alpha = 2$ and $\beta = 1$ and $q(x, \cdot) = \mathcal{N}(x, \delta t)$ with $\delta t = 1$.

We can see that the samples are distributed according to the Gamma distribution.

3.2.2 Gibbs Sampling

The Gibbs sampling is a MCMC algorithm for sampling from a probability distribution. The algorithm is defined as follows:

Algorithm 7: Gibbs Sampling

1. Choose an initial state x_0
2. For $n = 0, 1, 2, \ldots$
   a. Sample $x_{n+1}^{(1)}$ from $p(x_1 | x_n^{(2)}, x_n^{(3)}, \ldots, x_n^{(d)})$
   b. Sample $x_{n+1}^{(2)}$ from $p(x_2 | x_{n+1}^{(1)}, x_n^{(3)}, \ldots, x_n^{(d)})$
   c. $\ldots$
   d. Sample $x_{n+1}^{(d)}$ from $p(x_d | x_{n+1}^{(1)}, x_{n+1}^{(2)}, \ldots, x_{n+1}^{(d-1)})$
3. Return $(x_n)_{n \geq 0}$

where $p(x_i|x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ is the conditional probability distribution of x_i given all other variables.

3.2.2.1 Example: Sampling from a Bivariate Normal Distribution

Let's see an example of the Gibbs sampling algorithm. We will use the Gibbs sampling algorithm to sample from a bivariate normal distribution. The bivariate normal distribution is a probability distribution of two random variables X and Y that satisfy the following conditions:

$$\begin{cases} X \sim \mathcal{N}(\mu_x, \sigma_x^2) \\ Y \sim \mathcal{N}(\mu_y, \sigma_y^2) \\ X \perp Y \end{cases}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the parameters of the distribution. The probability density function of the bivariate normal distribution is:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

where ρ is the correlation coefficient between X and Y. The Figure 38 shows how the bivariate normal distribution looks like, this picture is extracted from [4].

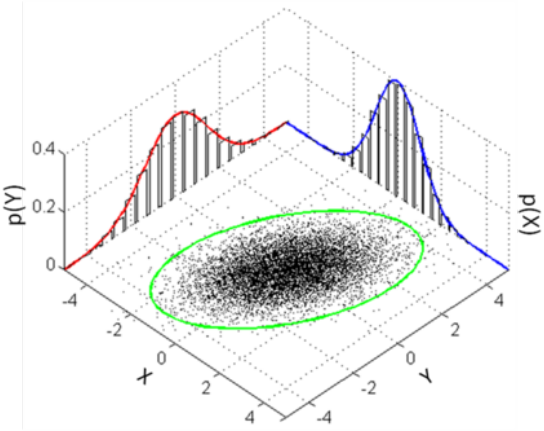

Figure 37: Bivariate normal distribution.

Executing the algorithm with the following parameters: $\mu_x = 0$, $\mu_y = 0$, $\sigma_x = 1$, $\sigma_y = 1$ and $\rho = 0.5$, we get the following results:

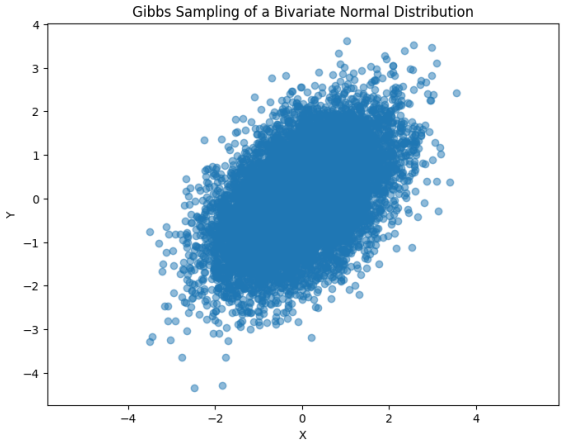

Figure 38: Sampling from a bivariate normal distribution using the Gibbs sampling algorithm with $\mu_x = 0$, $\mu_y = 0$, $\sigma_x = 1$, $\sigma_y = 1$ and $\rho = 0.5$.

We can see that the samples are distributed according to the bivariate normal distribution.

3.2.3 Müller-Brown Potential

In this section, we will introduce the Müller-Brown potential. The Müller-Brown potential is a function of two variables that is often used as a test function for optimization algorithms. The Müller-Brown potential is defined as:

$$U(x_1, x_2) = s \sum_{k=1}^{4} A_k \exp(\alpha_k (x_1 - a_k)^2 + \beta_k (x_1 - a_k)(x_2 - b_k) + \gamma_k (x_2 - b_k)^2)$$

where A_k is the height of the k^{th} Gaussian, (a_k, b_k) is the location of the k^{th} Gaussian and $(\alpha_k, \beta_k, \gamma_k)$ is the shape of the k^{th} Gaussian. In our case, the parameters that we will use are:

$$\begin{cases} (A_1, A_2, A_3, A_4) = (-200, -100, -170, 15) \\ (\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (-1, -1, -6.5, 0.7) \\ (\beta_1, \beta_2, \beta_3, \beta_4) = (0, 0, 11, 0.6) \\ (\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (-10, -10, -6.5, 0.7) \\ (a_1, a_2, a_3, a_4) = (1, 0, -0.5, -1) \\ (b_1, b_2, b_3, b_4) = (0, 0.5, 1.5, 1) \\ s = 0.05 \end{cases}$$

The Figure 39 shows how the Müller-Brown potential looks like with these parameters.

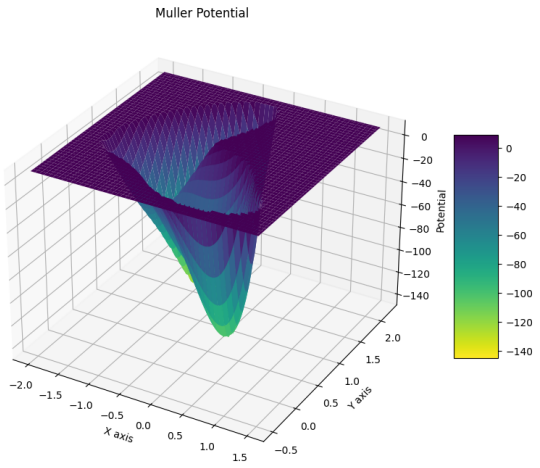

Figure 39: Müller-Brown potential.

The Müller-Brown potential is often used as a test function for optimization algorithms because it has many local minimun and one global minimun, therefore it is common that the optimization algorithms get stuck in a local minimun.

3.2.3.1 Metropolis-Hastings Algorithm for the Müller-Brown Potential

Now, we want to know how the Metropolis-Hastings algorithm behaves when we use the Müller-Brown potential as the desired distribution. The Figure 40 shows the results of the Metropolis-Hastings algorithm for the Müller-Brown potential.

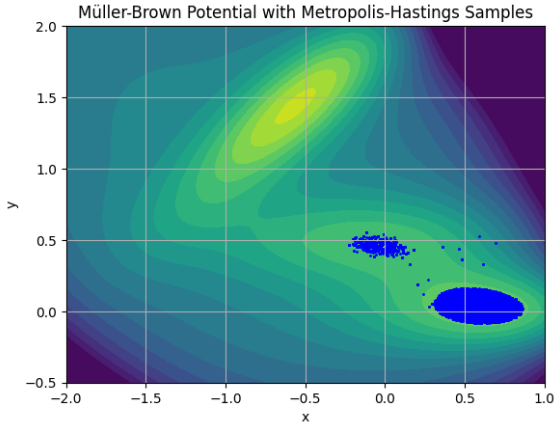

Figure 40: Sampling from the Müller-Brown potential using the Metropolis-Hastings algorithm with a temperature of 1 and 100000 samples.

We can see that the samples instead of beign distributed all around the Müller-Brown potential, they are concentrated in a local minimun. But what happens if we increase the temperature? The idea of increasing the temperature is to increase the probability of accepting a sample that is worse than the current one, therefore we will be able to escape from the local minimun. The Figure 41 shows the results of the Metropolis-Hastings algorithm for the Müller-Brown potential with a temperature of 40.

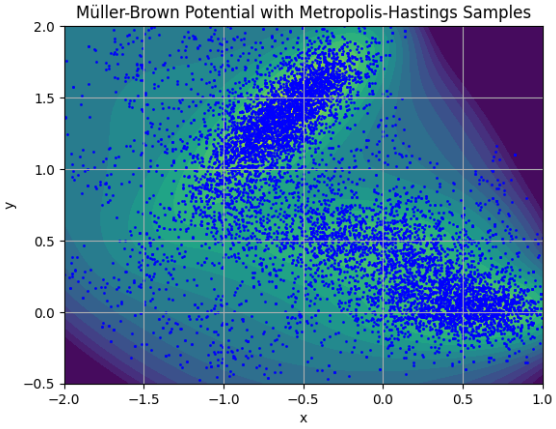

Figure 41: Sampling from the Müller-Brown potential using the Metropolis-Hastings algorithm with a temperature of 40 and 100000 samples.

We can see that the samples are distributed all around the Müller-Brown potential, which is the expected behaviour of the Metropolis-Hastings algorithm. But one could think wether this sampling is fitting the distribution or not. This will be checked in a future section after introducing the Gibbs sampling algorithm.

3.2.3.2 Gibbs Sampling Algorithm for the Müller-Brown Potential

Now, we want to know how the Gibbs sampling algorithm behaves when we use the Müller-Brown potential as the desired distribution. The Figure 42 shows the results of the Gibbs sampling algorithm for the Müller-Brown potential.

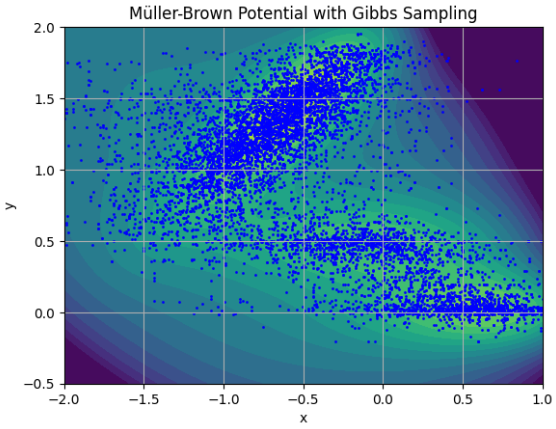

Figure 42: Sampling from the Müller-Brown potential using the Gibbs sampling algorithm with 100000 samples.

We can see that the samples are distributed all around the Müller-Brown potential, which is the expected behaviour of the Gibbs sampling algorithm.

3.2.3.3 Comparison between Metropolis-Hastings and Gibbs Sampling

This section is built with two aims. First we want to check if the samples obtained with the Metropolis-Hastings algorithm are distributed according to the Müller-Brown potential. Second, we want to compare the Metropolis-Hastings algorithm and the Gibbs sampling algorithm.

For these aims we will use the Kullback-Leibler divergence. The Kullback-Leibler divergence is a measure of how one probability distribution is different from a second, reference probability distribution. The Kullback-Leibler divergence of a probability distribution P from a probability distribution Q is defined as:

$$D_{KL}(P||Q) = \sum_{x \in \chi} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

where χ is the state space of the probability distributions.

This measure gives us a way to compare two probability distributions. If the Kullback-Leibler divergence is small, then the two probability distributions are similar. If the Kullback-Leibler divergence is big, then the two probability distributions are different.

The Figure 43 shows the Kullback-Leibler divergence between the samples obtained with the Metropolis-Hastings algorithm and the Müller-Brown potential as the number of samples increases.

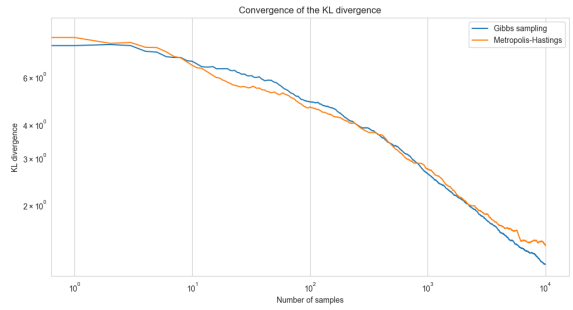

Figure 43: Kullback-Leibler divergence between the samples obtained with the Metropolis-Hastings algorithm, with temperature 40, and the Müller-Brown potential as the number of samples increases.

From this figure we can get two important results, first, and maybe the most important one, is that the Kullback-Leibler divergence is decreasing as the number of samples increases, therefore the samples obtained with the Metropolis-Hastings algorithm are distributed according to the Müller-Brown potential. Second, we can see that both algorithms, Metropolis-Hastings and Gibbs sampling, are converging almost at the same rate. This is a very important result because it shows that both algorithms are equally good for sampling from a distribution.

As a final experiment, we wanted to plot the heatmap of both distributions as well as the heatmap of the theoretical distribution, to see that indeed the samples obtained with the Metropolis-Hastings algorithm and the Gibbs sampling algorithm are distributed according to the Müller-Brown potential. The Figure 44 shows the heatmaps of the theoretical distribution, the Metropolis-Hastings algorithm and the Gibbs sampling algorithm.

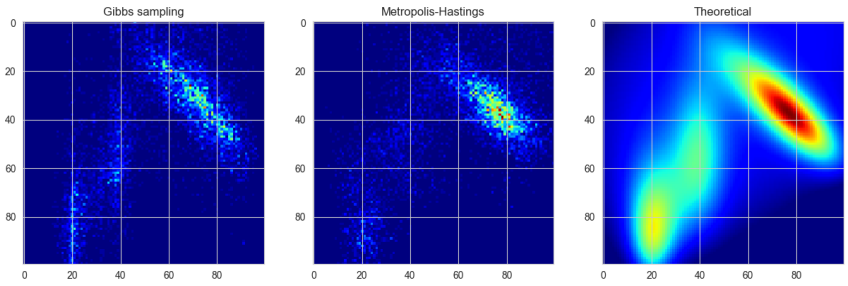

Figure 44: Heatmaps of the theoretical distribution, the Metropolis-Hastings algorithm, with tempeature 40, and the Gibbs sampling algorithm. The heatmaps are computed with 100000 samples.

3.3 Simulation of a 2D molecular system

This final section aims to show how the MCMC algorithms can be used to simulate a 2D molecular system. This subsection is divided in two parts. In the first part, we will introduce the potential we have used to simulate the ecosystem and we will delve a bit into the theory of the molecular dynamics. In the second part, we will do the implementation of the Gibbs sampling algorithm for the molecular system.

3.3.1 Theoretical Framework and Lennard-Jones Potential

At the heart of our molecular simulation lies the Lennard-Jones potential, a simplistic yet powerful model to represent interactions between a pair of atoms or molecules. It is mathematically expressed as:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$

where $U(r)$ represents the potential energy between two particles separated by a distance r, ϵ signifies the depth of the potential well (indicating the strength of attraction), and σ is the distance at which the potential is zero. The r^{-12} term models the Pauli exclusion principle, responsible for the repulsive force at short distances due to overlapping electron orbitals. Conversely, the r^{-6} term accounts for the attractive van der Waals forces, dominant at longer ranges. This dualistic nature of the potential captures the essence of molecular interactions, balancing attraction and repulsion.

The parameters ϵ and σ are not merely mathematical constructs but hold profound physical meanings. ϵ relates to the energy scale of interactions, dictating how strongly particles attract or repel each other. σ, on the other hand, sets the spatial scale, determining the effective size of the particles. These parameters are often determined empirically or through quantum mechanical calculations for specific substances, making the Lennard-Jones potential a versatile tool for simulating a wide range of molecular systems.

3.3.2 Implementation using the Metropolis Hastings Algorithm

When simulating the dynamics of molecular systems, accurately replicating the environmental context is crucial. To this end, we have conceptualized a 10x10 unit square space as the primary stage for the molecular interactions. However, to more closely mimic an infinite environment, where energy is not artificially confined or lost at the boundaries, we adopt a toroidal geometry. A torus,

which is essentially a surface of revolution generated by revolving a circle in three-dimensional space around a coplanar axis, offers an elegant solution. This spatial configuration is achieved computationally through the application of modular arithmetic at the boundaries, ensuring a seamless and continuous space. This approach is crucial for maintaining the integrity of the system's energy dynamics and is visually represented in the Figure 45.

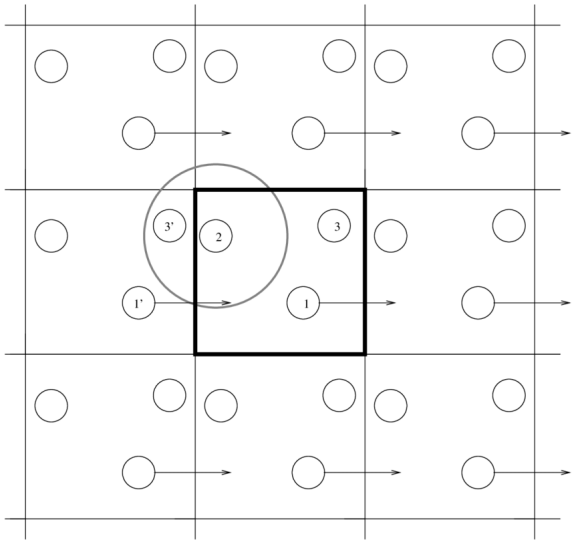

Figure 45: Illustration of boundary conditions demonstrating the toroidal topology, adapted from [9].

In our simulation, the initial placement of 30 particles within this toroidal space is generated using a uniform distribution, ensuring an unbiased starting point for the system. Given that these particles interact via the Lennard-Jones potential, we utilize the Metropolis-Hastings algorithm to simulate their dynamics. This algorithm, celebrated for its effectiveness in exploring complex probabilistic spaces, allows us to observe the evolution of the system under thermal fluctuations and intermolecular forces. For this whole experiment, we will be using the following parameters:

$$\begin{cases} n = 84 \\ \epsilon = 1 \\ \sigma = 1 \\ \Delta t = 0.01 \\ p^* = 0.844 \\ T^* = 0.71 \end{cases}$$

With ϵ and σ the parameters of the Lennard-Jones potential, Δt the time step, p^* the reduced pressure and T^* the reduced temperature and n the number of particles. Note that since $\epsilon = 1$ and $\sigma = 1$ then $p^* = p$ and $T^* = T$.

The Figure 46 depicts the spatial distribution of the particles within the torus at the outset and after 1000 iterations of the Metropolis-Hastings algorithm, showcasing the system's dynamic evolution.

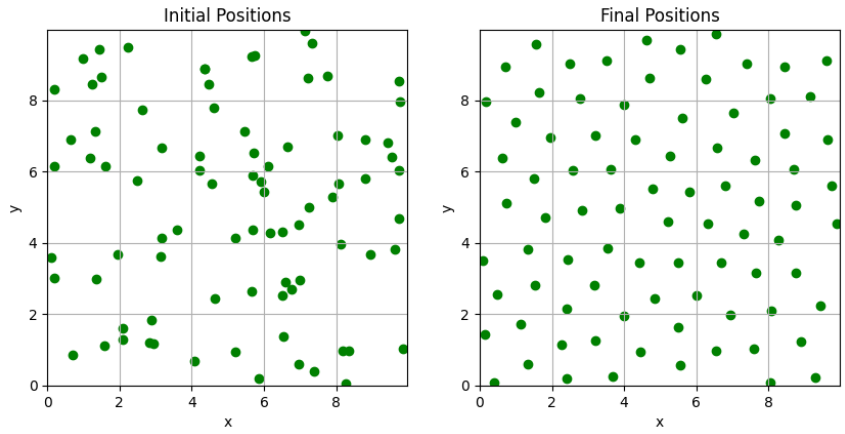

Figure 46: Spatial distribution of particles in the torus before and after 1000 iterations of the Metropolis-Hastings algorithm, demonstrating the system's evolution.

One thing that is extremly interesting of this simulation that can be observed in the Figure 46 is that, at the first case, we have distributed the particles randomly around the torus, but after 1000 iterations of the Metropolis-Hastings algorithm, the particles have been distributed in a way in which we can see that there are a minimun distance between the particles. This is a very interesting result because it shows that the model is working as expected.

An integral aspect of our analysis is the examination of the system's energy profile over the course of the simulation. The Metropolis-Hastings algorithm, in theory, should conserve the total energy of the system, a hypothesis we put to the test. The Figure 47 captures the energy of the system at the beginning and after 1000 iterations, providing insights into the algorithm's efficacy in maintaining energy conservation.

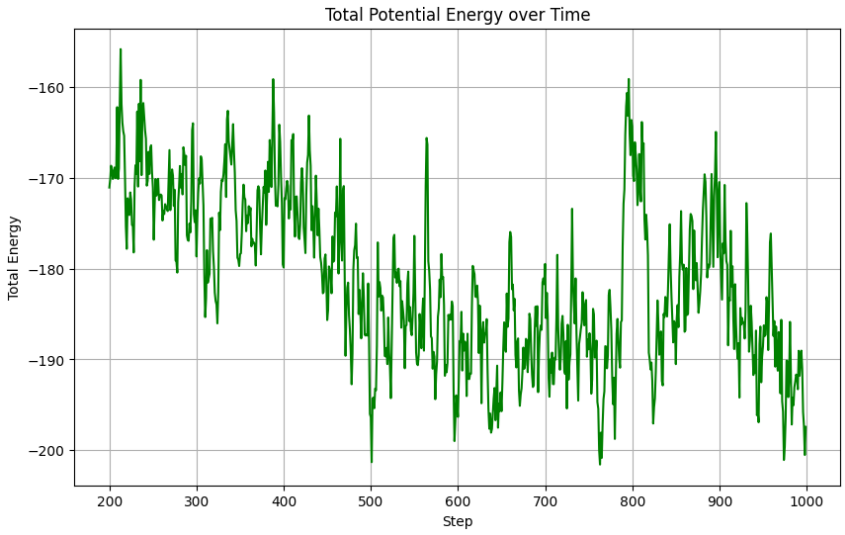

Figure 47: Energy profile of the system before and after 1000 iterations of the Metropolis-Hastings algorithm, illustrating the conservation of energy.

As observed, the energy of the system remains approximately constant, validating the expected behavior of the Metropolis-Hastings algorithm in preserving the thermodynamic properties of the

molecular system.

We have seen before the interesting result of the minimun distance between the particles, so now we want to see if indeed the theoretical aspects of the Lennard-Jones potential are being fulfilled. For this aim we can move on to the concept of the radial distribution function. The radial distribution function is a function that describes how density varies as a function of distance from a reference particle. We already know that this distribution should have the shape represented in the Figure 48, obtained from [5].

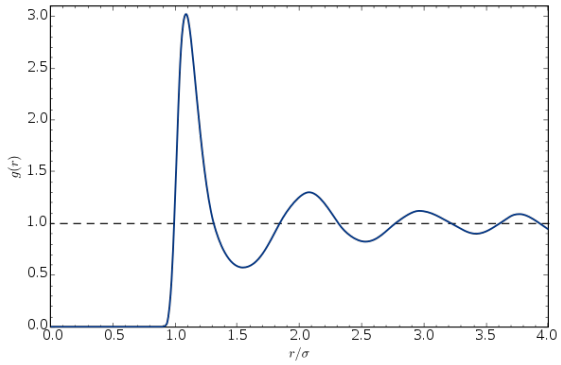

Figure 48: Radial distribution function for a liquid.

If we try to compute the radial distribution function for our system, we will get the Figure 49.

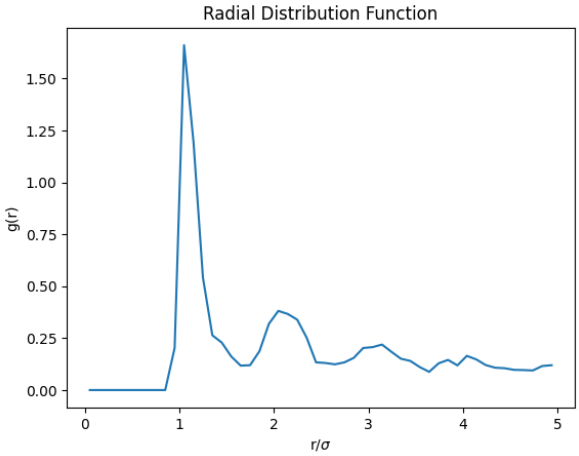

Figure 49: Radial distribution function for the system before and after 1000 iterations of the Metropolis-Hastings algorithm.

We can see that the radial distribution function is almost of the same shape as the theoretical one, even though there is not as smooth as the other one since we have a finite number of particles. Therefore, we can conclude that it seems that the Metropolis-Hastings algorithm is working as expected.

References

[1] Planetary fact sheet, 2023. Accessed: 2023-11-02.

[2] Wikipedia contributors. Inverse transform sampling, 2023. Accessed: 2023-11-24.

[3] Wikipedia contributors. Metropolis–hastings algorithm - wikipedia. Wikipedia, The Free Encyclopedia, 2023. Accessed: 2023-11-29.

[4] Wikipedia contributors. Multivariate normal distribution - wikipedia. Wikipedia, The Free Encyclopedia, 2023. Accessed: 2023-11-29.

[5] Wikipedia contributors. Radial distribution function - wikipedia. Wikipedia, The Free Encyclopedia, 2024. Accessed: 2024-01-20.

[6] Persi Diaconis. The markov chain monte carlo revolution. *Bulletin of the American Mathematical Society*, 46:179–205, 2008. Accessed: 2023-11-29.

[7] V. Frayssé. Hdr thesis, 2018. Accessed: 2023-11-02.

[8] Nicholas J. Higham. What is backward error?, 3 2020. Accessed: 2023-11-15.

[9] Thomas A. Hunt and Billy Dean Todd. On the arnold cat map and periodic boundary conditions for planar elongational flow. *Molecular Physics*, 101(23):3445–3454, 2003. Accessed: 2023-12-29.

[10] Benjamin Jourdain. *Probabilités et statistique*. Ellipses, 2 edition, 2016. Bibliogr. p. 185. Index.

[11] K. I. M. McKinnon. Convergence of the nelder–mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9(1):148–158, 1998.

[12] Roger D. Peng. 6.3 rejection sampling, 2023. Accessed: 2023-11-25.

[13] Julien Reygner. Méthodes numériques probabilistes, 2023. Course material for Master de Mathématiques et Applications, Spécialité Mathématiques de la Modélisation, Sorbonne Université et École des Ponts ParisTech, Academic Year 2023/2024.

[14] Gabriel Stoltz. *An Introduction to Computational Statistical Physics*. Master de Mathématiques et Applications Spécialité Mathématiques de la Modélisation, 1 2023.