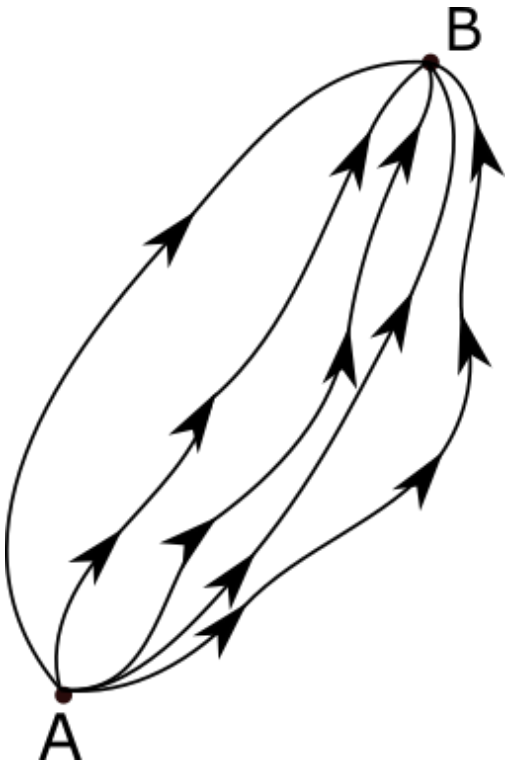
Introduction to Lattice Field Theory

Summer Project

Pablo Morandé



School of Physics And Astronomy
The University of Edinburgh
United Kingdom
2023

Personal Statement

During the course of this project, I have developed a basic understanding of the main tools in Lattice Field Theory. In particular, I have implemented the Hybrid Monte Carlo Algorithm to analyse the Quantum Harmonic Oscillator and the Metropolis Algorithm to study Lattice Quantum Chromodynamics in the Quenched approximation. I have also spent a large amount of time understanding the theory behind Markov Chains and MCMC in general.

Summary

This report offers a brief introduction to the main tools in Lattice field theory. It starts with an in-depth analysis of the theory behind Monte Carlo Methods for numerical integration. After this, the report continues with an example of how Markov Chain Monte Carlo can be used to analyse the Quantum Harmonic Oscillator using the Path Integral Formalism. Finally, the report concludes with an introduction to Lattice Field Theory which uses many of the concepts introduced in the example of the Harmonic Oscillator. The final chapter includes some calculations of Lattice QCD in the Quenched approximation.

Preface

> If I seem to wander, if I seem to stray, remember that true stories seldom take the straightest way.
>
> Patrick Rothfuss, The Name of the Wind

My intention with this report is to provide a beginner's guide to Lattice Field Theory. It encapsulates most of the knowledge I acquired during the Summer of the 2022/2023 Academic Year in which I started learning the fundamentals of this field under the supervision of Prof. Anthony Kennedy. Therefore, the paper's main aim is to give an introduction to the topic in the way that I would have liked to be taught if I were to start learning it now. Given that my learning process might differ from yours, you may find the report either excessively long or excessively short. However, I have tried to include all the aspects that first puzzled me and the best explanations I could find or derive from them. I sincerely hope you find this introduction to Lattice Field Theory useful.

Pablo Morandé, September 2023

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Anthony Kennedy, for his support and guidance throughout the project. I would also like to extend my heartfelt appreciation to my family and friends for their endless love, understanding, and encouragement.

Contents

CHAPTER 1

Markov Chain Monte Carlo

The best steel doesn't always shine the brightest.

Joe Abercrombie, The Blade Itself

1.1 Introduction

Let us assume that we are trying to determine the numerical value of the following 1-dimensional integral:

$$I = \int_{x_a}^{x_b} f(x)dx \tag{1.1}$$

$f(x)$ is a smooth function in the interval $[x_a, x_b]$. Several methods allow us to calculate an approximate integral value to a given degree of accuracy, such as the midpoint or Simpson's rule.

To apply these methods, we divide the interval into n sub-intervals (we will assume that they are of equal length), and we define $h = \frac{x_b - x_a}{n}$ The midpoint rule establishes that:

$$I = \int_{x_a}^{x_b} f(x)dx \approx h \sum_{r=1}^{n} f(x_r) + \mathcal{O}(h^2) \tag{1.2}$$

Where $x_r = x_a + (r - \frac{1}{2})h$
On the other hand, the Simpson's rule on the other hand establishes that:

$$I = \int_{x_a}^{x_b} f(x)dx \approx \frac{h}{3}\left((f(x_a) + f(x_b) + 4\sum_{r=1}^{\frac{n}{2}} f(x_{2r-1}) + 2\sum_{r=1}^{\frac{n}{2}-1} f(x_{2r}) \right) + \mathcal{O}(h^4) \tag{1.3}$$

where $x_r = x_a + rh$.
The error introduced by the midpoint rule is of the order $\mathcal{O}\left(\frac{1}{n^2}\right)$ and in Simpson's rule is of the order $\mathcal{O}\left(\frac{1}{n^4}\right)$, where n is the number of subdivisions made in the interval $[x_a, x_b]$. In 1 Dimensions, the number of points we need to consider (N) equals the number of subdivisions. While this algorithm can be easily generalized to

perform D-dimensional integrals, they are of little use as the number of points we need to consider to obtain a certain degree of accuracy scales with the number of dimensions. In the case of D-dimensional integrals, we need to divide each axis into n intervals to use the method, so the number of points we considered is $N = n^d$. The error of the algorithms is still $\frac{1}{n^2}$ and $\frac{1}{n^4}$ but n is not a good measure of performance (it is the number of points that define the number of operations that we need to perform), in D dimensions the errors are $\frac{1}{N^{\frac{2}{d}}}$ and $\frac{1}{N^{\frac{4}{d}}}$. This method becomes unfeasible when the number of dimensions is high (as in QFT calculations).

Therefore, we need to find an alternative method to estimate such integrals that scales reasonably with the dimension of the problem.

1.2 The Monte Carlo Method

1.2.1 Revision of Probability

Before introducing an alternative method to estimate the numerical values of integrals when the number of dimensions is high, we need to review some basic concepts about probability and random processes.

A random variable is a variable whose value is determined by the output of a random process. Designating the random variables by uppercase letters like X is customary. Random variables can be classified into two categories: continuous and discrete random variables. On the other hand, the sample space is the set of possible values that a random variable can take. This classification depends on whether the sample space is countable or not.

In the case of a discrete random variable, we have a Probability Mass Function (PMF) denoted as $p_X(x)$, which gives the probability of the random variable taking a particular value x, represented as $P(X = x)$. Therefore, the PMF is normalized by the condition that the sum of the probabilities must equal unity, as shown below:

$$\sum_x P(X = x) = 1 = \sum_x p_X(x) \tag{1.4}$$

In contrast, for continuous random variables, we are concerned with the probability of the random variable X being within a range $x_1 \leq X \leq x_2$. In this case, we define a probability density function $p(x)$ such that:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)dx \tag{1.5}$$

The normalization condition for continuous random variables is that the probability that the random variable spans from $-\infty$ to ∞ must be equal to unity, as shown below:

$$P(-\infty \leq X \leq \infty) = 1 = \int_{-\infty}^{\infty} p(x)dx \tag{1.6}$$

From this point on, we will only consider continuous random variables.

The mean or expectation value of a random variable X is denoted as $E[X]$ or μ, representing the typical values of the random variable. It is defined as follows:

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx \tag{1.7}$$

On the other hand, the variance (usually denoted as σ^2 or $Var[X]$) of the probability distribution measures the dispersion or spread of the values and is defined as:

$$Var[X] = \int_{-\infty}^{\infty} (x^2 - E[X]^2)p(x)dx \tag{1.8}$$

It is also possible to define the expectation value and the variance of any measurable function of X, not only the expectation value of X itself:

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx \tag{1.9}$$

$$Var[f(X)] = \int_{-\infty}^{\infty} (f(x)^2 - E[f(x)]^2)p(x)dx \tag{1.10}$$

In practice, however, it is often impossible to determine these values exactly because we do not know the underlying probability distribution, or the integrals cannot be computed directly. Therefore, they must be estimated by sampling the distribution.

To estimate these values, we must draw N samples from the random process associated with the PDF that we are analyzing. Let us label the samples as x_i with i going from 1 to N. Then, the sample average of a measurable function of the random variable denoted as $\hat{f}$, and its sample variance $\hat{\sigma}_f^2$ are defined as follows:

$$\hat{f} = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{1.11}$$

$$\hat{\sigma}_f^2 = \frac{1}{N-1}\sum_{i=1}^{N} (f(x_i) - \hat{f})^2 \tag{1.12}$$

The sample mean and average are known as estimators of the true average and variance, and they are also random variables as they are composed of the sum of random variables. Therefore, it is possible to calculate the expectation value and the variance of these new random variables. It turns out that for the sample average, one obtains

$$E[\hat{f}] = E[f(X)] \tag{1.13}$$

$$Var[\hat{f}] = \frac{Var[f(X)]}{N} \tag{1.14}$$

In the case of the sample variance:

$$E[\hat{\sigma}_f^2] = Var[f(X)] \tag{1.15}$$

$$Var[\hat{\sigma}_f^2] = \frac{2Var[f(X)]^2}{N-1} \tag{1.16}$$

Therefore, in both cases, the distribution of the sample average and the sample variance will accumulate towards the correct population values as the variance of both distributions falls as N increases. In practice, this means that it is possible to use the sample average and the sample variance as estimators of the true parameters of the distribution if we include sufficient data points.

More importantly, the Central Limit Theorem implies that the distribution of $\hat{f}$ will approach a normal distribution with $\mu = E[f(x)]$ and $\sigma^2 = \frac{Var[f(X)]}{N}$. Therefore, we can add statistical bounds to our predictions. It is possible to say that $E[f] = \hat{f} \pm \sqrt{\frac{Var[f(\hat{X})]}{N}}$ with 68% confidence. In practice, we usually do not know $Var[f(\hat{X})]$ directly, so we estimate it by the sample variance $\hat{\sigma}_f^2$.

$$E[f(X)] = \hat{f} \pm \frac{\hat{\sigma}_f}{\sqrt{N}} \tag{1.17}$$

with 68% confidence.

1.2.2 The Method

It is now time to introduce the Monte Carlo method to estimate integrals. Let us again assume that we are trying to estimate the numerical value of the following 1-dimensional integral:

$$I = \int_{x_a}^{x_b} f(x)dx \tag{1.18}$$

It is possible to modify the presentation of the problem such that it looks closer to a statistical problem:

$$I = \int_{x_a}^{x_b} f(x)dx = \int_{x_a}^{x_b} \tilde{f}(x)p(x)dx \tag{1.19}$$

Where $\tilde{f}(x) = \frac{f(x)}{p(x)}$ and $\int_{x_a}^{x_b} p(x) = 1$.
If we now consider X to be a random variable distributed according to $p(x)$, then the integral that we are trying to calculate is just $E[\tilde{f}(X)]$. Let us assume that it is possible to draw (pseudo)random samples x_i from the distribution $p(x)$. We can use all the theory of probability that we have developed to estimate the expectation value:

$$I \approx \frac{1}{N}\sum_{i=1}^{N} \tilde{f}(x_i) \pm \frac{\hat{\sigma}_{\tilde{f}}}{\sqrt{N}} \tag{1.20}$$

The advantage of the Monte Carlo Method is that all the results that we have developed generalize nicely to the D-dimensional problem. In such case, we have:

$$I_D = \int_{\Omega} f(\underline{x})dx^D = \int_{\Omega} \tilde{f}(\underline{x})p(\underline{x})dx^D \tag{1.21}$$

Where Ω is the region of integration, $\tilde{f}(\underline{x}) = \frac{f(\underline{x})}{p(\underline{x})}$ and $p(\underline{x})$ is a multivariate probability density function such that $\int_{\Omega} p(\underline{x})dx^D = 1$. We then generate a sequence of N D-dimensional vectors $\underline{x}_k$ drawing samples from $p(\underline{x})$ and we obtain an estimate of the integral using the usual estimators, and the final result is the same:

$$I_D \approx \frac{1}{N}\sum_{i=1}^{N} \tilde{f}(\underline{x}_i) \pm \frac{\hat{\sigma}_{\tilde{f}}}{\sqrt{N}} \tag{1.22}$$

It is important to note that the statistical error does not depend on the dimension; it is $\mathcal{O}(\frac{1}{\sqrt{N}})$ regardless of D.

The first obvious choice for $p(x)$ is a uniform distribution, as it is easy to draw samples from it. In such cases, we talk about Naive Monte Carlo. On the other hand, the choice of a uniform distribution is usually sub-optimal. We want $\tilde{f}$ to be as flat as possible, as that would decrease the error in our measurement (decreasing the variance). In these cases, we talk about Importance Sampling Monte Carlo.

1.3 Markov Chains and MCMC

We have now shown that the Monte Carlo method becomes highly useful when the problem's dimensionality is high. However, the only problem with the Monte Carlo method is that it might be difficult to sample from the distribution $p(x)$. In some instances, $p(x)$ might be an easy distribution, such as a Gaussian or a uniform distribution, but it will have a much more complicated form in many instances. Additionally, in physics, we do not have a choice over the probability distribution as the problems that we will be solving take the form of:

$$\langle f \rangle = \frac{\int_{\Omega} f(\underline{q})\Pi(\underline{q})d^D q}{\int_{\Omega} \Pi(\underline{q})d^D q} \tag{1.23}$$

Where both f and Π are given, we need to sample from a distribution $\boldsymbol{\pi} \propto \Pi$. Therefore, we must introduce a tool to sample from complicated probability distributions.

1.3.1 Markov Chains

A Markov Chain is a sequence of Random variables $\{X_i\}$ which have the Markov property [1], that is:

$$P(X_{n+1} = s_{n+1}|X_n = s_n, X_{n-1} = s_{n-1}..., X_0 = s_0) = P(X_{n+1} = s_{n+1}|X_n = s_n) \tag{1.24}$$

In other words, this means that the probability of the next states of the Chain depends only on the present state and not on the past states of the Chain. The values that the X_i are allowed to take form a set S known as state space, which can be countable (finite or infinite) or continuous. The state of the Markov chain at time t is the value of the random variable X_t, and a trajectory is a particular set of values from the state space for the $\{X_i\}$ sequence. On the other hand, we say that the Chain is time-homogeneous if the transition probabilities are independent of n, that is:

$$\forall n, P(X_{n+1} = b|X_n = a) = P(X_1 = b|X_0 = a), a, b \in S \tag{1.25}$$

Since Markov Chains follow the Markov property, it is reasonably easy to represent them using transition diagrams if the sample space is countable and finite. Each possible state s of X_i is represented as a node in the diagram, and the probability and the $P(X_{n+1} = s_a|X_n = s_b), s_a, s_b \in S$ is represented as an arrow from state s_a to state s_b (if it is zero we do not include an arrow). If we are constructing a trajectory and we have $X_i = s_a$, then we say that we are in the node of s_a, and to get the next value of the Chain, we need to consider the probabilities going out of the s_a node to determine to which node to go, and that will be the value of X_{i+1}. Consider the following example of a transition diagram for a Markov Chain with a state space of 3 elements, $S = \{s_1, s_2, s_3\}$:

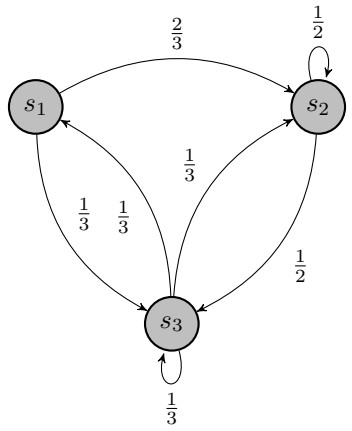


Fig. 1.1: example of the transition diagram of a Markov Chain with a space state of 3 elements and particular transition probabilities

Another way to represent a discrete Markov Chain is through a transition Matrix. Let the elements of S be represented by the natural numbers $1, 2, 3...$ then we can express the conditional probability $P(X_{n+1} = j|X_n = i)$ as the entry of a matrix p_{ij}. This Matrix must have the property:

$$\sum_{j \in S} p_{ij} = 1, \forall i \in S \tag{1.26}$$

Given that the following random variable in the Chain must take a value in S. Of course, it is possible to generalize this to the case of continuous S. In such case, instead of a matrix, we will have a Kernel $p(\mathbf{x}, A) = P(X_{i+1} \in A|X_i = x)$ where $A \subset S$ and $x \in S$. It is also important to note that the transition matrix will generally depend on n, the position in the Chain. However, there will be no dependence in this case as we are considering only time-homogeneous chains.

The Matrix introduced above describes the one-step transitions, but a more general question would be how to represent the m-step probability, that is:

$$p_{ij}^{(m)} = P(X_{n+m} = j | X_n = i) \tag{1.27}$$

It turns out that following the definition of the Markov chain, it can be shown that the m-step conditional probability can be expressed in terms of the one-step transition probability as:

$$p_{ij}^{(m)} = p_{ij}^m \tag{1.28}$$

On the other hand, in the case of a discrete state space, it is possible to express the probability distribution of any of the random variables of the Chain (X_n) as a probability vector:

$$v_i^{(n)} = P(X_n = i), i \in S \tag{1.29}$$

If the probability distribution for X_n is represented by the probability vector $v_i^{(n)}$ then the probability distribution of X_{n+1} is given by:

$$v_j^{(n+1)} = \sum_i v_i^{(n)} p_{ij} = (\mathbf{v}^T p)_j \tag{1.30}$$

Consequently, if the probability distribution of X_n is given by $\nu^{(n)}$ then the probability distribution of X_{n+m} will be given by $\nu^{(n+m)} = \sum_i v_i^{(n)} p_{ij}^m$. In practice, when defining the Markov chain, we start with an initial distribution for X_0, which we directly denote as $\nu_i = \nu_i^{(0)}$. Therefore, the probability distribution of X_n is given by:

$$v_j^{(n)} = \sum_i v_i p_{ij}^n = (\mathbf{v}^T p^n)_j \tag{1.31}$$

Definition 1. *In the context of Markov Chains, a Stationary Distributions is a distribution $\boldsymbol{\pi}$ such that [1]:*

$$\boldsymbol{\pi}_j = \sum_i \boldsymbol{\pi}_i p_{ij}. \forall j \in S \tag{1.32}$$

Essentially, this means that once the Chain reaches a stationary distribution for some n, it will stay there as by definition $\boldsymbol{\pi}_j = \sum_i \boldsymbol{\pi}_i p_{ij}^n$ for all n [1]. Therefore, we want to build a Markov chain that has as stationary distribution, the one that we want to sample from, and we also obviously want the Chain to reach the stationary distribution.

Definition 2. *Let there be two states i, j from the State Space S of a Markov Chain. We say that state j is accessible from state i if and only if:*

$$\sum_{n=0}^{\infty} P_{ij}^n > 0 \tag{1.33}$$

This means that there is a non-vanishing probability of the Chain reaching state j if it has started at state i in a finite number of steps.

Definition 3. *Let there be two states i, j from the State Space S of a Markov Chain, then i, j communicate if i is accessible from j and j is accessible from i*

It should be obvious that Communication constitutes an equivalent relation and can be used to partition the set S in disjoint equivalent classes.

Definition 4. *A Markov Chain with only one communication class is known as an Irreducible Markov Chain*

That is to say that every state can be reached from any other initial state with a non-zero probability in a finite number of steps.

Definition 5. *The period of a state i Markov Chain with a discrete State Space is defined as:*

$$d_i = gcd\{n : p_{ii}^n > 0\} \tag{1.34}$$

The State i is said to be aperiodic if $d_i = 1$ and periodic otherwise.

Theorem 1. *If states i, j communicate then $d_i = d_j$ [1]*

Meaning that all the states in a communicating class will have the same period. Furthermore, if the Chain is irreducible, then all the states of the Chain will have the same period. An aperiodic chain is one in which all its states are aperiodic.

Definition 6. *We define the first hitting time of state i of a Markov Chain with a discrete Stete Space S as:*

$$T_i = min[n \geq 1 | X_n = 1] \tag{1.35}$$

Definition 7. *A state is recurrent if and only if:*

$$P(T_i < \inf | X_0 = i) = 1 \tag{1.36}$$

And transient if and only if:

$$P(T_i < \inf | X_0 = i) < 1 \tag{1.37}$$

An alternative but equivalent way of defining recurrent stat is that the sum $\sum_n = 1^\infty p_{i,i}^n$ diverges if and only if i is recurrent and converges if and only if i is transient.

Theorem 2. *Recurrence is a class property, meaning that if a state in a communication class is recurrent, then all states in such class are recurrent, and the same happens if the state is transient. Therefore, in an irreducible Markov Chain, all states are either all recurrent (in which case the Chain is said to be recurrent) or all transient (in Which the case is said to be transient) [1]*

Definition 8. *Let i be a recurrent state of a discrete Markov Chain with Sample Space S, then i is positive recurrent if and only if:*

$$E[T_i | X_0 = i] < \infty \tag{1.38}$$

And null recurrent if and only if:

$$E[T_i | X_0 = i] = \infty \tag{1.39}$$

Essentially, positive recurrent states are recurrent by a wide margin.

Theorem 3. *Positive and Null Recurrence are class properties, meaning that if a state in a communication class is Positive recurrent, all states in such class are Positive Recurrent. The same happens if the state is Null Recurrent. Therefore, in a recurrent irreducible Markov Chain, all states are either all Positive Recurrent (in which case the Chain is said to be Positive Recurrent) or all Null Recurrent (in Which the case is said to be Null Recurrent) [1]*

It is now time to see how all these properties relate to the problem that we want to solve, that is, how to create a Markov Chain that has as stationary distribution a target distribution and that converges to such distribution regardless of the initial conditions.

Theorem 4. *An irreducible Markov Chain is positive Recurrent if and only if it has a stationary distribution $\boldsymbol{\pi}$, in which case $\boldsymbol{\pi}$ is the unique stationary distribution of the Markov Chain [1].*

Theorem 5. *Basic Limit Theorem*
Let $\{X_i\}$ be an irreducible and aperiodic Discrete Markov Chain with a countable Set Space S and a stationary distribution $\boldsymbol{\pi}$. Let X_0 have a probability distribution denoted as $\boldsymbol{\pi}_0$ and let the probability distribution of X_n be denoted as $\boldsymbol{\pi}_n$, then it follows that:

$$\lim_{n \to \infty} \boldsymbol{\pi}_n = \boldsymbol{\pi} \tag{1.40}$$

Therefore, if we can build such a Markov Chain, it will be guaranteed to converge to the correct distribution regardless of the initial configuration. We can force our Markov Chain to be both aperiodic and irreducible, but how do we force it to have the distribution that we like as a stationary distribution? One way of doing so is through detailed balance:

Theorem 6. *Let there be a Markov Chain with a discrete state Space S with transition probabilities given by $p_{ij} = P(X_{n+1} = j | X_n = i)$ and let $\boldsymbol{\pi}$ be a distribution vector, we say that $\boldsymbol{\pi}$ satisfies detailed balance with respect to the Matrix p if:*

$$\boldsymbol{\pi}_i p_{ij} = \boldsymbol{\pi}_j p_{ji}, \forall i, j \in S \tag{1.41}$$

Then $\boldsymbol{\pi}$ is a stationary distribution of the Markov Chain. It is important to note that Detailed balance is just a sufficient condition for $\boldsymbol{\pi}$ being the Markov Chain. This means there are stationary distributions of Markov Chains that do not satisfy detailed balance [2].

Proof. The proof of Theorem 6 is trivial. We need to sum over i in both sides of the equation:

$$\sum_i \boldsymbol{\pi}_i p_{ij} = \sum_i \boldsymbol{\pi}_j p_{ji} \tag{1.42}$$

$$\sum_i \boldsymbol{\pi}_i p_{ij} = \boldsymbol{\pi}_j \sum_i p_{ji} \tag{1.43}$$

$$\sum_i \boldsymbol{\pi}_i p_{ij} = \boldsymbol{\pi}_j, \text{ as } \sum_i p_{ji} = 1 \tag{1.44}$$

The last line is exactly the definition of a Stationary Distribution. ■

Therefore, we need an algorithm that allows us to create an Irreducible and aperiodic Markov Chain that satisfies the detailed balance equation with respect to the target distribution. This will allow us to take samples from that distribution. The abovementioned theory applies particularly to Markov Chains with a countable State Space. However, in most cases, we will be interested in creating Markov Chains with a non-countable State Space as we will want to sample from continuous distributions. The basic concepts generalize nicely to the case of the non-countable State Space, but the definitions of irreducibility, aperiodicity and recurrence must be adapted. For example, classic irreducibility is impossible, so we are left with a weaker restriction known as ϕ irreducibility. Additionally, it is not sufficient for the Chain to be both ϕ irreducible aperiodic and have a stationary distribution to converge to the stationary distribution completely. For this, we need a further requirement known as Harris Recurrence.

In the case of a General State Space, a Markov Chain that is aperiodic, ϕ irreducible and positive Harris Recurrent with a stationary probability distribution $\boldsymbol{\pi}$ is known as Harris Ergodic. For such chains, the distribution of X_n will converge to the stationary distribution $\boldsymbol{\pi}$ with probability one [3]. The basic idea is the same as before: we need an algorithm that allows us to create a (ϕ) irreducible, aperiodic (and positive Harris Recurrent) and that satisfies detailed balance with respect to a desired probability distribution so that we can sample from it. The Metropolis Algorithm can be used in both the case of a Countable State Space and the case of a non-countable State Space.

1.3.2 Algorithms in MCMC

One possibility that creates an irreducible (ϕ in case of Non-countable S) aperiodic (and Harris Recurrent in case of Non-countable S) and that satisfies detailed balance with respect to a probability distribution $\boldsymbol{\pi}$ is the Metropolis Algorithm, the pseudo-code for it is given below: The Conditional probability distribution $g(x|y)$ is responsible for suggesting new states given the previous state of the Chain and is an arbitrary distribution. However, we need it to be symmetric, that is, $g(x|y) = g(y|x)$, and it needs to be easy to sample. Additionally, given that $\boldsymbol{\pi}$ appears only in a ratio, it does not need to be strictly a probability distribution but rather proportional to one as the ratio takes care of the normalization. This Algorithm will create a Markov Chain such that in the large t limit, the probability distribution of X_t converges to

Algorithm 1 Metrpolis Algorithm

procedure METROPOLIS(T, g(x|y), $\boldsymbol{\pi}$)
  $X_0 \leftarrow$ *Initialize* X_0 *with a value* $\in S$
  for $t = 0$ *up to* $t = T$ **do**
    $X' \leftarrow$ *Generate Candidate X' by sampling from the Distributiong*$(X'|X_t)$
    $\alpha \leftarrow min\left(1, \frac{\boldsymbol{\pi}(x')}{\boldsymbol{\pi}(x)}\right)$
    $r \leftarrow$ *Generate Uniform Random Number [0,1]*
    if $r \leq \alpha$ **then**
      $X_{t+1} \leftarrow X'$
    else
      $X_{t+1} \leftarrow X_t$

a probability distribution proportional to $\boldsymbol{\pi}$. This Algorithm is tailored to satisfy detailed balance. To see this, we need to start considering the detailed balance condition:

$$\boldsymbol{\pi}(x)p(x,x') = \boldsymbol{\pi}(x')p(x',x) \tag{1.45}$$

If we consider that the candidate state is first proposed and then either accepted or rejected, then:

$$p(x,x') = g(x,x')A(x,x') \tag{1.46}$$

$g(x,x')$ is the probability distribution of proposing state x' being in state x, and $A(x,x')$ is the probability of accepting the new state being in state x. For this formulation, the detailed balance condition becomes:

$$\frac{\boldsymbol{\pi}(x')g(x',x)}{\boldsymbol{\pi}(x)g(x,x')} = \frac{A(x,x')}{A(x',x)} \tag{1.47}$$

We must choose an $A(x,x')$ that satisfies that condition. It is clear that this is immediately satisfied by:

$$A(x,x') = min\left(1, \frac{\boldsymbol{\pi}(x')g(x',x)}{\boldsymbol{\pi}(x)g(x,x')}\right) \tag{1.48}$$

As $A(x,x')$ or $A(x',x)$ will be one using this condition, the other will give the correct ratio. This is the metropolis choice, which is simplified to the Algorithm shown in the pseudo-code if we require a symmetric proposal distribution, that is, $g(x',x) = g(x',x)$

1.3.3 Getting Results in MCMC

Now that we have an algorithm allowing us to sample from the distribution, it is time to ask ourselves if the samples we create can be used for Monte Carlo Integration.

Theorem 7. *Let X be a Harris Ergodic Markov Chain* $\{X_n\}$ *with a General Set Space S with stationary distribution* $\boldsymbol{\pi}$ *and f a real-valued function which is* $\boldsymbol{\pi}$*-integrable on S. Then if* $E_{\boldsymbol{\pi}}[f] < \infty$*[3]:*

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} f(X_i) = E_{\boldsymbol{\pi}}[f]\right) = 1 \tag{1.49}$$

Which is a good start as this means that we can use $\hat{f}_n = \frac{1}{n}\sum_{i=0}^{n} f(X_i)$ as estimator of the true expectation value. However, we also want some Central Limit Theorem to hold. It turns out that under certain conditions (Which are outside the scope of this review) [3], the CLT applies for Markov Chains, and therefore we have that:

$$\lim_{n\to\infty} \sqrt{n}(\hat{f}_n - E_{\boldsymbol{\pi}}[f]) \to_d N(0,\sigma^2) \tag{1.50}$$

Where the d-subscript denotes convergence in distribution and $\sigma^2 = \lim_{n\to\infty} Var[\sqrt{n}\hat{f}]$. Therefore, it is possible to conclude that with 68% confidence, it is true that:

$$E_{\boldsymbol{\pi}}[f] = \hat{f}_N \pm \frac{\sigma}{\sqrt{N}} \tag{1.51}$$

However, we still need to determine a way of estimating the variance σ as, in this case, the variables are correlated and using the naive variance estimator would result in underestimating the errors. First, we must look at the true variance of $\hat{f}$. Suppose that have Harris Ergodic Markov Chain X_t and a stationary distribution $\boldsymbol{\pi}$. We assume that we have been able to sample the first N random variables of the Chain, such that t is going from 0 to N. We will further assume that the Chain has thermalized, that is to say, that $X_0 \propto \boldsymbol{\pi}$

$$\sigma^2 = \lim_{n\to\infty} Var[\sqrt{n}\hat{f}_n] \tag{1.52}$$

$$= \lim_{n\to\infty} nVar[\frac{1}{N}\sum_{t=0}^{n} f(X_t)] \tag{1.53}$$

$$= Var[f(X_0)] + 2\sum_{k=1}^{\infty} Cov(X_s, Xs+k) \tag{1.54}$$

$$= \sigma^2 + 2\sum_{k=1}^{\infty} Cov(X_s, X_{s+k}) \tag{1.55}$$

$$= \sigma_f^2\left(1 + 2\sum_{k=1}^{\infty}\frac{Cov(X_s, X_{s+k})}{\sigma_f^2}\right) \tag{1.56}$$

$$= \sigma_f^2\left(1 + 2\sum_{k=1}^{\infty} p_k\right), p_k = \frac{Cov(X_s, X_{s+k})}{\sigma_f^2} \tag{1.57}$$

$$= \sigma_f^2\tau_{int}, \tau_{int} = 1 + 2\sum_{k=1}^{\infty} p_k \tag{1.58}$$

Therefore, in order to estimate σ^2, we use the usual unbiased estimator for σ_f^2, mainly $\hat{\sigma}_f^2$ and then we need to estimate τ_{int} Known as integrated autocorrelation time, IAT. To estimate the IAT, we first estimate the autocorrelation function p_k. For which a natural estimator is:

$$\hat{p}_k = \frac{1}{\hat{\sigma}_f^2(N-k)}\sum_{t=1}^{N-k}(f(X_n) - \hat{f}_N)(f(X_{n+t}) - \hat{f}_N) \tag{1.59}$$

It is tempting to estimate the IAT as

$$\hat{\tau}_{int} = 1 + 2\sum_{k=1}^{N}\hat{p}_k \tag{1.60}$$

However, for Markov Chains, we expect the autocorrelation p_k to follow an exponential so that for big k, the process is dominated by noise. Therefore, to avoid adding noise it is common to add a cutoff, M such that $M << N$ so that the noise effect is reduced. Therefore, our estimator for the IAT is:

$$\hat{\tau}_{int} = 1 + 2\sum_{k=1}^{M}\hat{p}_k \tag{1.61}$$

The Cutoff is chosen such that τ_{int} is flat; a general criterion is that $M = 4\hat{\tau}_{int}$. In practice, we calculate $\hat{\tau}_{int}$ for different M until the condition $M \geq 4\hat{\tau}_{int}$ is satisfied. The truncation of the sum induces an error in our estimate of the IAT, which is given by:

$$Var[\hat{\tau}] = \frac{2(2M+1)}{N}\hat{\tau}_{int} \tag{1.62}$$

Given that this quantity is expensive to compute, it is useful not to record all values when doing metropolis, calling metropolis several times before recording the value so that the autocorrelation between the samples is reduced. Additionally, the Fast Fourier Transform can be used to reduce the computational load of computing the IAT.

1.3.4 Example Using MCMC

As an example, we can look again at the Gaussian Distribution using the Metropolis Algorithm. The objective is to compute the integral:

$$\langle x^m \rangle = \frac{\int_{-\infty}^{\infty} x^m \exp\left\{-\frac{x^2}{2}\right\}}{Z} \tag{1.63}$$

$$Z = \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} \tag{1.64}$$

The results of this computation can be easily obtained analytically, but we will use it as a test of the theory and algorithms explained in this chapter. Implementing the Metropolis algorithm for the case of a Normal Distribution $\mathcal{N}(0,1)$ is straightforward. First, we select a probability distribution $g(x,x')$ from which we will sample candidate states given the current state. A simple choice for $g(x,x')$ is a uniform distribution centred in x:

$$g(x,x') = U[x-\Delta, x+\Delta] \begin{cases} \frac{1}{2\Delta} & |x'-x| \leq \Delta \\ 0 & Otherwise \end{cases} \tag{1.65}$$

Where Δ is a parameter of the program, this distribution satisfies the Metropolis algorithm's symmetry requirement since $g(x,x') = g(x',x)$. This choice of distribution and the form of the Metropolis algorithm will guarantee that the Markov Chain is Harris ergodic, which is a necessary condition for the convergence to the stationary distribution.

After deciding the proposal probability distribution, the only thing left to do is decide on the Algorithm's parameters. The first parameter of the Algorithm is Δ, which controls the width of the proposal distribution. If Δ is chosen too big, it will result in almost all new states being rejected in the Metropolis state. This is unreasonable as the Chain will not converge to the desired probability distribution quickly enough, and we will constantly be sampling the same values. On the other hand, if Δ is chosen to be too small, then almost all proposed states will be accepted but will be highly correlated as they will represent slight variations of the initial state. Therefore, Δ must be tuned with care and a value of $\approx 60\%$ acceptance is desirable. Next, deciding how many values of the Markov Chain will be discarded is essential. Discarding the first values the Chain gives is vital as it is necessary to give the chain time to reach the stationary distribution. This will accelerate the convergence of the measured quantities to their expected values. Note that the thermalization is optional as it only accelerates the convergence; Without it, the values would still converge to the expected ones, but it would take longer. A good choice is to discard the first 10% iterations. Finally, we need to decide the length of the Chain that we will consider, that is, the number of iterations in the Metropolis algorithm, but this depends on the precision we require for our measurements.

An excellent initial check for our program is to check the distribution of x when the Chain has thermalized. If everything is working, x should be distributed according to $\mathcal{N}(0,1)$. Therefore, a normalized histogram of the variable x generated by the metropolis algorithm is shown below.

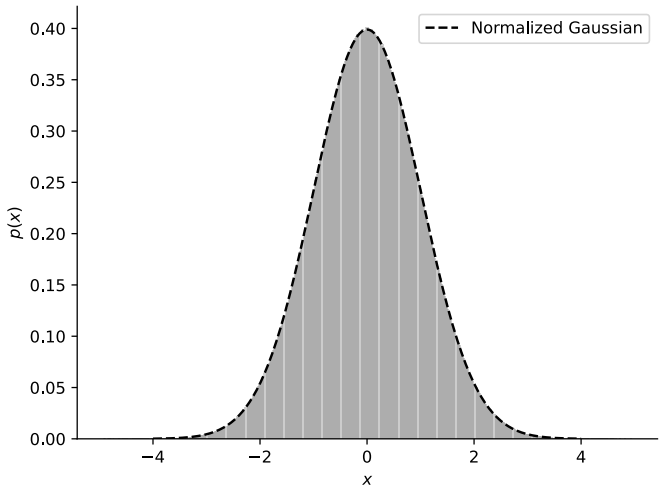


Fig. 1.2: Histogram obtained by implementing Metropolis algorithm for a Gaussian Distribution. The parameters of this particular run were: $\Delta = 0.5$, Number of Measurements: 10^7, Thermalization iterations: 10% of Measurement iterations. The red line indicates the normalized Gaussian distribution.

It is clear from Figure 1.2 that the random variable x follows a Gaussian Distribution after thermalization, which means that we can use Monte Carlo Methods to Estimate the Integrals $\langle x^m \rangle$. The Estimated values of $\langle x^m \rangle$ for $m = \{1, 2, 4\}$ and different numbers of Measurements are shown below. The errors were computed by estimating the IATs of the measurements as explained in the section above.

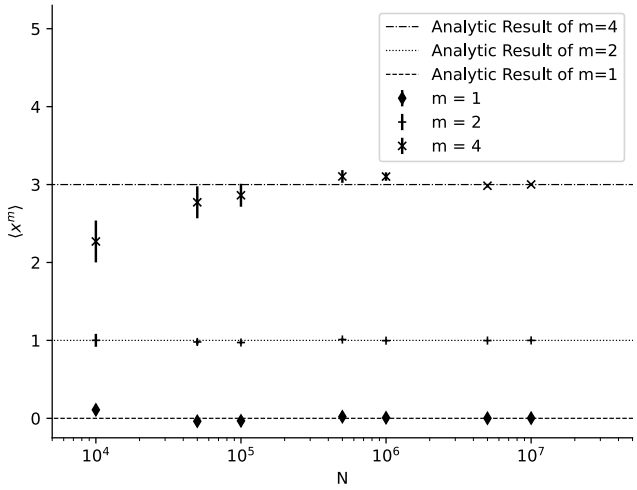


Fig. 1.3: Numerical Estimation of $\langle x^m \rangle$ for $m = 1, 2, 4$ for a Gaussian Distribution using Metropolis Algorithm. The Algorithm's parameters (other than N) are the same as in 1.2.

It is clear from the plot that the result of the numerical simulation approaches the values of the analytical solution for all the investigated values of m. An interesting point is that higher powers take more time to converge, and the error bars decrease more slowly.

CHAPTER 2

Path Integrals in Quantum Mechanics and MCMC for the QHO

> Language was just difference. A thousand different ways of seeing, of moving through the world. No; a thousand worlds within one. And translation – a necessary endeavour, however futile, to move between them
>
> R.F. Kuang, Babel: An Arcane History

2.1 Path Integrals In Quantum Mechanics

It is helpful to begin with a brief review of the Path Integral formulation of Quantum Mechanics.
Let us assume that we have a Hermitian position operator $\hat{x}$ with a complete set of eigenstates such that:

$$\hat{x}\,|x\rangle = x\,|x\rangle \tag{2.1}$$

$$\hat{1} = \int dx\,|x\rangle\,\langle x| \tag{2.2}$$

Then, we need to define the transition amplitude between the states $\langle x_b, t_b|x_a, t_a\rangle = \langle x|\exp\left\{-i\hat{H}(t_b - t_a)\right\}|x\rangle$.

$$\langle x_b, t_b|x_a, t_a\rangle = \int_{x(t_a)=x_a}^{x(t_b)=x_b} \mathcal{D}x \exp\{iS\} \tag{2.3}$$

Where $S = \int_{t_a}^{t_b} \mathcal{L}(x(t), \dot{x}$. Equation 2.3 only carries a meaningful meaning when we discretize the theory. We divide time in $N+1$ points t_n such that $t_n = t_a + n\epsilon, \epsilon = \frac{t_a - t_b}{N}$ such that $t_0 = t_a$ and $t_b = t_N$. Additionally, we denote $x_n = x(t_n)$ with the constraints $x_a = x_0$ and $x_b = x_N$. Under this construction Equation 2.3 is:

$$\langle x_b, t_b|x_a, t_a\rangle \propto \lim_{N\to\infty} \int \Pi_{n=1}^{N-1} dx_n \exp\left\{ i\epsilon \sum_{i=0}^{N-1} \left(\frac{m}{2}\left(\frac{x_{i+1} - x_i}{\epsilon}\right)^2 - V(x_i)\right)\right\} \tag{2.4}$$

We now define the trace of $\langle x_b t_b | x_a t_a \rangle$ as the integral over all paths with periodic boundary conditions $x_b = x_a = x$. Such that:

$$Z = \int dx \, \langle x, t_a | x, t_b \rangle \propto \int \Pi_{n=0}^{N-1} dx_n \exp\left\{ i\epsilon \sum_{i=0}^{N-1} \left(\frac{m}{2} \left(\frac{x_{i+1} - x_i}{\epsilon} \right)^2 - V(x_i) \right) \right\} \tag{2.5}$$

Z is also known as the partition function, and it is also possible to express it in a different form:

$$\begin{aligned} Z &= \int dx \, \langle x, t_b | x, t_a \rangle && (2.6) \\ &= \sum_{n=0}^{\infty} \int dx \, \langle x | \exp\{-iH(t_b - t_a)\} | n \rangle \langle n | x \rangle && (2.7) \\ &= \sum_{n=0}^{\infty} \int dx \exp\{-iE_n t\} |\psi_n(x)|^2 && (2.8) \\ &= \sum_{n=0}^{\infty} \exp\{-iE_n(t_b - t_a)\}| && (2.9) \\ &= Tr(\exp\{-iH(t_b - t_a)\}) && (2.10) \end{aligned}$$

On the other hand, we need to define the expectation value of a general operator $\hat{A}$:

$$\langle \hat{A} \rangle \equiv \frac{\int dx \, \langle x, t_b | \, \hat{A} \, | x, t_a \rangle}{Z} \tag{2.11}$$

The Normalization factor of $\frac{1}{Z}$ assures that $\langle 1 \rangle = 1$. It is possible to proceed similarly as before to express $\langle \hat{A} \rangle$ as a trace:

$$\frac{\int dx \, \langle x, t_b | \, \hat{A} \, | x, t_a \rangle}{Z} = \frac{\sum_{n=0}^{\infty} \exp\{-iE_n(t_b - t_a)\} \, \langle n | \, \hat{A} \, | n \rangle}{Z} = \frac{Tr(\exp\{-iH(t_b - t_a)\}\hat{A})}{Tr(\exp\{-iH(t_b - t_a)\}} \tag{2.12}$$

2.2 MCMC in Quantum Mechanics

To perform numerical calculations, we first need to analytically continue to imaginary time [4]. That is, we let $\tau = it$, also known as Wick Rotation [4]. This rotation produces a change in the action such that $iS \to -S_E$ where S_E is known as Euclidean action [4]. In general, the action is defined as:

$$S = \int_{t_a}^{t_b} dt \mathcal{L} \tag{2.13}$$

If we now let both $\tau_b = it_b, \tau_a = it_a$ we have that $S \to S_E = i \int_{\tau_a}^{\tau_b} d\tau' \mathcal{L}_E$. If the Lagrangian $\mathcal{L}$ is defined as $\mathcal{L} = \frac{m}{2}\dot{x}^2 - V(x)$, then $\mathcal{L} \to \mathcal{L}_E = \frac{m}{2}\dot{x}^2 + V(x)$ [4].

By doing this, we have that:

$$\langle \hat{A} \rangle = \frac{\int dx \, \langle x, \tau_b | \, \hat{A} \, | x, \tau_a \rangle}{Z} = \frac{\sum_{n=0}^{\infty} \exp\{-E_n(\tau_b - \tau_a)\} \, \langle n | \, \hat{A} \, | n \rangle}{\sum_{n=0}^{\infty} \exp\{-E_n(\tau_b - \tau_a)\}} \tag{2.14}$$

Therefore, if we define $\tau = \tau_b - \tau_a$ and we obtain that for large τ:

$$\lim_{\tau \to \infty} \frac{\int dx \, \langle x, \tau_b | \, \hat{A} \, | x, \tau_a \rangle}{Z} = \langle 0 | \, \hat{A} \, | 0 \rangle \tag{2.15}$$

In the Large τ limit, the path integral formalism can measure quantities relative to the vacuum state $|0\rangle$. From now on, we will directly denote $\langle 0|\,\hat{A}\,|0\rangle$ as $\langle\hat{A}\rangle$ [4].

Let us assume that we are interested in computing the expectation values of an observable $\hat{O}$ relative to the vacuum:

$$\langle\hat{O}\rangle = \lim_{\tau\to\infty} \frac{\int \mathcal{D}x O(x)\exp\{-S_E\}}{Z} \tag{2.16}$$

In this case, the integration is made over all paths with periodic boundary conditions. Therefore, it is possible to express Equation 2.16 in the following way:

$$\langle\hat{O}\rangle = \lim_{\tau\to\infty, N\to\infty} \frac{\int \Pi_{n=0}^{N-1} dx_n O(x_0,...x_{N-1})\exp\{-S_E\}}{\int \Pi_{n=0}^{N-1}\exp\{-S_E\}} \tag{2.17}$$

Where $\epsilon = \frac{\tau}{N}$, $\tau_0 = 0$ and $\tau_n = \tau_0 + \epsilon n$ and $x_i = x(\tau_i)$ and the constraint $x_0 = x_N$. On the other hand, the Euclidean action is expressed as:

$$S_E = \lim_{N\to\infty} \sum_{i=0}^{N-1} \epsilon\cdot m\left(\frac{(x_{i+1}-x_i)^2}{2\epsilon^2} + V(x_i)\right) \tag{2.18}$$

Therefore, we can estimate the Expectation value of an observable of a quantum system if we can perform the integral in Equation 2.17 for finite (but large) τ and small ϵ (which for fixed τ involves a large N). This is precisely where Markov Chain Monte Carlo comes in. We have effectively reduced the problem to solving an N dimensional integral (with N being large), so the only method that can be used to estimate such an integral is Monte Carlo [2]. Additionally, the problem is already formulated in a way where we can interpret the set of $\{x_i\}$ as being distributed according to $\exp\{-S_E\}$. Hence, it is possible to build an algorithm that generates a Markov Chain that allows us to draw samples of $\{x_i\}$ distributed as $\exp\{-S_E\}$ and then use those samples to estimate the integral with the observable:

$$\langle\hat{O}\rangle_E = \sum_{t=0}^{T} O(\{x_i\}_t) \pm \frac{\sigma}{\sqrt{T}} \tag{2.19}$$

σ is calculated with the integrated autocorrelation time [2]. It is important to note that there are three sources of errors at this point. The most obvious is the statistical error due to the Markov Chain Monte Carlo, expressed in Equation 2.19. It concerns the value of $\langle\hat{O}\rangle_E$, that is, the value of the integral of Equation 2.17 when N and τ are finite. However, if we want to compare this quantity with the true value of $\langle\hat{O}\rangle$, we must consider that having both N and τ finite will introduce an error in addition to the statistical one. In general, we will only have that $\langle O\rangle = \lim_{\tau,N\to\infty}\langle\hat{O}\rangle_E$ [2].

2.3 The Quantum Harmonic Oscillator with MCMC

We will now use the abovementioned methods to study the Quantum Harmonic Oscillator in Euclidean time. This is a good starting point as it allows us to test the validity of our theory and methods as the analytic solution is available for this system. The Discrete Euclidean action for this system is given by [5]:

$$S_E = \sum_{i=0}^{N-1} \epsilon\cdot m\left(\frac{(x_{i+1}-x_i)^2}{2\epsilon^2} + \frac{1}{2}w^2x_i^2\right) \tag{2.20}$$

2.3.1 HMC for QHO

We will use this opportunity to implement the Hybrid Monte Carlo (HMC) algorithm to estimate the integrals of this system:

Algorithm 2 Hybrid Monte Carlo

procedure GENERATE CHAIN($N, N_\tau, \epsilon_\tau, \epsilon, m, w$)
  $\phi_0 = Initialize\ \phi_0$
  for $n = 1$ *Up to* $n = N$ **do**
    $\phi_n = UPDATE(\phi_{n-1}, N_\tau, \epsilon_\tau, \epsilon, m, w)$
procedure UPDATE(ϕ,N_τ,$\epsilon_\tau, \epsilon, m, w$)
  $\phi(0) \leftarrow \phi$
  $\pi(0) \leftarrow$ *Array of N random conjugate Momenta distributed* $\propto \exp\left\{-\frac{\pi^2}{2}\right\}$
  $H = \frac{1}{2}\sum \pi(0)_i^2 + S(\phi(0), \epsilon, m, w)$
  $\tau \leftarrow N_\tau \epsilon$
  $\pi(\tau), \phi(\tau) = Molecular_dyanmics(\pi(0), \phi(0), \epsilon_\tau, N_\tau, \epsilon, m, w)$
  $H_{new} = \frac{1}{2}\sum \pi(\tau)_i^2 + S(\phi(\tau), \epsilon, m, w)$
  $r \leftarrow$ *Generate Uniform Random Number [0,1]*
  $\alpha \leftarrow min\left(1, \frac{\boldsymbol{\pi}(x')}{\boldsymbol{\pi}(x)}\right)$
  $\alpha \leftarrow min\left(1, \exp\{-(H_{new} - H)\}\right)$
  $r \leftarrow$ *Generate Uniform Random Number [0,1]*
  if $r \leq \alpha$ **then return** $\phi(t)$
  elsereturn $\phi(t)$

Algorithm 3 Hybrid Monte Carlo II

procedure ACTION(ϕ, ϵ, m, w)
  $S \leftarrow 0$
  $L \leftarrow len(\phi)$
  for $i = 0$ *Up to* $i = L - 1$ **do**
    $S \leftarrow S + \frac{\epsilon \cdot m}{2}\left(\left(\frac{\phi[(i+1)\%L] - \phi[i]}{\epsilon}\right)^2 + w^2\phi[i]\right)$
  return S
procedure MOLECULAR_DYNAMICS($\pi(0), \phi(0), \epsilon, N_\tau, \epsilon, m, w$)
  for $i = 0$ *Up to* $i = L - 1$ **do**
    $\pi(\frac{\epsilon}{2})[i] \leftarrow \pi(0)[i] - \frac{\epsilon}{2}D_S(\phi(0), i, \epsilon, m, w)$
    $\phi(\epsilon_\tau)[i] \leftarrow \phi(0)[i] + \epsilon\pi\left(\frac{\epsilon_\tau}{2}\right)[i]$
  for $n = 1$ *Up to* $N - 1$ **do**
    for $i = 0$ *Up to* $i = L - 1$ **do**
      $\pi(n\epsilon_\tau + \frac{\epsilon_\tau}{2})[i] \leftarrow \pi(n\epsilon_\tau - \frac{\epsilon_\tau}{2})[i] - \frac{\epsilon_\tau}{2}DS(\phi(n\epsilon_\tau), i, a, m, w)$
      $\phi(n\epsilon_\tau + \epsilon_\tau)[i] \leftarrow \phi(n\epsilon_\tau)[i] + \epsilon_\tau\pi\left(n\epsilon_\tau + \frac{\epsilon_\tau}{2}\right)[i]$
  for $i = 0$ *Up to* $i = L - 1$ **do**
    $\pi(N\epsilon_\tau)[i] \leftarrow \pi(N\epsilon_\tau - \frac{\epsilon_\tau}{2})[i] - DS(\phi(L\epsilon_\tau), i, \epsilon, m, w)[i]$
procedure DS(ϕ, i, ϵ, m, w)
  return $\frac{m}{\epsilon}(2\phi[i] - \phi[(i-1)\%L] - \phi[(i+1)\%L]) + \epsilon \cdot m \cdot w^2 \cdot \phi[i]$

This algorithm Generates a set of lattices $\phi_t = \{x_n | n = 0, 1...L-1\}_t$, which can then be used to estimate the expectation value of any operator in the Vacuum state:

$$\langle O \rangle \approx \frac{1}{N}\sum_t O(\phi_t) \pm \frac{\sigma}{\sqrt{N}} \tag{2.21}$$

Where σ is estimated by calculating the IAT.

2.3.2 Results

We can now begin to make some numerical estimations on the observables. The first observable we could consider is $O(\{x_n|n = 0, 1...L-1\}_t) = x_i$ for a particular value of i. However, given that we are measuring the observables relative to the vacuum state (An energy Eigenstate and, therefore, a stationary state), the average of an observable at a particular time is equal to the average of that same observable at a different time (we can shift the index by n and the result would be the same), we can instead calculate $O(\{x_n|n = 0, 1...L-1\}_t) = \frac{1}{N}\sum_{i=0}^{L-1} x_i = \langle x \rangle$ which will converge faster. The results for the numerical estimate of $\langle x \rangle$ are shown below. the results were computed for $L = 100$, and a total of 10^5 Measurements were made. A 100 thermalization updates were made before taking any measurements.

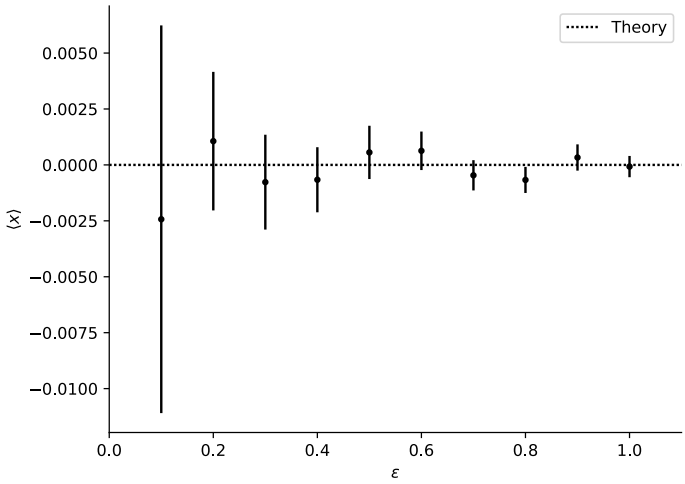


Fig. 2.1: The plot shows the estimation for the observable $\langle x \rangle$ in the QHO using HMC to generate the samples. The observable is estimated for different values of the lattice spacing, and the dotted line indicates the Theoretical Analytic value for a QHO

From the results shown in Figure 2.1, it is clear the numerical estimates $\langle x \rangle$ converge to the expected value (0 for all possible a). The plot reveals another important feature of the Markov Chain Monte Carlo: the error bars of the measured quantities increase with decreasing a. This happens because the auto-correlation between measurements increases as we approach the continuum limit, and therefore, the IAT is much higher in the limit $\epsilon \to 0$. Therefore, the number of measurements needed to reach a determined level of statistical certainty increases as the lattice spacing is reduced. This problem is usually referred to as critical slowdown.

A perhaps more interesting observable is $\langle x^2 \rangle$, which has a non-zero value. The plot in Figure 2.2 shows the numerical estimates of $\langle x^2 \rangle$ for different values of a with fixed $L = 100$. The plot of 2.2 is particularly educating as it provides a perfect example of the difference between statistical and Discretization errors. The Quantum Harmonic Oscillator Can be solved exactly in both the continuum limit and the Discrete case. Hence, the theoretical values for $\langle x^2 \rangle$ are available for the different values of a (with fixed L) and not only in the case of the Continuum Limit.

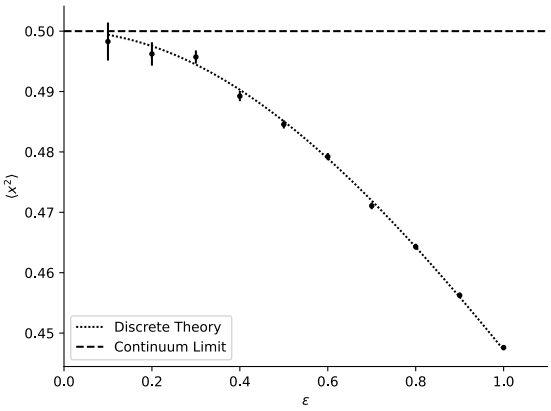


Fig. 2.2: The plot shows the estimation for the observable $\langle x^2 \rangle$ in the QHO using HMC to generate the samples. The observable is estimated for different values of the lattice spacing, and the dotted line indicates the Theoretical Analytic value for a discrete QHO with those characteristics

As the figure shows, the values obtained for $\langle x^2 \rangle$ appear to converge adequately to the theoretical values of the Discrete Theory, which itself converges to the continuous result when $\epsilon \to 0$. Therefore, when using Markov Chain Monte Carlo for estimating the values of certain observables, it is vital to bear in mind that our estimates will tend to converge (in the statistical sense, reducing the error bars) to the values of the observables in a Discrete Model, which will differ from the Continuum values (which are the ones that we are interested in calculate) and will converge to them as the lattice spacing is reduced. Generally, neither the Theoretical values for the Discrete theory nor the values in the continuum limit can be known analytically.

It is also possible to estimate the form of the square of the Wavefunction $|\langle x|0\rangle|^2 = |\psi_0(x)|^2$ by looking at the distribution of x in all the time-slices along the Markov chain. The normalized histogram of x (all the time slices were taken into account to make the histogram) along the Markov chain is shown in Figure 2.3.

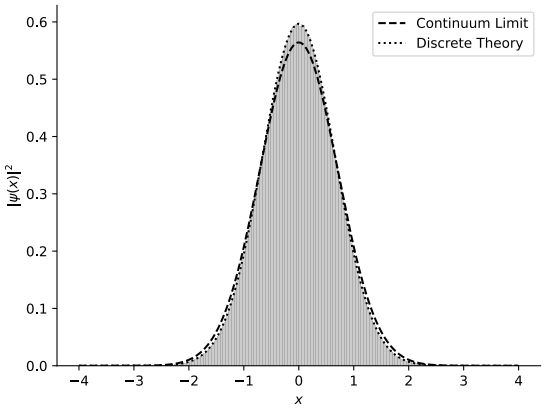


Fig. 2.3: The plot shows a histogram of the x variable along the chain.. The distribution of x is an estimation of the square of the ground state wave function, $|\psi_0(x)|^2$ for $\epsilon = 1, L = 100$. The analytic values are given in [6]

.

The Histogram shown in Figure 2.3 shows how the distribution of $x's$ along the Markov Chain tends to

$|\psi(x)^2|$ from the discrete theory, in the continuum limit (taking both $N \to \infty, \epsilon \to 0$ the discrete theory approaches the continuum theory and therefore the results of the MCMC would agree with the theoretical ones.

Finally, it is interesting to look at two-point correlation functions:

$$G(\tau) = \langle x(\tau_0)x(\tau_0+\tau)\rangle - \langle x(\tau_0)\rangle\langle x(\tau_0+\tau)\rangle \tag{2.22}$$

It is important to note that $G(\tau)$ is independent of the choice of τ_0. Therefore, it is possible to exploit the periodic boundary conditions of the lattice to improve the statistics on our estimate. In the lattice simulation, to estimate the two-point correlation function, we need to take samples of:

$$x(0)x(\tau_i) = x_0 x_i = \frac{1}{N}\sum_j x_j x_{(j+i)\%L} \tag{2.23}$$

Using the usual notation $x(0) = x_0$ and $x(\tau_i) = x_i$. Two-point correlation functions are useful as they can be used to obtain the Energy Gap between the vacuum and the first excited energy level. To see this, we start with the definition of the Two-point Correlation function in Equation 2.22:

$$G(t) = \sum_{n=0} \langle 0|\, x(\tau_0)\, |n\rangle \langle n|\, x(\tau_0+\tau)\, |0\rangle - \langle x(\tau_0)\rangle\langle x(\tau_0+\tau)\rangle \tag{2.24}$$

We have introduced a complete set of Energy Eigenstate and used the fact that $\hat{1} = \sum_{n=0} |0\rangle\langle 0|$. We will first inspect the first term on Equation 2.24 for simplicity. We use the definition of x in the Heisenberg Picture in Euclidean time, that is: $x(\tau) = \exp\{-H\tau\}x(0)\exp\{H\tau\}$

$$\sum_{n=0} \langle 0|\exp\{-H\tau_0\}x(0)\exp\{H\tau_0\}|n\rangle\langle n|\exp\{-H(\tau+\tau_0)\}x(0)\exp\{H(\tau+\tau_0)\}|0\rangle \tag{2.25}$$

We Now need to act the exponential operators on the Energy eigenstates to obtain:

$$\sum_{n=0} \langle 0|\exp\{-E_0\tau_0\}x(0)\exp\{E_n\tau_0\}|n\rangle\langle n|\exp\{-E_n(\tau+\tau_0)\}x(0)\exp\{E_0(\tau+\tau_0)\}|0\rangle \tag{2.26}$$

After reordering and simplifying, we are left with the following:

$$\sum_{n=0} \exp\{-\tau(E_n - E_0)\}|\langle 0|\, x(0)\, |n\rangle|^2 \tag{2.27}$$

The $n = 0$ term of this expression is just $|\langle 0|\, x(0)\, |0\rangle|^2$, which is exactly equal to the second term in the definition of $G(t)$, which we had been ignoring. Therefore, the second term in $G(t)$ cancels the $n = 0$ term of our expression. Therefore we are left with:

$$G(t) = \sum_{n=1} \exp\{-\tau(E_n - E_0)\}|\langle 0|\, x(0)\, |n\rangle|^2 \tag{2.28}$$

It is clear from Equation 2.28 that for sufficiently large τ, the dominant contribution will be given by the exponential of the Energy difference between the first excited state and the vacuum, that is $G(\tau) \approx A\exp\{-\tau(E_1 - E_0)\}$ in the large τ limit. On the other hand, this derivation has not considered the periodic boundary conditions in our lattice. If we include the periodic boundary conditions, we expect that in the large τ limit, the two-point correlation function will take the form of [2]:

$$G(\tau) = A\left(\exp\{-a\cdot d(E_1 - E_0)\} + \exp\{-a(N-d)(E_1 - E_0)\}\right) = B\cosh\left(a\Delta E\left(d - \frac{N}{2}\right)\right) \tag{2.29}$$

Where N is the number of lattice points, and d ($\tau = da$) goes from 0 to $N-1$. Therefore, it is possible to fit our numerical result to the cosh function to obtain a numerical estimate for $\Delta E = E_1 - E_0$. Figure

2.4 shows an example of the two-point correlation function estimate for $a = 0.2$ and $L = 100$. The plot also includes the best fit to the cosh model.

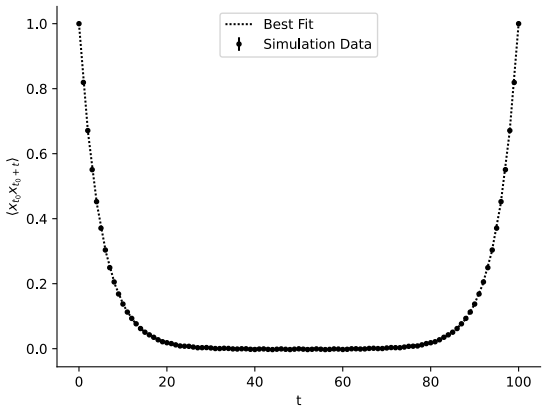


Fig. 2.4: The plot shows the estimation of the observable $\langle x_{t_0} x_{t_0+t} \rangle$ for all the allowed values of t of a Discrete QHO System with lattice spacing ϵ equal to 0.2 and 100 lattice points. On the other hand, the dashed line represents the best fit of the Cosh function model to our data, which can be used to obtain the Energy gap between E_0 and E_1 of this system.

On the other hand, in Figure 2.5, we can see the estimated energy difference for various a with fixed L. To obtain these estimations, the data of the two-point correlation function was fitted to the cosh model using Scipy. Scipy directly gave the error on the Estimated parameter. Additionally, the Theoretical curve of the Discrete theory is obtained from the results of [6]. As expected, the values converge to the ones given by the Discrete Theory.

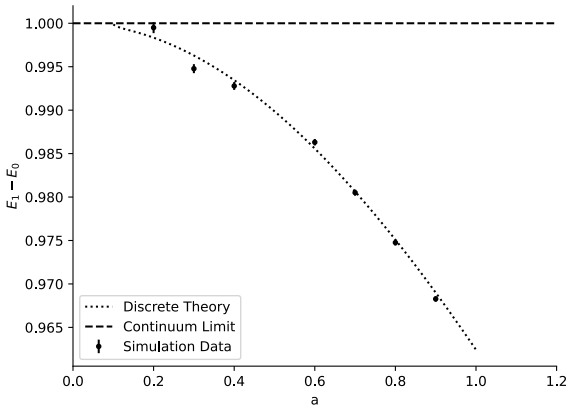


Fig. 2.5: The plot shows the estimation of the Energy difference $\Delta E = E_1 - E_0$ for different values of lattice spacing (a) of a Discrete QHO System with 100 lattice points. On the other hand, the dotted line represents the theoretical discrete values, and the dotted line represents the value of ΔE In the continuum limit. The analytic results of the discrete theory are given in [6]

.

CHAPTER 3

MCMC in Quantum Field Theory

Don't Panic.

Douglas Adams, The Hitchhiker's Guide to the Galaxy

3.1 Introduction

It is now time to generalize all the work we have done with the Quantum Harmonic Oscillator to the case of Quantum Field theory. In QFT, the main object of concern is not the position but a field $\phi(\underline{x}, t)$, which are functions of both space and time. For convenience, if x is not underlined, we will refer to the four components $x = (ct, \underline{x})$. We will use scalar fields for this derivation, but this can easily be expanded to other fields.

Therefore, the field plays the same role as position in quantum mechanics, meaning they are promoted to operators. That is, we have field operators $\hat{\phi}(\underline{x}, t)$ with a complete set of eigenstates $|\phi_b, t\rangle$ such that:

$$\hat{\phi}(\underline{x}, t) |\phi', t\rangle = \phi'(\underline{x}) |\phi', t\rangle \tag{3.1}$$

$$\hat{1} = \int \mathcal{D}\phi' |\phi', t\rangle \langle \phi', t| \tag{3.2}$$

In this case, the integration measure $\int \mathcal{D}\phi'$ is just given by:

$$\int \mathcal{D}\phi' = \int \Pi_{\underline{x}} d\phi'(\underline{x}) \tag{3.3}$$

The integral covers all possible field configurations at a given time t.

As before, the time evolution of an operator is given by:

$$\hat{\phi}(\underline{x}, \tau) = \exp\{H\tau\} \hat{\phi}(\underline{x}) \exp\{-H\tau\} \tag{3.4}$$

We are directly working with imaginary time (Wicks Rotation) such that $\tau = it$. We begin by defining the transition elements $\langle \phi_\tau, \tau | \phi_0, 0\rangle$ in terms of path integrals:

$$\langle \phi_\tau, \tau | \phi_0, 0\rangle = \langle \phi_\tau | \exp\{-H\tau\} |\phi_0\rangle = \int_{\phi_0}^{\phi_\tau} \mathcal{D}\phi \exp\{-S_E[\phi]\} \tag{3.5}$$

In this case, the Integration integration measure goes over all field configurations with boundary conditions $\phi(\underline{x}, 0) = \phi_0(\underline{x})$ and $\phi(\underline{x}, \tau) = \phi_\tau(\underline{x})$:

$$\int_{\phi_0}^{\phi_\tau} \mathcal{D}\phi \propto \int_{\phi_0}^{\phi_\tau} \Pi_x d\phi(x) \tag{3.6}$$

Additionally, The Euclidean Action S_E is obtained from integrating over all space the Lagrangian Density and then integrating over time up to τ. That is:

$$S_E = \int_0^\tau d\tau' \int dx^3 \mathcal{L}_E \tag{3.7}$$

The Euclidean Action can be obtained from the Real Action through the Wick Rotation $\tau = it$.

As before, it is possible to define the Partition function, which is given by integrating all possible paths with periodic boundary conditions $\phi_0 = \phi_\tau$:

$$Z \equiv Tr[\exp\{-\tau H\}] = \int \mathcal{D}\phi_0 \, \langle \phi_0, \tau | \phi_0, 0 \rangle = \int \mathcal{D}\phi \exp\{-S_E[\phi]\} \tag{3.8}$$

In the last step, we have omitted boundary conditions since we are integrating all possible periodic field boundary conditions. Furthermore, for a general Observable that depends on the field variables, we have:

$$\langle \hat{O} \rangle_\tau \equiv \frac{Tr(\exp\{-\tau H\}\hat{O})}{Z} = \frac{\int \mathcal{D}\phi O \exp\{-S_E[\phi]\}}{Z} \tag{3.9}$$

Where we are integrating over all possible periodic conditions in the path integral. Therefore, in the $\tau \to \infty$ limit, we will obtain the expectation value of the observable relative to the vacuum state $|\Omega\rangle$:

$$\langle \hat{O} \rangle \equiv \langle \Omega | \, \hat{O} \, | \Omega \rangle = \lim_{\tau \to \infty} \frac{\int \mathcal{D}\phi O \exp\{-S_E[\phi]\}}{Z} \tag{3.10}$$

It is important to note that in the $\tau \to \infty$ limit, The Euclidean Action S_E can be expressed as an integral over all 4 Space:

$$\lim_{\tau \to \infty} S_E = \lim_{\tau \to \infty} \int_0^\tau d\tau' \int dx^3 \mathcal{L}_E = \int dx^4 \mathcal{L}_E \tag{3.11}$$

3.2 Lattice Field Theory

Therefore, we need to find a way of numerically estimating the result of Equation 3.10 so that we can obtain estimates of the Expectation values of observables relative to the vacuum for different field theories.

In order to do so, we need to discretize all four dimensions, creating a grid of points or a Lattice. Therefore, we start by considering a cubic-like Region of Spacetime with 4 volume $V = L^4$ where L is the maximum amount we can move in each direction (from 0 to L). Then, we divide each dimension into N slices such that, in total, we obtain N^4 points. $x_\mu = n_\mu a$ where $\{n_\mu = 0, 1...N-1, \forall \mu\}$ $a = \frac{L}{N}$ with n going 0 to $N-1$. In summary, our lattice Λ is defined as:

$$\Lambda = \{x = (x_0, \underline{x}) \in aZ^4 | 0 \leq x_\mu \leq a(N-1)\} \tag{3.12}$$

Additionally, we impose periodic boundary conditions in all directions such that the region becomes a toroid, that is:

$$\phi(x) = \phi(x + Na\hat{\mu}), \forall \mu \tag{3.13}$$

where $\hat{\mu}$ is a unit vector aligned with the μ axis. Then, the Measure of the path integral is just given by:

$$\int \mathcal{D}\phi = \int \Pi_{x \in \Lambda} d\phi_x \tag{3.14}$$

The x index goes over all the possible points in the Lattice. Therefore, we will be calculating quantities like:

$$\langle \hat{O} \rangle_{L,a} = \frac{\int \Pi_{x \in \Lambda} d\phi_x O(\phi_x) \exp\{-S_E[\phi]\}}{\int \Pi_{x \in \Lambda} d\phi_x \exp\{-S_E[\phi]\}} \tag{3.15}$$

Where the Euclidean Action S_E needs to be discretized in the Lattice. If we can calculate $\langle \hat{O} \rangle_{L,a}$, then we will be able to obtain the physical quantity $\langle \hat{O} \rangle$ by taking the limit $L \to \infty$ and $a \to 0$. As in the case of Quantum Mechanics, it is possible to use MCMC to estimate the values of $\langle \hat{O} \rangle_{L,a}$.

3.3 Lattice QCD

Quantum Chromodynamics (QCD) is the theory of quarks and strong interactions, and it can be formulated in terms of Path Integrals in Euclidean time [7].

$$Z = \int \mathcal{D}A \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\{-S_{QCD}\} \tag{3.16}$$

Where the A is a vector field with components A_μ, which are three by three Hermitian traceless matrices, and $\psi, \bar{\psi}$ are quark fields with 12 components (3 for colours times 4 for spin) [7].
The QCD action can be expressed as the sum of the gluon action and the fermion action:

$$S_{QCD} = S_G + S_F = S_G + \bar{\psi} M \psi \tag{3.17}$$

Where M is the Dirac operator. It is important to note that S_G does not depend on the Quark fields [7], so the Integration over them can be performed:

$$Z = \int \mathcal{D}A \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\{-S_{QCD}\} = \int \mathcal{D}A \lceil et(M) \exp\{-S_G\} \tag{3.18}$$

The quenched approximation consists of approximating $det(M) \approx C$ where C is a constant, and it corresponds to ignoring all the quarks of the theory and the problems they involve [8]. Nevertheless, QCD without quarks is an interesting and highly non-trivial theory mainly due to the Self-interaction of gluons [8].

3.3.1 Pure Gauge Theory

The continuum Euclidean Action for a Pure Gauge Theory (no quarks) is given by:

$$S_G = \frac{1}{2g^2} \int dx^4 Tr(F_{uv} F^{uv}) \tag{3.19}$$

Where g^2 is the coupling constant and Einstein's summation convention is assumed. It is important to note that as we are working in Euclidean time (τ), there is no difference between covariant and contravariant indices, that is, $F_{uv}(x)F^{uv}(x) = F_{uv}(x)F_{uv}(x)$[8]. The components of the tensor field F are three-by-three traceless Hermitian matrices [8]; those components (F_{uv}(x)) are defined as:

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_v A_\mu(x) + i[A_\mu(x), A_\nu(x)] \tag{3.20}$$

Where the A constitutes a gauge vector field with components A_μ, which are three-by-three hermitian traceless matrices [8]. One of the properties of this Action is that it is Gauge invariant. Under a Gauge Transformation, the Field Tensor Transforms as:

$$F_{\mu\nu}(x) \to F'_{\mu\nu}(x) = \Omega(x) F_{\mu\nu}(x) \Omega^\dagger(x) \tag{3.21}$$

Where $\Omega(x)$ is an arbitrary element of $SU(3)$ [8]. From the definition of S_E, it is clear that:

$$S'_E = \frac{1}{2} \int dx^4 Tr(F'_{uv}(x) F^{uv}(x)') = S_E \tag{3.22}$$

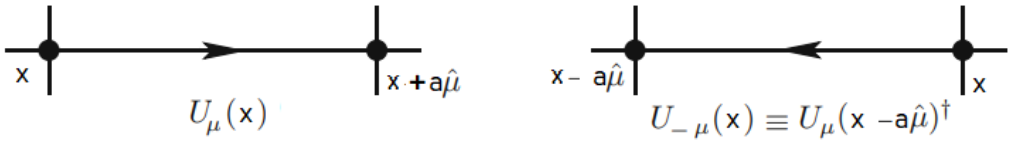
Therefore, the partition function for Lattice QCD in the quenched approximation is given by:

$$Z = \int \mathcal{D}A \exp\left\{-\frac{1}{2}\int dx^4 Tr(F_{uv}(x)F^{uv}(x))\right\} \tag{3.23}$$

We now need to build a lattice formulation of this theory. As before, we define the Lattice Λ with a total of N^4 points. It turns out that it is impossible to formulate an exact Gauge invariant Lattice theory in terms of $A_\mu(x)$. Therefore, instead of specifying the values of $A_\mu(x)$ at the lattice sites, we define the link variables $U_\mu(x)$.

$$U_\mu(x) \equiv \mathcal{P}\exp\left\{-i\int_x^{x+a\hat{\mu}} A\cdot dy\right\} \approx \exp\{-iagA_\mu(x)\} \tag{3.24}$$

Where $\mathcal{P}$ is an operator that path orders A along the integration line. The link variable $U_\mu(x)$ can be pictorially represented by a line from site x to site $x + a\hat{\mu}$. On the other hand, $U_\mu^\dagger(x)$ is represented by a line from site $x + a\hat{\mu}$ to x. The line from x to $x - a\hat{\mu}$ is given by $U_\mu^\dagger(x - a\hat{\mu})$, but for convenience, we will, in some instances, use the notation $U_{-\mu}(x) \equiv U_\mu^\dagger(x - a\hat{\mu})$



(a) $U_\mu(x)$ (b) $U_{-\mu}(x)$

Fig. 3.1: Graphical Representation of the Link Variables, adapted from Reference [9]

The Link variable $U_\mu(x)$ belongs to the SU(3) group, and under a Gauge transformation, the Links transform like:

$$U_\mu(x) \to U'_\mu(x) = \Omega(x)U_\mu(x)\Omega(x + a\hat{\mu}) \tag{3.25}$$

Additionally, we can define a Wilson loop along a closed path along the Lattice as:

$$W(\zeta) = \mathcal{P}\exp\left\{-i\oint_\zeta A\cdot dy\right\} \tag{3.26}$$

Which can be expressed as the product of link variables in the Lattice. For example, we can consider the Wilson Loop along a square in the μ, ν plane, which is known as Plaquette:

$$U_{uv} = \mathcal{P}\exp\left\{-i\oint_{\square_{\mu\nu}} A\cdot dy\right\} = U_\mu(x)U_\nu(x + a\hat{\mu})U_\mu^\dagger(x + a\hat{\nu})U_\mu^\dagger(x) \tag{3.27}$$

We now want to expand $U_{\mu\nu}$ in terms of a. To do so, we first need to use the Baker-Campbell-Hausdorff formula [8] for the exponential of matrices:

$$\exp\{A\}\exp\{B\} = \exp\{A + B + [A, B] + ...\} \tag{3.28}$$

Where A and B are matrices. Doing this with $U_{\mu\nu}$ gives:

$$U_{\mu\nu} = \exp\{iaA_\mu(x) + iaA_\nu(x + a\hat{\mu}) - \frac{a^2}{2}[A_\mu(x), A_\nu(x + a\hat{\mu})] \tag{3.29}$$

$$- iaA_\mu(x + a\hat{\nu} - iaA_\nu(x) - \frac{a^2}{2}[A_\mu(x + a\hat{\nu}), A_\nu(x)]) \tag{3.30}$$

$$+ \frac{a^2}{2}[A_\nu(x + a\hat{\mu}), A_\mu(x + a\hat{\nu})] + \frac{a^2}{2}[A_\mu(x), A_\nu(x)] \tag{3.31}$$

$$+ \frac{a^2}{2}[A_\mu(x), A_\mu(x + a\hat{\nu})] + \frac{a^2}{2}[A_\nu(x + a\hat{\mu}), A_\nu(x)] + O(a^3)\} \tag{3.32}$$

If we now Taylor expand $A_\nu(x+a\mu) = A_\nu(x) + a\partial_\mu A_\nu + \mathcal{O}(a^2)$ And collect all the terms up to a^2 we obtain:

$$U_{\mu\nu} = \exp\{ia^2(\partial_\mu A_\nu(x) - \partial_\nu A\mu(x) + i[A_\mu(x), A_\nu(x)]) + \mathcal{O}(a^3)\} = \exp\{ia^2 F_{\mu\nu}(x) + \mathcal{O}(a^3)\} \tag{3.33}$$

Now we define the Wilson action $S_W[U]$:

$$S_W[U] = \frac{2}{g^2}\sum_{x\in\Lambda}\sum_{\mu<\nu} ReTr(\mathbf{1} - U_{\mu\nu}) = \frac{a^4}{2g^2}\sum_{x\in\Lambda}\sum_{\mu,\nu}(tr[F_{uv}^2] + O(a^2)) \tag{3.34}$$

Therefore, in the continuum limit with $a \to 0$ and $a^4\sum_{x\in\Lambda} \to \int dx^4$, The Wilson Action $S_W[U]$ converges to the QCD action in the Quenched Approximation $S_G[A]$ [8]. Usually, the Wilson action is given in terms of $\beta = \frac{6}{g^2}$, $S_W[U] = \frac{\beta}{3}\sum_{x\in\Lambda}\sum_{\mu<\nu} ReTr(\mathbf{1} - U_{\mu\nu})$ [9].
Therefore, we need to estimate path integrals with the form:

$$Z = \int \mathcal{D}U \exp\{-S_W[U]\} \tag{3.35}$$

With $\mathcal{D}U = \Pi_{x\in\Lambda}\Pi_{\mu=0}^3 dU_\mu(x)$

3.3.2 MCMC in Lattice QCD

We are interested in calculating:

$$\langle\hat{O}\rangle_{L,a} = \frac{\int \mathcal{D}U \exp\{-S_W[U]\}}{Z} \tag{3.36}$$

Whith $\mathcal{D}U = \Pi_{x\in\Lambda}\Pi_{\mu=0}^3 dU_\mu(x)$. Markov Chain Monte Carlo is necessary in order to estimate the Integral of Equation 3.36). In order to create a Markov Chain that allows us to sample field configurations, a version metropolis algorithm was implemented. Note that it is also possible to implement a version of Hybrid Monte Carlo, but its implementation is non-trivial and outside the scope of this paper.

At the start of the program, a number M (which is a parameter given by the user) of random elements of $SU(3)$ which deviate from the identity by a factor of ϵ (Another parameter given by the user). Therefore, the new candidate states will be generated by multiplying the to-be-updated link U by a random element of the generated Set. Given that the metropolis algorithm requires that the probability of proposing a U' given state U must be the same as the probability of proposing link U given U' when generating the M random elements of SU(3), we must also store the inverses (the hermitian conjugates for SU(3) elements), giving a total set of $2M$ possible matrices to choose from. Let χ denote the Set of all these matrices. For the chain to be able to reach all possible states (to be ergodic), it is necessary that the χ contains sufficient elements such that all the group elements are reachable by successive multiplication with χ elements. It is also a good idea to refresh the Set χ at some point during the run, but this was not implemented in the code.

Once the candidate Matrices have been generated, the metropolis algorithm begins. We iterate sequentially over all possible links U_μ in the Lattice for each link U_μ. We propose a new candidate link through the multiplication with one of the elements $K \in \chi$, that is:

$$U'_\mu = KU_\mu \tag{3.37}$$

As both M and U_μ are elements of $SU(3)$, so is U'_μ (group axiom). Then, the Action of the system is calculated with the new link U'_μ. The Action is given by:

$$S_E = \frac{\beta}{3}\sum_{x\in\Lambda}\sum_{\mu,\nu<\mu} ReTr(\mathbf{1} - P_{\mu\nu}(x)) \tag{3.38}$$

$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x + a\hat{\mu})U_\mu^\dagger(x + a\hat{\mu} + a\hat{\nu})U_\nu^\dagger(x) \tag{3.39}$$

The total change in the Action ΔS produced by the change $U_\mu(x) \to U'_\mu(x)$ can be calculated by looking at the local contribution of the updated link to the Action. That is:

$$S[U_\mu(x)]_{loc} = \frac{\beta}{3}\sum_{i=1}^{6} ReTr(\mathbf{1} - U_\mu(x)P_i) = \frac{\beta}{3} ReTr(6\cdot\mathbf{1} - U_\mu(x)A), A = \sum_{i=1}^{6} P_i \tag{3.40}$$

Where The P_i are the products of the other three links that constitute the Plaquette together with $U_\mu(x)$, each link contributes to six plaquettes. Therefore, the sum goes from 1 to 6. These products are also referred to as Staples, and it is possible to express A explicitly as a sum over all possible Staples for a given link U_μ(x):

$$A = \sum_{\nu\neq\mu}(U_\nu(x+a\hat{\mu})U_{-\mu}(x+a\hat{\mu}+a\hat{\nu})U_{-\nu}(x+\hat{\nu}) + U_{-\nu}(x+a\hat{\mu})U_{-\mu}(x+a\hat{\mu}-a\hat{v})U_\nu(x-a\hat{\nu})) \tag{3.41}$$

Alternatively, Using the hermitian conjugate for the links with a negative index, we find that:

$$A = \sum_{\nu\neq\mu}(U_\nu(x+a\hat{\mu})U^\dagger_\mu(x+a\hat{\nu})U^\dagger_\nu(x) + U^\dagger_\nu(x+a\hat{\mu}-a\hat{\nu})U^\dagger_\mu(x-a\hat{v})U_\nu(x-a\hat{\nu})) \tag{3.42}$$

Therefore, the total change in Action due to the change $U_\mu(x) \to U'_\mu(x)$ is given by:

$$\Delta S = S[U'_\mu(x)]_{loc} - S[U_\mu(x)]_{loc} = -\frac{\beta}{3} ReTr((U'_\mu(x) - U_\mu(x))A) \tag{3.43}$$

After calculating the change action, we perform the usual metropolis step to decide whether to accept or reject the new state. However, given that calculating A is the most expensive part of the process and that this matrix is independent of the particular value of $U_\mu(x)$ (it only depends on x and μ), it is more efficient that each time we visit a particular $U_\mu(x)$, we do not just update it once but a fixed number of times specified by the user, which helps $U_\mu(x)$ to equilibrate with its surroundings. We call a sweep when the algorithm has gone once over all the links of the system. We record the state of the Lattice when a sweep has been completed. This is repeated for as many sweeps as it is required to obtain a significant result. It is important to note that as we are going over the links sequentially, this algorithm does not satisfy detailed balance. However, detailed balance is only a sufficient condition for the stationary distribution to exist, not a necessary condition. That is, our algorithm still has the correct distribution as a stationary distribution while not satisfying detailed balance. The pseudo-code for the algorithm is shown below.

```
1:  procedure METROPOLIS(Sweeps, M, ϵ ,hits)
2:      χ ← Initialize The set of 2M matrices that deviate from the identity by ϵ
3:      Set of Links ← Initialize all links to the identity
4:      for t = 1 up to t = Sweeps do
5:          for for U_μ(x) in Set of Links do
6:              for hit = 1 up to hit = hits do
7:                  K ← Random Element of χ
8:                  U'_μ ← K U_μ
9:                  S_old ← Action with U_μ
10:                 S_new ← Action with U_μ
11:                 r ← Generate Uniform Random Number [0,1]
12:                 α ← min(1, exp{−(S_new − S_old)})
13:                 if r ≤ α then
14:                     U_μ ← U'_μ
15:                 else
16:                     U_μ ← U_μ
17:         Store Lattice links with Markov time t
```

Now that we have a working algorithm, we can verify if it is working by measuring the expectation value of some observable and comparing it with its literature value. The easiest operator is probably the average Plaquette, which is defined as:

$$O_P \equiv \frac{\beta}{N_{plaquettes}} \sum_{x \in \Lambda} \sum_{\mu,\nu<\mu} (1 - P_{\mu\nu}) \tag{3.44}$$

For the case of a four-dimensional lattice, the number of plaquettes is given by $N_{plaquettes} = 6|\Lambda|$ where $|\Lambda|$ is the number of points in the Lattice. Reference [10] suggests that for $|\Lambda| = 8^4$ and $\beta = 5.5$, the average plaquette should be near 0.5. The value obtained for that particular configuration using the procedure outlined above with $2 \cdot 10^4$ groups of one measurement sweep and one discarded sweep with an initial 100 sweeps for thermalization is 0.4969 ± 0.0001, which validates our implementation. On the other hand, Figure 3.2 shows the measured average Plaquette for different values of β. If we look at the IAT for the different data points, it is easy to see that it increases as β increases. Given that we know about the problem of Critical Slowing Down, it is possible to infer just from the data that the continuous limit is obtained when increasing β, which gives finer lattices (smaller lattice spacing).

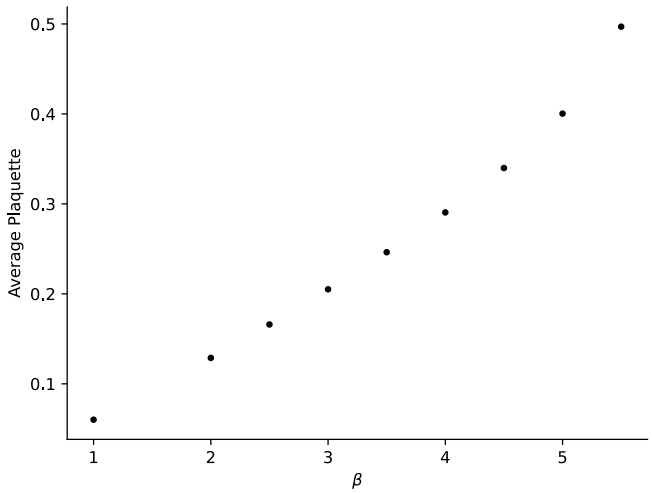
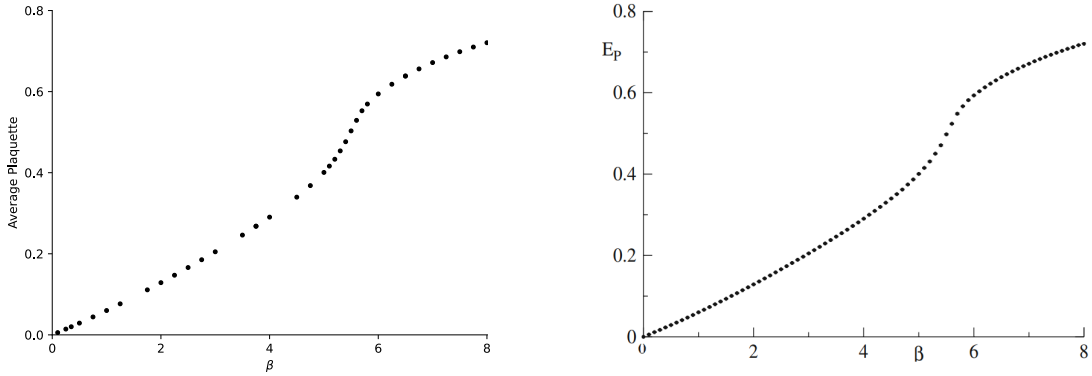


Fig. 3.2: The plot shows the estimation of the Average Plaquette for a pure Gauge SU(3) theory for different values of β in a Lattice with 8^4 points. One hundred thermalization runs were made, followed by $2 \cdot 10^4$ groups of one measurement update and one discarded update for each data point. The Error bars are too small to be appreciated.

To verify our implementation further, the Average Plaquette was calculated for different values of β for a 4-dimensional lattice with $|\Lambda| = 12^4$ points. The results can then be compared to the ones obtained by Reference [9]. The results are shown in Figure 3.3. It can be seen that the two graphs show the same behaviour and values for the range of β that was investigated. It is important to mention that the error bars in the collected data points of Figure 3.3 (which cannot be appreciated in the figure) might be underestimated since our estimation for the IAT is valid for Markov Chains with length $\approx 1000\tau_{IAT}$, and our chains are too small to fulfil this criterion. Other error analysis methods should be implemented for short chains, such as Jackknife or Bootstrap. While this might be the case, the close resemblance between the values and behaviour of the two plots can still be used to validate our implementation.



(a) Collected Data

(b) Reference Plot

Fig. 3.3: Plot a) shows the estimation of the Average Plaquette for a pure Gauge SU(3) theory for different values of β in a Lattice with 12^4 points. Fifty thermalization runs were made, followed by 200 groups of one measurement update and one discarded update for each data point. The Error bars are too small to be appreciated. On the other hand, the same plot is shown in b), but the data was obtained from Reference [9]

CHAPTER 4

Conclusion

> Mysteries are always more exciting than truths.
>
> V.E. Schwab, A Gathering of Shadows

4.1 Future Work

The relatively short duration of this project means that there are many ways in which it could be expanded. Additionally, some features that were implemented, such as the Static Quark Potential, were intentionally left out of this report as it was oriented as a first introduction to lattice field theory, and it already exceeds by far the typical extent of Summer Project Reports. This section will provide a small guide to some possible ways of extending the project.

Firstly, more work could be done in the Lattice QCD area. The first step would be to set the scale of the theory, that is, extract the value of a, the lattice spacing, from β, the coupling constant, as in this theory, the value of a is not a parameter of the algorithm. One way of working this out is by calculating the Static Quark potential, that is, the potential between two infinitely heavy quarks. Doing this would involve calculating the values of the average Rectangular Wilson loop with two parallel sides on a spatial direction and length R and the other two along the time axis with length T. By calculating this for different values of R and T, it is possible to extract the Potential between two infinitely heavy (static) quarks, and then the value of a can be extracted from it (given a particular value of β).

Once the scale of the theory has been defined with the use of the Static Quark potential, it would be interesting to look at the possible masses of Glueballs (Excited States). As in the case of Quantum Mechanics, to do this, we just need to look at the two-point function of two operators. A possible option would be to look at the two-point function (or correlation function) of the plaquette-plaquette. However, this would probably give a very noisy signal in such a small lattice (8^4); a better option would be first to reduce the number of dimensions so that this can be studied with more care.

Additionally, the next logical step would be to leave the quenched approximation and to include Fermions in the lattice calculations so that it is possible to make a simulation with the complete theory. However, including fermions does not come without its own complications, which is why it was not attempted during the relatively short duration of the Summer Project. A First step towards achieving this goal would be to

implement Hybrid Monte Carlo (HMC) for the SU(3) Gauge Theory. However, this implementation is not as trivial as in the case of the Quantum Harmonic Oscillator.

Finally, on a more general note, it would be interesting to study further Critical Slowing Down and to investigate methods to suppress it, such as Fourier Acceleration. This method could be first implemented in a system with a known solution, such as the Quantum Harmonic Oscillator. Even more, it would be interesting to compare how HMC with Fourier Acceleration performs against plain HMC.

4.2 Final Thoughts

The intention behind this report was to create a simple introduction to Lattice Field Theory, being as thorough as possible in all the areas that this field encapsulates. In the First Chapter, we introduced the Method Of Markov Chain Monte Carlo to perform multi-dimensional integrals and particular care was taken in explaining the theory behind Markov Chains. The Second Chapter continued with a simple example in which Markov Chain Monte Carlo is used to solve the Quantum Harmonic Oscillator in non-relativistic Quantum Mechanics. That chapter allowed us to build the basic language of Lattice Calculations and to become familiar with some of its features, such as the critical slowdown, two-point correlation functions or the projection of the path integral into the vacuum. That chapter also allowed us to develop a simple code for HMC, which can be tested against the analytic solution which is available for this system. Finally, chapter 3 Generalised many of the concepts of the previous chapter to the case of Quantum Field Theory. The last calculations on Lattice QCD are the culmination of this project and are intended to look forward to what can be done using these methods in Modern QCD.

Bibliography

Elend: I kind of lost track of time
Breeze: For two hours?
Elend: There were books involved.

Brandon Sanderson, The Well of Ascension

[1] Douc, R., Moulines, E., Priouret, P. & Soulier, P. Markov Chains on a Discrete State Space. In *Markov Chains*, 145–164 (Springer International Publishing, Cham) (2018). Series Title: Springer Series in Operations Research and Financial Engineering.

[2] Joseph, A. *Markov Chain Monte Carlo Methods in Quantum Field Theories: A Modern Primer*. SpringerBriefs in Physics (Springer International Publishing, Cham) (2020).

[3] Jones, G. L. On the Markov chain central limit theorem. *Probability Surveys* **1**(none), 299 (2004).

[4] Montvay, I. & Münster, G. Introduction. In *Quantum Fields on a Lattice*, 1–53 (Cambridge University Press) (2009).

[5] Westbroek, M. J. E., King, P. R., Vvedensky, D. D. & Dürr, S. User's guide to Monte Carlo methods for evaluating path integrals. *American Journal of Physics* **86**(4) (2018).

[6] Creutz, M. & Freedman, B. A statistical approach to quantum mechanics. *Annals of Physics* **132**(2), 427 (1981).

[7] Davies, C. Lattice QCD - A guide for people who want results (2005). ArXiv:hep-lat/0509046.

[8] Gattringer, C. & Lang, C. B. Pure gauge theory on the lattice. In Gattringer, C. & Lang, C. B. (editors), *Quantum Chromodynamics on the Lattice: An Introductory Presentation*, Lecture Notes in Physics, 43–71 (Springer, Berlin, Heidelberg) (2010).

[9] Gattringer, C. & Lang, C. B. Numerical simulation of pure gauge theory. In Gattringer, C. & Lang, C. B. (editors), *Quantum Chromodynamics on the Lattice: An Introductory Presentation*, Lecture Notes in Physics, 73–101 (Springer, Berlin, Heidelberg) (2010).

[10] Lepage, G. P. Lattice QCD for Novices (2005). ArXiv:hep-lat/0506036.