Bachelor's Thesis

The Ground State Wavefunction Of The Universe

Department of Physics
Lahore University Of Management Sciences

Malaik Kabir

November 27th, 2023

Submitted in partial fulfillment of the requirements for the degree of B.S Physics
Supervised by Dr. Syed Moeez Hassan

For Amma and Baba
Without your support I would never have had the privilege of thinking about everything that follows

Contents

1 Acknowledgments

I would first and foremost like to thank Dr. Moeez for being patient with me throughout my Sproj and helping me everytime I got stuck. I don't think I would have gotten as much work done had I been doing it under somebody else's supervision and for that, I am grateful.

This thesis is the culmination of a year's worth of work. Over the course of that entire year I have had the constant support of many friends that I would like to name. Rehan, for all of the conversations we've had where I have been forced to take my eyes off of the screen or the paper and look at the bigger picture. Daniel for all of the love that he has shown me. Hamza, for all of his jokes. Hania, for helping me in situations I'd rather not mention. Abeera and Irtaza for playing bomber robot with me on the computer. Manahil for making us all go out based on impromptu plans. And then all of the friends outside of the Physics department. Hassan, for always being by my side no matter what. Aimal, for hyping me up when I thought I did not have it in me. Arhum, for being the best roommate these past four years, Usman, for meeting me every week. Alishba, for being the biggest fan of my sproj. Waseem, for checking up on me even though I sometimes forget to check up on him. Baria, for being somebody I can count on consistently for four years. Taneer, Ayesha, Adeen, Yazook, Maira, Muneeb and everybody who has shaped me into the person I am in my senior year. From the bottom of my heart, thank you!

2 Path Integrals

Two roads diverged in a yellow wood,
And yes, good sir, I travelled both
- the electron (probably)

2.1 Why let go of classical mechanics?

Quantum mechanics is a deeply un-intuitive theory. At the face of it, it seems to require us to completely discard all of the tenets of classical mechanics: the notion that a particle could have a definite position and momentum, that it could take a particular path in space etc. Take, for example, the infamous double slit experiment with electrons as the particles under consideration. We label the slits as 1 and 2 as shown in the figure:

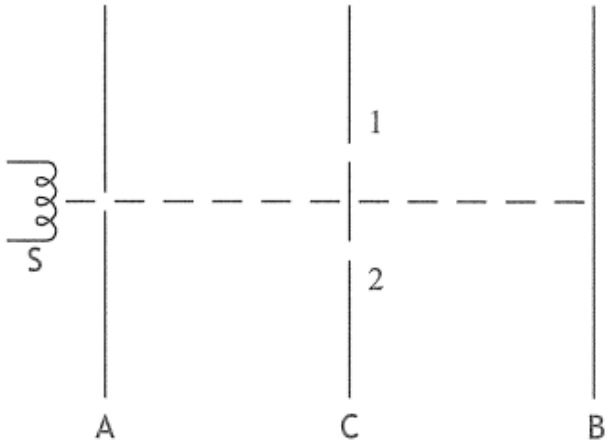

Figure 1: The double-slit experiment

The most jarring fact about the entire experiment is of course that the probability distribution $\phi(x)$ for the particle to land at some position x on the screen when both slits are open, is **not** the sum of the individual probability distributions $P_1(x)$ and $P_2(x)$, where we have defined $P_i(x)$ to be the the probability distribution observed on the screen when only slit i is open. That is, if Fig. 2(b) is $P_1(x)$ and Fig. 2(c) is $P_2(x)$ - with x plotted in the vertical direction - then we expect the distribution to look like Fig. 2(d), the sum of the individual probability distributions. In case you're late to the party and don't know already, it actually looks like Fig. 2(a), which is $P(x)$.

A straightforward solution to the problem is of course to completely disregard the notion that the particle is "going through" one or the other slit prior to being observed. This is exactly the route taken by Schrodinger, Heisenberg and many of the other founders of quantum mechanics. This is also exactly the route that most textbooks on quantum mechanics follow. In the context of such an interpretation, before observing the particle the most you can talk about is its "wavefunction". And since the

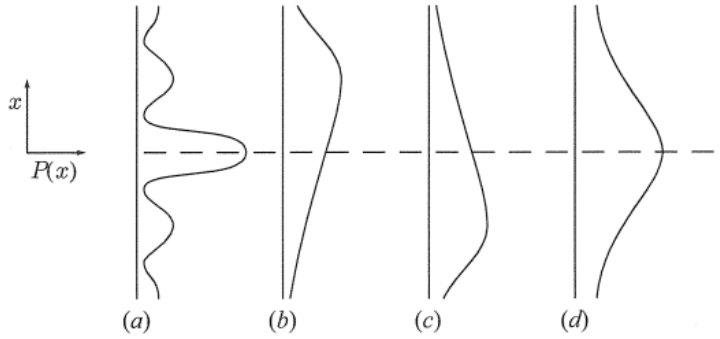

Figure 2: Probability Distributions

wavefunction is **not** the probability distribution, but rather the *probability amplitude* (which we will henceforth denote using $\phi(x)$), it makes sense that the probability distribution observed on the screen is the one given by Fig. 2(a): it is the *probability amplitudes* that add, not the *probability distributions*. Then by the definition of the probability distribution as the modulus squared of the probability amplitude, we get:

$$\begin{aligned} P(x) &= |\phi_1(x) + \phi_2(x)|^2 \\ &= |\phi_1(x)|^2 + |\phi_2(x)|^2 + 2[(\phi_1\phi_2^*)(\phi_1\phi_2^*)^*] \\ &= P_1 + P_2 + 2Re(\phi_1\phi_2^*) \end{aligned} \tag{1}$$

Were it not for the last term, no interference would be observed, and $P(x)$ would have been the sum of $P_1(x)$ and $P_2(x)$.

But lets step back for a moment. In the process of defining the wavefunction we had to let go of the sweet sweet handholds of classical mechanics. We had to leave the familiar territory of definite positions and trajectories and tread into a world of uncertainty. Is there a way to still keep some of these handholds? Is there a middle-ground between the certainty of classical mechanics and the absurdly un-intuitive concept of the wavefunction? In other words, is there an interpretation of quantum mechanics that can incorporate the idea of particles following trajectories, something that we are so accustomed to in classical mechanics, while still describing reality accurately?

The answer to this question is Richard P. Feynman's Path Integral Formulation of Quantum Mechanics. Such a formulation actually arises quite naturally from an extension of the double slit experiment. Here's what this *extension* looks like:

Let us ignore the double slit experiment for a moment and instead try to calculate the amplitude for a particle starting at the source to end up at some point x on the screen. To do so let us imagine putting more screens in between the source and the first screen, and then drilling holes into each of these (Fig. 3). As discussed previously, there is an amplitude associated with each combination of holes that the electron passes through. For example, there is a probability amplitude associated with the path where the electron passes through the first hole in screen E, fourth in screen D and second in

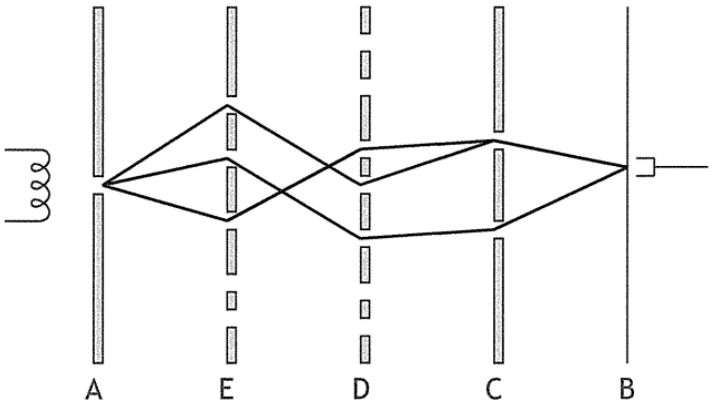

Figure 3: Infinite Screens

screen C. Each of these amplitudes must be added together in order to get the total final amplitude. In fact, if we label the holes in screens E, D and C using x_{Ei}, x_{Dj} x_{Ck} (with the indices i, j and k being stand-ins for the hole numbers), then the total final amplitude is given by the following triple sum:

$$\phi(x) = \sum_i \sum_j \sum_{k=1}^{2} \phi(x_{Ei}, x_{Dj}, x_{Ck}) \tag{2}$$

Where $\phi(x_{Ei}, x_{Dj}, x_{Ck})$ is the amplitude associated with the particular path associated with the combination (i, j, k) for holes in screens (E, D, C). In the limit of an infinite number of holes in all three screens, the sums turn into integrals over the continuous variables x_E, x_D and x_C and ϕ becomes a *continuous* multivariable function of these variables:

$$\phi(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(x_E, x_D, x_C)\, dx_E\, dx_D\, dx_C \tag{3}$$

What if we keep adding more intermittent screens? It seems, in the beginning, like we're in a bit of a rut. Why do we need to keep adding screens and holes. Aren't we complicating things further? Well, it turns out that we can imagine space as being made of exactly such screens. An infinite number of them, in fact, with an infinite number of holes. In this limit, our triple integral actually becomes an *infinite* integral. At the same time, our combination (i,j,k) becomes an infinite vector, or to put it in more bland words *a function*. On the other hand the elusive ϕ now turns into a *functional*: a function of a function. In this case, the function in the argument is the path x(y) that the particle takes (with y pointing in the horizontal direction). We can introduce, instead, a parameter t that turns the path into the vector $(x(t), y(t))$. Then our final expression for the amplitude becomes the following infinite integral:

$$K(b,a) = \lim_{N\to\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} .... \int_{-\infty}^{+\infty} \phi[(x(t), y(t))]\, dx_1\, dx_2 \,...\, dx_N \tag{4}$$

Here x_i denotes the position of the particle at the ith screen. Notice that we have renamed our amplitude, it is now called $K(b,a)$. This is intentional. What we have obtained is called the Kernel or amplitude for going from point a to point b. Later in this section, we will make this statement more precise and also discuss the kind of information that the Kernel provides us.

While we do have the general form of our path integral, we are far from finished. As an example: we are completely unaware about the form of the amplitude functional $\phi[(x(t), y(t))]$. Surely not all paths contribute equally towards the final amplitude. If they did we would have a sum of a finite functional an infinite amount of times which is divergent and wouldn't make sense. There is yet another problem, the path integral is not normalized. It turns out that the solution to the first problem is the following: the weight given to each path is a function of the *action* of that path. Lets briefly review the concept of classical action.

2.2 A short aside - The action

Classical mechanics is deterministic. Given an initial state (position and velocity) of a particle and the potential it experiences, one can find - in theory - the path that the particle will follow for all time. One way of finding such a path is by solving Newton's second law equation:

$$F = m\ddot{x} \tag{5}$$

The force in this equation can be found by the negative of the gradient of the potential:

$$F = -\nabla V \tag{6}$$

From this, one can find the acceleration and therefore the path that the system follows for all subsequent times. Yet another way of doing this is through the Lagrangian formalism. In this formalism, one considers the quantity:

$$\mathcal{L}(x, \dot{x}, t) = T - V \tag{7}$$

known as the Lagrangian of the system. Where t is a parameter, usually time. T is the kinetic energy and V the potential energy. Then, one associates with each possible path that can be followed by the particle a quantity:

$$S = \int_{t_A}^{t_B} \mathcal{L}(x, \dot{x}, t) dt \tag{8}$$

known as the action for a particular path. The action is a *functional*, it is a function of the path chosen. It turns out that the path actually followed by a particle is a *critical point*, of the action. That is, for a small change in path, the action remains unchanged. This is known as the **principle of least action**. From this requirement, one can derive what is known as the Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = \frac{\partial \mathcal{L}}{\partial x} \tag{9}$$

Any path of *least action* (or more accurately *critical action*) is also one that satisfies the above equation at all times.

2.3 Constructing the path integral carefully

Having understood the gist of the entire process, we now attempt to construct the path integral a little more carefully, taking care of the normalization and the weights of the paths along the way. During the entire process, we will take inspiration from a remotely similar, but much more familiar mathematical construct: The Riemann Sum.

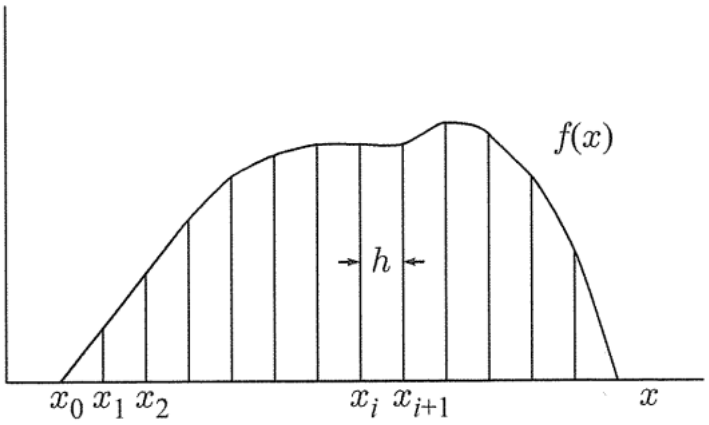

Figure 4: The Riemann Sum

Recall that the Riemann sum is an *approximation* to a value A, the area under a curve $f(x)$ as shown in Figure 4. In the case of a finite sum, the area under the curve goes as the sum of the y values (or heights) of each of the rectangles i.e if for rectangle i, the x value is x_i and the height is $f(x_i)$, then we can say that:

$$A \sim \sum_i f(x_i) \tag{10}$$

Then, in order for to approximate the area more accurately, we require that the width of the rectangles h decreases, while simultaneously increasing the number of rectangles N so that we always have:

$$Nh = \text{constant} \tag{11}$$

In the limit that N approaches infinity, the Riemann sum becomes the area of the integral. But there's a catch. And this catch will be important to us when constructing the path integral in a similar manner. As it stands, the relation (11) is not complete. As we increase the number of rectangles, the number of terms in the sum increases and it diverges. What we require is some *normalization factor*. It just so happens that in the case of the Riemann sum, this normalizing factor is the width of the rectangles itself. Thus we finally get the equality:

$$A = \lim_{N\to\infty}\left[\sum_i f(x_i)\right] \tag{12}$$

The inspiration that we will take from this is the following. We will first consider a discrete *lattice* of N time steps t_i. Each of these time steps will be spaced ϵ units apart. At each time step, the particle will be allowed to take any value of position. We will denote the position at the time step t_i by the continuous variable x_i. In such a situation, a path from some spatial point x_a (at time t_a) to some other spatial point x_b (at time t_b) becomes a simple interpolation of these spatial variables at different time steps as shown in Figure 5.

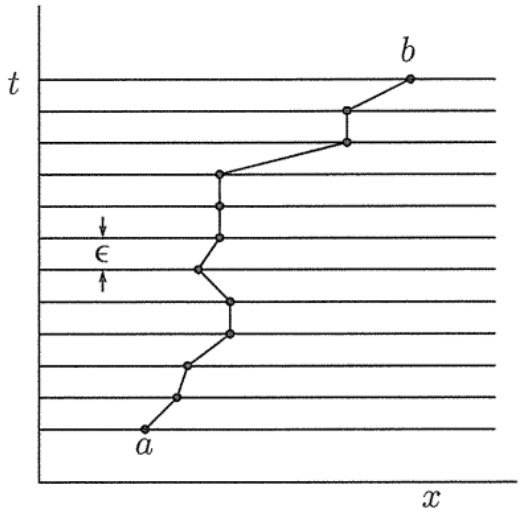

Figure 5: The Discrete Path Integral

From our discussion in section 2.1, it is quite immediately obvious that the analog to relation (11) in this case should be:

$$K(b,a) \sim \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} .... \int_{-\infty}^{+\infty} \phi[(x(t))]\, dx_1\, dx_2\, ...\, dx_{N-1} \tag{13}$$

Notice how even though there are N time steps, there are only N-1 integrals. This is because we are considering paths that start from x_a and end at x_b which fixes one of the variables (in this case x_N).

The next step follows quite naturally. In a similar fashion to what we did in the Riemann Sum case, in order to make our jagged interpolations more similar to a realistic path that might be followed by a particle in space, we increase the number of time steps while simultaneously decreasing the interval ϵ between them, all the while making sure that:

$$N\epsilon = t_b - t_a \tag{14}$$

However, note that we are still not quite done. As in the Riemann sum case, relation (14) is incomplete. Increasing the number of time steps will only increase the number of terms in the sum and it will not converge to any sensible value as it stands. What we require, yet again, is something to tame the beast: a normalization constant. It turns out that there is no one normalization constant for path integrals as there was in the Riemann Sum case. Whatever the normalization constant might be, however, it can be written in the form A^{-N}. Then, applying the limit, our relation turns into an equality:

$$K(b,a) = \lim_{\epsilon \to 0} \frac{1}{A} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} .... \int_{-\infty}^{+\infty} \phi[(x(t))]\, \frac{dx_1}{A}\, \frac{dx_2}{A}\, ...\, \frac{dx_{N-1}}{A} \tag{15}$$

All that remains is the form of the amplitude $\phi(x)$. It turns out that the amplitude of a path is given by the following function:

$$\phi[(x(t))] = \exp\left\{ \frac{i}{\hbar} S[b,a] \right\} \tag{16}$$

where of course:

$$S[b,a] = \int_{t_a}^{t_b} \mathcal{L}(x, \dot{x}, t) dt \tag{17}$$

We have arrived at our final expression for the path integral:

$$\boxed{K(b,a) = \int_a^b \exp\left\{\frac{i}{\hbar}S[b,a]\right\}\mathcal{D}x(t)} \tag{18}$$

Where we have absorbed the normalization constants into our integration measure $\mathcal{D}x$ defined as:

$$\mathcal{D}x = \lim_{N\to\infty} \frac{1}{A}\prod_i^N \frac{dx_i}{A} \tag{19}$$

And the deceptively simple integral sign actually contains an infinite amount of integrals carried out over all dx_i from $-\infty$ to $+\infty$.

2.4 What is the Kernel?

While this is all well and good, we still don't really understand the nature of what we have computed. What exactly is the Kernel and how is it related to the wavefunction? Let us address each of these concerns in turn.

Recall that it was vaguely stated in section 2.1 that the Kernel is the "amplitude for going from point a to point b". In precise terms, this means that the kernel measures the probability amplitude of the system ending up in state x_b at time t_b given that it starts from the definite position state x_a at time t_a. In mathematical terms what this means is that the Kernel is equivalent to the following inner product:

$$\begin{aligned}\langle x_b, t_b | x_a, t_a\rangle &= \langle x_b|\, e^{\frac{i}{\hbar}Ht_b}e^{-\frac{i}{\hbar}Ht_a}\,|x_a\rangle \\ &= \langle x_b|\, e^{\frac{i}{\hbar}H(t_b-t_a)}\,|x_a\rangle\end{aligned} \tag{20}$$

Regarding the question of how the Kernel is related to the wavefunction: the Kernel tells us a little more about the system than the wavefunction does. While the wavefunction contains only information about any subsequent probability amplitudes, the kernel also provides us information about the state that the system started from. However, for the purposes of calculations concerning subsequent times, the wavefunction and the Kernel are equivalent. In other words, one can say that:

$$K(b,a) = K(x_b,t_b;x_a,t_a) = \psi(x_b,t_b) \tag{21}$$

2.5 Rule for events occurring in succession

Let us consider a situation where the path of a particle has been split into two parts. In other words, its journey consists of two stretches. In the first stretch the particle goes

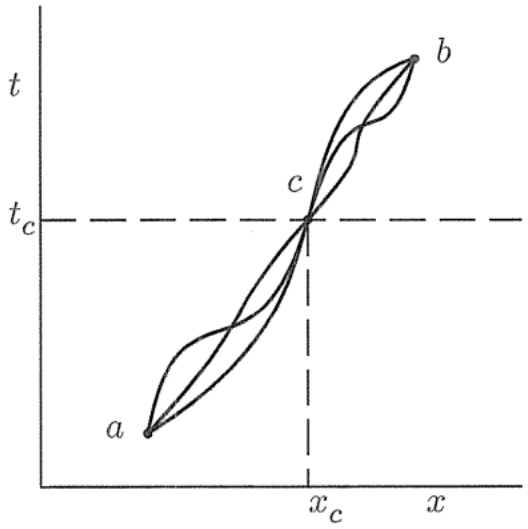

Figure 6: Events occurring in succession

from (x_a,t_a) to some intermediate point (x_c,t_c) and then finally carries on to the final point (x_b,t_b). How does one evaluate the total Kernel?

For one we know that the action in going from a to b $S[b,a]$ can be split as follows:

$$S[b,a] = S[b,c] + S[c,a] \tag{22}$$

This follows naturally from the definition of the action as an integral of the Lagrangian. All we have done is split the integral into two parts:

$$S[b,a] = \int_{t_a}^{t_b} \mathcal{L}(x,\dot{x},t)dt = \int_{t_c}^{t_b} \mathcal{L}(x,\dot{x},t)dt + \int_{t_a}^{t_c} \mathcal{L}(x,\dot{x},t)dt \tag{23}$$

With this done, we can now rewrite the Kernel in (18) as:

$$\begin{aligned} K(b,a) &= \int_a^b \exp\left\{\frac{i}{\hbar}(S[b,c] + S[c,a])\right\}\mathcal{D}x \\ &= \int_a^b \exp\left\{\frac{i}{\hbar}S[b,c]\right\} \times \exp\left\{\frac{i}{\hbar}S[c,a]\right\}\mathcal{D}x \end{aligned} \tag{24}$$

Then we can imagine the path integral as being carried out from a to c and then from c to b i.e the integral becomes:

$$K(b,a) = \int_c^b \int_a^c \exp\left\{\frac{i}{\hbar}S[b,c]\right\} \times \exp\left\{\frac{i}{\hbar}S[c,a]\right\}\mathcal{D}x \tag{25}$$

There is, however, something important that should be pointed out here. It makes intuitive sense that if we are to calculate the amplitude in going from a to b, then we need to consider all of the possibilities for what x_c could be. In other words, if we are to join the paths $a \to c$ and $c \to b$ then it makes sense to integrate our result over all possible x_c values, since the particle could arrive at b after having taken *any* path. It is also apparent that if we carry out the inner path integral first, the factor $\exp\{\frac{i}{\hbar}S[b,c]\}$ is a constant and therefore can simply be pulled out of the inner integral. We can therefore rewrite (25) as:

$$K(b,a) = \int_{-\infty}^{+\infty} \int_c^b \exp\left\{\frac{i}{\hbar}S[b,c]\right\} K(c,a)\mathcal{D}x\, dx_c \tag{26}$$

We can now carry out the inner integration again to find the second kernel and our final expression becomes:

$$\boxed{K(b,a) = \int_{-\infty}^{+\infty} K(b,c)\, K(c,a)\, dx_c} \tag{27}$$

This gives a simple rule for combining propagators. If you have two propagators, one for going from point a $\to$ c and another for going from $c \to b$ then all you need to do is put them together, and integrate over the all spatial values at the point connecting them. It is for this very reason that the kernel is sometimes also called the "propagator" in the context of the schrodinger equation. If we interpret the second kernel $K(c,a)$ as a wavefunction (that is we ignore the initial state of the system), then the equation can be rewritten as follows:

$$\boxed{K(b,a) = \int_{-\infty}^{+\infty} K(x_b,t_b;x_c,t_c)\, \psi(x_c,t_c)\, dx_c} \tag{28}$$

If you know the wavefunction at some time t_c, you can use $K(c,a)$ to *propagate* it forward in time.

2.6 The classical limit

This is a good point to discuss the classical limit. That is, how does the path integral of quantum mechanics *select* in the classical limit, just a single path: the path of least action. Consider for a second, the path integral as given in (18):

$$K(b,a) = \int_a^b \exp\left\{\frac{i}{\hbar}S[b,a]\right\}\mathcal{D}x(t) \tag{29}$$

Notice that there is a phase $e^{\frac{i}{\hbar}S[x(t)]}$ associated with each path. In the context of the spatial and temporal scales involved in classical mechanic (usually of the order of meters and seconds respectively), $\hbar$ is a miniscule number. Therefore, for a very small change in the action results in an extremely large change in the phase. Recall that the classical path extremizes the action. Another way of stating this is that for first order deviations from the classical path, the action does not change. As a result of this, only the paths in the immediate vicinity of the classical path contribute in phase. As we move farther and farther away from this extremum of the action, small changes in S cause huge changes in the phase. Thus there are on average just as many phases pointing in a direction as there are pointing in a direction opposite to it. Therefore contributions from paths significantly far away from the classical path (on the classical scale) cancel out, and we are only left with the classical path.

3 Monte Carlo Techniques

He deals the cards to find the answer,
the sacred geometry of chance,
the hidden law of a probable outcome,
the numbers lead a dance
- Sting

3.1 Monte Carlo Integration

Let us consider a situation where we have a continuous probability distribution $p(\theta)$. One interesting thing we might want to do with this probability distribution is calculate some of its quantities of interest. For example we might want to calculate the mean, standard deviation, variance so on and so forth. The way that you usually go about this is by computing an integral of the form:

$$G = \int g(\theta)\, p(\theta)\, d\theta \tag{30}$$

Where the integral is over all values that your parameter θ can take. Here G and $g(\theta)$ change based on the quantity you want to calculate. For example if you want to calculate the expectation value $E(\theta)$ or the mean, you replace $g(\theta)$ by θ and the equation becomes:

$$E(\theta) = \int \theta\, p(\theta)\, d\theta \tag{31}$$

In a similar manner if you want to calculate the variance $\mathrm{Var}(\theta)$, you might use, instead, the integral:

$$\mathrm{Var}(\theta) = \int (E(\theta) - \theta)^2\, p(\theta)\, d\theta \tag{32}$$

This is all well and good and in an ideal world, you're done. But suppose that for some reason, you cannot solve the integral in this way. For example, it might be that your parameter space Θ consists of vectors $\vec{\theta}$ with a large number of entries and it is computationally cumbersome to do the integral by hand. Or it might be that the integral just cannot be solved analytically. In any case, you can do the next best thing: you can approximate the integral using a process known as **Monte Carlo Integration**.

In a typical Monte Carlo Integration one simulates draws from the probability distribution $p(\theta)$ by randomly generating numbers distributed according to that distribution. Then, for a very large number of such draws M one can approximate the integral using the following sum:

$$G_{\text{approx}} = \frac{1}{M}\sum_{i}^{M} g(\theta^{(i)}) \tag{33}$$

Where $\theta^{(i)}$ corresponds to the ith draw in the Monte Carlo simulation. As the number of draws M gets larger, this approximation approx gets closer and closer to the actual value G. In fact:

$$G = \lim_{M\to\infty} \frac{1}{M}\sum_{i}^{M} g(\theta^{(i)}) \tag{34}$$

That is, in the limit that the number of draws approaches infinity, the approximation G_{approx} approaches the actual value G. We know this is true because of the Strong Law Of Large Numbers.

Strong Law Of Large Numbers:

The Strong Law Of Large Numbers (SLLN) states that if we have a sequence of M independent and identically distributed (we will henceforth refer to this condition by the abbreviation IID) random variables $X_1, X_2...X_M$, with mean $E(X)$, then:

$$\lim_{M\to\infty} \frac{X_1 + X_2 + ... + X_M}{M} = E(X) \tag{35}$$

In fact, the SLLN makes an even stronger statement, since any function of an IID random variable is also an IID random variable, the above relation holds true not just for the mean but for any $g(\theta)$ such as the variance, standard deviation etc.

Great, but what if we have a probability distribution but can't sample **independent** draws from it?

This situation might seem contrived and highly unlikely, but is actually quite common. For example, let us say we have a function $f(\theta)$ and we want to make a probability distribution $p(\theta)$ out of it so that we can sample independent draws from it. The first step would be to normalize it,

$$p(\theta) = \frac{f(\theta)}{N} \tag{36}$$

But what if we don't know the normalization constant N? It might be that the distribution requires us to compute some convoluted multidimensional integral. Or maybe calculating the normalization constant is computationally expensive. Enter the Markov Chain Monte Carlo method. To understand it, let us first briefly discuss Markov Chains.

3.2 Markov Chains

Since a markov chain is an example of a stochastic process, we need to define what a stochastic process is.

Stochastic process:

A stochastic process is a sequence of random values defined on a state space Θ. Since a stochastic process is a sequence, each random value or state is indexed. This index can be interpreted as a time step or an iteration. At each step or iteration the system jumps from some state θ in the state space, to some other state θ' using some random (non-deterministic) rule.

A markov chain is a special type of stochastic process:

Markov Chain:

A Markov Chain is a stochastic process whereby each state in the random sequence only depends on the state immediately preceding it. That is to say that if at time step t the system is in the state θ^t, then state θ^{t+1} only depends on the state θ^t and none of the preceding states such as θ^{t-1}, θ^{t-2} and so on. In other words, it is a stochastic process which follows the following condition:

$$P(\theta^t|\theta^{t-1}, \theta^{t-2}, ..., \theta^0) = P(\theta^t|\theta^{t-1}) \tag{37}$$

This is sometimes referred to as the **Markov property**. Here $p(A|B)$ simply stands for the probability that A will happen given that we know B has happened. Equation (36) therefore says that barring the current state, all previous states of the system are irrelevant in determining what state the system will end up during the next iteration. Our system is an amnesiac, all that it remembers is the last state that it was in.

This is all well and good, but we still haven't really explained how we mathematically encapsulate the rule for jumping from one state to another. As it turns out, for the discrete case - that is, when the parameter space Θ consists of discrete points - the rule for jumping from one state to another is given by a matrix. In the continuous case, we instead have an operator. In either case, this mathematical object is called the transition Kernel. Let us consider the discrete case.

3.2.1 The Discrete Case

In the discrete case, the transition Kernel is a matrix, and the *distribution* of the system at some time/iteration is given by a vector Π. Since each iteration causes the system to jump to a state **non-deterministically**, all we can talk about is the *probability* of ending up in a particular state θ_i. This combination of probabilities for ending up in each state is precisely the distribution given by Π.

In general if there are N states in the state space Θ then the distribution of the system is given by an N x 1 vector. Each entry in this N x 1 vector is populated by the

probability of the system ending up in one of the N states. Thus if we consider a situation where there are a total of 3 states, then at some time step t, the vector Π looks like:

$$\Pi = \begin{bmatrix} P(\theta_1) \\ P(\theta_2) \\ P(\theta_3) \end{bmatrix} \tag{38}$$

The transition Kernel K, on the other hand, is an N x N matrix populated by a set of conditional probabilities. In other words, , the probability of ending up in state j given that the system is in state i. This ends up looking as follows:

$$K = \begin{bmatrix} P(\theta_1|\theta_1) & P(\theta_1|\theta_2) & P(\theta_1|\theta_3) \\ P(\theta_2|\theta_1) & P(\theta_2|\theta_2) & P(\theta_2|\theta_3) \\ P(\theta_3|\theta_1) & P(\theta_3|\theta_2) & P(\theta_3|\theta_3) \end{bmatrix} \tag{39}$$

If we start from some distribution $\Pi_{(0)}$, we can obtain $\Pi_{(1)}$ by applying the Kernel on the vector $\Pi_{(0)}$ once. In fact for all successive distributions:

$$\begin{aligned} \Pi_{(1)} &= K \times \Pi_{(0)} \\ \Pi_{(2)} &= K \times \Pi_{(1)} \\ &... \\ \Pi_{(n)} &= K \times \Pi_{(n-1)} \\ \Pi_{(n)} &= K^n \times \Pi_{(0)} \end{aligned} \tag{40}$$

Where the $\times$ represents a matrix product. That is, starting from the initial distribution $\Pi_{(0)}$, we can obtain the distribution at iteration n by applying the nth power of the Transition Kernel K onto the vector $\Pi_{(0)}$. It is intuitively clear from the form of the matrix in equation (38) why this should be the case. If we take the matrix product of the transition kernel and the distribution vector once, we obtain:

$$\begin{bmatrix} P(\theta_1|\theta_1) & P(\theta_1|\theta_2) & P(\theta_1|\theta_3) \\ P(\theta_2|\theta_1) & P(\theta_2|\theta_2) & P(\theta_2|\theta_3) \\ P(\theta_3|\theta_1) & P(\theta_3|\theta_2) & P(\theta_3|\theta_3) \end{bmatrix} \begin{bmatrix} P(\theta_1) \\ P(\theta_2) \\ P(\theta_3) \end{bmatrix} = \begin{bmatrix} P(\theta_1|\theta_1)P(\theta_1) + P(\theta_1|\theta_2)P(\theta_2) + P(\theta_1|\theta_3)P(\theta_3) \\ P(\theta_2|\theta_1)P(\theta_1) + P(\theta_2|\theta_2)P(\theta_2) + P(\theta_2|\theta_3)P(\theta_3) \\ P(\theta_3|\theta_1)P(\theta_1) + P(\theta_3|\theta_2)P(\theta_2) + P(\theta_3|\theta_3)P(\theta_3) \end{bmatrix}$$

In calculating the probability for ending up in each state θ_i the matrix product takes into account all of the contributions, that is the conditional probabilities $P(\theta_i|\theta_j)$ for all states θ_j. It turns out that most Markov Chains converge to an **equilibrium** or **stationary distribution** Π_{stat}. Such a distribution is characterized by the condition:

$$\Pi_{\text{stat}} = K \times \Pi_{\text{stat}} \tag{41}$$

That is, any subsequent application of the transition Kernel on the stationary distribution has no bearing on the vector. Once the distribution has achieved equilibrium, it stays in equilibrium. Additionally, it doesn't matter what distribution you start the Markov Chain from, eventually the Markov chain will cause the distribution to achieve equilibrium.

3.3 Markov Chain Monte Carlo

Now that we have introduced the concept of Markov Chains, we can return to our original problem: how does one calculate quantities of interest if the draws/samples are not completely independent.

One way to go about it is to set up our Markov Chain intelligently. That is, we can make sure that the Transition Kernel is such that given an initial distribution, the chain causes it to eventually converge to our original distribution $p(\theta)$. How one goes about doing this will be discussed shortly. For now we have another problem at hand. Since our draws are no longer independent, we can no longer use the crutches afforded to us by the Strong Law Of Large Numbers: the law only holds for IID random variables, and our variables are no longer independent - each draw from a Markov Chain depends on the previous draw. Luckily we have an even stronger condition, known as the Ergodic Theorem.

Ergodic Theorem:

The Ergodic theorem states that for a sequence of M draws, $\theta_1, \theta_2, ..., \theta_M$ from a Markov Chain that is *irreducible*, *aperiodic* and *positive recurrent*:

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i}^{M} g(\theta^{(i)}) = \int g(\theta)\, \Pi_{\text{stat}}(\theta)\, d\theta \tag{42}$$

Where Π_{stat} is the stationary distribution of the Markov Chain under consideration.

Evidently the minor dependence of consecutive entries is unimportant, once convergence has occurred the Markov Chain starts *behaving* in some sense, as if it were a probability distribution. Before we proceed any further, however, we need to define what *irreducible*, *aperiodic* and *positive recurrent* means. Let us define each of these in turn:

Aperiodicity

A Markov chain is said to be aperiodic if it doesn't keep repeating a cycle of values or states in a finite amount of time. This definition can be made more precise with the help of a counter-example. Consider the Markov Chain shown in Figure 7.

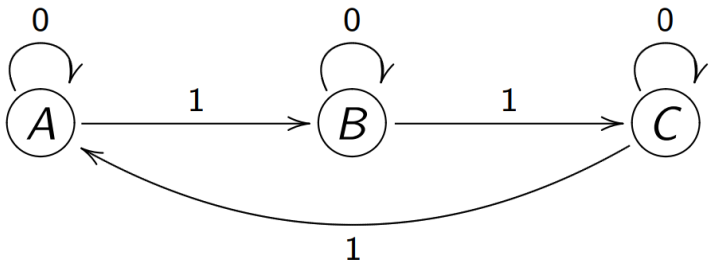

Figure 7: An example of a periodic Markov Chain

The numbers above the arrows represent the transition probabilities for going from one state to another. Evidently such a Markov chain keeps repeating the same set of states with a period of 3. This chain does not classify as aperiodic and therefore does not satisfy the conditions of the Ergodic theorem. If a Markov Chain doesn't trivially keep repeating the same sequence of states over and over again then it is aperiodic.

Irreducibility

A Markov chain is said to be irreducible if it is possible to go from any state to any other state in the state space. For example consider the Markov Chain in Figure 8:

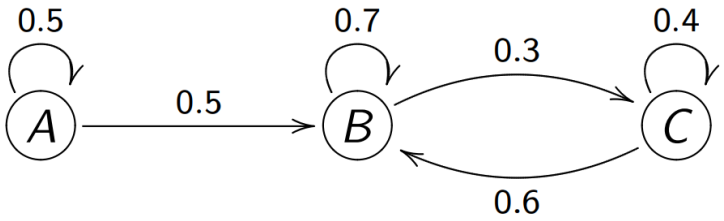

Figure 8: An example of a reducible Markov Chain

In this Markov Chain, the system cannot go to state A from either state B or state C. This is an example of a reducible chain (or a non-irreducible chain but that's a mouthful). As long as the Markov chain has some path leading from every state to every other state it will be considered irreducible.

Positive Recurrence

A positive recurrent Markov chain is one where if the system starts in state i, then the system will definitely return to state i in a finite amount of time. This is in contrast to *null recurrence* where the system returns to the state i in an infinite amount of time.

Thus if a Markov chain sastifies all of these criteria, then it satisfies the Ergodic Theorem and is said to be Ergodic. Such a Markov chain can be chosen for the purposes of carrying out Markov Chain Monte Carlo.

3.4 Choosing the Kernel Intelligently - The Metropolis Hastings Algorithm

We have built up the entire background necessary to perform Markov Chain Monte Carlo. But the most important part of the entire process still remains to be addressed: How does one go about choosing a Kernel so that the Markov chain converges to the required probability distribution?

We do this through a sampling algorithm first devised by Nicholas Metropolis and others, and then generalized by W.K. Hastings, called the Metropolis-Hastings algorithm. The general idea of the algorithm is quite simple. We start with an initial draw (or an initial Markov state) $\theta^{(t)}$. In order to get the next draw, we define a "Jumping distribution" $J(\theta^*|\theta^{(t)})$ which generates a new value (or state) θ^* given the current state $\theta^{(t)}$. We then compute a ratio r given by:

$$r = \frac{p(\theta^*)}{p(\theta^{(t)})} \frac{J(\theta^{(t)}|\theta^*)}{J(\theta^*|\theta^{(t)})} \tag{43}$$

If this ratio is equal to or greater than 1, we choose the state θ^* as our next draw $\theta^{(t+1)}$. On the other hand, if the ratio is less than 1, we generate a uniform random between 0 and 1. If this random number is smaller than r, we once again choose θ^* as our next draw. If the random number turns out to be greater than r, we keep $\theta^{(t)}$ as our next draw.

The essential idea behind the ratio r, still remains to be explained. The first term in this ratio was devised by Metropolis himself, the second term is the generalization provided by Hastings. The first term in the ratio calculates whether the probability of our distribution $p(\theta)$, generating the value θ^* is less than, greater than or equal to the probability of generating the current value $\theta^{(t)}$. Ignoring the second term, the first term selects all values that have a greater probability of being generated than the previous term. This in turn *moves* our Markov Chain to the more probable part of the distribution.

On the other hand, the prescription that a new value θ^* be selected with some non-zero probability even if the probability of it being generated is less than that of the current value θ^t, makes sure that the Markov chain does not get *stuck* so to speak, at the maximum of the distribution. In order to be able to parse the entire distribution, the Kernel should have *some leniency* towards less probable values.

The second term is less important, and is actually 1 if the Jumping Distribution J is *symmetric* (which mostly **is** the case). For Jumping distributions that are *asymmetric*:

$$J(\theta^{(t)}|\theta^*) \neq J(\theta^*|\theta^{(t)}) \tag{44}$$

the second term provides a sort of *compensation*, since the distribution is biased in one direction.

4 A Jewel In Time

What good is a jewel that ain't still precious?
- Frank Ocean

We are now in the position to actually start implementing some of the techniques that we have learnt. In this section, we will devise a scheme to approximate the path integrals for the quantum harmonic and anharmonic oscillators using path integration carried out using Markov Chain Monte Carlo.

4.1 Wick Rotation

Our first order of business will be to convert our time parameter (which takes up real values) into an imaginary number. In other words we consider a new parameter τ:

$$\tau = it \tag{45}$$

Geometrically, this amounts to a *rotation* in the argand plane of $\frac{\pi}{2}$ radians. This is known as a Wick Rotation. As we shall see , this process makes it possible for us to extract some quantities of interest from certain sums since the phases turn into exponential decay factors.

Consider for example, our Kernel from section 2. After wick rotation it becomes:

$$Z = \int_a^b \exp\Big\{ -\frac{S[b,a]}{\hbar} \Big\} \mathcal{D}x(t) \tag{46}$$

We will henceforth denote the kernel by Z. Note also that due to a change in our time parameter, the prescription for calculating the action also changes:

$$S = \int_0^T \left[\frac{1}{2} m \left[\frac{dx}{d\tau} \right]^2 + V(x) \right] \tag{47}$$

Where ofcourse τ is defined using eq (45). As discussed previously, this can be written in the bra-ket notation as the following transition amplitude:

$$Z = \langle x_b | \, e^{-\frac{1}{\hbar} H(t_b - t_a)} \, | x_a \rangle \tag{48}$$

$$= \langle x_b | \, e^{-\frac{1}{\hbar} HT} \, | x_a \rangle \tag{49}$$

where we have rewritten the time interval between t_b and t_a as T. Let us insert a completeness relation in the energy basis into this expression as follows:

$$Z = \sum_n \langle x_b | \, e^{-\frac{1}{\hbar} HT} \, |n\rangle \, \langle n|x_a\rangle \tag{50}$$

$$= \sum_n e^{-\frac{1}{\hbar} E_n T} \, \langle x_b|n\rangle \, \langle n|x_a\rangle \tag{51}$$

Notice how due to Wick rotation, the pre-factors $e^{-\frac{E_n T}{\hbar}}$ which were originally phases, have now turned into exponential decay factors. This is important because it allows us to extract the first term from the sum in the limit that $T \to \infty$. In other words:

$$\lim_{T\to\infty} Z = e^{-\frac{1}{\hbar} E_0 T} \, \langle x_b|0\rangle \, \langle 0|x_a\rangle \tag{52}$$

This happens because the exponential decay factors associated with all of the subsequent terms cause them to die down much more quickly compared to the first term. This allows us to approximate the ground state energy of the system.

4.2 The Jewel

We consider yet again a lattice of time steps t_i where the position of the particle at the time step t_i is given by the continuous variable x_i. As in section 2, we once again set the spacing between subsequent time steps to be ϵ. That is:

$$t_{i+1} - t_i = \epsilon \tag{53}$$

If there are N such lattice sites, we can imagine the path integral as being approximated by the following set of N integrals carried out in succession:

$$Z_{\text{approx}} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_{i=1}^{N-1} dx_i \exp\left(\frac{-1}{\hbar} S[x_i]\right) \tag{54}$$

Note that we always have:

$$T = N\epsilon \tag{55}$$

Ofcourse in the limit that $N \to \infty$ the approximate kernel $Z_{\text{approx}} \to Z$. Throughout all of this since the path consists of discrete time steps, we calculate the action using the following **sum** instead of carrying out an actual integral as in (47):

$$S = \sum_{i=1}^{N} a \left[\frac{1}{2} m \frac{(x_{i+1} - x_i)^2}{a^2} + V(x_i) \right] \tag{56}$$

Where we have defined:

$$a = i\epsilon \tag{57}$$

We now make the following claim: "The Kernel in equation (54) is mathematically equivalent to the partition function of a crystal lattice of N sites."

In this partition function, the Boltzmann factor is replaced by the amplitude for each path:

$$e^{-\frac{E}{k_B T}} \rightarrow e^{-\frac{S}{\hbar}} \tag{58}$$

On each lattice site there is a continuous variable denoted by x_i, which can take up values from $-\infty$ to $+\infty$. The action couples the nearest neighbours x_i and x_{i+1} and the N integrals:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_{i=1}^{N-1} dx_i \tag{59}$$

simply represent a sum over all possible lattice configurations given the range of possible values of x_i. The lattice points, however, are not spatial points, but rather *temporal* points. What we have constructed is a "Jewel in time".

4.3 Quantities of interest

4.3.1 Energy difference between E_0 and E_1

We can now briefly discuss some of the quantities that we can calculate for our system as it stands. Recall that the expectation value for an operator $\hat{A}$ in statistical mechanics is calculated using the expression:

$$\langle \hat{A} \rangle = \frac{\mathrm{Tr}(e^{-\frac{HT}{\hbar}} \hat{A})}{\mathrm{Tr}(e^{-\frac{HT}{\hbar}})} \tag{60}$$

Ofcourse one can take the trace in the energy basis and the result is the following sum:

$$\langle \hat{A} \rangle = \frac{\sum_n \langle n| \, e^{-\frac{HT}{\hbar}} \hat{A} \, |n\rangle}{\sum_n \langle n| \, e^{-\frac{HT}{\hbar}} \, |n\rangle}$$

$$= \frac{\sum_n e^{-\frac{E_n T}{\hbar}} \langle n| \, \hat{A} \, |n\rangle}{\sum_n e^{-\frac{E_n T}{\hbar}}} \tag{61}$$

While this isn't exactly what we need it gives us a blueprint for the actual expression. Notice that we have a scalar quantity $\langle n| \, \hat{A} \, |n\rangle$ times the Boltzmann factor being summed over all of the possible *configurations* of the system. This is important, because in the context of our path integral lattice, the configurations are actually multiple integrals carried out over all of space. We have already established that our sum is given by (59). This now lets us set up an expression for calculating the expectation value for our system of lattice configurations:

$$\langle \hat{A} \rangle = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_{i=1}^{N-1} \, dx_i \exp\left(\frac{-1}{\hbar} S[x_i]\right) A(x_1, x_2, ..., x_N)}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} \prod_{i=1}^{N-1} \, dx_i \exp\left(\frac{-1}{\hbar} S[x_i]\right)} \tag{62}$$

Where now the A inside the integral in (62) is not an operator but a scalar function of the lattice configuration $(x_1, x_2, ..., x_N)$. This makes total sense: we sum over the contributions from A for all possible lattice configurations, then divide by a suitable normalizing factor to get our quantity of interest.

Another quantity of interest are the connected two-point correlation functions denoted by $\Gamma_c^{(2)}$ and defined as:

$$\Gamma_c^{(2)} = \langle x(\tau_1) x(\tau_2) \rangle - \langle x(\tau_1) \rangle \langle x(\tau_2) \rangle \tag{63}$$

Where each of the terms can be calculated exactly as prescribed in equation (62). The first term in this expression $\langle x(\tau_1) x(\tau_2) \rangle$ is called a correlation function, and is of interest in and of itself.

Yet another quantity of interest is the energy gap between the first and the second excited states. This can also be calculated in the limit that $T \to \infty$. Consider equation (61) in this limit. Just as in equation (52), only the first terms in the sums survive due to the decay factors and we are left with:

$$\langle \hat{A} \rangle = \frac{\cancel{e^{-\frac{E_0 T}{\hbar}}} \langle 0| \, \hat{A} \, |0\rangle}{\cancel{e^{-\frac{E_0 T}{\hbar}}}}$$

$$= \langle 0| \, \hat{A} \, |0\rangle \quad (\lim T \to \infty) \tag{64}$$

Then using this equation for our two-point correlation function (setting $\tau_1 = 0$ and $\tau_2 = \tau$) in (63) we get:

$$\lim_{T\to\infty} \Gamma_c^{(2)} = \langle 0| \, x(0)x(\tau) \, |0\rangle - |\, \langle 0| \, x \, |0\rangle \, |^2 \tag{65}$$

Putting a completeness relation in the energy basis in-between $x(\tau)$ and $|0\rangle$, then expanding $x(\tau)$ in the Heisenberg picture we get:

$$\lim_{T\to\infty} \Gamma_c^{(2)} = \sum_{n\neq 0} e^{\frac{-1}{\hbar}(E_n - E_0)\tau} |\, \langle 0| \, x \, |n\rangle \, |^2 \tag{66}$$

Then, if we consider a situation where we have $\tau' > \tau$ and then take the limit that $\tau \to \infty$, using the exponential decay argument again we can say that:

$$\frac{\Gamma_c^{(2)}(\tau')}{\Gamma_c^{(2)}(\tau)} = e^{-(E_1 - E_0)(\tau' - \tau)} \tag{67}$$

By taking the logarithm of this equation, we can find the energy gap between the ground state and the first excited state (given that we know the time interval between τ and τ').

4.3.2 Ground State Energy

We can also calculate the ground state energy, the procedure is as follows. We know that the ground state energy is the eigenvalue associated with one of the Eigenstates of the Hamiltonian, so we can write:

$$\begin{aligned} E_0 &= \langle 0| \, H \, |0\rangle \\ &= \lim_{T\to\infty} \left[\frac{\mathrm{Tr}(e^{\frac{-HT}{\hbar}} H)}{\mathrm{Tr}(e^{\frac{-HT}{\hbar}})} \right] \\ &= \lim_{T\to\infty} \frac{-1}{T} \frac{\partial}{\partial(\hbar^{-1})} \ln Z \end{aligned} \tag{68}$$

Where we have used our definition of expectation values in the second equality, and the third equality is just a tidying up of the expression using the fact that our 'partition function' is defined as:

$$\int_{-\infty}^{+\infty} \langle x| \, e^{\frac{-HT}{\hbar}} \, |x\rangle \, dx = \mathrm{Tr} \; e^{\frac{-HT}{\hbar}} \tag{69}$$

and then we can use the expression for Z from eq (54) to rewrite (68) as follows:

$$E_0 = \lim_{T\to\infty} \frac{-1}{T} \frac{\partial}{\partial(\hbar^{-1})} \ln \int e^{\frac{-1}{\hbar}S[x]}[dx] \tag{70}$$

We can also calculate the ground state energy directly using the virial theorem. According to the virial theorem, the expectation value of kinetic energy $< T >$ is related to the potential energy as follows:

$$< T >= \frac{1}{2} < x.V' > \tag{71}$$

Which we can then substitute this into the expression for the expectation value of the total energy to get:

$$E_0 = \lim_{T\to\infty} \left[\frac{\int e^{\frac{-1}{\hbar}S[x]}(\frac{1}{2}xV' + V)[dx]}{\int e^{\frac{-1}{\hbar}S[x]}[dx]} \right] \tag{72}$$

4.3.3 Ground State Wavefunction

Arguably the most important quantity of interest that can be calculated through the techniques expounded upon in the preceding sections is the ground state wavefunction corresponding to a given potential.

The procedure is fairly straightforward, once the Markov Chain has 'thermalized' (converged to the required probability distribution), the paths being sampled are the ones that the particle is most likely to take (i.e deviations around the classical path). By keeping track of the positions that the particle is more likely to take, we can in essence 'reconstruct' the wavefunction of the particle. In order to keep track of the position of the particle, however, we need to split the real line into bins of finite size (due to the finite precision of our computer). Then, every time the Markov Chain generates a new sample path, we can decide which bin each element of our path array falls in. As the number of paths approaches infinity (or a sufficiently large number in the case of computer simulations), this distribution of binned positions approaches the wavefunction (mod squared) of the particle, which can then be appropriately normalized.

4.3.4 Correlation functions

Another quantity of interest that can be calculated through our MCMC simulations are correlation functions, specifically correlation functions between positions at two different times. To achieve this we simply compute the product of positions at two separate time points τ and $\tau + \delta\tau$ for each sample and then average over all samples.

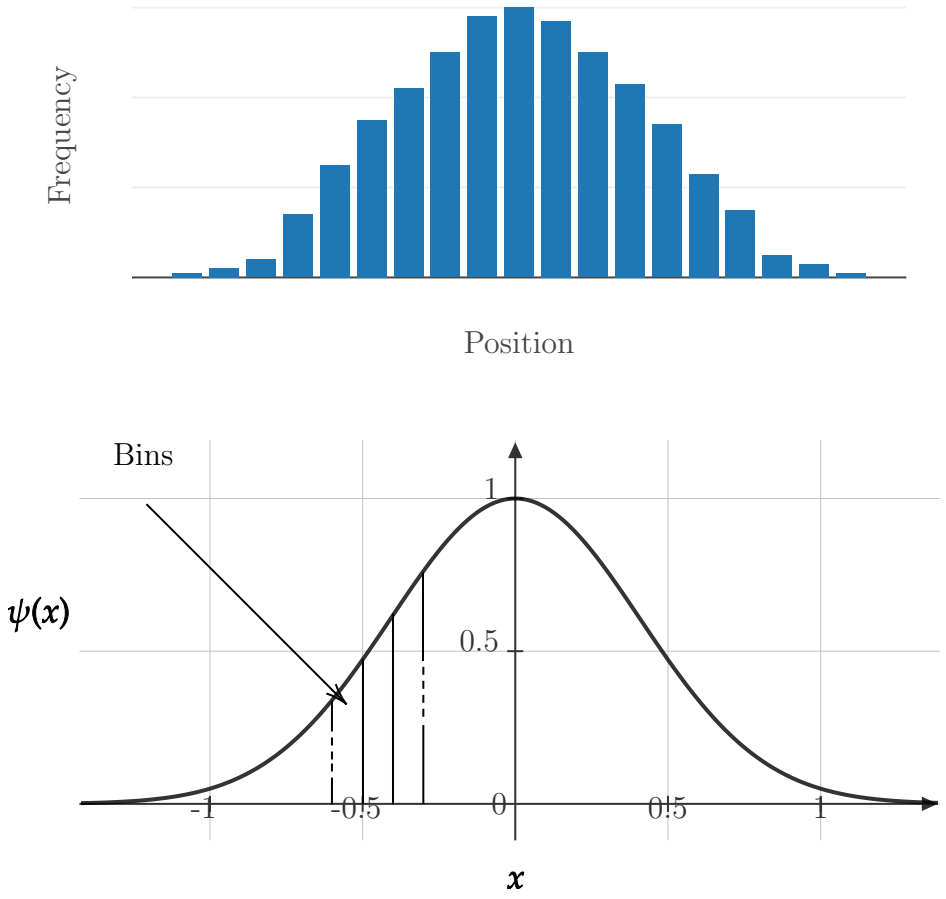

Figure 9: A hypothetical distribution of positions obtained by sampling from a thermalized markov chain (top). The distribution approaches an analytic wavefunction (mod squared) (bottom) as the number of sampled paths and the number of bins approaches infinity.

5 Simulations and Results - Quantum Mechanics

Anything can be a harmonic oscillator if you're bold enough
- Some physicist

All of the simulations were programmed in Python using only the NumPy library. The entirety of the code has been provided in the appendix. Most of this section is a recreation of the results from the paper by Creutz and Freedman.

5.1 Metropolis-Hastings applied

Before proceeding any further, we write down the set of rules which defines the Kernel for our Markov chain. The rules are based on the Metropolis-Hastings algorithm elaborated upon previously:

1. Start with a random array of positions generated within a given range, this is your path
2. For each element of the array, one after the other, generate a new position value within some chosen distance Δ of the original position
3. Compute the action (S_f) for the path array given this new position and compare it to the action (S_i) for the path array given the original position
4. If $S_f < S_i$ accept the new path
5. If $S_f > S_i$ accept the new path with a probability of $e^{\Delta S}$ (where $\Delta S = S_f - S_i$) else reject it

This entire process defines one Monte Carlo iteration. All of the results that follow were noted once the system had achieved a suitable degree of thermalization (that is, the action did not change by a considerable degree between samples).

5.2 Harmonic Oscillator

The harmonic oscillator serves as a good starting point and proof of concept for the application of MCMC techniques. This is because the system can be solved analytically, both directly through the Schrodinger equation and also through the path integral method. The values of interest calculated through simulations can therefore be compared to analytical results to evaluate the accuracy of the numerical approximations. Recall that the harmonic oscillator potential is given by,

$$V(x) = \frac{1}{2}\mu^2 x^2 \tag{73}$$

where for our simulations we chose $\mu = 1$.

5.2.1 Wavefunction

We simulated paths with $N = 1000$ lattice points, with a lattice spacing a of 0.1. The number of Monte Carlo samples generated (N_E) was 100,000. The number of 'hits' $\bar{n}$ for each lattice site was chosen to be 10. This meant that each lattice site was changed to a new random site a distance $\Delta = 0.4$ around the original site 10 times before the new path was either accepted or rejected. In all simulations we assumed $\hbar = 1$. Since the real line had to be binned for the wavefunction, we chose a bin size of 0.04 on either side (so 0.08 in total). The wavefunction obtained is shown in Figure 10, with the result from theory superimposed as a dotted line.

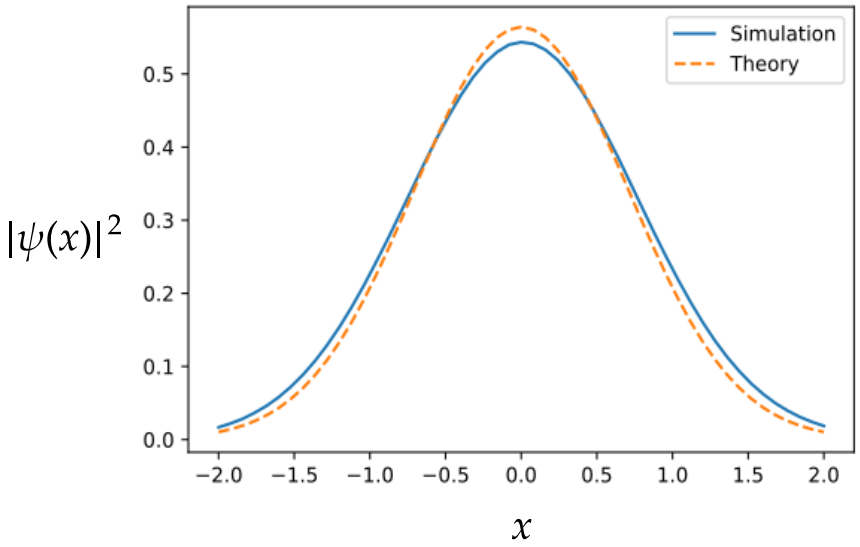

Figure 10: The wavefunction for the harmonic oscillator obtained by binning 100,000 monte carlo samples of paths with 1000 lattice points (blue). The dotted line (orange) is the wavefunction obtained from theory.

5.2.2 Correlation function

Using exactly the same parameters as mentioned in the previous section, we calculated the correlations between the position at time zero and subsequent time points, the results are shown in Figure 11. The results are consistent with the results from the paper by Creutz and Freedman.

5.2.3 Expectation of x^2

The expectation value of x^2 was also calculated through MCMC, and the value obtained was 0.485, which was close to the value of 0.5 predicted by theory.

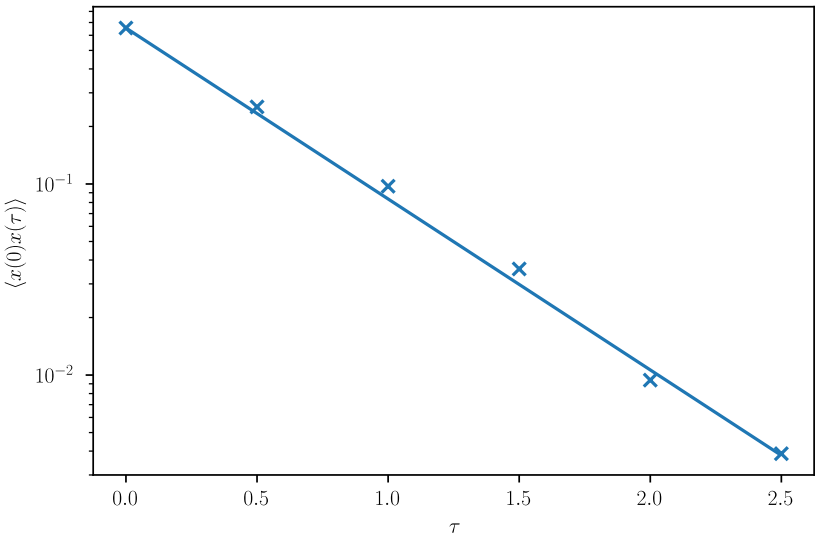

Figure 11: Semi-log plot of the position correlation function obtained for the harmonic oscillator

5.3 Anharmonic Oscillator

The Anharmonic oscillator is a natural next step when testing MCMC techniques since it is not solvable analytically, but has a potential function which bears some resemblance to the quadratic potential of the harmonic oscillator. The Anharmonic oscillator potential is given by:

$$V(x) = \frac{1}{2}\mu^2 x^2 + \lambda x^4 \tag{74}$$

where for our simulations we chose $\mu = 1$ and $\lambda = 1$.

5.3.1 Wavefunction

The parameters for this simulation were as follows: $N = 200$ lattice sites with a lattice spacing of $a = 0.25$. The number of monte carlo samples generated was once again $N_E = 100000$ with an $\bar{n}$ (hits per site) of 10. New sample positions were generated around the original positions in a range of $\Delta = 0.4$ on either side. The bin size was chosen to be 0.04. The wavefunction obtained was qualitatively identical to the one obtained by Creutz and Freedman and is shown in Figure 12.

It was noted that the simulations had to be run for a significantly longer period of time before the paths explored the entire space. This was because of the presence of two minima (as is evident from the two peaks in the wavefunction) with a region of lower probability (and thus higher potential) between them.

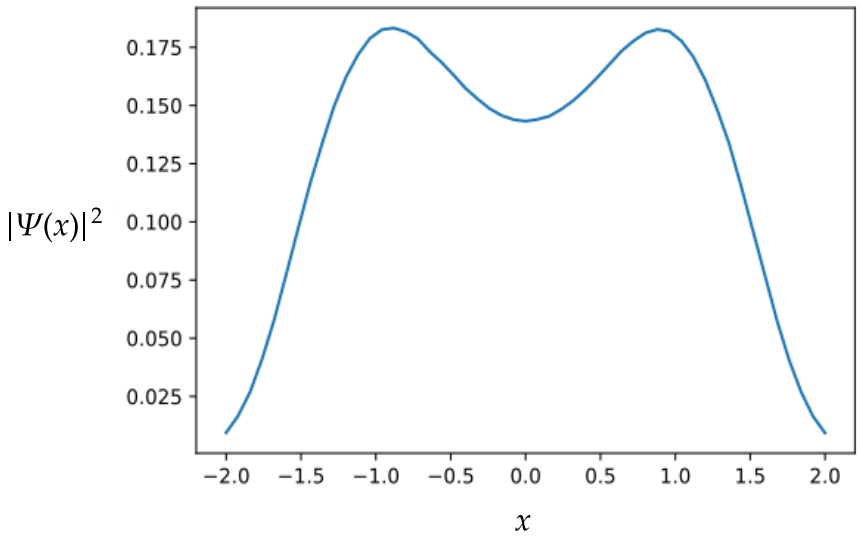

Figure 12: Wavefunction obtained for the anharmonic oscillator. $N = 200$, $a = 0.25$ and $N_E = 100,000$

5.3.2 Correlation function

A calculation of the correlation function was repeated for the anharmonic oscillator with the same parameters as mentioned in the previous section. A semi-log plot of the results obtained is shown in Figure 13. This result, too, was in agreement with the results in the literature.

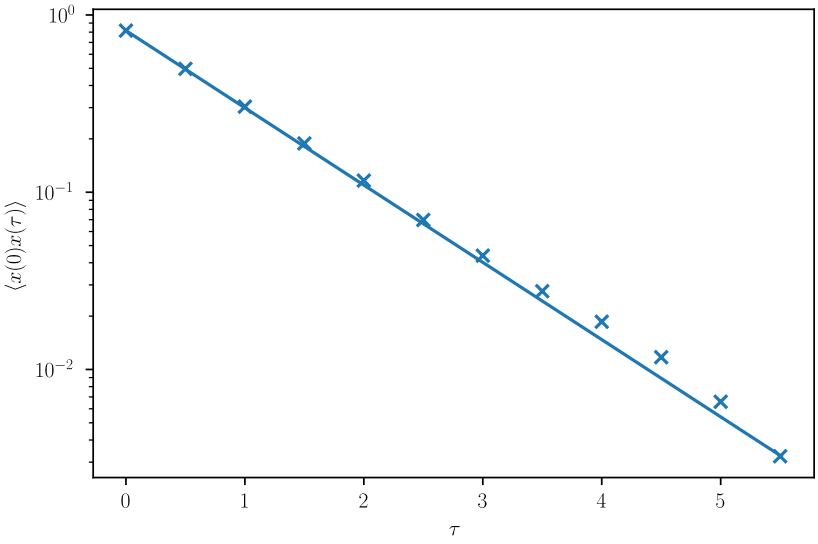

Figure 13: Semi-log plot of the position correlation function obtained for the anharmonic oscillator

6 Cosmology

Isn't lightning beautiful precisely because it is momentary? Like you. Like me. Like this collection of binary codes that hold my thoughts and the duration of our collective wonder against the unfathomable aeons of cosmological evolution.

We now move on towards the most important part of this thesis. Namely, that which pertains to the application of PIMC (Path Integral Monte Carlo) techniques to approximate path integrals in cosmology/quantum gravity.

It is widely known that the primary goal of quantum gravity is to merge two of Physics' most successful theories. These are, of course, general relativity and quantum mechanics, which explain the universe in the domains of the very large (planetary systems, cosmological evolution, black hole dynamics so on and so forth) and the very small (electrons, photons, quarks so on and so forth) respectively.

While most field theories give in very easily to the tools of quantization, it has proven exceedingly difficult to quantize general relativity. One of the problems that makes it notoriously difficult to quantize gravity is the 'Problem Of Time'. This is because in the Einstein field equations:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{75}$$

time does not appear as an external parameter but is instead just as much of a dynamical variable as, say, the matter/energy distribution. This can be validated from simply noting that g_{00} appears on the left hand side in this equation and depends on one of the components of the stress-energy tensor. Many different schemes have been proposed to quantize general relativity over the past century. These include:

- String Theory
- Loop Quantum Gravity
- Causal Dynamical Triangulations
- Asymptotic Safety

These can be further classified into groups based on the overarching approach that each scheme takes. Loop quantum gravity falls under the broad category of approaches labeled 'Canonical Quantum Gravity'.The name comes from the approach known as 'Canonical Quantization' which is ubiquitous in Physics due to the ease with which it can be used to convert a classical theory into a quantum theory. Since this approach is relevant to our present discussion, we discuss it in detail in the following section.

6.1 Canonical Quantization

Canonical quantization proceeds in the following manner. Suppose you have the Lagrangian formulation of a classical theory. This implies the existence of a phase space with generalized coordinates and their respective momentum conjugates represented by q and p respectively. These are related as follows:

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \tag{76}$$

Where $\mathcal{L}$ is the Lagrangian. Assume also that the Hamiltonian of the system is represented by $H(p, q)$. The dynamics of any function $f(p, q)$, then, is dictated by the Poisson Bracket of the Hamiltonian with said function:

$$\frac{d}{dt} f(p, q) = \{f, H\} \tag{77}$$

where the Poisson bracket $\{f, H\}$ is defined as:

$$\{f, H\} = \frac{\partial f}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial H}{\partial q} \tag{78}$$

and it has been assumed in equation (76) that f does not have any explicit time dependence (that is an additional term of $\frac{\partial f}{\partial t}$ has been ignored).

The systematic way to quantize such a theory is as follows. Take the phase space from the classical theory and promote it to an appropriate Hilbert space and take the position and momentum variables and promote them to the operators $\hat{q}_i$ and $\hat{p}_i$ acting on vectors within this Hilbert space. As before if the Hamiltonian (now also an operator) of the system is assumed to be $\hat{H}$, then the time evolution of any operator $\hat{A}$ within this Hilbert space is dictated by the **commutator** of the operator with the Hamiltonian as follows:

$$\frac{d}{dt}\hat{A} = \frac{i}{\hbar}[\hat{A}, \hat{H}] \tag{79}$$

where we have once again ignored any explicit time dependence that $\hat{A}$ might have. The commutator is defined as follows:

$$[\hat{A}, \hat{H}] = \hat{A}\hat{H} - \hat{H}\hat{A} \tag{80}$$

For a Field theory the quantization process proceeds in more or less the same manner. In a classical field theory, position and momentum variables take the shape of fields $\phi(x,t)$ whose momentum conjugates $\pi(x,t)$ are defined as follows:

$$\pi = \frac{\partial \mathcal{L}}{\partial \dot{\phi}} \tag{81}$$

Where $\mathcal{L}$ in this case is the Lagrangian density. The time evolution of the fields is once again dictated by the Poisson bracket of the field with the Hamiltonian as follows:

$$\frac{\partial \phi(x,t)}{\partial t} = \{\phi(x,t), H\} \tag{82}$$

where H is the Hamiltonian and is related to the Hamiltonian density $\mathcal{H}$ of the system as follows:

$$H = \int d^3y \, \mathcal{H} \tag{83}$$

that is, the Hamiltonian H is just the integral of the Hamiltonian density $\mathcal{H}$ over all space. Thus the Poisson bracket in (82) in terms of the hamiltonian density is given by:

$$\{\phi(x,t), H\} = \int d^3y \, \{\phi(x,t), \mathcal{H}\} \tag{84}$$

To quantize this kind of classical field theory, the fields $\phi(x,t)$ and their conjugate momenta $\pi(x,t)$ are promoted to operator fields $\hat{\phi}(x,t)$ and $\hat{\pi}(x,t)$ and their time evolution is dictated by the commutator with the hamiltonian operator.

$$\frac{\partial \hat{\phi}(x,t)}{\partial t} = [\hat{\phi}(x,t), \hat{H}] \tag{85}$$

6.2 ADM Formalism

As is evident from equation (75), general relativity in its standard form is not an initial value formulation. As mentioned previously, this is precisely due to the fact that time, in general relativity, factors into the spacetime metric and is not an external parameter. One can, however, reframe general relativity in a manner that is more akin to an initial-value/Hamiltonian formulation and is therefore primed for the process of canonical quantization. This initial-value formulation is termed the Arnowitt-Deser-Misner

or ADM formalism of general relativity, named after the three physicists who made the most significant contributions towards its development.

In very rough terms, the ADM formalism involves 'slicing' or 'foliating' the entire spacetime manifold into non-overlapping space-like hypersurfaces each of which is associated with a certain value of the time parameter. This, in some sense, 'separates' the three dimensions of space and one dimension of time, which is why it is also sometimes referred to as the 3+1 formalism of GR.

As a result of the foliation, we are left with the following variables which allow us to reconstruct the full metric:

- The induced metric q_{ab} on the space-like hypersurfaces (since they are submanifolds)
- The conjugate momenta associated with the induced metric π^{ab}
- The shift vector N^a which encapsulates the change in co-ordinates from one hypersurface to the next
- The lapse function $\mathcal{N}$ which encapsulates the rate of change of proper time from one hypersurface to the next

In terms of these variables, the original metric can be reconstructed in the following manner:

$$ds^2 = -\mathcal{N}^2 dt^2 + q_{ab}\left(dx^a + N^a dt\right)\left(dx^b + N^b dt\right) \tag{86}$$

The action which needs to be extremized in this hamiltonian formulation is, in turn, given by (disregarding the cosmological constant),

$$S = \int dt\ d^3x \left(\pi^{ab}\dot{q}_{ab} - \mathcal{N}\mathcal{H} - N^a\mathcal{H}_a\right) \tag{87}$$

Where $\mathcal{H}$ is the Hamiltonian constraint and $\mathcal{H}_a$ is the momentum constraint. Now that we have the action we can finally apply our PIMC techniques to solve for the quantum analog to the theory. But we have yet to address one major issue previously mentioned, namely the problem of time.

6.3 Time Gauge Fixing

The problem of time is a major unresolved problem in the field of quantum gravity. The crux of the issue lies in the diffeomorphism invariance of general relativity, namely that any space and time re-parametrization leaves the theory unchanged. The time reparametrization part is manifest in the hamiltonian constraint, namely the requirement that:

$$\mathcal{H} = 0 \tag{88}$$

A very naive application of canonical quantization would require that such a Hamiltonian be promoted to an operator in an appropriate Hilbert space such that its action on any vector in such a space leads to zero, i.e:

$$\hat{\mathcal{H}} \left|\psi\right\rangle = 0 \tag{89}$$

This is known as the Wheeler-DeWitt equation, whose solutions we will attempt to numerically approximate. For now it should be noted that this is equation is in stark contrast to the time evolution equation of any standard quantum theory since according to the Schrodinger equation:

$$\hat{H} \left|\psi\right\rangle = i\frac{\partial}{\partial t} \left|\psi\right\rangle \tag{90}$$

where the derivative is with respect to time which is not a part of the Hilbert space but is instead an external parameter. Clearly any quantum theory of gravity would require a reinterpretation of time in order to be consistent with the hamiltonian constraint.

One way of circumventing the problem of time is by choosing a phase space variable to signify the passage of time. That is, by using the dynamical evolution of said variable as a clock. Let us call this phase space variable t and its conjugate momentum p_t. Then for a system with four degrees of freedom (t, p_t) and (x, p_x) we have the following action:

$$S = \int (p_t \dot{t} + p_x \dot{x} - NH) d\lambda \tag{91}$$

Where N is a lagrange multiplier and the dots indicate derivatives with respect to λ. If we now reparametrize $\lambda \rightarrow t$ and simultaneously enforce the Hamiltonian constraint, we are left with the following equation:

$$S = \int (p_x \dot{x} + p_t \dot{t}) dt \tag{92}$$

Comparing this with the action of a theory involving just two degrees of freedom (x, p_x),

$$S = \int (p_x \dot{x} - H) dt \tag{93}$$

Evidently our choice of time guage implies that:

$$H = -p_t \tag{94}$$

that is, the conjugate momentum to our chosen time variable acts as the (negative of) the hamiltonian for the remaining phase space variables. The only thing required of our chosen time variable is that it should have a non-trivial time evolution, in other words:

$$\{t, \mathcal{H}\} \neq 0 \tag{95}$$

It should be noted that this choice of a physical time gauge also circumvents the problem which could potentially arise upon quantization of the theory. The physical Hamiltonian belongs to the phase space in the classical case and so would be promoted to an operator in the quantum case and would therefore also be a part of the Hilbert space.

7 Simulations and Results - Cosmology

In this section, we discuss the results of simulations carried out on a cosmological system described below. The simulations were carried out in exactly the same manner as those in section 5, the only difference being, ofcourse, the choice of action.

7.1 The Model

We consider a system whose action is given by:

$$S = \int d^4x\sqrt{-g}(R - 2\Lambda) + \int d^4x\sqrt{-g}M(g^{ab}\partial_a\phi\partial_b\phi + 1) \tag{96}$$

The integral on the left comes from the gravitational degrees of freedom within the theory itself whereas the term on the right is the contribution from a scalar dust field ϕ with an energy density M. Now in the ADM formalism, the canonical action of such a system can be written as (using eq (87)):

$$S = \int dt\ d^3x(\pi^{ab}\dot{q}_{ab} - \mathcal{N}(\mathcal{H}_G + \mathcal{H}_D) - N^a(\mathcal{C}_G + \mathcal{C}_D)) \tag{97}$$

Some clarifications are in order here. $\mathcal{H}_G$ is the portion of the total Hamiltonian density corresponding to the gravitational degrees of freedom and is given by,

$$\mathcal{H}_G = \frac{1}{\sqrt{q}}\left(\pi_{ab}\pi^{ab} - \frac{1}{2}\pi^2\right) + \sqrt{q}(\Lambda - R) \tag{98}$$

In a similar manner, $\mathcal{H}_D$ is the portion of the total Hamiltonian density corresponding to the dust degrees of freedom and is given by,

$$\mathcal{H}_D = \frac{p_\phi^2}{2Ma^3} + \frac{Ma^3}{2}(1 + q^{ab}\partial_a\phi\partial_b\phi) \tag{99}$$

similarly, $\mathcal{C}_G$ is the portion of the momentum constraint corresponding to the gravitational degrees of freedom,

$$\mathcal{C}_G = -D_b\pi_a^b \tag{100}$$

where D represents the covariant derivative and $\mathcal{C}_D$ is the portion corresponding to the dust degrees of freedom,

$$\mathcal{C}_D = -p_\phi \partial_a \phi \tag{101}$$

As discussed in the last section of the previous chapter, we fix the time gauge by choosing our time t to be:

$$t = \epsilon\phi \tag{102}$$

where we say that $\epsilon^2 = 1$. Then, by solving the Hamiltonian constraint, we can write our physical Hamiltonian (using eq (94)) as follows:

$$H_p = -\epsilon \int d^3x \; p_\phi = \epsilon \int d^3x \; \mathcal{H}_G \tag{103}$$

For the simulations, we consider homogenous and isotropic metrics of the form:

$$ds^2 = -dt^2 + \frac{a(t)^2}{1 + kr^2/4}(dr^2 + r^2 d\Omega^2) \tag{104}$$

where k is the curvature and $a(t)$ is known as the scale factor. We define:

$$f(r) = 1 + \frac{kr^2}{4} \tag{105}$$

$$h_{ab} = f(r) e_{ab} \tag{106}$$

where e_{ab} is the 3 dimensional euclidean metric and then define the following re-parametrization:

$$q_{ab} = \frac{3}{8} A^{4/3}(t) h_{ab} \tag{107}$$

$$\pi^{ab} = 2A^{-1/3}(t) p_A(t) \sqrt{h} h^{ab} \tag{108}$$

This simplifying assumption makes things significantly easier for us because now, instead of accounting for all of the degrees of freedom inside of the tensors q_{ab} and π^{ab}, we only have to account for the newly defined phase space variable A and its conjugate

momentum p_A. Note that this new variable A is related to our original scale factor by the relation:

$$a(t) \sim A^{2/3}(t)$$

If we now apply the following rescalings:

$$\Lambda \to \frac{3}{4}\Lambda$$
$$k \to (\frac{3}{8})^{1/3}k$$

fixing $\epsilon = -1$ we get the following final physical Hamiltonian:

$$H_p = -\big(-\frac{p_A^2}{2} + \frac{\Lambda}{2}A^2 - kA^{2/3}\big) \tag{109}$$

and so the final action corresponding to a system with this Hamiltonian is given by:

$$S_E = \int dt\big(\frac{\dot{A}^2}{2} - \frac{\Lambda}{2}A^2 + kA^{2/3}\big) \tag{110}$$

Now that we have the action for our theory, and a suitable variable A to iterate over, we can finally apply the PIMC techniques we used in section 5 on our cosmological model.

7.2 Case 1: $\Lambda < 0$

7.2.1 Wavefunction averaged over all times

For negative Λ the integral converges for all values of T, i.e it is bounded, since the second term creates an upward opening parabola. We can therefore take indefinitely large values of T in our simulation.

For the first set of simulations, we computed the no-boundary wavefunction for the following set of parameters: The total time T was chosen to be 100, the initial path was generated randomly in the range $[-10, 10]$, $\Delta = 0.2$, $a = 0.01$, $N_E = 50,000$ and $N = 1000$ and $k = 1$. Λ was chosen to be -1. The wavefunction obtained is shown in Figure 14.

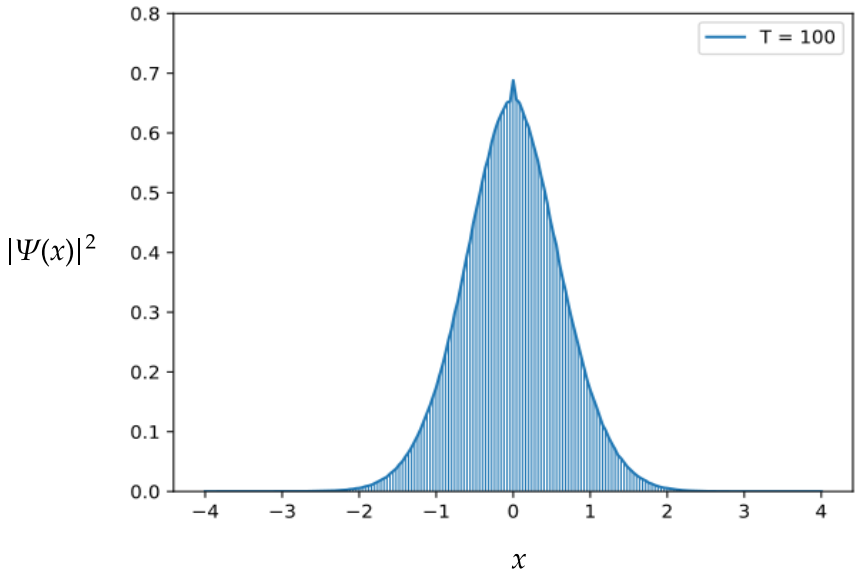

Figure 14: No-boundary wavefunction averaged over all times for $\Lambda < 0$

7.2.2 Dynamical evolution of the wavefunction

For the second set of simulations, we constructed the wavefunction for some specific values of time by binning elements at specific indices of our path array. Using this approach, we tracked the evolution of the wavefunction of our cosmological model. The results are shown in Figure 15. The parameters used were identical to the ones used for the average wavefunction.

7.2.3 Correlation function

As in the case of the harmonic and anharmonic oscillators, we also computed the correlation function for our model for the same set of parameters. The results are shown in Figure 16. In this case the number of monte carlo samples generated N_E was 200,000.

7.3 Case 2: $\Lambda > 0$

For the case where $\Lambda > 0$ the integral is not bounded for all values of T because the second term behaves as a downward opening parabola. It can be shown analytically that the integral only for the case where:

$$T < \frac{\pi}{\sqrt{\Lambda}} \tag{111}$$

thus in our simulations, the T values were chosen to be smaller than $\frac{\pi}{\sqrt{\Lambda}}$ for the positive Λ case.

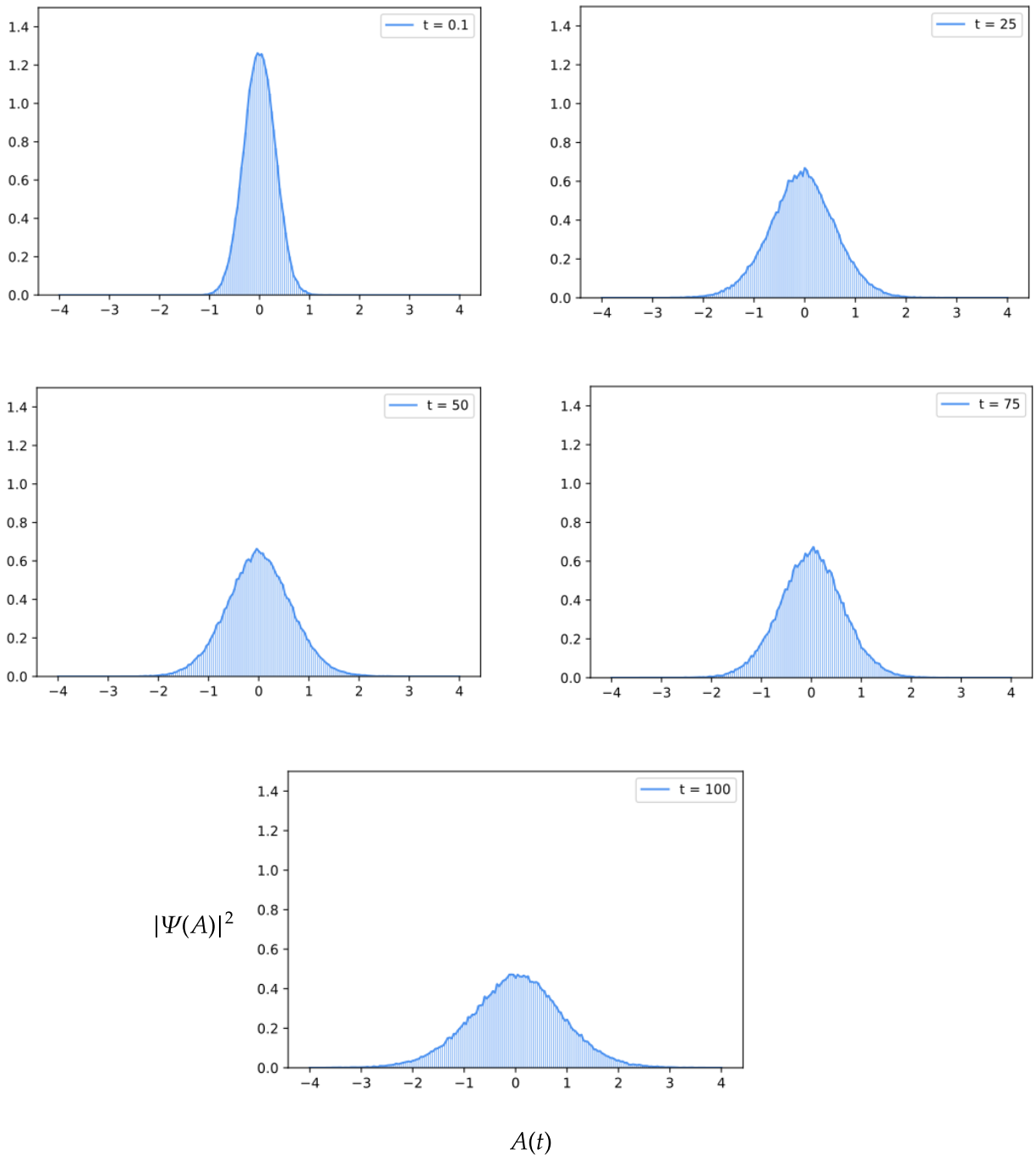

Figure 15: No-boundary wavefunctions computed for 5 different time points t = 0.1, 25, 50, 75 and 100. Note how the gaussian is peaked at smaller values of A in the beginning but then spreads out.

7.3.1 Wavefunction averaged over all times

In this case the parameters chosen were as follows $\Lambda = 0.01$, $k = 1$, $T = 10$. All other parameters were kept the same. This set of simulations took quite a bit longer to ther-

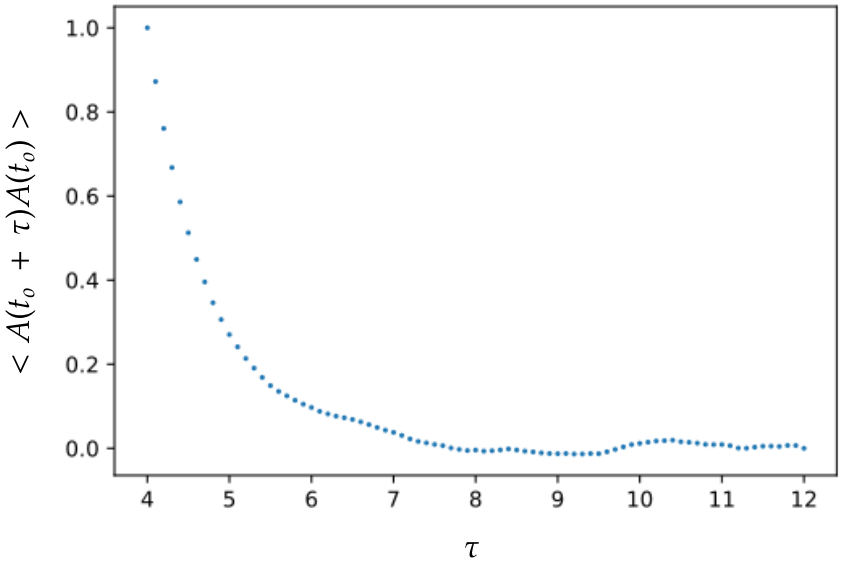

Figure 16: Correlation function for A computed for the negative λ case. Note how the correlations exponentially decay as the time interval increases

malize. The resulting wavefunction is shown in figure 17.

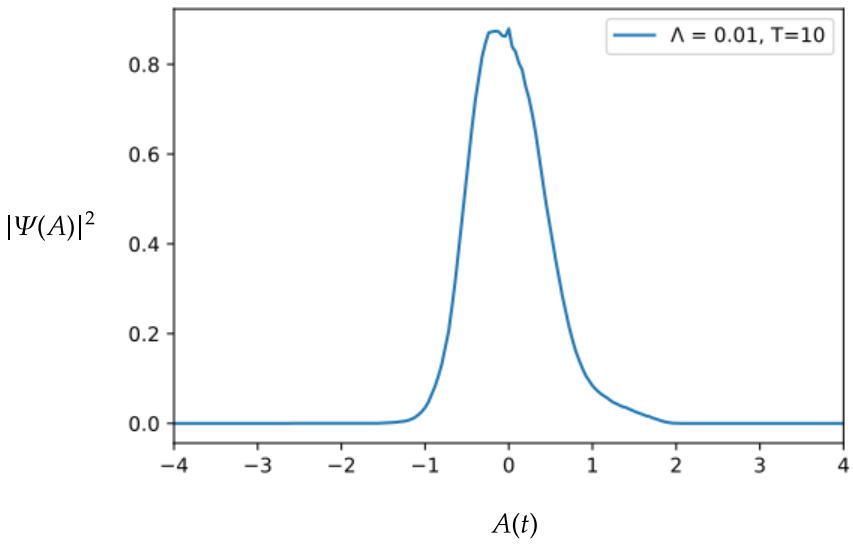

Figure 17: No-boundary wavefunction for the case for the $\Lambda > 0$ case

A Python Code

The Python code used for our simulations is provided below. This particular piece of code is for the harmonic oscillator wavefunction. The function SimulationRun can be modified appropriately to calculate correlation functions and Ground State Energies. In a similar manner, the Action function can be modified as appropriate for systems with different action.

```
import numpy as np
import random
import matplotlib.pyplot as plt

## Variables

N = 200 #Number of lattice points
a = 0.25 #Lattice spacing
NE = 300 #Number of Monte Carlo samples
nbar = 10 #Number of Monte Carlo hits per lattice site
M = 0.5 #The mass factor in the kinetic energy term
delta = 0.4 #The range within which new values should be generated
binsize = 0.02 #The size of the bins on either side
Nbins = 100 #Number of bins for the wavefunction
lamb = 1 #Lambda for the anharmonic perturbation
fsquared = 2 #Rearranged term f for the anharmonic oscillator

xs = np.random.rand(N) ## The original path that we start from (random numbers)

## A function for calculating the action (in this case for the harmonic
    oscillator)

def Action(inp):
    term1 = M*np.sum((np.square(np.diff(inp))))/(2*a)
    term2 = ((mu**2*np.sum((np.square(inp))))/2)*a

    return (term1 + term2)**2


## A function for performing a set number of Monte Carlo "hits" on a
    particular lattice point


def MonteCarloHit(x, index, xs, hits):
    currentx = x
    currentxs = xs

    for j in range(hits):
```

```
37
38          newx = random.uniform(currentx - delta, currentx + delta)
39          orgAction = Action(currentxs)
40          currentxs[index] = newx
41          newAction = Action(currentxs)
42          currentxs[index] = currentx
43
44          if orgAction > newAction:
45              currentx = newx
46
47          elif random.uniform(0,1) <= np.exp(-(newAction - orgAction)):
48              currentx = newx
49
50      return currentx
51
52  def Thermalize(init):
53      currentxs = init
54      current = init[0]
55      for i in range(100):
56          for j in range(N):
57              current = MonteCarloHit(currentxs[j], j, currentxs, nbar)
58              currentxs[j] = current
59
60          if not (i % 10):
61              print(str((i/100)*100) + " PERCENT THERMALIZED")
62
63
64      return currentxs
65
66
67  def SimulationRun(init):
68      ground = 0
69      pos = np.arange(-2,2 + binsize*2,binsize*2)
70      wavefunc = np.zeros(pos.size)
71      currentxs = init
72      for i in range(NE):
73          for j in range(N):
74              current = MonteCarloHit(currentxs[j], j, currentxs, nbar)
75              currentxs[j] = current
76              for k in range(pos.size):
77                wavefunc[k] += np.count_nonzero((currentxs >= pos[k] - binsize) &
                      (currentxs <= pos[k] + binsize))
78
79          if not (i % 10):
80
81              print(str((i/NE)*100) + " PERCENT DONE")
82              print(wavefunc)
83
```

```
84
85      return ground/(NE), wavefunc/(N*NE*2*binsize), currentxs
86
87
88  thermalized = Thermalize(xs)
89  ground, wave, xs = SimulationRun(thermalized)
90
91  plt.plot(np.arange(-2,2 + binsize*2,binsize*2), wave)
```

References

- Richard Phillips Feynman, and A R Hibbs. Quantum Mechanics and Path Integrals [By] R. P. Feynman [And] A. R. Hibbs. 1965

- Creutz, M, and B Freedman. “A Statistical Approach to Quantum Mechanics.” Annals of Physics, vol. 132, no. 2, Apr. 1981, pp. 427–462, https://doi.org/10.1016/0003-4916(81)90074-9. Accessed 10 Nov. 2022.

- Ali, Masooma and Hassan, Syed and Husain, Viqar. (2019). Monte Carlo simulation of cosmologies with dust. Classical and Quantum Gravity. 36. 10.1088/1361-6382/ab32cb.