Quantum Path Integral Sampling using Stochastic Gradient Hamiltonian Monte Carlo

Scott Hanna – shanna7@jh.edu

Final Paper – 625.744 Simulation, Modeling, and Monte Carlo – Johns Hopkins University

# Abstract

The path integral formalism of quantum mechanics is a method to model the dynamics of a quantum system that lends itself well to discretization and therefore to computational approximation methods. Unlike in classical mechanics, where a single path that extremizes the action describes the system evolution, in quantum mechanics all possible paths must be considered to appropriately estimate the dynamics of a system. Therefore, a probability distribution defined by the Feynman-Kac formula models the distribution of the path space. Markov Chain Monte Carlo (MCMC) is a common approach to draw samples from such a distribution. In this paper we consider an extension to MCMC called Hamiltonian Monte Carlo with a stochastic gradient estimation method (SG-HMC) that uses simultaneous perturbation stochastic approximation (SPSA) as the driver of the state dynamics. We apply MCMC and SG-HMC to the problem of the quantum harmonic oscillator. The SG-HMC method achieves similar estimates of the ground state energy and comparable convergence to that of MCMC.

# Introduction

In classical mechanics, given the positions and momenta of a system at a particular point in time, we can make predictions about the future state of the system by use of deterministic dynamics (Forsyth, 1960). One method used to analyze the dynamics of a physics system is through the calculation of paths of least action. Action is defined at the path integral over the Lagrangian: the difference between the kinetic and potential energy of the system. The states of a system will evolve along a path that minimize the action (Forsyth,1960).

In practice, analytically computing an outcome may be an intractable problem due to the large number of interacting particles within the system, or the complexity of the underlying potential. Therefore, statistical methods are employed to determine averaged behavior of the system. One of the benefits of the path integral formulation is the ease which paths can be discretized and therefore approximated for complex systems. Several statistical optimization techniques have been employed to optimize paths of least action for physical systems (Dornheim, 2019, Gonzalez, 2020).

In quantum mechanics paths of least action are an important tool for determining the dynamical behavior of quantum systems. Unlike in classical mechanics, however, quantum systems do not evolve under the dynamics of a single realized path, but instead are better modeled as a distribution of all possible paths (Joseph, 2020). Therefore, optimization tools like those employed for single paths of least action are not well suited for quantum mechanical systems. Instead, approaches that aim to average over the probabilistic path space are utilized.

One such method of approximating the probability distribution of the path space is Markov chain Monte Carlo (MCMC). Using this method, paths are discretized into a sequence such that the transition between subsequent paths are dependent on the action dynamics. This transition probability can then be sampled over and accepted or rejected on the Metropolis-Hasting (MH) criterion. Given some thermalization (bake-in) time, the distribution will converge to the target distribution of the path space (Mittal, 2020).

Another more recent technique that builds on the MCMC framework is called Hamiltonian Monte Carlo (HMC). In this method dynamic paths in the phase space of the path are integrated over by means of Hamilton’s equations of motion. This integration method has traditionally been determined by means of evolving the Newtonian dynamics of the state. Accepting or rejecting such an evolved path can be achieved by the MH criterion as in standard MCMC. The HMC method has been shown to improve sampling efficiency of the MCMC algorithm (Joseph, 2020).

In this paper, an extension of HMC is considered called stochastic gradient Hamiltonian Monte Carlo (SG-HMC). This method evolves the path through phase space based on dynamics determined by a stochastic gradient algorithm. With appropriate conditions, this method preserves phase-space volume and therefore can avoid the necessity of accepting or rejecting proposed paths using the MH criterion (Chen, 2014). We use a SG-HMC algorithm that makes use of the simultaneous perturbation stochastic approximation (SPSA) algorithm as the stochastic gradient estimator.

In this paper two algorithms are analyzed, MCMC, and SG-HMC, on a well-known quantum mechanical problem: the quantum harmonic oscillator. The ground state solution of the quantum harmonic oscillator has an analytical and experimentally supported solution based on the Schrodinger equation. This allows us to compare the performance of the two algorithms against an accepted benchmark.

# Path Integral Formalism in Quantum Mechanics

In classical mechanics, a particle or system will evolve along a path in space-time that minimizes (or, more generally, extremizes) the action along the path. The action is defined as the difference between the kinetic and potential energy of the system summed over the path. The action, , is therefore a functional, mathematically defined as

(1)

Where is the path defined over times to . Assuming proper extremizing of the action, the resulting path represents the realized space-time path of the system.

In quantum mechanics, however, this no longer holds. Every path between the initial location and target location represents a possible and realizable path of the system (Joseph, 2020). The goal therefore is no longer to find the correct realized path, but to find a probability distribution of possible paths that a system could traverse. One approach to determine this distribution is by use of a time evolution operator acting on the state of the system. This operator moves the probability space of the system forward in time and describes the time evolution of the probability amplitude of the wave function which is a state wave function defined over space and time. The Schrodinger equation is the dynamical equation that describes the time evolution of the probability amplitude for non-relativistic masses:

The probability distribution is found as where represents the complex conjugate of the wave function. However, solving the Schrodinger equation is tractable for only certain problems, often when a set of stationary eigenfunctions can be determined with respect to a simple potential distribution (Griffiths, 2004). An alternative formulation of the quantum dynamics using path integrals can be utilized that will allow us to discretize the problem and find approximate solutions for more complicated potentials. The Feynman-Kac equation describes the probability amplitude over a set of paths dependent on the action of each path:

Where the integration over means we integrate over all paths for to , is defined as in Equation 1, is Plank’s constant, and . The notation denotes the probability amplitude between the initial and final location. The imaginary exponent makes summing over discretization of the probability distribution difficult and is not well suited to optimization techniques. A solution to avoid this problem is to perform a transformation into imaginary time called a Wick rotation (Joseph, 2020). That is, we let which transforms the Feynman-Kac path equation into

(2)

and allows us to reformulate the quantum dynamics into a statistical mechanics problem.

In this paper we consider the specific problem of the harmonic oscillator which has the defined potential and gives the action functional:

.

The quantum harmonic oscillator has well known solutions governed by the Schrodinger equation (Griffiths, 2004). In the next section we will compare our approximation techniques of the path integral formalism to the analytical solutions for the ground state energy level of the quantum harmonic oscillator.

# Markov Chain Monte Carlo for Path Integrals

One important method with which we can sample from the path integral distribution in equation 2 is the Markov chain Monte Carlo (MCMC) technique. In MCMC we consider a random process in which the probability at each state of the process only depends on the previous state. That is

It has been shown that given an appropriate choice of proposal distribution, that defines the probability of transition from state to , the Markov process will converge in probably to a target distribution under irreducibility and stationarity of the process (Spall, 2003).

We can therefore use MCMC to draw samples from the probability amplitude which represents our target distribution:

We consider a proposal perturbation to the current state given by the transition probability . Then we accept or reject based on the Metropolis-Hastings criterion:

There are many viable proposal distributions, but we can simplify the acceptance ratio by choosing one that is symmetric such that .

For the target distribution the MH algorithm only requires that we choose a distribution that is proportional to (Joseph, 2020) Thus, we can choose which results in an acceptance probability of

where we define , the difference in action between the newly proposed path and the original path. In this paper we will also assume that for simplification purposes. To find this action difference, we will need to approximate the action over some discretized path. We therefore break up the path into equally sized time slices of length :

where is the path and is the average value between two successive times. For the quantum harmonic oscillator of frequency the potential is given by,

In summary, for the standard MCMC algorithm with MH acceptance probability we follow these steps (at iteration ):

1. Sample from the proposal distribution giving
2. Determine the new path,
3. Accept the new path if where the standard uniform distribution.
4. Repeat steps 1-3 for some predefined number of thermalization steps. We assume that after thermalization the algorithm have converged near the target distribution.
5. Repeat steps 1-3 again for the number of sweeps desired. The statistics of the target distribution can be gathered from analysis of the accepted samples.

The standard MCMC is quite robust and requires only a few parameters but suffers from the disadvantages of other accept-reject algorithms. First, many potential proposed jumps are thrown out when rejected. Some fine tuning usually needs to be done to achieve an acceptance ratio of around 50% (Joseph, 2020). Secondly, the random direction of proposed jumps takes no information from the underlying action function. A method that considers gradient information in the path space may reduce the number of steps required to converge to the target distribution.

One method to explore the path space dynamically is through the Hamiltonian Monte Carlo (HMC) technique. In HMC, paths are propagated through the Hamiltonian pair of differential equations:

Here, corresponds to the momentum of the system but can be thought of as an intermediary quantity in the dynamic equations. The Hamiltonian has some important properties that we can make use of to apply it to Monte Carlo methods. First, Hamiltonian dynamics are reversible, a requirement for MCMC algorithms. Second in the noiseless case which implies that, in such cases, remains invariant under perturbations to the path and the acceptance ratio of the MH algorithm is 1. Meaning, we could avoid the MH accept-reject step altogether. Finally, Hamiltonian dynamics preserves volume of the phase space (Joseph, 2020).

The Hamiltonian equations can be used to derive the dynamic evolution of the state space. However, solving them analytically is not always achievable. Therefore, an integration technique is often used called leap-frog integration that approximates updates to the path-space dynamics over some number of iterations. This leap-frog technique is shown in step 3 of the HMC steps below. These steps outline the algorithm at the ’th iteration:

1. Let . Draw . Where is the zero vector with the same dimension (*d*) as and is the identity matrix.
2. With MH step: Let
3. For iterations update and using Hamiltonian dynamics as follows (this step is often called the Leapfrog method of integration).
4. Let
5. With MH step: Let . Let Then accept if for .
6. With no MH step:
7. Repeat steps 1-6 for the number of chosen Thermalization Steps and Sweeps.

One requirement of the standard implementation of HMC is knowing the gradient of . For many problems, finding such a gradient for a given quantum mechanical action function can be untenable, so the question arises whether gradient estimate methods can be utilized. Chen, 2014 proposes a stochastic gradient approximation HMC implementation. They find that naively replacing = where is a batch stochastic gradient approximation performs arbitrarily poorly. They also show that due to the random nature of the gradient approximation it does not preserve the invariance of which means we will not see convergence to the target distribution.

Their solution, and the one adopted in this paper, is to add in a friction correction term into the Hamiltonian dynamics. This friction term helps reduce the energy added to by the effects of additional noise from the stochastic gradient approximation (Chen, 2014). The corrected step three therefore becomes:

3. (stochastic correction): For iterations update and using Hamiltonian dynamics as follows:

where is a friction coefficient parameter. Chen, 2014 showed that adding this friction correction preserves the convergence to the stationary distribution, . Furthermore, under certain conditions related to the choice of gradient approximation method and , the MH-step is not required (Chen, 2014). Essentially, the friction term, when properly tuned, prevents the additional noise in the stochastic gradient approximation from overwhelming the dynamics approximation.

For this paper, the simultaneous perturbation stochastic approximation (SPSA) method is used as the stochastic gradient approximation. This method has the advantage of only requiring two action measurement per iteration regardless of the dimension of the parameter space. Thus,

where is a chosen perturbation distribution. Note that this is a modification to SPSA in (Spall, 2003) where the gradient difference is multiplied by the element-wise inverse of . For the implementation of SPSA in this paper perturbations are chosen such that . For this Bernoulli random perturbation multiplying by is equivalent to multiplying by its element-wise inverse (Chin, 1997).

# Analysis

In this section we analyze the problem of a quantum system in a harmonic oscillator potential in one dimension. We apply two algorithms, MCMC and SG-HMC to approximating the resulting wavefunction and compare that to the known solution from the Schrodinger equation. The Schrodinger equation for the ground state of the harmonic oscillator is given by (Griffiths, 2004),

In this paper , therefore the ground state probability distribution is a normal distribution with mean and variance and the ground state energy .

The paths were split into discretized components. Thus, the path is a 48-dimensional parameter space. Both Monte Carlo methods were run with a thermalization (bake-in) of iterations and sweeps. For standard MCMC a symmetrical uniform proposal distribution of was used for each component. The SPSA perturbation distribution was a uniform for each component with parameter . The number of Hamiltonian dynamic iterations with and friction coefficient . The SG-HMC algorithm was performed with no MH accept-reject step.

To analyze the data, we first want to know the spatial distribution that results over the realized paths. Each sample path oscillates over the imaginary-time sequence. Three such samples are displayed in Figure 1.

Chart, line chart, histogram

Description automatically generated

Figure : Three sample paths taken from the SH-HMC algorithm. Aggregating the frequency by which the paths cross a specific location, x, gives a probability distribution of the spatial wave function (shown in Figure 2).

To simulate the probability distribution, we aggregate the frequency that a path crosses a position, *x*, for every sample path of the 10,000 sweeps and find the distribution using a normalized histogram as is done in (Mittal, 2020). Figure 2 shows a histogram of the results for both algorithms.

Chart, line chart

Description automatically generated

Figure : Histogram of results for MCMC and SGHMC for 10,000 Sweeps (after 1000 thermalization steps). The orange curve is the analytical Schrodinger solution.

Both algorithms show good alignment visually with the analytical solution. An important statistic we can use to analyze the simulated results is an analysis of the ground state energy estimates. The ground state energy can be estimated by taking the expected value (Mittal, 2020). This can be compared to the accepted value of based on the analytical Schrodinger equation solution. For the 10,000 sweeps, every 10 estimates were averaged and found. Both results converge to the accepted value with similar confidence intervals (with a slight advantage to MCMC). The results are displayed in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **MCMC** | **SGHMC** | **Accepted Value** |
|  | 0.501 | 0.495 | 0.5 |
| 95% Confidence Interval | [0.496,0.505] | [0.490,0.501] | - |

Table 1: Ground state energy estimates. is shown for both the MCMC and SGHMC algorithms based on 10,000 sweeps. The 0.95% T-test confidence intervals are also shown.

Another important consideration for the algorithms is the rate of convergence of the Markov chain. Figure 2 shows the ground state energy estimates for the first 400 function evaluations of each algorithm (during the thermalization stage). The SG-HMC algorithm required 10 times as many function evaluations per iteration as the MCMC algorithm (20 compared to 2). This is due to the necessary Hamiltonian dynamic steps for the HMC algorithm. Using fewer than steps degraded the accuracy of SG-HMC algorithm. Despite this large difference in function evaluations, SG-HMC convergence is on par with MCMC convergence as can be seen in Figure 3.

Chart, line chart

Description automatically generated

Figure 3: In orange and green are the energy level estimates () over 400 function evaluations. This shows relative convergence to the accepted value. Despite requiring times more function evaluations per iteration, the SGHMC algorithm convergence is on par with MCMC convergence.

# Discussion

**Results**

The SGHMC algorithm showed similar results to the MCMC algorithm both in terms of convergence rate and accuracy after 10,000 sweeps. However, unlike the MCMC algorithm, the SGHMC algorithm required careful fine tuning of many parameters. These parameters were necessarily part of the Hamiltonian dynamics which are approximated through the leap-frog integration. The parameters and took careful tuning to ensure the algorithm did not diverge.

Both algorithms saw equivalent convergence despite the much greater number of function evaluations (10x) required by the SGHMC algorithm. This higher number of evaluations is likely balanced by the 100% acceptance ratio and improved sampling of the action function using a gradient descent algorithm.

While there was no advantage to the SG-HMC algorithm, this paper represents a first attempt at implementation. There are several improvements that can be made to HMC algorithms were not tried due to time constraints (see Joseph, 2020). Furthermore, because we have refocused the sampling problem into a stochastic gradient problem, we could potentially utilize the vast amount of work done in the field of stochastic gradient optimization (see Spall, 2003) to further improve the algorithm. Some of these modifications are mentioned in the next section.

**Further Research**

The work in this paper represents a first-order stochastic gradient descent approximation method using SPSA. It is possible that using a higher order method like 2nd order SPSA (Spall, 2003) could improve results, especially considering the Hamiltonian dynamics use a second order update step. There could be some important interplay between the dynamics update and the gradient descent algorithm that may reduce the number of dynamic steps and thus the total number of function evaluations for SG-HMC.

Another possible avenue to improve the efficiency of the stochastic gradient step is to transform the path space to a lower dimensional space by means of a linear transformation. I used this method in my previous paper on classical path integral optimization and it improved convergence and terminal mean squared error results. For the classical problem in that paper a linear transformation based on Bezier curves was used (see González, 2020), but for a quantum mechanical problem, other types of transformations may be prudent. This method could decrease auto-correlation in the Markov chain process by increasing the distance between perturbations to the path space (Joseph, 2020).

As an extension, the algorithms utilized here were all written to be dimension independent. It may be interesting to look at 2- or 3-dimensional quantum systems. Quantum systems in higher spatial dimensions are interesting as energy levels are further split by degeneracy (more than one state per energy level). This is an important quantum property for the study of real-world quantum systems (Griffith, 2004). Additionally, changing the potential energy function is quite easy and it would be interesting to see other quantum potentials such as that of the hydrogen atom or higher order molecular dynamical systems.

Finally, the tools developed here can also be applied to other domains. The Feynman-Kac formula is well known outside of physics as a tool for describing distributions related to diffusions and Ito processes (Øksendal,2007). Additionally, path integral simulations have been extended to fields such as finance (Øksendal,2007) and machine learning (Chen, 2014).

# References

Chen, T., Fox, E., Guestrin, C. (2014), “Stochastic Gradient Hamiltonian Monte Carlo,” Proceedings of Machine Learning Research, 1683-1691, Vol. 32, No. 2.

Chin, D.C. (1997),"Comparative Study of Stochastic Algorithms for System Optimization Based on Gradient Approximations," IEEE Transactions on Systems, Man, and Cybernetics — B, vol. 27, pp. 244-249 (theoretical and numerical efficiency analysis).

Dirac, Paul A. M. (1933), “The Lagrangian in Quantum Mechanics,” Physikalische Zeitschrift der Sowjetunion. 3 (1): 64–72.

Dornheim, T., Groth, S., Filinov, V., and Bonitz, M. (2019), “Path integral Monte Carlo simulation of degenerate electrons: Permutation-cycle properties,” The Journal of Chemical Physics 151, 014108.

Forsyth, A. R. (1960), “Calculus of Variations”. New York: Dover, pp. 17-20 and 29.

González Diaz, D., Davis, S., & Curilef, S. (2020), “Solving Equations of Motion by Using Monte Carlo Metropolis: Novel Method Via Random Paths Sampling and the Maximum Caliber Principle,” Entropy (Basel, Switzerland), 22(9), 916.

Griffiths, D. (2004), “Introduction to Quantum Mechanics,” Prentice Hall, 2nd Ed.

Joseph, A. (2020), “Markov Chain Monte Carlo Methods in Quantum Field Theories: A Modern Primer,” Cham: Spring International Publishing, 1st Edition.

Marx, D., and Parrinello, M. (1996), “Ab initio path integral molecular dynamics: Basic ideas,” The Journal of Chemical Physics 104, 4077-4082.

Mittal, S., Westbroek, M., King, P., and Vvendensky, D. (2020), “Path integral monte carlo method for the quantum anharmonic oscillator,” European Journal of Physics, 41(5):055401.

Øksendal, B. (2007), Stochastic Differential Equations: an Introduction with Applications. Berlin: Springer. 6th Ed.

Spall, J. C. (2003), “Introduction to Stochastic Search and Optimization,” Hoboken, NJ, USA: Wiley-Interscience.

Code for project available at: <https://github.com/sphanna/SGHMC>

12/9/21