Reweighting and the Complex Langevin Algorithm

Peter Tsimiklis

Master of Science

Department of Physics

McGill University

Montreal, Quebec

Fall 2016

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science

Abstract

The complex Langevin algorithm was developed to evade the sign problem by extending the degrees of freedom into a complex space. The validity of the results depends on the fall off of the resulting probability distribution. To explore the falloff, we attempted to reweight the probability distribution by introducing an acceptance probability at each step in the algorithm, and by direct manipulation of the Fokker-Planck equation. The acceptance probability did not lead to the target reweighted distribution because the complex Langevin algorithm does not follow detailed balance. Manipulating the Fokker-Planck equation gave a differential equation for the reweighted distribution, but we could not find a way to simulate it.

Abstrait

AbstraitL'algorithme complexe de Langevin a été développé pour échapper la problme de la signe en étendant les degrés de liberté dans un espace complexe. La validité des résultats de l'algorithme dépend de la décroissance de la distribution de probabilité résultante. Pour analyser cette décroissance, nous avons tenté de modifier la distribution de probabilité en introduisant une probabilité d'approbation à chaque étape de l'algorithme et en manipulant directement l'équation de Fokker-Planck. La probabilité d'approbation n'a pas reproduit à la distribution desirée puisque l'algorithme Langevin complexe ne suit pas le principe la balance détaillé. La manipulation de l'équation de Fokker-Planck a donné une équation différentielle pour la distribution modifier, mais nous n'avons pas trouvé un moyen de la simuler.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Guy Moore. He was always available to discuss problems and was always willing to give guidance.

I would also like to thank my fellow students especially Jeremy Rothschild and Eric Hanson for being open to conversation regarding work and for their moral and essential technical support.

Finally I'd like to thank Krypto for his years of support and generally being a good dog.

Thank you!

TABLE OF CONTENTS

CHAPTER 1
Introduction and Background

A common strategy for determining the value of observables such as correlation functions in quantum field theories is to write them as an integral over all trajectories through spacetime with a complex weight involving the action of the theory [1]. In practice, solving this so called path integral is not possible for all but the simplest quantum field theories. In some cases, proceeding numerically is the most effective method of obtaining useful results from the path integral formulation. Before applying numerical methods, the problem must be discretized and reformulated onto a lattice while preserving symmetries.

Once discretized, the path integral becomes a finite number of integrals over each lattice site. These finite integrals still involve a complex weight, making typical Markov Chain Monte Carlo techniques ineffective in many cases. In order to circumvent the complex phase in the integral, it is common to perform a Wick rotation, which is essentially replacing the time variable with a complex time variable. This transformation changes the complex weight into a real one and in some cases, Monte Carlo techniques become available [2]. However for many theories the Wick rotation does not completely eliminate the complex phase.

The phase diagram of QCD is of particular fundamental importance. At nonzero chemical potential, the phase diagram is almost completely unverified numerically except in some limiting cases. Large regions of parameter space are currently unexplored by simulations because of the so called sign problem, which arises from a complex determinant in the action [3]. Other systems

which exhibit the sign problem include the XY chain and the relativistic Bose gas at finite temperature. These simpler systems have been used as testing grounds for methods to circumvent the sign problem [4][5].

One proposed method of averting the sign problem is the complex Langevin algorithm. In practice, this is similar to the standard Langevin algorithm, but uses the complexity of the action to drive the real degrees of freedom into the complex plane. The advantage of such an algorithm is that it doesn't rely on the weight as a probability distribution. The complex Langevin equation has so far shown promise, giving accurate results when studying systems such as the relativistic Bose gas and the XY model [4]. Though the algorithm shows potential, it fails in some cases, making it hard to believe any predictive results that could not be verified by other techniques.

In this thesis, we will go through the theory of Markov chains and Monte Carlo techniques, as well as detail the causes of the sign problem. The main body of this text will be reviewing prior works which have developed the complex Langevin algorithm and have created formal arguments for its validity. Finally, we will detail some attempts we have made at incorporating reweighting into the algorithm.

Before exploring the sign problems and the algorithm to avert it, we will review the process of deriving the path integral formulation.

1.1 Path Integrals

The path integral formulation of quantum theory is indispensable for constructing theories on a lattice and gives an important and unique perspective into quantum theory. To illustrate the physical reasoning behind the formulation, consider performing a series of multi slit experiments [6]. In such an experiment, a particle travels from a source to a screen. The screen is partially blocked by an opaque wall which has two narrow slits in it. The particle can

reach any point on the screen via two paths; one for each slit. The amplitude of a particle reaching our point on the screen is simply the sum of the amplitudes of each path.

Now imagine adding another double slit between the source and the screen. There are now four amplitudes which need to be summed in order to determine the probability that the particle hits a particular point on the screen. Consider replacing the double slit walls with walls which have three slits. In this case, there are nine amplitudes to consider.

The path integral formulation is a more extreme version of this multi slit experiment. Imagine having a million (or more) walls with a million (or more) slits in them between the source and the screen. There are now a million-million paths and amplitudes to consider. In fact, a wall with so many slits in it looks like empty space. What this extreme experiment is really suggesting is to consider all the paths a particle can take through empty space, and consider the quantum mechanical amplitude for each. The accounting of all the many (infinite) trajectories a particle can take are encompassed in the path integral as we will see. The question of how to compute such an elaborate sum is an important one in the study of lattice quantum field theories.

In quantum mechanics, the state $|\psi\rangle$ of a system evolves from initial time t_0 to a later time t via the Schrödinger equation

$$|\psi(t)\rangle = e^{-iH(t-t_0)}|\psi(t_0)\rangle,$$

where H is the Hamiltonian of the system. Consider a system with degrees of freedom q_α where α is an index running from 1 to n. We denote the set of degrees of freedom by q and denote the simultaneous eigenstates of the system by $|q\rangle$. Quantum mechanical amplitudes such as the amplitudes described above

in the multi slit experiment are written as $\langle q', t'|q, t\rangle = \langle q'|e^{-iHt'}e^{iHt}|q\rangle$. Inserting a complete set of energy eigenstates $\sum_n |E_n\rangle\langle E_n|$ in the matrix element gives

$$\langle q', t'|q, t\rangle = \sum_n e^{-i(t'-t)E_n}\psi_n(q')\psi_n^*(q),$$

where E_n is the energy of the eigenstate $|E_n\rangle$ and $\psi_n(q)$ is the wavefunction $\langle q|E_n\rangle$. At this point, it is convenient to transform this equation to imaginary time. In practice, this amounts to making the substitution $t \to -i\tau$ and $t' \to -i\tau'$. This substitution, called a Wick rotation, simplifies the calculations by eliminating the complex exponential in the problem.

With this transformation done, the interval $[\tau', \tau]$ can be divided into subintervals of length $\epsilon = \frac{\tau'-\tau}{N}$ for a large integer N. The endpoints of the subintervals are denoted by τ_i with i ranging from 0 to N and $\tau_0 = \tau$, $\tau_N = \tau'$. The matrix element in question can be written as

$$\begin{aligned}\langle q', t'|q, t\rangle &= \langle q'|e^{-H(\tau'-\tau)}|q\rangle \\ &= \langle q'|e^{-H(\tau'-\tau_{N-1})}e^{-H(\tau_{N-1}-\tau_{N-2})}\dots e^{-H(\tau_1-\tau)}|q\rangle.\end{aligned}$$

At each intermediate time, a complete set of eigenstates can be inserted:

$$\begin{aligned}\langle q', t'|q, t\rangle = \int \mathrm{d}q^{(N-1)} \dots \int \mathrm{d}q^{(1)}\, &\langle q'|e^{-H\epsilon}|q^{(N-1)}\rangle\langle q^{(N-1)}|e^{-H\epsilon}|q^{(N-2)}\rangle \\ &\langle q^{(N-2)}|\dots|q^{(1)}\rangle\langle q^{(1)}|e^{-H\epsilon}|q\rangle.\end{aligned}$$

The integral $\int \mathrm{d}q$ represents an integral over all coordinate degrees of freedom q_α. More concisely, the many integrals are written

$$\langle q', t'|q, t\rangle = \prod_{l=1}^{N-1} \int \mathrm{d}q^{(l)}\, \langle q'|e^{-H\epsilon}|q^{(N-1)}\rangle\langle q^{(N-1)}|e^{-H\epsilon}|q^{(N-2)}\rangle\langle q^{(N-2)}|\ldots$$
$$\ldots|q^{(1)}\rangle\langle q^{(1)}|e^{-H\epsilon}|q\rangle. \tag{1.1}$$

At this step, it is apparent that the amplitude $\langle q', t'|q, t\rangle$ is the product of many intermediate amplitudes, representing the multiple trajectories through coordinate space the system can explore. This is in line with our multi slit experiment described above.

Proceeding further, we choose a specific form for the Hamiltonian. The typical example of a Hamiltonian is $\frac{1}{2}P_iP_i + V(Q)$, where the operators Q_α are the operators corresponding to the q_α degree of freedom ($Q_\alpha|q\rangle = q_\alpha|q\rangle$). The operators P_i are the momentum operators, which are conjugate to Q_i. The repeated indices are summed over.

With this Hamiltonian, and using the Baker-Campbell-Hausdorff theorem,

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\ldots},$$

any one of the amplitudes in equation (1.1) can be written as

$$\langle q^{(l+1)}|e^{-\epsilon\frac{1}{2}P_iP_i-\epsilon V(Q)}|q^{(l)}\rangle \approx \langle q^{(l+1)}|e^{-\frac{\epsilon}{2}P_iP_i}|q^{(l)}\rangle e^{-\epsilon V(q)},$$

where the terms which are proportional to ϵ^2 and higher powers of ϵ have been disregarded because ϵ is arbitrarily small. In order to evaluate the momentum operators, a complete set of momentum eigenstates are placed to the right of the exponential.

$$
\begin{aligned}
e^{-\epsilon V(q)}\langle q^{(l+1)}|e^{-\frac{\epsilon}{2}P_iP_i}|q^{(l)}\rangle &= \int \mathrm{d}p^{(l)}\, e^{-\epsilon V(q)}\langle q^{(l+1)}|e^{-\frac{\epsilon}{2}P_iP_i}|p^{(l)}\rangle\langle p^{(l)}|q^{(l)}\rangle \\
&= \int \mathrm{d}p^{(l)}\, e^{-\epsilon V(q)-\frac{\epsilon}{2}{p^{(l)}}^2}\langle q^{(l+1)}|p^{(l)}\rangle\langle p^{(l)}|q^{(l)}\rangle \\
&= \int \frac{\mathrm{d}p^{(l)}}{(2\pi)^n}\, e^{-\epsilon V(q)} e^{ip_i^{(l)}q_i^{(l+1)}} e^{-ip_i^{(l)}q_i^{(l)}} \\
&= \int \frac{\mathrm{d}p^{(l)}}{(2\pi)^n} e^{-\epsilon V(q)} \exp\left\{-\epsilon\left(\frac{1}{2}{p^{(l)}}^2\right.\right. \\
&\qquad \left.\left. - ip_i^{(l)}\left(\frac{q_i^{(l+1)}-q_i^{(l)}}{\epsilon}\right)\right)\right\}.
\end{aligned}
\tag{1.2}
$$

In equation 1.2, the canonical eigenfunction was used,

$$
\langle q|p\rangle = \frac{1}{\sqrt{2\pi}}e^{ipq}, \tag{1.3}
$$

and $\mathrm{d}p$ is shorthand for $\prod_{i=1}^{n}\mathrm{d}p_i$. This integral can be calculated by completing the square to turn it into a simple Gaussian integral:

$$
\begin{aligned}
&\int \frac{\mathrm{d}p^{(l)}}{(2\pi)^n} e^{-\epsilon V(q)} \exp\left\{-\epsilon\left(\frac{1}{2}{p^{(l)}}^2 - ip_i^{(l)}\left(\frac{q_i^{(l+1)}-q_i^{(l)}}{\epsilon}\right)\right)\right\} \\
&\quad = e^{-\epsilon V(q)} \exp\left\{-\frac{\epsilon}{2}\left(\frac{q^{(l+1)}-q^{(l)}}{\epsilon}\right)^2\right\} \int \frac{\mathrm{d}p^{(l)}}{(2\pi)^2} \exp\left\{-\frac{\epsilon}{2}\left(p^{(l)}\right.\right. \\
&\qquad\qquad \left.\left. - i\frac{q^{(l+1)}-q^{(l)}}{\epsilon}\right)^2\right\} \\
&\quad \propto e^{-\epsilon V(q)} \exp\left\{-\frac{\epsilon}{2}\left(\frac{q^{(l+1)}-q^{(l)}}{\epsilon}\right)^2\right\}.
\end{aligned}
\tag{1.4}
$$

The term in the exponential can be rewritten. The states $|q^{(l)}\rangle$ were inserted as intermediate states. In other words, integrating over $q^{(l)}$ is really integration over possible states at the (imaginary) time $i\tau_l$. As such, the term $\frac{q^{(l+1)}-q^{(l)}}{\epsilon}$ begins to look a lot like a (imaginary) time derivative. This terms is then denoted $\dot{q}^{(l)}$.

The total exponential in equation(1.4) now resembles the Lagrangian from classical mechanics. The (Euclidean) Lagrangian is defined by

$$L_E(q^{(l)}, \dot{q}^{(l)}) = V(q^{(l)}) + \frac{1}{2}\dot{q}^{(l)}. \tag{1.5}$$

The amplitude we are calculating then becomes

$$\begin{aligned} \langle q|e^{-H(\tau'-\tau)}|q\rangle &\approx \int \prod_{l=1}^{N-1} \frac{\mathrm{d}q^{(l)}}{\sqrt{2\pi\epsilon}} e^{-\epsilon L_E(q^{(l)},\dot{q}^{(l)})} \\ &= \int \left(\prod_{l=1}^{N-1} \frac{\mathrm{d}q^{(l)}}{\sqrt{2\pi\epsilon}} \right) e^{-\sum_{l=0}^{N-1} \epsilon L_E(q^{(l)},\dot{q}^{(l)})} \end{aligned} \tag{1.6}$$

The sum in the exponential is the (Euclidean) action,

$$S_E = \sum_{l=0}^{N-1} \epsilon L_E(q^{(l)}, \dot{q}^{(l)}) \tag{1.7}$$

The integration measure in equation (1.6) is an integral over all the possible intermediate states of the system. Such an integral is called a path integral and is denoted by $\int_q^{q'} \mathcal{D}[q]$. In all, equation (1.6) is written in a continuum form as

$$\langle q|e^{-H(\tau'-\tau)}|q\rangle = \int_q^{q'} \mathcal{D}\left[q\right] e^{-S_E(q,\dot{q})}, \tag{1.8}$$

where the continuum action is

$$S_E(q, \dot{q}) = \int_\tau^{\tau'} \mathrm{d}\tau'' L_E(q(\tau''), \dot{q}(\tau'')).$$

The paths which are important in the integral are those for which $\delta S_E = 0$. Deviations from these extremum paths are exponentially suppressed. Expanding around the extremum of the action gives the usual Euler-Lagrange equations.

The path integral can only be solved exactly for very few cases in quantum mechanics. Most problems in quantum mechanics are solved simply without this formulation. In the next section, we will look at the path integral in the

context of quantum field theory, where the formalism is more useful. The path integral allows for the derivation of the Feynman rules and Feynman diagrams.

1.2 Path Integrals for Fields

In the previous section we managed to write a quantum mechanical amplitude as a path integral. Through various integral solving techniques, the path integral can be used to obtain an actual value for the amplitudes. In this section we will use the same line of reasoning for quantum fields. The simplest operator we can consider is the scalar field. Using the Heisenberg picture, the time evolution is given by

$$\phi(\mathbf{x}, t) = e^{iHt}\phi(\mathbf{x}, 0)e^{-iHt}. \tag{1.9}$$

In field theory, the vacuum expectation values are the focus of interest:

$$G(x_1, x_2, ..., x_l) = \langle\Omega|T\{\phi(x_1)...\phi(x_l)\}|\Omega\rangle, \tag{1.10}$$

where $\langle\Omega|$ is the ground state (vacuum) and T represents time ordering, which ensures that the fields are sorted in order of their time component. The transition to imaginary time for Heisenberg operators gives, for the operator Q_i:

$$\hat{Q}_i(\tau_i) = e^{H\tau_i}Q_i e^{-H\tau_i} \tag{1.11}$$

The vacuum expectation value can be obtained from the matrix element

$$\langle q'|e^{-H\tau'}\hat{Q}_1...\hat{Q}_l e^{-H\tau}|q\rangle. \tag{1.12}$$

To show this, first insert energy eigenstates to the left and right of the exponentials. This yields

$$\langle q'|e^{-H\tau'}\hat{Q}_1...\hat{Q}_l e^{-H\tau}|q\rangle = \sum_{\kappa,\kappa'} e^{-E_{\kappa'}\tau'}e^{E_\kappa\tau}\psi_{\kappa'}(q')\psi_\kappa(q)\langle E_{\kappa'}|\hat{Q}_1...\hat{Q}_l|E_\kappa\rangle \tag{1.13}$$

In the limit where τ' and τ approach infinity and negative infinity respectively, only the lowest energy term remains relevant in the sum. This leads to

$$\langle q'|e^{-H\tau'}\hat{Q}_1...\hat{Q}_l e^{-H\tau}|q\rangle \to e^{E_0(\tau'-\tau)}\psi_0(q')\psi_0^*(q)\langle E_0|\hat{Q}_i(\tau_1)...\hat{Q}_l(\tau_l)|E_0\rangle. \quad (1.14)$$

If the operators Q are the unit operators we can readily see that

$$\langle q', \tau'|q, \tau\rangle \to e^{E_0(\tau'-\tau)}\psi_0(q')\psi_0^*(q). \quad (1.15)$$

Going back to the desired matrix element, we have

$$\frac{\langle q', \tau'|\hat{Q}_1(\tau_1)...\hat{Q}_l(\tau_l)|q, \tau\rangle}{\langle q', \tau'|q, \tau\rangle} \to \langle E_0|\hat{Q}_1(\tau_1)...\hat{Q}_l(\tau_l)|E_0\rangle. \quad (1.16)$$

A path integral expression for the denominator on the right hand side of equation (1.16) is identical to the one computed in the previous section. Returning to equation (1.12), we write out the time dependence explicitly

$$\begin{aligned}\langle q'|e^{-H\tau'}\hat{Q}_1...\hat{Q}_l e^{-H\tau}|q\rangle =& \langle q'|e^{-H(\tau'-\tau_1)}\hat{Q}_1 e^{-H(\tau_1-\tau_2)}\hat{Q}_2 \dots \\ & e^{-H(\tau_{l-1}-\tau_l)}\hat{Q}_l e^{-H(\tau_l-\tau)}|q\rangle.\end{aligned} \quad (1.17)$$

On either side of each $\hat{Q}_i$, we insert a complete set of the operators respective eigenstates. This results in the following

$$\langle q'|e^{-H\tau'}\hat{Q}_1...\hat{Q}_l e^{-H\tau}|q\rangle = \int_q^{q'} \mathcal{D}q q_1(\tau_1)...q_l(\tau_l)e^{-\int_\tau^{\tau'} \mathrm{d}\tau L}. \quad (1.18)$$

The exponential of the action comes from the amplitude we derived before, $\langle q_i, \tau_i|q_j, \tau_j\rangle$. Putting our two results together yields an expression for the vacuum expectation value of our operator,

$$\langle E_0|\hat{Q}_1(\tau)\hat{Q}_l(\tau_l)|E_0\rangle = \frac{\int \mathcal{D}q q_1(\tau_1)...q_l(\tau_l)e^{-\int_\tau^{\tau'} \mathrm{d}\tau L}}{\int \mathcal{D}q e^{-\int_\tau^{\tau'} \mathrm{d}\tau L}} \quad (1.19)$$

From the above we can see a probabilistic interpretation for this expectation value, with the probability being the exponential of the Euclidean action. This is also important for the following sections. The thought experiment described at the start of this chapter also gives a physical motivation for discretizing a system. Of course, discretization is also necessary when dealing with problems numerically.

CHAPTER 2
The Sign Problem

In this chapter we will investigate exactly why standard Monte Carlo techniques cannot be used when there are nonpositive weights.

As we will see in more detail in the next chapter, a Monte Carlo technique is an algorithm which yields a set of system configurations according to a given probability distribution. Once a large number, M, of such configurations are attained, the expectation value of an observable, A, can be approximated by the sample mean

$$\langle A \rangle \approx \frac{1}{M} \sum_{i=1}^{M} A(c_i). \tag{2.1}$$

If the probability distribution, or Boltzmann weight used in the Monte Carlo is positive, then the problem is solvable in polynomial time. In other words if the size of the system (number of particles, for example) is N, then there is a number n such that the time needed to estimate the observable to any desired accuracy is bounded by N^n [7].

If the Boltzmann factor is not positive its interpretation as a probability distribution does not hold. We can instead write the problem in the so called phase quenched theory [4][7]. This is done by writing the Boltzmann factor as $p(c) = |p(c)|s(c)$, where $s(c)$ is the phase. Writing it in this way the expectation value of the observable A is

$$\begin{aligned} \langle A \rangle &= \frac{\sum A(c)p(c)}{\sum p(c)} \\ &= \frac{\sum A(c)s(c)|p(c)|}{\sum s(c)|p(c)|}. \end{aligned} \tag{2.2}$$

Written this way, we can see that the desired quantity can be rewritten in terms of expectation values calculated with the positive probability distribution, $|p(c)|$. If we denote the expectation value with respect to the positive probability distribution with primed angle brackets, then we see that

$$\langle A \rangle = \frac{\langle As \rangle'}{\langle s \rangle'}. \tag{2.3}$$

This is an example of reweighting, a technique which generally involves factoring the probability distribution and distribute factors between the observable and a new probability distribution which is more easily sampled. Using this reweighting, Monte Carlo simulations can be used but the error grows exponentially with the system size, making such simulations infeasible. In order to see the exponential growth of the error, let's examine the expectation value of the sign itself,

$$\begin{aligned} \langle s \rangle &= \frac{\sum s(c)p(c)}{\sum p(c)} \\ &= \frac{\sum s(c)^2|p(c)|}{\sum p(c)} \\ &= \frac{\sum |p(c)|}{\sum p(c)} \\ &= \frac{Z'}{Z}, \end{aligned} \tag{2.4}$$

where Z and Z' are the partition functions of the original system and the reweighted system respectively. The partition function of a system is the exponential of the free energy, so the uncertainty of the measurement is

$$\begin{aligned} \frac{\Delta s}{\langle s \rangle} &= \frac{\sqrt{\langle s^2 \rangle - \langle s \rangle^2}}{\sqrt{M}\langle s \rangle} \\ &\sim \frac{e^{\beta \Delta F}}{\sqrt{M}}, \end{aligned} \tag{2.5}$$

where ΔF is the difference of the free energy and β is the inverse temperature. Since the free energy is an extensive quantity, the fractional error increases

exponentially in the size of the system [7]. Given the large errors inherent in this method, it is not a feasible solution to the sign problem.

The sign problem is not an intrinsic property of a system. If the Hamiltonian of a quantum system were expressed in its diagonal basis, there would not be a sign problem [7]. Of course, diagonalizing Hamiltonians of large systems is also computationally challenging and errors also grow exponentially.

Perhaps the most important appearance of the sign problem is in QCD with a finite chemical potential. In this theory, when the different fermion species are integrated over, a factor of $\det(\not{D} + m + \mu\gamma_0)$ appears, where $\not{D} = \gamma^\nu D_\nu$ and the γ^ν are the Dirac, or gamma matrices. Using the anti-commutativity relations of the gamma matrices we find

$$\gamma_5(\not{D} + m + \mu\gamma_0)\gamma_5 = (\not{D} + m - \mu^*\gamma_0)^\dagger, \tag{2.6}$$

where $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$. Taking the determinant on both sides of the above equation yields $\det(\not{D} + m + \mu\gamma_0) = \det(\not{D} + m - \mu^*\gamma_0)^*$. If μ is zero or purely imaginary, this determinant is positive and there is no sign problem. However, to reproduce expected physics the determinant must be complex and there is a sign problem [3].

CHAPTER 3
The Complex Langevin Equation

As mentioned in the previous sections, systems with sign problems cannot be solved by usual importance sampling techniques such as Metropolis-Hastings algorithm because of the nonpositive measure. For systems without the sign problem, there are algorithms which do not rely on the factor e^{-S} being a probability density. One such algorithm, which will be the focus of this section is the Langevin algorithm. In this section, standard Langevin algorithm and the corresponding Fokker-Planck equation will be studied. An extension of the algorithm to systems with nonpositive weights will be made, along with some formal arguments for the correctness of the complex Langevin algorithm.

3.1 Markov Chains

The goal of Monte Carlo (MC) simulations is to generate configurations of a system according to a given probability distribution. Such MC algorithms are realized by the construction of a Markov chain.

Consider a system with possible states $\{C_i\}$[1] . If this system can change states in a way that only depends on its current configuration, the system and the stochastic process is called a Markov chain. This thesis will deal with

[1] In this section we will consider the set of configurations to be countable, but it is possible to have an uncountable number of configurations. In the uncountable case, the summations should be replaced with integrals where appropriate.

irreducible Markov chains. These are processes which can take the system from one state to any other in finitely many steps

The art of Monte Carlo Markov chain techniques is in defining the right possible states and transition probabilities in order to obtain the correct distribution of final states after enough steps in the stochastic process have taken place. Once a large number, N, of realizations of the Markov chain have occurred, the expectation value of a desired observable $\mathcal{O}$ is simply the arithmetic mean,

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(C(t_i)). \tag{3.1}$$

A sufficient condition for a Markov chain to give configurations with the desired distribution is what is known as the detailed balance equation,

$$e^{-S(C)} P(C \to C') = e^{-S(C')} P(C' \to C), \tag{3.2}$$

where the probability to change states from C to C' is denoted $P(C \to \ C')$ and $e^{-S(C)}$ is the probability distribution that we wish to sample, up to normalization. Let's now prove that any Markov chain which satisfies equation (3.2) will result in configurations which follow the distribution $e^{-S(C)}$.

From this point forward, the state of the system at the i^{th} step in the Markov process will be denoted by C_i, with the initial starting configuration being denoted by C_0. Let $P_N(C)$ be the probability of finding the system in state C after N steps in the Markov chain. The probability $P_N(C)$ is equal to the the sum of probabilities of paths which start at C_0 and end at C and consist of N steps.

$$P_N(C) = \sum_{\{C_j\}_{j=1}^{N-1}} P(C_0 \to C_1) P(C_1 \to C_2) \dots P(C_{N-1} \to C). \tag{3.3}$$

From equation (3.3), an expression for $P_{N+1}(C)$ can be obtained recursively,

$$\begin{aligned}
P_{N+1}(C) &= \sum_{\{C_j\}_{j=1}^{N}} P(C_0 \to C_1)P(C_1 \to C_2)\dots P(C_N \to C) \\
&= \sum_{C_N} \sum_{\{C_j\}_{j=1}^{N-1}} P(C_0 \to C_1)P(C_1 \to C_2)\dots P(C_N \to C) \\
&= \sum_{C_N} \left[\sum_{\{C_j\}_{j=1}^{N-1}} P(C_0 \to C_1)P(C_1 \to C_2)\dots \right. \\
&\qquad \left. \dots P(C_{N-1} \to C) \right] P(C_N \to C) \\
&= \sum_{C_N} P_N(C_N)P(C_N \to C).
\end{aligned} \tag{3.4}$$

In order to determine whether the Markov chain gives configurations with the desired distribution, which is denoted P_{eq}, the variation σ_N is defined,

$$\sigma_N = \sum_C \left| P_N(C) - P_{eq}(C) \right|. \tag{3.5}$$

The variation measures the difference between the target distribution and the distribution at the N^{th} step in the Markov process. For a useful Markov chain, it is expected that this variation decreases as the chain progresses [2]. Indeed,

$$\begin{aligned}
\sigma_{N+1} &= \sum_C \left| P_{N+1}(C) - P_{eq}(C) \right| \\
&= \sum_C \left| \sum_{C'} P_N(C')P(C' \to C) - P_{eq}(C) \right| \\
&= \sum_C \left| \sum_{C'} \left[P_N(C')P(C' \to C) - P_{eq}(C)P(C \to C') \right] \right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_C \sum_{C'} |P_N(C')P(C' \to C) - P_{eq}(C)P(C \to C')| \qquad (3.6)\\
&= \sum_C \sum_{C'} |P_N(C')P(C' \to C) - P_{eq}(C')P(C' \to C)|\\
&= \sum_C \sum_{C'} |P_N(C') - P_{eq}(C')|\, P(C' \to C)\\
&= \sum_{C'} |P_N(C') - P_{eq}(C')|\\
&= \sigma_N.
\end{aligned}
$$

In the third and second to last line, the fact that the transition probabilities are normalized was used,

$$\sum_C P(C \to C') = 1. \qquad (3.7)$$

The detailed balance condition (equation (3.2)) was used to obtain the fifth line in equation (3.6). Since the sequence of σ are bounded from below by 0 and they are decreasing, the limit exists. The difference between σ_N and σ_{N-1} can be made arbitrarily small by choosing a large enough value for N. If we ignore the arbitrarily small error involved in setting $\sigma_N = \sigma_{N+1}$ for large N, and look back on the steps in equation (3.6), we see that the triangle inequality used in the fourth line should be an equality. The triangle inequality is only equality when the vectors $P_N(C)$ and $P_{eq}(C)$ are scalar multiples of one another. However, since both of these are normalized there is no scalar to multiply by and it must be true that $P_N(C) = P_{eq}(C)$.

Any Markov chain construction which satisfies detailed balance equation (3.2) will give configurations according to the desired distribution, provided the irreducibility condition mentioned before is met. The transition probabilities are not unique. The choice of the transition probabilities is the difference

between the various types of Markov chain Monte Carlo algorithms. The next section will detail one such algorithm.

3.2 Langevin Algorithm

The most important Monte Carlo technique for the purposes of this thesis is the Langevin algorithm and its extension to systems which have a sign problem. Consider a system which has an action S which is a real function of the system's coordinate variables q_i. The starting point for this algorithm is in the Langevin equation

$$\frac{\mathrm{d}q_i}{\mathrm{d}\tau} = -\frac{\mathrm{d}S}{\mathrm{d}q_i} + \eta_i(\tau) \tag{3.8}$$

where η is a Gaussian random variable. In this case, the time variable τ labels the evolution of the system as it progresses through the Markov process, and is not the physical time. This variable is called the Langevin time.

In simulations, it is necessary to discretize the Langevin equation. In doing so, we discretize the Langevin time and set the difference between Langevin time points to be ϵ, that is, $\tau_{i+1} - \tau_i = \epsilon$. Equation (3.8) then becomes

$$q_i(\tau_{n+1}) = q_i(\tau_n) + \epsilon\left(-\frac{\mathrm{d}S}{\mathrm{d}q_i(\tau_n)} + \eta_i(\tau_n)\right). \tag{3.9}$$

The random variables η_i is given a Gaussian distribution with variance $\frac{2}{\epsilon}$, that is, the probability of a set of noises $\{\eta_i\}_i$ is given by

$$P(\{\eta_i\}) = \prod_i \sqrt{\frac{\epsilon}{4\pi}} \exp\left(-\frac{\epsilon}{4}\eta_i^2\right) \tag{3.10}$$

For simplicity, we can replace the random variable η with a random variable ξ with variance 1 and mean 0 by making the substitution $\eta = \sqrt{\epsilon/2}\xi$ so that equation (3.9) becomes

$$q_i(\tau_{n+1}) = q_i(\tau_n) - \epsilon\frac{\mathrm{d}S}{\mathrm{d}q_i(\tau_n)} + \sqrt{2\epsilon}\xi_i(\tau_n) \tag{3.11}$$

To confirm that this is a valid MC algorithm, we will check that this satisfies detailed balance. The probability of transitioning from a state q to q' is equal to the probability of the random variable ξ to be equal to $\frac{q'-q}{\sqrt{2\epsilon}} + \sqrt{\frac{\epsilon}{2}} S'(q)$. So the transition probability is

$$P(q \to q') \propto \exp\left\{ -\frac{1}{2}\left(\frac{q'-q}{\sqrt{2\epsilon}} + \sqrt{\frac{\epsilon}{2}} S'(q) \right)^2 \right\} \tag{3.12}$$

If we ignore terms which are at least proportional to ϵ, the ratio of this transition probability and the reverse transition is

$$\frac{P(q \to q')}{P(q' \to q)} = \exp\left\{ -\frac{1}{2}((q'-q)S'(q) - (q-q')S'(q')) \right\} \tag{3.13}$$

Using a Taylor series and again taking the limit where ϵ goes to zero we can say that $S(q') - S(q) = (q'-q)S'(q)$ and we find

$$\frac{P(q \to q')}{P(q' \to q)} = \frac{e^{-S(q')}}{e^{-S(q)}}, \tag{3.14}$$

which is the detailed balance condition. This means that an algorithm which updates as in equation (3.11) will give the desired probability distribution [2].

When we discuss extending this algorithm to complex actions, we will not use detailed balance to justify the validity of the algorithm. Instead, the justification will come directly from the Fokker-Planck equation, which is described in the next section.

3.3 The Fokker-Planck Equation

The result of the Langevin algorithm is a series of configurations which follow a specified probability distribution. The Langevin equation is equivalent to the so-called Fokker-Planck equation, which is essentially an equation for the evolution of the probability distribution. This equation will be of vital

importance when we extend the algorithm to the complex plane to investigate systems which have a sign problem because the underlying probability distribution is not obvious in most cases.

We begin our derivation of the Fokker-Planck equation by considering a smooth function f of the random variables q_i whose dynamics are determined by the Langevin equation. Using equation (3.11), we write

$$f(q(\tau_{n+1})) = f\left(q(\tau_n) - \epsilon\frac{\mathrm{d}S}{\mathrm{d}q} + \sqrt{2\epsilon}\xi(\tau_n)\right). \tag{3.15}$$

For small ϵ we can expand this function,

$$\begin{aligned} f\left(q(\tau_{n+1})\right) &= f(q(\tau_n)) + \frac{\mathrm{d}f}{\mathrm{d}q}(q(\tau_n))\left(-\epsilon\frac{\mathrm{d}S}{\mathrm{d}q} + \sqrt{2\epsilon}\xi(\tau_n)\right) \\ &\qquad + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}q^2}f(q(\tau_n))\left(-\epsilon\frac{\mathrm{d}S}{\mathrm{d}q} + \sqrt{2\epsilon}\xi(\tau_n)\right)^2 + \mathcal{O}(\epsilon^{3/2}) \\ &= f(q(\tau_n)) + \sqrt{2\epsilon}\frac{\mathrm{d}f}{\mathrm{d}q}(q(\tau_n))\xi(\tau_n) \\ &\quad + \epsilon\left(-\frac{\mathrm{d}f}{\mathrm{d}q}(q(\tau_n))\frac{\mathrm{d}S}{\mathrm{d}q}(q(\tau_n)) + \frac{\mathrm{d}^2 f}{\mathrm{d}q^2}f(q(\tau_n))\xi^2(\tau_n)\right) + \mathcal{O}(\epsilon^{3/2}). \end{aligned} \tag{3.16}$$

Rearranging and dividing by ϵ gives that

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}\tau}(q(\tau)) &\approx \frac{f(q(\tau_{n+1})) - f(q(\tau_n))}{\epsilon} \\ &= -\frac{\mathrm{d}f}{\mathrm{d}q}(q(\tau_n))\frac{\mathrm{d}S}{\mathrm{d}q}(q(\tau_n)) + \frac{\mathrm{d}^2 f}{\mathrm{d}q^2}(q(\tau_n))\xi^2(\tau_n) + \sqrt{\frac{2}{\epsilon}}\frac{\mathrm{d}f}{\mathrm{d}q}(q(\tau_n))\xi(\tau_n) \end{aligned} \tag{3.17}$$

Now, we can average over the noise. The last term in equation (3.41) vanishes with the average since ξ follows a Gaussian distribution that has 0 mean. Then we have,

$$\left\langle\frac{\mathrm{d}f}{\mathrm{d}\tau}\right\rangle = -\left\langle\frac{\mathrm{d}f}{\mathrm{d}q}\frac{\mathrm{d}S}{\mathrm{d}q}\right\rangle + \left\langle\frac{\mathrm{d}^2 f}{\mathrm{d}q^2}\right\rangle \tag{3.18}$$

In order to find a differential equation for the underlying probability distribution, $P(q,\tau)$, we may write this equation in integral form,

$$\int \mathrm{d}q\, f(q)\frac{\mathrm{d}P}{\mathrm{d}\tau}(q,\tau) = -\int \mathrm{d}q\, \frac{\mathrm{d}f}{\mathrm{d}q}(q)\frac{\mathrm{d}S}{\mathrm{d}q}(q)P(q,\tau) + \int \mathrm{d}q\, \frac{\mathrm{d}^2 f}{\mathrm{d}q^2}P(q,\tau). \quad (3.19)$$

Assuming integration by parts holds, we arrive at the Fokker-Planck equation,

$$\begin{aligned} \int \mathrm{d}q\, f(q)\frac{\mathrm{d}P}{\mathrm{d}\tau}(q,\tau) &= \int \mathrm{d}q\, f(q)\left[\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{\mathrm{d}S}{\mathrm{d}q}P(q,\tau)\right) + \frac{\mathrm{d}^2 P}{\mathrm{d}q^2}(q,\tau)\right] \\ \implies \frac{\mathrm{d}P}{\mathrm{d}\tau}(q,\tau) &= \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{\mathrm{d}S}{\mathrm{d}q}P(q,\tau)\right) + \frac{\mathrm{d}^2 P}{\mathrm{d}q^2}(q,\tau). \end{aligned} \quad (3.20)$$

The second line in equation (3.20) follows since f was chosen to be arbitrary.

The same general procedure holds in higher dimensions as well, where the Langevin equation is

$$q_i(\tau_{n+1}) = q_i(\tau_n) - \epsilon\partial_i S(q) + \sqrt{2}M_{ij}\xi_i(\tau) \quad (3.21)$$

with M being a constant matrix and ξ being multidimensional Gaussian noise. In this case the Fokker-Planck equation becomes

$$\partial_\tau P(q,\tau) = \partial_i(\partial_i S(q)P(q,\tau)) + \frac{1}{2}M_{ik}M_{jk}\partial_i\partial_j P(q,\tau). \quad (3.22)$$

3.4 The Complex Langevin Equation

As described in a previous section, regular Monte Carlo techniques such as the Langevin algorithm are not effective when the action has a non-zero imaginary part. The complex Langevin algorithm aims to resolve this. The starting point is again the Langevin equation

$$\frac{\mathrm{d}q_i}{\mathrm{d}\tau} = -\frac{\mathrm{d}S}{\mathrm{d}q_i} + \eta(\tau). \quad (3.23)$$

In this case, S is non real [8]. Since the right hand side of equation (3.23) is not purely real, we must include an imaginary part to the fields q_i and the

equation of motion that was the Langevin equation is now two equations; one for each of the real and imaginary parts of the field. If we define $q_i = q_i^R + iq_i^I$ then the equations of motion become

$$\frac{\mathrm{d}q_i^R}{\mathrm{d}\tau} = -\Re\left(\frac{\mathrm{d}S}{\mathrm{d}q_i}\right) + \Re\left(\eta_i(\tau)\right) \tag{3.24}$$

$$\frac{\mathrm{d}q_i^I}{\mathrm{d}\tau} = -\Im\left(\frac{\mathrm{d}S}{\mathrm{d}q_i}\right) + \Im\left(\eta_i(\tau)\right) \tag{3.25}$$

In most cases, the noise will be purely real and the second term on the right hand side of equation (3.25) will be zero. However, non-zero imaginary noise will be explored in a later section as well.

From these Langevin equations, a Fokker-Planck equation can be derived using similar methods as in the previous section. Most importantly, the probability distribution which evolves according to the Fokker-Planck equation shall be a real function. The idea behind the complex Langevin algorithm is that the ensemble average with respect to the complex measure will be the same as the average with respect to the probability distribution given by the Fokker-Planck equation corresponding to the Langevin equations in equations (3.24) and (3.25). If we name the complex weight $\rho(q,t) = e^{-S}$ and the real probability distribution obtained by solving the Fokker-Planck equation $P(q^R, q^I, t)$ then for the complex Langevin equation to be applicable, we would like for the following expressions to be equal [9]

$$\begin{aligned}\langle\mathcal{O}\rangle_\rho &= \frac{\int \mathrm{d}q\, \mathcal{O}(q)\rho(q,t)}{\int \mathrm{d}q\, \rho(q,t)} \\ \langle\mathcal{O}\rangle_P &= \frac{\int \mathrm{d}q^R \mathrm{d}q^I\, \mathcal{O}(q^R + iq^I)P(q^R, q^I, t)}{\int \mathrm{d}q^R \mathrm{d}q^I P(q^R, q^I, t)},\end{aligned} \tag{3.26}$$

where $\mathcal{O}$ is an observable which is a holomorphic function of the field configuration. For the remainder of this thesis we shall assume that our observables

are holomorphic. In the following section, we will look into the conditions for which the above expressions are equal.

3.5 The Applicability of Complex Langevin Algorithm

In this section, we will examine the formal arguments of the convergence of the complex Langevin algorithm to the correct limit. Apart from the issue of whether or not the limit is the correct one, there is also the question of whether or not the algorithm will converge at all, or if it will instead lead to a runaway solution. This second question will be addressed in another section.

In this section, we will consider a slight generalization to the Langevin equations by introducing a complex noise, η_I [9]. The introduction of this extra noise term can be justified if we only consider observables $\mathcal{O}$ which are holomorphic, and therefore the Laplacian , $\Delta\mathcal{O}$ vanishes. In this case, we can add a term $\int \mathrm{d}q^R \mathrm{d}q^I \, n_I \Delta\mathcal{O} P(q^R, q^I)$ to the integral form of the Fokker-Planck equation (equation (3.20)) with no consequence. After doing integration by parts we find that $n_R = 1 + n_I$. The complex noise Langevin equations are

$$\dot{q}^R = \epsilon K_R + \sqrt{n_R}\eta_R \tag{3.27}$$

$$\dot{q}^I = \epsilon K_I + \sqrt{n_I}\eta_I. \tag{3.28}$$

For notational convenience, the symbols K_i have been introduced to represent the derivative drift terms

$$K_R = -\Re\left(\frac{\partial S}{\partial q^R}(q^R + iq^I)\right), \quad K_I = -\Im\left(\frac{\partial S}{\partial q^R}(q^R + iq^I)\right). \tag{3.29}$$

The corresponding Fokker-Planck equation is

$$\partial_t P(q^R, q^I) = L^T P(q^R, q^I), \tag{3.30}$$

where the operator L^T is given by

$$L^T = \nabla_R(n_R\nabla_R - K_R) + \nabla_I(n_I\nabla_I - K_I). \tag{3.31}$$

These equations will determine the evolution of the real probability distribution. The complex measure, ρ, follows the complex Fokker-Planck equation,

$$\frac{\partial}{\partial t}\rho(q,t) = L_0^T\rho(q,t), \tag{3.32}$$

where $L_{y_0}^T$ is the operator

$$L_{y_0}^T = \nabla_x(\nabla_x + \nabla_x S(x+iy_0)). \tag{3.33}$$

In equation (3.32), though we have a Fokker-Planck equation, we cannot use it to create an algorithm to sample ρ directly as is possible in the real case. However, this complex Fokker-Planck equation will still be valuable in the formal justification of the complex Langevin algorithm.

In the long time limit, the stationary solution to equation (3.32) is $\exp(-S(x))$, which is our complex weight. Given these evolution equations for our complex and real distributions, we can determine what condition(s) must be met for the expressions in equation (3.26) to hold in the long time limit.

For as long as we are only considering holomorphic observables and actions, the Cauchy-Riemann equations and use the Langevin operator

$$\tilde{L}^T = \nabla_z(\nabla_z + \nabla_z S(z)). \tag{3.34}$$

This operator is used to introduce time evolution of observables via the equation

$$\frac{\partial}{\partial t}\mathcal{O}(z,t) = \tilde{L}^T\mathcal{O}(z,t), \tag{3.35}$$

where the initial condition is $\mathcal{O}(z,0) = \mathcal{O}(z)$. The formal solution to this equation is

$$\mathcal{O}(z,t) = \exp(t\tilde{L})\mathcal{O}(z), \tag{3.36}$$

with $\tilde{L}^T$ being the adjoint or transpose of $\tilde{L}$ with respect to the bilinear product

$$\langle f, g\rangle = \int f(x,y)g(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{3.37}$$

In order to determine how different our two expectation values in equation (3.26) are, we define a function which interpolates between the two [10]:

$$F(t,\tau) = \int P(x,y,t-\tau)\mathcal{O}(x+iy,\tau)\,\mathrm{d}x\mathrm{d}y, \quad \text{for } 0 \le \tau \le t. \tag{3.38}$$

The function in equation (3.38) is defined in such a way that as we vary the parameter τ from 0 to t, we can go from $\langle\mathcal{O}\rangle_{P(t)}$ to $\langle\mathcal{O}\rangle_{\rho(t)}$. In order to assure that the expectation values coincide at initial time, we impose the initial condition that

$$P(x,y,0) = \rho(x,0)\delta(y-y_0). \tag{3.39}$$

With this initial condition, we can show that $F(t,0) = \langle\mathcal{O}\rangle_{P(t)}$ and $F(t,t) = \langle\mathcal{O}\rangle_{\rho(t)}$. The first equality follows directly from the definition of $\langle\mathcal{O}\rangle_{P(t)}$. To show the second equality, we use the time evolution discussed above;

$$\begin{aligned}
F(t,t) &= \int P(x,y,0)\mathcal{O}(x+iy,t)\,\mathrm{d}x\mathrm{d}y \\
&= \int \rho(x,0)\delta(y)e^{tL}\mathcal{O}(x+iy,0)\,\mathrm{d}x\mathrm{d}y \\
&= \int \rho(x,0)e^{tL_0}\mathcal{O}(x,0)\,\mathrm{d}x \\
&= \int \mathcal{O}(x,0)e^{tL_0^T}\rho(x,0)\,\mathrm{d}x \\
&= \int \mathcal{O}(x,0)\rho(x,t)\,\mathrm{d}x \\
&= \langle\mathcal{O}\rangle_\rho.
\end{aligned} \tag{3.40}$$

To get to the fourth line, we use integration by parts and assume there are no boundary terms. If $F(t,\tau)$ were in fact independent of τ, this would show that the two expectation values are the same at all time. In order to determine the τ dependence, we take the derivative

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}F(t,\tau) = &\int P(x,y,t-\tau)L\mathcal{O}(x+iy,\tau)\,\mathrm{d}x\mathrm{d}y \\ &- \int L^T P(x,y,t-\tau)\mathcal{O}(x+iy,\tau)\,\mathrm{d}x\mathrm{d}y.\end{aligned} \tag{3.41}$$

If integration by parts in both x and y is justified with no boundary terms, then the two terms in equation (3.41) cancel for any imaginary noise. However, if the boundary term does not vanish, there is no longer a formal argument for the correctness of the complex Langevin algorithm.

In [9], it was shown that a weaker condition than requiring equation (3.41) to vanish may be sufficient for the algorithm to hold. Instead of equation (3.41), consider

$$\lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}\tau}F(t,\tau)\bigg|_{\tau=0}. \tag{3.42}$$

In the long t limit, $L^T P(x,y,t)$ goes to 0 by the Fokker-Planck equation since the long time distribution is stationary. Then the expression in equation (3.42) becomes

$$\lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}\tau}F(t,\tau)\bigg|_{\tau=0} = \int P(x,y,\infty)L\mathcal{O}(x+iy)\,\mathrm{d}x\mathrm{d}y. \tag{3.43}$$

To show that equation (3.42) is sufficient, assume that we have found a distribution Q which solves

$$\langle Q, L\mathcal{O}\rangle = \int Q(x,y)L\mathcal{O}(x+iy)\mathrm{d}x\mathrm{d}y = 0 \tag{3.44}$$

for any observable $\mathcal{O}$ in a dense subset of continuous functions from the (real) configuration space. If the inner product defined in equation (3.44) is bounded, that is if there is a constant C such that $\langle Q,\mathcal{O}\rangle \leq C||\mathcal{O}||$, then we can use

the Reisz-Markov representation theorem to find a complex measure μ such that $\langle Q, \mathcal{O} \rangle = \int \mathcal{O}(x)\mu(x)\mathrm{d}x$. With this measure we can conclude that

$$\langle Q, L\mathcal{O} \rangle = \int L\mathcal{O}(x)\mu(x)\mathrm{d}x = 0 \tag{3.45}$$

By using the fact that integration by parts is valid in the real variables, and that the above equation is valid for any choice of observable in our dense set, we find $L_0^T \mu(x) = 0$, which has the (maybe unique) solution e^{-S}. Thus that assumption that equation (3.41) is zero implies that the two expectation values of interest are equal.

The formal arguments for the validity of the algorithm fails in some scenarios. For example, in the case of the $U(1)$-link model, the time dependent operators $\mathcal{O}(x + iy; t)$ grow faster than any exponential as the imaginary part of the field becomes large for $t > 0$. This means that the expression in equation (3.38) is not integrable and is therefore ambiguous [9].

3.6 Imaginary Noise

In this section we will discuss some of the implications of having imaginary noise in the complex Langevin algorithm. First, we should check when the independence of the complex noise, n_I is justified. To do so we will look at the expectation value of a general observable $\langle \mathcal{O} \rangle$, and compute how the quantity depends on n_I by taking the derivative. We use the fact derived earlier, $n_R = 1 + n_I$ to write the operator L as

$$L = L_{n_I=0} + n_I \Delta, \tag{3.46}$$

where Δ is the Laplacian and $L_{n_I=0}$ is the operator L which has $n_I = 0$ and therefore $n_R = 1$. We use the general expression for the derivative of the exponential map,

$$\frac{\partial}{\partial n_I} e^{tL^T} = \int_0^t \mathrm{d}\tau e^{\tau L^T} \Delta e^{(t-\tau)L^T}. \tag{3.47}$$

Taking the derivative gives,

$$\begin{aligned}
\frac{\partial}{\partial n_I}\langle \mathcal{O}(x+iy)\rangle &= \frac{\partial}{\partial n_I}\int P(x,y,t)\mathcal{O}\,\mathrm{d}x\mathrm{d}y \\
&= \frac{\partial}{\partial n_I}\int e^{tL^T} P(x,y,0)\mathcal{O}(x+iy)\,\mathrm{d}x\mathrm{d}y \\
&= \int \mathrm{d}x\mathrm{d}y \int_0^t \mathrm{d}\tau\, e^{\tau L^T}\Delta e^{(t-\tau)L^T} P(x,y,0)\mathcal{O}(x+iy) \\
&= \int \mathrm{d}x\mathrm{d}y \int_0^t \mathrm{d}\tau e^{\tau L^T}\Delta P(x,y,t-\tau)\mathcal{O}(x+iy) \\
&= \int \mathrm{d}x\mathrm{d}y \int_0^t \mathrm{d}\tau P(x,y,t-\tau)\Delta e^{\tau L^T}\mathcal{O}(x+iy) + X \\
&= \int \mathrm{d}x\mathrm{d}y \int_0^t \mathrm{d}\tau P(x,y,t-\tau)\Delta\mathcal{O}(x+iy,t) + X.
\end{aligned} \tag{3.48}$$

The term X in the above calculation represents the possible boundary terms from the integration by parts. The first term in the last line of equation (3.48) will vanish by the Cauchy-Reimann equations since $\mathcal{O}$ was chosen to be holomorphic.

The boundary term will in general not vanish. There is no conclusion we can make in general, as this term may be finite, divergent, or zero. The expected form of the boundary term after carrying out the integral in the real dimensions are expected to involve terms containing combinations of $P\mathcal{O}$, $\mathcal{O}\nabla_y P$, and $P\nabla_y\mathcal{O}$. The question of whether or not the imaginary noise is relevant depends on whether or not the decay of these products causes the integral X to vanish [10].

In practice, determining the effect of the boundary terms is difficult. It requires the introduction of a cut off in the imaginary direction and periodic

boundary conditions. These boundary conditions will invalidate assumptions made about holomorphicity and continuity.

3.7 Runaways

In the previous section, we discussed the criterion for the probability distribution which is obtained from the Fokker-Planck equation to yield the same expectation values as the complex measure. To do this we assumed that the algorithm used would in fact converge. In this section we will address the possibility of the algorithm not converging at all. These cases are called runaways.

Runaways can be predicted, at least in the case of a one dimensional field, by examining the classical flow diagram. For example, the three dimensional XY model is a common system to which complex Langevin methods are applied [11]. The complex action for this system is

$$S = -\beta \sum_x \sum_{\nu=0}^{2} \cos(\phi_x - \phi_{x+\nu} - i\mu\delta_{\nu,0}). \tag{3.49}$$

The flow diagram is constructed by placing arrows which indicate the drift terms, K_x and K_y. The diagram is shown in figure (3–1), taken from [11]. The drifts are represented by arrows which have been normalized. In the XY model, the real part of the field is restricted to a compact domain, $-\pi \leq \phi_R < \pi$. From the diagram, we see that the imaginary part of the field ϕ_I is unbounded, leading to potential runaway solutions. The runaway paths occur at approximately $\phi_R \sim -0.7$ and at $\phi_R \sim 2.4$. When a trajectory comes close to these lines, there is a chance that the imaginary part will go to infinity.

In the diagram, the arrows have been normalized for clarity. In reality, the drift forces grow larger the further the system is from a fixed point, leading to configurations reaching infinity in finite time. One way of dealing with these runaways is to use a small stepsize when discretizing the system. This

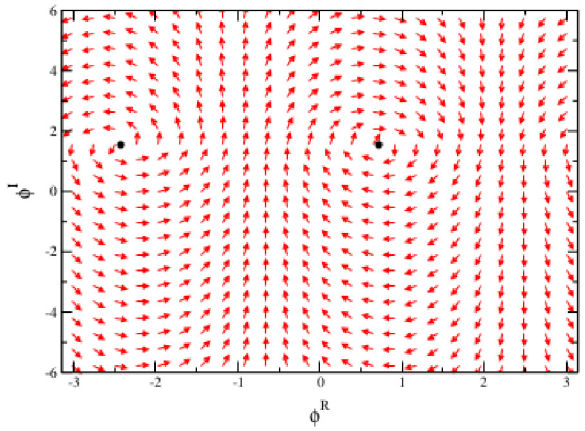

Figure 3–1: The flow diagram of the XY chain. The fixed points are marked by black dots. The runaways appear at $\phi_R \ -0.7$ and ϕ_R 2.4, where the flow becomes vertical.

is sufficient in some cases, but fails in other and can also lead to decreased efficiency of the evolution [11]. To deal with this, we can introduce an adaptive stepsize and take smaller steps in regions where there is a danger of runaways, and larger steps in safe regions.

To regulate the stepsize we would like to keep ϵK_n^{max} bounded, where K_n^{max} is the largest drift vector at the n^{th} time step:

$$K_n^{max} = \max_x \sqrt{K_x^{R2}(n) + K_x^{I2}(n)}. \tag{3.50}$$

At each step we choose the timestep ϵ to be such that

$$\frac{1}{p}\mathcal{K} \leq \epsilon K^{max} \leq p\mathcal{K}. \tag{3.51}$$

The reference value, $\mathcal{K}$ and p are chosen beforehand. If the above inequality is not met for the current value of ϵ, it is changed by a factor of p until the condition is met. This method forces the stepsize to be small as the drift grows larger. It also does not force small steps in regions where the drift is not problematic.

Another technique to solve runaways comes from altering the Langevin equation. By making a clever choice of coordinate transformation, an unstable fixed point which would lead to runaways may become a stable fixed point. In general when undergoing the coordinate change,

$$x \to x(u) \tag{3.52}$$

the Jacobian of this transformation, $J(u) = \frac{\mathrm{d}x(u)}{\mathrm{d}u}$, is absorbed into the action:

$$S(x) \to \tilde{S}(u) = S(x(u)) - \ln(J(u)). \tag{3.53}$$

A simple example of the applicability of changing variables is with a simple Gaussian model with action given by

$$S(x) = \frac{1}{2}\sigma x^2, \qquad \sigma = a + ib. \tag{3.54}$$

For positive a, this action is readily sampled using complex Langevin. However, when $a < 0$, the fixed point at 0 becomes repulsive. By choosing variables $x(u) = u^3$ for example, the system becomes stable. The flow diagrams for each choice of coordinates is shown in Fig. 3–2[12]. With this coordinate transformation, the Langevin algorithm gives the correct result for moments, for example

$$\langle u^6 \rangle = \langle x^2 \rangle = \frac{1}{\sigma}, \tag{3.55}$$

which is the analytic continuation of the case of real a. The change of variable techniques can be used to solve integrals of this type. The same transformation as above can be used for a complex quartic integral, $S(x) = i\frac{\lambda}{4!}x^4$.[12]

The addition of an extra term in the effective action makes this coordinate change resemble reweighting, which we will talk about in more detail further on. However, it is not our desired method to alter the probability distribution. Although coordinate transformations may lead to stable evolution, it

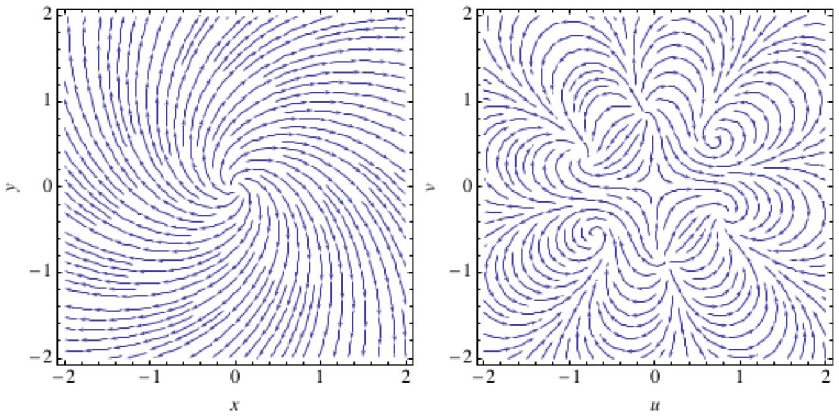

Figure 3–2: Flow diagram of the Gaussian model with $\sigma = -1 + i$. The left shows the real and imaginary axes of the unchanged coordinate description including the repulsive fixed point at the origin. The right diagram is the same system under the transformation $x \to u^3$.

amounts to a change of the complex weight, whereas our goal is to reweight the corresponding real probability distribution.

CHAPTER 4
Higher Order Algorithm

In this chapter we will look at an improvement to the Langevin algorithm which is developed in [13]. When we verified that detailed balance was satisfied for the Langevin equation, we were required to take the limit of vanishing stepsize. Taking this limit introduces a systematic error, since in practice the stepsize must be finite. The simplest way to use the Langevin equation to generate a new configuration for the system is

$$\phi(n+1) = \phi(n) + \epsilon K(\phi(n)) + \sqrt{2\epsilon}\eta(n). \tag{4.1}$$

This is the lowest order algorithm and carries a systematic error which is of order ϵ. In order to reduce this error, the noise must be changed.

A higher order algorithm is proposed. The algorithm is constructed by first defining the following variables [13]

$$\begin{aligned} \psi(n) &= \phi(n) + \frac{1}{2}\epsilon K(\phi) \\ \tilde{\psi}(n) &= \phi(n) + \frac{1}{2}\epsilon K(\phi) + \frac{3}{2}\sqrt{2\epsilon}\tilde{\alpha}(n) \\ \tilde{\alpha}(n) &= \frac{1}{2}\alpha(n) + \frac{\sqrt{3}}{2}\xi(n). \end{aligned} \tag{4.2}$$

Each of α and ξ are Gaussian random variables with zero mean and a variance of 1. The update step in the algorithm is as follows

$$\phi(n+1) = \phi(n) + \frac{\epsilon}{3}(K(\psi) + 2K(\tilde{\psi})) + \sqrt{2\epsilon}\alpha. \tag{4.3}$$

For a system which is one dimensional the error in the algorithm is of order ϵ^2. For a coupled system, the error is of order $\epsilon^{3/2}$[14].

Let's check the stepsize errors analytically. For simplicity, we will take a one dimensional system with a linear drift, $K(\phi) = -\omega\phi$. With this drift the field which solves the discretized Langevin equation is $\phi(n) = \sqrt{2\epsilon}\sum_{i=0}^{n-1}(1-\epsilon\omega)^{n-1-i}\eta(i)$. Inserting this in the right hand side of equation (4.1) will confirm that this is in fact a solution:

$$\begin{aligned}
\phi_(n) - \epsilon\omega\phi(n) + \sqrt{2\epsilon}\eta(n) &= (1-\epsilon\omega)\phi(n) + \sqrt{2\epsilon}\eta(n) \\
&= (1-\epsilon\omega)\sqrt{2\epsilon}\sum_{i=0}^{n-1}(1-\epsilon\omega)^{n-1-i}\eta(i) \\
&\quad + \sqrt{2\epsilon}(1-\epsilon\omega)^0\eta(n) \\
&= \sqrt{2\epsilon}\sum_{i=0}^{n}(1-\epsilon\omega)^{n-i}\eta(i) \\
&= \phi(n+1).
\end{aligned} \tag{4.4}$$

We will examine the expectation value $\langle\phi^2\rangle = \lim_{n\to\infty}\langle\phi_n\phi_n\rangle$.

$$\begin{aligned}
\lim_{n\to\infty}\langle\phi(n)\phi(n)\rangle &= \lim_{n\to\infty} 2\epsilon\left\langle\left(\sum_{i=0}^{n-1}(1-\epsilon\omega)^{n-1-i}\eta(i)\right)^2\right\rangle \\
&= \lim_{n\to\infty} 2\epsilon\left(\sum_{i=0}^{n-1}(1-\epsilon\omega)^{2n-2-2i}\langle\eta(i)\eta(i)\rangle\right. \\
&\quad \left. + \sum_{i\neq j}(1-\epsilon\omega)^{n-1-j}(1-\epsilon\omega)^{n-1-i}\langle\eta(j)\eta(i)\rangle\right)
\end{aligned} \tag{4.5}$$

The second term vanishes because the noise at each time step are not correlated. We are left with a geometric sum

$$\begin{aligned}
\lim_{n\to\infty}\langle\phi(n)\phi(n)\rangle &= \lim_{n\to\infty} 2\epsilon(1-\epsilon\omega)^{2n-2}\sum_{i=0}^{n-1}(1-\epsilon\omega)^{-2i} \\
&= \lim_{n\to\infty} 2\epsilon(1-\epsilon\omega)^{2n-2}\frac{1-(1-\epsilon\omega)^{-2n}}{1-(1-\epsilon\omega)^{-2}}.
\end{aligned} \tag{4.6}$$

After multiplying through and taking the limit, we find

$$\begin{aligned}\lim_{n\to\infty}\langle\phi(n)\phi(n)\rangle &= \frac{1}{\omega}\frac{1}{1-\frac{\epsilon\omega}{2}} \\ &= \frac{1}{\omega}\left(1-\frac{\epsilon\omega}{2}+\mathcal{O}(\epsilon^2)\right).\end{aligned} \tag{4.7}$$

This shows that the obvious way to implement Langevin updates have an error of order ϵ. To arrive at the algorithm asserted in equation (4.2), we propose the following

$$\begin{aligned}\psi(n) &= \phi(n)+\frac{1}{2}\epsilon K(\phi(n))+k\sqrt{2\epsilon}\tilde{\alpha}(n) \\ \tilde{\psi(n)} &= \phi(n)\frac{1}{2}\epsilon K(\phi(n))+l\sqrt{2\epsilon}\tilde{\alpha}(n) \\ \phi(n+1) &= \phi(n)+\epsilon[aK(\psi(n))+bK(\tilde{\psi(n)})]+\sqrt{2\epsilon}\alpha(n),\end{aligned} \tag{4.8}$$

The values of a, b, l, and k are to be determined, and $\tilde{\alpha}$ is as defined above. After making the same simplifying assumptions on K and rearranging terms, the last line in equation (4.8) becomes

$$\phi(n+1) = \phi(n)-\epsilon\tilde{\omega}\phi(n)+\sqrt{2\epsilon}\tilde{\eta}(n), \tag{4.9}$$

where

$$\begin{aligned}\tilde{\omega} &= \omega(a+b)\left(1-\frac{1}{2}\epsilon\omega\right) \\ \tilde{\eta}(n) &= \alpha(n)-\omega\epsilon(ak+bl)\tilde{\alpha}(n).\end{aligned} \tag{4.10}$$

Equation (4.9) is identical to our simple case we dealt with before. The steps taken in that case can be used here as well. The result is

$$\langle\phi^2\rangle = \frac{1}{\tilde{\omega}}\frac{\langle\tilde{\eta}^2\rangle}{1-\frac{\epsilon\tilde{\omega}}{2}}. \tag{4.11}$$

Once we expand the expectation value and Taylor expand as before, we get

$$\langle \phi^2 \rangle = \frac{1}{\omega} \left(\frac{1}{a+b} + \frac{1+a+b-2(ak+bl)}{2(a+b)} \epsilon\omega + \mathcal{O}(\epsilon^2) \right). \tag{4.12}$$

To achieve the correct result, we require that $a+b=1$. To have the linear ϵ term vanish, we have $ak+bl=1$. By considering a more complicated drift than our linear example, a further condition can be derived, $ak^2+b^2=\frac{3}{2}$ [14]. The values used in equation (4.2) satisfy these equations. They are

$$a = \frac{1}{3,} \quad b = \frac{2}{3}, \quad k = 0, \quad l = \frac{3}{2}. \tag{4.13}$$

Here we see that this algorithm has smaller errors associated with the step size is smaller than for the simper algorithm. For higher dimensional systems, the calculations are more complicated, but similar. In that case, the result is an error of order $\mathcal{O}(\epsilon^{3/2})$.

CHAPTER 5
Reweighting

5.1 Reweighting

As we saw in the previous chapter, the usefulness of the complex Langevin algorithm depends on the decay of the real probability distribution P in the imaginary direction. In order to study the decay, we wish to make the behaviour of the probability distribution in this region more apparent. To do this, we will try to use the technique of reweighting.

Reweighting was briefly mentioned in a previous chapter, when we discussed the sign problem. This method, which we will review shortly, is useful when the probability distribution has features which may be hard to detect directly. In this case, since we wish to study the falloff of the probability distribution we are interested in a region which is unlikely to be sampled without some sort of reweighting.

A simple one-dimensional application of this technique which illustrates its usefulness is when the sample probability, P has two peaks which are widely separated in configuration space, as is sometimes the case when dealing with phase transitions. If the probability distribution is near zero between the two peaks, it is unlikely that a Monte Carlo simulation will cross this trough if it only takes small steps. This means that once the system reaches one local maximum of the distribution, it will probably not leave and will fail to sample the entire relevant space of configurations. In order to explore past these unlikely regions, we may chose a reweight function, W, such that the reweighted probability WP no longer has peaks which are very isolated. For example, W can be chosen such that instead of isolated peaks, WP has a

much shallower valley between the two maxima. This allows a Monte Carlo sampling algorithm to explore the entire configuration space [15].

When reweighting, the observable must be altered as well to offset the fact that the probability distribution being sampled has been reweighted. Our expectation value in terms of the reweighted probability is

$$\begin{aligned}\langle \mathcal{O} \rangle &= \frac{\int \mathrm{d}x P(x)\mathcal{O}(x)}{\int \mathrm{d}x P(x)} \\ &= \frac{\int \mathrm{d}x P(x)W(x)W^{-1}(x)\mathcal{O}(x)}{\int \mathrm{d}x W(x)W^{-1}(x)P(x)} \\ &= \frac{\langle W^{-1}\mathcal{O}\rangle_W}{\langle W^{-1}\rangle_W}.\end{aligned} \tag{5.1}$$

where $\langle\cdot\rangle_W$ denotes an expectation value with respect to the reweighted probability distribution. In simulations this translates to

$$\langle \mathcal{O} \rangle \approx \frac{\sum_{i=1}^{M} \mathcal{O}(x_i)W^{-1}(x_i)}{\sum_{i=1}^{M} W^{-1}(x_i)}. \tag{5.2}$$

In the case of the complex Langevin algorithm, the target distribution is not known, making it difficult to reweight directly, by changing the drift terms for example. Furthermore, we shall see that since the Langevin algorithm does not maintain detailed balance, techniques which can be used to reweight in usual algorithms may not apply here.

5.2 Lattice Gauge Theory

In our attempt to combine the concepts of reweighting and complex Langevin we will use a simple one dimensional scalar field with a gauge symmetry as our toy model. Before we try to use reweighting, we should review the process for discretizing a theory with an abelian gauge symmetry.

Suppose we wish to create a Lagrangian which is invariant under the action of some group G. In the simple example of a complex scalar field, the group G is the abelian group $U(1)$. In this example, we are to construct a

Lagrangian which is invariant under the transformation

$$\psi(x) \to e^{i\alpha(x)}\psi, \tag{5.3}$$

where ψ is our field variable. The physical reasoning behind this symmetry is simply that our Lagrangian should be unchanged when our field is multiplied by any arbitrary phase, even if the phase is spacetime dependent. Now we can ask what terms can be included in our Lagrangian. Clearly, the combination of ψ with its conjugate is allowed since it is invariant under the $U(1)$ transformation. Including a derivative term proves to be more complicated.

For example, recall the limit definition of the derivative:

$$n^\mu \partial_\mu \psi(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(\psi(x + \epsilon n) - \psi(x)). \tag{5.4}$$

Note that each term in the limit may pick up a different phase under the gauge transformation, causing the transformation law of the derivative to be complicated. The field can pick up a different factor at neighbouring points in spacetime, making an interpretation of this difference unclear. The derivative must be altered so that both terms in the difference have the same transformation rules.

To reconcile the need for a derivative and gauge invariance, we define the covariant derivative [16]

$$n^\mu D_\mu \psi = \lim_{\epsilon \to 0}(\psi(x + \epsilon n) - U(x + \epsilon n, x)\psi(x)). \tag{5.5}$$

The comparator $U(x, y)$ is introduced as a pure phase with the transformation law

$$U(x, y) \to e^{i\alpha(x)} U(x, y) e^{-i\alpha(y)} \tag{5.6}$$

With this definition, the quantities $U(x+\epsilon n, x)\psi(x)$ and $\psi(x+\epsilon n)$ transform in the same way and so the covariant derivative is a sensible choice for a derivative replacement .

Since we have defined $U(x, y)$ as a phase, we can Taylor expand the comparator in equation (5.5)

$$U(x+\epsilon n, x) = 1 - i\epsilon n^{\mu} A_{\mu}(x) + \mathcal{O}(\epsilon^2) \tag{5.7}$$

The coefficient A_{μ} is the connection. By imposing the transformation for $U(x, y)$, we find that under a gauge transformation, the connection must transform as

$$A_{\mu}(x) \to A_{\mu}(x) - \partial_{\mu}\alpha(x), \tag{5.8}$$

and the covariant derivative is simplified as

$$D_{\mu}\psi(x) = \partial_{\mu}\psi(x) + iA_{\mu}(x)\psi(x) \tag{5.9}$$

The covariant derivative now transforms the same way that ψ does. Thus contracted combination of covariant derivatives will be a valid term in the Lagrangian. Writing the comparator as a pure phase again, we have the approximation

$$U(x+\epsilon n, x) = e^{-i\epsilon \cdot A} \tag{5.10}$$

When we discretize an action, the same logic will be used; gauge invariance should be imposed at every step in the process.

In this thesis we will be using a one dimensional Bose gas, which is very similar to a complex scalar field. The Lagrangian for this system is

$$\mathcal{L} = \partial_{\mu}\phi\partial^{\mu}\phi^* - m^2|\phi|^2 \tag{5.11}$$

So, the naive discretization of the action would then be

$$\int \mathrm{d}\tau(\partial_\tau\phi\partial_\tau\phi^* + m^2|\phi|^2) \rightarrow$$
$$x\sum_x[(2+m^2)\phi(x)\phi^*(x) - (\phi^*(x)\phi(x+\hat{e}) + \phi(x)\phi^*(x+\hat{e}))] \tag{5.12}$$

Clearly, however, this suffers from the same problem as the derivative term from above, namely the terms which multiply ϕ and ϕ^* at two different points are not gauge invariant. In order to amend this, we add a comparator $U_{x,x+\hat{e}}$. The comparator is the same as in the continuous case, but we must think about how the gauge field changes in a Wick rotation. Recall how the time variable and temporal derivatives are transformed:

$$\begin{aligned} x_0 &\rightarrow -ix_4 \\ \partial_0 &\rightarrow i\partial_4 \end{aligned} \tag{5.13}$$

In order for the gauge transformation to be consistent with the transformation of the derivative, i.e. $A_0 \rightarrow A_0 - \partial_0\alpha$ the Wick rotation of the gauge field must be

$$A_0 \rightarrow iA_4. \tag{5.14}$$

So, the comparator changes into,

$$U(x+n,x) = e^{-iA_0} \rightarrow e^{A_4} := e^{\mu}. \tag{5.15}$$

The field μ will be called the chemical potential. To understand how the gauge field and the usual notion of a chemical potential are related, we look at the conserved current which arises from the phase invariance of the Lagrangian. Under the phase transformation, the Lagrangian changes to

$$\mathcal{L} \rightarrow \mathcal{L}' = \partial_\mu\phi\partial^\mu\phi^* + i\partial_\mu\alpha[\phi\partial^\mu\phi^* - \phi^*\partial^\mu\phi] + \partial_\mu\alpha\partial^\mu\alpha|\phi|^2 - m^2|\phi|^2. \tag{5.16}$$

Using the Euler-Lagrange equations gives an equation of motion for α.

$$\partial^\mu \left(\frac{\partial \mathcal{L}'}{\partial(\partial^\mu \alpha)} \right) = \frac{\partial \mathcal{L}'}{\partial \alpha} \tag{5.17}$$

The right hand side of equation (5.17) is zero. By choosing α as a constant we find that the conserved current density is

$$j^\mu = i[\phi \partial^\mu \phi^* - \phi^* \partial^\mu \phi] \tag{5.18}$$

Recall the partition function from a previous section came from the amplitude $\int \mathrm{d}\phi \langle \phi | e^{-\beta H} | \phi \rangle$. When there is a conserved charge, the Hamiltonian is replaced with $H + \mu Q$, where Q is the conserved current expressed as a function of the field and the conjugate momentum and μ is the chemical potential. The appearance of a chemical potential is the same as in statistical mechanics. In our case, the partition function translates to [17]

$$\begin{aligned} Z = \int \mathcal{D}[\phi_1, \phi_2, \pi_1, \pi_1] \exp \Bigg\{ \int \mathrm{d}\tau \, \Big(& i\pi_i \dot{\phi}_i \\ & - \frac{1}{2}(\pi_i \pi_i + m^2 \phi_i \phi_i + \mu(\phi_1 \pi_2 - \pi_1 \phi_2) \Big) \Bigg\}. \end{aligned} \tag{5.19}$$

The fields ϕ_1 and ϕ_2 are the real and imaginary parts of ϕ,

$$\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}}, \tag{5.20}$$

and π_i are there respective canonical momenta.

After completing the square and integrating out the momenta in the usual way, the partition function simplifies to

$$\int \int \mathcal{D}[\phi_1, \phi_2] \exp \left\{ - \int \mathrm{d}\tau \, (\partial_\tau - \mu)\phi(\partial + \mu)\phi^* + m^2 |\phi|^2) \right\} \tag{5.21}$$

We can see that the derivative terms look like covariant derivatives with gauge field μ. This is the reason why we called our gauge field a chemical potential above.

5.3 Attempts at Reweighting a Complex Action

In this section, we will look at a few techniques which we attempted to use to reweight a distribution that was attained via the complex Langevin algorithm.

When we reweight in a standard Langevin algorithm, we can simply change the drift forces and know exactly how the resulting distribution will differ from the original. In the complex case, we do not actually know the target distribution so changing the drift forces will lead to hard to predict changes in the resulting sample set. To overcome this, we attempted to reweight by adding an accept reject step to the algorithm and by altering the Fokker-Planck equation.

5.3.1 Metropolis

The first technique we will use to attempt reweighting is by combining it with the Metropolis algorithm. The Metropolis algorithm is one of the most popular MCMC because of its simplicity to implement [18][2].

The first step in Metropolis is to propose an update for the system. It is common to propose an update in a symmetric way. For example, the proposed state may be drawn from a normal distribution centred at the current state.

The second step imposes detailed balance. Once a proposed state x' is obtained, it is accepted as the new configuration for the system with a probability $\min\{1, e^{-S(x')+S(x)}\}$. The proposal is always accepted if the proposal is more likely than the current configuration. The probability of the system exploring less probable configurations is chosen to ensure detailed balance. To check that detailed balance is maintained, assume without loss of generality that $e^{-S(x')} < e^{S(x)}$. Then the probability of making the transition from a state x to x' is

$$P(x \to x') = P_0(x \to x')\frac{e^{-S(x')}}{e^{-S(x)}}, \tag{5.22}$$

and the reverse transition probability is

$$P(x' \to x) = P_0(x' \to x), \tag{5.23}$$

where P_0 is the probability of proposing the new state. We chose this to be symmetric and so the function P_0 in equations (5.22) and (5.23) are the same, and so these two equations give use detailed balance.

We can use a Metropolis step to reweight any sample obtained by any algorithm which maintains detailed balance [19]. For example, assume we have sampled a probability distribution, $P(x)$ by Langevin algorithm. Then we have detailed balance,

$$P(x)P(x \to x') = P(x')P(x \to x'). \tag{5.24}$$

If we wish to sample a distribution $P(x)W(x)$, then simply including an acceptance rate at the end of each step with the form $\min\{1, W(x')/W(x)\}$ then we see that our algorithm maintains detailed balance for the distribution $P(x)W(x)$ as desired. This makes adding a Metropolis step a valid method of reweighting.

To test if the Metropolis step is a valid way to reweight a sample attained from the complex Langevin algorithm, we looked at some simple systems which are solvable analytically, such as the complex scalar field described in a previous section, or the simpler case of a quadratic action with complex coefficients [4]. The complex Langevin algorithm is known to work in these examples. The algorithm was used as described in a previous section, but with the additional accept reject which was the ratio of our reweight function, W at the proposed and current configuration.

Since the distribution attained from complex Langevin steps is not known analytically, to test the Metropolis reweighting we compared the observables

$$\frac{\langle \mathcal{O} W^{-1} \rangle_{RW}}{\langle W^{-1} \rangle_{RW}} \qquad \texttt{and} \qquad \langle \mathcal{O} \rangle$$

where the subscript on the angle brackets indicate the sample mean after doing the Metropolis step. If these two calculations are equal, then the reweighting will have been a success.

In simulations, we have found that the Metropolis step addition is not a successful way to reweight. The reweight functions used were simple Gaussians, and the observables that we attempted to calculate were the expectation values of various powers of the field variables of the complex Bose gas.

The failure of this technique should not be a surprise. We motivated it by showing that it was valid when we had detailed balance to begin with. The complex Langevin algorithm never claims to maintain detailed balance, it only uses the Fokker-Planck equation and Ito calculus. The failure of this method is just further evidence that complex Langevin doesn't hold detailed balance.

5.3.2 Altering the Fokker-Planck

Since we cannot use detailed balance to reweight, we will use the Fokker-Planck equation directly. All we know about the target distribution is that is solves the Fokker-Planck equation, we can try use that to determine a differential equation which is solved by the reweighted distribution. We will then attempt to use this differential equation to directly sample the reweighted distribution.

First we will examine the method of using the Fokker-Planck equation to directly sample a distribution. This is equivalent to using the Langevin equation but we hope that it allows for easier manipulation of the probability distribution.

The starting point is the master equation

$$P(\mathbf{x}, t+dt) = \sum_y P(\mathbf{y} \to \mathbf{x}) P(\mathbf{y}, t). \tag{5.25}$$

Each term on the right hand side represents the probability of going from state $\mathbf{y}$ to state $\mathbf{x}$ in a time window dt. By discretizing the Fokker-Planck equation using finite differences and comparing it to the master equation, we can identify the necessary transition probabilities. The transition probabilities are

$$\begin{aligned} P((x+dx, y) \to (x, y)) = P_{x,-} &= \frac{dt\, n_R}{dx^2} \left(1 - K_x \frac{dx}{2n_R}\right) \\ P((x-dx, y) \to (x, y)) = P_{x,+} &= \frac{dt\, n_R}{dx^2} \left(1 + K_x \frac{dx}{2n_R}\right), \end{aligned} \tag{5.26}$$

and the probability for no transition in the real coordinate is

$$P_{x,0} = \left(1 - \frac{2n_R dt}{dx^2}\right). \tag{5.27}$$

The transition probabilities are the same for the imaginary coordinate, but with K_x replaced with K_y and the real noise replaced by imaginary noise, n_I. At each step in the algorithm, the real and imaginary part of each field is changed by $dx = \sqrt{2\, dt\, n_R}$ and $dy = \sqrt{2\, dt\, n_I}$ respectively according to the above probabilities. With these choices of stepsizes, the probability of having the system not change vanishes.

We have reviewed a method for taking a differential equation for a probability distribution and turning it into an algorithm to generate a sample set. Now, we would like to derive an equation which is solved by the reweighted distribution $R = PW$, given that P solves the Fokker-Planck equation. We take our reweight function $W(\mathbf{x})$ to be independent of time, since we wish to reweight the stationary distribution anyway. We also take W to have the form e^{-M}, where M is a real differentiable function. Both sides of the Fokker-Planck

equation are multiplied by W. After rearranging terms and using product rule, we reach the following differential equation

$$\begin{aligned}\partial_t R = n_R \partial_x^2 R - \partial_x((K_x - 2n_R M_x)R) + (n_R M_x^2 - n_R M_{xx} - M_x K_x)R \\ + (n_R \leftrightarrow n_I, x \leftrightarrow y).\end{aligned} \tag{5.28}$$

Once the equation is discretized, we may read off the coefficients and attempt to interpret them as transition probabilities. Unfortunately, the coefficients are not normalized, nor are they necessarily non-negative. For example, the sum of all the coefficients gives

$$1 + n_R \mathrm{d}t(M_x^2 - M_{xx} - M_x K_x) + n_I \mathrm{d}t(M_y^2 - M_{yy} - M_y K_y). \tag{5.29}$$

There is no choice of the noise or timestep which allows the second and third terms in parentheses to cancel out. There is also no way to guarantee that they will be positive.

We have derived a differential equation followed by the reweighted distribution we wish to sample from, however, we cannot come up with a master equation with which to sample it.

CHAPTER 6
Conclusion

In this final chapter we will summarize the complex Langevin algorithm and mention a couple of alternatives for evading the sign problem.

The path integral formulation opens up problems in quantum field theory to many possible numerical solutions, including the Metropolis and Langevin algorithms. There is a large class of problems, including QCD with finite chemical potential, which cannot be solved by simple application of these algorithms because of the sign problem.

In an attempt to evade the sign problem, the complex Langevin algorithm was developed. We looked at how the algorithm can be applied and how formal arguments for validity may fail depending on the fall off of the probability distribution and the observable.

In order to explore the fall off we attempted to reweight the probability distribution. To accomplish this, we attempted to combine the Langevin equation with the Metropolis algorithm. This method was not effective, as the complex Langevin equation lacked detailed balance. Altering the Fokker-Planck equation into a differential equation for the reweighted distribution also did not lead to anything which was easily simulated.

In addition to what was explored in this thesis, there are a number of alterations that can be made to the standard real Langevin equation which do away with detailed balance altogether [20]. Though these methods look different from the complex Langevin equation, learning to manipulate these may give some insight on how to manipulate to complex case.

A promising alternative solution to the complex Langevin equation is the method of Lefschetz thimbles [21] [22] . In this method, the contour of integration is changed to the path of steepest descent, on which the imaginary part of the action is constant. The path of steepest descent gives and equation which is conjugate to the complex Langevin equation. This method may introduce new sign problems, but these may be more mild than the original theory's sign problem.

REFERENCES

[1] Richard P Feynman; Albert R Hibbs; Daniel F Styer. *Quantum Mechanics and Path Integrals*. Dover books on physics. Dover Publications, emended ed edition, 2010.

[2] H. J. Rothe. Lattice gauge theories: An Introduction. *World Sci. Lect. Notes Phys.*, 43:1–381, 1992. [World Sci. Lect. Notes Phys.82,1(2012)].

[3] P. de Forcrand. Simulating QCD at finite density, May 2010. [arXiv:1005.0539[hep-lat]].

[4] Gert Aarts. Can stochastic quantization evade the sign problem? the relativistic bose gas at finite chemical potential. *Phys. Rev. Lett.*, 102:131601, Apr 2009.

[5] Gert Aarts and FrankA. James. On the convergence of complex langevin dynamics: the three-dimensional xy model at finite chemical potential. *Journal of High Energy Physics*, 2010(8), 2010.

[6] John S. Townsend. *A Modern Approach to Quantum Physics*. University Science Books, 2 edition, 2012.

[7] Matthias Troyer and Uwe-Jens Wiese. Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations. *Phys. Rev. Lett.*, 94:170201, May 2005.

[8] Giorgio Parisi. On complex probabilities. *Physics Letters B*, 131(46):393 – 395, 1983.

[9] Gert Aarts, FrankA. James, Erhard Seiler, and Ion-Olimpiu Stamatescu. Complex langevin: etiology and diagnostics of its main problem. *The European Physical Journal C*, 71(10), 2011.

[10] Gert Aarts, Erhard Seiler, and Ion-Olimpiu Stamatescu. Complex langevin method: When can it be trusted? *Phys. Rev. D*, 81:054508, Mar 2010.

[11] Gert Aarts, Frank A. James, Erhard Seiler, and Ion-Olimpiu Stamatescu. Adaptive stepsize and instabilities in complex langevin dynamics. *Physics Letters B*, 687(23):154 – 159, 2010.

[12] Gert Aarts, FrankA. James, JanM. Pawlowski, Erhard Seiler, Dnes Sexty, and Ion-Olimpiu Stamatescu. Stability of complex langevin dynamics in effective models. *Journal of High Energy Physics*, 2013(3), 2013.

[13] Gert Aarts and Frank A. James. Complex Langevin dynamics in the SU(3) spin model at nonzero chemical potential revisited. *JHEP*, 01:118, 2012.

[14] Chien Cheng Chang. Numerical solution of stochastic differential equations with constant diffusion coefficients. *Mathematics of computation*, 49(180):523–542, 1987.

[15] Bernd A. Berg and Thomas Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267(2):249 – 253, 1991.

[16] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[17] Charles Gale and Joseph I. Kapusta. *Finite-Temperature Field Theory Principles and Applications*. Cambridge University Press, 2 edition, 2006.

[18] W. K. HASTINGS. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[19] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[20] Masayuki Ohzeki and Akihisa Ichiki. Mathematical understanding of detailed balance condition violation and its application to langevin dynamics. *Journal of Physics: Conference Series*, 638(1):012003, 2015.

[21] Marco Cristoforetti, Francesco Di Renzo, and Luigi Scorzato. New approach to the sign problem in quantum field theories: High density qcd on a lefschetz thimble. *Phys. Rev. D*, 86:074506, Oct 2012.

[22] Gert Aarts, Lorenzo Bongiovanni, Erhard Seiler, and Dénes Sexty. Some remarks on lefschetz thimbles and complex langevin dynamics. *Journal of High Energy Physics*, 2014(10):159, 2014.