Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Manuel Scherzer

born in: Marburg-Wehrda, Germany
Oral examination: June 4th, 2019

Phase transitions in lattice gauge theories: From the numerical sign problem to machine learning

Referees: Prof. Dr. Ion-Olimpiu Stamatescu
Prof. Dr. Tilman Plehn

Phasenübergänge in Gittereichtheorien: Vom numerischen Vorzeichenproblem zu maschinellem Lernen

Gittersimulationen der Quantenchromodynamik (QCD) sind ein wichtiges Werkzeug der modernen Quantenfeldtheorie. Derartige Simulationen der grundlegenden Theorie ergeben höchst präzise Ergebnisse und ermöglichen somit den Vergleich von Theorie und Experiment. Bei endlicher baryonischer Dichte sind solche Simulationen nicht mehr möglich. Dies liegt am sogenannten numerischen Vorzeichenproblem, welches auftritt, wenn die Wirkung der Theorie komplexwertig wird und somit zu Integralen über stark oszillierenden Funktionen führt. Wir untersuchen zwei Ansätze dieses Problem zu lösen. Zunächst verwenden wir die komplexe Langevin Methode, ein komplexifizierter stochastischer Prozess, und untersuchen ihre Eigenschaften. Wir wenden diese Methode auf QCD an und simulieren eine Parameterregion in der andere Methoden unzuverlässig sind, wir simulieren Parameter bis $\mu/T_c \approx 5$. Wir untersuchen schlussendlich die Anwendbarkeit der Methode für SU(2) Realzeitsimulationen. Weiterhin untersuchen wir die Methode der sogenannten Lefschetz Thimbles, welche das Vorzeichenproblem durch eine Deformierung der Integrationsmannigfaltigkeit löst, sodass keine Oszillationen mehr auftreten. Wir diskutieren Aspekte der Methode anhand einfacher Modelle und entwickeln Algorithmen für höhere Dimensionen.
Zum Abschluss wenden wir neuronale Netze auf Daten aus Gittersimulationen and und benutzen diese um den Ordnungsparameter des Phasenübergangs im Ising Modell und in SU(2) Eichtheorie zu finden. Wir entschlüsseln somit die Grösse, die das neuronale Netz lernt.

Phase transitions in lattice gauge theories: From the numerical sign problem to machine learning

Lattice simulations of Quantum chromodynamics (QCD) are an important tool of modern quantum field theory. They provide high precision results from first principle computations and as such allow for comparison between experiment and theory. At finite baryon density, such simulations are no longer possible due to the numerical sign problem which occurs when the action of the theory becomes complex, leading to integrals over highly oscillatory functions. We investigate two approaches to solve this problem. We employ complex Langevin method, which is a complexified stochastic process and investigate its properties. We apply it to QCD in a region where other methods are unreliable, we go up to $\mu/T_c \approx 5$. We finally investigate its applicability for SU(2) real-time simulations. We also investigate the Lefschetz Thimble method, which solves the sign problem by deforming the manifold of integration, such that there is no more oscillatory behavior. We discuss aspects of the method in simple models and develop algorithms for higher dimensions.
Finally, we apply neural networks to lattice simulation data and use them to extract the order parameter for the phase transition in the Ising model and SU(2) gauge theory. Thus, we uncover what the neural network learns.

Contents

1 Introduction

Knowledge of the fundamental interactions between the constituents of matter is imperative to our understanding of how nature works. Those interactions are largely described by the standard model of particle physics, which is a staggering success of modern high energy physics. It describes the fundamental interactions of particles via the electroweak and strong force and has been able to explain many phenomena in very precise agreement to experimental results. The most prominent recent success is the discovery of the Higgs boson [1, 2]. In this thesis, we are interested in the strong sector of the standard model, which is described by quantum chromodynamics (QCD). While QCD at high energies is well described by perturbative calculations, it is strongly coupled at low energies and thus requires non-perturbative methods. QCD at $T \geq 0$ can be simulated with high-performance lattice simulations to yield high precision results, which are in very good agreement with experimental data [3]. Hence, nowadays lattice QCD (LQCD) is the go-to method for precision computations of QCD, it will be the method of choice in this thesis as well.

1.1 Motivation

QCD describes the dynamics of quarks which are interacting via gluons, both carry a charge called "color". Its vacuum physics is well understood in the high energy regime via perturbative calculations and in the low energy regime via lattice QCD simulations. In nature only color neutral objects exist, thus quarks and gluons are strongly coupled and form bound states called hadrons. On the other hand, we know that scattering processes of QCD at high energy collider experiments are weakly coupled in the so-called quark-gluon plasma. It is thus described very well by perturbation theory. This weak coupling tends to zero for high energies, which is known as asymptotic freedom [4, 5]. Its discovery led to the physics Nobel prize in 2004. The transition from the quark-gluon plasma to hadronic matter is described by two phenomena. In Yang-Mills theory this transition is a proper phase transition, which goes from a deconfined phase at high temperature into a confined phase at low temperatures. Similarly, QCD with massless quarks exhibits chiral symmetry breaking from high temperatures where quarks are massless to low temperatures where the constituents of hadrons have acquired mass. In practice both of those transitions turn into a crossover, i.e. the transition is continuous and smooth since neither pure Yang-Mills theory nor massless quarks are realized in nature. This crossover is known as the QCD crossover transition. It is generally accepted that the transition is a crossover and that deconfinement and chiral symmetry breaking happen simultaneously [6]. At vanishing chemical potential, the crossover is acces-

sible to high precision from first principle lattice QCD simulations [6]. Alternative first principle approaches are functional methods such as the functional renormalization group (FRG) [7–10] and Dyson-Schwinger equations, for a review, see [11].
The situation changes for finite baryon density or baryochemical potential μ[1]. Here lattice simulations are no longer applicable due to the so-called sign problem [12] – the Boltzmann factor is no longer real and positive and thus loses its interpretation as a probability distribution. However, model computations show that there are interesting phenomena at finite μ. For instance models of QCD specifically focusing on chiral symmetry and deconfinement show a critical endpoint of the transition where the crossover turns into a first-order transition, see e.g. [13, 14]. The same is seen in computations via the Dyson-Schwinger equations. Thus, it is important to confirm or dispute the existence of the critical endpoint of QCD from first principle methods. Despite the sign problem for lattice simulations at finite μ it is, possible to use expansion or extrapolation techniques to obtain results at small finite μ anyway. Hence, there is a strong focus of several collaborations on investigating the phase diagram with large success at $\mu/T < 1$, for recent works see [6, 15–19]. For an overview addressing the QCD phase diagram in general and other phases see [20–22].
Experimentally the phase diagram can be probed via heavy-ion collisions, which is done at the Large Hadron Collider (LHC) [23] and the Relativistic Heavy Ion Collider (RHIC) [24]. Here the transition cannot be directly observed, instead so-called chemical freeze-out – the region where hadron content no longer changes – is investigated and the corresponding freeze-out temperature and chemical potential are extracted, see [25] for a recent article.
Nowadays results from lattice computations and experiment agree well at $\mu/T < 1$. Hence, it is highly desirable to directly simulate at finite chemical potential in order to go to higher μ/T to determine the precise location critical endpoint. In this thesis will investigate two of the approaches to the sign problem, the complex Langevin method and the Lefschetz Thimble method with the goal of simulating first principle QCD at finite T and μ. We are able to use the complex Langevin method to simulate regions of the phase diagram which so far have been inaccessible in lattice simulations.

Lattice simulations typically require large statistics to yield high precision results. Hence it is natural to apply techniques designed for "big data". Recently there has been a lot of progress in applying machine learning and in particular neural networks to physics problems, see [26] for a review. Oftentimes these techniques are treated as black boxes. I.e. there is no general understanding of what exactly they learn. As such they are unsuitable for purely theoretical predictions. However, an understanding of the learning process and structure of machine learning algorithms could make them suitable for such tasks. An application could be the sign problem

[1] Chemical potential is the energy that is related to a change in particle number. Baryochemical potential refers specifically to baryons.

e.g. by finding coordinate transformations or manifolds of integration where the sign problem is weaker [27, 28]. We will take a different approach here and apply neural networks to data from Monte Carlo simulations in order to detect phase transitions and understand what the neural network learns in the process – thus partly decoding the black box.

1.2 Publications

While the compilation of this thesis was done solely by the author, the vast majority of the work presented here was done together with my collaborators. The yet unpublished parts of this work are based on collaborations with

- E. Seiler, D. Sexty and I.-O. Stamatescu for the improved method for boundary terms in complex Langevin.
- D. Sexty and I.-O. Stamatescu for everything related to complex Langevin and QCD.
- J.M. Pawlowski, C. Schmidt, I.-O. Stamatescu, F.P.G. Ziegler and F. Ziesché for the local thimble algorithm.

All results from published works were obtained with the corresponding co-authors. Results and figures from these published works are not marked explicitly. The corresponding publications are:

- [29] **Complex Langevin and boundary terms**
  Manuel Scherzer, Erhard Seiler, Dénes Sexty and Ion-Olimpiu Stamatescu
  Published in Phys.Rev. D99 (2019) no.1, 014512.
  Eprint: arXiv:1808.05187
  Comment: The contents of this work are contained in section 3.1.1.
- [30] **Reweighting Lefschetz Thimbles**
  Stefan Bluecher, Jan M. Pawlowski, Manuel Scherzer, Mike Schlosser, Ion-Olimpiu Stamatescu, Sebastian Syrkowski and Felix P.G. Ziegler
  Published in SciPost Phys. 5 (2018) no.5, 044.
  Eprint: arXiv:1803.08418
  Comment: The contents of this work are contained in section 4.2.
- [31] **Machine Learning of Explicit Order Parameters: From the Ising Model to SU(2) Lattice Gauge Theory**
  Sebastian Johann Wetzel and Manuel Scherzer
  Published in Phys.Rev. B96 (2017) no.18, 184410.
  Eprint: arXiv:1705.05582
  Comment: The contents of this work are contained in section 5.2.1.

There are also some proceedings from the annual lattice conference, which present some of the results discussed in this thesis.

- [32] **Getting even with CLE**
  Gert Aarts, Kirill Boguslavski, Manuel Scherzer, Erhard Seiler, Dénes Sexty, Ion-Olimpiu Stamatescu
  Published in EPJ Web Conf. 175 (2018) 14007.
  Eprint: arXiv:1710.05699
  Comment: Parts of the contents of this work are contained in sections 3.2.1 and 3.3.

There is another proceeding from the 2018 lattice conference [33] which contains parts of 3.2.3, however, the review process of those proceedings is not yet finished, hence the work is not officially published.
I am currently also working on a project within Prof. Dr. Jan M. Pawlowskis group on the application of neural networks to extract spectral functions from propagator data, which is an ill-posed inverse problem. The progress of this project is not contained in this thesis.

1.3 Outline

This thesis is organized as follows. Chapter 2 introduces the basics of the phase diagram of quantum chromodynamics (QCD), lattice QCD and the sign problem. The purpose of this chapter is mainly to set the stage and establish notation, the content does not contain new research and may thus be skipped by a reader familiar with those topics.
In chapter 3 we introduce the complex Langevin method as an approach to the sign problem, discuss its convergence properties and apply it to QCD and SU(2) lattice gauge theory on a real-time contour.
Chapter 4 introduces another approach to the sign problem, namely Lefschetz thimbles. We investigate aspects of the method in simple models and discuss its applicability in higher dimensional theories.
In Chapter 5 we briefly introduce feed-forward neural networks and show how they can be used to find order parameters from lattice configurations while also finding out what the neural network learns, thus decoding the black box for specific cases.
We summarize our findings and give an outlook in chapter 6.
Finally, the appendices A-C contain more technical information on chapters 3-5.

2 QCD, the lattice, and the sign problem

In this part, we briefly introduce the relevant physics and basic methods that are necessary throughout this thesis.

A brief reminder of continuum QCD

The QCD Lagrangian describes the coupling of quarks and gluons. With the inclusion of finite baryochemical potential it reads

$$\mathcal{L}_{\mathrm{QCD}} = \frac{1}{4} F^a_{\mu\nu} F^a_{\mu\nu} + \sum_\alpha \bar{\psi}^\alpha \left(\gamma_\mu D_\mu + m^\alpha + i\mu\gamma_0\right) \psi^\alpha \,. \tag{2.1}$$

We work in Euclidean space, hence the metric is just $\delta_{\mu\nu}$. We also use $c = k_B = \hbar = 1$ throughout this thesis. In equation (2.1) as well as throughout this thesis the subscript Greek letters refer to space-time indices, i.e. $\mu \in 0 \ldots d-1$ with d being the number of dimensions, superscript Latin letters represent the adjoint color index $a = 1 \ldots N^2 - 1$, where N refers to the gauge group $SU(N)$. The superscript α stands for flavors, i.e. $\alpha \in \{\mathrm{u}, \mathrm{d}, \mathrm{s}, \mathrm{c}, \mathrm{b}, \mathrm{t}\}$. $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$ is the field strength tensor with f^{abc} being the structure constants of SU(3). $D_\mu = \left(\partial_\mu + i g A^a_\mu t^a\right)$ is the covariant derivative with t^a being the generators of SU(3). For a detailed derivation of those quantities we refer to standard textbooks such as [34, 35]. QCD in contrast to Quantum electrodynamics (QED), which describes the interaction between particles with electric charge, is a nonabelian gauge theory, meaning that the elements of the gauge group SU(3) do not commute. A consequence of this is the term $g f^{abc} A^b_\mu A^c_\nu$ in the field strength tensor. Since the action is quadratic in the field strength there are self-interactions between the gluons. This is in contrast to QED with the abelian gauge group U(1). There this term does not occur and hence there is no interaction between photons. Possible interactions from the QCD Lagrangian in equation (2.1) are between three gluons, four gluons, and two quarks and a gluon. Those interactions are described by vertex terms and together with the gluon and quark propagators they make up the essential quantities needed to describe QCD in the continuum.

There is one property of QCD that is of particular importance for its lattice discretization, namely the invariance under local gauge transformations

$$\psi \to U\psi A_\mu \to U A_\mu U^\dagger - \frac{i}{g} U \left(\partial_\mu U^\dagger\right) \,, \tag{2.2}$$

where $U(x) \in \mathrm{SU}(3)$. There are many intricate details related to gauge symmetry such as the need for gauge fixing in continuum calculations, see e.g. [34, 35]. On

the lattice, gauge fixing is not necessary since lattice QCD is formulated inherently gauge invariant, see section 2.3. Thus, we will not discuss gauge fixing.
In this thesis, we are interested in QCD at finite chemical potential μ and finite temperature T. We will thus briefly introduce the important phenomena occurring in the corresponding phase diagram.

2.1 The phase diagram of QCD

One of the best-known property of QCD is asymptotic freedom [4, 5] whose discovery lead to the physics Nobel prize in 2004. Asymptotic freedom is the property of QCD that at large energies the coupling strength between particle becomes small, such that they effectively behave like free particles. It is the reason that perturbative calculations for QCD at arbitrarily high energies are possible. The state of QCD matter at high energies is called the quark-gluon plasma. In contrast, at low energies the coupling strength becomes large. Hence, perturbation theory is no longer applicable and particles are strongly interacting. Due to this strong interaction confinement occurs. Confinement is the phenomenon that at low energies quarks and gluons can exist only in color-neutral bound states, such as protons and neutrons. Hence, it is an interesting question to ask for the properties of the transition between those regions.

The deconfinement transition

Confinement is usually defined via the potential between a static quark-antiquark pair. To be more precise, at large distances the potential rises linearly,

$$\lim_{r\to\infty} V_{q\bar{q}}(r) = \sigma r\,. \tag{2.3}$$

Thus an infinite amount of energy is required to separate two static quarks. In nature another quark-antiquark pair will be created at sufficiently large energies introducing a screening effect for the potential, this is referred to as string breaking. In the deconfined phase the potential is Coulomb-like

$$\lim_{r\to 0} V(r) = \frac{a}{r}\,. \tag{2.4}$$

The behavior of the potential can be shown best in certain limits in the lattice discretization of Yang-Mills theories, see e.g. [36, 37] for a derivation. Lattice simulations have confirmed that this behavior holds also outside those limits, see e.g. [38–40] for some examples. The deconfinement transition is a phase transition only in pure Yang-Mills theory since it relies on quarks being static. This is not the case in nature such that the transition becomes a crossover. In pure Yang-Mills the order parameter for the transition is the so-called Polyakov loop $\langle L\rangle$. Physically, it is related to the free energy of a static quark-antiquark pair

$$e^{-\beta F_{q\bar{q}}(\vec{n}-\vec{m})} = \left\langle L(\vec{n})L^\dagger(\vec{m})\right\rangle\,. \tag{2.5}$$

If the Polyakov loops decouple for large distances $|\vec{n} - \vec{m}| \to \infty$ we have

$$\left\langle L(\vec{n})L^{\dagger}(\vec{m})\right\rangle = |\langle L\rangle|^2 \ . \tag{2.6}$$

If $\langle L\rangle = 0$ this means that $F_{q\bar{q}} \to \infty$ as the distance increases. On the other hand if $\langle L\rangle > 0$, we have $F_{q\bar{q}} = \text{const} < \infty$, in which case the energy becomes constant for infinite distances. This can be interpreted as the deconfinement transition, since in the confined phase it takes infinitely much energy to separate a bound state, while in the deconfined phase a finite amount of energy suffices. To conclude

$$\begin{aligned} &\langle L\rangle = 0 \Rightarrow \ \text{confinement} \\ &\langle L\rangle \neq 0 \Rightarrow \ \text{deconfinement} \, . \end{aligned}$$

In the strong coupling limit of lattice QCD one can compute the potential of a quark antiquark pair directly and show that in what we called the confined phase it rises linearly with distance, while it has a Coulomb like behavior in the deconfined phase, see e.g. [36]. With the inclusion of fermions the deconfinement transition becomes a crossover, however the Polyakov loop can still be used to give a transition temperature. Due to the inherent noisiness of the Polyakov loop it is often simpler to look at its absolute value, i.e. $\langle |L| \rangle$.
There is a symmetry connected to the deconfinement transition, the so-called center symmetry [41]. The center of a group is given by elements which commute with all elements of the group. Pure Yang-Mills theory is invariant under center symmetry transformations

$$U = zU \, , \tag{2.7}$$

where $U \in \mathrm{SU(N)}$ is the exponential of the gluon field A_μ and z an element of the center of SU(N) center. The Polyakov loop is not invariant under the same transformation $L \to zL$, such that center symmetry holds only when $\langle L\rangle = 0$. In that sense center symmetry breaking signals deconfinement.

Chiral symmetry breaking

Another transition related to the QCD crossover is the chiral phase transition. The QCD Lagrangian from equation 2.1 in the limit of vanishing masses – the chiral limit – exhibits chiral symmetry. The quark field can be decomposed as $q = q_R + q_L$ with projectors

$$q_{L/R} = P_{\pm} q = \frac{1}{2}\left(1 \pm \gamma_5\right) \ . \tag{2.8}$$

Under this decomposition the Lagrangian is invariant under

$$\begin{aligned} q_L &\to e^{i\alpha} q_L \\ \bar{q}_L &\to \bar{q}_L e^{-i\alpha} \\ q_R &\to e^{i\beta} q_R \\ \bar{q}_R &\to \bar{q}_R e^{-i\beta} \, , \end{aligned} \tag{2.9}$$

where $\alpha, \beta \in \mathbb{R}$ are constants. This symmetry is spontaneously broken at low temperatures, an order parameter is given by the chiral condensate $\langle q\bar{q}\rangle$, which corresponds to the mass term in equation (2.1) The symmetry breaking pattern is the following

$$\begin{aligned} U(N_f)_L(N_f)_R &= SU(N_f)_V \times U(1)_V \times SU(N_f)_A \times U(1)_A \\ &\xrightarrow{\text{anomalous}} U(N_f)_V \times U(1)_V \times SU(N_f)_A \\ &\xrightarrow{\text{chiral}} U(N_f)_V \times U(1)_V \,. \end{aligned} \tag{2.10}$$

The anomalous breaking of $U(1)_A$ is related to the axial anomaly [42–45], which leads to the mass splitting of the η and η' mesons [46]. Chiral symmetry breaking is described by the breaking of $SU(N_f)_A$. In the chiral limit, it leads to massless pions, the Goldstone bosons of chiral symmetry. In nature the mass of the up- and down quarks are only approximately zero, hence the pions have a small mass as well.

The QCD phase transition and chemical freeze-out

The above transitions are very well studied in the limit of vanishing chemical potential by means of lattice QCD and by now it is accepted that the transition is a crossover [6] and that the deconfinement and chiral phase crossovers happen at the same point. The transition temperature in full QCD with $N_f = 2 + 1$ is $T_c \approx 155$ MeV [16, 47] and about 25 MeV lower in the chiral limit [48].
The situation is different at finite chemical potential μ. Due to the sign problem, which will be discussed below in section 2.4, lattice QCD is not easily applicable at finite μ. Standard lattice simulations are restricted to either extrapolation from imaginary chemical potential, see e.g. [18, 19, 49] or Taylor expansion around $\mu = 0$, see e.g. [15, 16, 19]. The results from those suggest that the transition temperature $T_c(\mu)$ is well described by a quadratic behavior with curvature $\kappa_2 \approx 0.014$. There are many results from continuum computations from Dyson-Schwinger equations, see [11] for a review or from model computations, see e.g. [13, 14]. The phase diagram in those computations typically shows a critical endpoint. The critical endpoint of QCD is the point where the crossover turns into a first order transition. Unfortunately, the prediction of the position of the critical point between those methods varies wildly. It remains to be seen whether standard lattice computations can prove or disprove the existence of a critical point since extrapolations and expansions typically have a radius of convergence and thus may not be able to reach it, so far they yield reliable results up to $\mu/T < 1$. Hence, it is important to look for alternatives which allow for direct simulation at finite μ, such as the complex Langevin method which we will discuss in chapter 3 or the Lefschetz thimble method which we discuss in chapter 4.
Experimentally there is no way to directly probe the phase transition. Instead, the so-called chemical freeze-out parameters can be measured. Chemical freeze-out describes the region in the phase diagram at which the hadron content in a heavy-ion collision no longer changes, a comparison for criteria of chemical freeze-out can be found in [50]. It is clear that this transition has to be below the QCD

phase transition since hadrons do not exist above the transition. Unfortunately in heavy-ion collisions, one cannot just tune μ and T since the center-of-mass energy of a nucleon-nucleon pair $\sqrt{s_{\mathrm{NN}}}$ is the only adjustable parameter. What is typically done to extract the transition parameters is a fit of the so-called hadron resonance gas model to experimental quantities such as hadron yields [25]. The hadron resonance model is motivated from statistical physics, it is a mixed gas of baryons and mesons, i.e.

$$\begin{aligned}\log Z(T,V,\mu_B,\mu_Q,\mu_S) \approx &\sum_{i\in\text{mesons}} \log Z_{m_i}(T,V,\mu_Q,\mu_S) \\ &+ \sum_{i\in\text{baryons}} \log Z_{m_i}(T,V,\mu_B,\mu_Q,\mu_S)\,, \end{aligned} \tag{2.11}$$

where m_i is the mass of the hadron. The sum over mesons and baryons incorporates all stable particles and resonances listed in the Particle Data Booklet [51]. Those fits then yield μ_B and T. Note however that they can only be trusted below the phase transition, where hadrons exist. The chemical freeze-out temperature at $\mu_B \approx 0$ agrees within errors with the phase transition temperature, suggesting that the phenomenon occurs close to the QCD transition [25]. The freeze-out temperature and chemical potential are usually described by rather simple parametrizations of $\mu_B(\sqrt{s_{\mathrm{NN}}})$ and $T(\sqrt{s_{\mathrm{NN}}})$, e.g. from [25]

$$T_{\mathrm{CF}} = \frac{158.4\ \mathrm{MeV}}{1 + e^{2.60 - \frac{\log\sqrt{s_{\mathrm{NN}}}}{0.45}}} \tag{2.12}$$

$$\mu_B = \frac{1307.5\ \mathrm{MeV}}{1 + 0.288\sqrt{s_{\mathrm{NN}}}}\,, \tag{2.13}$$

which describes the experimental data remarkably well in the measured region down to $T_{\mathrm{CF}} \approx 60$ MeV and up to $\mu_B = 800$ MeV. There are also recent efforts to extract the temperature by comparing lattice simulations to experiment [52].
Concerning the critical endpoint, so far there has been no experimental evidence for its existence.

Nuclear liquid-gas transition

Another better-understood transition is the nuclear liquid-gas transition. At vanishing temperature $T = 0$, a baryon can only be created if there is enough energy, i.e. when $\mu_B = m_B$. While this is physically intuitive, it is not simple to deduce this from the Dirac operator. This phenomenon has been coined the Silver-Blaze problem [53]. At $T = 0$ the sudden creation of a particle corresponds to a first order phase transition, where the density is the order parameter. At sufficiently small $T > 0$, the transition is still first order by continuity and shows a critical endpoint. There is no first principle computation for the location of the critical endpoint of this transition, however models and experiment largely agree, see e.g. [54, 55]. It is found at $T \approx 10-20$ MeV. It is also desirable to have a first principle determination

of the equation of state at $T = 0$ as a function of μ, since this is an important input e.g. for the mass-radius relation of neutron stars, for a review see [56].

There are other phases in the QCD phase diagram such as a color-superconducting phase, see e.g. [20] for a review. Such exotic phases are not the topic of this thesis.

2.2 Equilibrium quantum field theory in real time

Since we are interested in doing lattice QCD simulations, we typically work in Euclidean space, i.e. imaginary time, such that the Boltzmann factor is $\exp(-S_E)$. There are many quantities such as spectral functions or transport coefficients, which are inaccessible in imaginary time and require continuation to real time.
In the case of imaginary time, the time variable is traded for inverse temperature due to the analogy to statistical physics. Bosonic (fermionic) fields are (anti-)periodic in inverse temperature, see e.g. [57, 58] for a derivation of the corresponding path integral. Correlation functions can be taken along the imaginary time or temperature axis, i.e. along a path from 0 to $-i/T$. If real-time t comes into play, this path has to be deformed. The common way to do this is the closed time path formalism, i.e. the integration path along which the action is integrated is deformed according to

$$\int_0^{\beta} d\tau \int d\vec{x}\mathcal{L}_E \rightarrow \int_0^{-i\beta} dt \int d\vec{x}\mathcal{L}_M \rightarrow \int_{\mathcal{C}} dt \int d\vec{x}\mathcal{L}_M \tag{2.14}$$

where in the first step we rephrased the imaginary time action in terms of real time to visualize that the contour of integration in the imaginary time formalism goes along the negative imaginary axis. In the second step we changed the integration path according to

$$\int_{\mathcal{C}} = \int_{-\infty}^{\infty} dt_+ + \int_{\infty}^{-\infty} dt_- + \int_{-\infty}^{-\infty-i\beta} . \tag{2.15}$$

Physically the situation is the following: the field starts in its asymptotic past at $t = -\infty$, we measure correlation functions at some intermediate time t, let the field evolve to the asymptotic future at ∞, The contour in equation (2.15) is called the Schwinger-Keldysh contour. back to $-\infty$ and finally to $-i\beta$. Thus we have included all finite time and finite temperature effects. For the purpose of this thesis details of the formalism are not important, see [58] for Greens functions and a proper path integral definition along the path. It is important to note that deformations of the path are also possible [59], which we will make use of in section 3.3 when setting up real-time lattice simulations along such paths.

2.3 Lattice discretization of QCD

Lattice gauge theory was originally introduced by Wegner [60] for Z_2 gauge theory and by Wilson [61] for SU(N). It quickly gained popularity and nowadays is the go-to method for high precision computations in high energy physics and especially QCD, where applicable. Due to this popularity, there exist many good textbooks on the matter. Some of the more recent ones are [36, 37, 62–64]. Since our introduction in this section will be rather brief, we refer to those textbooks for details.

We start the lattice discretization by introducing a four-dimensional lattice Λ with index $n = t, x, y, z$ and periodic boundary conditions. For now, we work without chemical potential. On each lattice site there is a fermion, which we demand to transform under local gauge transformations, which we now call Ω as

$$\psi(n) \to \Omega(n)\psi(n) \text{ and } \bar{\psi}(n) \to \bar{\psi}(n)\Omega^\dagger(n)\,. \tag{2.16}$$

There is an apparent problem when looking at the discretization of parts of the derivative term

$$\bar{\psi}\partial_\mu\psi \to \bar{\psi}(n)\frac{1}{2a}\left(\psi(n+\hat{\mu}) - \psi(n-\hat{\mu})\right)\,, \tag{2.17}$$

where $\hat{\mu}$ is the unit vector in direction μ, i.e. it refers to the neighboring point in direction μ. This term is not invariant under gauge transformations

$$\bar{\psi}(n)\psi(n+\hat{\mu}) \to \bar{\psi}(n)\Omega^\dagger(n)\Omega(n+\hat{\mu})\psi(n+\hat{\mu})\,, \tag{2.18}$$

unless an additional field $U_\mu(n)$ which transforms as

$$U_\mu(n) \to \Omega(n)U_\mu(n)\Omega^\dagger(n+\hat{\mu})\,, \tag{2.19}$$

is introduced to connect the fermions. The fermionic derivative term is replaced according to

$$\bar{\psi}(n)\psi(n+\hat{\mu}) \to \bar{\psi}(n)U_\mu(n)\psi(n+\hat{\mu})\,, \tag{2.20}$$

which again is gauge invariant. Note that the field $U_\mu(n)$, which is called the gauge link has a direction. It points from n to $n+\hat{\mu}$, this will be important when going back to the continuum. The naive fermion action on the lattice then reads

$$S_F = a^4 \sum_{n\in\Lambda} \bar{\psi}(n)\left(\sum_{\mu=1}^{4}\gamma_\mu \frac{U_\mu(n)\psi(n+\hat{\mu}) - U_{-\mu}(n)\psi(n-\hat{\mu})}{2a} + m\psi(n)\right)\,. \tag{2.21}$$

To get the continuum action back, we define

$$U_\mu(n) = e^{iaA_\mu(n)} = 1 + iaA_\mu(n) + \mathcal{O}(a^2)\,. \tag{2.22}$$

By plugging this into the fermionic lattice action it is straight forward to see that up to $\mathcal{O}(a)$ this yields the continuum action.
The gauge part of the action is represented by

$$S_G = -\beta \sum_{U_p} \left(\frac{1}{2N_c} \left(\mathrm{Tr}(U_p) + \mathrm{Tr}(U_p^{-1}) \right) - 1 \right) \tag{2.23}$$

with the plaquette U_p defined via

$$U_p = U_{\mu\nu}(n) = U_\mu(n) U_\nu(x+\hat{\mu}) U_\mu^{-1}(n+\hat{\nu}) U_\nu^{-1}(n)\,, \tag{2.24}$$

which is the smallest possible closed loop on the lattice. This action is called the Wilson gauge action. Without providing much detail, we note that one can regain the continuum gauge action by inserting $U_\mu(n) = e^{iaA_\mu(n)}$ as in the fermionic case, see e.g. [37]. Note that in the standard lattice literature the link variables are in SU(N) and hence further simplifications due to $U^{-1} = U^\dagger$ can be made. Since in a large part of this thesis we are interested in the group $\mathrm{SL(N,\mathbb{C})}$ we cannot make those simplifications.

Fermion doubling and Wilson fermions

Unfortunately the naive fermion action in equation (2.21) has a problem. We start by defining the lattice Dirac operator, which we will call M, to avoid confusion later

$$M(n,m)_{ab}^{\alpha\beta} \sum_\mu (\gamma_\mu)^{\alpha\beta} \frac{U_\mu(n)_{ab}\delta_{n+\hat{\mu},m} - U_{-\mu}(n)_{ab}\delta_{n-\hat{\mu},m}}{2a} + m\delta_{ab}\delta^{\alpha\beta}\delta_{nm}\,, \tag{2.25}$$

with Dirac indices α, β, color indices a, b and space-time positions n, m. In Fourier space it becomes

$$\tilde{M}(p,q) = \delta(p-q) \left(m\mathbb{1} + \frac{i}{a} \sum_{\mu=1}^{4} \gamma_\mu \sin(p_\mu a) \right)\,. \tag{2.26}$$

Its inverse can be shown to be

$$\tilde{M}(p)^{-1} = \frac{m\mathbb{1} - ia^{-1}\sum_\mu \gamma_\mu \sin(p_\mu a)}{m^2 + a^{-2}\sum_\mu \sin(p_\mu a)^2}\,. \tag{2.27}$$

When Fourier transforming back into real space, this is the lattice equivalent of the quark propagator. Looking at the massless limit, we observe that we have $\sin(p_\mu a) = 0$ for $p_\mu = 0$ as well as $p_\mu = \pi/a$ and its periodic images. Hence, we do not only get the physical mass pole at $p_\mu = 0$ but 15 additional poles [2]. There are

[2] There are 16 poles in total since the momentum p is a four-vector and each of the entries can be either π/a or zero.

different approaches to get rid of those so-called doublers. Here we discuss so-called Wilson fermions since we will use them in our simulations. For Wilson fermions to get rid of those doublers, one adds an extra term

$$\tilde{M}(p) \to \tilde{M}(p) + \mathbb{1}\frac{1}{a}\sum_{\mu=1}^{4}\left(1-\cos\left(p_\mu a\right)\right) , \tag{2.28}$$

which vanishes in the case of $p = 0$ but is finite for $p = \pi/a$. This term gives an additional mass to the doublers, which now have mass $m + 2l/a$, with l being the number of momentum components π/a. In the continuum limit, $a \to 0$ the doublers become infinitely heavy and hence decouple from any dynamics. The additional term can be Fourier transformed back and upon insertion of link variables in appropriate places yields an expression very similar to the naive fermion action on the lattice. Both can be combined to yield the full Dirac operator for Wilson fermions

$$\begin{aligned} M(n,m)^{\alpha\beta}_{ab} = &\left(m+\frac{4}{a}\right)\delta^{\alpha\beta}\delta_{ab}\delta_{nm} \\ &-\frac{1}{2a}\sum_{\mu=1}^{4}\left((\mathbb{1}-\gamma_\mu)^{\alpha\beta}U_\mu(n)\delta_{n+\hat{\mu},m}+(\mathbb{1}+\gamma_\mu)^{\alpha\beta}U_{-\mu}(n)\delta_{n+\hat{\mu},m}\right) . \end{aligned} \tag{2.29}$$

Finally, let us slightly reparametrize the Dirac operator by introducing the so called hopping parameter

$$\kappa = \frac{1}{2(am+4)} , \tag{2.30}$$

which allows us to rewrite the Dirac operator as

$$M = C(\mathbb{1}-\kappa H) , \tag{2.31}$$

with $C = m + 4/a$ and

$$H(n,m)^{\alpha\beta}_{ab} = \sum_{\mu=1}^{4}\left((\mathbb{1}-\gamma_\mu)^{\alpha\beta}U_\mu(n)\delta_{n+\hat{\mu},m}+(\mathbb{1}+\gamma_\mu)^{\alpha\beta}U_{-\mu}(n)\delta_{n+\hat{\mu},m}\right) . \tag{2.32}$$

Note that $U_{-\mu}(n) = U^{-1}_{\mu}(n-\hat{\mu})$, which we will use in some equations later. We finally redefine the quark fields $\psi \to \sqrt{C}\psi$ such that we can forget about C. This formulation is convenient since κ corresponds to inverse mass and becomes large for small m. It can hence be used as an expansion parameter, a fact we will use in section 3.2.1.

Introduction of chemical potential

So far the lattice formulation has been introduced without a chemical potential for simplicity. It can be introduced by going from the canonical to the grand canonical

ensemble. As seen in equation (2.1), only the temporal direction relates to chemical potential. Then H in the Dirac operator from equation (2.32) is replaced by

$$H(n,m)_{ab}^{\alpha\beta} = \sum_{\nu=1}^{4} \left(e^{\delta_{\nu,4}\mu} (\mathbb{1} - \gamma_\nu)^{\alpha\beta} U_\nu(n) \delta_{n+\hat{\nu},m} + e^{-\delta_{\nu,4}\mu} (\mathbb{1} + \gamma_\nu)^{\alpha\beta} U_{-\nu}(n) \delta_{n+\hat{\nu},m} \right) . \tag{2.33}$$

Introducing chemical potential μ in this way ensures that the correct continuum limit is reached, see e.g. [37] for details.
Given a lattice configuration from a simulation of QCD discretized as explained above, we can compute physical observables on the configuration.

Observables on the lattice

The fact that observables generally should be gauge invariant limits our freedom to construct them. The simplest observables are traces of closed loops made up of link variables.

Plaquettes

The simplest of those is the plaquette, which already occurred in the definition of the Wilson gauge action. It is the trace of the smallest possible closed loop starting on a given site, which is the product of four links. It is sometimes convenient to look at the spatial and temporal plaquette separately. The temporal plaquette is defined as the average over all plaquettes on a given lattice configuration which contain a temporal link. The spatial plaquette only consists of spatial links.

Polyakov loop

Another useful observable is the Polyakov loop, which we already introduced in 2.1. On the lattice, it is the product of all temporal link variables at a spatial point

$$L = \frac{1}{N_c} \mathrm{Tr} \prod_{t=0}^{N_t - 1} U(t, \vec{n}) \, . \tag{2.34}$$

Similarly, one can define the inverse Polyakov loop

$$L^{-1} = \frac{1}{N_c} \mathrm{Tr} \prod_{t=N_t-1}^{0} U^{-1}(t, \vec{n}) \, . \tag{2.35}$$

Chiral condensate

There are several fermionic observables that are useful. One of them is the chiral condensate, which is given by

$$\langle \bar{\psi}\psi \rangle = \frac{1}{N_s^3 N_t} \frac{\partial \ln Z}{\partial m} = -\frac{2\kappa N_f}{N_s^3 N_t} \mathrm{Tr}(D^{-1}(D-1)) = -\frac{2\kappa N_f}{N_s^3 N_t} \mathrm{Tr}(D^{-1}) \tag{2.36}$$

in the last equality we dropped a constant, which we can do since the chiral condensate renormalizes additively [65]. It is the order parameter for the chiral transition in the limit of vanishing quark masses, which turns into a crossover at finite quark masses.

Quark number density
Derivatives of the quark number density

$$n = \frac{1}{N_s^3 N_t}\frac{\partial \ln Z}{\partial \mu} = \frac{N_f}{N_s^3 N_t}\left\langle \mathrm{Tr}\left(M^{-1}\frac{\partial}{\partial \mu}M\right)\right\rangle , \tag{2.37}$$

e.g. the susceptibility

$$\begin{aligned}\chi_n = \frac{1}{N_s^3 N_t}\frac{\partial^2 \ln Z}{\partial_\mu^2} &= \frac{N_f^2}{N_s^3 N_t}\left\langle \mathrm{Tr}\left(M^{-1}\frac{\partial M}{\partial \mu}\right)^2\right\rangle \\ &- \frac{N_f}{N_s^3 N_t}\left\langle \mathrm{Tr}\left(M^{-1}\frac{\partial M}{\partial \mu}M^{-1}\frac{\partial M}{\partial \mu}\right)^2\right\rangle \\ &+ \frac{N_f}{N_s^3 N_t}\left\langle \mathrm{Tr}\left(M^{-1}\frac{\partial^2 M}{\partial \mu^2}M^{-1}\right)\right\rangle - N_s^3 N_t n^2 .\end{aligned} \tag{2.38}$$

also signal the phase transition.

We will discuss the usefulness of those observables when we discuss the phase diagram in the context of complex Langevin.

Representation of SU(2) and SU(3) matrices and their complex generalizations

The Lie algebra su(n) which corresponds to the group SU(N) has N^2-1 generators $\lambda_a/2$, with λ_a the Gell-Mann matrices. Thus, in principle, it is enough to store N^2-1 real numbers per matrix. Since exponentiation is needed to get from the algebra into the group this is expensive. Instead, we save the first two rows u and v of the matrix, which allows us to reconstruct the third row via the cross product $w = \bar{u}\times\bar{v}$, this is convenient on GPUs, where a memory operation is much more expensive than a few floating point operations, we will do this for QCD. CPU simulations typically are not limited by memory bandwidth and we can just save the full matrix, i.e. 9 complex or 18 real numbers we will do this for HDQCD and the Polyakov chain. Both representations are also useful since in the case of complexification of the manifold i.e. when we consider SL(3,$\mathbb{C}$) – as we will do in chapter 3 – nothing needs to be changed, even though the coefficients of the generators become complex.
For SU(2) there is a real representation, it can be written in terms of Pauli matrices σ_i as

$$U = x_0\mathbb{1} + i\sum_{i=1}^{3} x_i\sigma_i , \tag{2.39}$$

with $x_i \in \mathbb{R}$ and $\det U = \sum x_i^2 = 1$. After complexification the coefficient x_i can be a complex number, the rest stays the same.

From the lattice to physics

On the lattice, all quantities are formulated in a dimensionless way. To get back to physical units one has to multiply by appropriate powers of the lattice spacing, which has the dimension of length. E.g. temperature, which is given by the temporal lattice extent as $1/N_t$ in physical units reads $1/(aN_t)$. The appropriate power of a can be obtained by dimensional analysis of the quantity at hand.
Getting to physical values also requires an actual value for the lattice spacing. This can be done in several ways, e.g. by measuring some observables with good precision and fixing them to the corresponding physical value. However, a particularly convenient and cheap way is the gradient flow [66]. The gradient flow is a method which smoothes a given lattice configuration. If too much smoothing is applied, it drives the lattice towards the classical theory. It is a way to get rid of strong UV fluctuations, for an alternative way to do this see [67]. For the gradient flow observables typically monotonously tend towards their classical values as flow time increases. This is used for scale setting by monitoring the flow of $1/4F^{\mu\nu}F^{\mu\nu}$. When it reaches a certain value which has to be predetermined by other scale fixing methods, the corresponding flow time can be used as a scale. The advantage of this approach is that it is precise and cheap to do. We have used a high precision value from [68] to measure the scale in our QCD simulations. Since this value is for $N_f = 2 + 1$ at physical pion masses, our physical values are just approximate, since we have $N_f = 2$ flavors and heavy pions. The details of the procedure can be found in the corresponding references, they are not the focus of this thesis.

Whether a lattice simulation is close to the physical point or not is typically determined by the pion rest-mass, since it is the lightest particle of the theory. Therefore, we briefly explain how hadron masses can be obtained from the lattice, see [37] for more details. In this thesis we are mainly interested in the pion, hence we use it as an example. The idea is to find an observable which has the same quantum numbers as the pion. We then measure the correlation function of the creation operator and annihilation operator between different lattice points. Intuitively this means that we create a pion at some lattice site, let it propagate to another point and annihilate it there. The Euclidean correlator of such observables is given by

$$\left\langle \mathcal{O}(t)\bar{\mathcal{O}}(0)\right\rangle \sim \left\langle 0|e^{-(T-t)H}\mathcal{O}(t)e^{-tH}\bar{\mathcal{O}}(0)|0\right\rangle \xrightarrow{T\to\infty} \sum_n \left\langle 0|\mathcal{O}|n\right\rangle \left\langle n|\bar{\mathcal{O}}|0\right\rangle e^{-tE_n}\,, \quad (2.40)$$

where we inserted $1 = \sum_n |n\rangle\langle n|$ and let the Hamiltonian act on the states. Thus for large temporal lattice extent, we can extract the lowest lying energies, which are given by the rest mass of the observable, via an exponential fit. For the pion, we look at the pion interpolator given by

$$\pi = \bar{q}_1\gamma_5 q_2\,, \quad (2.41)$$

with $q_1 = u$ and $q_2 = d$ for the π_+, vice versa for π_- and we have a sum of two terms with $q_1 = u$, $q_2 = u$ and $q_1 = d$, $q_2 = d$ for the π_0. Note that for degenerate quark masses – which we have since we do not take into account the U(1) charge – the quarks and pions are not distinguishable. Therefore all three choices give the same result. One can show that this interpolator transforms just like the pion under parity and charge conjugation, see e.g. [37] for details. The correlator for a pion created at position m and annihilated at n is given by

$$\left\langle \mathcal{O}(n)\bar{\mathcal{O}}(m)\right\rangle = \left\langle \bar{d}(n)\gamma_5 u(n)\bar{u}(m)\gamma_4 d(m)\right\rangle = -\mathrm{Tr}\ \left[\gamma_5 M_u^{-1}(n,m)\gamma_4 M_d^{-1}(m,n)\right]\,, \tag{2.42}$$

where we contracted the quarks according to Wicks theorem. This correlator can be measured on the lattice. As a final step, we project the correlators to zero momentum. This is done by observing that the hadron interpolator can be Fourier transformed with respect to spatial momentum

$$\tilde{\mathcal{O}}(\vec{p}, n_t) \sim \sum_{\vec{n}} \mathcal{O}(\vec{n}, n_t) e^{-ia\vec{n}\cdot\vec{p}}\,, \tag{2.43}$$

such that for $\vec{p} = 0$ we only need to sum over all points in a given time-slice.

Monte Carlo simulations and importance sampling

In statistical physics the partition function is given over a sum of all possible energy states via

$$Z = \sum_i e^{-\frac{1}{T}E_i}\,. \tag{2.44}$$

Many interesting quantities can be directly derived from it. If there is a small finite amount of states this sum can often be evaluated analytically. However, typically those sums are too large and need to be computed by simulations. In lattice QCD we deal with infinite integrals instead of finite sums, which have to be approximated by a finite simulation. This is done by setting up a simulation which produces field configurations distributed according to the Boltzmann factor e^{-S}. Observables can be computed from a finite number N of those configurations as

$$\langle \mathcal{O}\rangle = \frac{1}{N}\sum_i \mathcal{O}_i \approx \frac{1}{Z}\int d\phi\, \mathcal{O} e^{-S(\phi)}\,. \tag{2.45}$$

This goes under the name of importance sampling. There are many different algorithms to produce such configurations, the simplest one being the Metropolis algorithm [69]

Since we will need variations of the Metropolis algorithm in chapter 4, we briefly describe it here. Given the Euclidean action S in terms of fields ϕ, the fields are updated sequentially – or in parallel if allowed by symmetry. The updating of a single site goes as follows

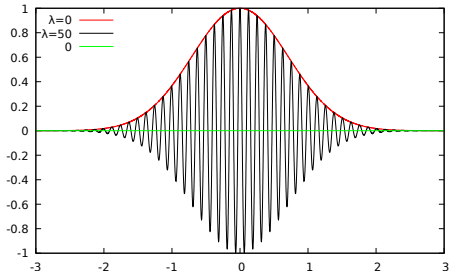

Figure 2.1: Illustration of the sign problem in a simple model from [71]. See text for an explanation.

1. Propose a new value for the field ϕ. This is typically done by adding a small random number from a symmetric probability distribution to the current value of a field.
2. Compute the difference in the action $\Delta S = S(\phi_{\text{new}}) - S(\phi_{\text{old}})$.
3. Draw a uniform random number u from $[0, 1]$. If $u \leq e^{-\Delta S}$ accept the update and set the field to its new value. Otherwise, reject and keep the old value.

This algorithm is repeated for all fields. The last step of the algorithm is called an accept-reject step and we will make use of it in our algorithms in chapter 4.
We will use the Metropolis algorithm for simulations of the Ising model and the heat bath algorithm [70] for SU(2) Yang-Mills simulations. For everything related to complex actions, we will use a complexified Langevin process, which will be discussed in much more detail in section 3. In the case of Lefschetz thimbles, as will be described in the corresponding section, we will use specialized algorithms to do importance sampling.

2.4 The sign problem in QCD

Consider the simple integral [71]

$$Z(\lambda) = \int_{-\infty}^{\infty} e^{-x^2 + i\lambda x} dx = \sqrt{\pi} e^{-\frac{\lambda^2}{4}} , \tag{2.46}$$

which is entirely real. Here Z has the form of a simple partition function. The real part of the integrand is plotted in figure 2.1, it is a smooth positive function for $\lambda = 0$ and the partition function takes the value $\sqrt{\pi}$ while it wildly oscillates for $\lambda = 50$ and the partition function evaluates to $Z(50) = 6.5 \times 10^{-272}$. Hence, when Monte Carlo sampling is applied contributions of the order $\mathcal{O}(1)$ have to be added up to yield a final result of almost 300 orders of magnitude lower. Getting a reliable value which can be distinguished from zero would require too many Monte Carlo samples. The numerical sign problem, therefore, is the problem of the precise evaluation

of highly oscillatory functions by sampling methods. In many theories – also in QCD at finite chemical potential, see equation (2.1) – oscillatory integrands are complex, prohibiting the usual probability interpretation of the measure in which the Boltzmann factor is positive. Thus the usual simulation algorithms no longer work. Both problems, oscillations and complex integrands mathematically speaking are not an issue since in principle one can shift parts of the measure into the observable by taking averages only with respect to some smooth and real $\rho_R(x)$ of the measure according to

$$\langle \mathcal{O}(x) \rangle = \frac{\int \mathcal{O}(x)\rho(x)dx}{\int \rho(x)dx} = \frac{\int \mathcal{O}(x)\frac{\rho(x)}{\rho_R(x)}\rho_R(x)dx}{\int \frac{\rho(x)}{\rho_R(x)}\rho_R(x)dx} = \frac{\left\langle \mathcal{O}(x)\frac{\rho(x)}{\rho_R(x)} \right\rangle_R}{\left\langle \frac{\rho(x)}{\rho_R(x)} \right\rangle_R}. \tag{2.47}$$

This technique goes by the name of reweighting, see e.g. [37]. The denominator is typically referred to as the average sign. The standard choice is $\rho_R = |\rho|$. Since we are discussing a statistical physics system, we can rewrite the average sign according to

$$\left\langle \frac{\rho(x)}{\rho_R(x)} \right\rangle_R = \frac{\int \rho(x)dx}{\int \rho_R(x)dx} = \frac{Z}{Z_R} = e^{-\frac{V}{T}\Delta f(T)} \tag{2.48}$$

where we used that in statistical physics $Z = e^{-V/Tf(T)}$ with the free energy density $f(T)$. Thus the average sign approaches zero exponentially fast with increasing volume V for fixed temperature. This is a problem since we are usually interested in the infinite volume limit to get rid of finite size effects. If the average sign is close to zero the cancellation of large contributions is again an issue. Hence, to have a constant numerical accuracy the computation cost grows exponentially with volume.

For specific instances the sign problem has been shown to be NP hard[3] [72], it is hence not to be taken lightly. There exist a variety of approaches to solving the sign problem. Here we list just a few of those that are applicable to QCD or QCD like theories.

- Reweighting (see above) and its variation multi-parameter reweighting [73].
- Dual formulations make use of the fact that the sign problem is representation dependent. Hence, one can find model specific coordinate transformations such that the action is real or at least such that the sign problem is weaker. Recent examples include QCD in the strong coupling limit [74, 75] and several other lattice field theories [76–79].
- The density of states method consists of first simulating the theory with (typically) the energy fixed and then numerically integrating over the resulting distribution. This method relies on very precise simulations in the first step. For a recent review see [80], for applications to gauge theories see [81, 82].

[3] NP hardness of an algorithm means that it does scale worse than polynomially with system size, i.e. the complexity of the algorithm cannot be reduced to $O(N^p)$ with $p \in \mathbb{N}$.

- In the complex Langevin method a complex stochastic process is simulated. The equilibrium distribution of this process mimics the Boltzmann factor e^{-S} under certain conditions. It has been successfully applied to different parameter regions of QCD [83, 84]. It is, however, difficult to quantify convergence of the method. We will investigate complex Langevin in much more detail in chapter 3.

- The Lefschetz Thimble method and its generalizations consist of deforming the real manifold in such a way that the integrand of the path integral stays real [85] or imaginary fluctuations are minimal [86, 87]. Due to its complexity so far there has been no application to QCD. We will investigate this method in more detail in chapter 4.

- Much progress exists for the Taylor expansion around $\mu = 0$. This method has brought forward many high precision calculations at the physical point for recent examples see [15, 16, 19]. The Taylor expansion is limited by its radius of convergence. It is expected to break down at the critical point. Currently, it is not reliable beyond $\mu/T = 1$.

- Analytic continuation from imaginary chemical potential also allows for high precision computations at the physical point. With $i\mu \in \mathbb{R}$, the action once again is real and simulations can be performed. One can then use analyticity arguments at $\mu = 0$ to extrapolate from $\mu^2 \leq 0$ to $\mu^2 > 0$. For recent progress see [18, 19, 49]. This method so far is also only reliable in the region $\mu/T < 1$.

Despite this multitude of approaches, there is no general solution to "the sign problem". The general consensus amongst the community is that each sign problem has its own solution. One can only hope to find solutions for classes of theories. The aim of this thesis is to investigate different aspects of the sign problem.

3 Complex Langevin

One popular approach to the sign problem is the complex Langevin method. Historically, it was introduced by Klauder [88–90] and Parisi [91]. However, problems of wrong convergence – as will be investigated further below – showed up rather soon [92–95]. Despite some initial success (see e.g. [96]) and some analysis on the failure of complex Langevin in specific situations [97–101] its popularity stayed limited for a long time. The main issue was the lack of a criterion to identify convergence to the wrong distribution. It has since regained popularity and has successfully been applied to some real time problems [59, 102, 103], models of QCD at finite chemical potential [83, 104–110], the complex Bose gas [111, 112], some fermionic theories [113–117] and ultimately some parameter regions of QCD at finite chemical potential [84, 118, 119]. Simultaneously proofs of convergence and resulting criteria have been developed [29, 120–123]. However, there are some cases in which complex Langevin simply does not work. Popular cases are the 3d XY-model at small chemical potentials [124] as well as QCD at low lattice couplings β [125–127]. While it is usually clear if it does not work, there is no general recipe to make it work. The failure of complex Langevin is related either to poles in the action or to insufficient decay of the distribution of observables, see section 3.1 for a thorough discussion. Some methods circumventing those problems have been developed for QCD [105, 128] and non-relativistic fermionic theories [115].

In this section, we will give a brief and simple introduction to complex Langevin. In section 3.1 we discuss convergence properties of the Complex Langevin method. Section 3.2 contains results and discussions for the application of Complex Langevin to QCD. Finally, in section 3.3 we briefly discuss the applicability of the method to real-time simulations of SU(2) lattice field theory.

Stochastic quantization

The complex Langevin method is the complex generalization of stochastic quantization which was introduced in [129], for an overview, see [130, 131]. The idea behind stochastic quantization is to obtain the Euclidean Boltzmann factor as an equilibrium distribution of a stochastic process. This will be briefly introduced here. First, a fifth coordinate t – called Langevin time – in addition to the four Euclidean coordinates $x = (x_0, x_1, x_2, x_3)$ is introduced. The quantum field then depends on all five variables $\phi(x, t)$. One can define a stochastic process in form of a Langevin

equation as

$$\frac{\partial\phi(x,t)}{\partial t} = -\frac{\delta S}{\delta\phi(x,t)} + \eta(x,t) = K_\phi + \eta(x,t)\,, \tag{3.1}$$

where K_ϕ is called a drift term, η is Gaussian noise with $\langle\eta(x,t)\rangle = 0$ and $\langle\eta(x,t)\eta(x',t')\rangle = 2\delta(x-x')\delta(t-t')$. By stochastic calculus[4] one can derive the corresponding Fokker-Planck equation, which is a differential equation for the probability distribution corresponding to a stochastic process. It is easy to see that in this case it has a stationary solution of the form

$$\lim_{t\to\infty} P(\phi,t) = \frac{e^{-S(\phi)}}{\int \mathcal{D}\phi e^{-S(\phi)}}\,, \tag{3.2}$$

i.e. the Euclidean Boltzmann factor as long as the action is well behaved. Hence, stochastic quantization is an alternative way of quantization. Here we focus on lattice simulations of quantum field theory, a general discussion of stochastic quantization/Langevin dynamics for lattice field theories can be found in [132, 133].
So far everything in this chapter has been discussed for the case of real variables, in the remainder of this chapter, we will refer to this as real Langevin.

Complex Langevin dynamics

In contrast, complex Langevin is obtained by complexifying the field variable. We will demonstrate this for a simple system with only one variable x. In the real case, the integral

$$\langle\mathcal{O}\rangle = \frac{1}{Z}\int \mathcal{O}(x)e^{-S(x)}dx \tag{3.3}$$

can be simulated using stochastic quantization as described above. The Langevin equation simply reads

$$\dot{x} = \frac{\partial x(t)}{\partial t} = -\frac{\partial S}{\partial x} + \eta(t) = K_x + \eta(t)\,. \tag{3.4}$$

This stochastic process gives an estimate of the integral in 3.3 after averaging the observable $\mathcal{O}$ over the whole process after equilibration. Note that $x \in \mathbb{R}$, hence this process is not properly defined if the action $S(x)$ can take complex values. When this happens, the whole space needs to be complexified, i.e. $x \in \mathbb{R} \to z \in \mathbb{C}$. The complex Langevin equation is the same as (3.4) but with the replacement $x \to z = x + iy$. By splitting into real- and imaginary part it reads

$$\begin{aligned}\dot{x} &= -\operatorname{Re}\frac{\partial S}{\partial z} + \eta = K_x + \eta \\ \dot{y} &= -\operatorname{Im}\frac{\partial S}{\partial z} = K_y\,,\end{aligned} \tag{3.5}$$

[4] In this work everything is formulated via Itô-calculus.

where we chose a purely real noise, since complex noise usually introduces stability issues [121]. This is a coupled system of stochastic differential equations, which has an equilibrium distribution of $P(x, y)$. The question that remains to be answered is whether this leads to the same expectation values as the desired path integral measure with the Euclidean action.

A note on discretization and step size

When discretizing the Langevin equation (3.5), a discretization parameter ϵ, which represents the step size in t direction must be introduced. We will typically perform the easiest discretization, i.e. Euler-Maruyama discretization. Only in full QCD will we make use of an improved scheme, since it will speed up the simulations. In principle, this parameter then needs to be extrapolated to $\epsilon = 0$ by performing simulations at different values of ϵ and observing some converging behavior. Mostly we will not be concerned with this limit since there are enough issues to be discussed already before this extrapolation.
A pressing issue in complex Langevin are runaway trajectories. They occur in many models no matter how simple or complex they are when the simulation escapes into an unstable direction. The natural way to circumvent those is to make the step size ϵ very small, however, this drastically increases simulation time and can lead to ergodicity issues if bottleneck regions exist. A better way to tackle this problem is adaptive step size [104, 134], which makes the step size small if a runaway trajectory occurs and forbids the process to go large steps in such directions. We will make use of adaptive step size when necessary, especially in the case of QCD.

3.1 Convergence requirements of Complex Langevin

Let us first give a definition of what correct convergence of Complex Langevin means. There are two probability distributions which are of importance. On the one hand there is the complex path integral measure $\rho(x, t)dx = e^{-S(x)}dx$ with $S \in \mathbb{C}$, which gives rise to the correct evolution for the expectation value

$$\langle O(t)\rangle_{\rho(t)} = \frac{1}{Z}\int \mathcal{O}(x)\rho(x;t)\,dx\,. \tag{3.6}$$

On the other hand there is the equilibrium distribution of the process after complexification $P(x, y; t)$, which gives rise to

$$\langle \mathcal{O}(t)\rangle_{P(t)} = \frac{1}{Z}\int \mathcal{O}(x+iy)P(x,y;t)\,dx\,dy\,. \tag{3.7}$$

Correctness now means that

$$\langle \mathcal{O}\rangle_{P(t)} = \langle \mathcal{O}\rangle_{\rho(t)}\ \forall t\,, \tag{3.8}$$

i.e. the expectation value of any observable agrees under both distributions, provided the initial conditions agree. We choose initial conditions according to equation (A.4). More specifically we choose

$$P(x, y; 0) = \frac{1}{2\pi}\delta(y) \tag{3.9}$$

There are a few requirements which if not fulfilled can spoil this equivalence. Those requirements can be split into two groups. The first group are requirements, which if not fulfilled do not lead to a Langevin time independent equilibrium distribution. While this can be a problem it is also easy to detect. In lattice simulations, we want to simulate in thermal equilibrium and hence, if this is not fulfilled it is clear that some condition was not met. In the real case, those requirements are usually fulfilled and they are established by theorems, see e.g. [131]. In the complex case, it is usually not clear. This has been discussed in [121, 135]. We will not be concerned with those criteria here since it is easy to see a posteriori if the simulation does not converge at all. The other group of requirements covers those, which if not fulfilled lead to convergence to a wrong limit. It is not so easy to detect this and hence will be discussed in more detail. Those requirements are

1. The process must be ergodic.
2. The distribution, observable, drift and action must be holomorphic. If instead, they are meromorphic poles can spoil the convergence.
3. The distribution, observable and drift must decay fast enough far away from the real manifold. We will see below what fast enough means.

The first point is a problem not only in complex Langevin but also for real Langevin and similar simulation methods such as the more commonly used Hybrid Monte Carlo algorithm [136]. See [100, 137–139] for some discussions on ergodicity in real as well as complex Langevin simulations. We will not go into further detail here.

The second point is one of the current issues still plaguing complex Langevin simulations. The occurrence of poles has received lots of attention in recent years [106, 107, 110, 137, 140–143] and it has been suggested that the problem with poles is similar to the problem of slow decay, i.e. if all quantities decay fast enough towards the pole, they are no issues [144]. If such fast decay is not observed, poles typically lead to separated regions in the manifold, spoiling ergodicity. Some possible solutions such as the deformation technique [118, 145] have been proposed. However so far there are not enough results to make a final judgment on a general procedure for dealing with poles.

The approach we take here is to carefully monitor our simulations for the occurrence of poles and avoid parameter sets where the poles affect the simulation. Hence, poles will not be an issue for us, see section 3.2.3.

The third point has been discussed mainly in [120, 121]. There a criterion for correctness was based on so-called “consistency conditions”, which are a measure for

convergence of complex Langevin. However, neither this condition nor the related Schwinger-Dyson equations [120, 146], which are a slightly stronger condition [147] are enough to guarantee convergence without further control. See [59] for a case where the Schwinger-Dyson equations are fulfilled but the evolution converges to the wrong measure. This can only happen if the solution of the Fokker-Plank equation is not unique.

Below we will take a more practical approach. Namely, we will focus on finding a criterion which can be measured without much additional effort in a realistic lattice simulation.

3.1.1 Convergence of complex Langevin and boundary terms

We start by briefly sketching the proof of convergence for the complex Langevin method from [120, 121]. The idea of the proof is to introduce an interpolating quantity $F_{\mathcal{O}}(t,\tau)$ between the two sides of equation (3.8) and to show that its derivative vanishes, here $\tau \in [0,t]$. To that end we define

$$F_{\mathcal{O}}(t,\tau) = \int P(x,y;t-\tau)\mathcal{O}(x+iy;\tau)dxdy\,, \tag{3.10}$$

which gives $\langle\mathcal{O}\rangle_{P(t)}$ at $\tau = 0$ and $\langle\mathcal{O}\rangle_{\rho(t)}$ at $t = \tau$, see appendix A.1 for a proof. Hence, if $F_{\mathcal{O}}(t,\tau)$ is constant for all t and τ equation (3.8) is clearly fulfilled. In other words, we require that

$$\begin{aligned}\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = &\int \left(\partial_\tau P(x,y;t-\tau)\right)\mathcal{O}(x+iy;\tau)dxdy \\ &+ \int P(x,y;t-\tau)\partial_\tau\mathcal{O}(x+iy;\tau)dxdy = 0\,.\end{aligned} \tag{3.11}$$

In appendix A.1 we show via integration by parts of equation (3.11) that this boils down to the vanishing of a boundary term

$$\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = \lim_{Y\to\infty} B_{\mathcal{O}}(Y;t,\tau)\,, \tag{3.12}$$

where

$$\begin{aligned}&B_{\mathcal{O}}(Y;t,\tau) \equiv \\ &\int [K_y(x,Y)P(x,Y;t-\tau)\mathcal{O}(x+iY;\tau) - K_y P(x,-Y;t-\tau)\mathcal{O}(x-iY;\tau)]dx\,.\end{aligned} \tag{3.13}$$

A simple model

In order to gain a deeper understanding of this boundary term we investigate the model

$$\rho(x) = \frac{1}{Z}e^{i\beta\cos(x)}\,, \tag{3.14}$$

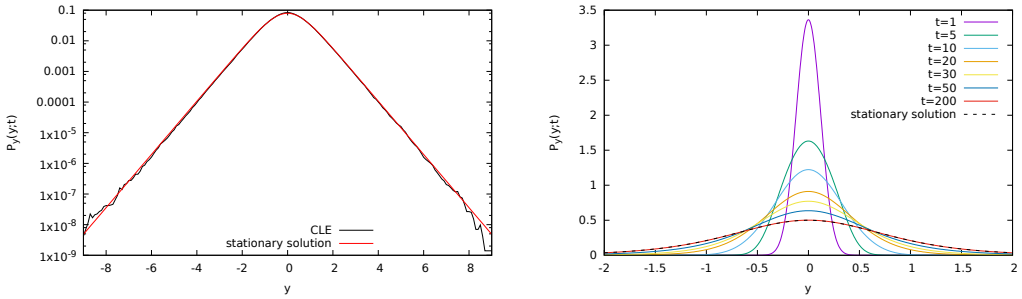

Figure 3.1: Left: Comparison of the equilibrium distribution for the model in equation (3.14) with $\beta = 1$ from complex Langevin simulations (black) with the known solution from equation (3.14) (red). Right: Marginal distribution $P_y(y;t)$ from solving the Fokker-Planck equation for different times t (solid lines) compared with the known solution from equation (3.14) (dashed line).

with $\beta \in \mathbb{R}$. This model has been studied with complex Langevin in [103, 148] and does give wrong results under complex Langevin evolution. The correct expectation values for exponential modes can be obtained in terms of Bessel functions of the first kind

$$\int e^{ikx}\rho(x)dx = (-i)^k \frac{J_k(\beta)}{J_0(\beta)} \neq 0\,. \tag{3.15}$$

For this model the equilibrium distribution $P(x,y)$ is known [149]

$$P(x,y) = \frac{1}{4\pi}\frac{1}{\cosh(y)^2}\,. \tag{3.16}$$

It is remarkable that this solution does neither depend on the coupling β nor on x. By comparison to equation (3.15) it is immediately clear that P cannot give correct results. Furthermore P decays like $e^{-2|y|}$ for large $|y|$ and hence, the integral for all modes with $k > 1$ is not absolutely convergent.
We study this model in two ways: By simulating the complex Langevin equation and by solving the Fokker-Planck equation for $P(x,y;t)$ on a grid. Details on the latter can be found in appendix A.2. The drifts of the complex Langevin equation for the measure (3.14) read

$$\begin{aligned} K_x &= -\beta\cos(x)\sinh(y) \\ K_y &= \beta\sin(x)\cosh(y)\,, \end{aligned} \tag{3.17}$$

the simulation is then done using Newton discretization for equation (3.5). In figure 3.1 left we show that the equilibrium marginal distribution $P_y(y;t\to\infty)$ for $\beta = 1.0$ obtained by complex Langevin agrees with the exact distribution from equation (3.14) over several orders of magnitude.

$\mathcal{O}$	e^{ix}	e^{-ix}	e^{2ix}	e^{-2ix}
CL	0.004(3)	0.002(3)	1.027(22)	1.001(20)
correct	-0.575081i	-0.575081i	-0.150162	-0.150162

Table 3.1: Complex Langevin (real part, imaginary part negligible) and correct results for the model in equation (3.14) with $\beta = 1$.

To show that the simulation gives wrong results, we list expectation values for the first few modes in table 3.1. The expectation values are measured from 100 different runs with randomly chosen starting points and Langevin time up to $t \approx 2500$ and step size $dt \approx 5 \times 10^{-6}$. For $|k| > 2$ we find that expectation values are dominated by noise, but already for $|k| = 2$ the result depends on how the averages are computed, taking measurements only every 0.01 Langevin time also leads to a result dominated by noise, while measuring every dt yields the results in table 3.1. This is because the integral for the higher modes is not absolutely convergent as mentioned above. Hence, in principle we would expect a correct result only from the first mode. However we see that this is also wrong.

Investigation of boundary terms

Since the initial conditions in the complex Langevin and Fokker-Planck simulations were chosen such that equation (3.8) is fulfilled (see (A.4) and equation (3.9)), the question is where both evolutions start do deviate. The right plot in figure 3.1 shows the marginal distribution $P_y(y;t)$ of the real process for different times for $\beta = 0.1$. The distribution starts out as a δ-peak at $t = 0$, which is correct by definition. It then broadens over time and finally converges to the "wrong" distribution in equation (3.14). Looking at the other marginal distribution $P_x(x;t)$ in figure 3.2 we see that $P(x,y)$ does have a small x-dependence at intermediate times, which vanishes for larger times.

In order to see at which time the correct distribution transitions to the wrong one we need to see when the boundary term in equation (3.13) becomes non-negligible, which in this simple model we can see by either looking at the time evolution of $\langle \exp(ix) \rangle$ or at the boundary terms directly. Figure 3.3 left shows Fokker-Planck evolution of the first mode $\langle \exp(ix) \rangle$ (solid black line) compared to the correct evolution (solid black line) [5] (dashed black line). The expectation values from correct- and Fokker-Planck evolution agree up to $t \approx 20$, which suggests that up to that time the deviation of the probability distributions is minor. The boundary term is hard to properly visualize since it depends on three variables already in this simple model. Hence, we make the simplification that we only look at $\tau = 0$. We will

[5] See appendix A.3 for how the correct evolution is obtained.

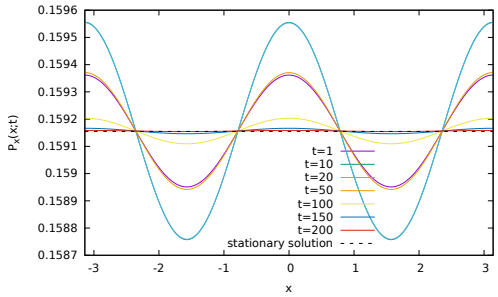

Figure 3.2: Marginal distribution $P_x(x;t)$ for $\beta = 0.1$ at different times t. By choice of the initial condition the distribution starts out constant. It then develops a nontrivial structure and in the limit of large times t becomes constant again.

justify this below. The boundary term at $\tau = 0$ for the first mode is shown in the right plot of figure 3.3 for $\beta = 0.1$ and $Y = 5$. It starts to deviate from 0 at around $t \approx 20$ and hence confirms the suspicion that up to this point the evolution seems to agree. Note that the time at which the evolution starts to deviate depends on the choice of coupling. A larger coupling seems to lead to an earlier deviation, see [29]. For the wrong equilibrium distribution in equation (3.14) the boundary term can be evaluated explicitly. Using equations (3.14) and (3.17) we obtain the boundary term from equation (3.12) at $\tau = 0$ in the limit $t \to \infty$

$$B_k(Y;\infty,0) = -2\beta \int_{-\pi}^{\pi} dx \frac{\sin(x)\cosh(Y)e^{ikx}\sinh(Y)}{4\pi\cosh^2(Y)} \,. \tag{3.18}$$

Which does converge only for $|k| = 1$

$$B_{\mp 1}(Y;\infty,0) = \mp\frac{i\beta}{2}\tanh(Y)\,, \tag{3.19}$$

and is also plotted in figure 3.3, where one can see that the Fokker-Planck evolution again converges to the stationary solution.
In the end, the limit $Y \to \infty$ has to be taken, see equation (3.12). To this end we repeat the analysis of the boundary term for different values of Y to look for convergence. This is demonstrated in the left plot of figure 3.4. We see that the deviation between $Y = 5$ and $Y = 6$ is rather small, and hence our earlier choice of $Y = 5$ can be regarded as already being in the limit $Y \to \infty$. This can also be seen by looking at equation (3.19) which is very close to its asymptotic value for $Y = 5$. The right plot of figure 3.4 shows that the Langevin simulation actually yields the boundary terms following equation (3.19). This can be obtained by looking at the histogram of $K_y\mathcal{O}$, since the histogram already encodes the distribution.

To justify only looking at the boundary term for $\tau = 0$, we need to look at $F_{\mathcal{O}}(t,\tau)$ directly as shown in figure 3.5. First of all, as expected $F_{\mathcal{O}}(t,\tau)$ is approximately

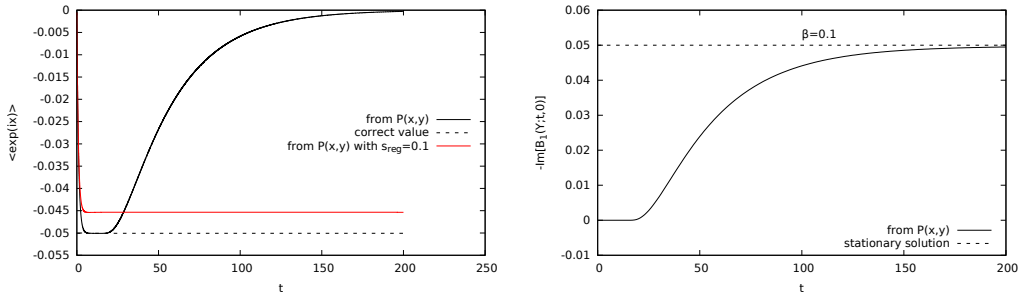

Figure 3.3: Left: Comparison of Im $[\langle \exp(ix) \rangle]$ using the Fokker-Planck evolution of P (solid black line) equation (3.7) with the L_c evolution of the observables (dashed black line – equation (3.6) for $\beta = 0.1$. For small times $t < 20$ both evolutions are practically indistinguishable. The red line shows a regularized version of the action. Right: Boundary term for $\beta = 0.1$ and $Y = 5$ directly computed from the Fokker-Planck evolution (solid line) and the asymptotic value for $t \to \infty$ from equation (3.19) (dashed line).

constant for times $t < 20$. Secondly we see that its slope – i.e. the boundary term – is always maximal at $\tau = 0$. This observation is the reason why we are allowed to choose $\tau = 0$ in equations (3.18) and (3.19). Note that the evaluation of $F_{\mathcal{O}}(t, \tau)$ from equation (3.10) requires knowledge of $\mathcal{O}(z; t)$ from the exact evolution as described in appendix A.3. This evolution reaches it asymptotic limit $\langle \mathcal{O} \rangle_c$ approximately at $\tau > 7$. Hence, for large τ it is just a constant that can be pulled out under the integral and for large τ we arrive at

$$F_{\mathcal{O}}(t, \tau) \approx \langle \mathcal{O} \rangle_c \int P(x, y) dx dy = \langle \mathcal{O} \rangle_c \,, \tag{3.20}$$

which is $\langle \mathcal{O} \rangle_c = 0.500626i$ for the first mode at $\beta = 0.1$.

Relation to tails in the distribution

Some works stated that wrong convergence can be identified by looking for large tails in the distribution of the observable [120, 121]. This can be related to the boundary terms in the following way: We can regard equation (3.18) in the limit of large Y as the probability density of

$$v(Y) \equiv \mathrm{Im} \int dx K_y(x, Y) \mathcal{O}_1(x - iY) \sim e^{2Y} \,. \tag{3.21}$$

Then the statement

$$\lim_{Y \to \infty} v(Y) P_y(Y) \neq 0 \,, \tag{3.22}$$

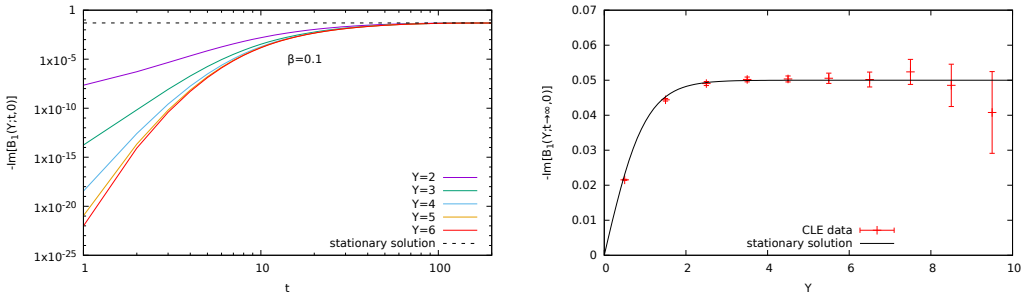

Figure 3.4: Left: Boundary term from equation (3.13) at $\tau = 0$ for different values of Y. Right: Stationary behavior of the boundary term as a function of Y from the analytic solution in equation (3.19) (solid line) compared to measurements from Langevin simulations (red error bars).

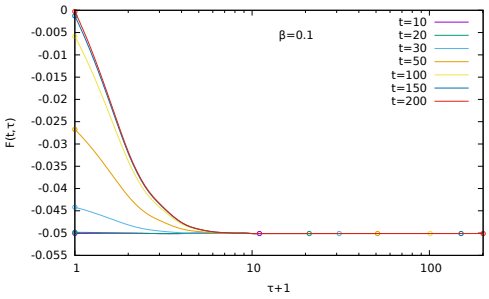

Figure 3.5: $\mathrm{Im}\left[F_{\mathcal{O}}(t,\tau)\right]$ from equation (3.10) for the observable $\mathcal{O} = \langle \exp(ix) \rangle$ and $\beta = 0.1$ for different times t. The small circles denote the beginning and end of the respective curves.

encodes the non-vanishing of the boundary term. By instead looking at the distribution of $v(Y)$ itself

$$p(v) = P_y(Y(v))\frac{dY}{dv}\,, \tag{3.23}$$

we realize that

$$p(v) \sim v^{-2}\,. \tag{3.24}$$

Hence, $vp(v)$ is not integrable. This means that a "skirt" in the distribution of $K_y\mathcal{O}$ which has power -2 or lower will have a non-vanishing boundary term. This statement was only possible in the case of no x dependence, otherwise the whole discussion becomes two dimensional.

We can also directly look at the distribution of the observable $\mathrm{Im}\left(\exp(i(x+iy))\right)$

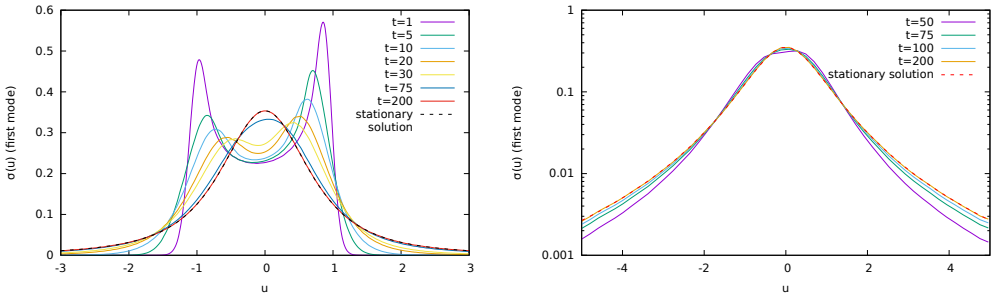

Figure 3.6: Evolution of the distribution of the first mode $\sigma(u)$ for different times t. Left: Linear scale. Right: Log-scale, the approach to the stationary solution is more visible here.

for our model by looking at its histogram

$$\sigma(u;t) \equiv \int dxdyP(x,y;t)\delta(\sin(x)e^{-y} - u)\,, \tag{3.25}$$

which can be evaluated for equation (3.16)

$$\sigma(u;\infty) = \frac{1}{2(1+u^2)^{3/2}}\,. \tag{3.26}$$

It can also be computed from the Fokker-Planck evolution by introducing a binning in the observable, i.e. we sum over all $P(x,y)$ that correspond to a certain value $u = \mathcal{O}(x,y)$. The result for the first mode is shown in figure 3.6. Initially, the observable has a double peak structure and decays exponentially, however for larger t the double peak structure disappears and tails develop. Hence, we conclude that the occurrence of boundary terms and tails in the distribution of observables are intimately linked.

Comparison the drift distribution

Another criterion that has been derived directly in the discretized formulation in [122] is concerned with the distribution of the drift term K_y. It states that the distribution of the drift term has to decay at least exponentially strong. Since this criterion is quite simple, we can easily check it from the Fokker-Planck evolution as well as from complex Langevin simulations directly using the same methodology that was used to compute the histogram of the observable. The result is shown in figure 3.7, where it is clear that the drift decays with a power law for all times t and hence violates the criterion from [122]. It is interesting though that initially, the drift decays with a large power – which corresponds to the small boundary terms in that region. Hence, it seems that both criteria – the decay of the drift and the boundary term – signal a lack of correct convergence in this model.

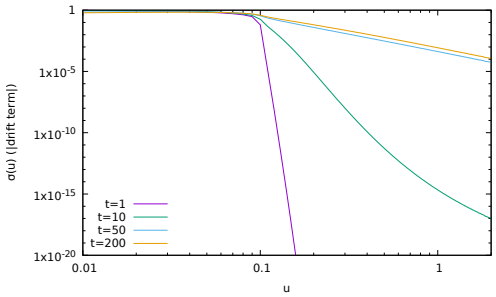

Figure 3.7: Histogram of the Langevin drift for different times at $\beta = 0.1$. See text for a discussion.

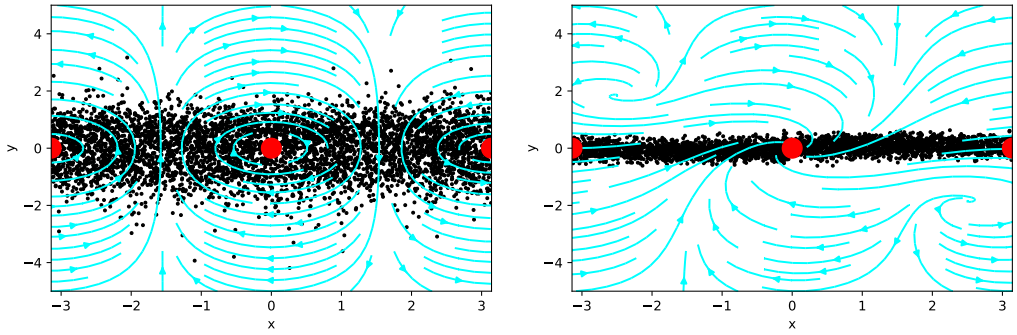

Figure 3.8: Visualization of the Langevin flow (stream lines), the fixed points (red dots) and the actual Langevin evolution (black dots) for $\beta = 0.1$. Left: Without regularization. Right: With regularization parameter $s_z = 0.1$.

Restoring convergence in the simple model

It is known quite well that failure of convergence is usually related to large deviations or runaways in the complex direction of complex Langevin simulations, see e.g. [121]. Those deviations occur if there are repulsive direction close to the real manifold. In some cases this can be resolved by using adaptive step size [104, 134], such that the drift force adapts itself if a strong repulsion exists. In figure 3.8 left we visualize the trajectory the Langevin process takes. The blue flow lines represent the classical flow without noise. There are no attractive fixed points near the real axis but some repulsive directions. The fixed points (red dots) are also not attractive, but circular. This means that the process has no incentive to stay near the real manifold and will make excursions further from it. In this model the excursions can be circumvented by adding a regularizing term to the action [29, 115], which is similar in spirit to dynamical stabilization [128] – which will be discussed further in the context of gauge theories below. One can regularize here in different ways. Adding $K_y^R = -s_y y$ to the drift in the imaginary direction introduces an attractive force

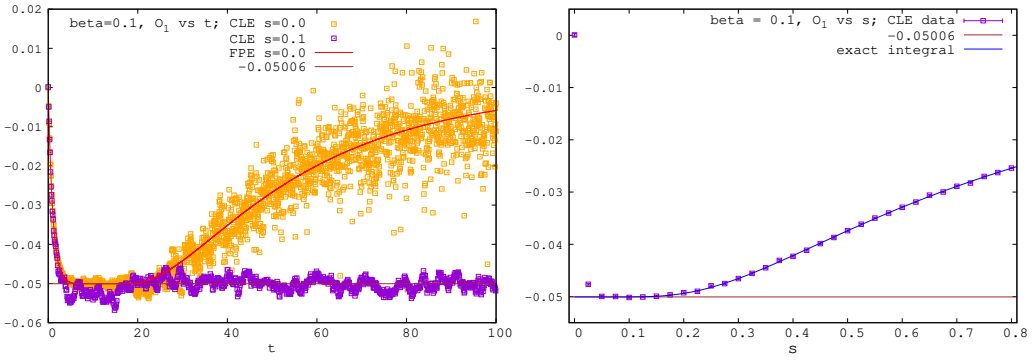

Figure 3.9: Left: Comparison of the complex Langevin evolution of $\mathrm{Im}(\exp(ix))$ with (purple) and without (orange) regularization ($s_z = 0.1$) for $\beta = 0.1$ as a function of time. The red line shows the Fokker-Planck solution without regularization, the purple line shows the correct result for the regularized model. Here the regularization leads to correct convergence. Right: Longtime expectation value of $\mathrm{Im}(\exp(ix))$ for $\beta = 0.1$ for different regularization strengths.

towards the real manifold. This term also leaves the periodicity of the real manifold intact but cannot be expressed as part of a holomorphic action with $-\partial_z S = K$. We did solve the Fokker-Planck equation with such regularization for $\beta = 0.1$ and $s_y = 0.1$, see the red curve in figure 3.3 left. This regularization term makes the dynamics stable but introduces a large deviation from the original theory. Another way of regularizing is to add a term $S_R = -s_z z$. This leaves the action holomorphic but destroys the periodicity of the real manifold. We illustrate the effect of such a term in the right plot of figure 3.8. As one can see such regularization removes the repulsive directions from the real manifold and makes the fixed points attractive and as a result, the evolution is confined to a much thinner strip around the real manifold. Depending on the choice of parameters this can make the expectation values correct. In the left plot of figure 3.9 we show the Langevin evolution of the regularized and the unregularized model from many different Langevin trajectories. The evolution of the regularized model at $\beta = 0.1$ closely follows the correct expectation value in the case of $s = 0.1$. However, at larger regularization parameters this is no longer the case. The evolution still is stable, but the expectation values deviate, see the right plot of figure 3.9.
For larger couplings β, the plateau region in the unregularized model becomes much smaller or vanishes at all. In this case, the regularization does also stabilize but the expectation values of the regularized model no longer agree with the original model for any choice of regularization parameter [29].

The comparison of the regularized and unregularized model is a good test-bed for the boundary term and drift criterion. Figure 3.10 shows both criteria and compares

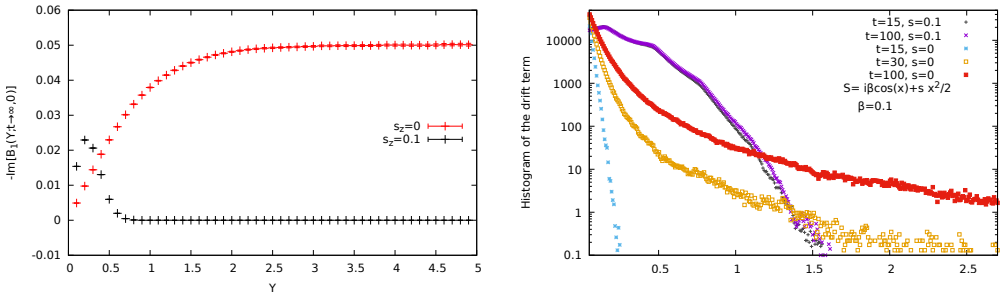

Figure 3.10: Left: Boundary term for $\beta = 0.1$ at $t \to \infty$, while there is a boundary term for $s_z = 0$ it vanishes for $s_z = 0.1$. Right: Histogram of the drift at different times t for $s_z = 0$, where the decay is slower than exponential and $s_z = 0.1$ where the decay is stronger than exponential.

the regularized case $s_z = 0.1$ with the unregularized case $s_z = 0$ for $\beta = 0.1$. In both cases, the boundary term goes to a plateau and hence the limit $Y \to \infty$ can be taken easily. It becomes clear that the regularization removes boundary terms completely. Similarly, the drift term decays weaker than exponential for the unregularized case suggesting the failure of complex Langevin. For $t = 15$ this is not so clear, however comparing with the drift histogram from the Fokker-Planck evolution (figure 3.7) shows that the decay should be with a high power polynomial. In case of regularization, the decay of the drift decays stronger than exponential suggesting that complex Langevin works.
The conclusion here is that both criteria perform equally well.

3.1.2 A more practical way to compute boundary terms

Computing boundary terms according to equation (3.13) is practically impossible in high dimensional models. This can, however, be simplified. In a practical simulation, we are always interested in expectation values after thermalization, essentially $t \to \infty$. Hence, to compute the boundary terms according to equation (3.12) we have to take an additional limit. However, one may wonder about the order, i.e. whether the limit $Y \to \infty$ or $t \to \infty$ has to be taken first. The boundary term (equation (3.12) with (3.11)) without the limit $Y \to \infty$ can be rewritten as (cf. equation (A.12))

$$\begin{aligned}
&\frac{\partial}{\partial\tau} F_{\mathcal{O}}(t,\tau) = \\
&-\int_{-Y}^{Y} \left(\frac{\partial}{\partial t} P(x,y;t-\tau)\right) \mathcal{O}(x+iy;\tau)dxdy \\
&+\int_{-Y}^{Y} P(x,y;t-\tau) L_c \mathcal{O}(x+iy;\tau)dxdy = 0\,.
\end{aligned} \tag{3.27}$$

Where as before we will take $\tau = 0$, since there the boundary term, i.e. the slope in figure 3.5, is maximal. The first term vanishes if a stationary solution of the Fokker-

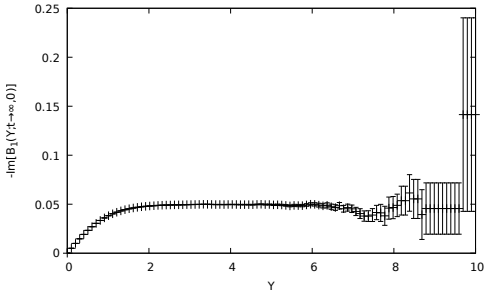

Figure 3.11: Boundary term of the U(1) model computed via $\langle L_c e^{ix} \rangle$.

Planck equation does exist. The second term is just the average of $L_c\mathcal{O}$ constrained to the interval $[-Y, Y]$ such that

$$B_{\mathcal{O}}(Y; t \to \infty, \tau = 0) = \langle L_c \mathcal{O} \rangle_Y \ , \tag{3.28}$$

with L_c defined in equation (A.7). In the limit $Y \to \infty$ this is just the "consistency condition" from [120]. A careful look at the Y dependence shows that in the limit the observable becomes very noisy to the extent that it is almost always consistent to zero within large error bars. Figure 3.11 shows the boundary term for the U(1) model and the observable e^{ix} for $\beta = 0.1$ without regularization. The equation of the observable for the boundary term is

$$L_c e^{iz} = (1 + i\beta \sin(z)) e^{iz} \, . \tag{3.29}$$

The boundary term computed this way agrees with figure 3.10. However, for large Y the boundary term becomes very noisy such that it is approximately consistent with zero. Hence, by the criterion from [120] one could conclude proper convergence without further control. This shows that not only the limit $Y \to \infty$ is important but also how the limit is approached. From this simple case, we conclude that the development of a plateau before large error bars kick in gives the value of the boundary term. If this is nonzero complex Langevin does not converge to the correct result.

Higher dimensions

So far everything was focused on a simple integral with one variable. Here we will compute the boundary terms in an exactly solvable higher dimensional model, the SU(3) Polyakov chain. The model consists of a chain of length N_t where points are connected with SU(3) matrices. Its action reads [105]

$$-S = c_+ \mathrm{Tr} U_1 \dots U_{N_t} + c_- \mathrm{Tr} U_{N_t}^{-1} \dots U_1^{-1} \, , \tag{3.30}$$

with $c_\pm = \beta + \kappa e^{\pm\mu}$. When $\mu \neq 0$ or any of the parameters are chosen to be complex, this will not stay in SU(3), hence for complex Langevin simulations one complexifies

to $\mathrm{SL}(3,\mathbb{C})$, which are all 3×3 matrices with $\det U = 1$. This model can be gauge fixed such that only one link remains. This one link can then be diagonalized, and only two variables remain[6]. In this case the remaining link has the form

$$U = \mathrm{diag}\left(e^{i\omega_1}, e^{i\omega_2}, e^{-i(\omega_1+\omega_2)}\right) , \tag{3.31}$$

and the action reads

$$-S = c_+\left(e^{i\omega_1} + e^{i\omega_2} + e^{-i(\omega_1+\omega_2)}\right) + c_-\left(e^{-i\omega_1} + e^{-i\omega_2} + e^{i(\omega_1+\omega_2)}\right) . \tag{3.32}$$

Since this is a variable transformation from $\mathrm{SL}(3,\mathbb{C}) \to \mathbb{C}^2$ one has to explicitly compute the Haar-measure and include it in the action. The additional term is

$$-S \to -S + \ln\left[\sin^2\left(-\frac{2\omega_1+\omega_2}{2}\right)\sin^2\left(\frac{\omega_1-\omega_2}{2}\right)\sin^2\left(\frac{\omega_1+2\omega_2}{2}\right)\right] . \tag{3.33}$$

Details on the Langevin simulation in both prescriptions can be found in appendix A.4.

For the gauge-fixed, diagonalized model it is still possible to compute the boundary terms in both ways. I.e. following equation (3.13) or (3.28). For the explicit computation equation (3.13) needs to be generalized. Dropping t and τ dependence for now the boundary term reads

$$\int\int \vec{\nabla}_y \left(\vec{K}_y P(\vec{x},\vec{y})\mathcal{O}(\vec{x}+i\vec{y})\right) d\vec{x}d\vec{y}, \tag{3.34}$$

where $\vec{x}$ and $\vec{y}$ are vectors containing the real- and imaginary part of the variables ω_i. This can be further simplified by using Gauss theorem

$$\int\int \left[\left(\vec{K}_y P(\vec{x},\vec{y})\mathcal{O}(\vec{x}+i\vec{y})\right)\cdot\vec{n}_y\right] d\vec{x}dS_y . \tag{3.35}$$

We choose the surface S_y to be a square with side length $2Y$ around the origin with $Y = \max(|y_1|,|y_2|)$. Hence, the boundary term becomes

$$\begin{aligned}&\int\int K_{y_1} P\mathcal{O}|_{y_1=+Y} d\vec{x}dy_2 - \int\int K_{y_1} P\mathcal{O}|_{y_1=-Y} d\vec{x}dy_2 \\ &+\int\int K_{y_2} P\mathcal{O}|_{y_2=+Y} d\vec{x}dy_1 - \int\int K_{y_2} P\mathcal{O}|_{y_2=-Y} d\vec{x}dy_1 .\end{aligned} \tag{3.36}$$

In practice this is done by looking at the histogram of Y, which encodes $P(Y)$, the x-dependence is dropped here. For a given interval $[Y - dY/2, Y + dY/2]$ one computes the probability of this interval, i.e. the ratio of the number of Y in this interval compared to the total number $N_{\mathrm{bin}}/(N_{\mathrm{total}}dY)$. Then one averages $PK_y\mathcal{O}$ over the bin, i.e.

$$\int K_Y P\mathcal{O} = \frac{1}{N_{\mathrm{bin}}}\sum_{\mathrm{bin}} P_{\mathrm{bin}} K_Y\mathcal{O} = \frac{1}{N_{\mathrm{total}}dY}\sum_{\mathrm{bin}} K_Y\mathcal{O} , \tag{3.37}$$

[6] There are three diagonal entries, but the condition for $U \in \mathrm{SU}(3)$ that $\det U = 1$ reduces the number of variables such that two remain.

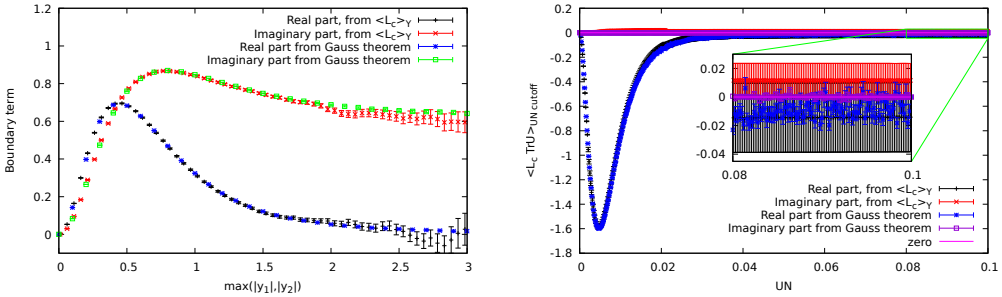

Figure 3.12: Boundary term for the gauge-fixed, diagonalized Polyakov Chain. Left: $\beta = i$, $\kappa = 0 = \mu$, complex Langevin does not work. Right: $\beta = 2$, $\kappa = 0.1$, $\mu = 1$. In both cases the discussed methods to compute the boundary terms agree.

The only subtlety is that there are terms with y_1 and y_2 in the boundary terms. This is solved by choosing a surface, we always choose $K\mathcal{O}$ belonging to the current $Y = \max(|y_1|, |y_2|)$. Having thus computed $\int (KP\mathcal{O})(Y) = F(Y)$ the only thing that remains is to compute $F(Y) - F(-Y)$ due to the difference in equation (3.36). The computation according to equation (3.28) is much easier. Choosing $\mathcal{O} = TrU = e^{i\omega_1} + e^{i\omega_2} + e^{-i(\omega_1+\omega_2)}$ one has to average has to average

$$\begin{aligned}\langle L_c \mathrm{Tr}U\rangle &= \left(\vec{\nabla} + \vec{K}\right)\vec{\nabla}\mathrm{Tr}U \\ &= -\left(e^{i\omega_1} + 2e^{-i(\omega_1+\omega_2)} + e^{i\omega_2}\right) \\ &+ iK_{\omega_1}\left(e^{i\omega_1} - e^{-i(\omega_1+\omega_2)}\right) + iK_{\omega_2}\left(e^{i\omega_2} - e^{-i(\omega_1+\omega_2)}\right), \end{aligned} \tag{3.38}$$

with the drifts K_{ω_i} given in appendix A.4. For this model one can easily check if complex Langevin yields correct results, since the two-dimensional integral can be done numerically. Choosing $\beta = 2$, $\kappa = 0.1$ and $\mu = 1.0$ The correct result is $\langle \mathrm{Tr}U\rangle = 2.0957$. The simulation yields $\langle \mathrm{Tr}U\rangle = 2.0955(13)$. Hence, for that set of parameters complex Langevin yields correct results. Accordingly, the boundary terms plateau to zero up to step size artifacts due to step size of 10^{-4}.
If instead we choose $\beta = i$, $\kappa = 0 = \mu$, the correct result is $\langle \mathrm{Tr}U\rangle = -0.664 + 0.793i$ whereas the simulation yields $\langle \mathrm{Tr}U\rangle = -0.4809(6) + 0.5968(5)i$. Here complex Langevin clearly fails to converge properly. This is also reflected in the boundary terms in figure 3.12. While the real part goes to zero, the imaginary part saturates at a nonzero value. The comparison between both methods for computing boundary terms shows that they agree, however, the computation via $\langle L_c\mathcal{O}\rangle$ becomes noisy faster at larger Y.

Next, we generalize the boundary terms according to equation (3.28) to the full Polyakov chain, which is a first test in a field theoretical setting and thus an important test in a more realistic case. Since the real manifold is SU(3) with the

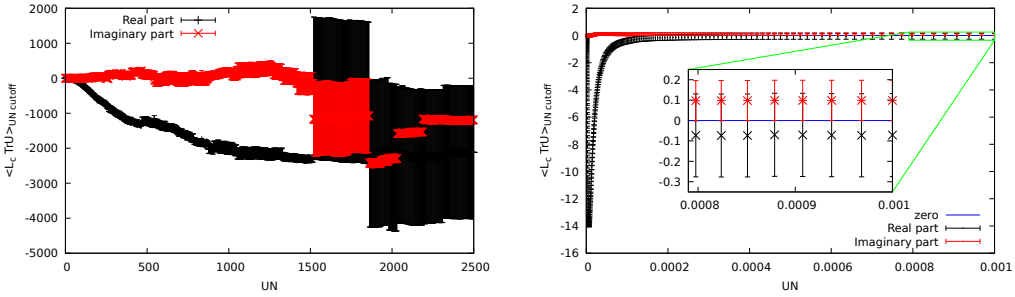

Figure 3.13: Boundary term for the full Polyakov-chain for $\beta = 2$, $\kappa = 0.1$ and $\mu = 1$. Left is without gauge cooling, where boundary terms are expected, right is with gauge cooling, where they should vanish.

property that for $U \in \mathrm{SU}(3)$ $U^\dagger U = 1$ but we lose this property when complexifying to $\mathrm{SL}(3, \mathbb{C})$, it makes sense to define the distance via the so called unitarity norm, which for SU(N) reads

$$d_U = \mathrm{Tr} U^\dagger U - N \,, \tag{3.39}$$

and becomes zero for SU(N) matrices. Choosing the unitary norm for Y we can apply equation (3.28). As an observable we choose the Polyakov loop P, which is just the trace of the product of all the links in order, yielding

$$\begin{aligned} L_c \mathrm{Tr} P &= \sum_{j=0}^{N_t - 1} \left(D_a^j + K_a^j \right) D_a^j \mathrm{Tr} P \\ &= -2 N_t \frac{N^2 - 1}{N} \mathrm{Tr} P + i \sum_{j=1}^{N_t} K_a^j \mathrm{Tr} \left(U_1 \dots \lambda_a U_j \dots U_{N_t} \right) , \end{aligned} \tag{3.40}$$

Remember that for $\beta = 2$, $\kappa = 0.1$ and $\mu = 1.0$ the correct result for the Polyakov loop P is $\langle \mathrm{Tr} P \rangle = 2.0957$. The simulation at a constant step size $\epsilon = 10^{-5}$ yields $\langle \mathrm{Tr} P \rangle = 6.09(2) - 0.04(1)i$. During the simulation the unitarity norm took values up to $d_U = 3e3$, so this deviation is to be expected. The left plot of figure 3.13 shows the resulting boundary terms which clearly signals wrong convergence.
For the Polyakov chain, it is hard to find a set of parameters which give correct results unless they are chosen such that the simulation stays on the real manifold. Hence, we resort to gauge cooling [105]. This procedure is described in A.7. The idea of the method is to do gauge transformations in the opposite direction of the gradient of the unitarity norm such that the rise of the unitarity norm is slowed down or stopped. With gauge cooling and a step size of 10^{-6} the expectation value of the Polyakov loop becomes $\langle \mathrm{Tr} P \rangle = 2.0961(9)$ suggesting that convergence has been restored. This is also reflected in the boundary terms in the right plot of figure 3.13, which is consistent with zero.

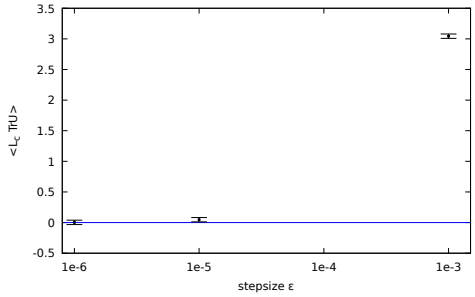

Figure 3.14: Step size dependence of the boundary term in a case where it should be identically zero.

Step size dependence of the boundary terms

In Langevin simulations of lattice field theories, in principle one has to extrapolate the step size to zero. Here we are not interested in the $\epsilon \to 0$ limit since our simulations have strong systematic errors due to a finite volume. Hence we simply use small step sizes. In the case of the boundary terms, this can be a problem, since at least in the Polyakov chain they are really noisy and fluctuate around zero up to $\mathcal{O}(10^2)$. This means that a deviation from a small step size artifact can be inflated considerably. We demonstrate this in figure 3.14 for $\beta = 2$, $\kappa = 0$, $\mu = 0$ where we regularly reunitarize the SU(3) matrices. This case is a real Langevin simulation and should not show boundary terms. The figure only shows the average over the full range, since Y is zero for all values, due to reunitarization. As one can see at large step size the boundary term has a large deviation from zero which vanishes for small enough step sizes. This needs to be taken into account when judging whether a boundary term is there or not.

3.1.3 Conclusion on boundary terms

We have given a detailed explanation of what happens when complex Langevin fails to converge due to the occurrence of boundary terms. To that end, we have thoroughly investigated a simple model and computed its boundary terms using complex Langevin as well as the Fokker-Planck formulation. We were also able to formulate an observable which can be measured on the lattice and signals wrong convergence, see equation (3.28). We related the boundary terms to skirts of the distribution – which is typically used to check for convergence – and found that a non-vanishing boundary term leads to skirts.

3.2 Application to QCD

Initially, lack of convergence prohibited the application of complex Langevin to QCD at finite chemical potential μ. However, since its revival, some methods to improve

convergence have been introduced. Notable methods are the use of adaptive step size [104, 134], gauge cooling [105] and dynamical stabilization[7] [128], see appendix A.7 for a brief explanation of the last two. In our simulations for QCD, we make use of adaptive step size and gauge cooling, which respects gauge symmetry and does not alter the theory. So far we have not used dynamical stabilization, which alters the dynamics similar to regularization in the U(1) model in the previous section.
This section is split into two parts. The first one is the use of an expansion scheme, which allows for fast simulations. The second part is concerned with the simulation of full QCD with $N_f = 2$ degenerate fermion flavors.
Note that for our simulations of gauge theories using complex Langevin we always monitor the unitarity norm given in equation (3.39) and discard all lattice configurations after it grows too large, i.e. after $d_U = 0.1$ is reached for the first time. We will explain why this suffices in the context of QCD in section 3.2.3.
For the simulations in this section in the expansion and full QCD, we used a GPU code provided by Dénes Sexty for the production of the lattice configurations. For the HDQCD simulations, we used a CPU code by the author.

3.2.1 HDQCD and hopping expansion

The simulation of full QCD is a numerically expensive endeavor. The bottleneck of lattice simulations with fermions is the inversion of the fermion matrix. Simulations may be sped up by using the hopping parameter expansion (see e.g. [62] for a textbook). There have been several works employing different version of such expansions with aim to study the phase diagram of QCD [83, 108, 150–156]. Here we will further investigate the systematic approach brought forward in [108]. Given the full action

$$S = S_{\mathrm{YM}} + S_{\mathrm{ferm}} = S_{\mathrm{YM}} - \log \det M \,, \tag{3.41}$$

the fermion matrix M from equation (2.31) and below can be rewritten as

$$M = 1 - \kappa_s S - R \,, \tag{3.42}$$

with the spatial part

$$S(n,m) = 2 \sum_{i=1}^{3} \left(\Gamma_{-i} U_i(n) \delta_{m,n+\hat{i}} + \Gamma_{+i} U_i^{-1}(m) \delta_{m,n-\hat{i}} \right) \,, \tag{3.43}$$

and the temporal part

$$R(n,m) = 2\kappa_t \left(e^{\mu} \Gamma_{-4} U_4(n) \delta_{m,n+\hat{4}} + e^{-\mu} \Gamma_{+4} U_4^{-1}(m) \delta_{m,n-\hat{4}} \right) \,, \tag{3.44}$$

[7] Dynamical stabilization is a modification of the Langevin process. A term is added to the drift, which makes the SU(N) manifold attractive. The only issue with dynamical stabilization is that the proof of convergence for complex Langevin no longer goes through since the drift is modified in a non-holomorphic fashion. However, first tests show a strong improvement in the stability of the simulations upon the use of this method.

where we introduced the projectors

$$\Gamma_{\pm\nu} = (1 \pm \gamma_\nu)\,. \tag{3.45}$$

A further rewriting

$$M = (1-R)\left(1 - \frac{1}{1-R}\kappa_s S\right) \tag{3.46}$$

allows for factorization of the determinant

$$\det M = \det(1-R)\exp\left(\sum_{n=1}^{\infty} -\frac{\kappa_s^n}{n}\mathrm{Tr}\left(\frac{1}{1-R}S\right)^n\right)\,, \tag{3.47}$$

where the spatial part was exponentiated, $\log\det A = \mathrm{Tr}\,\log A$ was used and the logarithm was expanded. The first determinant can be computed analytically [104, 150]. Both, the fermionic part as well as the terms in the expansion of the spatial determinant now only depends on Wilson lines and Wilson loops[8] and no longer require expensive matrix inversions. To zeroth order, this expansion is known as heavy dense QCD (HDQCD) which is the limit of heavy mass such that fermions cannot move in spatial directions on the lattice but temporal movement is allowed. In HDQCD complex Langevin in combination with gauge cooling allows for simulation of the full phase diagram [83].

Brief analysis of the failure of complex Langevin at low β

It is however known that at lattice coupling $\beta < 5.7$ complex Langevin no longer gives correct results even in HDQCD as was determined [105] by comparing to reweighting.
In figure 3.15 we show the evolution of the spatial part of the plaquette, the simplest observable on the lattice. Parameters are $\beta = 5.4$, $\kappa = 0.12$, $\mu = 0.85$ on a 6^4 lattice as in [105] where complex Langevin was shown to fail even with gauge cooling. The figure shows the long term complex Langevin average (green), the long term reweighting average (blue line) – which is expected to be correct in this case – and the evolution of the observable during the simulation. One can see that initially, the simulation approaches the correct value. However, as soon as the correct value is reached the process goes to another value and stays there. This behavior is very similar to the left plot in figure 3.3 for the simple U(1) model. Hence, in analogy, we conclude that even with gauge cooling the real manifold is not attractive enough and excursions in the complex direction lead to wrong convergence.

Note that in HDQCD this is generally not an issue since one can just simulate at larger β, where correct convergence is observed. We will, however, take this lesson to full QCD as well as higher orders in the expansion and stay at larger β.

[8] A Wilson line is the product of link variables that form a path. A Wilson loop is a closed Wilson line.

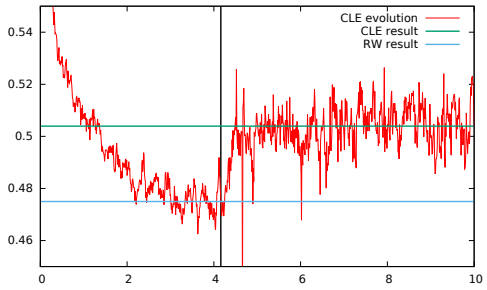

Figure 3.15: Langevin time evolution of the spatial plaquette at $\beta = 5.4$, $\kappa = 0.12$, $\mu = 0.85$ on a 6^4 lattice. The green line represents the long term complex Langevin result, the blue line shows the same but for reweighting, which in this parameter region is expected to give correct results.

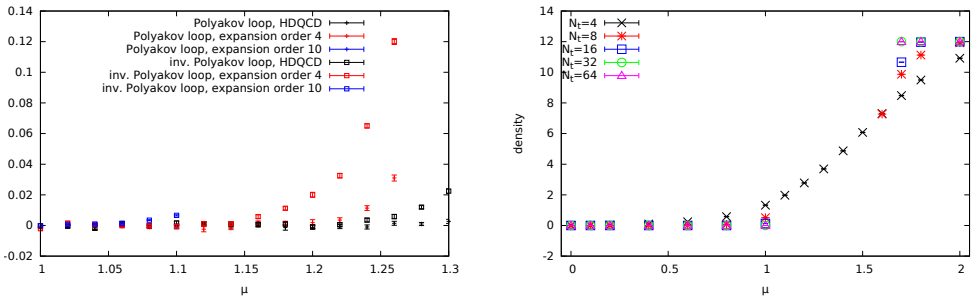

Figure 3.16: Left: Polyakov loop at $\beta = 6.0$, $\kappa_s = 0.12 = \kappa_t$, $\mu = 0.85$ on a $10^3 \times 32$ lattice for different orders of the expansion. The onset of the transition moves to the left for higher orders. Right: Density for same parameters all at order $N = 10$ for different inverse temperatures N_t. A non-resolvable region occurs at lower temperatures (see text for details).

Higher orders

We aim to simulate at low temperature to resolve the nuclear transition. In [108] it has been shown that the expansion for sensible parameters converges around order $N = 10$, which we will assume to be the case in general and investigate further below. To see the effect of the higher orders we simulate $\kappa_s = 0.12 = \kappa_t$, $\beta = 6.0$, spatial volume $N_s^3 = 10^3$ and temporal extend $N_t = 32$ for orders $N = 0, 4, 10$. The left plot of figure 3.16 shows the resulting Polyakov loops. One can see that there is a significant difference in the onset of the phase transition depending on the order. As a next step, we try to map out the phase diagram of κ_s expanded QCD at order $N = 10$ by varying N_t for different temperatures. Choosing the same parameters as above, we show the resulting density in the right plot of 3.16. Unfortunately for μ in the phase transition region, the simulation becomes very slow and it is practically unfeasible to get any results for those points.

Hence, we conclude that the κ_s expansion is not a good way to investigate the

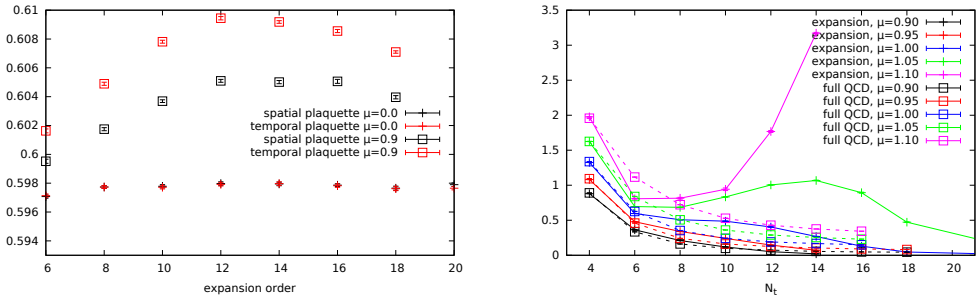

Figure 3.17: Lack of convergence of the κ_s expansion, see text for details. Left: Convergence of the κ_s expansion at $\beta = 6.0$, $\kappa_s = 0.12 = \kappa_t$ on a 10^4 lattice Right: Comparison of 10th order κ_s expansion to full QCD.

phase diagram of QCD. To see why, we take a look at how well the expansion converges. This is investigated in figure 3.17. The left plot shows the dependence of the plaquette on the order of the expansion. While for $\mu = 0$ it is practically converged for $N = 10$ in agreement to [108], for $\mu = 0.9$ convergence has not been reached at even $N = 20$. This suggests that one needs to go even higher or that convergence is not possible at all. The right plot shows a comparison of full QCD with 10th order κ_s expansion close to the non-resolvable region. One can see that the expansion describes QCD less and less while approaching the border. We will see later in the case of full QCD that the eigenvalues of the fermion matrix become rather small and it even becomes difficult to invert the fermion matrix in some cases. The breakdown of the expansion seems to be affected by that as well, just much earlier [9].

Complex Langevin at $\mu = 0$

Before we turn to full QCD it is instructive to compare the complex Langevin method at $\mu = 0$ to the real Langevin method, where the only difference is regular reunitarization of the SU(3) matrices in the case of real Langevin. We make this comparison due to repeated claims in the recent years, that complex Langevin does not even yield correct results at $\mu = 0$, where no complex values should occur [119, 126, 157]. The reason why the process even starts to deviate from the real manifold is numerical instability. In standard lattice simulations, regular reunitarization is a necessary feature to avoid this. In complex Langevin however, we are not allowed to reunitarize but only to use gauge cooling, which has a similar but weaker effect. As we have already established above, complex Langevin does not perform well at lattice coupling $\beta < 5.7$ in the case of HDQCD. The reason for the apparent failure of complex Langevin at $\mu = 0$ in other works is the fact that the

[9] Here the full theory is not holomorphic, but the expansion is. If there is a pole in the full theory, the convergence radius of the expansion is expected to not include that point. Hence, here the breakdown of the expansion signals the occurrence of the pole in the full theory.

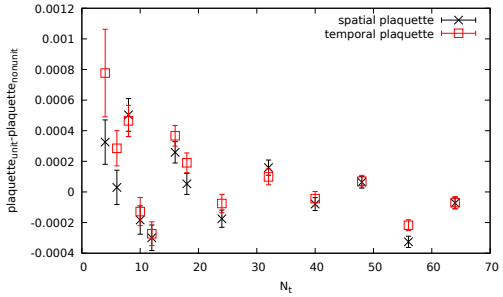

Figure 3.18: Comparison of complex Langevin and real Langevin for $\beta = 6.0$, $\kappa_s = 0.12 = \kappa_t$ in the κ_s expansion to 10th order. The plot shows the difference in the plaquettes between both methods at $\mu = 0$.

simulations have been done close to or below that boundary region[10]. Assuming that one can generalize from HDQCD – which is suggested by [125], where complex Langevin fails at low β in full QCD as well – this explains the failure and suggests that larger β is necessary. Hence, to show that complex Langevin and real Langevin agree for large enough coupling we simulate QCD in the κ_s expansion to 10th order with the same parameters as above. The difference in the plaquette between the simulations with and without reunitarization is shown in figure 3.18, where one can see that there is a good agreement with zero. Note that the range of the y-axis in this plot is rather small, hence everything is consistent with zero[11].

Since the κ_s expansion did not converge to full QCD, we now directly simulate full QCD.

3.2.2 Full QCD at low temperature

First, we continue to pursue low-temperature simulations. This is achieved by making N_t large, keeping β fixed and varying μ to find the nuclear liquid-gas transition. So far this has not been achieved in lattice QCD due to the sign problem, however, there are studies for gauge groups without a sign problem, see e.g. [158] for the group G_2. The transition should be at $m_N/3$, the quark mass from the nucleon. We start by simulating full QCD at the same parameters we used in the expansion, i.e. spatial volume $N_s^3 = 10^3$, $\beta = 6.0$, $\kappa = 0.12$. For those parameters, the simulation is well behaved and there are no issues for any value of μ. The density for $N_t = 10$ ($T \approx 237$MeV) and $N_t = 16$ ($T \approx 135$MeV) is shown in figure 3.19. First we note that the temperature is still rather high, however at $T \approx 135$ MeV at least a signal of the transition should be visible. We do only see the lattice go

[10] Of course this value depends on the choice of discretization for fermions, so in this context, it is to be taken as a rough guideline.

[11] At small N_t fluctuations are usually stronger, hence the larger errors.

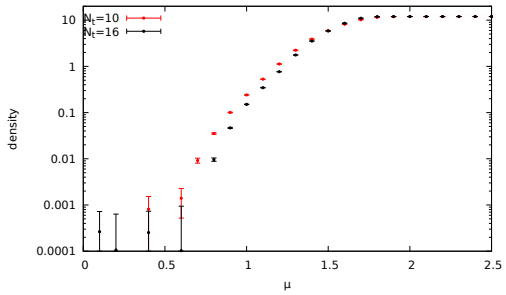

Figure 3.19: Density for full QCD at two different temperatures. $N_t = 10$ corresponds to $T \approx 237$MeV, $N_t = 16$ to $T \approx 135$MeV. This is at a pion mass above $m_\pi > 4.5$GeV.

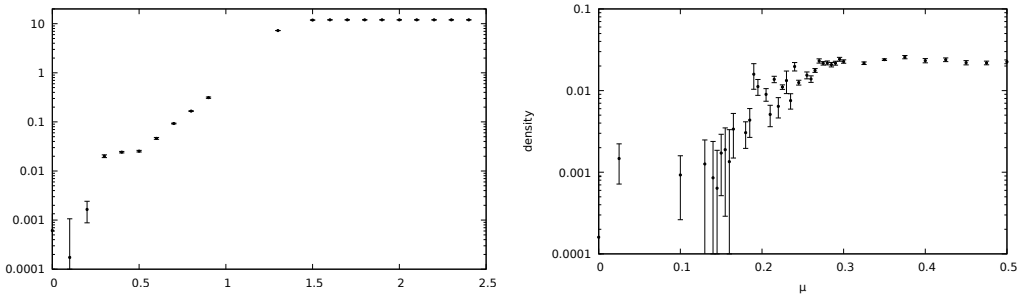

Figure 3.20: Density for full QCD for two different parameter sets. The simulations are at much lower masses than in figure 3.19 and show an onset which may be interpreted as the nuclear transition. Note that the onset is not at the correct position, see text for details.

into saturation, but no distinct plateaus as in the case of G_2 [158] occur. This is to be expected here since the pion mass is[12] $m_\pi > 4.5$GeV, see appendix A.6. Hence, the quark mass for the onset from the nucleon $m_N/3$ will be at least $m_\pi/2$ or even larger. In lattice units, this means that the earliest point where we expect the transition is at $\mu \approx 0.9$, which is already far into the rise to saturation. Hence, we cannot see the transition even if it is there. Therefore we need to go to lower masses.

Figure 3.20 shows the resulting density at smaller masses. The left plot is for $\beta = 5.9$ and $\kappa = 0.148$ on a $8^3 \times 20$ lattice. We list the lattice spacing and masses in appendix A.6. The corresponding pion mass is $am_\pi = 0.834(25)$ or $m_\pi \approx 3.1$GeV. We expect the onset of the transition to be at $m_N/3$, which is above $m_\pi/2$. However, even $m_\pi/2$ is already on the plateau, this means that the onset happens to early. In the right plot we zoom in on the plateau region. The data are for $\beta = 6.0$, $\kappa = 0.15$ on a $8^3 \times 24$ lattice. Here $am_\pi/2 = 0.370(9)$ ($m_\pi/2 \approx 1.3$GeV) and again the onset is already before that. A possible reason for this discrepancy is that the

[12] Here we only measured the pion mass on a rather large lattice. Finite size effects will make the mass even larger.

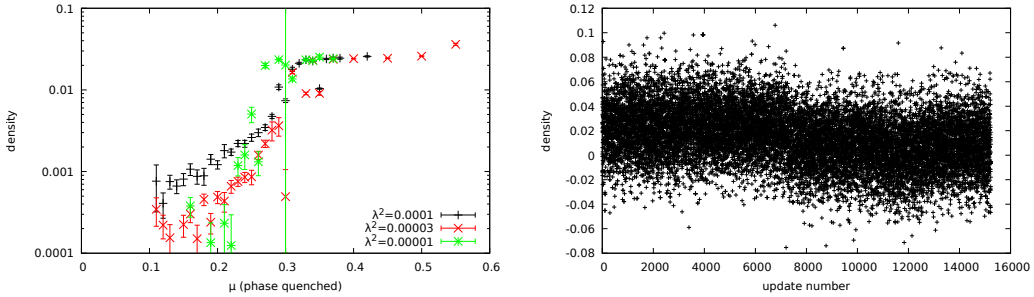

Figure 3.21: Left: Density for a phase quenched simulation with on a $8^3 \times 24$ lattice with $\kappa = 0.15$ and $\beta = 6.0$. Lower λ^2 is closer to the actual physics. Right: Evolution of the density for $\mu = 0.31$ and $\lambda^2 = 0.00001$. See text for a discussion.

temperature in those simulations is already moderately high. In physical units, the right plot of figure 3.20 has temperature $T \approx 141$MeV. In the left plot this value similar $T \approx 140$MeV. Physically the critical endpoint of this transition is found at $T \approx 10$MeV. Hence, at our high temperatures, this can only be a remnant of the transition. We are however not in the continuum limit, where things might change.

The more pressing concern, however, is a problem in the simulation itself, which we observe in phase quenched simulations.

Phase quenched simulations

Phase quenched QCD is full QCD but with the phase factor left out, it does not agree with full QCD and we only use it for diagnostic purposes. Its partition function reads

$$Z_{\text{pq}} = \int DU \left(\det M(\mu) M^\dagger(\mu) + \lambda^2\right)^{N_f/2} e^{-S_g} \tag{3.48}$$

and the determinant is real. Here λ acts as a regularization parameter to prevent too small eigenvalues in the fermion matrix M and hence speeds up the simulation, it has to be extrapolated to zero. Physically this is equivalent to the introduction of isospin chemical potential, which parametrizes the difference in different quark species. For lattice studies of phase quenched QCD or QCD at finite isospin chemical potential see e.g. [159–161]. In phase quenched simulations the onset of the transition is expected at $m_\pi/2$. The left plot of figure 3.21 shows the onset transition in from a phase quenched simulation at finite λ. Here the extrapolation $\lambda \to 0$ would move the onset of the plateau even further to the left. The onset should be at half the pion mass $am_\pi/2 = 0.371(9)$ ($m_\pi/2 \approx 1.31$GeV), however it happens too early again, just as in full QCD. Thus we conclude that this is not an issue of complex Langevin. Instead, the reason is a problem with the thermalization of the simulation. The right plot of the same figure shows the HMC trajectory at $\mu = 0.31$

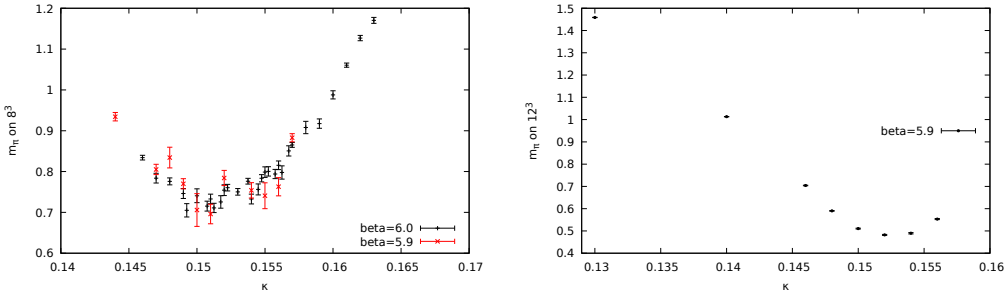

Figure 3.22: Pion mass at spatial volume 8^3 and 12^3. It was measured with $N_t = 32$.

for the smallest value of λ^2. After a long simulation time, the density goes to a different value, which should not happen for purely real simulations unless the system is not yet thermalized properly.

Assuming we have the same issue in complex Langevin simulations, this explains the occurrence of the transition at wrong values. The need for long runs is in conflict with the fast rise of the unitarity norm. Dynamical stabilization [128] may make long simulations and a proper investigation of this issue possible. Additionally for continuum extrapolated simulations, one needs to go to a much lower temperature, hence making lattices much larger. Such simulations are out of reach with our resources for now.

Let us briefly mention that in our simulations we see a clear difference between phase quenched and full QCD simulations. This can be seen e.g. for the Polyakov loop. In a phase quenched simulation the Polyakov loop and the inverse Polyakov loop are equal. In our full QCD simulations only observe this at $\mu = 0$, where phase quenched and full QCD agree either way.

Volume dependence of the mass

In the previous section, the pion mass was always above 2GeV. We want to get far below that, however, at small spatial volumes such as ours, the mass has a lower bound [162]. We measured the mass at a volume of $N_s^3 \times 32$ with $N_s = 8, 12$. The result is shown in figure 3.22. This confirms that our simulations have been done at approximately the lowest possible mass at this volume. Below we will perform simulations at $\beta = 5.9$ and $\kappa = 0.15$. The pion mass for $N_s = 8$ is $m_\pi > 2.1$ GeV, we did not measure m_N. For $N_s = 12$ $m_\pi \approx 1.55$ GeV and $m_N \approx 2.8$ GeV. See appendix A.6. Hence, $m_\pi/2 < m_N/3$ such that we are not in the limit of very heavy quarks. Hence, we see nontrivial effects from the quarks and the dynamics are not purely dominated by the Yang-Mills sector.

3.2.3 Curvature of the transition line from full QCD

Instead, we turn to another interesting region of the QCD phase diagram, the deconfinement transition and the search for the critical endpoint. The transition has been investigated to high precision in the Taylor expansion and continuation from imaginary chemical potential approaches [15, 18, 19, 19, 49]. However, it is commonly accepted that so far those methods cannot reliably give results at $\mu/T > 1$, see e.g. [163]. In this section, we aim to demonstrate that the complex Langevin method is able to do so and extract the curvature of the transition line.

Simulation setup

In order to scan the transition, we need to perform simulations at different μ and T. On the lattice typically T is varied by varying the lattice coupling β, as we have argued e.g. in section 3.2.1. Here we take another approach: We fix parameters β and κ and vary temperature by varying the temporal lattice size, since $T \sim 1/N_t$. This is different to usual simulations where one keeps the aspect ratio N_s/N_t fixed and we will see that this affects our simulations in one observable. Note also that our simulations will mainly be performed at spatial volumes $N_s^3 = 8^3$ and $N_s^3 = 12^3$ which as we will see still is affected by finite size artifacts.

We start by checking our region of interest for convergence of complex Langevin.

Boundaries and poles

As we have discussed in section 3.1, the two main issues plaguing complex Langevin simulations are poles in the complex plane and boundary terms which manifest as tails in the distribution of observables and the drift.
The only way poles can appear in QCD is via the fermion matrix M. Since the fermionic part of the action is $-\log \det M$, any zero eigenvalues in the fermion matrix will lead to a diverging logarithm. To look for poles, we compute the lowest lying eigenvalues of the fermion matrix using the Krylov-Schur algorithm, see e.g. [164], which is particularly useful for large sparse matrices. We do this for spatial volume of 8^3 for $\beta = 5.9$ and $\kappa = 0.15$ for different μ and T on 30 thermalized configurations per point. We only investigate eigenvalues for configurations with unitarity norm $d_U < 0.1$. If the unitarity norm grows too large occasionally eigenvalues which are much smaller than the rest occur, if we cut the unitarity norm appropriately this becomes an extremely rare event. We show distributions of eigenvalues in figure 3.23. For none of the data shown there, the eigenvalues are too close to zero. We observe that they move closer to zero for decreasing temperature (increasing N_t). Additionally, there is a trend of more smaller eigenvalues at increasing μ. However, all eigenvalues are sufficiently far from zero, such that simulations in the parameter regions shown are allowed from this point of view. Note that this changes, when we go to even higher μ. When we did the simulations in section 3.2.2, we observed that there is a region in which we can practically not simulate. For the parameters

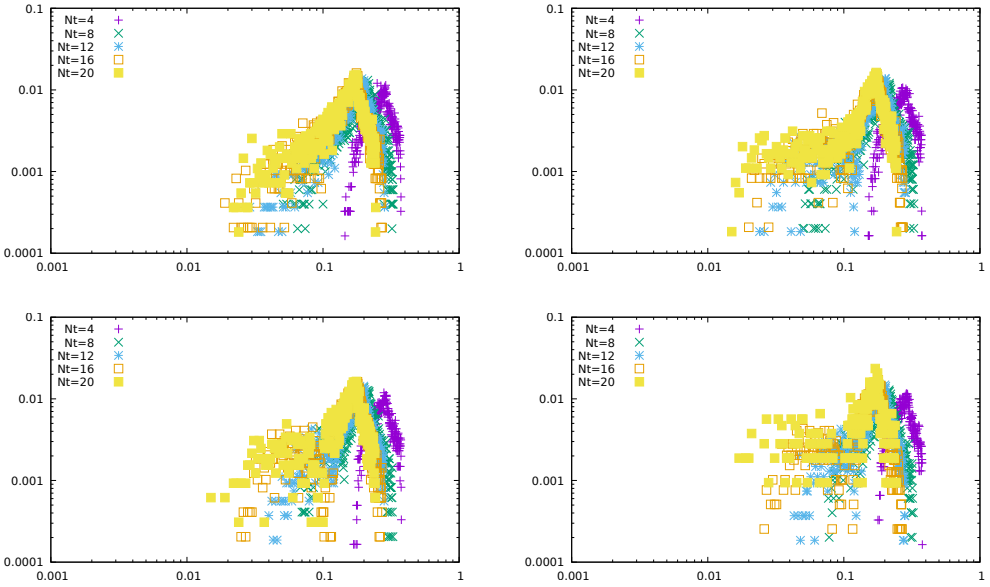

Figure 3.23: Normalized histogram of the absolute value of the lowest lying eigenvalues. The different plots are at $\mu = 0.0, 0.3, 0.9, 1.2$ from top left to bottom right.

investigated here, we also observe that the simulation slows down significantly when approaching the region of $\mu = 1$. This can be interpreted by an increasing number of low lying eigenvalues, since the conjugate gradient – which we use to invert the fermion matrix – is known to slow down significantly for too low eigenvalues. For now we stay away from too large μ and too low T, we have a maximal value of $\mu/T_c(0) \approx 5$ and our simulations do not hit any poles.

We investigate the boundary terms by looking at histograms of observables and check for tails in the distribution. If we observe exponential decay, we trust that there are no significant boundary terms. Figure 3.24 shows the histogram of $L_{\text{abs}} = \sqrt{LL^{-1}}$ for two different N_t at $\mu = 0.3$. The histograms decay exponentially and no change in behavior is seen when cutting the unitarity norm earlier than $d_U = 0.1$. We also show the expectation values of the Polyakov loop as we will use it, see equation (3.49) in the next section. This is shown in table 3.2, where no trend of more accuracy or deviations towards lower cutoff is visible. Here we focused on the Polyakov loop, since for the location of the phase transition we are mainly interested in this observable. Other observables such as the plaquette, the density, and the chiral condensate are equally well behaved, but not shown here. We conclude, that our simulations are not plagued by boundary terms.

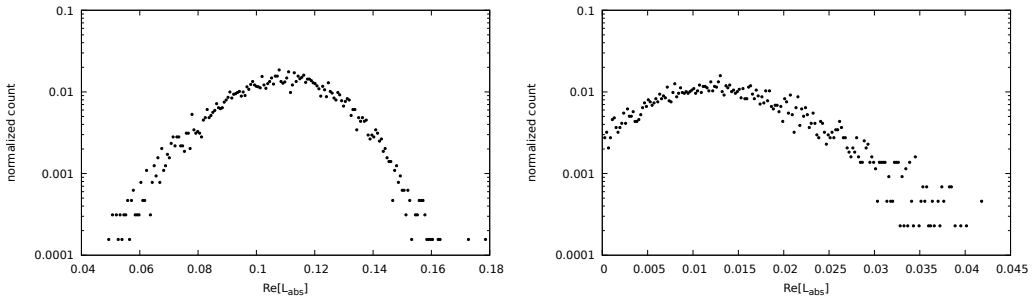

Figure 3.24: Normalized histogram of L_{abs} for $\mu = 0.3$ and $N_t = 7$ (left) and $N_t = 15$ (right). It is well behaved.

μ	N_t	L_{abs} cut at 0.1	L_{abs} cut at 0.01
0.0	7	0.081(1)	0.083(1)
0.0	15	0.0138(1)	0.0139(2)
0.3	7	0.0108(5)	0.1111(14)
0.3	15	0.0138(1)	0.0140(2)
0.5	7	0.1364(5)	0.1354(6)
0.5	15	0.0148(1)	0.0146(2)

Table 3.2: Comparing L_{abs} expectation values for different unitarity norm cutoffs.

A note on the drift criterion

For completeness, we also had a brief look at the criterion on the decay of the drift from [122]. This criterion states that the drift term has to decay at least exponentially fast for complex Langevin to be convergent. We show the drift histogram for two different μ and N_t[13] in figure 3.25. One can see that according to this criterion, complex Langevin results should not be trusted at much lower unitarity norm.
Since expectation values and histograms of the observables are well behaved as shown in the previous section, we do not cut earlier despite the drift histogram decaying polynomially in some cases. This can be viewed in analogy to the simple model, see figures 3.4, 3.7 and 3.3. There the drift histogram shows polynomial decay all the time, but the boundary terms initially are small enough such that the expectation values still are correct.

We conclude that we simulate in a region without boundary terms or poles. As long as we cut the unitarity norm early enough complex Langevin can be trusted.

[13]Note that there is no particular reason why those μ and N_t values are different from those in the previous section, we just did not measure the histogram in all cases, so here we chose runs which had large enough unitarity norm to show something. Many runs stay at low enough unitarity norm to never develop issues.

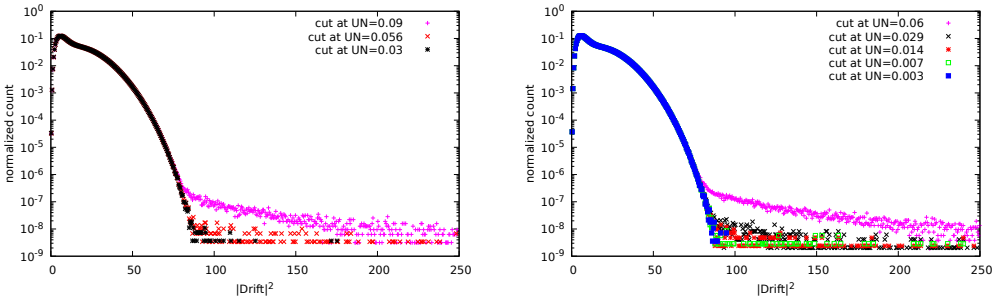

Figure 3.25: Histogram of the drift for $\mu = 0.15$ and $N_t = 10$ (left) and $\mu = 0.05$ and $N_t = 20$, see text for a discussion.

Observables for the transition

The typical observable that is chosen to probe the QCD crossover transition at physical pion mass is the susceptibility of the chiral condensate. Since our simulations are at heavier pion masses, it can be hard to get a signal for the crossover transition, see e.g. [165] for the mass dependence of the chiral susceptibility. The chiral condensate also needs to be renormalized, which requires additional expensive simulations at zero temperature. Hence, we discard the chiral condensate as an observable for this work.
Instead, we turn to the Polyakov loop, which at our masses gives a much better signal for the crossover. However, the Polyakov loop needs to be renormalized as well. Since it renormalizes purely multiplicatively [166], we can use ratios of Polyakov loops and not renormalize at all. Also, since the Polyakov loop itself is very noisy, we work with

$$L_{\text{abs}} = \sqrt{L L^{-1}} \tag{3.49}$$

instead. One could also use the sum of L and L^{-1}, but we find that within our setup the signal for the transition becomes much less clear in this case. We use the following observables to determine the phase transition

$$\mathcal{C}_3(L_{\text{abs}}) = \frac{\langle L_{\text{abs}}^3 \rangle}{\langle L_{\text{abs}}^2 \rangle^{1.5}} \tag{3.50}$$

$$\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle) = \frac{\left\langle (L_{\text{abs}} - \langle L_{\text{abs}} \rangle)^3 \right\rangle}{\left\langle (L_{\text{abs}} - \langle L_{\text{abs}} \rangle)^2 \right\rangle^{1.5}} \tag{3.51}$$

$$\mathcal{C}_4(L_{\text{abs}}) = \frac{\langle L_{\text{abs}}^4 \rangle}{\langle L_{\text{abs}}^2 \rangle^2} \,. \tag{3.52}$$

The quantity $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$ measures the difference of fluctuations in both phases. Commonly one extracts the phase transition temperature from such quantities as their intersection point from different volumes, see e.g. [167]. The Binder cumulant $\mathcal{C}_4(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$ is a more common observable for that, however within our setup it

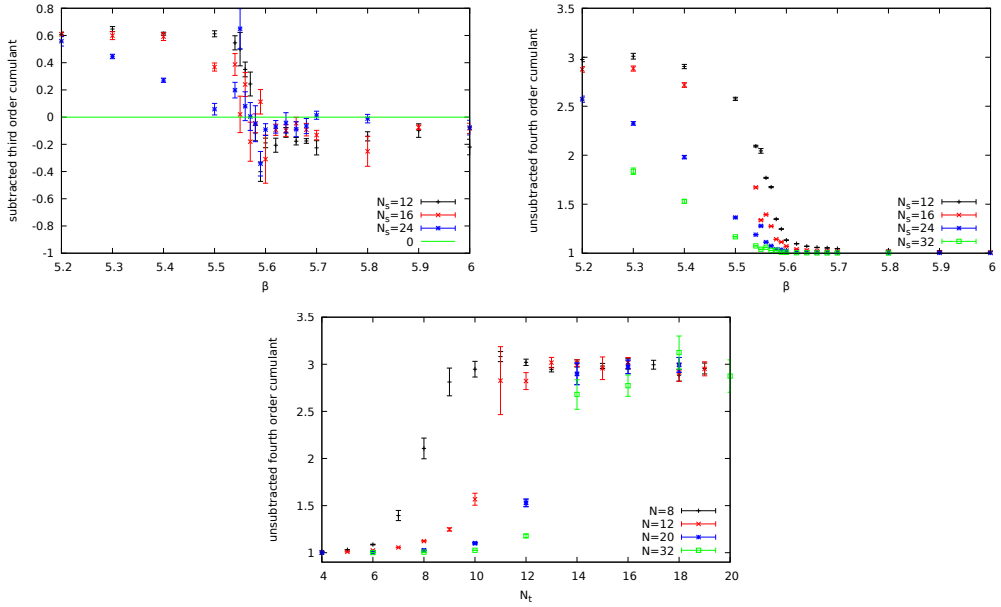

Figure 3.26: Top left: Subtracted third order cumulant $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$ as a function of β. Top right: Unsubtracted fourth order cumulant $\mathcal{C}_4(L_{\text{abs}})$. In both cases $\kappa = 0.15$, $N_t = 6$. Bottom: Unsubtracted fourth order cumulant as a function of N_t for different volumes. Here $\beta = 5.9$ and $\kappa = 0.15$. See text for a discussion.

does not show a strong enough signal, so we do not use it. Taking the same quantities but only with L_{abs} instead of $L_{\text{abs}} - \langle L_{\text{abs}} \rangle$ also shows a crossover between both phases. However, here it is not clear how to extract a phase transition temperature. We will argue how to get a transition temperature using a different method than for $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$, see below.
Before going to chemical potential, we look at the behavior of the third order cumulant from (3.51) at $\mu = 0$ and when varying T via β instead of N_t. We only perform this analysis to identify T_c, we will not use it further since complex Langevin does not work for low β. The result is shown in figure 3.26. One can see that the curves for different volumes meet at their zero crossing. Thus we define the transition temperature as the temperature at which $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle) = 0$. Figure 3.26 also shows the unsubtracted fourth order cumulant $\mathcal{C}_4(L_{\text{abs}})$ which does not show any distinct points which could be used to identify the phase transition, the same holds for the unsubtracted third order cumulant. Finally, 3.26 also shows the volume dependence of the unsubtracted fourth order cumulant. It shows strong finite size effects in the volumes we work with, hence strong effects in the results are expected.

The subtracted third order cumulant for $\beta = 5.9$, $\kappa = 0.15$ as a function of N_t for different μ is shown in figure 3.27 (top). From this figure, we see that larger μ tend to be less noisy. We extract the transition temperatures for $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$

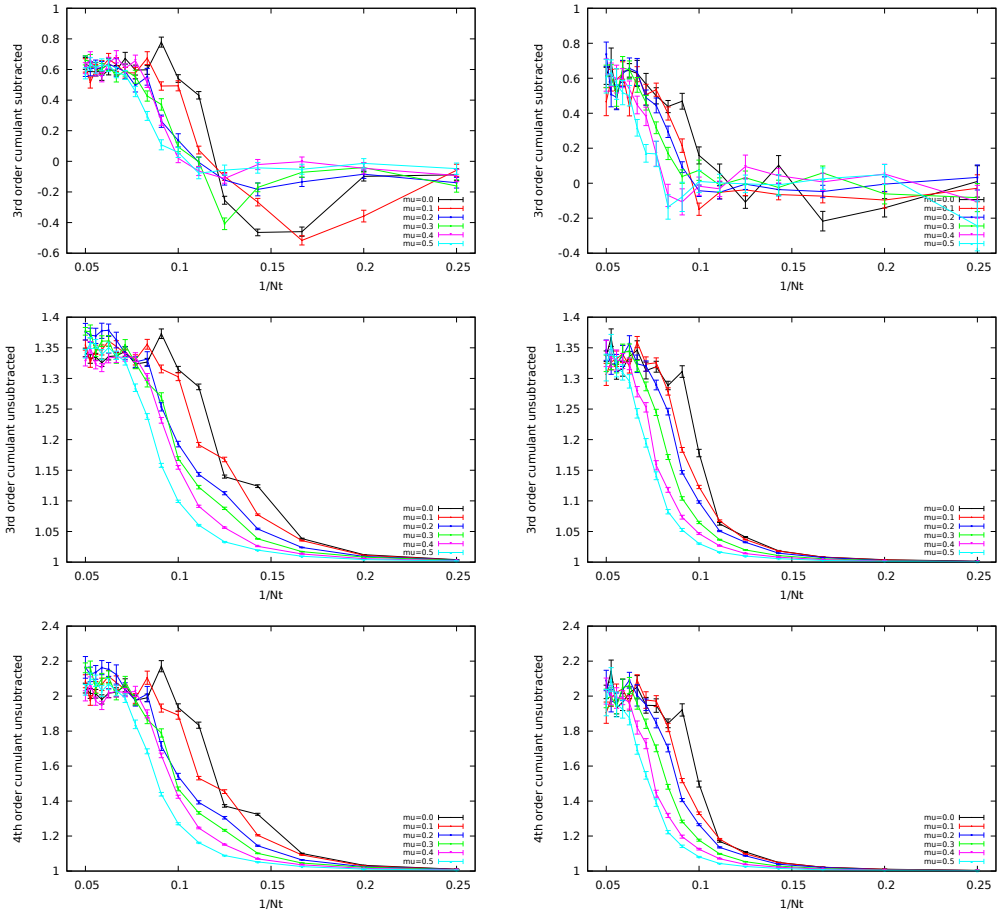

Figure 3.27: Observables as a function of $1/N_t$ for different μ. Parameters are $\beta = 5.9$, $\kappa = 0.15$ and $N_s^3 = 8^3$ (left) or $N_s^3 = 12^3$ (right). Top: Subtracted third order cumulant $\mathcal{C}_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$. Middle: Third order cumulant $\mathcal{C}_3(L_{\text{abs}})$. Bottom: Fourth order cumulant $\mathcal{C}_4(L_{\text{abs}})$. See text for a discussion.

from two different fits such that we can have an estimate for the systematic error.

1. A linear fit to the region close to the zero crossing.

2. A fit of

$$s - (s - b)e^{-kN_t}\,, \tag{3.53}$$

which is performed with N_t and not $1/N_t$ on the x axis. This fit takes into account all points with N_t larger than the first point where the cumulant is larger than 0.15.

Errors are estimated using the statistical bootstrap. I.e. given the original data set of size N, we draw N random samples from it, such that samples can occur multiple

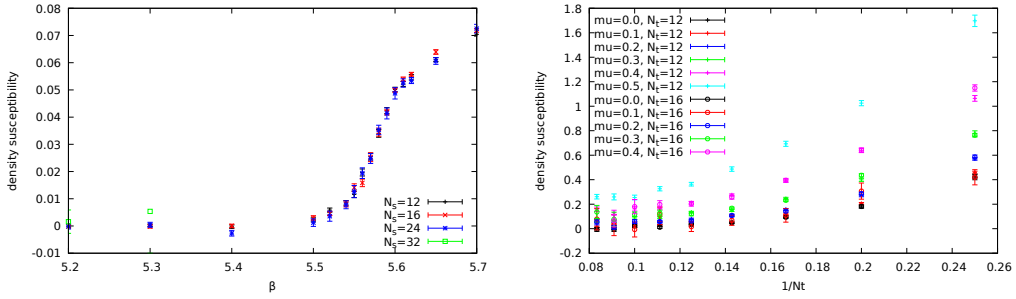

Figure 3.28: Density susceptibility. Left: From β-scan. Right: From N_t-scan. Their behavior differs, see text.

times. We then perform the fit on this random set. Averaging over many of those sets yields an estimate for the statistical error, see e.g. [37].

For the other two observables, there is no clear feature that can be used to identify a transition temperature, see figure 3.27. Instead, we use the method from [168], which was also applied and shown to agree with other definitions in the continuum in [19]. The idea is to obtain the transition temperature at finite μ from the one at $\mu = 0$ by shifting the observables. To be more precise, given an observable $\mathcal{O}_{ns}$ that does not show any identifiable behaviors in the transition region, we take the value of $T_c(0)$ from another observable $\mathcal{O}_s$, which gives a clear signal. We then extract the value of temperature at $\mu = 0$ at which $\mathcal{O}_{ns}^{\text{crit}} = T_c(0)$. Next, we look at $\mathcal{O}_{ns}$ at higher μ and check for the temperature $T_c(\mu)$ at which $\mathcal{O}_{ns} = \mathcal{O}_{ns}^{\text{crit}}$. This defines the critical temperature at higher μ. This defines the critical temperature and does not require any distinct features in the observable, as long as we have a good determination at $\mu = 0$. We use the $\mu = 0$ value from the analysis of the subtracted third-order cumulant. In our simulations, we use a smoothing cubic spline fit to get a continuous function from our data before applying the shifting method. We tuned the number of supports in the cubic spline fit such that a smooth behavior without any bumps was observed in the transition region. Errors again are estimated using bootstrap analysis.

Another observable that does not need to be renormalized is the density and its derivatives. We show the density susceptibility – see equation (2.38) – from a β scan and from our N_t scan in figure 3.28. While for the β scan a transition is clearly visible, this is not the case for the N_t scan. This is probably due to the fact that the aspect ratio of the lattice is not held constant. The advantage of this observable is that it does not depend on volume. Hence, we could use any value of N_s^3 and extract a transition temperature. Since we are unable to extract a distinct feature from our data and the difference between both plots in figure 3.28 is so large we discard this observable for now.

Curvature of the transition line for $\mu/T_c(0) < 5$.

Finally, we are interested in the curvature of the transition line, which is a commonly measured quantity from lattice simulations, see e.g. [19, 49] for recent values. Using $\mu_B = 3\mu$, the transition temperature is commonly parametrized as

$$T_c(\mu_B) = T_c(0) - \kappa_2 \frac{\mu_B^2}{T_c(0)} + \mathcal{O}(\mu_B^4)\,. \tag{3.54}$$

We fit this function to our data and normalize afterwards to

$$\frac{T_c(\mu_B)}{T_c(0)} = 1 - \kappa_2 \left(\frac{\mu_B}{T_c(0)}\right)^2 + \mathcal{O}(\mu_B^4)\,. \tag{3.55}$$

We report the resulting parameters in table 3.3. $T_c(0)$ is the fit parameter from the above fit, errors are from a bootstrap analysis by performing the fit on different bootstrap samples and averaging over those. Hence, even though we used the same T_c at $\mu = 0$ from the subtracted third-order cumulant for all other observables as well, the actual quoted value of $T_c(0)$ differs from that, since it comes from the fit and hence incorporates the effect of the other data points as well.

The resulting value for κ_2 is consistent for all observables from the Polyakov loop ratios we looked at.

Comparing to the values from [19, 49] which is approximately $\kappa_2 \approx 0.014$, the curvature from our data is much smaller. This is not surprising since in HDQCD – where quarks are heavy and cannot move in spatial directions – the transition line at small temperatures is flat [83]. Hence, our results nicely interpolate between heavy quarks and those with physical mass. Note, however, that there are a few other systematic sources of error which may have an effect. This is also reflected in the transition temperature which for physical quark masses should be $T_c \approx 155$ MeV, while in pure Yang-Mills or infinite masses it is at $T_c \approx 270$ MeV. Our simulations yield values above the Yang-Mills temperature due to finite size effects. This is also reflected in the fact that the curvature slightly differs between volumes, see table 3.3. We are also not in the continuum limit, while curvature values given in [19, 49] are properly continuum extrapolated.

3.2.4 Conclusion on complex Langevin for QCD

In this section, we found that complex Langevin seems to be applicable throughout the most part of the phase diagram. We found that the hopping expansion does not work well close to poles and thus should not be used to scan the phase diagram. Using the hopping expansion in a region where it behaves nicely, we showed that

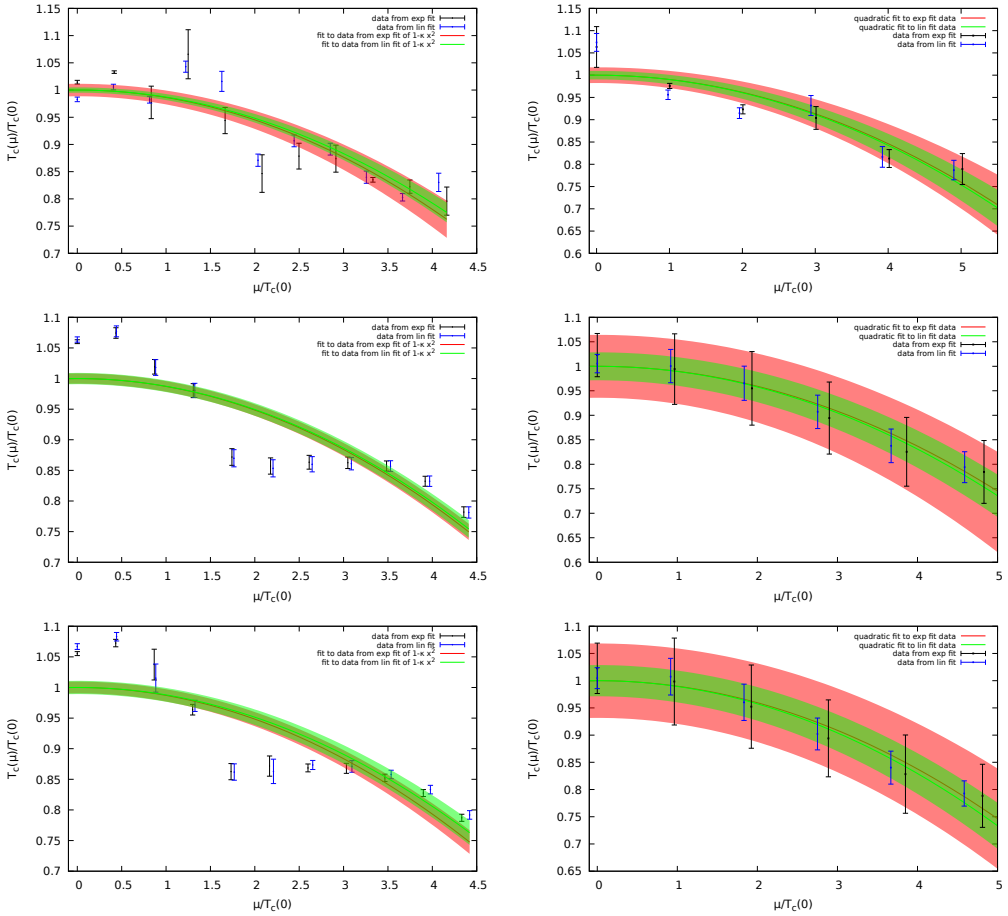

Figure 3.29: Curvature of the transition line. Parameters are as in figure 3.27.

Method	N_s	κ_2	$T_c(0) \times a$	$T_c(0)$/MeV
exp fit $B_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$	12	0.0011(2)	0.100(1)	305(3)
lin fit $B_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$	12	0.0011(1)	0.102(1)	311(3)
shift 1 $B_3(L_{\text{abs}})$	12	0.00113(7)	0.103(6)	314(18)
shift 2 $B_3(L_{\text{abs}})$	12	0.00118(7)	0.109(2)	332(6)
shift 1 $B_4(L_{\text{abs}})$	12	0.00112(10)	0.103(7)	314(21)
shift 2 $B_4(L_{\text{abs}})$	12	0.00118(5)	0.109(3)	332(9)
exp fit $B_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$	8	0.00152(14)	0.1201(13)	367(3)
lin fit $B_3(L_{\text{abs}} - \langle L_{\text{abs}} \rangle)$	8	0.00144(9)	0.1228(6)	374(2)
shift 1 $B_3(L_{\text{abs}})$	8	0.00140(3)	0.1148(10)	350(3)
shift 2 $B_3(L_{\text{abs}})$	8	0.00140(3)	0.1133(10)	346(3)
shift 1 $B_4(L_{\text{abs}})$	8	0.00144(6)	0.115(1)	351(3)
shift 2 $B_4(L_{\text{abs}})$	8	0.00135(5)	0.113(1)	345(3)

Table 3.3: Curvature and $T_c(0)$ from different methods. The curvature is given for $\mu_B = 3\mu$. Shift 1 and shift 2 are shifts (see text) with the data point for $T_c(0)$ taken from $B_3(P - \langle P \rangle)$.

complex Langevin simulations at $\mu = 0$ yield correct results.
For full QCD we were able to simulate at low temperatures and find first signals of the nuclear liquid-gas transition. While this transition was not at the correct position, we were able to trace this failure back to long thermalization times which also occur in phase quenched simulations. They are not a result of a failure of complex Langevin.
We were able to compute the transition temperature for moderately heavy pion masses of $m_\pi \approx 1.55$ GeV using the complex Langevin method. This was done up to $\mu/T_c(0) \approx 5$, which is much higher than the range of other methods. We carefully monitored the observables and poles to check for the convergence of the method and removed configurations after the first occurrence of boundary terms or poles.
Using ratios of Polyakov loops, see equations (3.50)-(3.52), we found that the transition line is well described by a quadratic behavior with curvature $\kappa_2 \approx 0.0011$ for our largest volumes, i.e. on a $N_s = 12$ lattice. We used two different methods to extract this value and both agreed. The curvature we found interpolates nicely between the value in HDQCD, where there is practically no curvature [83] and the value $\kappa_2 \approx 0.014$ at physical quark masses [19, 49]. Note that our simulations still show finite size effects, which is especially visible in the fact that the smallest possible masses are strongly restricted by the spatial volume.
To summarize, we found that complex Langevin does work for QCD if monitored appropriately. We will discuss lessons for complex Langevin from our QCD and SU(2) simulations at the end of the chapter.

3.3 Applicability of complex Langevin in SU(2) real-time simulations

Apart from simulations at finite chemical potential, the sign problem also prohibits simulations in Minkowski space directly, since there the Boltzmann factor is a pure phase. The application of complex Langevin to real-time problems was investigated in [59, 102, 103]. The focus there was put on real-time simulations at finite temperature. This is done by simulating on the Schwinger-Keldysh contour or on deformations thereof. For SU(2) gauge theory, it turned out that Langevin dynamics are very unstable and do not yield correct convergence unless rather unpractical parameters were chosen.
We aim to investigate the applicability of the complex Langevin method to SU(2) pure gauge theory by using gauge cooling [105] and dynamical stabilization [128] and investigate different choices of contours. For the simulations in this section, we used a CPU code written by the author.

The Wilson gauge action on a contour

We wish to investigate different contours, so a general expression for the lattice action on a contour is needed. Since the contour does only change the temporal part of the lattice gauge action, we have to split the action into two parts

$$\begin{aligned} S = &- \beta\gamma_t \sum_n \sum_i \left[\frac{1}{4} \left(\mathrm{Tr} U_{0i}(n) + \mathrm{Tr} U_{0i}^{-1}(n) \right) - 1 \right] \\ &+ \frac{\beta}{\gamma_t} \sum_n \sum_{ij, i<j} \left[\frac{1}{4} \left(\mathrm{Tr} U_{ij}(n) + \mathrm{Tr} U_{ij}^{-1}(n) \right) - 1 \right] , \end{aligned} \tag{3.56}$$

the first term contains loops containing temporal links, the second part contains only spatial loops. The anisotropy parameter $\gamma_t = a_s/a_t$ is the ratio of the spatial and temporal lattice spacing. When speaking of a_t in this part of the thesis, we will refer to the lattice spacing corresponding to the real-time formulation unless specified otherwise. It is necessary to yield the correct continuum limit. From equation (3.56), one can regain the usual Euclidean action in the usual way

$$iS_{\text{Minkowski}}(a_{t,\text{Minkowski}} = -ia_{t,\text{Euclidean}}) = -S_{\text{Euclidean}} \, . \tag{3.57}$$

In equation (3.56) this is done by replacing $a_t \to -ia_t$ in the anisotropy factor γ_t, yielding the usual Euclidean Wilson gauge action. Additionally, one has to take care when formulating Langevin dynamics, since in the Minkowski formulation the prefactor of the action changes. Hence, the Langevin drift changes accordingly

$$K_a(n) = -D_a(n)S \to K_a(n) = iD_a(n)S \, . \tag{3.58}$$

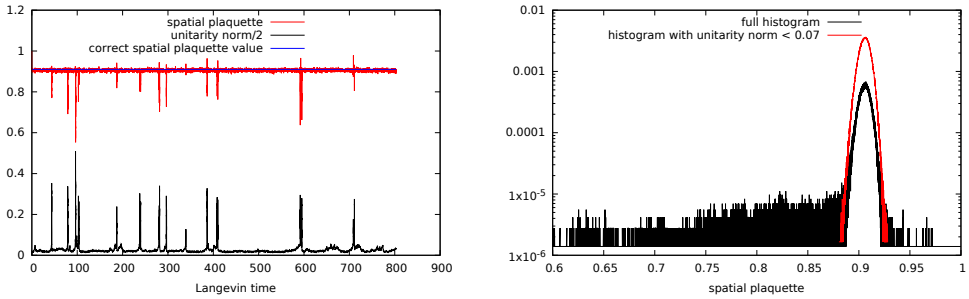

Figure 3.30: Left: Langevin evolution of the spatial plaquette and the unitarity norm. Note that the deviation of the Euclidean result from the apparent average of the complex Langevin trajectory is due to discretization artifacts. Right: Histogram of the spatial plaquette. See text for details.

Any given contour can be implemented via the anisotropy parameter. The contours are defined using the parameters τ which is the extent of the contour in the Euclidean direction and represents inverse temperature and t_0 which is the extent in Minkowski direction. All contours are deformations of the one described in section 2.2, they start at 0 and end at $-i\tau$ and both points are identified with periodic boundary conditions.
For a lattice with temporal extent N_t, in general, the anisotropy parameter is different for each pair of points. For example, given a lattice with N_t points in the temporal direction and using the action of equation (3.56), one regains the usual Euclidean simulation by putting $\gamma = N_t/(-i\tau)$ for all anisotropies.

Running the simulation

A good choice of the contour is a symmetric triangle [59, 103], we will investigate other choices below. We begin by simulating on a $4^3 \times 8$ lattice on a symmetric triangle with real-time extent $t_0 = 1$ and Euclidean extent $\tau = 1$. Following [103], we simulate at lattice coupling $\beta = 16.0$. Note that this value is extraordinarily large. The deconfinement transition in a standard Euclidean simulation happens at around $\beta \approx 2.5$ (for N_t around the values chosen in this part of the thesis). Instead of gauge fixing as in [103], we employ gauge cooling [105] to stabilize the simulation. We will compare different stabilization methods below. We employ adaptive step size [104, 134] for all simulations with a typical value for the time step of $\epsilon \approx 10^{-4}$ to 5×10^{-4} or lower. For some of the simulations we use the improved Langevin algorithm from [169].

As was discussed in section 3.1, in order for complex Langevin to converge, there must be no boundary terms in the observables or drift. An easy way to quantify this is to look at histograms of observables. Another way to check for correctness within this simulation setup is to compare it to Euclidean simulations. Fortunately, observables that depend only on spatial quantities do not have any dependence

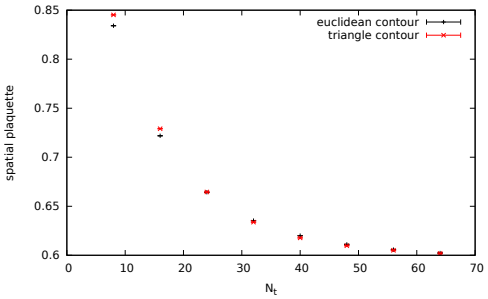

Figure 3.31: Visualization of discretization artifacts. We compare the spatial plaquette for the symmetric triangle contour with standard Euclidean simulations. Larger N_t means smaller steps. See text for details.

on the contour (except discretization artifacts) and can be used as a correctness measure. One such observable is the spatial plaquette. In figure 3.30 we show the evolution of the spatial plaquette and the unitarity norm (see equation (3.39)), which is a measure for the distance to SU(2). We also show the histogram of the spatial plaquette. One can see that the evolution seems to stay close to the correct value, however, occasionally there are spikes in the unitarity norm and plaquette evolution. Looking at the full histogram one can clearly observe a weaker than exponential decay when taking into account the full simulation, hinting at the failure of complex Langevin. This is fixed by leaving out all data points that are above $d_U = 0.07$. In this case, the histogram decays nicely and expectation values become correct. Typically instead of leaving out the parts of the simulation where the unitarity norm is small one would only take the initial part of the simulation before any spikes start to occur and do many of those simulations to average over. This requires the simulation to thermalize before the spikes set in. In general, this must be carefully monitored. In our case it is sufficient to compare to the spatial plaquette from Euclidean simulations.

A brief note on discretization artifacts

In figure 3.30 it looks like the spatial plaquette from complex Langevin and standard Euclidean simulations do not agree. This is due to discretization artifacts from the contour. In figure 3.31 we compare both simulations on a $8^3 \times N_t$ for $\beta = 3.0$. The triangle contour in this case has real time extent $t_0 = 3$ and inverse temperature $\tau = 16$. This triangle is really flat and hence should yield results close to the Euclidean ones. In the figure, one can see that by increasing the number of points on the contour in both triangle and Euclidean contour simulations leads to better agreement.

Comparison of contours

In principle, the interesting contour for real-time simulations is the Keldysh-Schwinger contour. However, on this contour complex Langevin does not converge at all. Instead, we take variations of the Keldysh-Schwinger contour on which the anisotropy parameter always has a real and imaginary part. In this case, one still has to do numerical analytic continuation to the pure Minkowski case via Bayesian methods [170–173]. These problems are typically numerically ill-posed due to statistical errors in the data. The hope when choosing a contour which is in between the pure Euclidean and Minkowski contour is that the problem becomes less ill-posed since the distance to Minkowski space is lower.
The contours we will compare are

- a triangle (symmetric and asymmetric), which we parametrize by splitting the inverse temperature into $\tau = \tau_+ + \tau_-$. The $+$ part lies above the edge of the triangle and the $-$ part below. We split the number of points in temporal direction N_t equally between the upper and lower branch of the triangle. The temporal lattice spacing at point n on the contour reads

$$\begin{aligned} a_{t,\mathrm{up}}(n) &= n\frac{(t_0 - i\tau_+)}{N_t/2} \\ a_{t,\mathrm{lw}}(n) &= n\frac{(-t_0 - i\tau_-)}{N_t/2}\,. \end{aligned} \tag{3.59}$$

  We assume periodic boundary conditions.

- an ellipse with points equidistant on the contour. This is parametrized by

$$a_t(n) = t_0 \sin\left(\frac{\pi}{N_t}n\right) + i\frac{\tau}{2}\left(-1 + \cos\left(n\frac{\pi}{N_t}\right)\right) \tag{3.60}$$

- an ellipse with points equidistant in Euclidean direction. Since the ellipse as in the previous point does not have equidistant points and hence in the Euclidean limit ($t_0 \to 0$) does not give the correct result unless N_t is really large. The equation in this case reads

$$a_t(n) = it_e(n) + 2t_0\sqrt{\frac{t_e(n)(t_e(n)+\tau)}{\tau^2}}\,, \tag{3.61}$$

  where $t_e(n) = -n\tau/N$ corresponds to an equidistant partitioning of the Euclidean contour.

Triangles have already been investigated in [59, 103], and it has been found that in the case of SU(2) the symmetric triangle performed best. To compare the contours we perform simulations on $8^3 \times 16$ lattices with inverse temperature $\tau = 16$. For the asymmetric triangle we choose $\tau_+ = 1$ and $\tau_- = 15$. This choice is close to the

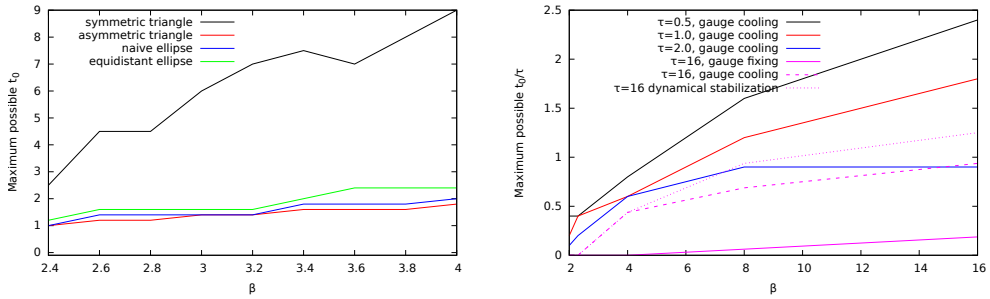

Figure 3.32: Left: Comparison of different contour choices for SU(2) real time simulations. Right: Investigation of the parameter range of the symmetric triangle contour. See text for a discussion.

real-time contour and if feasible will yield results close to Minkowski space physics as was shown for quantum mechanics in [59]. We vary β and t_0 to see how far in real time one can go. The criterion we choose for this is the following: The simulations are typically well thermalized after a Langevin time $t \approx 10$. To take averages one has to have enough data after thermalization to take averages, in this case, multiple runs can be done to produce higher statistics. We choose to say that the simulation works if the histogram of the simulation looks fine up to $t \approx 25$. Note that this is not an exact criterion since depending on the random numbers produced by the computer, deviations can occur earlier or later. In this case, we only did one run for each set of parameters and hence, the results for the possible range are a guideline and might differ slightly. Note also that while in figure 3.30 the simulation seems to be going on indefinitely, for some parameters the step size becomes very small and the simulation comes to a halt. The left plot in figure 3.32 shows the possible extent of the different contours using gauge cooling. All contours except the symmetric triangle seem to prohibit a larger real-time extent. In the case of the symmetric triangle, the range seems to increase with increasing β. We conclude that the symmetric triangle contour performs best.
Therefore we investigate the possible parameter range of this contour for the same parameters as before, the result is shown in the right plot of figure 3.32. There are two observations to be made. (i) for smaller lattice coupling β the possible real-time extent for given inverse temperature τ is rather limited. Especially in the interesting region around the critical beta of $\beta \approx 2.5$ barely any real time extent is possible. (ii) smaller τ allows for larger ratios t_0/τ. However, given that the ratio is plotted, the actual value of t_0 that is possible is larger for $\tau = 16$. In this figure, we also compare stabilization methods, which will be discussed below.

Comparison of stabilization methods

In [103] it was found that it is necessary to employ some method to keep the evolution close to the real manifold. The method that was chosen there was to use maximal tree gauge, i.e. the maximum possible number of gauge links is fixed to one, and

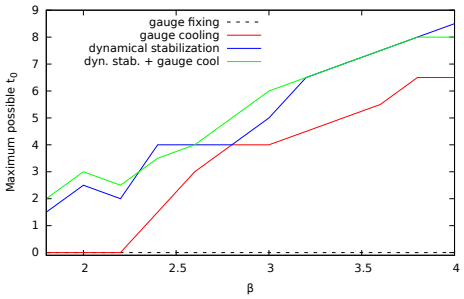

Figure 3.33: Comparison of different stabilization methods for SU(2) real-time simulations.

hence keeps the simulation somewhat close to the real manifold. Since then gauge cooling [105] and dynamical stabilization [128] have been shown to help with QCD simulations. Gauge cooling keeps the simulation close to the real manifold by gauge transformations while dynamical stabilization adds an attractive drift force towards the real manifold to the dynamics. For more details on both methods see appendix A.7. Here we compare four different methods

- maximal tree gauge fixing as in [103]
- gauge cooling where we found that a cooling parameter $\alpha_{\mathrm{gc}} = 1$ with five cooling sweeps is typically enough
- dynamical stabilization with $\alpha_{\mathrm{ds}} = 10$
- dynamical stabilization with one gauge cooling sweep.

Note that in the case of dynamical stabilization one usually compares different values of α_{ds} to find a region in which the additional drift does not change the physical outcome. Since in our simulations we have the additional advantage that we know the value of the spatial plaquette, we can fix a value for α_{ds} by simply choosing a value which makes the real manifold attractive enough and simultaneously keeps the spatial plaquette at its correct value. The comparison of those methods is shown in figure 3.33. We observe that in the investigated β region gauge fixing is not a valid option. The parameters chosen in [103] are such that they require currently unreachable lattice sizes to give a large enough physical lattice volume since larger β correspond to a smaller lattice spacing. The use of gauge cooling and dynamical stabilization opens up the possibility to simulate in a physically more interesting region. We also observe that dynamical stabilization outperforms gauge cooling though the difference is not too large.

3.3.1 Conclusion of SU(2) real-time simulations

We have shown that complex Langevin simulations employing gauge cooling and dynamical stabilization vastly improve the stability of complex Langevin simulation

compared to gauge fixing. We found that the best contour is a symmetric triangle. Unfortunately, the symmetric triangle is far enough away from the real-time contour that some kind of analytic continuation still has to be performed. The parameter range of the contour is also limited to short real-time extents t_0.

3.4 Summary, Outlook, and lessons for complex Langevin

The complex Langevin method has proved to be a powerful tool to attack the sign problem. Especially in SU(N) simulations and for QCD there has been much progress. It enabled us to systematically investigate the phase transition of full QCD for a large range of μ/T, which was not possible before. In particular, we were able to compute the curvature of the transition line for up to $\mu/T_c(0) \approx 5$, which is out of the reach of other approaches so far. We find values of the curvature of $\kappa_2 \approx 0.0011$ to 0.0015 depending on volume and observable. Since our pion masses are between 1 to 2 GeV this fits nicely between the results of HDQCD [83], where there appears to be no curvature and simulations at physical pion masses, where $\kappa_2 \approx 0.014$ [19, 49]. We were also able to see first signals of the liquid-gas transition at low temperature, though there were some systematic errors that prohibited a precise evaluation of the data. This should be improvable by applying dynamical stabilization. The next step should be simulations closer to the physical point to see if complex Langevin still performs well or if poles start to appear.
For SU(2) Yang-Mills on a contour, we were able to improve upon previous simulations and to simulate in physically more interesting regions than before. However, simulations directly on a real-time contour were not possible as well and there still are restrictions to the types of contours that are possible.

Lessons from QCD and SU(2): How to use complex Langevin

In addition to all systematics that have to be controlled in standard lattice simulations, complex Langevin adds a few points that have to be taken care of. Here we give a brief summary of those.

- Avoid regions of wrong convergence. They are rather simple to identify by looking at histograms, boundary terms or the unitarity norm, or –in the case of our SU(2) simulations– by comparing to the spatial plaquette. Many complex Langevin simulations have been done in inaccessible parameter regions with the conclusion that complex Langevin is not a valid method. In QCD we were able to go below the phase transition anyway by varying N_t instead of β as is usually done. Hence, unconventional ideas may be able to help in making complex Langevin work better.

- Avoid long runs. Even with gauge cooling, the unitarity norm tends to grow large after some time due to the accumulation of numerical errors. It is best to take measurements only as long as the unitarity norm stays small. Typically runs are already rather long before the rise of the unitarity norm. Longer runs may be possible using dynamical stabilization. It is best to find out at which value the unitarity norm is too large, i.e. signals boundary terms, and discard all configurations after that value was reached.

When those points are taken into account the complex Langevin method can be applied. Unfortunately, in some cases the interesting regions are inaccessible and it is not clear how to modify the simulation to make them accessible – such as QCD before gauge cooling existed. Hence, there is no "set it and forget it" recipe to apply complex Langevin to any theory. When complex Langevin works it is rather reliable and can solve many interesting physical phenomena.

4 Lefschetz Thimbles

An alternative approach to the sign problem is that of Lefschetz thimbles and generalized complex manifolds. It is related to the complex Langevin method since it also starts with complexification of the original integration space. The idea of the Lefschetz thimble method is to find paths in the complexified space on which there is no sign problem such that Monte Carlo methods can be applied. The method was introduced in a different context in [174] and was first applied to lattice simulations in [85]. It has since been applied to the complex Bose gas [175] and mainly to simple models [30, 176–183]. Some more algorithms also appeared in [184–187]. There have also been some efforts to combine the Lefschetz Thimble method with complex Langevin [30, 188]. One of the problems with Lefschetz thimbles is that a theory usually has many of them and that simulations are only possible on one at a time. This is circumvented in the generalized Lefschetz Thimble method [87] or path optimization method [189]. Both of these methods try to find manifolds which lie close to Lefschetz thimbles but are continuous deformations of the original integration domain such that they are homotopy equivalent to it. There are several approaches to find those manifolds. The most straight forward one is to use the steepest ascent flow which asymptotes towards the thimble. With this method models of QCD [190], real time quantum mechanics [191, 192] and actual field theories in low volumes [193–195] have been investigated. For algorithmic improvements see [196–198]. Another way to find such manifolds is a physical ansatz, which of course is model dependent, see [86, 199] for an application in the Thirring model. Finally, there are some works parametrizing such manifolds using neural networks [27, 28, 200–202] by either directly finding a manifold or by optimizing the parameters of a physical ansatz. The problem with those generalized approaches to Lefschetz thimbles is that they only ameliorate the sign problem such that reweighting (see section 2.4) is still necessary, hence they are not easily scalable to high volumes. They are also oftentimes rather expensive. Hence, so far simulations have been possible only in low volumes.

In this chapter, we will focus on Lefschetz thimbles and not their generalizations. We will briefly introduce the method in section 4.1. We then discuss its issues and give some ideas for solutions which we demonstrate in simple models in section 4.2. We briefly discuss the relation of complex Langevin and Lefschetz thimbles in section 4.3. Finally, we investigate the difficulties that arise when applying the method to field theories in section 4.4.

4.1 Introduction to Lefschetz Thimbles

The mathematics behind Lefschetz thimbles is described by Picard-Lefschetz theory, a complexification of Morse theory. We are interested in the application to oscillatory integrals [203]. Here we only state the important equations and do not go into mathematical details.
Consider an action $S(x) \in \mathbb{C}$ with $x \in \mathbb{R}$. As in complex Langevin, our first step is to complexify, i.e. $S(z) \in \mathbb{C}$ with $z \in \mathbb{C}$. We then compute the stationary points z_σ, which are the solution to

$$\left.\frac{\partial S}{\partial z}\right|_{z=z_\sigma} = 0\,. \tag{4.1}$$

Note that for the method to be applicable the z_σ need to be non-degenerate. The Lefschetz (anti-/unstable) thimbles are then the paths D_σ in the complex plain which follow the steepest (ascent/)descent equation

$$\frac{\partial z}{\partial \tau} = \pm\overline{\frac{\partial S}{\partial z}}\,. \tag{4.2}$$

and end in the stationary points. The plus/minus sign is for the steepest ascent/descent equation. Integrals of the form of the partition function can then be decomposed according to

$$Z = \int_{\mathbb{R}} e^{-S} = \sum_\sigma n_\sigma \int_{D_\sigma} e^{-S}\,, \tag{4.3}$$

where the sum goes over all stationary points. The number n_σ is the oriented intersection number of the corresponding anti-thimble with the original manifold. One can see the advantage of this formulation by differentiating the imaginary part of the action

$$\begin{aligned} 2i\frac{\partial}{\partial\tau}\text{Im } S(z) = \partial_\tau\left(S(z) - \bar{S}(z)\right) &= \frac{S(z)}{\partial z}\frac{\partial z}{\partial\tau} - \overline{\frac{S(z)}{\partial z}}\overline{\frac{\partial z}{\partial\tau}} \\ &= \frac{S(z)}{\partial z}\overline{\frac{S(z)}{\partial z}} - \overline{\frac{S(z)}{\partial z}}\frac{S(z)}{\partial z} = 0\,, \end{aligned} \tag{4.4}$$

where we inserted equation (4.2). Thus the imaginary part of the action along the thimbles and anti-thimbles is a constant and can be pulled out of the integral. The partition function becomes

$$Z = \sum_\sigma n_\sigma e^{-i\text{Im } S(z_\sigma)} \int_{D_\sigma} e^{-\text{Re } S(z)} \equiv \sum_\sigma n_\sigma e^{-i\text{Im } S(z_\sigma)} Z_\sigma\,, \tag{4.5}$$

which defines the partition function Z_σ on the thimble. The Boltzmann factor along the thimble now has the nice property that it is no longer oscillatory since it is purely real. Furthermore, it is maximal at the stationary point and decays

exponentially. Thus we have a well-defined integral which we can sample using Monte Carlo techniques. Observables on one thimble are computed via

$$\begin{aligned}\langle \mathcal{O} \rangle &= \frac{1}{Z} \sum_\sigma n_\sigma e^{-i\mathrm{Im}\ S(z_\sigma)} \int_{D_\sigma} dz\, \mathcal{O} e^{-\mathrm{Re}\ S(z)} = \frac{1}{Z} \sum_\sigma n_\sigma e^{-i\mathrm{Im}\ S(z_\sigma)} \langle \mathcal{O} \rangle_\sigma \\ &= \frac{\sum_\sigma n_\sigma e^{-i\mathrm{Im}\ S(z_\sigma)} Z_\sigma \langle \mathcal{O} \rangle_\sigma}{\sum_\sigma n_\sigma e^{-i\mathrm{Im}\ S(z_\sigma)} Z_\sigma} \,. \end{aligned} \tag{4.6}$$

While this method leads to an integration over a well defined probability measure, it has its own drawbacks

- There is a residual sign problem from the Jacobian for the transformation of the integral.
- In a general field theory it can be hard to find all stationary points and determine which of the thimbles contribute.
- In the case of multiple thimbles we need to know the partition function on the thimble, which cannot be computed via Monte Carlo methods.

For most applications the first point usually does not pose a problem, it seems that the sign problem from the Jacobian is much weaker than the sign problem of the original theory. The last two problems are challenging and there is no general solution. Some ideas on how to find stationary points and thimbles are described in appendix B.1. For now, we deal only with simple models, where finding the fixed points is simple and the partition functions can be computed by numerical integration such that we can focus on conceptual problems.

4.2 Monte Carlo simulations on Lefschetz Thimbles in simple models

We find the (anti-)thimble by starting arbitrarily close the the fixed point and numerically solving the normalized steepest (descent/)ascent equation

$$\frac{\partial z}{\partial \tau} = \pm \overline{\frac{\partial S}{\partial z}} \Big/ \left| \overline{\frac{\partial S}{\partial z}} \right| \,, \tag{4.7}$$

where the normalization has been done for convenience. Having a parametrization of the thimble in terms of the parameter τ the partition function on the thimble becomes

$$Z_\sigma = \int_{D_\sigma} dz\, e^{-\mathrm{Re} S(z)} = \int_a^b d\tau\, e^{-\mathrm{Re} S_\sigma(\tau)} J_\sigma(z(\tau)) \,, \tag{4.8}$$

where the Jacobian now introduces the residual sign problem mentioned above. We perform standard Monte Carlo simulations on the thimble sampling from $\exp(-\mathrm{Re}S(\tau))$ and deal with the Jacobian via reweighting, i.e.

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O}\, J_\sigma \rangle}{\langle J_\sigma \rangle} \,. \tag{4.9}$$

For multiple thimbles this equation becomes

$$\langle \mathcal{O} \rangle = \frac{\sum_\sigma n_\sigma e^{-i\mathrm{Im}S(z_\sigma)} Z_\sigma^r \langle \mathcal{O}\, J_\sigma \rangle_\sigma^r}{\sum_\sigma n_\sigma e^{-i\mathrm{Im}S(z_\sigma)} Z_\sigma^r \langle J_\sigma \rangle_\sigma^r} \,, \tag{4.10}$$

with

$$\begin{aligned} \langle \mathcal{O} \rangle_\sigma^r &= \frac{1}{Z_\sigma^r} \int_a^b d\tau\, e^{-\mathrm{Re}S_\sigma(\tau)} \mathcal{O} \\ Z_\sigma^r &= \int_a^b d\tau\, e^{-\mathrm{Re}S_\sigma(\tau)} \,. \end{aligned} \tag{4.11}$$

The evaluation of the full partition function requires knowledge about the partition functions on all contributing thimbles, which is a priori not known. In [177] it has been solved by normalizing the partition functions at a point where simulations are possible without a sign problem, which is not always possible. Instead we solve this problem by rewriting the ratio of the partition functions of two different thimbles

$$\begin{aligned} \frac{Z_1^r}{Z_2^r} &= \frac{\int_a^b d\tau\, e^{-\mathrm{Re}[S_1(\tau)]}}{\int_a^b d\tau\, e^{-\mathrm{Re}[S_2(\tau)]}} = \frac{\int_a^b d\tau\, e^{-\mathrm{Re}[S_1(\tau)+S_2(\tau)-S_2(\tau)]}}{\int_a^b d\tau\, e^{-\mathrm{Re}[S_2(\tau)]}} \\ &= \frac{\int_a^b d\tau\, e^{-\mathrm{Re}[S_2(\tau)]} e^{\mathrm{Re}[S_2(\tau)-S_1(\tau)]}}{\int_a^b d\tau\, e^{-\mathrm{Re}[S_2(\tau)]}} = \left\langle e^{\mathrm{Re}[S_2(\tau)-S_1(\tau)]} \right\rangle_2^r \,. \end{aligned} \tag{4.12}$$

This means that we can compute the ratio of partition functions on thimbles by computing a simple observable on one of the thimbles. There are some subtleties that have to be taken care of here. First, we implicitly assumed that the parameters τ on both thimbles can be identified. In our case this is automatically fulfilled since we normalized equation (4.7), otherwise, an additional Jacobian would occur. The second is that there has to be sufficient overlap of the distributions on the thimble. To some extent, this is fulfilled since in all cases the distribution on the thimble has its maximum at the fixed point. If one thimble ends at some finite value there may be no overlap in some region such that the ratio from equation (4.12) cannot be computed.
This can be solved by simply mapping the boundaries of all thimbles into the same interval, e.g. $\tau \in [0,1]$. If the thimble originally is defined on a finite interval $\tau \in [a,b]$ this can be done by a linear transformation. If $\tau \in [-\infty, a]$ a possible transformation is

$$x \to x' = 1 + \tanh\left(\xi\,(x-a)\right) \,, \tag{4.13}$$

and accordingly for $\tau \in [a, \infty]$

$$x \to x' = \tanh\left(\xi\left(x - a\right)\right) , \tag{4.14}$$

for $\tau \in [-\infty, \infty]$ the mapping becomes

$$x \to x' = \frac{1 + \tanh\left(\xi x\right)}{2} , \tag{4.15}$$

where ξ is a tunable parameter which can be used to maximize the overlap of the distributions.

4.2.1 Applications

We test our procedure to compute equation (4.10) in three simple models

1. A quartic action with a symmetry breaking term such that there is no $z \to -z$ symmetry

$$S(z) = \frac{\sigma}{2} z^2 + \frac{\lambda}{4} z^4 + hz , \tag{4.16}$$

with $\sigma = 1$, $\lambda = 1/3$ and $h = 1 + i$, here only one thimble contributes.

2. The same action with $\sigma = 1$, $\lambda = 1$ and $h = 1 + i$, where two thimbles contribute

3. A U(1) one link model, somewhat modeling the occurrence of a fermion determinant via the logarithm

$$S(x) = -\beta \cos(x) - \log\left(1 + \kappa \cos(x - i\mu)\right) , \tag{4.17}$$

where $\kappa = 2$, $\beta = 1$ and $\mu = 2$. This model is particularly interesting, since complex Langevin fails to give correct results due to the occurrence of poles [204].

Figure 4.1 shows the thimble structures of all models. The flow lines in the background are the negative gradient of the action $-\partial_z S$. This is the flow that describes complex Langevin, we will discuss the connection further below. (Anti-)thimbles are given by (dashed)solid lines. One can see that in the first model only one thimble contributes since only one anti-thimble has an intersection with the real axis. In the second model, two thimbles contribute. The third model is periodic in the real part such that the green thimbles should be identified. Here three thimbles contribute.
We apply our algorithm to those models. For the first model only one thimble contributes. We show the result for the observable $\langle z^2 \rangle$ in table 4.1, everything agrees nicely.
For the second model we map the distributions to the interval $\tau \in [0, 1]$ according to equation (4.15) with $\xi = 0.25$, since both thimbles end at infinity in both directions.

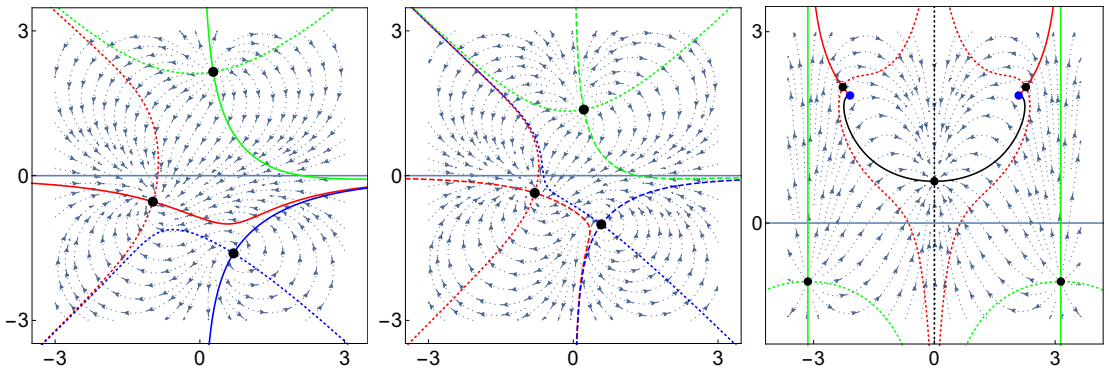

Figure 4.1: Complex plane with thimble structure and drift of the three models in order from left to right. See text for details.

We show the corresponding distributions in the left plot of figure 4.2. They have sufficient overlap and we can compute the ratio of partition functions. Results are shown in table 4.1, where again all observables agree with the analytical solutions. The third model is the most interesting case. Here the thimbles end in poles, such that we need to apply equations (4.13), (4.14) with $\xi = 1.5$ for the outer thimbles and a linear transformation for the middle thimble. The distributions resulting from the mappings are shown in the right plot of figure 4.2, the overlap is still sufficient for simulations. Observables are zero dimensional versions of the Polyakov loop and its inverse, the plaquette and the number density

$$\begin{aligned} \langle U \rangle &= \left\langle e^{ix} \right\rangle , \\ \left\langle U^{-1} \right\rangle &= \left\langle e^{-ix} \right\rangle , \\ \langle P \rangle &= \langle \cos(x) \rangle , \\ \langle n \rangle &= \left\langle \frac{i\kappa \sin(x - i\mu)}{1 + \kappa \cos(x - i\mu)} \right\rangle . \end{aligned} \tag{4.18}$$

The results are shown in table 4.1, again everything agrees. We conclude that our algorithm for combining thimbles works, at least in simple models. In principle, this is applicable also in higher dimensions as long as the variables of the parametrizations overlap. A similar idea has appeared in [177, 178].

4.3 Thimbles and complex Langevin

Intuitively the Lefschetz Thimble and complex Langevin method are related. Both start with the complexification of the integration space and in both methods the fixed points are important, in complex Langevin by driving the dynamics of the simulation and for Lefschetz thimbles by being their anchors. It has been shown [204] that the complex Langevin evolution typically clusters around contributing

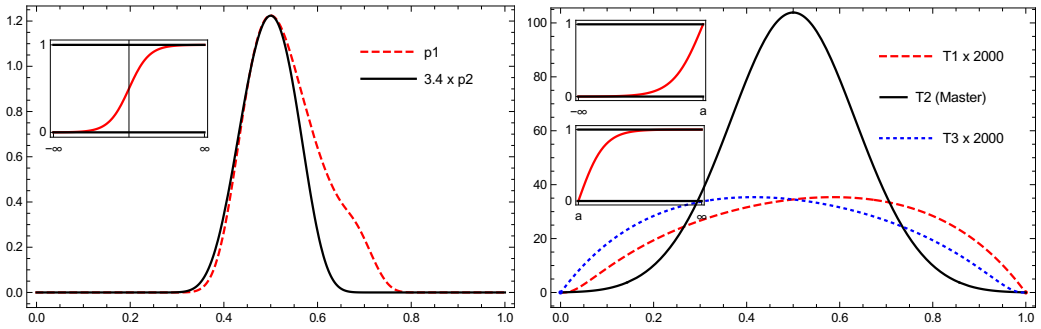

Figure 4.2: Boltzmann factor $\exp(-\mathrm{Re}(S))$ vs. flow parameter τ on the contributing thimbles. Left: Model 2, the thimble structure is shown in the middle of figure 4.1 $p1$ corresponds to the red thimble, $p2$ to the black one. Right: Model3, corresponding to the rightmost plot in figure 4.1. We call the black thimble the "master"-thimble, $T1$ indicates the left red thimble and $T3$ the right red thimble, they are connected by symmetry. The inlays correspond to the transformation that were used to map the thimbles to the interval $[0, 1]$, see equations (4.13)-(4.15)

.

$\mathcal{O}$	numerical	exact
z^4–1 thimble		
$\mathrm{Re}z^2$	0.73922(6)	0.73922
$\mathrm{Im}z^2$	0.63006(4)	0.630089
z^4–2 thimbles		
$\mathrm{Re}z^2$	0.509299(5)	0.509297
$\mathrm{Im}z^2$	0.305819(3)	0.305815
$Z_2/Z_1\|_{T_1}$	0.2253778(4)	0.2253779
$Z_1/Z_2\|_{T_2}$	4.436(12)	4.437
$\mathrm{Re}\langle U\rangle$	0.315217(3)	0.315219
$\mathrm{Re}\langle U^{-1}\rangle$	1.800941(3)	1.800939
$\mathrm{Re}\langle P\rangle$	1.058079(3)	1.058079
$\mathrm{Re}\langle n\rangle$	0.742861(1)	0.742860
$Z_2/Z_1\|_{T_1} \times 10^{-3}$	2.99378(3)	2.99382
$Z_1/Z_2\|_{T_2} \times 10^{4}$	3.34032(4)	3.34022
$Z_2/Z_3\|_{T_3} \times 10^{-3}$	2.99377(3)	2.99382
$Z_3/Z_2\|_{T_2} \times 10^{4}$	3.34026(9)	3.34022

Table 4.1: Numerical results and exact values of observables for the three models with statistical errors. Imaginary parts of some observables agree with zero and thus are not shown.

thimbles. Hence, the natural question is if one can combine both methods. Complex Langevin has been combined with the generalized Lefschetz thimble method in [188] in a simple model.
We briefly discuss our insights from trying to combine both methods, details can be found in appendix B.2. It is well known that complex Langevin can be improved by appropriate coordinate transformations [205], which is e.g. also the case for the Polyakov chain from 3.1.2, where the gauge fixed and diagonalized version is much more stable than the full- or only gauge fixed chain.
Hence we tried to use coordinate transformations to keep the complex Langevin evolution close to thimbles. I.e. we introduce a transformation $z \to z(u, \theta)$, where u is the new variable and θ is a parameter which is updated every Langevin step such that the transformation is always able to get close to thimbles.
In practice, this is done by introducing a constraint into the complex Langevin evolution. This constraint should be derived from some quantity which measures the distance of the complex Langevin evolution to the thimbles, which is not a well-defined quantity. We found that some measures for the distance can significantly destabilize the evolution by introducing poles, which is a problem for complex Langevin as we have seen in chapter 3. However, even without poles, it is hard to find a criterion to measure the distance to all thimbles, which also allows jumps between thimbles. While we found a way to push the distribution of the fields close to the thimbles and keep the simulation stable by trial and error – see appendix B.2 – we were unable to get correct values for observables. Thus we conclude that much more mathematical insight is needed to combine both methods, this is work in progress.

4.4 Towards higher dimensions

So far we simulated on the thimbles by explicitly finding a parametrization. In higher dimensions, this will become numerically unfeasible. Unfortunately, global simulation algorithms are also expensive. Hence, we would like to find local updating algorithms similar to the Metropolis algorithm to make simulations of interesting theories possible. We do this by using the fact that the imaginary part of the action does not change along the thimble, i.e.

$$\Delta S_I = 0\,. \tag{4.19}$$

Instead of writing down a global algorithm, which requires finding the thimble of the whole lattice, we would like to find an algorithm which allows local updates. We will explain this idea in more detail using an example.

4.4.1 U(1) gauge theory

Here we are interested in U(1) gauge theory on a $1+1d$ lattice with the action

$$S = -\beta \sum_{U_p} \left(\frac{1}{2} \left(U_p + U_p^{-1} \right) - 1 \right) , \tag{4.20}$$

where we chose $\beta \in \mathbb{C}$ to introduce a sign problem. Note that this is just the Wilson gauge action from equation (2.23). In the thermodynamic limit, the expectation value of the plaquette for this theory is known analytically, see e.g. [36], as

$$\langle U_p \rangle = \frac{I_1(\beta)}{I_0(\beta)}\,, \tag{4.21}$$

where $I_n(\beta)$ are the modified Bessel functions of the first kind.
In order to describe the full theory we now need to find all stationary points and thimbles. For a lattice theory, the stationary points are given by whole lattice configurations. For now, we will concentrate on the thimble corresponding to the classical fixed point, where $S = 0$. Hence, we proceed to describe how to do updates on a global thimble using local updates by keeping the imaginary part of the action constant. By writing the action as

$$S = \beta\Sigma = (\beta_R\Sigma_R - \beta_I\Sigma_I) + i(\beta_I\Sigma_R + \beta_R\Sigma_I)\,, \tag{4.22}$$

we see that this is equivalent to demanding

$$S_I = \beta_I\Sigma_R + \beta_R\Sigma_I = \text{const}\,. \tag{4.23}$$

The local action of a single link is given by

$$\begin{aligned}\frac{S(U_k(n))}{\beta} &= 2 - \frac{1}{2}\left(U_{ki}(n) + U_{kj}^{-1}(n-\hat{i}) + U_{kj}(n-\hat{i}) + U_{ki}^{-1}(n)\right)\\ &= 2 - \frac{1}{2}\Big[U_k(n)\left(V_i(n) + V_j^{-1}(n-\hat{i})\right)\\ &\quad + U_k^{-1}(n)\left(V_j(n-\hat{i}) + V_i^{-1}(n)\right)\Big]\\ &\equiv 2 - \left(U_k(n)V_{ij}(n) + U_k^{-1}V_{ji}(n-\hat{i})\right)\,,\end{aligned} \tag{4.24}$$

since each link appears in two different plaquettes. We split the link from the so-called staples V_i and V_j and introduced the sum of two staples V_{ij}. We drop the dependence on the site n from now on since it is unambiguously clear which to use for U_k and U_k^{-1}. In case of an update $U \to U'$ the difference will be

$$\frac{S(U_k'(n))}{\beta} - \frac{S(U_k(n))}{\beta} = V_{ij}(U_k - U_k') + V_{ji}(U_k^{-1} - U_k'^{-1}) \equiv \frac{\delta}{\beta}\,. \tag{4.25}$$

If we parametrize the link as $U_k = e^{i\phi_k}$ this leads to $U_k' = e^{i\phi_k}e^{i\Delta\phi_k} = U_k\Delta U_k$. With $P_{ij} \equiv U_kV_{ij}$ the difference thus becomes

$$\frac{\delta}{\beta} = P_{ij}(1 - \Delta U_k) + P_{ji}\left(1 - \Delta U_k^{-1}\right)\,, \tag{4.26}$$

if we now split $\Delta U = |\Delta U|\, e^{i\Delta\phi^{\mathrm{R}}}$, we see that

$$\text{Im}\ \delta = \text{Im}\ (\beta(P_{ij}+P_{ji})) - |\Delta U_k|\text{Im}\ (\beta P_{ij}e^{i\Delta\phi_k^R}) - |\Delta U_k|^{-1}\text{Im}\ (\beta P_{ji}e^{-i\Delta\phi_k^R})\,, \tag{4.27}$$

which has to be zero such that the imaginary part of the action stays constant. By multiplying through with $|\Delta U_k|$ this becomes a simple quadratic equation with solution

$$|\Delta U_k| = \frac{\mathrm{Im}(\beta(P_{ij}+P_{ji})) \pm \sqrt{[\mathrm{Im}(\beta(P_{ij}+P_{ji}))]^2 - 4\mathrm{Im}(\beta P_{ij}e^{i\Delta\phi_k^R})\mathrm{Im}(\beta P_{ji}e^{-i\Delta\phi_k^R})}}{2\mathrm{Im}(\beta P_{ij}e^{i\Delta\phi_k^R})}\,. \tag{4.28}$$

The update is done by choosing a random number for $\Delta\phi_k^R$, computing the corresponding $|\Delta U_k|$ and by doing an accept reject step in the end. There are some subtleties to be taken care of such as the regions where the square root is not defined, see appendix B.3 for the exact updating algorithm.
Note that by doing the updating in that way, we parametrize the partition function on the thimble in terms of the algebra fields

$$\tilde{\phi}_k(\phi_k^R) = \phi_k^R + i\phi_k^I(\phi_k^R)\,, \tag{4.29}$$

in the following way

$$Z_\sigma = \frac{1}{(2\pi)^{2|\Lambda|}} \int_{\min(\phi_1)(\phi^R)}^{\max(\phi_1)(\phi^R)} d\phi_1^R \cdots \int_{\min(\phi_{2|\Lambda|})(\phi^R)}^{\max(\phi_{2|\Lambda|})(\phi^R)} d\phi_{2|\Lambda|}^R e^{-S(\tilde{\phi}(\phi^R))} \det\left[\frac{\partial\tilde{\phi}_k}{\partial\phi_l^R}\right]\,, \tag{4.30}$$

where $|\Lambda|$ is the lattice volume. The boundaries of the integral may be finite due to fields only being defined in certain regions, see appendix B.3. The Jacobian is still global, it reads

$$\frac{\partial\tilde{\phi}_k}{\partial\phi_l^R} = \frac{\partial(\phi_k^R + i\phi_k^I(\phi^R))}{\partial\phi_l^R} = \delta_{k,l} + i\frac{\partial\phi_k^I(\phi^R)}{\partial\phi_l^R}\,. \tag{4.31}$$

We rewrite the last term according to

$$\begin{aligned}
\frac{\partial\phi^I(\phi^R)}{\partial\phi^R} &= \frac{\partial\phi^I(\phi^R+\Delta\phi^R)}{\partial\phi^R}\bigg|_{\Delta\phi^R=0} = \frac{\partial\phi^I(\phi^R+\Delta\phi^R)}{\partial\Delta\phi^R}\bigg|_{\Delta\phi^R=0} \\
&= \left(\frac{\partial\left(\phi^I(\phi^R+\Delta\phi^R) - \phi^I(\phi^R)\right)}{\partial\Delta\phi^R}\right)\bigg|_{\Delta\phi^R=0} \\
&= \left(\frac{\partial\Delta\phi^I\left(\Delta\phi^R\right)}{\partial\Delta\phi^R}\right)\bigg|_{\Delta\phi^R=0}\,.
\end{aligned} \tag{4.32}$$

This is convenient, since we can just take the derivative of $\Delta\phi^I = -\log|\Delta U|$, see equation (4.28).

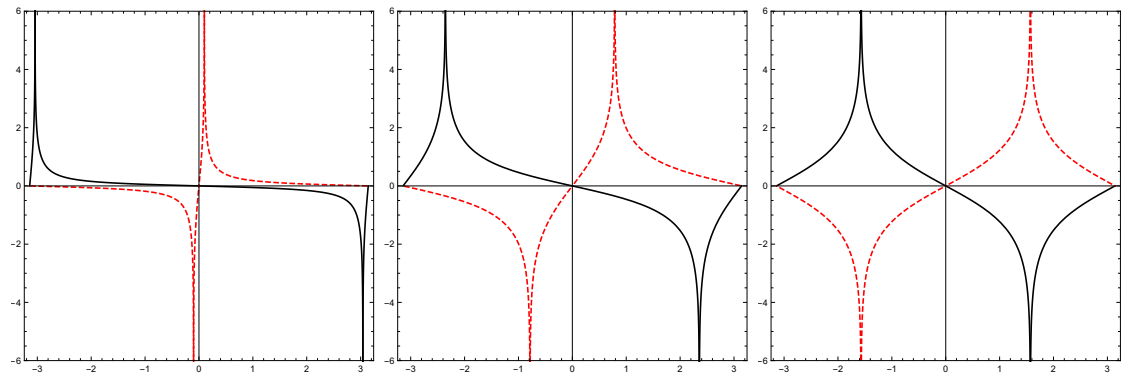

Figure 4.3: Complex plane with thimble for the model in equation (4.33) for $\beta = 1 + 0.1i$, $\beta = 1 + i$ and $\beta = i$ from left to right. Full black lines are the thimbles, red dashed lines the anti-thimbles. See text for a discussion.

Application to one link

We first test this procedure on one link, i.e. we set the staples to one. The action then reads

$$S = \beta \left(1 - \frac{1}{2}\left(U + U^{-1}\right)\right) = \beta \left(1 - \frac{1}{2}\left(e^{i\phi} + e^{-i\phi}\right)\right) = \beta(1 - \cos(\phi))\,, \quad (4.33)$$

where the constant part will drop out in the simulation, thus we ignore it. Note that for purely imaginary β this is the same action we used to investigate the boundary terms in section 3.1.1, see equation (3.14). We show the thimbles structure for different β in figure 4.3. For small imaginary part of β the thimble is close to the real axis, ultimately becoming the real axis at Im $\beta = 0$. For Re $\beta = 0$ both thimbles look the same just shifted by π. In this case, they both are equally important, i.e. they contribute equally to the partition function. We tested our algorithm in this model. Here the Jacobian is given by

$$J = 1 + i\frac{-d\log\,|\Delta U|}{d\Delta\phi^R} = 1 + i\frac{\text{Im}\; i\beta U + \text{Im}\; i\beta U^{-1}}{\text{Im}\; \beta U - \text{Im}\; \beta U^{-1}}\,, \quad (4.34)$$

see equation (4.32). We can compute observables according to the reweighting prescription in equation (4.9). On the main thimble the simulation yields $\langle e^{i\phi}\rangle = 0.840(5) + 0.060(15)i$ for $\beta = 3 + i$, while numerical integration along the thimble yields $\langle e^{i\phi}\rangle = 0.839 + 0.064i$. The algorithm yields correct results within error bars in the simple model.

Before we apply the algorithm to higher dimensions, we look at the relative importance of the thimbles in the one link model. Figure 4.4 shows the situation for different β. The left and center plot show that for fixed β_I the thimble corresponding to the stationary point at zero will yield the main contribution if β_R is large enough, i.e. when $\beta_R > \beta_I$. However, for small β_R there is no convergence in β_I. This means that in the most interesting case $\beta = i$, which corresponds to a real-time simulation,

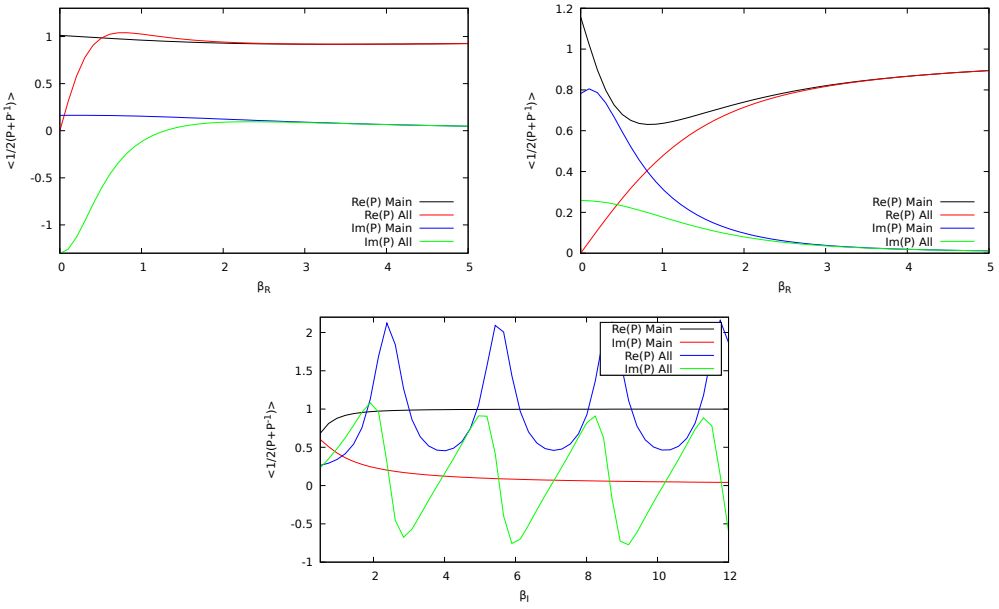

Figure 4.4: Expectation value of the plaquette for fixed β_I with $\beta_I = 0.5$ (left) and $\beta_I = 3$ (center) as a function of β_R. The same for fixed $\beta_R = 0.5$ as a function of β_I (right). The relevance of the main thimble is discussed in the text.

all thimbles contribute equally. This is interesting since here complex Langevin fails to yield correct results, see section 3.1.1.

U(1) in 1+1 dimensions

The same algorithm can straightforwardly be applied to higher dimensions. Unfortunately, the algorithm no longer is stable in 1+1 dimensions. First of all, even though initially the imaginary part stays constant, the simulation is driven away from the unitary manifold very fast. This can be seen by monitoring the unitarity norm, see equation (3.39). Since the imaginary part of the action stays constant, this is a priori not a problem. For gauge theories, the simulation can stay on the complexified gauge orbit of the thimble and still yield correct results. In practice, this introduces large differences in orders of magnitude of the variables thus destabilizing the simulation. This problem can be dealt with by means of gauge cooling [105] see appendix A.7, which slows down this rise of the unitarity norm considerably.
Even then the simulation is unstable, especially at small Reβ. We found that after some time configurations with $S = 0$ can no longer be gauge transformed back to the fixed point, suggesting that the simulation was able to jump on a different manifold with the same S_I at some point during the simulation. However, for such a complicated theory, we lack insight into the underlying structure of the manifold. Hence, we look at a simpler lattice theory to find out how to improve our algorithm.

4.4.2 Scalar theory on two lattice points

The action for a free scalar field on a lattice with two points and periodic boundary conditions reads

$$S = -i\left[(\phi_1 - \phi_2)^2 - \frac{m^2}{2}(\phi_1^2 + \phi_2^2)\right], \tag{4.35}$$

where the fields ϕ_i live on the sites i with $i = 1, 2$ and the factor $-i$ mimics a sign problem from real time simulations. In this theory, one can show that the tangent space at the critical point is identical to the thimble, which allows for a convenient parametrization of the action. See appendix B.4 for a detailed computation. In the end we find an analytical expression for an observable, namely

$$\langle\phi^2\rangle = \left\langle\frac{\phi_1^2 + \phi_2^2}{2}\right\rangle = -\frac{i}{2}\left(\frac{1}{m^2} + \frac{1}{m^2 - 4}\right), \tag{4.36}$$

which is useful to compare to results from our simulations

Application of our algorithm

We apply the same algorithm as for U(1), i.e. we demand that $\Delta S_I = 0$ and vary the fields along that constraint. We parametrize the fields as $\phi_k = \phi_k^R + i\phi_k^I$ and propose an update according to $\phi_k^R \to \phi_k^R + \Delta\phi_k^R$. We use the condition $\mathrm{Im}[S(\phi') - S(\phi)] = 0$ to compute $\Delta\phi_k^I$. In this theory this turns out much simpler than the algorithm in the U(1) model[14], which is described in appendix B.3. Here we only have one stationary point, hence the corresponding thimble gives results for the full theory. Since the action is quadratic, the equation for $\Delta\phi_k^I(\Delta\phi_k^R)$ still contains a square root

$$\Delta\phi_k^I = \frac{-\mu\phi_k^I - 2\phi_j^I \pm \sqrt{\mu^2\Delta\phi_k^R + \left(\mu\phi_k^I + 2\phi_j^I\right)^2 + 2\mu\Delta\phi_k^R\left(\mu\phi_k^R + 2\phi_j^R\right)}}{\mu} \tag{4.37}$$

where $k \neq j$ and we introduced the abbreviation $\mu = (m^2 - 2)$. This equation still has undefined regions due to the square root. The boundaries can simply be found by numerically finding the root in $\Delta\phi_k^R$ for the term under the square root. The algorithm is then similar to the U(1) case. Here we find that the solution with a plus corresponds to the thimble when $\Delta\phi_k^R < 0$ and the solution with a minus corresponds to the thimble when $\Delta\phi_k^R > 0$. Note, however, that the algorithm works only for $m > 2$, see appendix B.4 for details. We update a single site according to

1. Draw a Gaussian random number with mean zero and standard deviation one. If it is smaller than zero, choose the plus solution, otherwise choose the minus solution in (4.37).

[14] The reasons are twofold: First, here we do not have an absolute value and thus do not get a constraint from a quantity being positive. Second, the theory is not periodic and hence the space of possible values extends to infinity. Thus there are fewer constraints on the proposable values for $\delta\phi$.

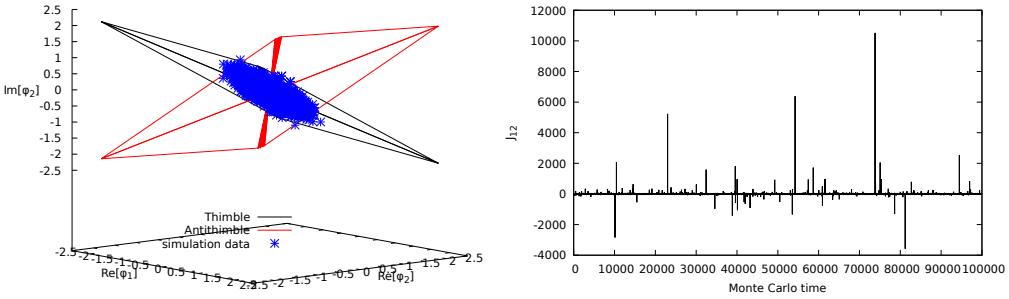

Figure 4.5: Left: Projection of the thimble structure with simulation data. Right: Evolution of part of the Jacobian. See text for details.

2. Find the boundary where the square root becomes real.

3. Add the random number to the position of the boundary $\Delta\phi_k^R = \text{boundary} + \text{random number}$, this gives the proposal for the field.

4. Do a Metropolis accept-reject step.

We simulated this for $m^2 = 6$. The left plot of figure 4.5 shows a projection of the thimble structure, the simulation lies on the thimble and is stable. To get expectation values we also need the Jacobian, which is computed according to (4.31) and below, i.e. it is just the derivative of equation (4.37) with respect to $\Delta\phi^R$. We show the Monte Carlo evolution of one of the entries of the Jacobian in the right plot of figure 4.5. The equation for that entry of the Jacobian is

$$J_{12}(\phi) = \frac{4\phi_{2,R}}{\sqrt{(4\phi_{1,I} - 4\phi_{2,I} + 2m^2\phi_{2,I})^2}}\,. \tag{4.38}$$

The enormous spikes in figure 4.5 occur when numerator and denominator have largely different orders of magnitude and often even lead to numerical divergences, i.e. "NaN".
Alternatively we can use the Takagi vectors v_i – see appendix B.4 – as a basis. I.e. we update a generic field configuration as

$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \to \begin{pmatrix}\phi_1 + \Delta\phi_1\\ \phi_2 + \Delta\phi_2\end{pmatrix} = \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} + \xi_1 v_1 + \xi_2 v_2 \tag{4.39}$$

where the ξ_i are just Gaussian random numbers. We finalize the update with a Metropolis accept-reject step. In this basis, the Jacobian is just a constant and thus drops out of the reweighting, see equation (4.9). Results from this procedure are shown in figure 4.6. They agree with the analytical result.

Note that the scalar model is much simpler than U(1) theory in 1+1d. It is just a free Gaussian theory, which is usually highly symmetric and easier to solve. However, our original algorithm turned out to be unstable in both theories. In the

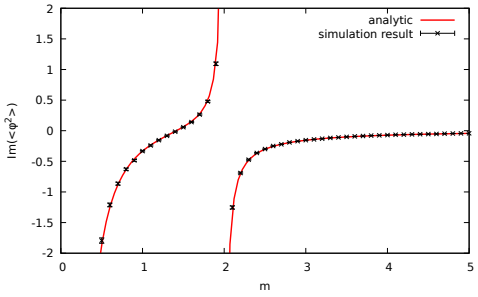

Figure 4.6: Results from simulating in the space spanned by the Takagi vectors, they agree with the analytical results.

scalar case, we were able to recover correct results by choosing the correct basis. Hence, finding a good basis for U(1) on the lattice may enable local updating in the future.

4.5 Summary and conclusions on Lefschetz thimbles

In this chapter, we have investigated so-called Lefschetz thimbles which are integration paths in the complexified space of a given theory. Integrands which are oscillatory in the original manifold are Gaussian-like along the thimbles and thus enable normal Monte Carlo simulations. We have examined some simple integrals and found a way to compute the ratio of the partition functions on different thimbles, which trivially generalizes to higher dimensions. This is highly desirable since so far there is no way to find the relative weights for contributing thimbles.
Finally, we have investigated higher dimensional theories and found a way to perform local updates which sample the global thimble. While this algorithm does sample the thimble in a simple scalar theory, it does not yield correct expectation values due to numerical deviations. The same algorithm is not stable in a complexified U(1) gauge theory. In the scalar theory, we were able to find a better basis for the simulation manifold, which yields correct results.

We conclude that except for simple models, a thorough understanding of the underlying manifolds is required to properly sample Lefschetz thimbles.

5 Finding order parameters with Neural Networks

In the last decade, there has been a huge increase in the use of machine learning. This is mainly due to the increase in computing power, which enabled the now popular deep learning using neural networks.
The first instance where deep learning decidedly outperformed other methods was AlexNet [206], a neural network which was entered into ImageNet Large Scale Visual Recognition Challenge [207], where the machine learning algorithm is tasked to correctly classify pictures into hundreds of categories. By now most machine learning applications use neural networks since they are the best machine learning technique for hierarchical and nonlinear learning tasks. They are being used in image classification [206, 208], Language processing [209], beating humans in complex games [210] and medical diagnosis [211, 212] to give just a few examples.
In high energy physics standard neural networks are used e.g. to find signals in detector data [213–215]. By now there are many other applications of machine learning in physics including quantum many-body systems [216–218], simulating statistical physics systems [219, 220], solving inverse problems relevant to physics [221–223], detecting and investigating phase transitions [31, 224–237], improving Monte Carlo simulations [238–243] and more. There have even been attempts to solve the sign problem using neural networks [244]. So far applications in lattice QCD are [31, 245, 246] For a review on applications in physics see [26].
Despite their wide usage, there is no generic way to find out what neural networks learn. In theoretical physics, however, it is crucial to understand the underlying mechanisms of the tools at hand in order to gain a deeper understanding of physics. In this chapter, we will set neural networks on physical grounds by explicitly showing that they learn order parameters when they are trained on phase transitions. We thus open the black box for a specific case.

We give a brief introduction to feed-forward neural networks in 5.1 show how to find order parameters with neural networks in section 5.2.

5.1 Brief introduction to feed-forward neural networks

By now there is a huge variety of many different neural network architectures. Hence, for a complete understanding of neural networks, we refer to books such as [247, 248]. Machine learning tasks are often classified via the learning method.

- Supervised learning relies on labeled data sets. A data point corresponds to a certain class and the algorithm is tasked to correctly classify the data set.
- Unsupervised learning is used for unlabeled data sets. Here the algorithm is tasked to find structures in a large data set without further input.
- Reinforcement learning is used to let an agent find a strategy to solve a task. Typically the agent does not get any information a priori but is punished or rewarded based on its actions.

Here we will focus on supervised learning and introduce a simple neural network architecture known as a feed-forward network.
The use of neural networks is motivated by the architecture of the brain, where many rather simple neurons work together to perform complex tasks. As the name suggests, neural networks are formed by a network of neurons. A neuron is a function $f(x)$ which takes an input vector x and maps it to a an output z

$$f(\vec{x}) = a(\vec{w} \cdot \vec{x} + b)\,. \tag{5.1}$$

The weights $\vec{w}$ represent the relative importance of each input to the neuron, the bias b introduces a minimum threshold the input has to reach and the activation function $a(x)$ produces the scalar output of the neuron depending on the input. Popular activation functions include the sigmoid function

$$a(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{5.2}$$

and the rectified linear unit (ReLU)

$$a(x) = \max(0, x)\,. \tag{5.3}$$

Typically neural networks perform well with a whole range of activation functions. A single neuron by itself is not very powerful. Thus they are typically grouped into layers. A layer $F(\vec{x})$ consists of many neurons, which are not interconnected in a feed-forward network. Those layers are then connected, i.e. the output of the neural network is

$$y = F_M(F_{M-1}(\dots(F_1(F_0(\vec{x}))))\,, \tag{5.4}$$

where F_0 is the input layer and F_M is the output layer. In a binary classification task, for example, the input could be a picture – and the input layer would have as many neurons as there are pixels – and the output layer would be a single neuron outputting 0 or 1. This setup typically has several hundred to several millions of parameters, i.e. the weights and biases. Given a data point as input, it is fed through the network and produces an output. If the weights and biases are chosen correctly, the data point will be classified correctly. Since choosing millions of parameters by hand is a near impossible task, this is done in an automatic way as described below. It has been shown that such a simple network is able to approximate any continuous function arbitrarily close [249, 250]. Thus a correctly trained neural network is a powerful tool for regression and classification tasks.

Training the feed-forward network

In the case of supervised learning, we can use a so called loss-function to determine how far the output of the network was from the correct classification, we will use the binary cross-entropy

$$C(Y,P) = -\frac{1}{N}\sum_i (y_i \log p_i + (1-y_i)\log(1-p_i))\,, \tag{5.5}$$

where $y_i \in Y$ are the correct labels of the data point, and $p_i \in P$ are the predictions of the network and the sum now goes over N data points. Note that the label here can be only 0 or 1 such that $\log(p_i)$ (or $\log(1-p_i)$) is added to the sum if $y_i = 1$ (or $y_i = 0$). The logarithm is maximal if $p_i = 1$, since $p_i \in [0,1]$ is a probability. A similar argument holds for the second term. The global minus sign then turns maxima into minima such that the binary cross-entropy needs to be minimized.
This minimization is done by some variant of gradient descent, which is a simple algorithm to find the minimum of a function. It does so by making small steps in the direction of steepest descent. Thus it requires a step size, which in the context of neural networks is referred to as the learning rate η. Each of those steps is then used to adjust the parameters of the neural network using backpropagation. In gradient descent, we take the derivative of the loss function with respect to the parameters of the network. Since the network has a hierarchical structure, the derivative in earlier layers depends on the parameters in later layers by the chain rule. The derivative for the output layer can be computed directly. The derivative for the second to last layer then depends on the derivative of the output layer via the chain rule, and so on. Thus, the change in parameters is propagated back through the network. A full description of the algorithm can be found in any standard textbook on neural networks, see e.g. [247, 248]. State-of-the-art algorithms use adaptive learning rates, compute the gradient just in a random subset of all directions (stochastic gradient descent) in order to decrease the chance of ending up in bad local minima, and have ways to not get stuck in shallow minima, e.g. momentum terms.
Training is then done by feeding the training set through the network multiple times, each of which is called a training epoch. Usually, in order to avoid over-fitting, i.e. to prevent the neural network from just memorizing the training data, regularization is used. Popular regularization schemes include dropout, where only part of the network is updated in each learning epoch or to add regularization terms to the loss function. Widely used are L^1 and L^2 regularization, where the L^n norm of the weights is added to the loss function in order to keep their values low. We now introduced everything needed to train a simple neural network. We will not need much more in the application to lattice configurations and thus will introduce additional techniques when they are needed.

5.2 Machine Learning of Explicit Order Parameters

Binary classification tasks as introduced above can give very precise predictions. Hence, they are a wonderful tool to look for phase transitions and are able to find the position of the transition to high accuracy see [224]. A neural network in a binary classification task can be viewed as an approximation of the so-called decision function D, which assigns a probability $p_i \in [0,1]$ to each sample S to be in class one. The decision boundary is a hyperplane in the hyperspace of configurations with $D(S) = 0.5$, i.e. the boundary where the neural network is unsure about its prediction. Thus, if the neural network learns a single function $Q(S)$ which distinguished between phases, we expect that $\nabla Q || \nabla D$ near the decision boundary. In our networks we will use the sigmoid activation function – see equation (5.2) – to map the latent prediction $\xi(S)$ (the output of the last ReLU layer) to a probability[15]. Hence, we can write $D(S) = \sigma(\xi(S))$ and $\nabla Q || \nabla \xi$ which allows us to write down an approximate linear relation

$$\xi(S) = wQ(S) + b\,, \tag{5.6}$$

in the vicinity of the decision boundary. In the following sections we will explicitly determine ξ and hence D as functions of the lattice configurations S and thus ultimately find the quantity Q, which turns out to be the order parameter or energy of the theory.

5.2.1 The Correlation Probing Neural Network

Neural networks encode their decision function in highly elusive non-linear ways. Since we are interested in applications to Monte Carlo samples, we would like to find out if the leading contributions to the decision function can be expressed by combining just a few lattice points. Typically many interesting observables in lattice theories can be expressed by local combinations of variables. Motivated by this observation, we construct a neural network which only gets to see a subset of the full lattice. If we successively lower the number of variables the network gets to see we can find the minimum amount of variable the neural network requires to still correctly predict the phase transition. This is the idea behind the correlation probing neural network which is depicted in figure 5.1. The lattice configuration is fed into the localization network, which simply is a fully connected convolutional neural network. A convolutional neural network consists of receptive fields which can only see a part of the configuration. This is typically used in image recognition, where a receptive field is only allowed to see parts of a picture, which is useful to extract information in a hierarchical way. Many convolutional layers are used to successively

[15] We will use the ReLU activation function everywhere else, hence the output of the last ReLU layer $\xi(S)$, which in our case just consists of one neuron, is generally not in $[0,1]$ and thus cannot be interpreted as a probability. The only reason for the use of $\sigma(x)$ is this mapping to $[0,1]$.

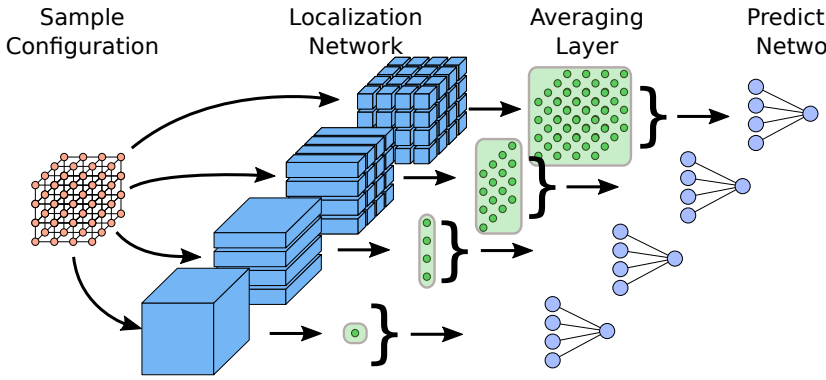

Figure 5.1: Schematic depiction of the correlation probing neural network. It consists of three different neural networks stacked on top of each other. The localization network is a fully connected convolutional neural network which allows connections only in within the receptive fields. The averaging layer averages over all receptive fields. Finally, the prediction network transforms the output of the averaging layer into a prediction probability.

combine the receptive fields. In our case, We do not combine the receptive fields and keep them separate all the time. The output of the convolutional network is fed into an averaging layer, which averages over all receptive fields. Finally, a fully connected prediction network is used to map the output of the averaging layer to a prediction.
We train the network with different receptive field sizes to find the minimum required size such that the phase transition is still found. In the following, we will describe the application of the correlation probing neural network to the Ising model and SU(2) lattice gauge theory. Technical details on the network setup and training can be found in appendix C.1.

Application to the Ising model

We first do this for the Ising model in two dimensions. The Ising model is given by the Hamiltonian

$$H(S) = -\sum_{\langle i,j\rangle_{\mathrm{nn}}} s_i s_j + h \sum_i s_i \,. \tag{5.7}$$

The spin $s_i \in \{-1, 1\}$ variables are aligned on a regular lattice. The first sum runs over nearest neighbors of spins. For simplicity we set $h = 0$ in the following. One spin configuration is given by $S = (s_1, \ldots, s_n)$. It is a simple model for ferromagnetism, its partition function is

$$Z = \sum_i e^{-\beta H(S_i)} \,, \tag{5.8}$$

Receptive Field Size	Train Loss	Validation Loss
28×28	6.1588e-04	0.0232
1×2	**1.2559e-04**	**1.2105e-07**
1×1	**0.2015**	**0.1886**
baseline	0.6931	0.6931

Table 5.1: Ising model: Losses of neural networks with different receptive field sizes in the localization network. Smaller losses mean better performance. The baseline classifier corresponds to a prediction of $p = 0.5$ for each phase, it is the worst possible loss.

where the sum runs over all possible spin configurations and $\beta = 1/T$ is the inverse temperature. The energy is given by

$$\langle E \rangle = \frac{\partial \log Z}{\partial \beta}\,, \tag{5.9}$$

and its derivative

$$C_V = \frac{\partial \langle E \rangle}{\partial \beta} = \frac{\partial^2 \log Z}{\partial \beta^2}\,, \tag{5.10}$$

diverges at the critical temperature, signaling a second order phase transition by the Ehrenfest classification. Thus, the energy can be used to distinguish phases. By Landau-Ginzburg theory phases are classified via an order parameter, which for the Ising model is the magnetization

$$\langle M \rangle = -\frac{1}{\beta}\frac{\partial(\log Z)}{\partial h}|_{h=0}\,, \tag{5.11}$$

its derivative with respect to T also diverges at T_c. The critical temperature is known analytically [251] $T_c = 2/\log(1+\sqrt{2})$).
We apply the correlation probing neural network to a data set of 55000 spin configurations of size 28×28 equally distributed over eleven equidistant temperatures in the range $T \in [0, 5]$. The neural network is trained on all configurations below $T = 1.6$ and above $T = 2.9$. It is tasked to predict the phase of a given configuration. We train it by using the binary cross entropy loss-function, see equation (5.5). After the network has been properly trained it is asked to predict the phase over the full range of T. Results are shown in table 5.1. The network correctly classifies the configurations when the receptive field size is equal to the lattice size. Even when the receptive field size is lowered to 1×2 it does not lose accuracy. For a receptive field size of 1×1 the prediction becomes less accurate. See figure 5.2f for a visualization of different receptive field sizes in the Ising model. Figure 5.2e shows the average classification probability for the 1×1 and 1×2 net. We find the phase transition temperature to be $T_c = 2.5(5)$ for the 1×1 network and $T_c = 2.25(25)$

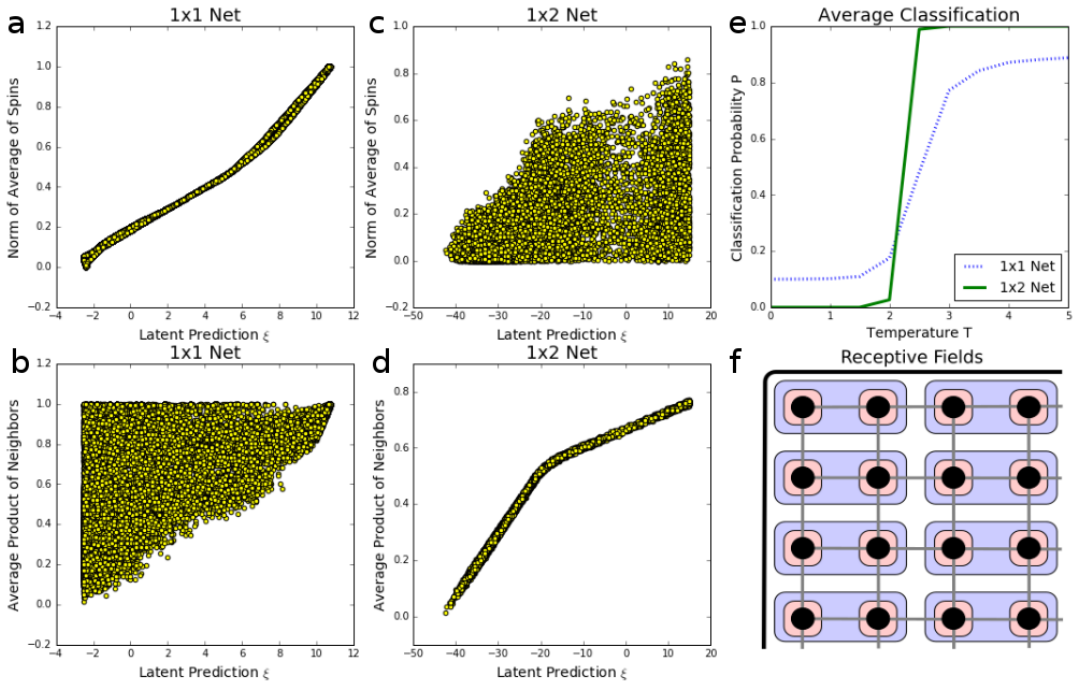

Figure 5.2: Results of the correlation probing neural network for the Ising Model. The latent prediction ξ is the argument of the sigmoid σ function, which we use as the activation function for the last layer of the prediction network. **a**: Correlation of the latent prediction with the absolute value of the average spins, i.e. the magnetization in the 1×1 network. **b**: Same as (a), but with the average product of neighboring spins, i.e. the energy instead of the magnetization. **c**: Correlation of the latent prediction with the magnetization for the 1×2 network. **d**: Same as (c), but with the energy instead of the magnetization. **e**: Average classification probability as a function of temperature. **f**: Visualization of receptive fields 1×1, 1×2 and 28×28 of the localization network output neurons.

for the 1×2 network. Both agree with the correct value within error bars which suffices for our task, see [224] for more accurate results for T_c from neural networks. In our task to find an analytical expression for the decision function, we first turn to the 1×1 net. By construction its decision function can be expressed as

$$D(S) = F\left(\frac{1}{N}\sum_i f(s_i)\right) = \sigma\left(\xi\left(\frac{1}{N}\sum_i f(s_i)\right)\right), \tag{5.12}$$

where f is the output of the localization network and F is the output of the prediction network. Due to the small receptive field size, we can expand the function f in terms of a single spin variable

$$f(s_i) = f_0 + f_1\ s_i + f_2\ \underbrace{s_i^2}_{1} + f_3\ \underbrace{s_i^3}_{s_i} + \ldots, \tag{5.13}$$

where higher powers of spins can be absorbed into the lowest two orders due to $s_i^2 = 1$. The constants f_0 and f_1 can be absorbed into weights and biases of the prediction network approximating F. This significantly simplifies the decision function to

$$D(S) = F\left(\frac{1}{N}\sum_i s_i\right). \tag{5.14}$$

Now we need to find an expression for F, which we do by comparing the latent prediction ξ with the argument of F. Figure 5.2a suggests a linear relation

$$\xi(S) \approx w\left|1/N\sum_i s_i\right| + b\,. \tag{5.15}$$

Thus employing equation (5.12), the decision function becomes

$$D(S) \approx \sigma\left(w\ \left|\frac{1}{N}\sum_i s_i\right| + b\right), \tag{5.16}$$

where w and b are weight and bias of the prediction neuron. Since we used no physical information except Monte Carlo configurations, we can conclude that the network learns the quantity

$$Q(S) = \left|\frac{1}{N}\sum_i s_i\right|, \tag{5.17}$$

the magnetization.
We can give similar arguments for the 1×2 network. Here, by construction

$$D(S) = F\left(\frac{1}{N}\sum_{<i,j>_T} f(s_i, s_j)\right), \tag{5.18}$$

where the sum goes only over transversal neighbors. The expansion of f

$$f(s_i, s_j) = f_{0,0} + f_{1,0}\, s_i + f_{0,1}\, s_j + f_{2,0}\, s_i^2 + f_{1,1}\, s_i\, s_j + f_{0,2}\, s_j^2 + \dots , \tag{5.19}$$

contains only three notable terms, again since $s_i^2 = 1$. Terms linear in one spin s_i represent the magnetization, which – according to the drop in performance seen in table 5.1 – contain less information than what we are looking for. Hence, the leading term must be $f_{1,1} s_i s_j$, reducing the decision function to

$$D(S) \approx F\left(\frac{1}{N} \sum_{<i,j>_T} s_i s_j\right) . \tag{5.20}$$

Figure 5.2d shows a perfect correlation between the latent prediction ξ and the sum over transversal nearest neighbors $1/N \sum_{<i,j>_T} s_i s_j$. Here, the final expression for the decision function is

$$D(S) \approx \sigma\left(w\left(\frac{1}{N} \sum_{<i,j>_T} s_i s_j\right) + b\right) . \tag{5.21}$$

Note that the sum over transversal nearest neighbors was sufficient. It can be generalized to all nearest neighbors by rotational and translational symmetry. Thus, the neural network learns

$$Q(S) = \frac{1}{N} \sum_{<i,j>_{\mathrm{nn}}} s_i s_j , \tag{5.22}$$

which corresponds to the energy per site.
We find that the energy works better in our context, see table 5.1. Figure 5.2 also shows the correlation of the latent predictions of both the 1×1 and 1×2 network with the energy and magnetization respectively. No correlation is found suggesting that the network really learns the order parameters that are dominant for the corresponding field size.

Application to SU(2) lattice gauge theory

Motivated by the success of the method in the Ising model, we apply it to SU(2) gauge theory on the lattice. We already introduced lattice gauge theory for SU(N) in section 2.3, however here we will change notation slightly to be more clear. Given a four dimensional lattice with $N_\tau \times N_s^3$ points, where in the following we choose $N_s = 8$ and $N_\tau = 2$, the gauge fields are parametrized via the link variables, which connect the lattice sites. For the site $x = (\tau, x, y, z)$ the SU(2) link pointing in direction $\mu \in \{t, x, y, z\}$ reads

$$U_\mu^x = a_\mu^x \mathbb{1} + i\left(b_\mu^x \sigma_1 + c_\mu^x \sigma_2 + d_\mu^x \sigma_3\right) , \tag{5.23}$$

where the σ_i are the Pauli matrices and the coefficients obey $(a_\mu^x)^2 + (b_\mu^x)^2 + (c_\mu^x)^2 + (d_\mu^x)^2 = 1$. Since the Pauli matrices are traceless. $\mathrm{Tr}U_\mu^x = 2a_\mu^x$. We write one lattice configuration as the collection of all link variables

$$S = \left(\{U_\mu^x |\, x \in N_\tau \times N_s \times N_s \times N_s,\, \mu \in \{\tau, x, y, z\}\}\right). \tag{5.24}$$

Since we hope that our neural network finds the order parameter, the Polyakov loop, we write it in the current notation as

$$\begin{aligned} L(\vec{x}) &= \mathrm{tr}\left(\prod_{x_0=0}^{N_\tau-1} U_\tau^x\right) \overset{N_\tau=2}{=} \mathrm{tr}\left(U_\tau^{0,\vec{x}} U_\tau^{1,\vec{x}}\right) \\ &= 2\left(a_\tau^{0,\vec{x}} a_\tau^{1,\vec{x}} - b_\tau^{0,\vec{x}} b_\tau^{1,\vec{x}} - c_\tau^{0,\vec{x}} c_\tau^{1,\vec{x}} - d_\tau^{0,\vec{x}} d_\tau^{1,\vec{x}}\right). \end{aligned} \tag{5.25}$$

For the following analysis we produced a total of 15600 decorrelated configurations using the Wilson gauge action (2.23). The configurations are equally distributed over 26 values of $\beta = 4/g^2 \in [1, 3.5]$. It is important that the configurations are decorrelated, see e.g. [246] for a case where correlations between the samples contaminated the learning task.

Since our aim is to have a completely blind method, we do not assume any knowledge except the gauge configurations themselves. I.e. we do not assume any knowledge about where the phase transition is. Thus, as a first step, we employ an unsupervised learning algorithm to find a rough approximation for the transition. Note that due to the gauge freedom in lattice QCD, networks like autoencoders [230] are not a good match for this task, since the reconstruction of the configuration from the latent variables is not unique if the latent variables represent a gauge invariant quantity such as the Polyakov loop. Instead, we use principal component analysis, which is an orthogonal linear transformation to a set of variables which are ordered by their variance. The first few principal components contain the most information. This algorithm is perfect in our case, since in the ordered phase there should not be many fluctuations and hence much lower variance than in the disordered phase if a phase transition exists. Note, however, that PCA is a linear transformation and as such can only find linear approximations. In the Ising model PCA was able to correctly predict the transition temperature [229]. Figure 5.3 shows that with two principal components there is a signal for a phase transition around $\beta \approx 2.0$. However, no clear correlation between the Polyakov loop and the principal components is found, which was expected, since the Polyakov loop is a nonlinear function.

PCA suggest a phase transition at $\beta \approx 2$. Hence, we train the correlation probing neural network with different receptive field sizes to predict the phase transition. To that end we train the neural network at $\beta \in [1, 1.2]$ and $\beta \in [3.3, 3.5]$. It is able to predict the phase transition with the receptive field size incorporating the whole lattice, i.e. $2 \times 8 \times 8 \times 8$ and finds a critical value of $\beta_c = 1.97(1)$. For a $2 \times 1 \times 1 \times 1$ network we find $\beta_c = 1.99(1)$, this is depicted in figure 5.4c. A direct lattice computation of the critical temperature yields $\beta_c = 1.880(25)$. The small deviation is

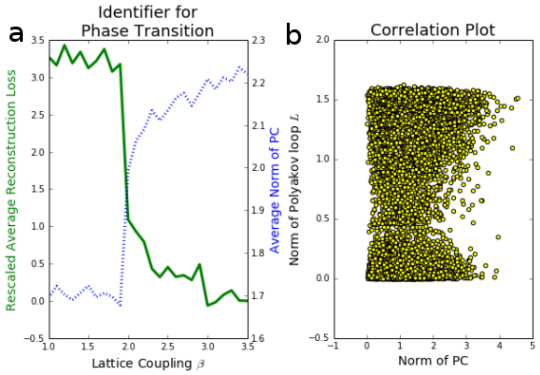

Figure 5.3: **a**: The average mean squared error reconstruction loss of PCA as a function of temperature signals a phase transition. It was calculated in 100 independent PCA runs with two principal components (PC), measured in units of $\times 10^{-5}$ and shifted by the value at $\beta = 3.5$. The average norm of the PC also indicates a phase transition. **b**: Correlation between the Polyakov loop and the principal components.

most likely due to a systematic error. When producing the gauge configurations, we always initialized at the same point, all links were set to one. In that case, it is hard for the simulation to cross over to having all links equal to minus one. This biases the network, hence the difference. We are however not interested in getting an exact value for the critical coupling and are satisfied with the given precision.
Since we want to find the order parameter, we look at different receptive field sizes in table 5.2. The minimal receptive field size is $2 \times 1 \times 1 \times 1$. Allowing only $1 \times 1 \times 1 \times 1$ is as bad as the network just guessing, i.e. the baseline classifier. Thus, the decision

Receptive Field Size	Train Loss	Validation Loss
$2 \times 8 \times 8 \times 8$	1.0004e-04	2.6266e-04
$2 \times 8 \times 8 \times 1$	8.8104e-08	1.3486e-07
$2 \times 8 \times 1 \times 1$	7.7671e-05	2.0394e-04
$2 \times 1 \times 1 \times 1$	**8.8104e-08**	**6.8276e-08**
$2 \times 1 \times 1 \times 1$ no hidden layers in prediction net	**2.2292e-07**	**4.2958e-07**
$1 \times 1 \times 1 \times 1$	0.6620	0.9482
baseline	0.6931	0.6931

Table 5.2: SU(2): Losses of neural networks with different receptive fields of the neurons in the localization network.

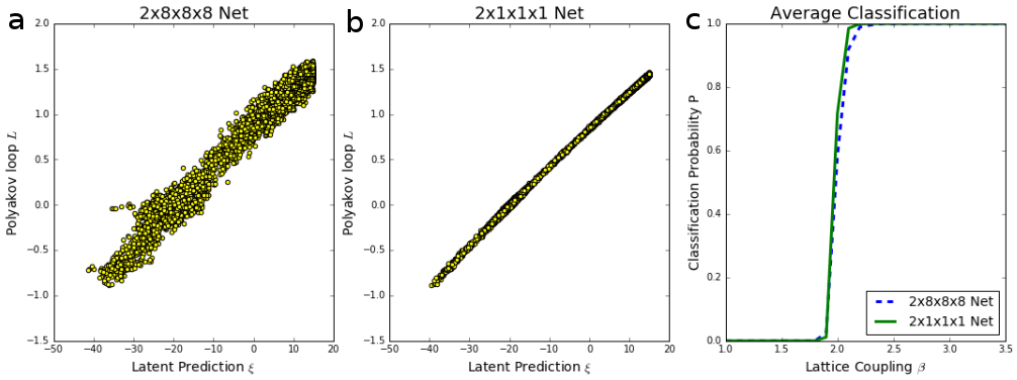

Figure 5.4: Results from the correlation probing neural network for SU(2). **a,b**: Correlation of the latent prediction with the Polyakov loop in both the $2 \times 8 \times 8 \times 8$ network and the $2 \times 1 \times 1 \times 1$ network. **c**: The average prediction probability of the two networks.

function must be constructed from two lattice points with four links each

$$D(S) = F\left(\frac{2}{N}\sum_{\vec{x}} f(U_\tau^{0,\vec{x}}, U_x^{0,\vec{x}}, U_y^{0,\vec{x}}, U_z^{0,\vec{x}}, U_\tau^{1,\vec{x}}, U_x^{1,\vec{x}}, U_y^{1,\vec{x}}, U_z^{1,\vec{x}})\right) . \tag{5.26}$$

Now f is a function of 32 variables, 2 lattice points times 4 links times 4 real variables per link and F is a function of a single variable as before.
We see in table 5.2, that it is enough for the prediction network to contain a single sigmoid neuron and no hidden layers, thus $D(S) = \sigma(wQ(S) + b)$. We also split the lattice configurations in patches of size $2 \times 1 \times 1 \times 1$. I.e. we view each of those patches as a single configuration, effectively enhancing our statistics by a factor of 512. Thus,

$$Q(S) = \frac{2}{N}\sum_{\vec{x}} f(U_\tau^{0,\vec{x}}, U_x^{0,\vec{x}}, U_y^{0,\vec{x}}, U_z^{0,\vec{x}}, U_\tau^{1,\vec{x}}, U_x^{1,\vec{x}}, U_y^{1,\vec{x}}, U_z^{1,\vec{x}}) . \tag{5.27}$$

In order to find an expression for f, we train a network only on the local patches. We then perform a polynomial regression on the latent prediction of the localization network on 1% of the data set, using another 1% for validation. To be more precise, the fit-function is given by all possible combinations of the 32 variables per patch up to the order of the polynomial. We find that a second-order polynomial fit suffices, it is a fit with 561 different terms. The results are shown in table 5.3. We find the

Order of Regression	Train Score	Validation Score
Polynomial Regression		
1	0.00128	-0.00042
2	**0.72025**	**0.72395**
3	0.75675	0.69129

Table 5.3: Scores of different regression algorithms. Higher is better.

coefficients with the larges absolute value from the fit

$$\begin{aligned} f(\{U_\mu^{x_0}\}) =& f(U_\tau^0, U_x^0, U_y^0, U_z^0, U_\tau^1, U_x^1, U_y^1, U_z^1) \\ =& f((a_\tau^0, b_\tau^0, ..., d_z^1)) \\ \approx& + 7.3816\ a_\tau^0 a_\tau^1 + 0.2529\ a_\tau^1 b_\tau^1 \\ & + \ ... \\ & - 0.2869\ d_\tau^0 c_\tau^1 - 7.2279\ b_\tau^0 b_\tau^1 \\ & - 7.3005\ c_\tau^0 c_\tau^1 - 7.4642\ d_\tau^0 d_\tau^1 \,, \end{aligned} \tag{5.28}$$

where the differences between the leading contributions are of the same order as the size of the sub-leading coefficient, and as such most likely originate from approximation errors in the network. Overall factors and constants can be absorbed into weights and biases and we are allowed to rescale the fit result. We conclude that

$$\begin{aligned} f((a_\tau^0, b_\tau^0, ..., d_z^1)) \approx& a_\tau^0 a_\tau^1 - b_\tau^0 b_\tau^1 - c_\tau^0 c_\tau^1 - d_\tau^0 d_\tau^1 \\ =& \left(U_\tau^0 U_\tau^1\right). \end{aligned} \tag{5.29}$$

Which is just the Polyakov loop (5.25) on the $2 \times 1 \times 1 \times 1$ patch as suspected. Due translation symmetry, we are allowed to generalize this to the whole lattice $f(\{U_\mu^{x_0}\}) \to f(\{U_\mu^{x_0,\vec{x}}\})$. The decision function of the full network thus becomes

$$D(S) \approx \sigma\left(w\left(\frac{2}{N}\sum_{\vec{x}} f(\{U_\mu^{x_0,\vec{x}}\})\right) + b\right) \ . \tag{5.30}$$

with the argument of σ begin $Q(S) = \frac{2}{N}\sum_{\vec{x}} f(\{U_\mu^{x_0,\vec{x}}\})$, i.e. the Polyakov loop on the full lattice. This is confirmed by the perfect correlation between the Polyakov loop and the latent prediction in figure 5.4a,b.

5.3 Conclusion on neural networks

We have applied neural networks to SU(2) gauge theory and the Ising model. By showing that the neural network explicitly learns the order parameter, we decoded the black box of neural networks, at least in this specific case. By doing so we also found a method which allows finding order parameters of systems of many variables.

This is in principle applicable anywhere but limited by computational power. In our case, all order parameters were local to some degree. While the Polyakov loop is a global observable, we chose the lattice size such that it was effectively local. We tried the same analysis on a 4×8^3 lattice and found that much higher statistics would be required to get equally good results. Hence, if the order parameter is completely global, as is the case e.g. for the chiral condensate, the amount of data required for our method grows quickly. Note that in principle this is not a problem. Early lattice simulations from the 70s and 80s have shown that qualitative features can be captured very well on even small lattices. Thus, one should try to find the minimal lattice size which shows the transition and perform the analysis on this lattice size. In the case of SU(2) gauge theory and the Ising model, the generalization of the order parameter to larger lattices is trivial. There would be no gain in performing the analysis on larger lattices.
In principle, this method can be used to find any somewhat local structure from lattice configurations, not only order parameters. Given any quantity we want to predict, we can use the correlation probing neural network and find out which quantity it learns.
Neural networks could also be useful for solving sign problems. A simple case has been investigated in [244]. One has to be careful however not to rely on extrapolations over phase transitions or similarly singular phenomena, since the structure may change significantly and if the network is not properly trained, the results will be unphysical since the network never would have seen such structures during training.

6 Summary and outlook

We have investigated the phase diagram of QCD by means of the complex Langevin method, investigated various aspects of the Lefschetz thimble method and finally showed that neural networks learn the order parameter when trained on lattice configurations.

Complex Langevin

Starting with the complex Langevin method, we gave a systematic way to check for its convergence, namely by looking for effects of poles and boundary terms. We were thus able to investigate several aspects of QCD with complex Langevin. We first investigated QCD using a hopping expansion and found that it does not converge properly in some interesting regions. In full QCD we found first signals of the nuclear liquid-gas transition from lattice QCD, though thermalization issues prohibited us from a thorough study of the transition. Finally, we focused on the QCD transition between hadronic matter and the quark-gluon plasma by simulating in the transition region up to $\mu/T_c(\mu = 0) \approx 5$, which is a much larger region than other methods can currently investigate. In our simulations transition temperature is a quadratic function of chemical potential μ in accordance with other simulations. Our simulations were still plagued by large finite size effects, which most prominently shows in our moderately heavy pion mass, which in our largest volume had a minimum possible value of $m_\pi \approx 1.55$ GeV. We have found that a value of $\kappa_2 \approx 0.0011 - 0.0015$ for the curvature of the transition line, see table 3.3. This value interpolates between a curvature of $\kappa_2 \approx 0.014$ from [19, 49] and QCD with heavy static quarks where there is basically no curvature [83]. All in all, we found that complex Langevin is applicable to a large portion of the QCD phase diagram if all convergence criteria are carefully monitored.
Finally, we investigated SU(2) gauge theory on a real-time contour and found that with the inclusion of recent improvements to complex Langevin, i.e. gauge cooling and dynamical stabilization, it is now applicable in a physically interesting region.
In terms of the QCD phase diagram, future research is straightforward. Simulations at smaller quark masses and larger volumes, as well as continuum extrapolated results, are highly desirable, especially to locate the critical endpoint. Here one has to be particularly careful with poles since low lying eigenvalues in the Dirac operator tend to occur for simulations at small masses.
In SU(2) gauge theory, the next step should be to compute any physical quantity on the contour, e.g. the shear viscosity, to see whether analytical continuation methods really perform better, when starting from a real-time contour instead of in Euclidean space.

In terms of the complex Langevin method, it is highly desirable to find a simpler way to look for poles, e.g. by computing boundary terms close to the poles.

Lefschetz Thimbles

We have thoroughly investigated the Lefschetz thimble method and discussed its issues in the case of simple models. There we have found a way to compute the relative weights of the separate thimbles, which is a crucial ingredient. This method is also applicable in higher dimensional theories. We formulated an algorithm to perform local updates in U(1) and scalar field theory. While being stable in simple models, the algorithm showed issues in higher dimensions, which we were able to trace back to a bad choice of basis. In a simplified scalar field theory, this insight allowed us to simulate the theory using Takagi vectors as a basis.
Current algorithms on Lefschetz thimbles become prohibitively expensive in higher dimensions. A large portion of current research seems to be focused on finding alternative manifolds, however here most algorithms are also expensive and only ameliorate the sign problem. Hence, there should be a focus to find better ways to simulate on Lefschetz thimbles or similar manifolds, e.g. by making local updates possible. Such algorithms are an important step towards solving the sign problem in QCD since there an application of Lefschetz thimbles or related methods is currently unfeasible.

Neural networks in Monte Carlo simulations

We applied a simple neural network architecture to lattice configurations from the Ising model and SU(2) gauge theory. This allowed us to explicitly show, that the neural network learns the order parameter of the theory. We thus also decoded the neural network black box in this particular application and simultaneously gave a way to find the order parameter from no prior knowledge except the Hamiltonian or action.
Decoding the neural network black box is an important step towards the integration of neural networks in the toolkit of lattice simulations. Theoretical predictions rely on a thorough understanding of the underlying theory, i.e. the neural network in this context. Applications of neural networks in the context of Monte Carlo simulation should be applied to help improve current algorithms, such as in the path optimization method [27], where they are used to find a better simulation manifold close to Lefschetz thimbles. Applications like the one in this work, in particular, prohibit the use of deep learning, since there is not much understanding of the learned quantities deep networks, due to their multilayered nature and complexity. Instead, the focus should be on simple neural networks for simple systems, where they can be understood.

A Complex Langevin simulation details

A.1 Detailed derivation of the convergence criterion and boundary terms for complex Langevin

Here we give a more detailed derivation of the criterion of correctness provided in chapter 3.1.1. The proof with full mathematical rigor can be found in [120, 121]. Before we investigate the interpolating quantity $F_{\mathcal{O}}(t,\tau)$ as defined in equation (3.10), we give some important definitions and equations, which follow from these definitions by Itô calculus, see e.g. [252]. The evolution equation for the distribution function $\rho(x)$ corresponding to the Langevin equation (3.4) is the Fokker-Planck equation

$$\frac{\partial}{\partial t}\rho(x;t) = \partial_x \left[\partial_x + S'(x)\right] \rho(x;t) = L_c^T \rho(x;t)\,. \tag{A.1}$$

We will refer to it as the complex Fokker-Planck equation in the case of a complex action. Similarly, when splitting the Langevin equation into real- and imaginary part (3.5) we get the real Fokker-Planck equation

$$\frac{\partial}{\partial t}P(x,y;t) = \left(\partial_x \left[\partial_x + \mathrm{Re}S'(x+iy)\right] + \partial_y \mathrm{Im}S'(x+iy)\right) P(x,y;t) = L^T P(x,y;t)\,. \tag{A.2}$$

The operators L_c^T and L^T are called the real and complex Fokker-Planck operators respectively. As is stated in the main text in equation (3.8) we want to show that both evolution agree, i.e.

$$\int dx \mathcal{O}(x)\rho(x;t) = \int dxdy \mathcal{O}(x+iy)P(x,y;t)\,, \tag{A.3}$$

provided

$$\begin{aligned} &P(x,y;0) = \rho(x;0)\delta(y) \geq 0 \\ &\int \rho(x;0)dx = 1\,. \end{aligned} \tag{A.4}$$

For the complex Fokker-Planck equation we know that if it has a spectrum in the left half plane with 0 a non-degenerate eigenvalue, see e.g. [131], that it has a unique equilibrium distribution $e^{-S(z)}$. Hence, if we can prove (A.3) we immediately know

that complex Langevin simulates the right physics in equilibrium.
With those prerequisites, we can start the proof for convergence from the interpolating quantity

$$F_{\mathcal{O}}(t,\tau) = \int P(x,y;t-\tau)\mathcal{O}(x+iy;\tau)dxdy\,. \tag{A.5}$$

This quantity interpolates between the real- and complex measures, which is trivial to see for $\tau = 0$

$$F_{\mathcal{O}}(t,0) = \int P(x,y;t)\mathcal{O}(x+iy)dxdy = \langle\mathcal{O}\rangle_{P(t)}\ , \tag{A.6}$$

For $\tau = t$ first note that an observable $\mathcal{O}$ follows the equation

$$\partial_t\mathcal{O}(z;t) = (\partial_z - S'(z))\,\partial_z\mathcal{O}(z;t) = L_c\mathcal{O}(z) \tag{A.7}$$

with formal solution

$$\mathcal{O}(z;t) = e^{tL}\mathcal{O}(z)\,, \tag{A.8}$$

where the initial condition is $\mathcal{O}(z;0) = \mathcal{O}(z)$. In a slight abuse of notation we used the same name for this operator as for the one in (A.2), since

$$(\partial_z S'(z))\,\partial_z O(z) = ([\partial_x + \mathrm{Re}S'(x+iy)]\,\partial_x + \mathrm{Im}S'(x+iy)\partial_y)\,\mathcal{O}(x+iy)\,, \tag{A.9}$$

if $\mathcal{O}$ is holomorphic. Using the evolution of an observable (A.8) as well as the initial condition (A.4), we obtain

$$\begin{aligned}
F_{\mathcal{O}}(t,t) &= \int P(x,y;0)\mathcal{O}(x+iy;t)dxdy \\
&= \int P(x,y;0)\left[e^{tL}\mathcal{O}\right](x+iy)dxdy \\
&= \int \rho(x;0)e^{tL_c}\mathcal{O}(x)dx \\
&= \int \mathcal{O}(x)e^{tL_c^T}\rho(x;0)dx \\
&= \langle\mathcal{O}\rangle_{\rho(t)}\ .
\end{aligned} \tag{A.10}$$

Here we also made the assumption that integration by parts in x is possible. This is correct if the action and observables are periodic in x, which is true for all theories we investigate in this thesis. Otherwise, locality of the integral is another requirement.
From this interpolating quantity, we can formulate a criterion for convergence, namely

$$\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = 0\,. \tag{A.11}$$

If this holds for all τ, then (A.3) is fulfilled and complex Langevin must converge to the correct probability distribution, provided all assumptions made in this derivation are correct. The derivative reads

$$\begin{aligned}&\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = \\ &-\int \left(L^T P(x,y;t-\tau)\right)\mathcal{O}(x+iy;\tau)dxdy \\ &+\int P(x,y;t-\tau)L_c\mathcal{O}(x+iy;\tau)dxdy = 0\end{aligned} \tag{A.12}$$

and can be rewritten further by writing out the Fokker-Planck operators. All terms containing only derivatives in x vanish via integration by parts since by assumption there are no boundary terms for x. This leaves

$$\begin{aligned}&\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = \\ &\int \mathcal{O}(x+iy;\tau)(\partial_x K_x + \partial_y K_y)P(x,y;t-\tau)dxdy \\ &-\int P(x,y;t-\tau)S'(x+iy)\partial_x\mathcal{O}(x+iy;\tau)dxdy\,.\end{aligned} \tag{A.13}$$

Since the observable is holomorphic for any τ [121], we can rewrite $-S'\partial_x = K_x\partial_x + K_y\partial_y$ and cancel the term with ∂_x. The remaining terms

$$\begin{aligned}&\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = \\ &\int \left(\partial_y K_y P(x,y;t-\tau)\right)\mathcal{O}(x+iy;\tau)dxdy \\ &+\int P(x,y;t-\tau)K_y\partial_y\mathcal{O}(x+iy;\tau)dxdy\,,\end{aligned} \tag{A.14}$$

can be interpreted as a total derivative in y, which by fundamental theorem of calculus can be integrated as

$$\frac{\partial}{\partial\tau}F_{\mathcal{O}}(t,\tau) = \lim_{Y\to\infty} B_{\mathcal{O}}(Y;t,\tau) \tag{A.15}$$

with

$$\begin{aligned}&B_{\mathcal{O}}(Y;t,\tau) \equiv \\ &\int [K_y(x,Y)P(x,Y;t-\tau)\mathcal{O}(x+iY;\tau) - K_y P(x,-Y;t-\tau)\mathcal{O}(x-iY;\tau)]dx\,,\end{aligned} \tag{A.16}$$

which is the boundary term that has to vanish for proper convergence of complex Langevin.

A.2 Details on the solution of the Fokker-Planck equation

The real Fokker-Planck equation for the model in equation 3.14 is

$$\begin{aligned}&\frac{\partial P(x,y;t)}{\partial t} = \left[\partial_x\left(\partial_x - K_x\right) - \partial_y K_y\right] P(x,y;t)\\ =&[\partial_x^2 + \beta(-2\sin x \sinh y + \cos x \sinh y \partial_x\\ &- \sin x \cosh y \partial_y)]P(x,y;t)\,.\end{aligned} \tag{A.17}$$

By discretizing (A.17) in x and y with symmetric derivatives we obtain

$$\begin{aligned}&P(x,y;t+dt) =\\ &\frac{1}{dx^2}\left(P(x_+,y;t) - 2P(x,y;t) + P(x_-,y;t)\right)\\ -&2\beta \sin x \sinh y P(x,y;t)\\ +&\frac{\beta}{2dx}\cos x \sinh y \left(P(x_+,y;t) - P(x_-,y;t)\right)\\ -&\frac{\beta}{2dy}\sin x \cosh y \left(P(x,y_+;t) - P(x,y_-;t)\right)\,,\end{aligned} \tag{A.18}$$

with $x_\pm = x \pm dx$ and $y_\pm = y \pm dy$.

With a regularization term in the y-drift $K_y \to K_y - s_y y$, additional terms occur

$$\begin{aligned}&P(x,y;t+dt) \to P(x,y;t+dt) + s_y P(x,y;t)\\ &+ s_y \frac{y}{2dy}\left(P(x,y_+;t) - P(x,y_-;t)\right)\,.\end{aligned} \tag{A.19}$$

We have solved the Fokker-Planck equation on an x-y-grid with parameters $dt = 10^{-6}$, $dx = 0.005 = dy$. A value of $Y = 5$ is sufficient for the cutoff in y-direction (see equation (3.18). In x-direction we cut off at $X = 3.14$, which is due to the 2π periodicity of the problem. Boundary conditions are periodic in both directions. Initial conditions are given by (A.4), with a smeared out δ-function to avoid numerical issues,

$$P(x,y;0) = \frac{1}{2\pi\sqrt{2\pi\sigma_y^2}} e^{-\frac{y^2}{2\sigma_y^2}}\,, \tag{A.20}$$

with $\sigma_y = 0.1$. In this discretization, it is difficult to resolve higher modes. This is easier in Fourier space, where the discretized evolution equation for P is given by

$$\begin{aligned}&P(k,y;t+dt) = -k^2 P(k,y;t)\\ &-\frac{i\beta}{2}\sinh(y)\left(k_- P(k_+,y;t) + k_+ P(k_-,y;t)\right)\\ &+\frac{i\beta}{4dy}\cosh(y)\left(P(k_+,y_+;t) - P(k_-,y_+;t)\right)\\ &+\frac{i\beta}{4dy}\cosh(y)\left(-P(k_+,y_-;t) + P(k_-,y_-;t)\right)\,,\end{aligned} \tag{A.21}$$

where $k_\pm = k \pm 1$ and similarly for y. Here we chose $dt = 0.5 \times 10^{-5}$, $k \in \{-19, \ldots, 20\}$, $dy = \sqrt{dt}$, $Y \approx 2.8$ with anti-periodic boundary conditions in k for the imaginary part of $P(k, y; t)$ and periodic boundary conditions for the real part in k as well as for y. After a time of $t \sim 30$ the result strongly depends on the choice of discretization. We thus use the k-y discretization to resolve the higher modes and the x-y-discretization for the rest.

A.3 The correct evolution

In order to compute the boundary term in equation (3.13) we require knowledge of $\mathcal{O}(x + iy; \tau)$ for $\tau > 0$ as described by equation (A.7). This can be done by going into Fourier space. For a general observable

$$\mathcal{O}(x) = \sum_k a_k \exp(ikx)\,, \tag{A.22}$$

the Fokker-Planck operator acts like

$$(L_c a)_k = -k^2 a_k - \frac{i\beta}{2}(k-1)a_{k-1} + \frac{i\beta}{2}(k+1)a_{k+1}\,, \tag{A.23}$$

which gives the matrix form of the operator L_c

$$(L_c)_{kl} = -k^2 \delta_{kl} - \frac{i\beta}{2}(k-1)\delta_{k-1,l} + \frac{i\beta}{2}(k+1)\delta_{k+1,l}\,. \tag{A.24}$$

One can compute the action of the exponential of the matrix as in equation (A.8) numerically by cutting the modes of at large $K = |k|$ which was chosen as $K = 50$. This solution gives correct expectation values for all modes as expected.

A.4 Polyakov chain

The action for the Polyakov chain with $\mathrm{SL}(3, \mathbb{C})$ links reads

$$-S = c_+ \mathrm{Tr} U_1 \ldots U_{N_t} + c_- \mathrm{Tr} U_{N_t}^{-1} \ldots U_1^{-1}\,, \tag{A.25}$$

where $c_\pm = \beta + \kappa e^{\pm\mu}$. Using the derivative

$$D_a f(U) = \frac{\partial}{\partial \alpha} f(e^{i\alpha} U)|_{\alpha=0}\,, \tag{A.26}$$

the Langevin at position n for generator a becomes

$$\begin{aligned} &K_{a,n} = -D_a S = \\ &ic_+ \mathrm{Tr}\left(U_1 \ldots U_{n-1} \lambda_a U_n \ldots U_{N_t}\right) - ic_- \mathrm{Tr}\left(U_{N_t}^{-1} \ldots U_n^{-1} \lambda_a U_{n-1}^{-1} \ldots U_1^{-1}\right)\,. \end{aligned} \tag{A.27}$$

The prescription for making updates in the algebra at position n is then

$$U \to \exp\left[\sum_{a=1}^{8} i\lambda_a \epsilon K_{a,n} + \sqrt{\epsilon}\eta\right] U\,, \tag{A.28}$$

where a first order discretization scheme was used for the Langevin equation, ϵ is the step size and η is the usual Gaussian noise.

After gauge fixing and diagonalization the action depends on two complex variables ω_1 and ω_2. The action including the Haar-measure reads

$$\begin{aligned} -S =& c_+\left(e^{i\omega_1} + e^{i\omega_2} + e^{-i(\omega_1+\omega_2)}\right) + c_-\left(e^{-i\omega_1} + e^{-i\omega_2} + e^{i(\omega_1+\omega_2)}\right) \\ &+ \ln\left[\sin^2\left(-\frac{2\omega_1+\omega_2}{2}\right)\sin^2\left(\frac{\omega_1-\omega_2}{2}\right)\sin^2\left(\frac{\omega_1+2\omega_2}{2}\right)\right]. \end{aligned} \tag{A.29}$$

The drift is

$$\begin{aligned} K_{\omega_1} =& -\partial_{\omega_1} S = ic_+(e^{i\omega_1} - e^{-i(\omega_1+\omega_2)}) + ic_-(-e^{-i\omega_1} + e^{i(\omega_1+\omega_2)}) \\ &+ \tan\left((\frac{\pi}{2} - \frac{(w1-w2)}{2}\right) + 2\tan\left(\frac{\pi}{2} - (\omega_1 + \frac{\omega_2}{2})\right) + \tan\left(\frac{\pi}{2} - (\frac{\omega_1}{2} + \omega_2)\right) \\ K_{\omega_2} =& -\partial_{\omega_2} S = ic_+(e^{i\omega_2} - e^{-i(\omega_1+\omega_2)}) + ic_-(-e^{-i\omega_2} + e^{i(\omega_1+\omega_2)}) \\ &- \tan\left((\frac{\pi}{2} - \frac{(w1-w2)}{2}\right) + \tan\left(\frac{\pi}{2} - (\omega_1 + \frac{\omega_2}{2})\right) + 2\tan\left(\frac{\pi}{2} - (\frac{\omega_1}{2} + \omega_2)\right), \end{aligned} \tag{A.30}$$

and the update prescription becomes

$$\omega_i \to \omega_i + \epsilon K_i + \sqrt{\epsilon}\eta\,, \tag{A.31}$$

with ϵ and η the step size and Gaussian noise.

A.5 Updating prescription for SU(N) and QCD

In the case of SU(N) and its complexifications, the update prescription for a single link following a Langevin process reads

$$U(n) \to e^{i\lambda_a\left(\epsilon K_{a,n,\nu} + \sqrt{\epsilon}\eta\right)U(n)}\,, \tag{A.32}$$

with the drift $K_{a,n,\nu} = -D_{a,n,\nu}S$ begin the derivative of the action with the derivative defined as in equation (A.26). For the Wilson gauge action (see equation (2.23)), the drift term reads

$$K^{\text{boson}}_{a,n,\nu} = i\frac{1}{2N_c}\sum_{\mu\neq\nu}\text{Tr}\left[\lambda_a U_\mu(n) C_{\mu\nu}(n) - \bar{C}_{\mu\nu}(n) U^{-1}_\mu(n)\right], \tag{A.33}$$

with

$$\begin{aligned} C_{\mu\nu}(n) =& U_\nu(n+\hat{\mu})U_\mu^{-1}(n+\hat{\nu})U_\nu^{-1}(n) \\ &+U_\nu^{-1}(n+\hat{\mu}-\hat{\nu})U_\mu^{-1}(n-\hat{\nu})U_\nu(n-\hat{\nu}) \\ \bar{C}_{\mu\nu}(n) =& U_\nu(n)U_\mu(n+\hat{\nu})U_\nu^{-1}(n+\hat{\mu}) \\ &+U_\nu^{-1}(n-\hat{\nu})U_\mu(n-\hat{\nu})U_\nu(n+\hat{\mu}-\hat{\nu})\,. \end{aligned} \tag{A.34}$$

The fermionic part of the action is given by

$$S_f = -\mathrm{Tr}\,\log M\,, \tag{A.35}$$

with the Dirac matrix M as defined in equation (2.31) and below. In the case of HDQCD and the hopping expansion the fermionic part can be simplified, see [83, 104, 108] for the drift terms. In full QCD the fermionic part of the drift term becomes

$$K_{a,n,\nu}^{\text{fermion}} = N_f \mathrm{Tr}\left(M^{-1} D_{a,n,\nu} M\right)\,. \tag{A.36}$$

An exact evaluation of the trace is numerically too expensive, since the inversion of a matrix of size (volume $\times\, d \times N_c \times 4$) is required[16]. Instead we use a stochastic estimator as was used in [84]. We write the drift as

$$\eta^\dagger M^{-1} D_{a,n,\nu} M\eta = \psi^\dagger D_{a,n,\nu} M\eta\,, \tag{A.37}$$

with $\psi^\dagger = \eta^\dagger M^{-1}$ is the solution of the linear equation $M^\dagger\psi = \eta$ and η is a noise vector with Gaussian noise with zero mean and variance one. If we have a way to compute this solution, we can just use one noise vector per update, since we average over long trajectories anyway. In the limit of Langevin step size $\epsilon \to 0$ there will be many noise vectors per time interval. The bottleneck of any lattice QCD simulation is the inversion of the Dirac matrix. We do so by using the standard conjugate gradient (CG) algorithm. It is particularly useful for sparse matrices, where it is too expensive to save the full matrix. Thus we do not need to save the full fermion matrix but only need to know how it multiplies a vector. The CG is defined for hermitian matrices, hence we invert $MM^\dagger$ to get $\tilde{\psi} = M^{-1}\left(M^\dagger\right)^{-1}\eta$. A simple multiplication by M then yields ψ.
In practice we use the bilinear noise scheme from [132, 253] for the fermionic drift force.

A.6 Lattice spacings and pion masses for QCD simulations

The lattice spacings and pion masses for our simulation parameters are given in table A.1. Everything was measured at $\mu = 0$. The lattice spacings were measured on $24^3 48$ lattices. The $|\Lambda|$ columns refers to the lattice sizes on which the masses were measured.

[16] The four comes from the Dirac gamma matrix per site.

κ	β	a/fm	$\lvert\Lambda\rvert$ for am	am_π	m_π/GeV	am_N	m_N/GeV
0.12	6.0	0.09146(78)	$16^3 \times 16$	1.869(39)	4.53(19)	-	-
0.148	5.9	0.0534(6)	$8^3 \times 32$	0.834(25)	3.08(9)	-	-
0.15	6.0	0.0555(6)	$8^3 \times 32$	0.741(16)	2.63(6)	-	-
0.15	5.9	0.0655(10)	$12^3 \times 32$	0.5104(18)	1.537(5)	0.9345(58)	2.88(2)

Table A.1: Lattice spacings and pion mass for QCD simulations with CLE.

A.7 Gauge cooling and dynamical stabilization

A.7.1 Gauge cooling

Gauge cooling was introduced in [105], we will briefly introduce it here. It is a method to keep the Langevin evolution close to the real SU(N) manifold in SL(N,$\mathbb{C}$) simulations. This is done by using gauge symmetry to move the evolution towards the unitary manifold. We use a more symmetric definition of the unitarity norm

$$d_U = \sum_n \mathrm{Tr}\left[U_\nu^\dagger(n)U_\nu(n) + \left(U_\nu^\dagger(n)\right)^{-1}\left(U_\nu(n)\right)^{-1} - 2\mathbb{1}\right] . \tag{A.38}$$

A gauge transformation in the opposite direction of the gradient of d_U is

$$\begin{aligned} U_\nu(n) \to & e^{-\sum_a \epsilon\alpha_{\mathrm{gc}}\lambda_a G_a(n)} U_\nu(n) \\ U_\nu(n-\hat{\nu}) \to & U_\nu(n-\hat{\nu}) e^{\sum_a \epsilon\alpha_{\mathrm{gc}}\lambda_a G_a(n)} \end{aligned} \tag{A.39}$$

with the cooling strength α_{gc}, the Langevin step size ϵ and

$$\begin{aligned} G_a(m) =& 2\mathrm{Tr}\ \lambda_a \left[U_\nu(m)U_\nu^\dagger(m) - U_\nu^\dagger(m-\hat{\nu})U_\nu(m-\hat{\nu})\right] \\ & + 2\mathrm{Tr}\ \lambda_a \left[-\left(U_\nu^\dagger(m)\right)^{-1}\left(U_\nu(m)\right)^{-1} - \left(U_\nu^\dagger(m-\hat{\nu})\right)^{-1}\left(U_\nu(m-\hat{\nu})\right)^{-1}\right] \end{aligned} \tag{A.40}$$

The particular value of α_{gc} and the number of cooling sweeps depend on the theory at hand. In QCD we typically use $\alpha_{\mathrm{gc}} \approx 1$ to 5 and between 10 and 20 sweeps. In SU(2) real-time simulations $\alpha_{\mathrm{gc}} = 1$ and just a few sweeps turned out to be sufficient. There are also adaptive ways to use gauge cooling [254].

A.7.2 Dynamical stabilization

Dynamical stabilization was investigated in [128]. Here no symmetry is used, but the drift is deformed to make the unitary manifold SU(N) attractive to the process. This is done via the substitution

$$K_\nu^a(n) \to K_\nu^a(n) + i\alpha_{\mathrm{ds}} i b^a(n) \left(\sum_c b^c(n)b^c(n)\right)^3 , \tag{A.41}$$

where b quantifies the distance to the unitary manifold

$$b^a(n) = \mathrm{Tr}\left[\lambda^a \sum_\nu U_\nu(n) U_\nu^\dagger(n)\right] . \tag{A.42}$$

This deformation was found heuristically. In principle, one has to be careful with dynamical stabilization since a non-holomorphic term is added to the drift. This deforms the original theory and in addition, the proof of convergence for complex Langevin does not go through without modification. In practice, it is useful to stabilize the simulation and the added term seems to vanish towards the continuum limit [128] thus restoring the applicability of the proof. The drawback of dynamical stabilization is that one has to study many values of α_{DS} and look for a stable region.

B Appendix to Lefschetz Thimbles

B.1 Some simple ideas to find stationary points and thimbles

Finding stationary points by scanning the real manifold

All contributing thimbles have a corresponding anti-thimble which intersects the original manifold. Hence if we start on the anti-thimble and solve the steepest ascent equation, we end up in the stationary point. One idea to find the fixed point is hence to discretize the real axis and flow from each point along the steepest ascent path [255]. We can measure the distance to a stationary point by looking at the derivative of the action, if it is very small we are close. We can then reach the fixed point by continuing the evolution with

$$\dot{z} = \begin{cases} -\frac{\partial S}{\partial z} & \text{FP attractive under } -\partial_z S \\ +\frac{\partial S}{\partial z} & \text{FP repulsive under } -\partial_z S \\ \pm e^{i\pi/2}\frac{\partial S}{\partial z} & \text{FP circular under } -\partial_z S \end{cases} . \tag{B.1}$$

This is visualized in the left plot of B.1.

Finding thimbles by random initialization in the whole space

Once we have found a fixed point we know the imaginary part of the action along the whole thimble. Hence we can find the thimble (and corresponding anti-thimble)

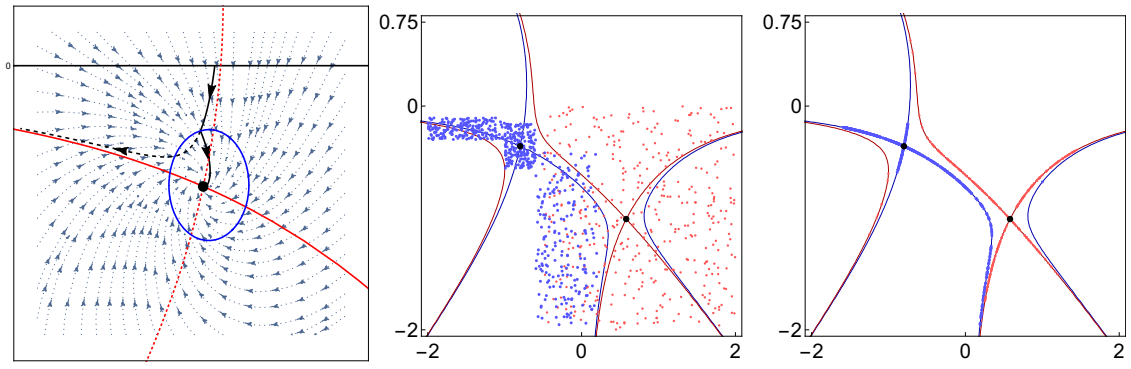

Figure B.1: Left: Visualization of the fixed point search. Center and right: Visualization of the Thimble search. See text for details.

by drawing random points close to the thimble (e.g. by successively going farther from the fixed point) or in the whole space and then minimizing the distance to the thimble using $M_\sigma(x,y) = |\mathrm{Im}S(x+iy) - \mathrm{Im}S(z_\sigma)|^2$ as a metric. The minimization is done using gradient descent

$$\begin{aligned} \dot{x} &= -\frac{\partial M_\sigma(x,y)}{\partial x}\,, \\ \dot{y} &= -\frac{\partial M_\sigma(x,y)}{\partial y}\,. \end{aligned} \tag{B.2}$$

We visualize the method for the z^4 model with parameters $\sigma = 1, \lambda = 1, h = 1 + i$ in figure B.1.

Note that both of those algorithms will become unfeasible in higher dimensions. The first method, because scanning a full high dimensional manifold is numerically unfeasible and the second one because typically thimbles are not the only manifolds having this particular imaginary part of the action. There are other curves with the same imaginary part, see figure B.1 which are typically dense in higher dimensions. The second algorithm can also be used to correct for numerical errors, i.e. drive the numerical evolution back onto the thimble if it is driven away by numerical uncertainties.

B.2 Combining Complex Langevin and Lefschetz Thimbles

Here we will discuss our insights from trying to combine Thimbles and complex Langevin [30].
We try to combine both methods by using coordinate transformations to restrict the complex Langevin evolution in a region close to thimbles. We formulate the complex Langevin equation in the transformed variables. We choose the transformation $z(u) = \exp(i\theta)u$, which will be used to rotate points onto the thimble. Here θ depends on Langevin time, it will be updated during the simulation. More general transformations are possible, but we found that more freedom leads to more instabilities. The transformed action reads

$$S_u := S(z(u)) - \log(z'(u))\,. \tag{B.3}$$

We are now free to formulate any condition which keeps us close to the thimble in order to get an equation for θ. Since the imaginary part is constant along thimbles, one possibility is to choose $\mathrm{Im}S_u = \text{const}$ for the transformed action or $\mathrm{Im}S = \text{const}$ for the original action. It turns out that both of those prescriptions lead to poles and cuts in the complex plane such that the dynamics of the process is severely restricted, see [30] for details.

Instead we combine the complex Langevin and steepest descent equations

$$\begin{aligned}\frac{\partial u}{\partial \tau} &= -\frac{\partial S_u}{\partial u} + \eta \\ \frac{\partial u}{\partial \tau} &= -\overline{\frac{\partial S_u}{\partial u}}\,,\end{aligned} \tag{B.4}$$

by demanding that on average their difference is zero

$$0 = \frac{\partial S_u}{\partial u} - \overline{\frac{\partial S_u}{\partial u}} \Rightarrow \partial_\tau \mathrm{Im}(u) = 0\,, \tag{B.5}$$

where we used the fact that the noise averages out. The idea is that complex Langevin already samples all relevant fixed points and that steepest descent keeps the evolution close to thimbles. Using equation (B.3), this leads to

$$\begin{aligned}\partial_\tau \mathrm{Re}\, u &= -\mathrm{Re}\,\frac{\partial S_u}{\partial u} + \eta \\ \partial_\tau \mathrm{Im}\, u &= -\mathrm{Im}\,\frac{\partial S_u}{\partial u} = -\mathrm{Im}\left(\frac{\partial S}{\partial z} z' - \frac{z''}{z'}\right) = 0\,,\end{aligned} \tag{B.6}$$

and we will use the second equation as a constraint to compute $\partial_\tau \theta$. In the Gaussian model with $S = \sigma z^2/2$ one can immediately see why a rotation is a good choice for the coordinate transformation, there $z' = \mathrm{e}^{i\,\theta}$, $z'' = 0$ leads to the constraint

$$\mathrm{Im}(\sigma\, z\, z') = \mathrm{Im}(\sigma\, \mathrm{e}^{2i\theta} u) = 0\,. \tag{B.7}$$

With $u \in \mathbb{R}$ and with $\sigma = \sigma_r\, \mathrm{e}^{i\theta_\sigma}$ this leads to the solution for equation (B.7)

$$\theta^* = -\frac{1}{2}\,\theta_\sigma\,. \tag{B.8}$$

This is just a rotation of the thimble in the original theory onto the real axis in the transformed theory. We take the derivative of the lower equation in (B.6), leading to

$$\dot{\theta} = \frac{\mathrm{Im}\left(\dfrac{\partial^2 S_u}{\partial u^2}\,\dfrac{\partial S_u}{\partial u}\right)}{\mathrm{Im}\left(\dfrac{\partial^2 S_u}{\partial u\,\partial \theta}\right)}\,. \tag{B.9}$$

Examining this in the Gaussian model, it turns out that this drives the evolution towards the anti-thimble instead of the thimble. Motivated by this heuristic finding we switch the prefactor in the equation for $\dot{\theta}$, leading to

$$\dot{\theta} = -\frac{\mathrm{Im}\left(\dfrac{\partial^2 S_u}{\partial u^2}\,\dfrac{\partial S_u}{\partial u}\right)}{\mathrm{Im}\left(\dfrac{\partial^2 S_u}{\partial u\,\partial \theta}\right)}\,. \tag{B.10}$$

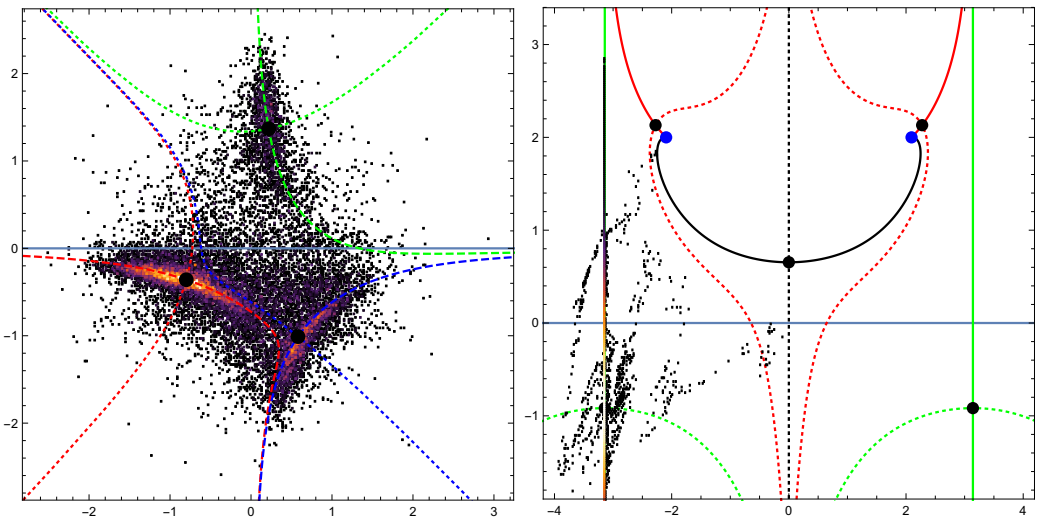

Figure B.2: Density plot of the evolution of the combination of complex Langevin with Lefschetz thimbles. Left: z^4 model. Right: U(1) model. A discussion can be found in the text.

.

Due to its heuristic nature, there is no mathematical reason why this equation should keep the evolution close to thimbles. However, simulations in the model from equation (4.16) with $\sigma = 1$, $\lambda = 1$ and $h = 1 + i$, i.e. the model 2 from section 4.2 show that the numerical evolution stays close to the thimbles and respects their hierarchy, i.e. it samples the thimble with the strongest contribution the most, see left plot of figure B.2. Unfortunately, observables come out wrong, also for different models [256]. Similarly, when applying the method to the U(1) model from equation (4.17), the evolution gets caught on a non-contributing thimble, see right plot of figure B.2. The reason is that the thimble is a straight line and hence no rotation is necessary once the process is caught on the thimble.

B.3 Local updating algorithm U(1)

Here we give the simulation algorithm for the U(1) part of section 4.4. We rewrite equation (4.28) as

$$|\Delta U_k| = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{B.11}$$

with $a = \text{Im } \beta P_{ij} e^{i\Delta\phi_k^R}$, $b = -\text{Im } \beta(P_{ij} + P_{ji})$, $c = \text{Im } \beta P_{ij} e^{-i\Delta\phi_k^R}$. Note that there are two solutions, one with a plus sign and one with a minus sign in front of the square root. If we start on the fixed point and go into positive $\Delta\phi^R$ direction, we find that the action becomes larger for the solution with the plus sign and thus corresponds

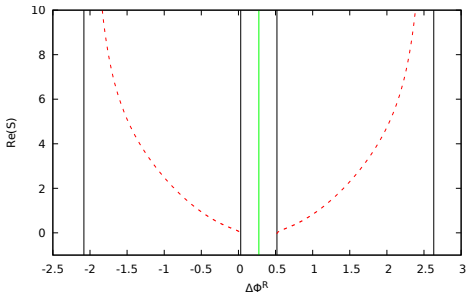

Figure B.3: Real part of the action in the defined region (red dashed line), vertical black lines are the boundaries of the range where $|\Delta U|$ is defined, the vertical green line gives the switching point.

to the thimble[17]. Similarly in negative $\Delta\phi^R$ direction the minus sign corresponds to the thimble. Both other combinations correspond to the anti-thimbles. In general, we do not start on the fixed point since we fix the local theory to the thimble, not the global one. Instead, we start at the switching angle which is the point where the plus and minus solution from the quadratic formula cross, i.e. the position of the minimum of the term under the square root, it can be written as

$$\phi^R_{\text{sw}} = \frac{\arg(P_j) - \arg(P_i)}{2} \,. \tag{B.12}$$

In general the square root in equation (B.11) can be undefined, since $|\Delta U| \in \mathbb{R}$. Hence, we need to choose the update proposals from a region where the term under the square root is larger or equal to zero. Additionally, $|\Delta U| > 0$, since it is an absolute value. This gives us additional constraints depending on the values of a,b and c. We will not explain this in detail here and instead give the algorithm to find the boundaries for the region where the solutions are defined, the intervals are $[\phi^-_L, \phi^-_R]$ for the solution with the minus sign and $[\phi^+_L, \phi^+_R]$ for the solution with the plus sign. Figure B.3 shows a typical situation. Here the real part of the action is plotted in the region where $|\Delta U|$ is defined and the boundaries, as well as the switching angle, are shown.

Note that the problem is 2π periodic and reflection symmetric about the switching angle, i.e. if we find solutions in $[\phi^R_{\text{sw}}, \phi^R_{\text{sw}} + \pi]$ we can reconstruct everything else from this. The algorithm for the boundaries reads

1. Find the switching angle ϕ^R_{sw} (see (B.12)).
2. Find a zero ϕ^+_L of $(b^2 - 4ac)$ in the interval $[\phi^R_{\text{Sw}}, \phi^R_{\text{Sw}} + \pi/2]$. If there is none, set $\phi^+_L = \phi^R_{\text{Sw}}$.
3. Set $\phi^+_R = 2\phi^R_{\text{Sw}} + \pi - \phi^+_L$ (this is the left boundary mirrored at $\phi^R_{\text{Sw}} + \pi/2$).

[17] Remember that e^{-S} has its maximum at the fixed point on the thimble, hence away from the fixed point the action has to become larger.

4. If $b < 0$:
   a) Calculate
   $$\phi_a = \arctan2(-\mathrm{Im}(\beta P_{ij}), \mathrm{Re}(\beta P_{ij})),$$
   and translate it into our interval by
   $$\tilde{\phi}_a = \phi_a - \lfloor \frac{\phi_a - \phi^R_{\mathrm{Sw}}}{\pi} \rfloor \pi.$$
   b) If $\tilde{\phi}_a \in [\phi^+_L, \phi^+_R]$ then calculate
   $$\mathrm{Im}(\beta P_{ij} e^{i(\tilde{\phi}_a - \pi/2)}).$$
   If it is smaller than zero, set $\phi^+_L = \tilde{\phi}_a$, otherwise set $\phi^+_R = \tilde{\phi}_a$.
5. If $b > 0$:
   a) Calculate
   $$\phi_c = \arctan 2(\mathrm{Im}(\beta P_{ji}), \mathrm{Re}(\beta P_{ji})),$$
   and translate it into our interval by
   $$\tilde{\phi}_c = \phi_c - \lfloor \frac{\phi_c - \phi^R_{\mathrm{Sw}}}{\pi} \rfloor \pi.$$
   b) If $\tilde{\phi}_c \in [\phi^+_L, \phi^+_R]$ then calculate
   $$\mathrm{Im}(\beta P_{ji} e^{-i(\tilde{\phi}_c - \pi/2)}).$$
   If it is smaller than zero, set $\phi^+_R = \tilde{\phi}_c$, otherwise set $\phi^+_L = \tilde{\phi}_c$.
6. Mirror the boundaries on the right side of the switching angle to the left side by
   $$\begin{aligned} \phi^-_L &= 2\phi^R_{\mathrm{Sw}} - \phi^+_R \\ \phi^-_R &= 2\phi^R_{\mathrm{Sw}} - \phi^+_L \end{aligned}$$

Having found the boundaries, the update prescription for a given link is rather simple

1. Choose with equal probability whether the update shall be done for the plus or minus solution of the quadratic equation (B.11).
2. Compute the boundaries $[\phi_{\min}, \phi_{\max}]$ for the corresponding branch.
3. Draw a uniform random number from the interval $[\phi_{\min}, \phi_{\max}]$ as a proposal for $\Delta\phi^R$.
4. Do a Metropolis accept/reject step with the real part of the action.

B.4 Tangent space and Takagi vector basis

In scalar theory on a two-point lattice given by equation (4.35), we can show that the thimble is the tangent space around the fixed point, which is convenient for choosing a good basis for the simulations and analytical integrals. Consider the Hessian matrix

$$H = -iM = -i\frac{\partial S}{\partial \phi_k \phi_l} = -i\begin{pmatrix}(2-m^2) & -2\\ -2 & (2-m^2)\end{pmatrix}. \tag{B.13}$$

The tangent space at the critical point is spanned by the Takagi vectors v, defined via

$$\bar{H}\bar{v} = \lambda v\,, \tag{B.14}$$

where $\lambda \in \mathbb{R}$ and the bar stands for complex conjugation. $\lambda > 0$ corresponds to the thimble and $\lambda < 0$ to the anti-thimble. We can bring this in the form of a normal eigenproblem, by writing $v = c\tilde{v}$, where c is a complex phase,

$$\bar{H}\bar{v} = i\bar{c}M\bar{\tilde{v}} = i\bar{c}\alpha\tilde{v} = i\bar{c}\alpha\frac{v}{c} \tag{B.15}$$

where we used that M is a real symmetric matrix and thus $\tilde{v}$ are real vectors. Thus, dividing both sides of equation (B.14) by $i\bar{c}$ we have

$$M\tilde{v} = \alpha\tilde{v}\,, \tag{B.16}$$

where α now are the eigenvalue of M. For the eigenvectors and eigenvalues we find

$$\tilde{v}_1 = \begin{pmatrix}1\\1\end{pmatrix} \text{ with } \alpha_1 = -m^2 \text{ and } \tilde{v}_2 = \begin{pmatrix}1\\-1\end{pmatrix} \text{ with } \alpha_2 = 4-m^2\,. \tag{B.17}$$

We can regain the Takagi vectors by multiplying with ic. Since c is just a phase, $\lambda = \pm\alpha$. Thus we find

$$c = \sqrt{\pm\text{sign}(\alpha)i}\,, \tag{B.18}$$

where again the positive sign corresponds to the thimble. We find

$$c_1 = e^{-i\pi/4} \text{ and } c_2 = e^{i\text{sign}(4-m^2)\pi/4} = \begin{cases} e^{i\pi/4}, & m < 2\\ e^{-i\pi/4}, & m > 2\,. \end{cases} \tag{B.19}$$

Note that when the phases are not equal c_{12}, i.e. for m<2, we cannot write down our local algorithm around equation (4.37). This is because this algorithm updates both lattice sites separately, i.e. the update is suggested according to

$$\begin{pmatrix}\Delta\phi_1^R + i\Delta\phi_1^I\\ 0\end{pmatrix} \text{ and } \begin{pmatrix}0\\ \Delta\phi_2^R + i\Delta\phi_2^I\end{pmatrix}, \tag{B.20}$$

which is not possible if the phases are unequal.
We can also use this parametrization and insert the fields in tangent space

$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = x_1 c_1 \tilde{v}_1 + x_2 c_2 \tilde{v}_2\,, \tag{B.21}$$

such that the action becomes

$$\begin{aligned} S =& -i\left((\phi_1-\phi_2)^2 - \frac{m^2}{2}(\phi_1^2+\phi_2^2)\right)\\ =& -i\left(4\text{sign}(4-m^2)ix_2^2 - \frac{m^2}{2}(-2ix_1^2+2\text{sign}(4-m^2)ix_2^2)\right)\\ =& m^2x_1^2 + |4-m^2|x_2^2\,, \end{aligned} \tag{B.22}$$

where the x_i are real variables. This action is completely real and positive and hence the tangent space and the thimble are equivalent. It is also simple enough to compute observables analytically by writing the partition function in terms of the x_i instead of ϕ_i. Since the action is quadratic and the Jacobian is just a constant matrix those are just simple Gaussian integrals. We find

$$\langle\phi^2\rangle = \left\langle\frac{\phi_1^2+\phi_2^2}{2}\right\rangle = -\frac{i}{2}\left(\frac{1}{m^2}+\frac{1}{m^2-4}\right)\,, \tag{B.23}$$

C Appendix to neural networks

C.1 Network architecture and training

We used a combination of the keras [257] and scikit-learn [258] packages for python to set up our machine learning pipeline.

The architecture of both networks is shown in tables C.1 and C.2. We chained one-dimensional convolutional layers to produce the correct receptive fields. For example in the SU(2) case with the $2\times1\times1\times1$ network, the first layer outputs patches of size $2 \times 8 \times 8 \times 1$. The second layer then outputs patches of $2 \times 8 \times 1 \times 1$ and the third layer has has $C = 1$ and outputs $2 \times 1 \times 1 \times 1$. Thus together they make up the receptive field. The average pooling layer just averages over all receptive fields, thus outputting a single averaged $2 \times 1 \times 1 \times 1$ patch. The flattening layer is just there for technical reasons and the dense layers correspond to the prediction network. A, B, C and n_A, n_B, n_C are the sizes and numbers of filters respectively, we always used $n_C = 1$.

We used ReLU activation functions, see equation (5.3) everywhere except for the last layer, where we used the sigmoid function, see equation (5.2). We chose the binary cross-entropy as in equation (5.5) as our loss function and employed RMSprop gradient descent as implemented in keras for the minimization of the loss and training. The networks were trained with a maximum of 300 epochs. We used 20% as a validation set in all cases.

Layer	Output Shape	Kernel Size
InputLayer	(784, 1)	
Convolution1D	(784/(A), n_A)	A
Convolution1D	(784/(A$\times B$), n_B)	B
Convolution1D	(784/(A$\times B \times C$), n_C)	C
Average Pooling	(1, n_C)	
Flatten	(n_C)	
Dense	(n_D)	
Dense	(1)	

Table C.1: Architecture of the Ising model neural network, see text for details

Layer	Output Shape	Kernel Size
InputLayer	(1024, 16)	
Convolution1D	(1024/(A), n_A)	A
Convolution1D	(1024/(A$\times B$), n_B)	B
Convolution1D	(1024/(A$\times B \times C$), n_C)	C
Average Pooling	(1, n_C)	
Flatten	(n_C)	
Dense	(n_D)	
Dense	(1)	

Table C.2: Architecture of SU(2) neural network, see text for details.

Acknowledgments

I am grateful to many people who supported me professionally and personally during my work on this thesis.

First and foremost I thank my supervisor, Prof. Dr. Ion-Olimpiu Stamatescu for the enjoyable and excellent supervision. While he always provided me with interesting topics to work on, he also let me pursue my own interests, which I find particularly important. I am also grateful for the many discussions in a pleasant atmosphere.

I am grateful to Prof. Dr. Tilman Plehn for being my second referee.

My special thanks go to all my collaborators. In particular, I would like to thank Dr. Dénes Sexty for providing me with his GPU code and for helping me with his lattice expertise. I thank Prof. Dr. Jan M. Pawlowski, Dr. Erhard Seiler, Dr. Christian Schmidt, Felix P.G. Ziegler and Felix Ziesché for the enjoyable collaboration and insightful discussions. I thank Dr. Sebastian Johann Wetzel for letting me join his work on neural networks and the many interesting discussions that arose from our collaboration.

I am also grateful many members of the extended group around Prof. Pawlowski and many other friends and colleagues at the Institute for making my time there enjoyable.

I thank my collaborators as well as Aaron Held for proofreading parts of this thesis.

On a personal level, I would like to thank my friends, family and my fiancee for the support in all respects during my time as a Ph.D. student.

I acknowledge support by the DFG from the project "QCD und andere Modelle bei endlicher Dichte, stochastische Quantisierung, Nichtgleichgewichtsphänomene in der Quantenfeldtheorie" with number STA 283/16-2.
I acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

Lists

List of Figures

List of Tables

Bibliography

[1] G. Aad *et al.* (ATLAS), "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," Phys. Lett. **B716**, 1–29 (2012), arXiv:1207.7214 .

[2] S. Chatrchyan *et al.* (CMS), "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," Phys. Lett. **B716**, 30–61 (2012), arXiv:1207.7235 .

[3] S. Aoki *et al.* (Flavour Lattice Averaging Group), "FLAG Review 2019," arXiv:1902.08191 .

[4] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories," Phys. Rev. Lett. **30**, 1343–1346 (1973).

[5] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?" Phys. Rev. Lett. **30**, 1346–1349 (1973).

[6] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, "The Order of the quantum chromodynamics transition predicted by the standard model of particle physics," Nature **443**, 675–678 (2006), arXiv:hep-lat/0611014 .

[7] A. K. Cyrol, M. Mitter, J. M. Pawlowski, and N. Strodthoff, "Nonperturbative finite-temperature Yang-Mills theory," Phys. Rev. **D97**, 054015 (2018), arXiv:1708.03482 .

[8] M. Mitter, J. M. Pawlowski, and N. Strodthoff, "Chiral symmetry breaking in continuum QCD," Phys. Rev. **D91**, 054035 (2015), arXiv:1411.7978 .

[9] A. K. Cyrol, L. Fister, M. Mitter, J. M. Pawlowski, and N. Strodthoff, "Landau gauge Yang-Mills correlation functions," Phys. Rev. **D94**, 054005 (2016), arXiv:1605.01856 .

[10] A. K. Cyrol, M. Mitter, J. M. Pawlowski, and N. Strodthoff, "Nonperturbative quark, gluon, and meson correlators of unquenched QCD," Phys. Rev. **D97**, 054006 (2018), arXiv:1706.06326 .

[11] C. S. Fischer, "QCD at finite temperature and chemical potential from Dyson-Schwinger equations," (2018), arXiv:1810.12938 .

[12] G. Aarts, "Introductory lectures on lattice QCD at nonzero baryon number," *Proceedings, 13th International Workshop on Hadron Physics: Angra dos Reis, Rio de Janeiro, Brazil, March 22-27, 2015*, J. Phys. Conf. Ser. **706**, 022004 (2016), arXiv:1512.05145 .

[13] B.-J. Schaefer, J. M. Pawlowski, and J. Wambach, "The Phase Structure of the Polyakov–Quark-Meson Model," Phys. Rev. **D76**, 074023 (2007), arXiv:0704.3234 .

[14] R.-A. Tripolt, N. Strodthoff, L. von Smekal, and J. Wambach, "Spectral Functions for the Quark-Meson Model Phase Diagram from the Functional Renormalization Group," Phys. Rev. **D89**, 034010 (2014), arXiv:1311.0630 .

[15] A. Bazavov *et al.*, "Chiral crossover in QCD at zero and non-zero chemical potentials," (2018), arXiv:1812.08235 .

[16] A. Bazavov *et al.*, "The QCD Equation of State to $\mathcal{O}(\mu_B^6)$ from Lattice QCD," Phys. Rev. **D95**, 054504 (2017), arXiv:1701.04325 .

[17] S. Borsanyi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, and K. K. Szabo, "Full result for the QCD equation of state with 2+1 flavors," Phys. Lett. **B730**, 99–104 (2014), arXiv:1309.5258 .

[18] S. Borsanyi, Z. Fodor, J. N. Guenther, S. K. Katz, K. K. Szabo, A. Pasztor, I. Portillo, and C. Ratti, "Higher order fluctuations and correlations of conserved charges from lattice QCD," JHEP **10**, 205 (2018), arXiv:1805.04445 .

[19] C. Bonati, M. D'Elia, F. Negro, F. Sanfilippo, and K. Zambello, "Curvature of the pseudocritical line in QCD: Taylor expansion matches analytic continuation," Phys. Rev. **D98**, 054510 (2018), arXiv:1805.02960 .

[20] K. Fukushima and T. Hatsuda, "The phase diagram of dense QCD," Rept. Prog. Phys. **74**, 014001 (2011), arXiv:1005.4814 .

[21] J. B. Kogut and M. A. Stephanov, "The phases of quantum chromodynamics: From confinement to extreme environments," Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **21**, 1–364 (2004).

[22] B. Friman, C. Hohne, J. Knoll, S. Leupold, J. Randrup, R. Rapp, and P. Senger, "The CBM physics book: Compressed baryonic matter in laboratory experiments," Lect. Notes Phys. **814**, pp.1–980 (2011).

[23] "Large hadron collider (lhc)," `https://home.cern/`.

[24] "Relativistic heavy ion collider (rhic)," `https://www.bnl.gov/rhic/`.

[25] A. Andronic, P. Braun-Munzinger, K. Redlich, and J. Stachel, "Decoding the phase structure of QCD via particle production at high energy," Nature **561**, 321–330 (2018), arXiv:1710.09425 .

[26] P. Mehta, M. Bukov, C.-H. Wang, A. r. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A high-bias, low-variance introduction to Machine Learning for physicists," (2018), arXiv:1803.08823 .

[27] Y. Mori, K. Kashiwa, and A. Ohnishi, "Application of a neural network to the sign problem via the path optimization method," PTEP **2018**, 023B04 (2018), arXiv:1709.03208 .

[28] F. Bursa and M. Kroyter, "A simple approach towards the sign problem using path optimisation," JHEP **12**, 054 (2018), arXiv:1805.04941 .

[29] M. Scherzer, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Complex Langevin and boundary terms," Phys. Rev. **D99**, 014512 (2019), arXiv:1808.05187 .

[30] S. Bluecher, J. M. Pawlowski, M. Scherzer, M. Schlosser, I.-O. Stamatescu, S. Syrkowski, and F. P. G. Ziegler, "Reweighting Lefschetz Thimbles," SciPost Phys. **5**, 044 (2018), arXiv:1803.08418 .

[31] S. J. Wetzel and M. Scherzer, "Machine Learning of Explicit Order Parameters: From the Ising Model to SU(2) Lattice Gauge Theory," Phys. Rev. **B96**, 184410 (2017), arXiv:1705.05582 .

[32] G. Aarts, K. Boguslavski, M. Scherzer, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Getting even with CLE," *Proceedings, 35th International Symposium on Lattice Field Theory (Lattice 2017): Granada, Spain, June 18-24, 2017*, EPJ Web Conf. **175**, 14007 (2018), arXiv:1710.05699 [hep-lat] .

[33] M. Scherzer, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Complex Langevin: Boundary terms and application to QCD," in *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018* (2018) arXiv:1810.09713 [hep-lat] .

[34] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995).

[35] S. Weinberg, *The quantum theory of fields. Vol. 2: Modern applications* (Cambridge University Press, 2013).

[36] H. J. Rothe, *Lattice Gauge Theories*, 4th ed. (World Scientific, 2012).

[37] C. Gattringer and C. B. Lang, "Quantum chromodynamics on the lattice," Lect. Notes Phys. **788**, 1–343 (2010).

[38] S. W. Otto and J. D. Stack, "The SU(3) Heavy Quark Potential with High Statistics," Phys. Rev. Lett. **52**, 2328 (1984).

[39] P. de Forcrand, "QCD from Chippewa Falls," in *Conference on Frontiers of Quantum Monte Carlo Los Alamos, New Mexico, September 3-6, 1985* (1985) pp. 1077–1094.

[40] H. Q. Ding, "Improved determination of the heavy quark potential in lattice QCD," Phys. Rev. **D44**, 2200–2202 (1991).

[41] G. 't Hooft, "On the Phase Transition Towards Permanent Quark Confinement," Nucl. Phys. **B138**, 1–25 (1978).

[42] J. S. Bell and R. Jackiw, "A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ model," Nuovo Cim. **A60**, 47–61 (1969).

[43] S. L. Adler, "Axial vector vertex in spinor electrodynamics," Phys. Rev. **177**, 2426–2438 (1969).

[44] S. L. Adler and W. A. Bardeen, "Absence of higher order corrections in the anomalous axial vector divergence equation," Phys. Rev. **182**, 1517–1536 (1969).

[45] K. Fujikawa, "Path Integral Measure for Gauge Invariant Fermion Theories," Phys. Rev. Lett. **42**, 1195–1198 (1979).

[46] G. 't Hooft, "Symmetry Breaking Through Bell-Jackiw Anomalies," Phys. Rev. Lett. **37**, 8–11 (1976).

[47] S. Borsanyi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, C. Ratti, and K. K. Szabo (Wuppertal-Budapest), "Is there still any Tc mystery in lattice QCD? Results with physical masses in the continuum limit III," JHEP **09**, 073 (2010), arXiv:1005.3508 .

[48] H. T. Ding *et al.*, "The chiral phase transition temperature in (2+1)-flavor QCD," (2019), arXiv:1903.04801 .

[49] R. Bellwied, S. Borsanyi, Z. Fodor, J. Günther, S. D. Katz, C. Ratti, and K. K. Szabo, "The QCD phase diagram from analytic continuation," Phys. Lett. **B751**, 559–564 (2015), arXiv:1507.07510 .

[50] J. Cleymans, H. Oeschler, K. Redlich, and S. Wheaton, "Comparison of chemical freeze-out criteria in heavy-ion collisions," Phys. Rev. **C73**, 034905 (2006), arXiv:hep-ph/0511094 .

[51] M. Tanabashi *et al.* (Particle Data Group), "Review of Particle Physics," Phys. Rev. **D98**, 030001 (2018).

[52] S. Borsanyi, Z. Fodor, S. D. Katz, S. Krieg, C. Ratti, and K. K. Szabo, "Freeze-out parameters from electric charge and baryon number fluctuations: is there consistency?" Phys. Rev. Lett. **113**, 052301 (2014), arXiv:1403.4576 .

[53] T. D. . Cohen, "Functional integrals for QCD at nonzero chemical potential and zero density," Phys. Rev. Lett. **91**, 222001 (2003), arXiv:hep-ph/0307089 .

[54] P. Chomaz, M. Colonna, and J. Randrup, "Nuclear spinodal fragmentation," Phys. Rept. **389**, 263–440 (2004).

[55] M. D'Agostino, M. Bruno, F. Gulminelli, F. Cannata, P. Chomaz, G. Casini, E. Geraci, F. Gramegna, A. Moroni, and G. Vannini, "Nuclear liquid-gas phase transition: Experimental signals," *Proceedings, 18th Nuclear Physics Division Conference of the EPS: Phase Transitions in Strongly Interacting Matter (NPDC 18): Prague, Czech Republic, August 23-29, 2004*, Nucl. Phys. **A749**, 55–64 (2005).

[56] F. Özel and P. Freire, "Masses, Radii, and the Equation of State of Neutron Stars," Ann. Rev. Astron. Astrophys. **54**, 401–440 (2016), arXiv:1603.02698 [astro-ph.HE] .

[57] J. I. Kapusta and C. Gale, *Finite-temperature field theory: Principles and applications*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2011).

[58] A. K. Das, *Finite Temperature Field Theory* (World Scientific, New York, 1997).

[59] J. Berges, S. Borsanyi, D. Sexty, and I. O. Stamatescu, "Lattice simulations of real-time quantum fields," Phys. Rev. **D75**, 045007 (2007), arXiv:hep-lat/0609058 .

[60] F. J. Wegner, "Duality in generalized ising models and phase transitions without local order parameters," Journal of Mathematical Physics **12**, 2259–2272 (1971).

[61] K. G. Wilson, "Confinement of Quarks," Phys. Rev. **D10**, 2445–2459 (1974).

[62] I. Montvay and G. Münster, *Quantum Fields on a Lattice*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 1994).

[63] T. DeGrand and C. DeTar, *Lattice Methods for Quantum Chromodynamics* (World Scientific, 2006).

[64] F. Knechtli, M. Günther, and M. Peardon, *Lattice Quantum Chromodynamics: Practical Essentials*, SpringerBriefs in Physics (Springer, 2017).

[65] L. Giusti, F. Rapuano, M. Talevi, and A. Vladikas, "The QCD chiral condensate from the lattice," Nucl. Phys. **B538**, 249–277 (1999), arXiv:hep-lat/9807014 [hep-lat] .

[66] M. Lüscher, "Properties and uses of the Wilson flow in lattice QCD," JHEP **08**, 071 (2010), [Erratum: JHEP03,092(2014)], arXiv:1006.4518 .

[67] J. M. Pawlowski, I.-O. Stamatescu, and F. P. G. Ziegler, "Cooling Stochastic Quantization with colored noise," Phys. Rev. **D96**, 114505 (2017), arXiv:1705.06231 [hep-lat] .

[68] S. Borsanyi *et al.*, "High-precision scale setting in lattice QCD," JHEP **09**, 010 (2012), arXiv:1203.4469 .

[69] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).

[70] M. Creutz, "Monte Carlo Study of Quantized SU(2) Gauge Theory," Phys. Rev. **D21**, 2308–2315 (1980).

[71] P. de Forcrand, "Simulating QCD at finite density," *Proceedings, 27th International Symposium on Lattice field theory (Lattice 2009): Beijing, P.R. China, July 26-31, 2009*, PoS **LAT2009**, 010 (2009), arXiv:1005.0539 .

[72] M. Troyer and U.-J. Wiese, "Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations," Phys. Rev. Lett. **94**, 170201 (2005), arXiv:cond-mat/0408370 .

[73] Z. Fodor and S. D. Katz, "A New method to study lattice QCD at finite temperature and chemical potential," Phys. Lett. **B534**, 87–92 (2002), arXiv:hep-lat/0104001 .

[74] P. de Forcrand, J. Langelage, O. Philipsen, and W. Unger, "Lattice QCD Phase Diagram In and Away from the Strong Coupling Limit," Phys. Rev. Lett. **113**, 152002 (2014), arXiv:1406.4397 .

[75] G. Gagliardi and W. Unger, "Towards a Dual Representation of Lattice QCD," in *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018* (2018) arXiv:1811.02817 .

[76] C. Gattringer, "New developments for dual methods in lattice field theory at non-zero density," *Proceedings, 31st International Symposium on Lattice Field Theory (Lattice 2013): Mainz, Germany, July 29-August 3, 2013*, PoS **LATTICE2013**, 002 (2014), arXiv:1401.7788 .

[77] Y. Delgado Mercado, C. Gattringer, and A. Schmidt, "Dual Lattice Simulation of the Abelian Gauge-Higgs Model at Finite Density: An Exploratory Proof of Concept Study," Phys. Rev. Lett. **111**, 141601 (2013), arXiv:1307.6120 .

[78] F. Bruckmann, C. Gattringer, T. Kloiber, and T. Sulejmanpasic, "Dual lattice representations for O(N) and CP(N1) models with a chemical potential," Phys. Lett. **B749**, 495–501 (2015), [Erratum: Phys. Lett.B751,595(2015)], arXiv:1507.04253 .

[79] C. Gattringer, D. Göschl, and T. Sulejmanpasic, "Dual simulation of the 2d U(1) gauge Higgs model at topological angle $\theta = \pi$: Critical endpoint behavior," Nucl. Phys. **B935**, 344–364 (2018), arXiv:1807.07793 .

[80] K. Langfeld, "Density-of-states," *Proceedings, 34th International Symposium on Lattice Field Theory (Lattice 2016): Southampton, UK, July 24-30, 2016*, PoS **LATTICE2016**, 010 (2017), arXiv:1610.09856 .

[81] K. Langfeld, B. Lucini, and A. Rago, "The density of states in gauge theories," Phys. Rev. Lett. **109**, 111601 (2012), arXiv:1204.3243 .

[82] N. Garron and K. Langfeld, "Tackling the sign problem with a moment expansion and application to Heavy dense QCD," *Proceedings, 34th International Symposium on Lattice Field Theory (Lattice 2016): Southampton, UK, July 24-30, 2016*, PoS **LATTICE2016**, 084 (2016), arXiv:1611.01378 .

[83] G. Aarts, F. Attanasio, B. Jäger, and D. Sexty, "The QCD phase diagram in the limit of heavy quarks using complex Langevin dynamics," JHEP **09**, 087 (2016), arXiv:1606.05561 .

[84] D. Sexty, "Simulating full QCD at nonzero density using the complex Langevin equation," Phys. Lett. **B729**, 108–111 (2014), arXiv:1307.7748 .

[85] M. Cristoforetti, F. Di Renzo, and L. Scorzato (AuroraScience), "New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble," Phys. Rev. **D86**, 074506 (2012), arXiv:1205.3996 .

[86] A. Alexandru, P. F. Bedaque, H. Lamm, S. Lawrence, and N. C. Warrington, "Fermions at Finite Density in 2+1 Dimensions with Sign-Optimized Manifolds," Phys. Rev. Lett. **121**, 191602 (2018), arXiv:1808.09799 .

[87] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, "Sign problem and Monte Carlo calculations beyond Lefschetz thimbles," JHEP **05**, 053 (2016), arXiv:1512.08764 .

[88] J. R. Klauder, "Stochastic quantization," *Recent developments in high-energy physics. Proceedings, 22. Internationale Universitätswochen für Kernphysik: Schladming, Austria, February 23 - March 5, 1983*, Acta Phys. Austriaca Suppl. **25**, 251–281 (1983).

[89] J. R. Klauder, "A Langevin Approach to Fermion and Quantum Spin Correlation Functions," J. Phys. **A16**, L317 (1983).

[90] J. R. Klauder, "Coherent State Langevin Equations for Canonical Quantum Systems With Applications to the Quantized Hall Effect," Phys. Rev. **A29**, 2036–2047 (1984).

[91] G. Parisi, "On complex probabilities," Phys. Lett. **131B**, 393–395 (1983).

[92] J. Ambjorn and S. K. Yang, "Numerical Problems in Applying the Langevin Equation to Complex Effective Actions," Phys. Lett. **165B**, 140 (1985).

[93] J. R. Klauder and W. P. Petersen, "Spectrum of certain nonselfadjoint operators and solutions of Langevin equations with complex drift," 10.1007/BF01007974.

[94] H. Q. Lin and J. E. Hirsch, "Monte carlo versus langevin methods for nonpositive definite weights," Phys. Rev. B **34**, 1964–1967 (1986).

[95] J. Ambjorn, M. Flensburg, and C. Peterson, "The Complex Langevin Equation and Monte Carlo Simulations of Actions With Static Charges," Nucl. Phys. **B275**, 375–397 (1986).

[96] F. Karsch and H. W. Wyld, "Complex Langevin Simulation of the SU(3) Spin Model With Nonzero Chemical Potential," Phys. Rev. Lett. **55**, 2242 (1985).

[97] H. Okamoto, K. Okano, L. Schulke, and S. Tanaka, "The Role of a Kernel in Complex Langevin Systems," Nucl. Phys. **B324**, 684–714 (1989).

[98] K. Okano, L. Schulke, and B. Zheng, "Kernel controlled complex Langevin simulation: Field dependent kernel," Phys. Lett. **B258**, 421–426 (1991).

[99] K. Okano, L. Schulke, and B. Zheng, "Complex Langevin simulation," Prog. Theor. Phys. Suppl. **111**, 313–347 (1993).

[100] K. Fujimura, K. Okano, L. Schulke, K. Yamagishi, and B. Zheng, "On the segregation phenomenon in complex Langevin simulation," Nucl. Phys. **B424**, 675–689 (1994), arXiv:hep-th/9311174 .

[101] L. L. Salcedo, "Spurious solutions of the complex Langevin equation," Phys. Lett. **B305**, 125–130 (1993).

[102] J. Berges and I. O. Stamatescu, "Simulating nonequilibrium quantum fields with stochastic quantization techniques," Phys. Rev. Lett. **95**, 202003 (2005), arXiv:hep-lat/0508030 .

[103] J. Berges and D. Sexty, "Real-time gauge theory simulations from stochastic quantization with optimized updating," Nucl. Phys. **B799**, 306–329 (2008), arXiv:0708.0779 .

[104] G. Aarts and I.-O. Stamatescu, "Stochastic quantization at finite chemical potential," JHEP **09**, 018 (2008), arXiv:0807.1597 .

[105] E. Seiler, D. Sexty, and I.-O. Stamatescu, "Gauge cooling in complex Langevin for QCD with heavy quarks," Phys. Lett. **B723**, 213–216 (2013), arXiv:1211.3709 .

[106] A. Mollgaard and K. Splittorff, "Complex Langevin Dynamics for chiral Random Matrix Theory," Phys. Rev. **D88**, 116007 (2013), arXiv:1309.4335 .

[107] J. Bloch, J. Glesaaen, J. J. M. Verbaarschot, and S. Zafeiropoulos, "Complex Langevin Simulation of a Random Matrix Model at Nonzero Chemical Potential," JHEP **03**, 015 (2018), arXiv:1712.07514 .

[108] G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Simulating QCD at nonzero baryon density to all orders in the hopping parameter expansion," Phys. Rev. **D90**, 114505 (2014), arXiv:1408.3770 .

[109] K. Nagata, J. Nishimura, and S. Shimasaki, "Testing the criterion for correct convergence in the complex Langevin method," JHEP **05**, 004 (2018), arXiv:1802.01876 .

[110] K. Nagata, J. Nishimura, and S. Shimasaki, "Gauge cooling for the singular-drift problem in the complex Langevin method - a test in Random Matrix Theory for finite density QCD," JHEP **07**, 073 (2016), arXiv:1604.07717 .

[111] G. Aarts, "Can stochastic quantization evade the sign problem? The relativistic Bose gas at finite chemical potential," Phys. Rev. Lett. **102**, 131601 (2009), arXiv:0810.2089 .

[112] G. Aarts, "Complex Langevin dynamics at finite chemical potential: Mean field analysis in the relativistic Bose gas," JHEP **05**, 052 (2009), arXiv:0902.4686 .

[113] J. M. Pawlowski and C. Zielinski, "Thirring model at finite density in 0+1 dimensions with stochastic quantization: Crosscheck with an exact solution," Phys. Rev. **D87**, 094503 (2013), arXiv:1302.1622 .

[114] J. M. Pawlowski and C. Zielinski, "Thirring model at finite density in 2+1 dimensions with stochastic quantization," Phys. Rev. **D87**, 094509 (2013), arXiv:1302.2249 .

[115] A. C. Loheac and J. E. Drut, "Third-order perturbative lattice and complex Langevin analyses of the finite-temperature equation of state of nonrelativistic fermions in one dimension," Phys. Rev. **D95**, 094502 (2017), arXiv:1702.04666 .

[116] A. C. Loheac, J. Braun, and J. E. Drut, "Polarized fermions in one dimension: density and polarization from complex Langevin calculations, perturbation theory, and the virial expansion," Phys. Rev. **D98**, 054507 (2018), arXiv:1804.10257 .

[117] L. Rammelmüller, W. J. Porter, J. E. Drut, and J. Braun, "Surmounting the sign problem in non-relativistic calculations: a case study with mass-imbalanced fermions," Phys. Rev. **D96**, 094506 (2017), arXiv:1708.03149 .

[118] K. Nagata, J. Nishimura, and S. Shimasaki, "Complex Langevin calculations in finite density QCD at large μ/T with the deformation technique," Phys. Rev. **D98**, 114513 (2018), arXiv:1805.03964 .

[119] J. B. Kogut and D. K. Sinclair, "Applying Complex Langevin Simulations to Lattice QCD at Finite Density," (2019), arXiv:1903.02622 .

[120] G. Aarts, F. A. James, E. Seiler, and I.-O. Stamatescu, "Complex Langevin: Etiology and Diagnostics of its Main Problem," Eur. Phys. J. **C71**, 1756 (2011), arXiv:1101.3270 .

[121] G. Aarts, E. Seiler, and I.-O. Stamatescu, "The Complex Langevin method: When can it be trusted?" Phys. Rev. **D81**, 054508 (2010), arXiv:0912.3360 .

[122] K. Nagata, J. Nishimura, and S. Shimasaki, "Argument for justification of the complex Langevin method and the condition for correct convergence," Phys. Rev. **D94**, 114515 (2016), arXiv:1606.07627 .

[123] K. Nagata, J. Nishimura, and S. Shimasaki, "Justification of the complex Langevin method with the gauge cooling procedure," PTEP **2016**, 013B01 (2016), arXiv:1508.02377 .

[124] G. Aarts and F. A. James, "On the convergence of complex Langevin dynamics: The Three-dimensional XY model at finite chemical potential," JHEP **08**, 020 (2010), arXiv:1005.3468 .

[125] Z. Fodor, S. D. Katz, D. Sexty, and C. Török, "Complex Langevin dynamics for dynamical QCD at nonzero chemical potential: A comparison with multiparameter reweighting," Phys. Rev. **D92**, 094516 (2015), arXiv:1508.05260 .

[126] J. Bloch and O. Schenk, "Selected inversion as key to a stable Langevin evolution across the QCD phase boundary," *Proceedings, 35th International Symposium on Lattice Field Theory (Lattice 2017): Granada, Spain, June 18-24, 2017*, EPJ Web Conf. **175**, 07003 (2018), arXiv:1707.08874 .

[127] D. K. Sinclair and J. B. Kogut, "Complex Langevin Simulations of QCD at Finite Density – Progress Report," *Proceedings, 35th International Symposium*

on Lattice Field Theory (Lattice 2017): Granada, Spain, June 18-24, 2017, EPJ Web Conf. **175**, 07031 (2018), arXiv:1710.08465 .

[128] F. Attanasio and B. Jäger, "Dynamical stabilisation of complex Langevin simulations of QCD," Eur. Phys. J. **C79**, 16 (2019), arXiv:1808.04400 .

[129] G. Parisi and Y.-s. Wu, "Perturbation Theory Without Gauge Fixing," Sci. Sin. **24**, 483 (1981).

[130] P. H. Damgaard and H. Hüffel, "Stochastic Quantization," Phys. Rept. **152**, 227 (1987).

[131] M. Namiki, I. Ohba, K. Okano, Y. Yamanaka, A. K. Kapoor, H. Nakazato, and S. Tanaka, "Stochastic quantization," Lect. Notes Phys. Monogr. **9**, 1–217 (1992).

[132] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, "Langevin Simulations of Lattice Field Theories," Phys. Rev. **D32**, 2736 (1985).

[133] A. Ukawa and M. Fukugita, "Langevin simulation including dynamical quark loops," Phys. Rev. Lett. **55**, 1854 (1985).

[134] G. Aarts, F. A. James, E. Seiler, and I.-O. Stamatescu, "Adaptive stepsize and instabilities in complex Langevin dynamics," Phys. Lett. **B687**, 154–159 (2010), arXiv:0912.0617 .

[135] E. Seiler, "Status of Complex Langevin," *Proceedings, 35th International Symposium on Lattice Field Theory (Lattice 2017): Granada, Spain, June 18-24, 2017*, EPJ Web Conf. **175**, 01019 (2018), arXiv:1708.08254 .

[136] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," Phys. Lett. **B195**, 216–222 (1987).

[137] G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Complex Langevin dynamics and zeroes of the fermion determinant," JHEP **05**, 044 (2017), [Erratum: JHEP01,128(2018)], arXiv:1701.02322 .

[138] M. Nagasawa, "Segregation of a population in an environment," Journal of Mathematical Biology **9**, 213–235 (1980).

[139] J. Flower, S. W. Otto, and S. Callahan, "Complex Langevin Equations and Lattice Gauge Theory," Phys. Rev. **D34**, 598 (1986).

[140] A. Mollgaard and K. Splittorff, "Full simulation of chiral random matrix theory at nonzero chemical potential by complex Langevin," Phys. Rev. **D91**, 036007 (2015), arXiv:1412.2729 .

[141] J. Greensite, "Comparison of complex Langevin and mean field methods applied to effective Polyakov line models," Phys. Rev. **D90**, 114507 (2014), arXiv:1406.4558 .

[142] K. Splittorff, "Dirac spectrum in complex Langevin simulations of QCD," Phys. Rev. **D91**, 034507 (2015), arXiv:1412.0502 .

[143] T. Ichihara, K. Nagata, and K. Kashiwa, "Test for a universal behavior of Dirac eigenvalues in the complex Langevin method," Phys. Rev. **D93**, 094511 (2016), arXiv:1603.09554 .

[144] J. Nishimura and S. Shimasaki, "New Insights into the Problem with a Singular Drift Term in the Complex Langevin Method," Phys. Rev. **D92**, 011501 (2015), arXiv:1504.08359 .

[145] Y. Ito and J. Nishimura, "The complex Langevin analysis of spontaneous symmetry breaking induced by complex fermion determinant," JHEP **12**, 009 (2016), arXiv:1609.04501 .

[146] C. Pehlevan and G. Guralnik, "Complex Langevin Equations and Schwinger-Dyson Equations," Nucl. Phys. **B811**, 519–536 (2009), arXiv:0710.3756 .

[147] L. L. Salcedo and E. Seiler, "Schwinger–Dyson equations and line integrals," J. Phys. **A52**, 035201 (2019), arXiv:1809.06888 .

[148] I.-O. Stamatescu, "unpublished notes (2007)," .

[149] L. L. Salcedo, "Does the complex Langevin method give unbiased results?" Phys. Rev. **D94**, 114505 (2016), arXiv:1611.06390 .

[150] R. De Pietri, A. Feo, E. Seiler, and I.-O. Stamatescu, "A Model for QCD at high density and large quark mass," Phys. Rev. **D76**, 114501 (2007), arXiv:0705.3420 .

[151] M. Fromm, J. Langelage, S. Lottini, and O. Philipsen, "The QCD deconfinement transition for heavy quarks and all baryon chemical potentials," JHEP **01**, 042 (2012), arXiv:1111.4953 .

[152] M. Fromm, J. Langelage, S. Lottini, M. Neuman, and O. Philipsen, "Onset Transition to Cold Nuclear Matter from Lattice QCD with Heavy Quarks," Phys. Rev. Lett. **110**, 122001 (2013), arXiv:1207.3005 .

[153] J. Langelage, M. Neuman, and O. Philipsen, "Heavy dense QCD and nuclear matter from an effective lattice theory," JHEP **09**, 131 (2014), arXiv:1403.4162 .

[154] T. C. Blum, J. E. Hetrick, and D. Toussaint, "High density QCD with static quarks," Phys. Rev. Lett. **76**, 1019–1022 (1996), arXiv:hep-lat/9509002 .

[155] G. Aarts, O. Kaczmarek, F. Karsch, and I.-O. Stamatescu, "1 / M correction to quenched QCD with nonzero baryon density," *Lattice field theory. Proceedings, 19th International Symposium, Lattice 2001, Berlin, Germany, August 19-24, 2001*, Nucl. Phys. Proc. Suppl. **106**, 456–458 (2002), arXiv:hep-lat/0110145 .

[156] I. Bender, T. Hashimoto, F. Karsch, V. Linke, A. Nakamura, M. Plewnia, I. O. Stamatescu, and W. Wetzel, "Full QCD and QED at finite temperature and chemical potential," *LATTICE 91: International Symposium on Lattice Field Theory Tsukuba, Japan, November 5-9, 1991*, Nucl. Phys. Proc. Suppl. **26**, 323–325 (1992).

[157] D. K. Sinclair and J. B. Kogut, "Complex Langevin for Lattice QCD at $T = 0$ and $\mu \geq 0$," *Proceedings, 34th International Symposium on Lattice Field Theory (Lattice 2016): Southampton, UK, July 24-30, 2016*, PoS **LATTICE2016**, 026 (2016), arXiv:1611.02312 .

[158] B. H. Wellegehausen, A. Maas, A. Wipf, and L. von Smekal, "Hadron masses and baryonic scales in G_2-QCD at finite density," Phys. Rev. **D89**, 056007 (2014), arXiv:1312.5579 .

[159] J. B. Kogut and D. K. Sinclair, "Lattice QCD at finite isospin density at zero and finite temperature," Phys. Rev. **D66**, 034505 (2002), arXiv:hep-lat/0202028 .

[160] J. B. Kogut and D. K. Sinclair, "Lattice QCD at finite temperature and density in the phase-quenched approximation," Phys. Rev. **D77**, 114503 (2008), arXiv:0712.2625 .

[161] B. B. Brandt, G. Endrodi, and S. Schmalzbauer, "QCD phase diagram for nonzero isospin-asymmetry," Phys. Rev. **D97**, 054514 (2018), arXiv:1712.08190 .

[162] F. Butler, H. Chen, J. Sexton, A. Vaccarino, and D. Weingarten, "Volume dependence of the valence Wilson fermion mass spectrum," *LATTICE 91: International Symposium on Lattice Field Theory Tsukuba, Japan, November 5-9, 1991*, Nucl. Phys. Proc. Suppl. **26**, 287–289 (1992).

[163] S. Kratochvila and P. de Forcrand, "The Canonical approach to finite density QCD," *Proceedings, 23rd International Symposium on Lattice field theory (Lattice 2005): Dublin, Ireland, Jul 25-30, 2005*, PoS **LAT2005**, 167 (2006), arXiv:hep-lat/0509143 .

[164] D. Kressner, *Numerical Methods for General and Structured Eigenvalue Problems*, Lecture Notes in Computational Science and Engineering (Springer-Verlag Berlin Heidelberg, 2005).

[165] H. T. Ding, P. Hegde, F. Karsch, A. Lahiri, S. T. Li, S. Mukherjee, and P. Petreczky, "Chiral phase transition of (2+1)-flavor QCD," *Proceedings, 27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2018): Venice, Italy, May 14-19, 2018*, Nucl. Phys. **A982**, 211–214 (2019), arXiv:1807.05727 .

[166] A. M. Polyakov, "Gauge Fields as Rings of Glue," Nucl. Phys. **B164**, 171–188 (1980).

[167] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay, "Monte carlo simulation in statistical physics," Computers in Physics **7**, 156–157 (1993).

[168] G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, "The QCD phase diagram at nonzero quark density," JHEP **04**, 001 (2011), arXiv:1102.1356 .

[169] A. S. Kronfeld, "Dynamics of Langevin simulations," Prog. Theor. Phys. Suppl. **111**, 293–312 (1993), arXiv:hep-lat/9205008 .

[170] M. Jarrell and J. E. Gubernatis, "Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data," Phys. Rept. **269**, 133–195 (1996).

[171] M. Asakawa, T. Hatsuda, and Y. Nakahara, "Maximum entropy analysis of the spectral functions in lattice QCD," Prog. Part. Nucl. Phys. **46**, 459–508 (2001), arXiv:hep-lat/0011040 .

[172] Y. Burnier and A. Rothkopf, "Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories," Phys. Rev. Lett. **111**, 182003 (2013), arXiv:1307.6106 .

[173] A. K. Cyrol, J. M. Pawlowski, A. Rothkopf, and N. Wink, "Reconstructing the gluon," SciPost Phys. **5**, 065 (2018), arXiv:1804.00945 .

[174] E. Witten, "Analytic Continuation Of Chern-Simons Theory," AMS/IP Stud. Adv. Math. **50**, 347–446 (2011), arXiv:1001.2933 .

[175] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato, "Monte Carlo simulations on the Lefschetz thimble: Taming the sign problem," Phys. Rev. **D88**, 051501 (2013), arXiv:1303.7204 .

[176] F. Di Renzo and G. Eruzzi, "Thimble regularization at work: from toy models to chiral random matrix theories," Phys. Rev. **D92**, 085030 (2015), arXiv:1507.03858 .

[177] F. Di Renzo and G. Eruzzi, "One-dimensional QCD in thimble regularization," Phys. Rev. **D97**, 014503 (2018), arXiv:1709.10468 .

[178] K. Zambello and F. Di Renzo, "Towards Lefschetz thimbles regularization of heavy-dense QCD," in *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018* (2018) arXiv:1811.03605 .

[179] Y. Tanizaki and M. Tachibana, "Multi-flavor massless QED_2 at finite densities via Lefschetz thimbles," JHEP **02**, 081 (2017), arXiv:1612.06529 .

[180] Y. Tanizaki, Y. Hidaka, and T. Hayata, "Lefschetz-thimble analysis of the sign problem in one-site fermion model," New J. Phys. **18**, 033002 (2016), arXiv:1509.07146 .

[181] K. Fukushima and Y. Tanizaki, "Hamilton dynamics for Lefschetz-thimble integration akin to the complex Langevin method," PTEP **2015**, 111A01 (2015), arXiv:1507.07351 .

[182] Y. Tanizaki, H. Nishimura, and K. Kashiwa, "Evading the sign problem in the mean-field approximation through Lefschetz-thimble path integral," Phys. Rev. **D91**, 101701 (2015), arXiv:1504.02979 .

[183] T. Kanazawa and Y. Tanizaki, "Structure of Lefschetz thimbles in simple fermionic systems," JHEP **03**, 044 (2015), arXiv:1412.2802 .

[184] H. Fujii, D. Honda, M. Kato, Y. Kikukawa, S. Komatsu, and T. Sano, "Hybrid Monte Carlo on Lefschetz thimbles - A study of the residual sign problem," JHEP **10**, 147 (2013), arXiv:1309.4371 .

[185] M. Cristoforetti, F. Di Renzo, G. Eruzzi, A. Mukherjee, C. Schmidt, L. Scorzato, and C. Torrero, "An efficient method to compute the residual phase on a Lefschetz thimble," Phys. Rev. **D89**, 114505 (2014), arXiv:1403.5637 .

[186] Y. Tanizaki, H. Nishimura, and J. J. M. Verbaarschot, "Gradient flows without blow-up for Lefschetz thimbles," JHEP **10**, 100 (2017), arXiv:1706.03822 .

[187] A. Alexandru, G. Basar, and P. Bedaque, "Monte Carlo algorithm for simulating fermions on Lefschetz thimbles," Phys. Rev. **D93**, 014504 (2016), arXiv:1510.03258 .

[188] J. Nishimura and S. Shimasaki, "Combining the complex Langevin method and the generalized Lefschetz-thimble method," JHEP **06**, 023 (2017), arXiv:1703.09409 .

[189] Y. Mori, K. Kashiwa, and A. Ohnishi, "Toward solving the sign problem with path optimization method," Phys. Rev. **D96**, 111501 (2017), arXiv:1705.05605 .

[190] C. Schmidt and F. Ziesché, "Simulating low dimensional QCD with Lefschetz thimbles," *Proceedings, 34th International Symposium on Lattice Field Theory (Lattice 2016): Southampton, UK, July 24-30, 2016*, PoS **LATTICE2016**, 076 (2017), arXiv:1701.08959 .

[191] A. Alexandru, G. Basar, P. F. Bedaque, S. Vartak, and N. C. Warrington, "Monte Carlo Study of Real Time Dynamics on the Lattice," Phys. Rev. Lett. **117**, 081602 (2016), arXiv:1605.08040 .

[192] A. Alexandru, G. Basar, P. F. Bedaque, and G. W. Ridgway, "Schwinger-Keldysh formalism on the lattice: A faster algorithm and its application to field theory," Phys. Rev. **D95**, 114501 (2017), arXiv:1704.06404 .

[193] A. Alexandru, G. Basar, P. Bedaque, G. W. Ridgway, and N. C. Warrington, "Study of symmetry breaking in a relativistic Bose gas using the contraction algorithm," Phys. Rev. **D94**, 045017 (2016), arXiv:1606.02742 .

[194] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, "Monte Carlo calculations of the finite density Thirring model," Phys. Rev. **D95**, 014502 (2017), arXiv:1609.01730 .

[195] A. Alexandru, G. Başar, P. F. Bedaque, H. Lamm, and S. Lawrence, "Finite Density QED_{1+1} Near Lefschetz Thimbles," Phys. Rev. **D98**, 034506 (2018), arXiv:1807.02027 .

[196] M. Fukuma and N. Umeda, "Parallel tempering algorithm for integration over Lefschetz thimbles," PTEP **2017**, 073B01 (2017), arXiv:1703.00861 .

[197] A. Alexandru, G. Basar, P. F. Bedaque, G. W. Ridgway, and N. C. Warrington, "Fast estimator of Jacobians in the Monte Carlo integration on Lefschetz thimbles," Phys. Rev. **D93**, 094514 (2016), arXiv:1604.00956 .

[198] A. Alexandru, G. Basar, P. F. Bedaque, and N. C. Warrington, "Tempered transitions between thimbles," Phys. Rev. **D96**, 034513 (2017), arXiv:1703.02414 .

[199] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, "Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds," Phys. Rev. **D97**, 094510 (2018), arXiv:1804.00697 .

[200] A. Alexandru, P. F. Bedaque, H. Lamm, and S. Lawrence, "Deep Learning Beyond Lefschetz Thimbles," Phys. Rev. **D96**, 094505 (2017), arXiv:1709.01971 .

[201] K. Kashiwa, Y. Mori, and A. Ohnishi, "Application of path optimization method to the model sign problem in QCD effective model with repulsive vector-type interaction," (2019), arXiv:1903.03679 .

[202] K. Kashiwa, Y. Mori, and A. Ohnishi, "Controlling the model sign problem via the path optimization method: Monte Carlo approach to a QCD effective model with Polyakov loop," Phys. Rev. **D99**, 014033 (2019), arXiv:1805.08940 .

[203] F. Pham, "Vanishing homologies and the n variable saddlepoint method," Proc.Symp.Pure Math **40**, 319 (1983).

[204] G. Aarts, L. Bongiovanni, E. Seiler, and D. Sexty, "Some remarks on Lefschetz thimbles and complex Langevin dynamics," JHEP **10**, 159 (2014), arXiv:1407.2090 .

[205] G. Aarts, F. A. James, J. M. Pawlowski, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Stability of complex Langevin dynamics in effective models," JHEP **03**, 073 (2013), arXiv:1212.5231 .

[206] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012) pp. 1097–1105.

[207] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," International Journal of Computer Vision (IJCV) **115**, 211–252 (2015), arXiv:1409.0575 .

[208] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018) arXiv:1709.01507 .

[209] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine **29**, 82–97 (2012).

[210] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," Nature **529**, 484–489 (2016).

[211] N. Strodthoff and C. Strodthoff, "Detecting and interpreting myocardial infarction using fully convolutional neural networks," Physiological Measurement **40**, 015001 (2019), arXiv:1806.07385 .

[212] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature **542**, 115– (2017).

[213] P. Baldi, P. Sadowski, and D. Whiteson, "Enhanced Higgs Boson to $\tau^+\tau^-$ Search with Deep Learning," Phys. Rev. Lett. **114**, 111801 (2015), arXiv:1410.3469 .

[214] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks," Phys. Rev. **D94**, 112002 (2016), arXiv:1607.08633 .

[215] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, "Deep-learning Top Taggers or The End of QCD?" JHEP **05**, 006 (2017), arXiv:1701.08784 .

[216] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," Science **355**, 602–606 (2017), arXiv:1606.02318 .

[217] X. Gao and L.-M. Duan, "Efficient representation of quantum many-body states with deep neural networks," Nature Communications **8**, 662 (2017), arXiv:1701.05039 .

[218] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, "Neural-network quantum state tomography," Nature Physics **14**, 447–450 (2018), arXiv:1703.05334 .

[219] G. Torlai and R. G. Melko, "Learning thermodynamics with boltzmann machines," Phys. Rev. B **94**, 165134 (2016), arXiv:1606.02718 .

[220] K.-I. Aoki and T. Kobayashi, "Restricted boltzmann machines for the long range ising models," Modern Physics Letters B **30**, 1650401 (2016), arXiv:1701.00246 .

[221] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, "Reconstructing parton distribution functions from Ioffe time data: from Bayesian methods to Neural Networks," (2019), arXiv:1901.05408 .

[222] R. Fournier, L. Wang, O. V. Yazyev, and Q. Wu, "An Artificial Neural Network Approach to the Analytic Continuation Problem," (2018), arXiv:1810.00913 .

[223] H. Yoon, J.-H. Sim, and M. J. Han, "Analytic continuation via domain knowledge free machine learning," Phys. Rev. B **98**, 245101 (2018), arXiv:1806.03841 .

[224] J. Carrasquilla and R. G. Melko, "Machine learning phases of matter," Nature Physics **13** (2017), arXiv:1605.01735 .

[225] L. Li, T. E. Baker, S. R. White, and K. Burke, "Pure density functional for strong correlation and the thermodynamic limit from machine learning," Phys. Rev. B **94**, 245129 (2016), arXiv:1609.03705 .

[226] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, "Machine learning phases of strongly correlated fermions," Phys. Rev. X **7**, 031038 (2017), arXiv:1609.02552 .

[227] Y. Zhang and E.-A. Kim, "Quantum loop topography for machine learning," Phys. Rev. Lett. **118**, 216401 (2017), arXiv:1611.01518 .

[228] F. Schindler, N. Regnault, and T. Neupert, "Probing many-body localization with neural networks," Phys. Rev. B **95**, 245134 (2017), arXiv:1704.01578 .

[229] L. Wang, "Discovering phase transitions with unsupervised learning," Phys. Rev. B **94**, 195105 (2016), arXiv:1606.00318 .

[230] S. J. Wetzel, "Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders," Phys. Rev. E **96**, 022140 (2017), arXiv:1703.02435 .

[231] E. L. van Nieuwenburg, Y.-H. Liu, and S. Huber, "Learning phase transitions by confusion," Nature Physics **13**, 435 EP – (2017), arXiv:1610.02048 .

[232] W. Hu, R. R. P. Singh, and R. T. Scalettar, "Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination," Phys. Rev. E **95**, 062122 (2017), arXiv:1704.00080 .

[233] C. Casert, T. Vieijra, J. Nys, and J. Ryckebusch, "Interpretable machine learning for inferring the phase boundaries in a nonequilibrium system," Phys. Rev. E **99**, 023304 (2019), arXiv:1807.02 .

[234] M. Matty, Y. Zhang, Z. Papic, and E.-A. Kim, "Multi-faceted machine learning of competing orders in disordered interacting systems," (2019), arXiv:1902.04079 .

[235] J. Greitemann, K. Liu, and L. Pollet, "Probing hidden spin order with interpretable machine learning," Phys. Rev. B **99**, 060404 (2019), arXiv:1804.08557 .

[236] P. Ponte and R. G. Melko, "Kernel methods for interpretable machine learning of order parameters," Phys. Rev. B **96**, 205146 (2017), arXiv:1704.05848 .

[237] D. Kim and D.-H. Kim, "Smallest neural network to learn the ising criticality," Phys. Rev. E **98**, 022138 (2018), arXiv:1804.02171 .

[238] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, "Self-learning monte carlo method," Phys. Rev. B **95**, 041101 (2017), arXiv:1610.03137 .

[239] H. Shen, J. Liu, and L. Fu, "Self-learning monte carlo with deep neural networks," Phys. Rev. B **97**, 205140 (2018), arXiv:1801.01127 .

[240] J. M. Urban and J. M. Pawlowski, "Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks," (2018), arXiv:1811.03533 .

[241] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, "Regressive and generative neural networks for scalar field theory," (2018), arXiv:1810.12879 .

[242] L. Huang and L. Wang, "Accelerated monte carlo simulations with restricted boltzmann machines," Phys. Rev. B **95**, 035105 (2017), arXiv:1610.02746 .

[243] N. Portman and I. Tamblyn, "Sampling algorithms for validation of supervised learning models for ising-like systems," Journal of Computational Physics **350**, 871 – 890 (2017), arXiv:1611.05891 .

[244] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, "Machine learning quantum phases of matter beyond the fermion sign problem," Scientific Reports **7**, 8823 (2017), arXiv:1608.07848 .

[245] B. Yoon, T. Bhattacharya, and R. Gupta, "Machine Learning Estimators for Lattice QCD Observables," (2018), arXiv:1807.05971 .

[246] P. E. Shanahan, D. Trewartha, and W. Detmold, "Machine learning action parameters in lattice quantum chromodynamics," Phys. Rev. **D97**, 094506 (2018), arXiv:1801.05784 .

[247] M. A. Nielsen, "Neural networks and deep learning," (2018).

[248] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016) `http://www.deeplearningbook.org`.

[249] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of Control, Signals and Systems **2**, 303–314 (1989).

[250] K. Hornik, "Approximation capabilities of multilayer feedforward networks," Neural Networks **4**, 251 – 257 (1991).

[251] L. Onsager, "Crystal statistics. i. a two-dimensional model with an order-disorder transition," Phys. Rev. **65**, 117–149 (1944).

[252] K. Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge University Press, 2010).

[253] M. Fukugita, Y. Oyanagi, and A. Ukawa, "Langevin Simulation of the Full QCD Hadron Mass Spectrum on a Lattice," Phys. Rev. **D36**, 824 (1987).

[254] L. Bongiovanni, G. Aarts, E. Seiler, D. Sexty, and I.-O. Stamatescu, "Adaptive gauge cooling for complex Langevin dynamics," *Proceedings, 31st International Symposium on Lattice Field Theory (Lattice 2013): Mainz, Germany, July 29-August 3, 2013*, PoS **LATTICE2013**, 449 (2014), arXiv:1311.1056 .

[255] M. Schlosser, *Langevin Dynamics on Lefschetz thimbles* (Bachelor thesis, 2017).

[256] S. Syrkowski, *Lefschetz cooled Stochastic Quantization* (Bachelor thesis, 2017).

[257] F. Chollet *et al.*, "Keras," `https://keras.io` (2015).

[258] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research **12**, 2825–2830 (2011).