Pion-Pion Scattering in the Non-Linear Sigma Model

Capstone Project Report

Ruaidhrí Campion
1933350
Theoretical Physics

Supervised by Prof. Mike Peardon

12th March, 2023

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Acknowledgements

Thanks to my supervisor, Professor Mike Peardon, for overseeing this project and consistently pointing me in the right direction whenever I got stuck.

Thanks to my parents and grandfather for coming to visit me and paying for dinner.

Thanks to the School of Maths library for being a home away from my home away from home throughout the year.

Thanks to my friends and classmates for helping me make the most out of the last four years.

Plagiarism Declaration

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at `http://www.tcd.ie/calendar`.

I have completed the online tutorial in avoiding plagiarism 'Ready, Steady, Write', located at `http://tcdie.libguides.com/plagiarism/readysteadywrite`.

Abstract

This project introduces the basic yet fundamental concepts behind studying quantum field theories on lattice models using numerical algorithms, and focuses on the calculation of the two-point correlation function for Klein-Gordon theory and the $SU(2)$ non-linear sigma model.

Klein-Gordon fields are explored with the intent of debugging and optimising computer programmes written to simulate fields on a lattice, by comparing two approaches to writing computer programmes in the context of runtime. It is deduced that the Gibbs sampler is preferable to the Metropolis-Hastings algorithm due to its comparable calculation speeds, higher accuracies, smaller intergated correlation times, and lack of fine-tuning of an external parameter. The Klein-Gordon study is concluded by determining relationships between the integrated correlation time and the parameters of the system.

A brief analysis of the $SU(2)$ non-linear sigma model Gibbs sampler is made to ensure efficient and accurate calculations, and the mass parameters of the model in two and three dimensions are determined from calculations of the two-point correlation for a small range of system parameters. A more detailed study of the model is carried out in three dimensions; the bare mass of the pion field is verified in the low-temperature limit, and the temperature range for a theoretical particle decay is determined. A measure of the efficiency of the Gibbs sampler for the model was obtained through a brief investigation of the accuracy and runtime for a variety of system parameters.

Conventions

This project uses natural units $c = \hbar = k_B = 1$.

Contents

1 Lattice Field Theory

1.1 Klein-Gordon Theory

The action of a real non-interacting scalar boson of mass m in d-dimensional Euclidean space is given by the Klein-Gordon action [1, 2]

$$\begin{aligned} S(\phi) &= \frac{1}{2}\int d^d x\, \phi(x)\left(-\Box + m^2\right)\phi(x), \\ \Box &\equiv \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}. \end{aligned} \tag{1}$$

Under the lattice parameterisation

$$x_i \equiv a\, n_i \tag{2}$$

for a regular d-dimensional spacetime lattice of size $L_1 \times \cdots \times L_d$ with grid spacing a, site vectors $\mathbf{e}_i$, and periodic boundary conditions, this action can be represented as a function of the fields $\phi(\mathbf{x})$ at each lattice site $\mathbf{x} = \sum_{i=1}^{d} x_i \mathbf{e}_i = (x_1, \ldots, x_d)$, and is given by (A.1)

$$\begin{aligned} S(\phi) &= \sum_{\mathbf{x}} \phi(\mathbf{x}) \left(\frac{2d + (am)^2}{2}\phi(\mathbf{x}) - \sum_{i=1}^{d} \phi(\mathbf{x} + \mathbf{e}_i) \right) \\ &= \sum_{\mathbf{x}} \frac{\phi(\mathbf{x})}{2} \left(\frac{\phi(\mathbf{x})}{\kappa^2} - \gamma(\mathbf{x}) \right), \end{aligned} \tag{3}$$

where

$$\begin{aligned} \kappa &\equiv \left(2d + (am)^2\right)^{-\frac{1}{2}}, \\ \gamma(\mathbf{x}) &\equiv \sum_{i=1}^{d} \left(\phi(\mathbf{x} - \mathbf{e}_i) + \phi(\mathbf{x} + \mathbf{e}_i)\right). \end{aligned} \tag{4}$$

1.2 Non-Linear Sigma Model

The non-linear sigma model [3] describes a scalar field $\boldsymbol{\phi} \equiv (\sigma, \boldsymbol{\pi})$ taking values in the manifold $\mathcal{M} = S^{N-1}$, i.e. satisfying the condition

$$\boldsymbol{\phi} \cdot \boldsymbol{\phi} = \sigma^2 + \boldsymbol{\pi} \cdot \boldsymbol{\pi} = 1, \tag{5}$$

with $\boldsymbol{\pi}$ an $(N-1)$-dimensional vector. This project focuses on the $N = 4$ case, i.e. $\mathcal{M} = S^3 \cong SU(2)$, known as the $SU(2)$ non-linear sigma model or the chiral model [4]. The action of the model is given by

$$S(\boldsymbol{\phi}) = \frac{\beta}{2}\int d^d x\ \mathrm{Tr}\left[\partial_\mu \boldsymbol{\phi}^\dagger(x)\partial^\mu \boldsymbol{\phi}(x) - \lambda_0 \boldsymbol{\phi}(x)\right], \tag{6}$$

where $\beta \equiv \frac{1}{T}$ is the inverse temperature, and λ_0 is a free parameter.

In general, elements u of $SU(2)$ are 2×2 unitary matrices ($u^\dagger u = u u^\dagger = \mathbb{I}$) with unit determinant, and can be represented as a 4-tuple of real numbers $\{u_0, u_1, u_2, u_3\}$ with

$$\begin{aligned} u = u_0 \mathbb{I} + i\sum_{k=1}^{3} u_k \tau_k &= \begin{pmatrix} u_0 + iu_3 & u_2 + iu_1 \\ -u_2 + iu_1 & u_0 - iu_3 \end{pmatrix} \in SU(2) \\ \implies u_0 u_0 + \sum_{k=1}^{3} u_k u_k = 1, &\qquad \mathrm{Tr}\, u = 2u_0, \end{aligned} \tag{7}$$

where the Pauli matrices τ_k are defined as

$$\tau_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \tau_2 \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \tau_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{8}$$

Therefore, with $u = \boldsymbol{\phi}$, $u_0 = \sigma$, and $(u_1, u_2, u_3) = \boldsymbol{\pi}$, the condition given in Equation 5 is met.

If $\lambda_0 = 0$, then the action (Equation 6) is invariant under an $SU(2)_L \times SU(2)_R$ chiral symmetry, i.e. under $\boldsymbol{\phi} \to U_L \boldsymbol{\phi} U_R^\dagger$; if $\lambda_0 \neq 0$, then the action is invariant under an $SU(2)$ isospin symmetry, i.e $U_L = U_R$ (A.2). Three pseudo-Goldstone bosons [5] of mass $m^2 = \beta\lambda_0$ appear when this chiral symmetry is explicitly broken, and correspond to the pion fields $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$ (A.2).

Under the same lattice parameterisation as before (Equation 2), Equation 6 can be written as (A.3)

$$\begin{aligned} S(\boldsymbol{\phi}) &= -\frac{\beta}{2} \sum_{\mathbf{x}} \mathrm{Tr}\left[\boldsymbol{\phi}(\mathbf{x}) \left(\lambda_0 \mathbb{I} + \sum_{i=1}^{d} \boldsymbol{\phi}^\dagger(\mathbf{x} + \mathbf{e}_i) \right) \right] \\ &= -\frac{\beta}{4} \sum_{\mathbf{x}} \mathrm{Tr}\Big(\boldsymbol{\phi}(\mathbf{x}) \boldsymbol{\Sigma}^\dagger(\mathbf{x}) \Big), \end{aligned} \tag{9}$$

where

$$\boldsymbol{\Sigma}(\mathbf{x}) \equiv \lambda_0 \mathbb{I} + \sum_{i=1}^{d} \left(\boldsymbol{\phi}(\mathbf{x} - \mathbf{e}_i) + \boldsymbol{\phi}(\mathbf{x} + \mathbf{e}_i) \right), \tag{10}$$

and the corresponding pion mass is now given by $(am)^2 = \lambda_0$.

1.3 Two-Point Correlation Function

The functions

$$\begin{aligned} \Phi(t) &\equiv \sum_{i=1}^{d-1} \sum_{x_i=1}^{L_i} \phi((x_1, \ldots, \ldots, x_{d-1}, t)), \\ c(\delta) &\equiv \sum_{t=1}^{T} \Phi(t)\, \Phi(t+\delta) \end{aligned} \tag{11}$$

can be defined (where $T \equiv L_d$), where $\Phi(t)$ corresponds to the t-th position in the 1-dimensional array resulting from summing over all spatial coordinates, and $c(\delta)$ represents the correlation between sites separated by a distance δ in the same (periodic) 1-dimensional array. $c(\delta)$ is called the two-point correlation/Green's function [6, 7], and is analogous to the probability amplitude for propagation over a time δ.

The Klein-Gordon two-point correlation function follows from Equation 11. For a non-periodic lattice, it is approximately proportional to $e^{-am\,\delta}$ [8]. Naturally, for a periodic lattice, it is then given by

$$c(\delta) \propto e^{-am\,\delta} + e^{-am\,(T-\delta)}. \tag{12}$$

For the non-linear sigma model, ϕ can be replaced by σ or π_i, $i = 1, 2, 3$ to obtain the two-point σ- or π_i-correlation functions $c_\sigma(\delta)$ or $c_{\pi_i}(\delta)$. The presence of the symmetry-breaking term λ_0 induces a preferred alignment of the field (similar to an external field in a statistical mechanics system, e.g. a magnetic field in an Ising model), which in turn imposes a "minimum correlation" between sites; in other words, the two-point correlation function is instead given by

$$c(\delta) \propto e^{-am\,\delta} + e^{-am\,(T-\delta)} + k, \tag{13}$$

for some constant k. As opposed to the Klein-Gordon action (Equation 3), the dimensionless mass term am does not explicitly appear in the non-linear sigma model action (Equation 9),

but rather implicitly through the inverse temperature β. By approximating the two-point correlation function as

$$c(\delta) \approx k_1 e^{-am\,\delta} + k_2 \tag{14}$$

for small δ, i.e. neglecting the contribution of the rising exponential term, the implicit mass am can be approximated as (B)

$$am \approx \ln\left\{ \frac{c(0) - c(2) \pm \sqrt{[c(0) - c(2)]^2 - 4\,[c(1) - c(2)]\,[c(0) - c(1)]}}{2\,[c(1) - c(2)]} \right\}. \tag{15}$$

2 Observable Calculation

This project deals with the calculation of observables of a field theory via simulation over a lattice. Before describing how one would go about simulating a statistical physics model, it is first important to define some fundamental concepts.

2.1 Statistical Background

For identically distributed random variables $A_1, \ldots, A_N$ with mean $\mu \equiv \mathrm{E}[A_i]$ and variance $\varsigma \equiv \mathrm{Var}(A_i)$, the sample mean $\hat{\mu}$ and sample variance $\hat{\varsigma}$ are defined as [9]

$$\begin{aligned} \hat{\mu} &\equiv \frac{1}{N} \sum_{i=1}^{N} A_i, \\ \hat{\varsigma} &\equiv \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \hat{\mu})^2 . \end{aligned} \tag{16}$$

The sample mean is an unbiased estimator of the mean, i.e. $\mathrm{E}[\hat{\mu}] = \mu$ (Lemma 1), and if the random variables are independent, then the sample variance is an unbiased estimator of the variance, i.e. $\mathrm{E}[\hat{\varsigma}] = \varsigma$ (Lemma 3).

The mean squared error $\mathrm{MSE}(\hat{\eta})$ of an estimator $\hat{\eta}$ of a random variable η is defined as

$$\mathrm{MSE}(\hat{\eta}) \equiv \mathrm{E}\Big[(\hat{\eta} - \eta)^2\Big]. \tag{17}$$

If $\hat{\eta}$ is an unbiased estimator, then $\mathrm{MSE}(\hat{\eta}) = \mathrm{Var}(\hat{\eta})$. For independent identically distributed random variables $A_1, \ldots, A_N$ with mean μ and variance ς, then the corresponding mean squared errors are given by (Lemma 2, [10])

$$\begin{aligned} \mathrm{MSE}(\hat{\mu}) &= \frac{\varsigma}{N}, \\ \mathrm{MSE}(\hat{\varsigma}) &= \frac{1}{N} \left(\mathrm{E}\Big[(A_i - \mu)^4\Big] - \frac{\varsigma^2 (N-3)}{N-1} \right). \end{aligned} \tag{18}$$

These have corresponding "naïve" estimators

$$\begin{aligned} \widehat{\mathrm{MSE}}_{\text{naïve}}(\hat{\mu}) &\equiv \frac{\hat{\varsigma}}{N}, \\ \widehat{\mathrm{MSE}}_{\text{naïve}}(\hat{\varsigma}) &\equiv \frac{1}{N} \left(\frac{1}{N} \sum_{i=1}^{N} (A_i - \hat{\mu})^4 - \frac{\hat{\varsigma}^2 (N-3)}{N-1} \right), \end{aligned} \tag{19}$$

i.e. simply replacing μ and ς in Equation 18 with $\hat{\mu}$ and $\hat{\varsigma}$. As is discussed in 2.3, these are only good estimators in the case of independent variables, and some adjustments need to be made for dependent variables.

2.2 Markov Chain Monte Carlo Methods

In short, Monte Carlo methods are numerical algorithms involving random sampling used to calculate expected values. For identically distributed random variables $A_1, A_2, \ldots$ with mean μ, then by the strong law of large numbers,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} A_i = \mu. \tag{20}$$

Thus, for a large enough N, the mean μ can be approximated by the sample mean $\hat{\mu}$ (Equation 16).

A Markov chain is a sequence of "memoryless" states, i.e. states which are generated based on the previous state alone. More formally, for a Markov chain of random variables $A_1, A_2, \ldots$, then

$$\begin{aligned} P\{A_{N+1} = a_i \,|\, A_N = a_j, A_{N-1} = a_{j_{N-1}}, \ldots, A_0 = a_{j_0}\} &= P\{A_{N+1} = a_i \,|\, A_N = a_j\} \\ &= P\{A_{M+1} = a_i \,|\, A_M = a_j\} \,\forall\, N, M. \end{aligned} \tag{21}$$

For a probability density function $\lambda(\xi)$ that is to be sampled from using a Markov chain, define the transition probability density function

$$p(\xi, \nu) \equiv P\{A_{N+1} = \xi \,|\, A_N = \nu\} \tag{22}$$

and the infinitesimal volume element $d\Gamma_\xi$ in the phase space centred around state ξ. $\lambda(\xi)$ and $p(\xi, \nu)$ are probability density functions, and so

$$\lambda(\xi) \geqslant 0, \qquad \int d\Gamma_\xi \, \lambda(\xi) = 1, \qquad p(\xi, \nu) \geqslant 0, \qquad \int d\Gamma_\xi \, p(\xi, \nu) = 1. \tag{23}$$

It is said that $p(\xi, \nu)$ satisfies detailed balance/time reversibility for $\lambda(\xi)$ if

$$p(\xi, \nu)\lambda(\nu) = p(\nu, \xi)\lambda(\xi) \,\forall\, \nu, \xi, \tag{24}$$

and if this is the case, then $\lambda(\xi)$ is the equilibrium distribution of $p(\xi, \nu)$, i.e. choosing $p(\xi, \nu)$ as the transition distribution leads to a Markov chain of values sampled from $\lambda(\xi)$ (Lemma 4).

Markov chains of random variables sampled using Monte Carlo methods are called Markov chain Monte Carlo (MCMC) methods. Typically when modelling a system, there exists a function of the system parameters that is minimised for a likely arrangement of the system, and maximised for an unlikely arrangement of the system; in the case of lattice field theories, this function is the action $S(\phi)$ of the field ϕ. Expectation values of observables of the field are given by

$$\langle A \rangle = \int \prod_{\mathbf{x}} d\phi(\mathbf{x}) \, A(\phi) \, \frac{e^{-S(\phi)}}{Z}, \tag{25}$$

where Z is a normalising factor. This gives rise to a natural choice of probability distribution

$$\lambda(\phi) = \frac{e^{-S(\phi)}}{Z} \tag{26}$$

from which states of the system can be sampled using MCMC methods, and corresponding observables can be approximated.

It is often the case where Equation 25 and Equation 26 are very difficult, or perhaps impossible, to solve analytically for complex systems. However, certain MCMC methods (such as those used in this project, and outlined below) allow for sampling without explicitly calculating these.

2.2.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm [11, 12] is perhaps the most well-known MCMC method to this day. It operates using an "accept/reject" process, outlined as follows [13].

Say $\lambda(\xi)$ is a probability density function from which values are to be sampled. Let the proposal density function $q(\xi, \nu)$ be a distribution corresponding to proposing a state ξ given the current state ν, and the acceptance probability $\alpha(\xi, \nu)$ be the probability that this proposed state is accepted as the next state, given by

$$\alpha(\xi, \nu) = \min\left(\frac{q(\nu, \xi)\lambda(\xi)}{q(\xi, \nu)\lambda(\nu)}, 1\right). \tag{27}$$

These expressions satisfy

$$\alpha(\xi, \nu)q(\xi, \nu)\lambda(\nu) = \alpha(\nu, \xi)q(\nu, \xi)\lambda(\xi), \tag{28}$$

and so if the transition probability density function is defined as

$$p(\xi, \nu) \equiv \alpha(\xi, \nu)q(\xi, \nu), \tag{29}$$

then $p(\xi, \nu)$ satisfies detailed balance for $\lambda(\xi)$.

The Metropolis-Hastings algorithm can easily be used to simulate a field theory over a lattice. For $\lambda(\phi)$ defined as in Equation 26, and the proposal density function chosen such that $q(\phi', \phi) = q(\phi, \phi')$ (e.g. symmetric distribution about the current state), then

$$\alpha(\phi', \phi) = \min\left(\frac{q(\phi', \phi)\lambda(\phi')}{q(\phi', \phi)\lambda(\phi)}, 1\right) = \min\left(e^{-\Delta S(\phi', \phi)}, 1\right), \tag{30}$$

and so only the difference $\Delta S(\phi', \phi) \equiv S(\phi') - S(\phi)$ between the proposed and current action is necessary to calculate for producing the next state. Notably, if the proposed state results in a lower action, then it is always accepted.

In most cases, the proposed field ϕ' differs to the current field ϕ at a single lattice site $\mathbf{x}$, i.e.

$$\phi' = \phi|_{\phi(\mathbf{x}) \to \phi'(\mathbf{x})}, \tag{31}$$

and the algorithm is repeated for each site in turn.

While any choice of $q(\xi, \nu)$ will eventually result in the Markov chain sampling from $\lambda(\xi)$, some intuition is still necessary; a good choice will result in proposed states being accepted sufficiently frequently such that the algorithm samples from the whole distribution of states, but also sufficiently infrequently such that likely states are sampled more often than less likely states.

2.2.2 Gibbs Sampler

The Gibbs sampler method [14] is a type of Metropolis-Hastings algorithm, where the proposed state is generated such that it is always accepted. In particular, the proposal density function is given by [13]

$$q(\xi, \nu) \propto \lambda(\xi), \tag{32}$$

and so

$$\alpha(\xi, \nu) = \min\left(\frac{\lambda(\nu)\lambda(\xi)}{\lambda(\xi)\lambda(\nu)}, 1\right) = 1. \tag{33}$$

2.3 Autocorrelation

By definition, any data produced using MCMC methods is correlated. As a result, when calculating any observables of a simulated system, it is important that autocorrelation is accounted for. Two notable consequences of autocorrelation are equilibrium convergence and effective independence.

2.3.1 Equilibrium Convergence

If the initial state of a system is chosen far away from its equilibrium state, then it may be the case that a calculation of a sample mean of an observable includes a disproportionate number of unlikely outliers. One simple remedy to this problem is to discard the first D states, and relabel $A_{D+1}, \ldots, A_N$ as $A_1, \ldots, A_{N-D}$ for calculations. A good choice of D requires some intuition, and can be found by noticing after how many iterations an observable sufficiently converges.

2.3.2 Effective Independence of Correlated Data

Equation 18 provides a notion of error in the calculation of the sample mean. However, this expression is only applicable in the case of independent sampling, and not for data obtained using MCMC methods. One method known as binning [15] can be implemented to approximate the true mean squared error of correlated data.

For identically distributed random variables $A_1, \ldots, A_N$ with mean μ and variance ς, define the sample mean $\hat{\mu}$ and sample variance $\hat{\varsigma}$ as in Equation 16. Introducing M bins of size $B = \frac{N}{M}$ produces a series of binned variables $A_1^{(B)}, \ldots, A_M^{(B)}$, with

$$A_i^{(B)} \equiv \frac{1}{B} \sum_{j=1}^{B} A_{j+(i-1)B} = \frac{1}{B} \sum_{j=(i-1)B+1}^{iB} A_j. \tag{34}$$

The corresponding binned sample variance estimator $\hat{\varsigma}^{(B)}$ and binned mean squared error estimator $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu})$ are defined as

$$\begin{aligned} \hat{\varsigma}^{(B)} &\equiv \frac{1}{M-1} \sum_{i=1}^{M} \left(A_i^{(B)} - \hat{\mu} \right)^2, \\ \widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu}) &\equiv \frac{\hat{\varsigma}^{(B)}}{M}. \end{aligned} \tag{35}$$

Consider the case where the random variables are not independent, such as those generated in a Markov chain. If B is chosen to be too small, then the binned sample variance $\hat{\varsigma}^{(B)}$ will not be an unbiased estimator of the variance ς, and so the binned mean squared error $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu})$ will not be an adequate estimator of the error. If B is chosen to be too large, then M will not be sufficiently large enough for $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu})$ to be a reliable estimator of the error. Thus a suitable B must be chosen such that there is a sufficient number of weakly correlated binned variables, i.e. that the binned mean squared error accurately represents the error of the sample mean $\hat{\mu}$. This can be done by plotting $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu})$ against B and noticing where the plot starts to level off, after it has increased from an underestimate (small B) and before it exhibits erratic behaviour (large B).

Defining the integrated correlation time $\tau_{\hat{\mu}}$ [16] introduces the notion of an effective number of independent variables of a Markov chain, i.e. $N_{\mathrm{eff},\hat{\mu}} \equiv \frac{N}{\tau_{\hat{\mu}}}$. The integrated correlation time can thus be thought of as the number of samples that need to be generated to effectively generate one independent sample. Defining the "naïve" mean squared error estimator as in Equation 19 allows an estimator $\hat{\tau}_{\hat{\mu}}^{(B)}$ for the integrated correlation time to be defined as

$$\hat{\tau}_{\hat{\mu}}^{(B)} \equiv \frac{\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu})}{\widehat{\mathrm{MSE}}_{\mathrm{naïve}}(\hat{\mu})}. \tag{36}$$

A suitable choice of B that results in $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{\mu}) \approx \mathrm{MSE}(\hat{\mu})$ will also result in $\hat{\tau}_{\hat{\mu}}^{(B)} \approx \tau_{\hat{\mu}}$. The integrated correlation time depends on both the observable being estimated and the sampling algorithm used.

2.4 Error Calculation

Equation 19 are sufficient estimators of the mean squared error of the sample mean/sample variance of independent identically distributed random variables. In the case of correlated random variables, the discarding and binning methods outlined in 2.3 can be used to obtain weakly correlated random variables, for which Equation 19 are again sufficient.

Now consider the case where a function $f(\mu)$ dependent on the mean μ of independent identically distributed random variables $A_1, \ldots, A_N$ is to be estimated. The estimator

$$\hat{f}' \equiv \frac{1}{N} \sum_{i=1}^{N} f(A_i) \tag{37}$$

does not converge to $f(\mu)$ for $N \to \infty$ [17]; instead, a more suitable estimator of $f(\mu)$ is given by

$$\hat{f} \equiv f(\hat{\mu})\,. \tag{38}$$

If N of these estimators $\hat{f}_i$ were independently generated, Equation 16 could be employed to estimate $f(\mu)$, with

$$A_i \to \hat{f}_i, \qquad\qquad \hat{\mu} \to \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i. \tag{39}$$

One method that achieves this is the jackknife approach [18, 19]. The jackknife estimator $\hat{f}_{\text{jack}}$ of $f(\mu)$, defined as

$$\begin{aligned} \hat{f}_{\text{jack}} &\equiv \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i, \\ \hat{f}_i &\equiv f(\hat{\mu}_i)\,, \\ \hat{\mu}_i &\equiv \frac{1}{N-1} \sum_{\substack{k=1 \\ k \neq i}}^{N} A_k, \end{aligned} \tag{40}$$

serves as an estimator for $f(\mu)$ with corresponding mean squared error estimate [17]

$$\widehat{\text{MSE}}\Big(\hat{f}_{\text{jack}}\Big) = \frac{N-1}{N} \sum_{i=1}^{N} \Big(\hat{f}_i - \hat{f}_{\text{jack}}\Big)^2. \tag{41}$$

If the random variables $A_1, \ldots, A_N$ are not independent, then a jackknife-binning hybrid method can be incorporated, i.e.

$$\begin{aligned} \hat{f}^{(B)}_{\text{jack}} &\equiv \frac{1}{M} \sum_{i=1}^{M} \hat{f}^{(B)}_i, \\ \hat{f}^{(B)}_i &\equiv f\Big(\hat{\mu}^{(B)}_i\Big)\,, \\ \hat{\mu}^{(B)}_i \equiv \frac{1}{M-1} \sum_{\substack{k=1 \\ k \neq i}}^{M} A^{(B)}_k &= \frac{1}{N-B} \sum_{\substack{k=1 \\ k \notin [(i-1)B+1, iB]}}^{N} A_k, \\ \widehat{\text{MSE}}\Big(\hat{f}^{(B)}_{\text{jack}}\Big) &\equiv \frac{M-1}{M} \sum_{i=1}^{M} \Big(\hat{f}^{(B)}_i - \hat{f}^{(B)}_{\text{jack}}\Big)^2, \\ \hat{\tau}_{\hat{f}^{(B)}_{\text{jack}}} &\equiv \frac{\widehat{\text{MSE}}\Big(\hat{f}^{(B)}_{\text{jack}}\Big)}{\widehat{\text{MSE}}\Big(\hat{f}^{(1)}_{\text{jack}}\Big)}, \end{aligned} \tag{42}$$

with M bins of size $B = \frac{N}{M}$. One notable difference between the sample mean binning method (2.3.2) and the jackknife-binning method is that the jackknife estimator $\hat{f}^{(B)}_{\text{jack}}$ depends on the choice of bin size B. However, a suitable choice of B that results in $\widehat{\text{MSE}}\Big(\hat{f}^{(B)}_{\text{jack}}\Big) \approx \text{MSE}\Big(\hat{f}_{\text{jack}}\Big)$ will result in $\hat{\tau}_{\hat{f}^{(B)}_{\text{jack}}} = \hat{\tau}_{\hat{f}_{\text{jack}}} \approx \tau_{\hat{f}_{\text{jack}}}$ as before, as well as $\hat{f}^{(B)}_{\text{jack}} = \hat{f}_{\text{jack}} \approx f(\mu)$. A suitable choice of the bin size B for the sample mean binning of $\text{MSE}(\hat{\mu})$ will also be a suitable choice for the jackknife-binning of $\text{MSE}\Big(\hat{f}_{\text{jack}}\Big)$; in other words, $\tau_{\hat{f}_{\text{jack}}} \leqslant \tau_{\hat{\mu}}$.

3 Results

3.1 Klein-Gordon Theory

The main purpose of simulating Klein-Gordon fields for this project was to establish a strong foundation for field theory calculations before approaching the non-linear sigma model. This was done by calculating the two-point correlation $c(\delta)$ (Equation 11) and

- ensuring calculated values matched the analytic theory,
- optimising programmes,
- comparing the autocorrelation and efficiency of the Metropolis-Hastings algorithm and Gibbs sampler, and
- obtaining possible relationships between the integrated correlation time $\tau_{\hat{c}(\delta)}$ and system parameters, i.e. size $L_1 \times \cdots \times L_{d-1} \times T$, dimensionless grid spacing/mass parameter am.

The Klein-Gordon Metropolis-Hastings algorithm (for a given proposal density function $q(\phi', \phi)$) is outlined as follows (E):

1. Choose a lattice site $\mathbf{x}$ and calculate $\gamma(\mathbf{x})$.
2. Generate a proposed value $\phi'(\mathbf{x}) \sim q(\phi', \phi)$.
3. Calculate $\psi(\mathbf{x}) = \phi(\mathbf{x}) - \kappa^2\gamma(\mathbf{x})$ and $\psi'(\mathbf{x}) = \phi'(\mathbf{x}) - \kappa^2\gamma(\mathbf{x})$.
4. If $|\psi'(\mathbf{x})| \leqslant |\psi(\mathbf{x})|$, then set $\phi(\mathbf{x}) = \phi'(\mathbf{x})$. Otherwise, generate $U \sim \mathcal{U}(0,1)$ and set $\phi(\mathbf{x}) = \phi'(\mathbf{x})$ if $U < \exp\Big[-\frac{(\psi'(\mathbf{x}))^2 - (\psi(\mathbf{x}))^2}{2\kappa^2}\Big]$.
5. Repeat steps 1-4 for each lattice site.

The proposal density function was chosen to be the uniform distribution $\mathcal{U}(\phi(\mathbf{x}) - \varepsilon, \phi(\mathbf{x}) + \varepsilon)$, with ε a free parameter.

The Klein-Gordon Gibbs sampler is outlined as follows (E):

1. Choose a lattice site $\mathbf{x}$ and calculate $\gamma(\mathbf{x})$.
2. Generate a proposed value $\phi'(\mathbf{x}) \sim \mathcal{N}\big(\kappa^2\gamma(\mathbf{x}), \kappa^2\big)$.
3. Set $\phi(\mathbf{x}) = \phi'(\mathbf{x})$.
4. Repeat steps 1-3 for each lattice site.

3.1.1 Analytic Comparison

The first step in the simulation of Klein-Gordon fields was to check that the calculated values of the two-point correlation function using the Metropolis-Hastings algorithm and Gibbs sampler matched the analytic values (Equation 12). The temporal direction of a lattice is arbitrary ($T \equiv L_d$ is purely conventional), and so the calculated correlations in each direction, i.e. $c_j(\delta)$, $T = L_j$, should be equivalent in theory.

The correlations $c_j(\delta)$ were calculated using both algorithms for a variety of lattices, and plotted against the corresponding analytic (normalised) correlation $c(\delta)$ (Figure 1).

The calculated two-point correlation in Figure 1 closely agrees with the analytic result. The error in the correlations were found using the binning method. Examples of plots of $\widehat{\mathrm{MSE}}_{\mathrm{bin}}^{(B)}(\hat{c}_j(\delta))$ are given in Figure 2 and Figure 3, where the characteristic shape of an increase from the naïve error (B too small), to a plateau (suitable B), to erratic behaviour (B too large) can be seen.

3.1.2 Optimisation

Before beginning an in-depth calculation of observables, autocorrelation times, and relationships between system parameters, it was important to first find the optimal approach to incorporating the MCMC algorithms used in this project. This was done by comparing two approaches to writing functions to execute an algorithm for an arbitrary lattice:

1. Creating separate functions for lattices of different dimensions, and so knowing exactly how to navigate the lattice and obtain nearest neighbours.

2. Having a single function for lattices of any shape, and thus needing to calculate how to navigate a lattice and where nearest neighbours are for an arbitrary lattice.

Each method has its own advantages and disadvantages; implementing the first method would be straightforward yet time-consuming, and the second method would not need any modifications once implemented but would initially be more difficult to develop.

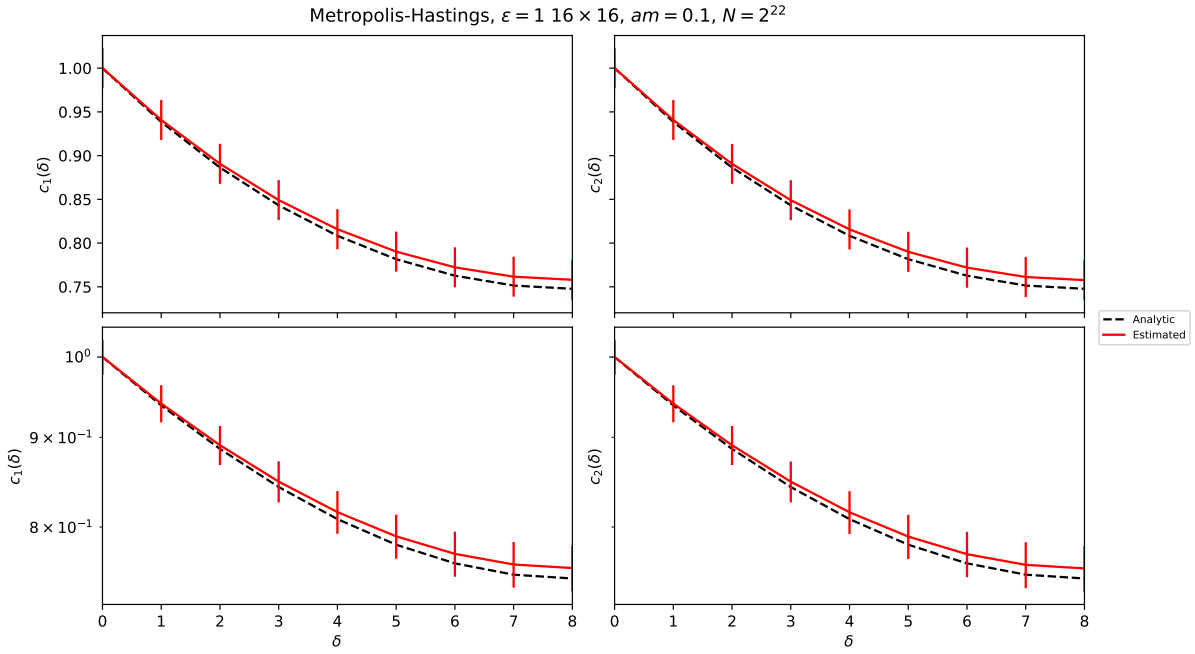
Both methods were written in Python and C++ for the Metropolis-Hastings algorithm and Gibbs sampler, and the time for each of these methods to perform N iterations was calculated for a range of N (Figure 4).

The significant difference between the runtimes of the Python and C++ programmes was expected, however the difference in method speed was much more negligible for C++ than for Python. For both Python and C++, the time to run the Metropolis-Hastings algorithm and Gibbs sampler for the same lattice were virtually identical. Thus, comparing the two algorithms only involved calculating mean squared errors and integrated autocorrelation times.

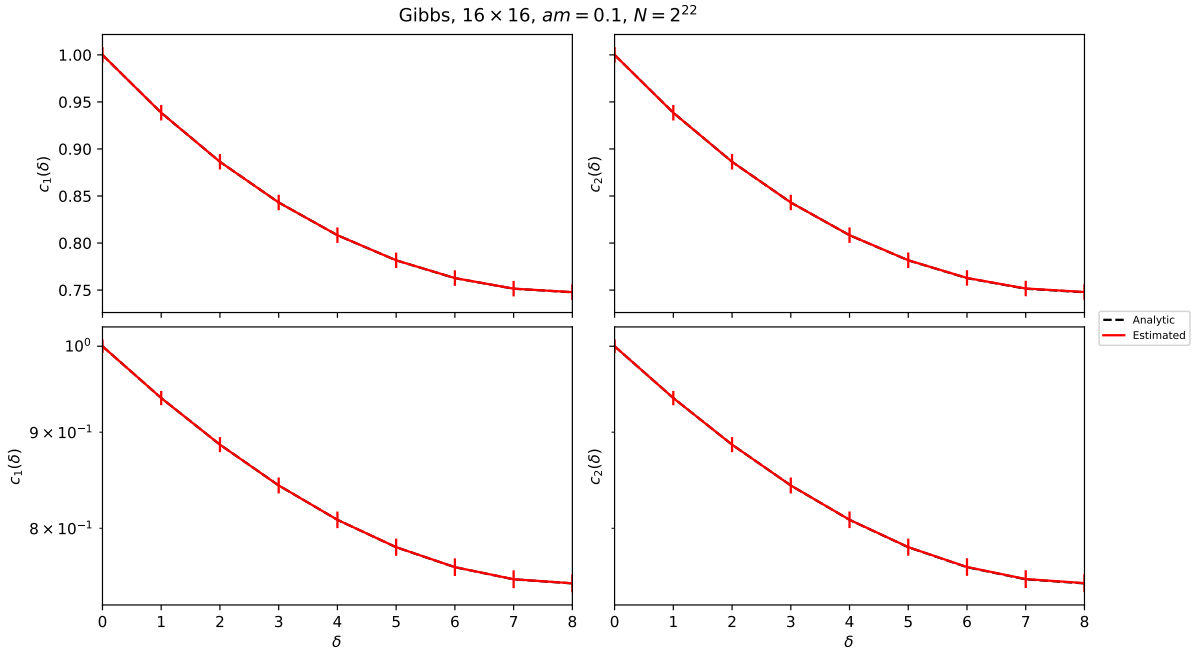
3.1.3 Algorithm Comparison

The two-point correlation $c(\delta)$ was calculated for a 16×16 lattice with $am = 10^{0.5}$ using the Metropolis-Hastings algorithm (for a variety of values of ε) and Gibbs sampler for $N = 2^{21}$ iterations (Figure 5). The corresponding mean squared error $\mathrm{MSE}(\hat{c}(\delta))$ and integrated correlation time $\tau_{\hat{c}(\delta)}$ were also estimated and compared between algorithms (Figure 6).

From Figure 5, the calculated correlations match the analytic values regardless of algorithm or choice of ε, as expected. Figure 6 shows that both the mean squared error and integrated correlation time for the Gibbs sampler were far lower than for the Metropolis-Hastings algorithm for a variety of values of ε. It also shows that the value of ε that reduces error and autocorrelation lies somewhere in the range $(1, 3.2)$. This is further amplified in Figure 7, where the average integrated correlation time for the Metropolis-Hastings algorithm is plotted against ε.

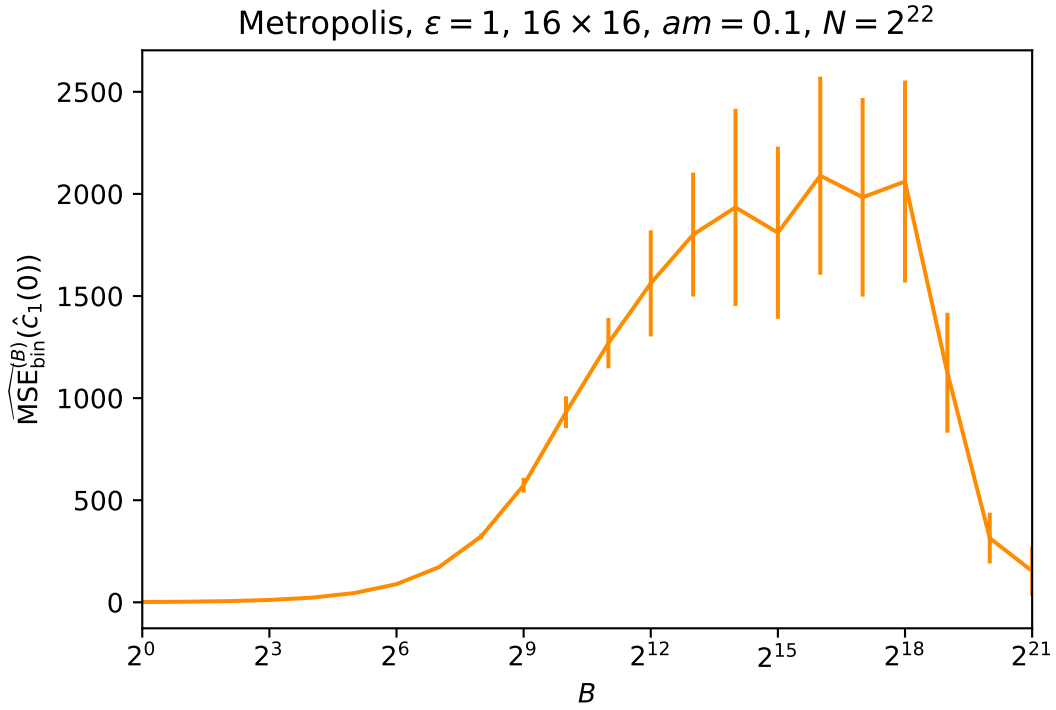


(a) Metropolis-Hastings algorithm, $\varepsilon = 1$.

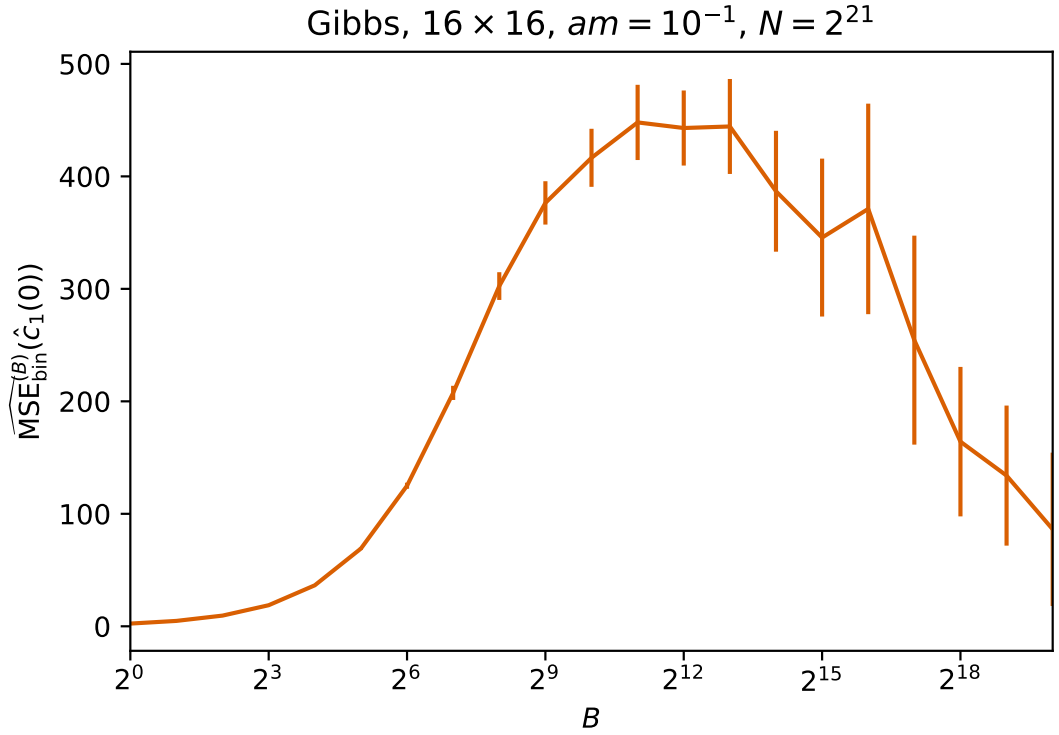


(b) Gibbs sampler.

Figure 1: Plots of (normalised) $\hat{c}_j(\delta)$ and $c(\delta)$ against δ for 16×16, $am = 0.1$, for both the Metropolis-Hastings algorithm and Gibbs sampler. The top and bottom rows show the plots on a linear and logarithmic scale, and the left and right columns correspond to $c_{1,2}(\delta)$.

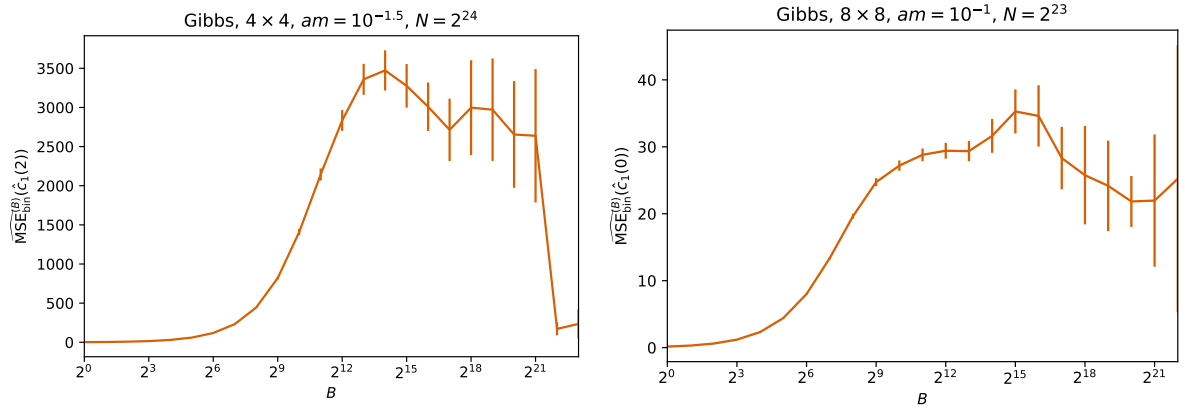


(a) Metropolis-Hastings algorithm, $\varepsilon = 1$.



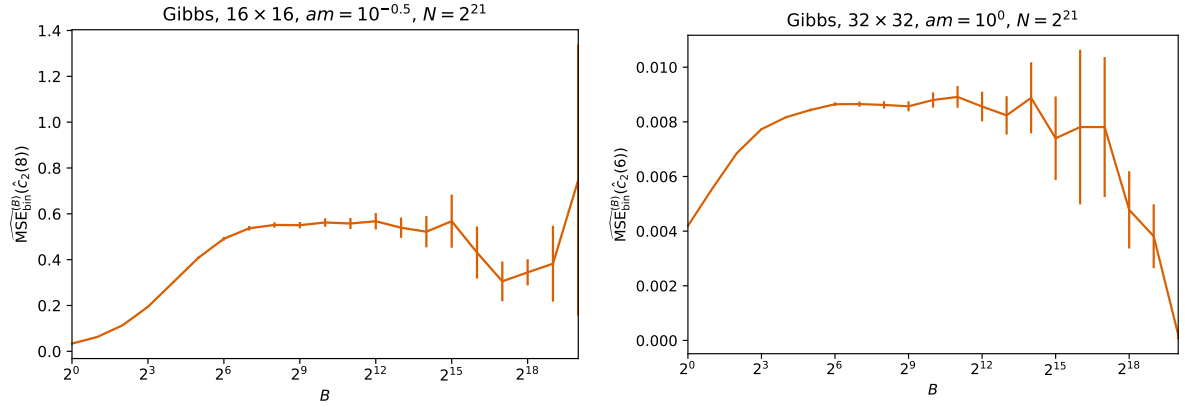
(b) Gibbs sampler.

Figure 2: Plots of $\widehat{\mathrm{MSE}}(\hat{c}_1(0))$ against B for 16×16, $am = 0.1$, for both the Metropolis-Hastings algorithm and Gibbs sampler.



(a) 4×4, $am = 10^{-1.5}$, $N = 2^{24}$, $j = 1$, $\delta = 2$.

(b) 8×8, $am = 10^{-1}$, $N = 2^{23}$, $j = 1$, $\delta = 0$.



(c) 16×16, $am = 10^{-0.5}$, $N = 2^{21}$, $j = 2$, $\delta = 8$.

(d) 32×32, $am = 10^{0}$, $N = 2^{21}$, $j = 2$, $\delta = 6$.

Figure 3: Plots of $\widehat{\mathrm{MSE}}(\hat{c}_j(\delta))$ against B for various lattice parameters and correlations $\hat{c}_j(\delta)$ using the Gibbs sampler.

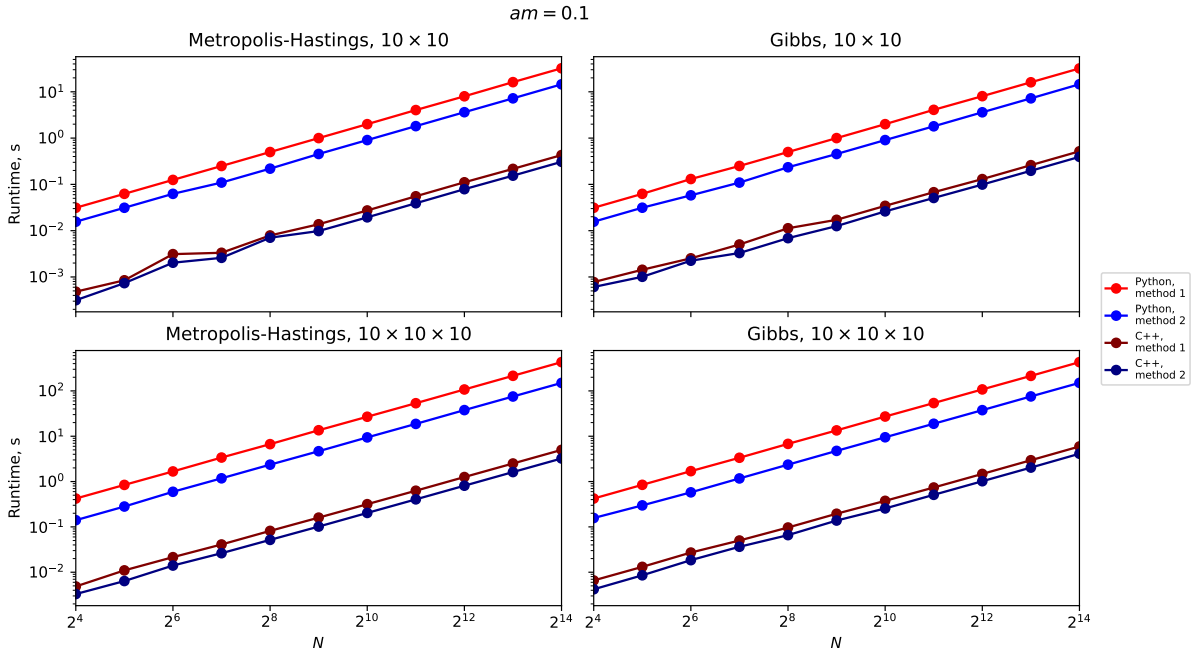


Figure 4: Time for each method to calculate the two-point correlation $c(\delta)$ for $am = 0.1$ using the Metropolis-Hastings algorithm and Gibbs sampler for 10×10 and $10 \times 10 \times 10$ lattices.

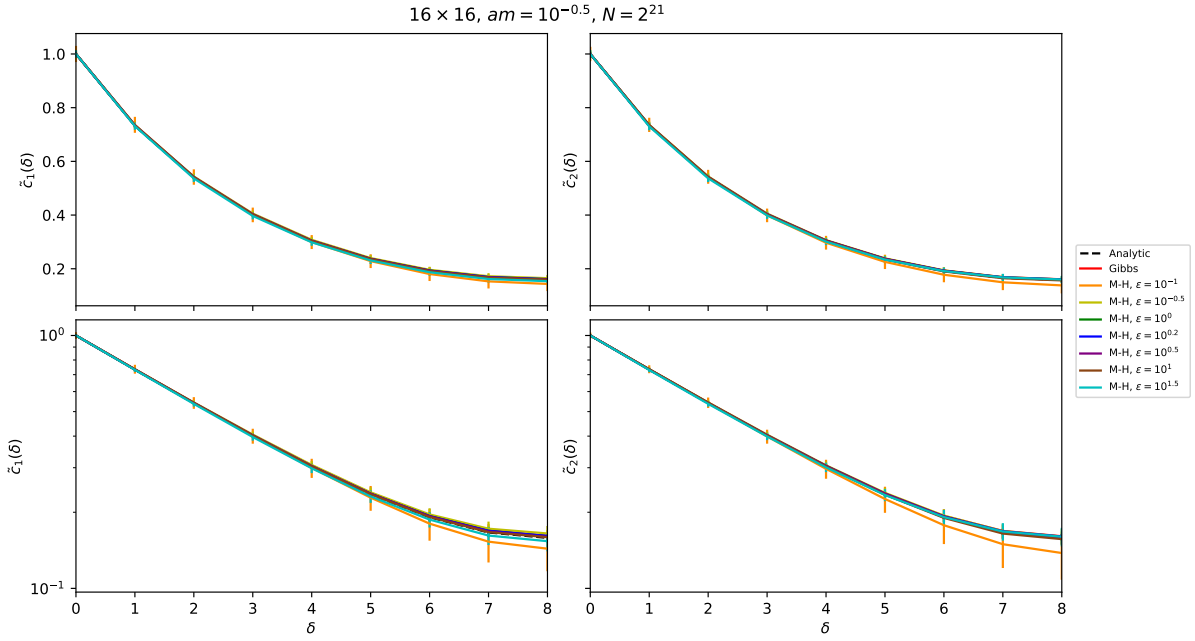


Figure 5: Two-point correlation $c_j(\delta)$ for a 16×16 lattice with $am = 10^{0.5}$ calculated using the Metropolis-Hastings algorithm and Gibbs sampler for $N = 2^{21}$ iterations. The top and bottom rows show the plots on a linear and logarithmic scale, and the left and right columns correspond to $c_{1,2}(\delta)$.

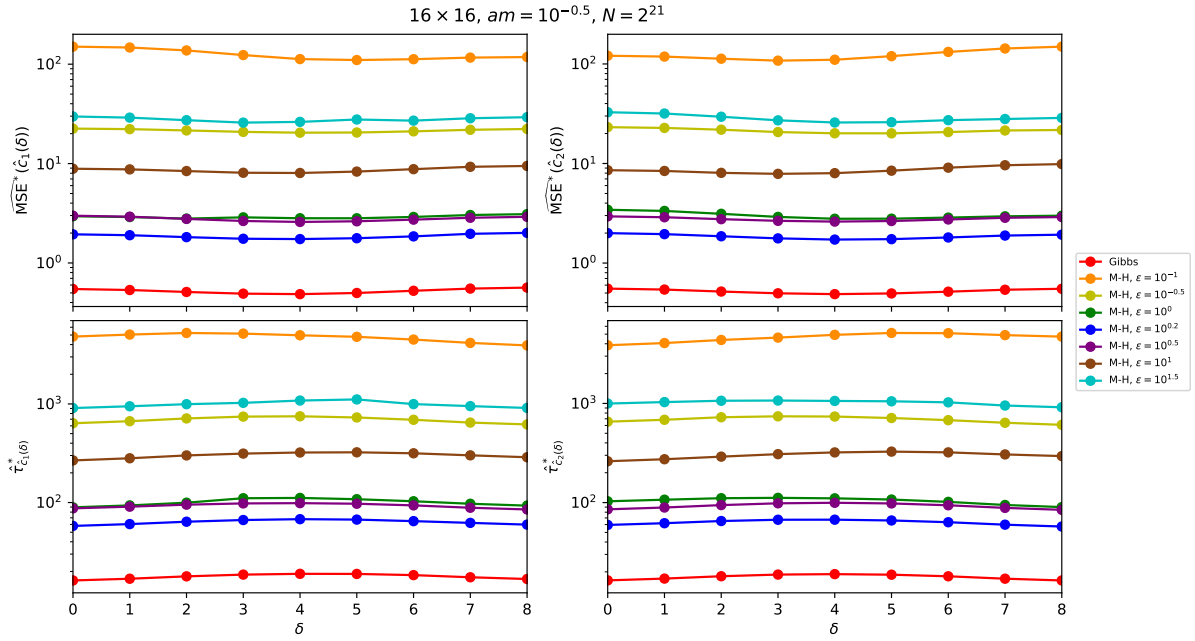


Figure 6: Estimates of $\mathrm{MSE}(\hat{c}_j(\delta))$ and $\tau_{\hat{c}_j(\delta)}$ for the Metropolis-Hastings algorithm and Gibbs sampler (16×16, $am = 10^{0.5}$, $N = 2^{21}$). The asterisk in $\hat{\tau}^*_{\hat{c}_j(\delta)}$ and $\widehat{\mathrm{MSE}}^*(c_j(\delta))$ corresponds to a suitable bin size B being chosen.

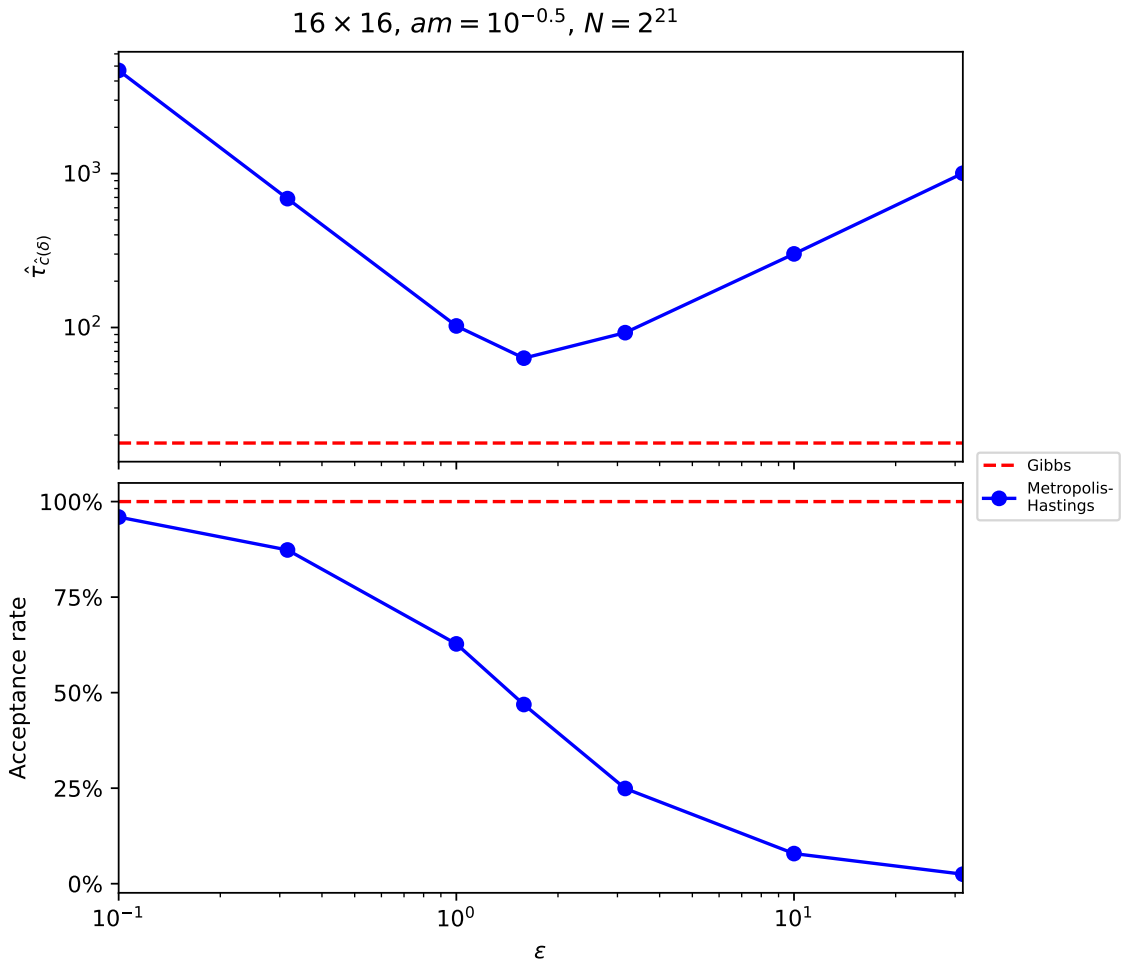


Figure 7: Estimated integrated correlation time $\hat{\tau}_{\hat{c}(\delta)}$ and the acceptance rate of the Metropolis-Hastings algorithm against ε (16×16, $am = 10^{0.5}$, $N = 2^{21}$). Also shown in red are the estimated integrated correlation time and the 100% acceptance rate of the Gibbs sampler.

Thus in terms of reducing autocorrelation, the Gibbs sampler was much more efficient than the Metropolis-Hastings algorithm. Although this has only been shown to be the case for this specific set of parameters, an attempt to find a relationship for the optimal ε that reduces correlation would require much time and calculation, with no guarantee that the Metropolis-Hastings algorithm performs better in any circumstances. This, combined with the facts that the Gibbs sampler does not require optimisation of an external parameter and that the two algorithms take the same time to compute, was conclusive enough to determine that the Gibbs sampler is the superior algorithm.

3.1.4 Parameter Relationships

To conclude the study of Klein-Gordon theory, the relationships between the Gibbs sampler integrated correlation time $\tau_{\hat{c}(\delta)}$ of the correlation estimator $\hat{c}(\delta)$ and the lattice length L of an $L \times L$ lattice (Figure 8) and mass parameter am (Figure 9) was studied.

As can be seen from Figure 8, the integrated correlation time for the correlation estimator does not depend on the lattice length L, whereas from Figure 9 there is an approximate power law relationship with the mass parameter am for $am < 1$, namely $\tau_{\hat{c}(\delta)} \propto am^{-k}$ with $k = 1.99(17)$. Although this was only shown for a 2×2 lattice, a similar relationship should hold for other lattice sizes due to the results in Figure 8.

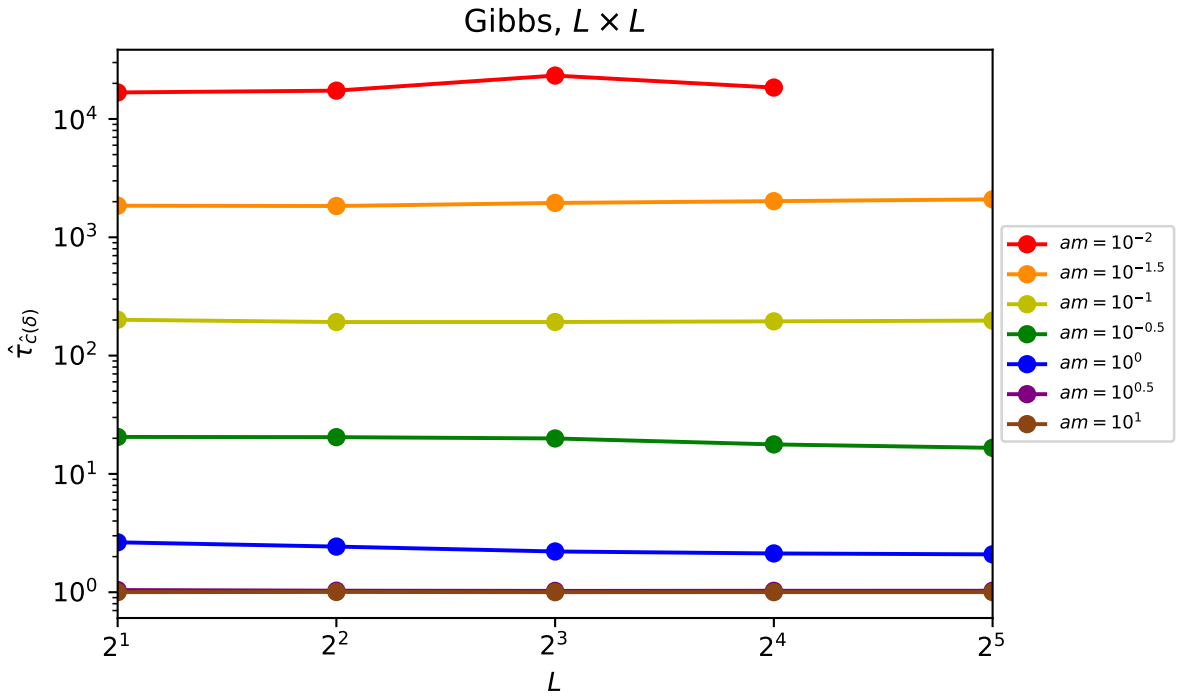


Figure 8: Estimate $\hat{\tau}_{\hat{c}(\delta)}$ of the integrated correlation time for the two-point correlation estimator against lattice length L, for a variety of mass parameters.

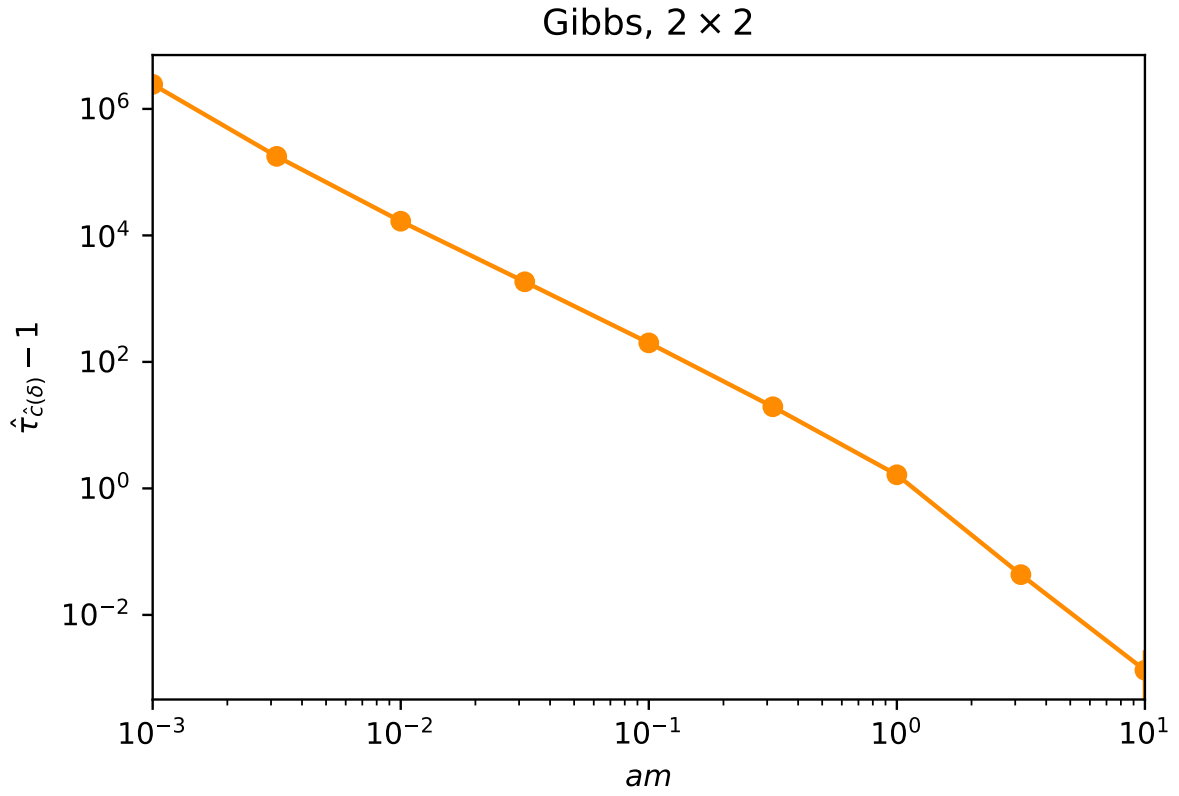


Figure 9: Estimate $\hat{\tau}_{\hat{c}(\delta)}$ of the integrated correlation time for the two-point correlation estimator against mass parameter am, for a 2×2 lattice.

3.2 Non-Linear Sigma Model

The non-linear sigma model Gibbs sampler is outlined as follows (F):

1. Choose a lattice site $\mathbf{x}$ and calculate $\boldsymbol{\Sigma}(\mathbf{x})$.
2. Generate $A \sim \mathcal{N}\big(0, \frac{1}{2}\big)$, $B \sim \text{Exp}(1)$, and $U \sim \mathcal{U}(0,1)$, and set $v_0 = 1 - \frac{A^2+B}{\beta\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})}}$.
3. If $2U^2 > v_0 + 1$, return to step 2.
4. Generate $n_1 \sim \mathcal{U}(-1,1)$ and $\theta \sim \mathcal{U}(0, 2\pi)$, set $n_2 = \sqrt{1-n_1^2}\cos\theta$ and $n_3 = \sqrt{1-n_1^2}\sin\theta$, and set $v_i = n_i\sqrt{1-v_0^2}$ for $i = 1, 2, 3$.
5. Calculate $v = v_0 + i\sum_{k=1}^{3} v_k\tau_k$ and set $\boldsymbol{\phi}(\mathbf{x}) = v \cdot \frac{\boldsymbol{\Sigma}(\mathbf{x})}{\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})}}$.
6. Repeat steps 1-5 for each lattice site.

3.2.1 $SU(2)$ Sampling

The non-linear sigma model Gibbs sampler involves sampling $v \in SU(2)$ from the distribution (F.1)

$$f(v)\,dv \propto \sqrt{1-v_0^2}\, e^{\rho v_0}\delta\big(n^2-1\big)\,dv_0 d^3n \tag{43}$$

using an accept/reject step, where $\rho \equiv \beta\sqrt{\det \boldsymbol{\Sigma}}$. Figure 10 shows graphs of the distribution of v_0 for a variety of values of ρ, and Figure 11 plots the number of rejections per sample against ρ.

From Figure 11, a larger value of ρ, and thus β, corresponds to fewer proposals needed for sampling v, and therefore a faster runtime of calculations.

3.2.2 $d = 2, 3$ Comparison

The $SU(2)$ non-linear sigma model was simulated for 10×20 and $10 \times 10 \times 20$ lattices, with $\beta = 1, 2$ and $\lambda_0 = 0, 0.1, 0.2$, and the corresponding two-point correlations $c_\sigma(\delta)$ and $c_{\pi_i}(\delta)$ were estimated.

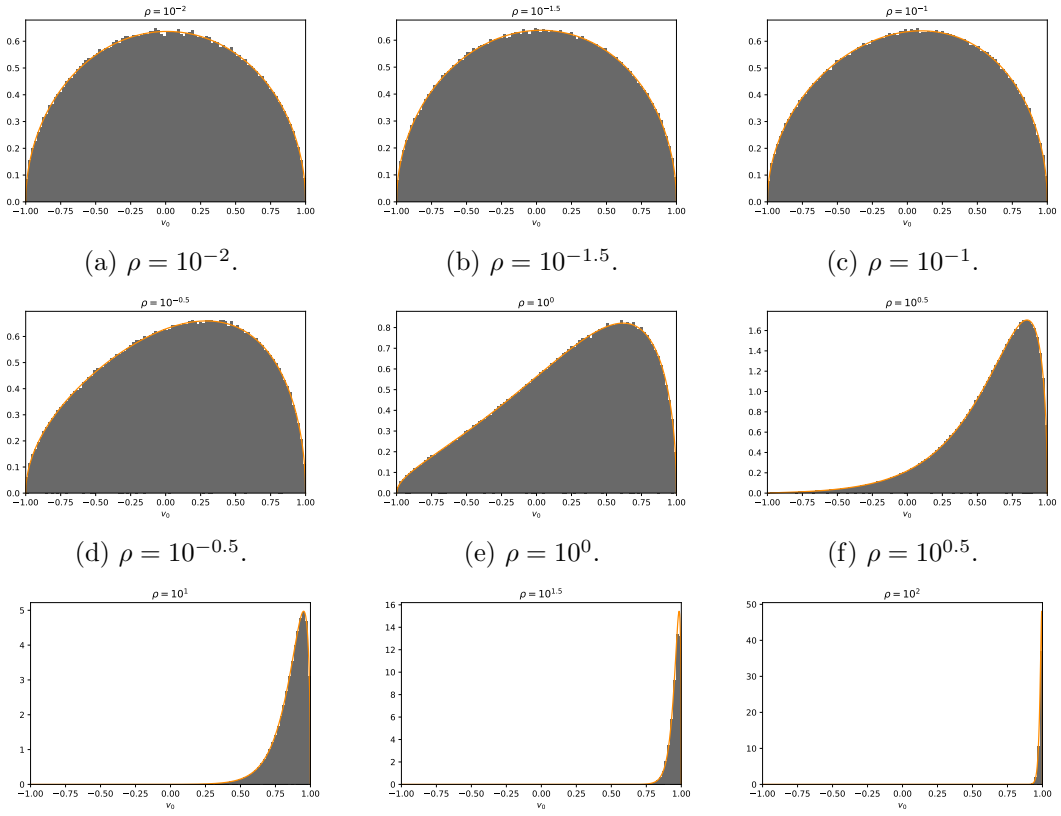
It was first important to check that each of the pion correlations $c_{\pi_i}(\delta)$ were equivalent for each set of system parameters, and thus confirm that the associated masses am_{π_i} were equal. This was done by plotting the three correlations and showing that they were equal within error (Figure 12). As this was shown to be true, only the average correlation

$$c_\pi(\delta) \equiv \frac{1}{3}\sum_{i=1}^{3} c_{\pi_i}(\delta) \tag{44}$$

and corresponding mass am_π were calculated for each system. The masses am_σ, am_π were estimated by using the jackknife method for the estimator (B)

$$\widehat{am} = \ln\left\{\frac{\hat{c}(0) - \hat{c}(2) + \sqrt{[\hat{c}(0)-\hat{c}(2)]^2 - 4\,[\hat{c}(1)-\hat{c}(2)]\,[\hat{c}(0)-\hat{c}(1)]}}{2\,[\hat{c}(1)-\hat{c}(2)]}\right\}, \tag{45}$$

as well as their corresponding integrated correlation times (Table 1, Table 2). As before, the error for the two-point correlations and masses was calculated using the binning method (Figure 13, Figure 14).



(a) $\rho = 10^{-2}$. (b) $\rho = 10^{-1.5}$. (c) $\rho = 10^{-1}$.

(d) $\rho = 10^{-0.5}$. (e) $\rho = 10^{0}$. (f) $\rho = 10^{0.5}$.

(g) $\rho = 10^{1}$. (h) $\rho = 10^{1.5}$. (i) $\rho = 10^{2}$.

Figure 10: Distribution of v_0 from Equation 43 for a variety of ρ values. In grey is a histogram of the sampled values, and orange is the corresponding density function.

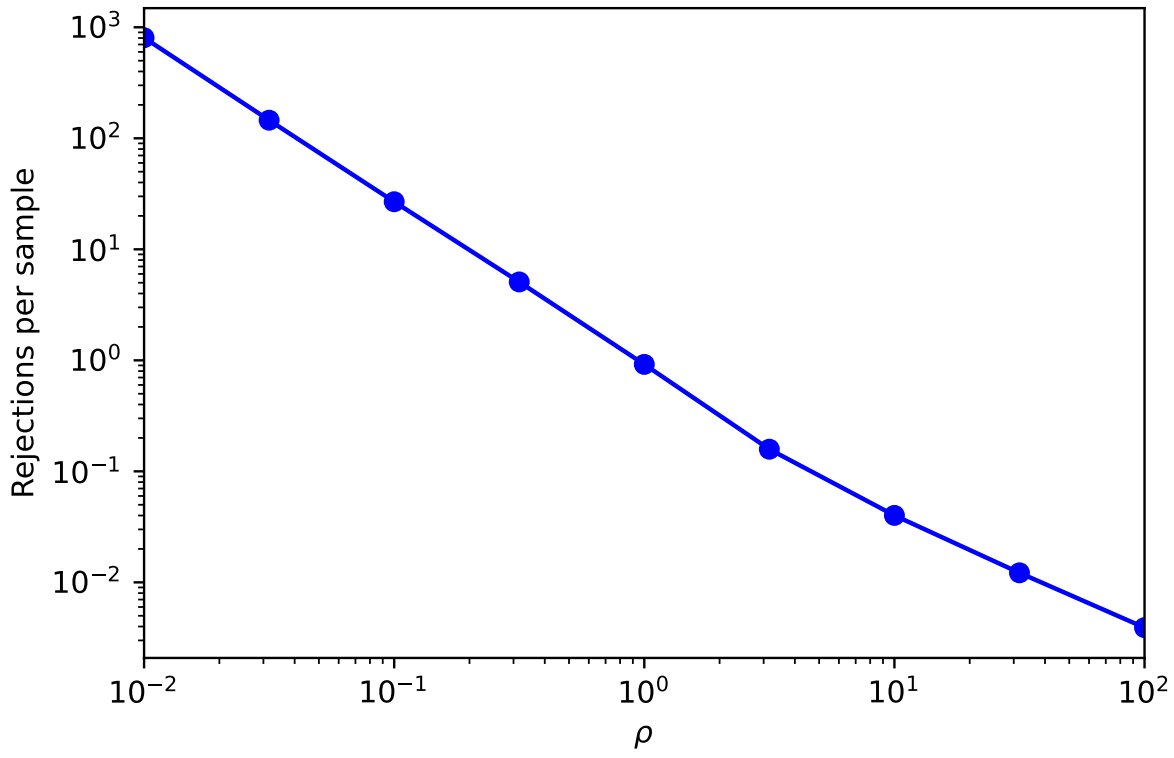


Figure 11: Number of rejections per value sampled from Equation 50.

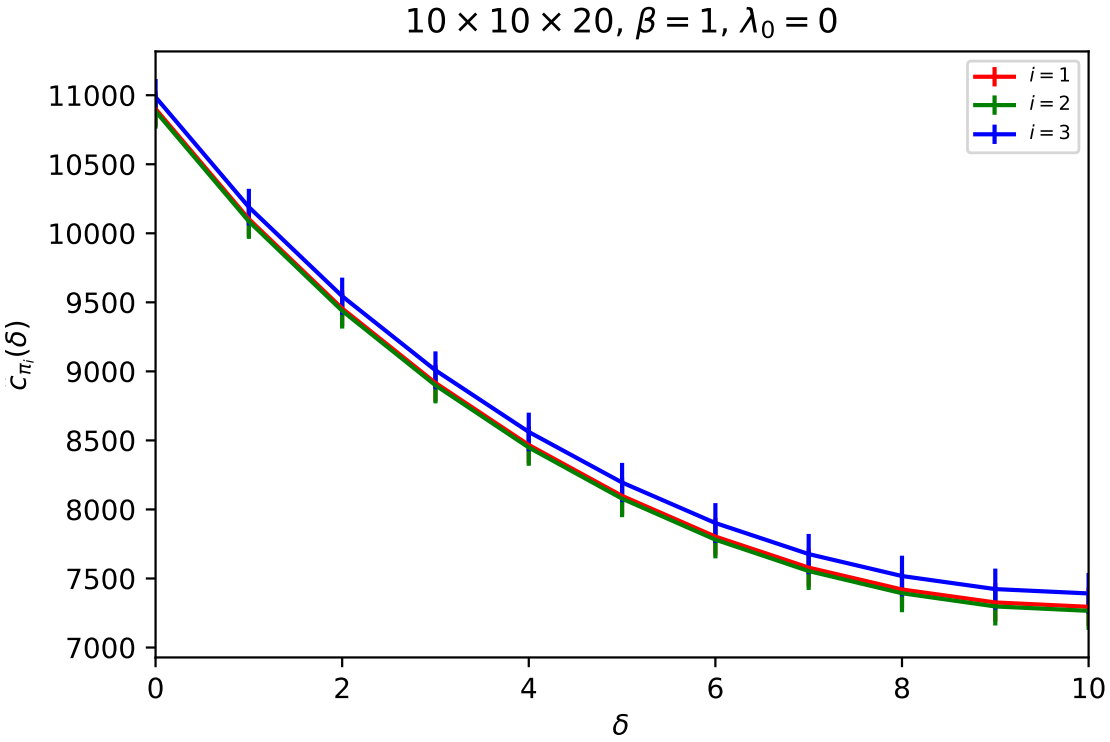


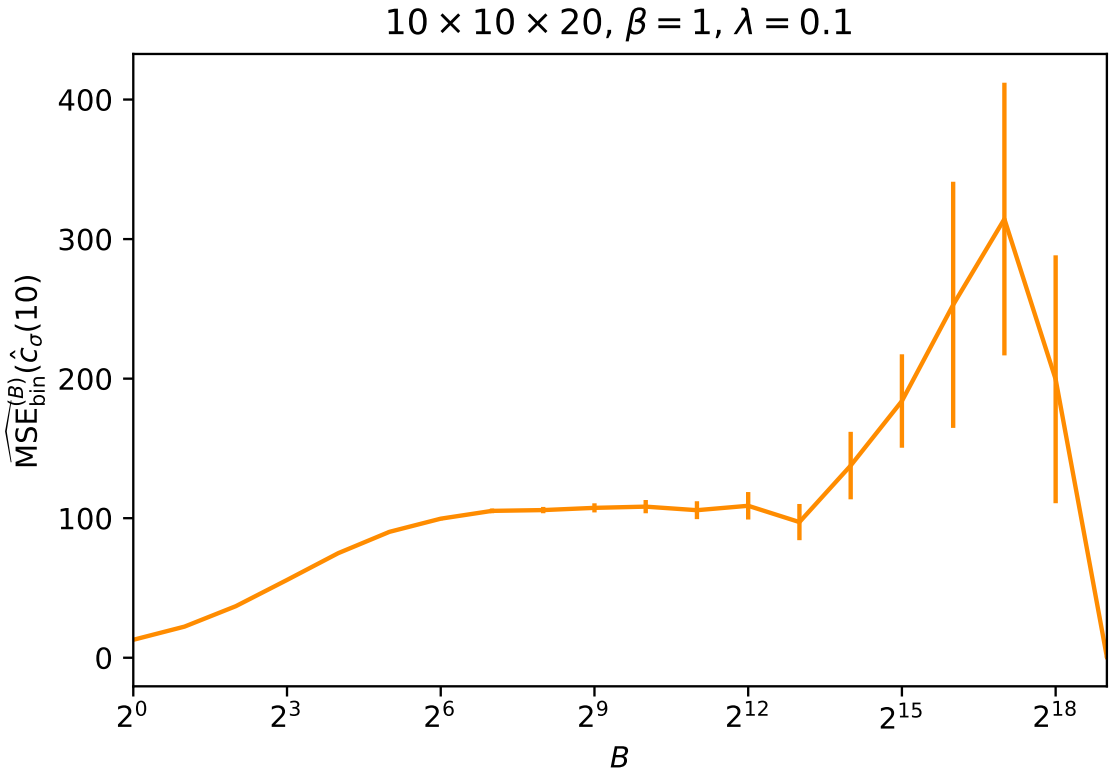
Figure 12: Plots of $\hat{c}_{\pi_i}(\delta)$ for $10 \times 10 \times 20$, $\beta = 1$, $\lambda_0 = 0$, for each of $i = 1, 2, 3$.

β	λ_0	$\widehat{am}_\sigma$	$\hat{\tau}_{\widehat{am}_\sigma}$	$\min \hat{\tau}_{\hat{c}_\sigma(\delta)}$	$\widehat{am}_\pi$	$\hat{\tau}_{\widehat{am}_\pi}$	$\min \hat{\tau}_{\hat{c}_\pi(\delta)}$	Rejection rate
1	0	0.9482(8)	1.29	2.02	0.9472(4)	1.25	1.94	0.3437
	0.1	0.9671(11)	1.26	3.31	0.9542(6)	1.24	1.93	0.3341
	0.2	1.0217(11)	1.24	3.57	0.9743(6)	1.26	1.96	0.2092
2	0	0.2321(9)	5.19	40.23	0.2322(5)	6.07	32.94	0.0729
	0.1	0.8069(13)	2.06	10.30	0.4063(5)	2.43	11.18	0.0674
	0.2	1.1003(13)	1.39	5.34	0.5365	1.88	6.16	0.0634

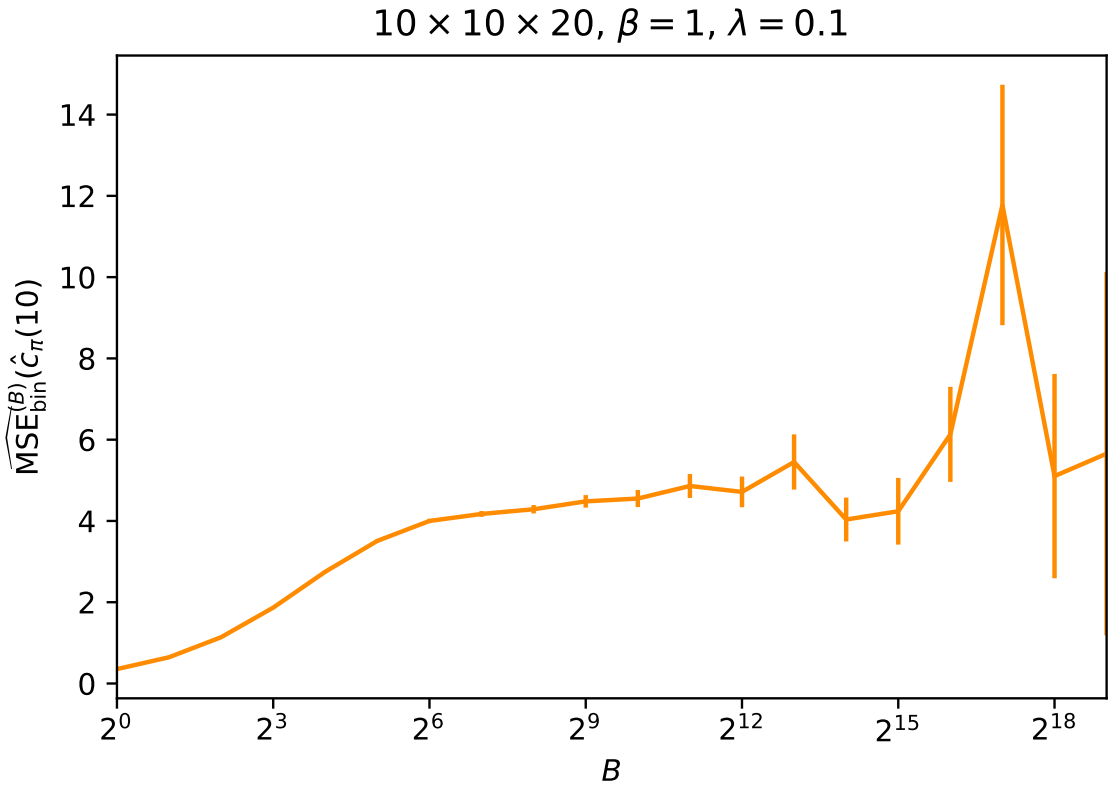
Table 1: Mass estimators $\widehat{am}_{\sigma,\pi}$ and their corresponding integrated correlation times, the minimum integrated correlation time of the estimators $\hat{c}_{\sigma,\pi}(\delta)$, $\delta = 0, 1, 2$ (i.e. the values involved in the calculation of $\widehat{am}$), and the average number of rejections per site for 10×20, $\beta = 1, 2$, $\lambda_0 = 0, 0.1, 0.2$.

β	λ_0	$\widehat{am}_\sigma$	$\hat{\tau}_{\widehat{am}_\sigma}$	$\min \hat{\tau}_{\hat{c}_\sigma(\delta)}$	$\widehat{am}_\pi$	$\hat{\tau}_{\widehat{am}_\pi}$	$\min \hat{\tau}_{\hat{c}_\pi(\delta)}$	Rejection rate
1	0	0.2107(20)	14.5	275.58	0.2092(11)	12.6	247.02	0.1418
	0.1	1.0741(13)	1.55	8.30	0.4380(6)	2.95	12.62	0.1243
	0.2	1.3315(15)	1.31	4.70	0.5843(6)	2.19	7.01	0.1140
2	0	0.1532(26)	25.8	2040.95	0.1523(10)	10.6	2037.12	0.0393
	0.1	1.5150(19)	1.4	10.58	0.3472(6)	3.71	20.21	0.0383
	0.2	1.7790(22)	1.2	4.41	0.4844(6)	2.56	9.76	0.0373

Table 2: Mass estimators $\widehat{am}_{\sigma,\pi}$ and their corresponding integrated correlation times, the minimum integrated correlation time of the estimators $\hat{c}_{\sigma,\pi}(\delta)$, $\delta = 0, 1, 2$ (i.e. the values involved in the calculation of $\widehat{am}$), and the average number of rejections per site for $10 \times 10 \times 20$, $\beta = 1, 2$, $\lambda_0 = 0, 0.1, 0.2$.

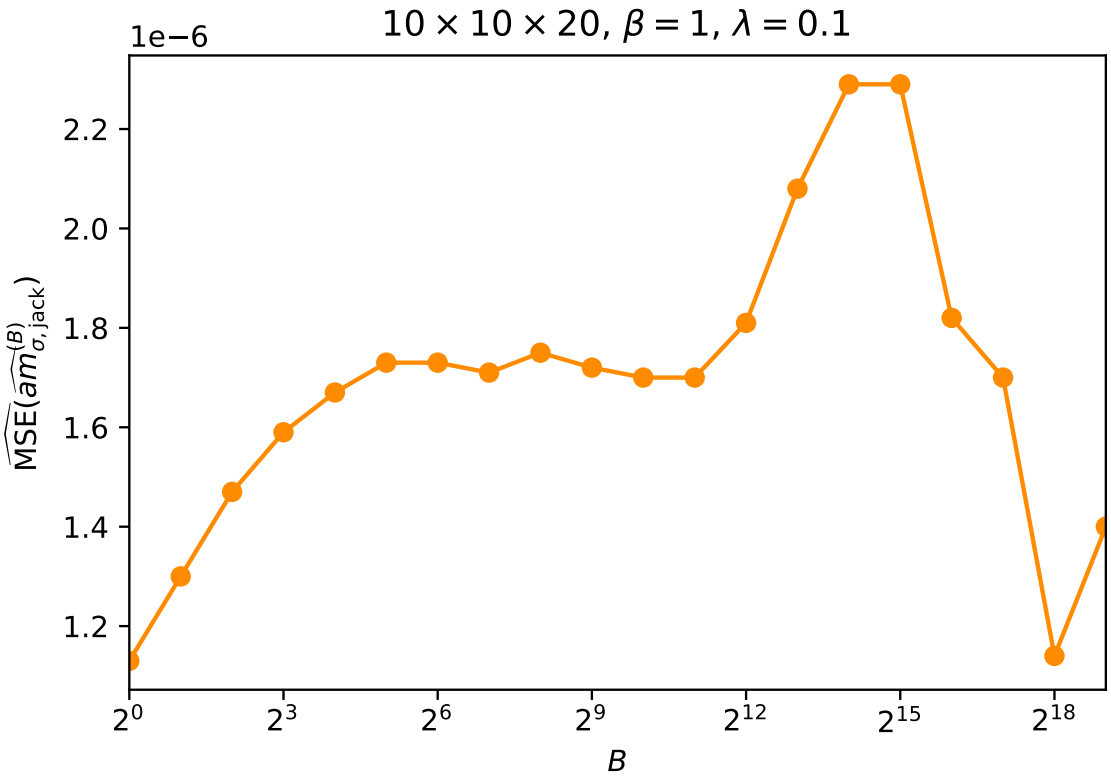


(a) $c_\sigma(10)$

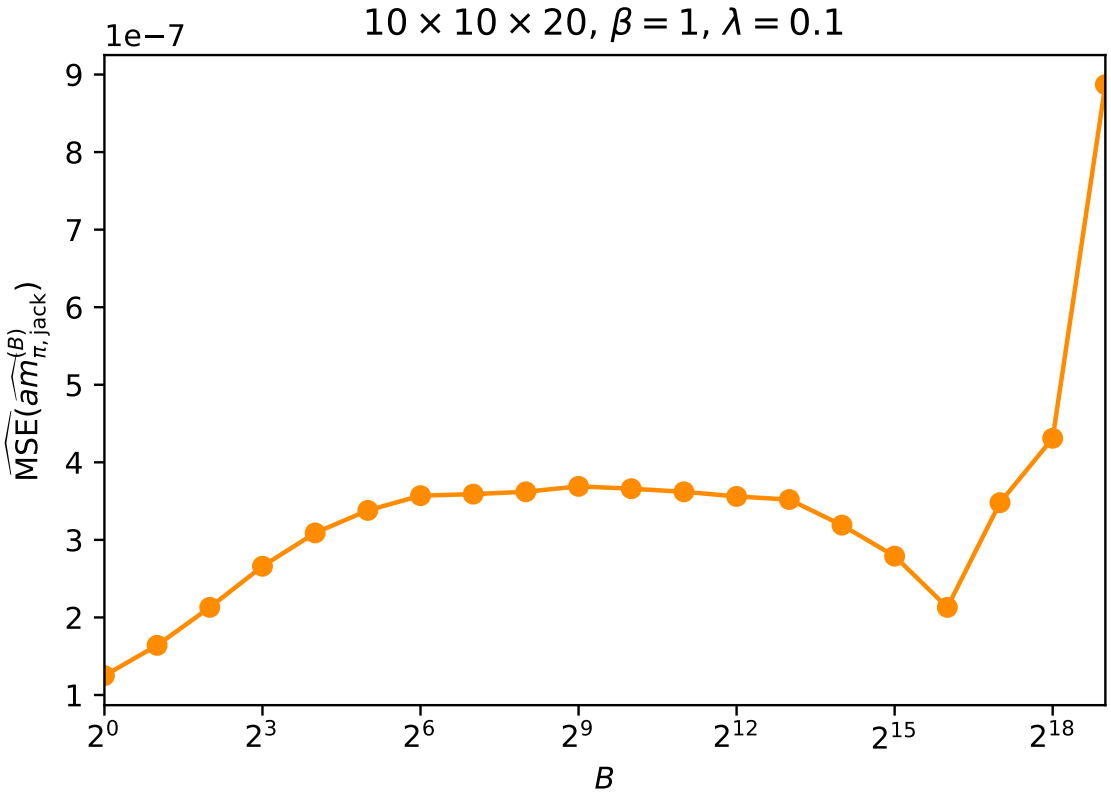


(b) $c_\pi(10)$

Figure 13: Plots of $\widehat{\mathrm{MSE}}^{(B)}_{\mathrm{bin}}(\hat{c}(10))$ against B for c_σ and c_π, for $10 \times 10 \times 20$, $\beta = 1$, $\lambda_0 = 0.1$.



(a) am_σ



(b) am_π

Figure 14: Plots of $\widehat{\mathrm{MSE}}\left(\widehat{am}_{\mathrm{jack}}^{(B)}\right)$ against B for am_σ and am_π, for $10 \times 10 \times 20$, $\beta = 1$, $\lambda_0 = 0.1$.

One notable result from Table 1 and Table 2 is that the mass parameters am_σ and am_π are roughly equal for $\lambda_0 = 0$. It is also evident that the integrated correlation times for the mass parameter estimators tend to be higher for $\lambda_0 = 0$. Increasing β, λ_0, or the lattice dimension also results in a lower rejection rate per site update. Although the error in the pion mass calculations is lower than in the sigma mass (due to the pion mass calculations involving three times as many data points), the corresponding integrated correlation times of the pion mass tend to be appreciably larger for $\lambda_0 \neq 0$. In each case, the integrated correlation time for the mass parameters were less than that for the two-point correlation; the factor by which they differed ranged from 1.6 to 192.2.

The remainder of the study of the non-linear sigma model focused mainly on the massive pions that appear as a result of the explicit symmetry breaking of the system, i.e. $\lambda_0 \neq 0$. As there do not exist Goldstone bosons in two dimensions due to the Mermin-Wagner theorem [20, 21], only 3-dimensional lattices were considered.

3.2.3 am_σ and am_π

The non-linear sigma model was simulated for a variety of system parameters ($10 \times 10 \times 20$, $10 \times 10 \times 10$, $13 \times 13 \times 13$ lattice sizes, $\lambda_0 = 0.1, 0.2, 0.3$, $\beta \in [0.5, 3]$). The mass parameters am_σ and am_π were calculated for each of these, and plotted against β (Figure 15).

As can be seen from Figure 15, the behaviour of each mass is similar across lattice sizes and value of λ_0; in particular, am_σ has a λ_0-dependent minimum with respect to β, whereas am_π monotonically decreases as β increases. It is also clear that, for the same set of parameters, $am_\sigma(\beta) > am_\pi(\beta)$.

3.2.4 $(am_\pi)^2$ and λ_0

The pion mass $(am_\pi)^2$ was also plotted against β (Figure 16) for the same set of lattice parameters as before.

As is clear from Figure 16, the pion mass $(am_\pi)^2$ for each set of parameters asymptotes to λ_0 as $\beta \to \infty$, i.e. as the temperature decreases to 0, indicating the bare mass of the pion.

3.2.5 $\sigma \to \pi + \pi$ Decay

It was also of interest to find the conditions under which a theoretical $\sigma \to \pi + \pi$ decay was possible, i.e. where $am_\sigma \geqslant 2am_\pi$, and so the difference in the masses $am_\sigma - 2am_\pi$ was plotted against β for the same lattice sizes and values of λ_0 (Figure 17).

Figure 17 shows that the temperature at which a $\sigma \to \pi + \pi$ decay is possible is $\beta \approx 0.9325$, for any lattice size or λ_0.

3.2.6 $\tau_{\widehat{am}}$ and Rejection Rate

To conclude the study of the non-linear sigma model, the integrated correlation times $\tau_{\widehat{am}_\sigma}$, $\tau_{\widehat{am}_\pi}$ and the average number of rejections per site update were calculated and plotted against β (Figure 18, Figure 19) to obtain a measure of the efficiency of the Gibbs sampler.

Although the calculation of $\widehat{am}_\sigma$ involves three times fewer data points than $\widehat{am}_\pi$, the corresponding integrated correlation time is appreciably lower across all parameters, as can be seen in Figure 18. Similarly to Figure 11, the rejection rate in Figure 19 significantly decreases as β increases, as expected. Figure 19 also shows that the rejection rate is much more dependent on β than on λ_0.

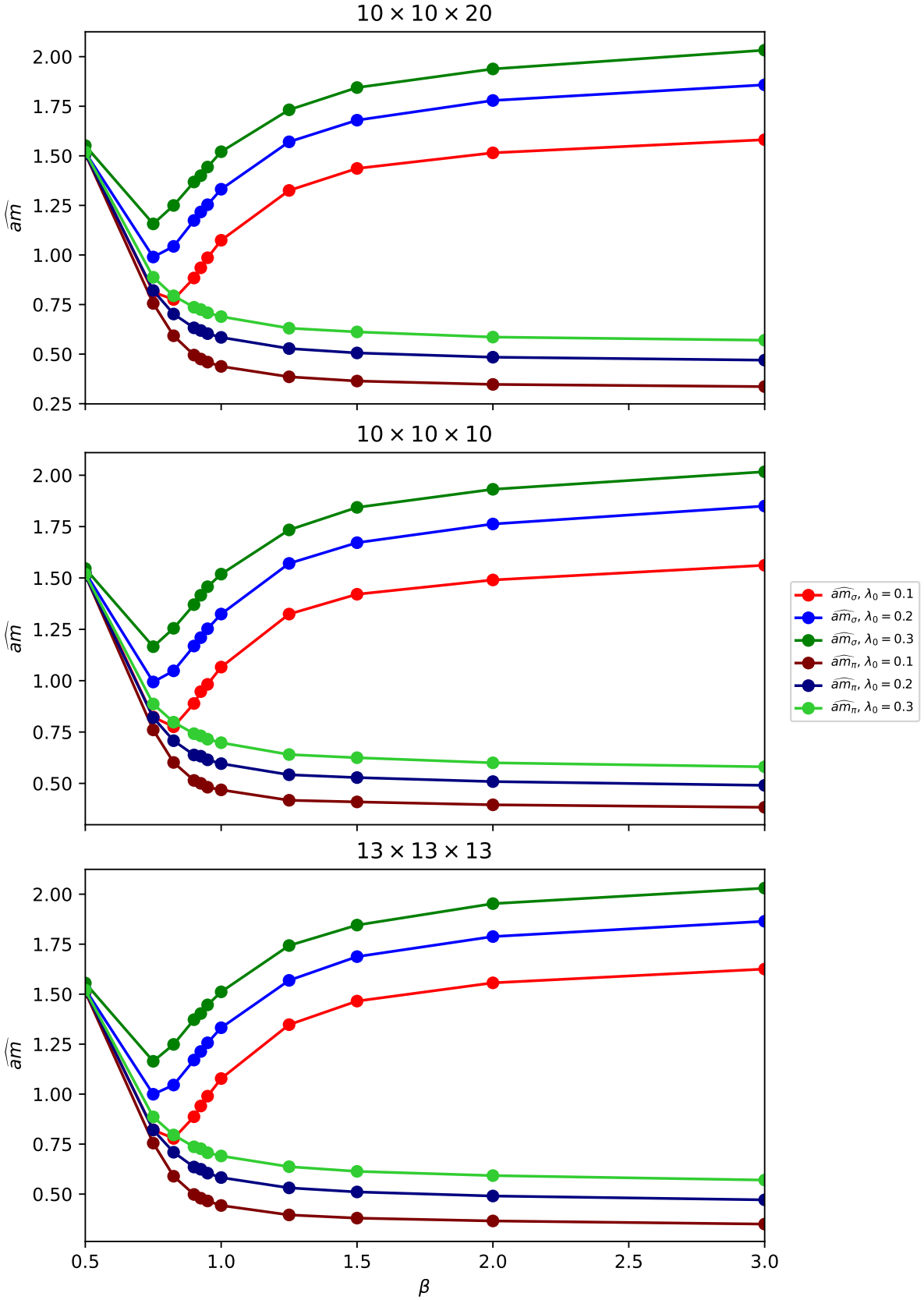


Figure 15: Plots of $\widehat{am}_\sigma$ and $\widehat{am}_\pi$ against β for $\lambda = 0.1, 0.2, 0.3$ and various 3-dimensional lattices.

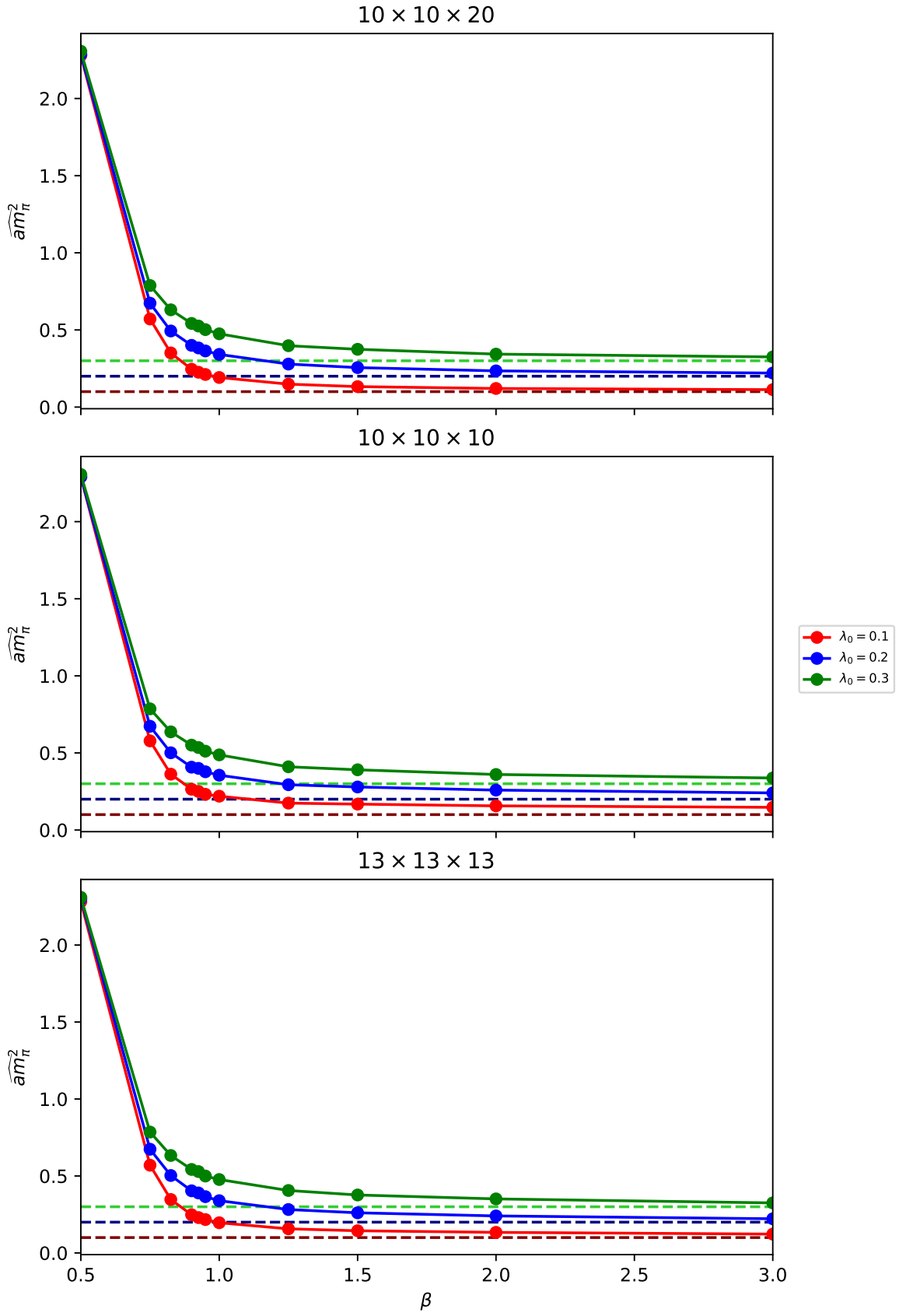


Figure 16: Plots of $(\widehat{am}_\pi)^2$ against β for $\lambda = 0.1, 0.2, 0.3$ and various 3-dimensional lattices. The dashed lines represent the asymptotic behaviour $(am_\pi)^2 \to \lambda_0$ as $\beta \to \infty$.

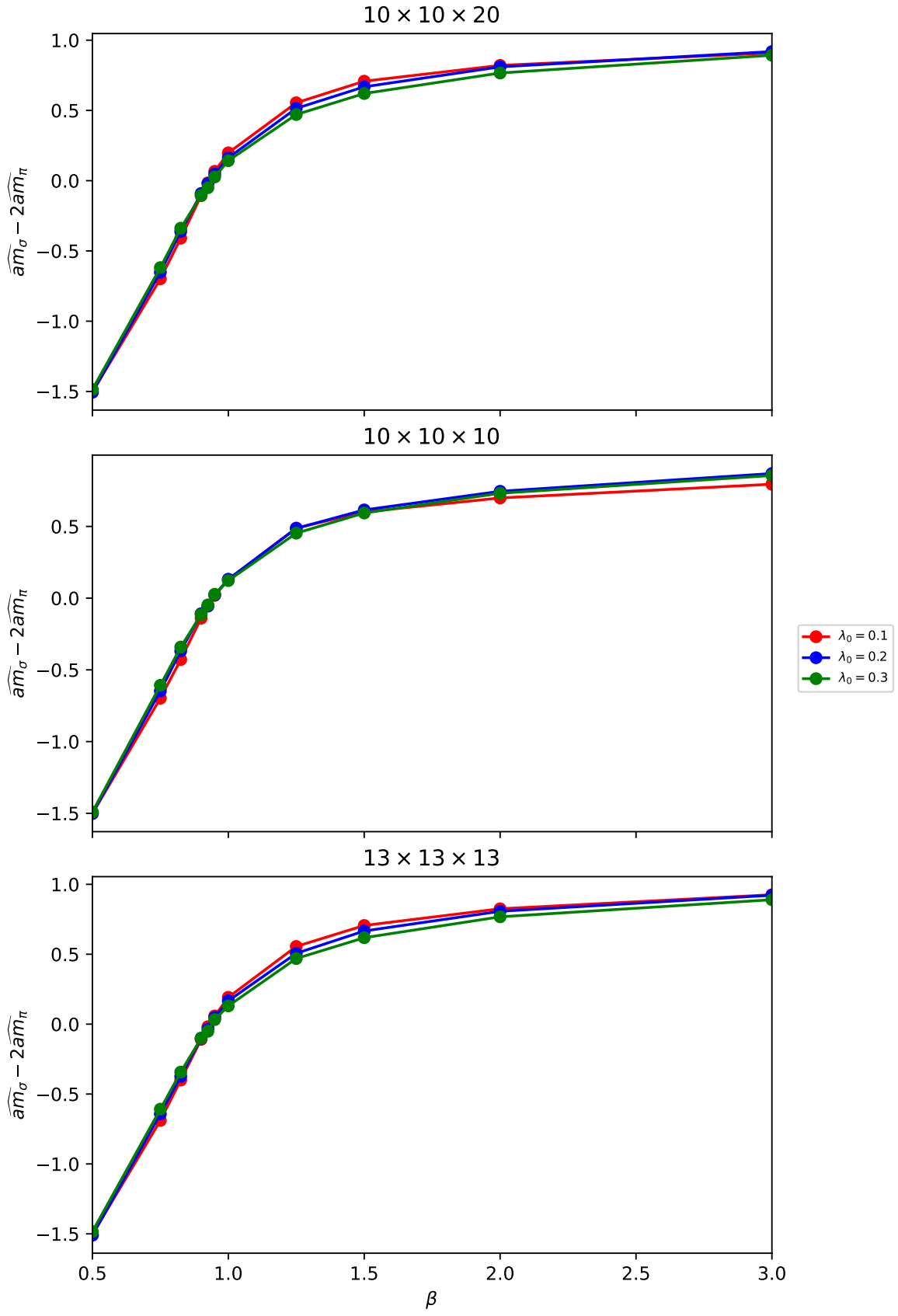


Figure 17: Plots of $\widehat{am}_\sigma - 2\widehat{am}_\pi$ against β for $\lambda = 0.1, 0.2, 0.3$ and various 3-dimensional lattices.

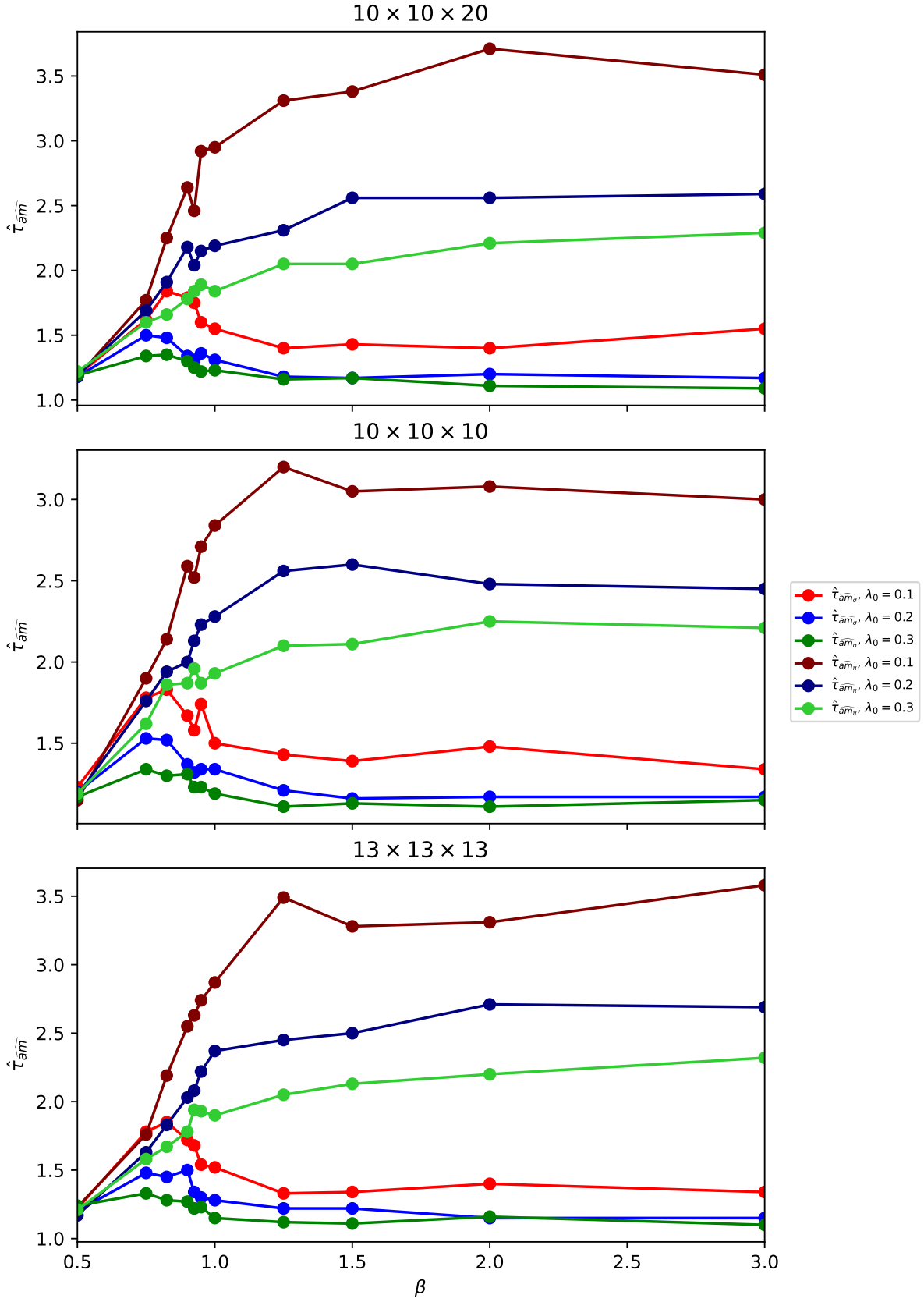


Figure 18: Plots of $\hat{\tau}_{\widehat{am}_\sigma}$ and $\hat{\tau}_{\widehat{am}_\pi}$ against β for $\lambda = 0.1, 0.2, 0.3$ and various 3-dimensional lattices.

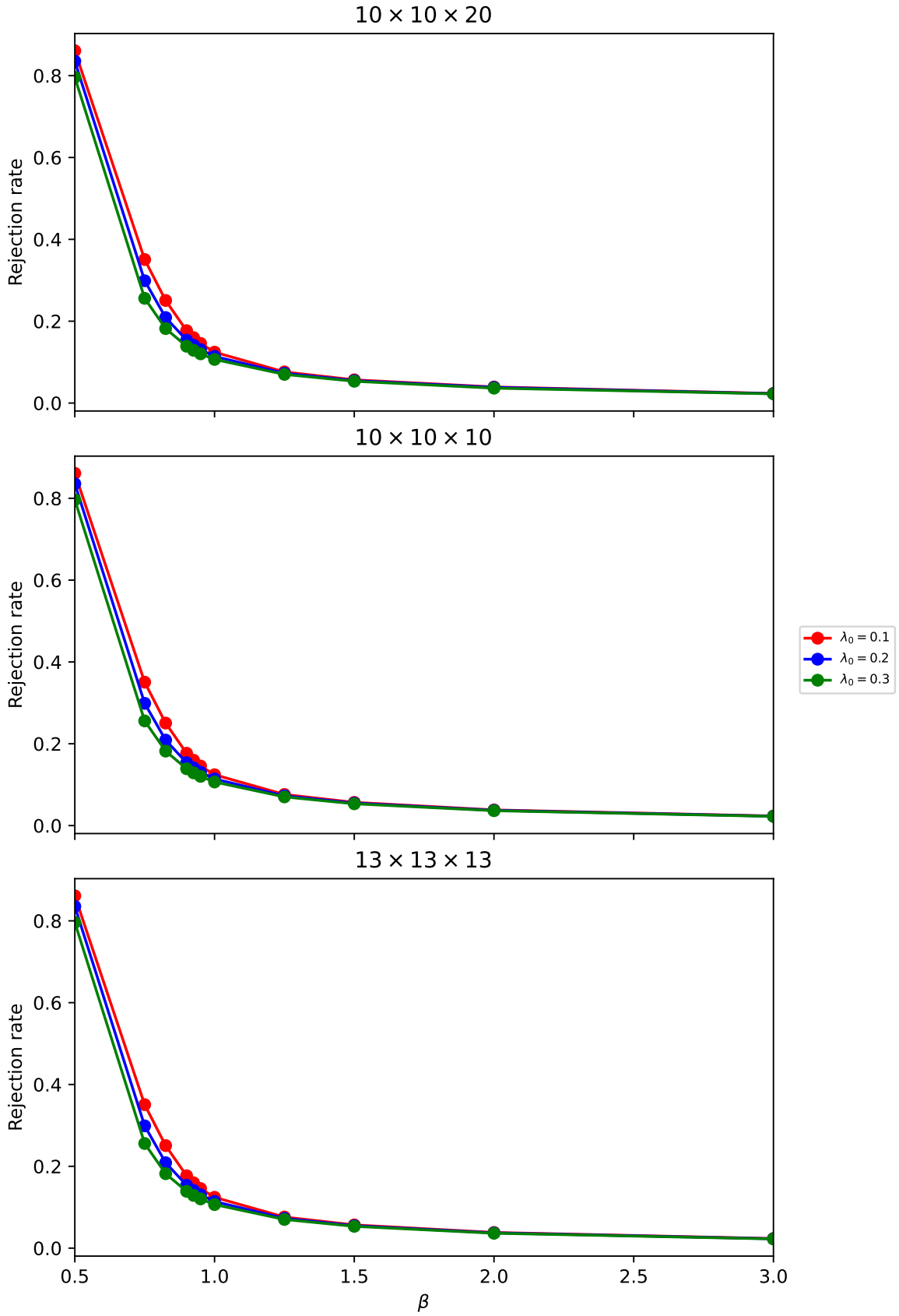


Figure 19: Plots of average rejections per site update against β for $\lambda = 0.1, 0.2, 0.3$ and various 3-dimensional lattices.

Conclusion & Further Work

Many areas of this project could have been expanded if time permitted. The conclusion of the Gibbs sampler being preferable to the Metropolis-Hastings algorithm in 3.1.3 would be far more concrete if a relationship between the optimal choice of ε and the system parameters was established and subsequently compared to the Gibbs sampler for a variety of systems. Original aims of this project that unfortunately did not get addressed included mapping the phase shift for scattering for a range of pion momenta using the Luescher method [22], and studying the decay of a heavy scalar boson coupled to the pion fields.

Much of the time and calculations involved in this project were spent on Klein-Gordon theory: optimising programmes, increasing accuracies, and relating algorithm efficiency and system parameters. While a more in-depth investigation of the non-linear sigma model would be more in line with the title of this project, the study of the more straightforward Klein-Gordon theory provided some useful information about the algorithms and approaches to numerical lattice field theory calculation, and was necessary to develop a general and optimised structure.

References

[1] O. Klein, "Quantentheorie und Fünfdimensionale Relativitätstheorie," *Zeitschrift für Physik*, vol. 37, no. 12, pp. 895–906, 1926.

[2] W. Gordon, "Der Comptoneffekt nach der Schrödingerschen Theorie," *Zeitschrift für Physik*, vol. 40, no. 1-2, pp. 117–133, 1926.

[3] M. Gellmann and M. Lévy, "The Axial Vector Current in Beta-Decay," *Nuovo Cimento*, vol. 16, no. 4, pp. 705–726, 1960.

[4] F. Gürsey, "On the Symmetries of Strong and Weak Interactions," *Nuovo Cimento*, vol. 16, no. 2, pp. 230–240, 1960.

[5] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Physical Review*, vol. 127, no. 3, pp. 965–970, 1962.

[6] M. E. Peskin and D. V. Schroeder, "4.2 Perturbation Expansion of Correlation Functions," in *An Introduction to Quantum Field Theory*, pp. 82–87, Boca Raton: CRC Press, 1995.

[7] H. J. Rothe, "2. The Path Integral Approach to Quantization," in *Lattice Gauge Theories: An Introduction*, World Scientific Lecture Notes in Physics, pp. 7–35, Singapore: World Scientific Publishing Company, 3rd ed., 2005.

[8] H. J. Rothe, "10.1 Energy Sum Rule for the Harmonic Oscillator," in *Lattice Gauge Theories: An Introduction*, World Scientific Lecture Notes in Physics, pp. 130–135, Singapore: World Scientific Publishing Company, 3rd ed., 2005.

[9] S. M. Ross, "8.1 The Sample Mean and Sample Variance," in *Simulation*, pp. 135–141, Cambridge: Academic Press, 5th ed., 2005.

[10] E. Cho and M. J. Cho, "Variance of the With-Replacement Sample Variance," *JSM Proceedings*, pp. 1291–1293, 2008.

[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[12] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[13] S. M. Ross, "12. Markov Chain Monte Carlo Methods," in *Simulation*, pp. 271–301, Cambridge: Academic Press, 5th ed., 2005.

[14] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[15] F. Knechtli, M. Günther, and M. Peardon, "2. Monte Carlo Methods," in *Lattice Quantum Chromodynamics: Practical Essentials*, SpringerBriefs in Physics, pp. 35–54, Dordrecht: Springer, 2017.

[16] A. D. Sokal, "Monte Carlo Methods in Statistical Mechanics: Foundations and Algorithms," Postgraduate Physics Courses, (Lausanne, Switzerland), Cargèse Summer School, 1996.

[17] B. A. Berg, "2.7 The Jackknife Approach," in *Markov Chain Monte Carlo Simulations and Their Statistical Analysis With Web-Based Fortran Code*, pp. 103–108, Singapore: World Scientific Publishing, 2004.

[18] M. H. Quenouille, "Notes on Bias in Estimation," *Biometrika*, vol. 43, no. 3-4, pp. 353–360, 1956.

[19] J. W. Tukey, "Bias and Confidence in Not-Quite Large Samples (abstract)," *Annals of Mathematical Statistics*, vol. 29, no. 2, p. 614, 1958.

[20] N. D. Mermin and H. Wagner, "Absence of Ferromagnetism or Antiferromagnetism in One- Or Two-Dimensional Isotropic Heisenberg Models," *Physical Review Letters*, vol. 17, no. 22, pp. 1133–1136, 1966.

[21] S. Coleman, "There Are No Goldstone Bosons in Two Dimensions," *Communications in Mathematical Physics*, vol. 31, no. 4, pp. 259–264, 1973.

[22] R. A. Briceno, J. J. Dudek, and R. D. Young, "Scattering Processes and Resonances from Lattice QCD," *Reviews of Modern Physics*, vol. 90, no. 2, p. 42, 2018.

[23] H. J. Rothe, "3. The Free Scalar Field on the Lattice," in *Lattice Gauge Theories: An Introduction*, World Scientific Lecture Notes in Physics, pp. 36–42, Singapore: World Scientific Publishing Company, 3rd ed., 2005.

[24] S. M. Ross, "4.4 The Acceptance-Rejection Technique," in *Simulation*, pp. 56–58, Cambridge: Academic Press, 5th ed., 2005.

[25] K. Fabricius and O. Haan, "Heat Bath Method for the Twisted Eguchi-Kawai Model," *Physics Letters B*, vol. 143, no. 4-6, pp. 459–462, 1984.

Appendices

A Lattice Field Theories

A.1 Klein-Gordon Lattice Action

The Klein-Gordon action in d-dimensional Euclidean space for a boson of mass m is given by [1, 2]

$$S(\phi) = \frac{1}{2}\int d^d x\, \phi(x)\left(-\Box + m^2\right)\phi(x). \tag{1}$$

Applying the lattice parameterisation (Equation 2) results in the transformations [23]

$$\begin{aligned}
&\int d^d x \to a^d \sum_n, \quad \Box\phi(x) \to \frac{1}{a^2}\sum_{i=1}^{d}\left[\phi(a\,n + a\,e_i) + \phi(a\,n - a\,e_i) - 2\phi(a\,n)\right], \\
\implies S(\phi) &\to \frac{a^d}{2}\sum_n \left\{ m^2\phi^2(a\,n) + \frac{2d}{a^2}\phi^2(a\,n) - \frac{\phi(a\,n)}{a^2}\sum_{i=1}^{d}\left[\phi(a\,n + a\,e_i) + \phi(a\,n - a\,e_i)\right]\right\} \\
&= \sum_n \left[a^{\frac{d}{2}-1}\phi(a\,n)\right]\left\{\frac{\left[(am)^2 + 2d\right]}{2} a^{\frac{d}{2}-1}\phi(a\,n) - \sum_{i=1}^{d} a^{\frac{d}{2}-1}\phi(a\,n + a\,e_i)\right\}.
\end{aligned}$$

The action S in Equation 1 must be dimensionless. Therefore the quantity $d^d x\, m^2\phi^2(x)$ must be dimensionless, and so $\phi(x)$ must have dimension $[l]^{-\frac{d}{2}}[m]^{-1} = [m]^{\frac{d}{2}-1}$. As $a^{\frac{d}{2}-1}$ has dimension $[l]^{\frac{d}{2}-1} = [m]^{1-\frac{d}{2}}$, $a^{\frac{d}{2}-1}\phi(x)$ is dimensionless. Relabelling

$$a\,e_i \to \mathbf{e}_i, \qquad a\,n \to \mathbf{x} \equiv \sum_{i=1}^{d} x_i \mathbf{e}_i, \qquad a^{\frac{d}{2}-1}\phi(x) \to \phi(\mathbf{x})$$

gives

$$S(\phi) = \sum_{\mathbf{x}} \phi(\mathbf{x})\left(\frac{2d + (am)^2}{2}\phi(\mathbf{x}) - \sum_{i=1}^{d}\phi(\mathbf{x} + \mathbf{e}_i)\right). \tag{3}$$

A.2 Non-Linear Sigma Model Symmetry Breaking

The non-linear sigma model action is given by

$$S(\boldsymbol{\phi}) = \frac{\beta}{2}\int d^d x\ \mathrm{Tr}\left[\partial_\mu \boldsymbol{\phi}^\dagger(x)\partial^\mu\boldsymbol{\phi}(x) - \lambda_0\boldsymbol{\phi}(x)\right]. \tag{6}$$

Consider the transformation $\boldsymbol{\phi}(x) \to U_L\boldsymbol{\phi}(x)U_R^\dagger$, where $U_L, U_R \in SU(2)$. Then

$$\begin{aligned}
S(\boldsymbol{\phi}) &\to \frac{\beta}{2}\int d^d x\ \mathrm{Tr}\left[U_R\left(\partial_\mu\boldsymbol{\phi}^\dagger(x)\right)U_L^\dagger U_L\left(\partial^\mu\boldsymbol{\phi}(x)\right)U_R^\dagger - \lambda_0 U_L\boldsymbol{\phi}(x)U_R^\dagger\right] \\
&= \frac{\beta}{2}\int d^d x\ \mathrm{Tr}\left[\partial_\mu\boldsymbol{\phi}^\dagger(x)\partial^\mu\boldsymbol{\phi}(x) - \lambda_0\boldsymbol{\phi}(x)U_R^\dagger U_L\right].
\end{aligned}$$

If $\lambda_0 = 0$ then the action is invariant under this transformation, i.e. an $SU(2)_L \times SU(2)_R$ chiral symmetry. If $\lambda_0 \neq 0$ then the action is invariant if $U_L = U_R$, i.e. an $SU(2)$ isospin symmetry.

The breaking of this chiral symmetry corresponds to the appearance of three massive pion fields. Writing $\boldsymbol{\phi}(x) = \sigma(x)\mathbb{I} + i\sum_{k=1}^{3}\pi_k(x)\tau_k$ gives

$$\begin{aligned}
S &= \frac{\beta}{2}\int dx\ \mathrm{Tr}\left[\partial_\mu\left(\sigma(x)\mathbb{I} - i\sum_{k=1}^{3}\pi_k(x)\tau_k\right)\partial^\mu\left(\sigma(x)\mathbb{I} + i\sum_{j=1}^{3}\pi_j(x)\tau_j\right)\right. \\
&\qquad\qquad \left. -\ \lambda_0\left(\sigma(x)\mathbb{I} + i\sum_{k=1}^{3}\pi_k(x)\tau_k\right)\right] \\
&= \beta\int dx\left[\partial_\mu\sigma(x)\partial^\mu\sigma(x) + \sum_{k=1}^{3}\partial_\mu\pi_k\partial^\mu\pi_k - \lambda_0\sigma(x)\right] \\
&\qquad\qquad (\mathrm{Tr}\,\tau_k = 0,\ \mathrm{Tr}(\tau_j\tau_k = 2\delta) = 2\delta_{jk},\ \mathrm{Tr}\,\mathbb{I} = 2) \\
&\approx \beta\int dx\left[\frac{\lambda_0}{2}\sum_{k=1}^{3}\pi_k\pi_k + \ldots\right], \qquad (\sigma(x) = \sqrt{1-\textstyle\sum_{k=1}^{3}\pi_k\pi_k} \approx 1 - \tfrac{1}{2}\textstyle\sum_{k=1}^{2}\pi_k\pi_k)
\end{aligned}$$

and so the mass of each pion field is given by $m_\pi^2 = \beta\lambda$.

A.3 Non-Linear Sigma Model Lattice Action

The non-linear sigma model action is given by

$$\begin{aligned}
S(\boldsymbol{\phi}) &= \frac{\beta}{2}\int d^dx\ \mathrm{Tr}\Big[\partial_\mu\boldsymbol{\phi}^\dagger(x)\partial^\mu\boldsymbol{\phi}(x) - \lambda_0\boldsymbol{\phi}(x)\Big]. \qquad (6)\\
&= -\frac{\beta}{2}\int dx\ \mathrm{Tr}\Big\{\boldsymbol{\phi}(x)\left[\partial_\mu\partial^\mu\boldsymbol{\phi}^\dagger(x) + \lambda_0\mathbb{I}\right]\Big\}. \qquad \text{(integrating by parts)}
\end{aligned}$$

Under the lattice parameterisation (Equation 2), this trnasforms into

$$\begin{aligned}
S(\boldsymbol{\phi}) &\to -\frac{\beta}{2}a^d\sum_n \mathrm{Tr}\left\{\boldsymbol{\phi}(a\,n)\left[\frac{1}{a^2}\sum_{i=1}^{d}\Big(\boldsymbol{\phi}^\dagger(a\,n + a\,e_i) + \boldsymbol{\phi}^\dagger(a\,n - a\,e_i) - 2\boldsymbol{\phi}^\dagger(a\,n)\Big) + \lambda_0\mathbb{I}\right]\right\} \\
&\simeq -\frac{a^{-1}\beta}{2}\sum_n \mathrm{Tr}\left\{a^{\frac{d-1}{2}}\boldsymbol{\phi}(a\,n)\left[a^{\frac{d-1}{2}}\sum_{i=1}^{d}\Big(\boldsymbol{\phi}^\dagger(a\,n + a\,e_i) + \boldsymbol{\phi}^\dagger(a\,n - a\,e_i)\Big) + a^{\frac{d+3}{2}}\lambda_0\mathbb{I}\right]\right\},
\end{aligned}$$

up to a constant term. Thus relabelling

$$\frac{\beta}{a}\to\beta, \qquad a^{\frac{d+3}{2}}\lambda_0\to\lambda_0, \qquad a^{\frac{d-1}{2}}\boldsymbol{\phi}(a\,n)\to\boldsymbol{\phi}(\mathbf{x})$$

gives

$$S(\boldsymbol{\phi}) = -\frac{\beta}{2}\sum_{\mathbf{x}}\mathrm{Tr}\left[\boldsymbol{\phi}(\mathbf{x})\left(\lambda_0\mathbb{I} + \sum_{i=1}^{d}\boldsymbol{\phi}^\dagger(\mathbf{x}+\mathbf{e}_i)\right)\right]. \qquad (9)$$

Straightforward dimension analysis (similar to what is done in A.1) shows that each of the relabelled values are dimensionless, as required.

B Two-point Correlation Function

For small δ, the non-linear sigma model two-point correlation function can be approximated as

$$c(\delta) \approx k_1 e^{-am\,\delta} + k_2, \tag{14}$$

i.e. an exponential decay with some constant scaling factor $k_1 > 0$ and shifting factor $k_2 \geqslant 0$. This expression can be manipulated to give an expression for am only in terms of $c(0)$, $c(1)$, and $c(2)$ as follows:

$$\begin{aligned}
c(0) \approx k_1 + k_2 \implies & c(\delta) \approx k_1 e^{-am\,\delta} + c(0) - k_1 \\
\implies & c(\delta) - c(0) \approx k_1 \left(e^{-am\,\delta} - 1\right) \\
\implies & \frac{c(2) - c(0)}{c(1) - c(0)} \approx \frac{e^{-2am} - 1}{e^{-am} - 1} \\
\implies & \left[c(2) - c(0)\right]\left(e^{am} - e^{2am}\right) \approx \left[c(1) - c(0)\right]\left(1 - e^{2am}\right) \\
\implies & e^{2am}\left[c(1) - c(2)\right] - e^{am}\left[c(0) - c(2)\right] + \left[c(0) - c(1)\right] \approx 0 \\
\implies & am \approx \ln\left\{ \frac{c(0) - c(2) \pm \sqrt{\left[c(0) - c(2)\right]^2 - 4\left[c(1) - c(2)\right]\left[c(0) - c(1)\right]}}{2\left[c(1) - c(2)\right]} \right\}
\end{aligned} \tag{15}$$

By considering different combinations of k_1, k_2, and am, it can be shown that the $\pm$ in the above expression should read $+$ for $k_1 > 0$, and $-$ for $k_1 < 0$. Thus an estimator $\widehat{am}$ for the mass parameter am can be given by

$$\widehat{am} = \ln\left\{ \frac{\hat{c}(0) - \hat{c}(2) + \sqrt{\left[\hat{c}(0) - \hat{c}(2)\right]^2 - 4\left[\hat{c}(1) - \hat{c}(2)\right]\left[\hat{c}(0) - \hat{c}(1)\right]}}{2\left[\hat{c}(1) - \hat{c}(2)\right]} \right\}. \tag{45}$$

As this estimator depends on the means $\hat{c}(\delta)$ of random variables $c(\delta)$, the corresponding mean squared error $\mathrm{MSE}(\widehat{am})$ can be approximated using the jackknife method.

C Statistical Proofs

Lemma 1. *For identically distributed random variables $A_1, \ldots, A_N$ with mean μ, the sample mean $\hat{\mu}$ (Equation 16) is an unbiased estimator of μ.*

Proof.

$$\begin{aligned}
\mathrm{E}[\hat{\mu}] &= \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N} A_i\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[A_i] \\
&= \frac{N\mu}{N} \\
&= \mu.
\end{aligned}$$

□

Lemma 2. *For identically distributed random variables $A_1, \ldots, A_N$ with mean μ and variance ς, the mean squared error $\mathrm{MSE}(\hat{\mu})$ of the sample mean $\hat{\mu}$ is given by*

$$\mathrm{MSE}(\hat{\mu}) = \frac{\varsigma}{N}. \tag{18}$$

Proof.

$$\begin{aligned}
\mathrm{MSE}(\hat{\mu}) &= \mathrm{E}\Big[(\hat{\mu} - \mu)^2\Big] \\
&= \mathrm{Var}(\hat{\mu}) && \text{(as } \mathrm{E}[\hat{\mu}] = \mu\text{)} \\
&= \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} A_i\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}(A_i) && \text{(by independence)} \\
&= \frac{N\varsigma}{N^2} \\
&= \frac{\varsigma}{N}
\end{aligned}$$

□

Lemma 3. *For independent identically distributed random variables $A_1, \ldots, A_N$ with mean μ and variance $\varsigma \equiv \mathrm{Var}(A_i)$, the sample variance $\hat{\varsigma}$ (Equation 16) is an unbiased estimator of ς.*

Proof.

$$\begin{aligned}
\mathrm{E}[\hat{\varsigma}] &= \mathrm{E}\left[\frac{1}{N-1}\sum_{i=1}^{N}(A_i - \hat{\mu})\right] \\
&= \frac{1}{N-1}\,\mathrm{E}\left[\sum_{i=1}^{N}\left(A_i^2 + \hat{\mu}^2 - 2\hat{\mu}A_i\right)\right] \\
&= \frac{1}{N-1}\,\mathrm{E}\left[\sum_{i=1}^{N} A_i^2 + N\hat{\mu}^2 - 2N\hat{\mu}\,\frac{1}{N}\sum_{i=1}^{N} A_i\right] \\
&= \frac{1}{N-1}\,\mathrm{E}\left[\sum_{i=1}^{N} A_i^2 + N\hat{\mu}^2 - 2N\hat{\mu}^2\right] \\
&= \frac{1}{N-1}\left(\sum_{i=1}^{N}\mathrm{E}\big[A_i^2\big] - N\,\mathrm{E}\big[\hat{\mu}^2\big]\right) \\
&= \frac{N}{N-1}\left(\mathrm{E}\big[A_i^2\big] - \mathrm{E}\big[\hat{\mu}^2\big]\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(A_i) &= \mathrm{E}\big[A_i^2\big] - \mathrm{E}[A_i]^2 \\
\implies \mathrm{E}\big[A_i^2\big] &= \mathrm{Var}(A_i) + \mathrm{E}[A_i]^2 \\
&= \varsigma + \mu^2
\end{aligned}
\qquad
\begin{aligned}
\mathrm{Var}(\hat{\mu}) &= \mathrm{E}\big[\hat{\mu}^2\big] - \mathrm{E}[\hat{\mu}]^2 \\
\implies \mathrm{E}\big[\hat{\mu}^2\big] &= \mathrm{Var}(\hat{\mu}) + \mathrm{E}[\hat{\mu}]^2 \\
&= \frac{\varsigma}{N} + \mu^2 \qquad \text{(shown in Lemma 2)}
\end{aligned}$$

$$\begin{aligned}
\implies \mathrm{E}[\hat{\varsigma}] &= \frac{N}{N-1}\left(\varsigma + \mu^2 - \frac{\varsigma}{N} - \mu^2\right) \\
&= \frac{1}{N-1}\left(N\varsigma - \varsigma\right) \\
&= \varsigma.
\end{aligned}$$

□

Lemma 4. *For a Markov process with transition probability density function $p(\xi, \nu)$ satisfying detailed balance for a probability density function $\lambda(\xi)$ (Equation 24), $\lambda(\xi)$ is the equilibrium distribution of $p(\xi, \nu)$.*

Proof.

$$\begin{aligned}\int d\Gamma_\nu \, p(\xi,\nu)\lambda(\nu) &= \int d\Gamma_\nu \, p(\nu,\xi)\lambda(\xi) \\ &= \lambda(\xi) \int d\Gamma_\nu \, p(\nu,\xi) \\ &= \lambda(\xi).\end{aligned}$$

□

D Distribution Notation

In this project, the following notation is used for some common probability distributions.

The continuous uniform distribution $\mathcal{U}(a, b)$ has probability density function

$$f_{\mathcal{U}}(x) = \begin{cases} \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0 & \text{otherwise} \end{cases},$$

and can be related to the standard uniform distribution $\mathcal{U}(0, 1)$ by

$$\mathcal{U}(a, b) = a + (b - a)\,\mathcal{U}(0, 1).$$

The normal distribution $\mathcal{N}(\mu, \sigma^2)$ has probability density function

$$f_{\mathcal{N}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

and can be related to the standard normal distribution $\mathcal{N}(0, 1)$ by

$$\mathcal{N}(\mu, \sigma) = \mu + \sigma\,\mathcal{N}(0, 1).$$

The exponential distribution $\mathrm{Exp}(\lambda)$ has probability density function

$$f_{\mathrm{Exp}}(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}.$$

E Klein-Gordon Field Sampling

The Klein-Gordon action (Equation 3) can be split up into $S_{\mathbf{x}}(\phi)$, the contribution to the action of the field $\phi(\mathbf{x})$ at the site $\mathbf{x}$, and $T_{\mathbf{x}}(\phi) \equiv S(\phi) - S_{\mathbf{x}}(\phi)$, the action to which $\mathbf{x}$ has no contribution. In particular,

$$\begin{aligned} S_{\mathbf{x}}(\phi) &\equiv \phi(\mathbf{x}) \left(\frac{\phi(\mathbf{x})}{2\kappa^2} - \gamma(\mathbf{x}) \right) \\ &= \frac{(\phi(\mathbf{x}))^2 - 2\kappa^2 \gamma(\mathbf{x}) \phi(\mathbf{x})}{2\kappa^2} \\ &= \frac{\big(\phi(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})\big)^2 - \kappa^4 (\gamma(\mathbf{x}))^2}{2\kappa^2} \\ &= \frac{1}{2} \left(\frac{\phi(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})}{\kappa} \right)^2 - \frac{\kappa^2 (\gamma(\mathbf{x}))^2}{2}. \end{aligned}$$

This gives

$$\begin{aligned} \lambda(\phi) &\propto e^{-S(\phi)} \\ &= e^{-S_{\mathbf{x}}(\phi)} e^{-T_{\mathbf{x}}(\phi)} \\ &= \exp\left[-\frac{1}{2} \left(\frac{\phi(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})}{\kappa} \right)^2 \right] \exp\left[\frac{\kappa^2 (\gamma(\mathbf{x}))^2}{2} - T_{\mathbf{x}}(\phi) \right]. \end{aligned} \tag{26}$$

For an algorithm that updates each site of the lattice in turn as in Equation 31, then only the distribution

$$\lambda_{\mathbf{x}}(\phi) \propto \exp\left[-\frac{1}{2} \left(\frac{\phi(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})}{\kappa} \right)^2 \right] \tag{46}$$

needs to be considered at each step, as this is the only part of $\lambda(\phi)$ depending on $\phi(\mathbf{x})$.

For the Metropolis-Hastings algorithm, proposing a field differing at a single site $\mathbf{x}$ as in Equation 31 results in

$$\frac{\lambda(\phi')}{\lambda(\phi)} = \frac{\lambda_{\mathbf{x}}(\phi')}{\lambda_{\mathbf{x}}(\phi)} = \exp\left[-\frac{(\psi'(\mathbf{x}))^2 - (\psi(\mathbf{x}))^2}{2\kappa^2} \right], \tag{47}$$

where $\psi(\mathbf{x}) \equiv \phi(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})$ and $\psi'(\mathbf{x}) \equiv \phi'(\mathbf{x}) - \kappa^2 \gamma(\mathbf{x})$. If a symmetric proposal density function $q(\phi', \phi) = q(\phi, \phi')$ is chosen, then by Equation 30, the new state is always accepted if $|\psi'(\mathbf{x})| \leqslant |\psi(\mathbf{x})|$, and accepted with probability as in Equation 47 otherwise.

For the Gibbs sampler, updating a single lattice site as before, the proposal density function $q(\phi', \phi)$ is given by Equation 46, and so each site is simply set to an element drawn from the normal distribution $\mathcal{N}\big(\kappa^2 \gamma(\mathbf{x}), \kappa^2\big)$.

F Non-Linear Sigma Model Sampling

F.1 Sampling from $f(u)$

Consider the distribution [15]

$$f(u)\,du \propto \exp\Big[\frac{\rho}{2}\operatorname{Tr}\Big(u\Pi^\dagger\Big)\Big]\,du \tag{48}$$

for some $\rho \in \mathbb{R}$, $\Pi \in SU(2)$, and where the Haar measure du for $u \equiv \{u_0, u_1, u_2, u_3\} \in SU(2)$ (Equation 7) is given in terms of the Euclidean metric d^4u as

$$du \propto d^4u\,\delta\big(u_0^2 + u_k u_k - 1\big)\,.$$

By the invariance of the Haar measure, this can be written as

$$f(v)\,dv \propto \exp\Big[\frac{\rho}{2}\operatorname{Tr} v\Big]\,dv,$$

where $v \equiv u\Pi^\dagger \in SU(2)$. It is convenient to write this density in terms of the density functions of each of v_i. The probability density for v_0 is given by

$$\begin{aligned}
f_0(v_0) &= \int dw\,\delta(w_0 - v_0) \\
&\propto \int d^4w\,\delta(w_0 - v_0)\,\delta\big(w_0^2 + w_k w_k - 1\big) \\
&= \int d^3w\,\delta\big(v_0^2 + w_k w_k - 1\big) \\
&= \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \int_0^{\infty} dr\, r^2\,\delta\big(v_0^2 + r^2 - 1\big) \\
&\propto \int_0^{\infty} dr\, r^2\,\delta\big(v_0^2 + r^2 - 1\big) \\
&\propto \sqrt{1 - v_0^2}.
\end{aligned}$$

This then gives

$$f(v)\,dv \propto \sqrt{1 - v_0^2}\,e^{\rho v_0}\delta\big(n^2 - 1\big)\,dv_0 d^3n, \tag{43}$$

where $v_i = n_i\sqrt{1 - v_0^2}$, $i = 1, 2, 3$. Using a change of variables $y \equiv \rho(1 - v_0)$ leads to the density function for y

$$f_Y(y) \propto \sqrt{2 - \frac{y}{\rho}}\sqrt{y}\,e^{-y}. \tag{49}$$

By the rejection method [24], sampling from $f_Y(y)$ can be achieved by generating Y from

$$g_{\text{trial}}(y) \propto \sqrt{y}\,e^{-y} \tag{50}$$

and $U \sim \mathcal{U}(0, 1)$, and accepting Y if $U \leqslant \sqrt{\dfrac{2 - \frac{Y}{\rho}}{2}}$. Sampling from $g_{\text{trial}}(y)$ is equivalent to generating $A \sim \mathcal{N}\big(0, \frac{1}{2}\big)$ and $B \sim \text{Exp}(1)$, and setting $Y = A^2 + B$ [25]. Once v_0 is generated via $f_Y(y)$, then by Equation 43, v_1, v_2, v_3 can be obtained by generating n_1, n_2, n_3 uniformly on the unit sphere. This can be done by generating $n_1 \sim \mathcal{U}(-1, 1)$ and $\theta \sim \mathcal{U}(0, 2\pi)$, setting $n_2 = \sqrt{1 - n_1^2}\cos\theta$ and $n_3 = \sqrt{1 - n_1^2}\sin\theta$, and letting $v_i = n_1\sqrt{1 - v_0^2}$. Finally, once v is generated from $f(v)$, u can be calculated from $v = u\Pi^\dagger \iff u = v\Pi$.

F.2 Proving $\frac{\mathbf{\Sigma(x)}}{\sqrt{\det \mathbf{\Sigma(x)}}} \in SU(2)$

Lemma 5. *For a sum*

$$\Sigma \equiv \sum_{i=1}^{n} U_i$$

of elements $U_i \in SU(2)$, then the quantity

$$\frac{\Sigma}{\sqrt{\det \Sigma}}$$

is also an element of $SU(2)$.

Proof. Clearly

$$\det \frac{\Sigma}{\sqrt{\det \Sigma}} = \frac{\det \Sigma}{\left(\sqrt{\det \Sigma}\right)^2} = 1,$$

and so it only remains to show that $\frac{\Sigma}{\sqrt{\det \Sigma}}$ is unitary. Denote

$$\Sigma^{(n)} = \sum_{i=1}^{n} U_i, \qquad\qquad U_i = u_i^{(0)}\mathbb{I} + i\sum_{j=1}^{3} u_i^{(j)}\sigma_j.$$

Consider $n = 1$. Then

$$\begin{aligned} \Sigma^{(1)}\Sigma^{(1)\dagger} &= U_1 U_1^\dagger \\ &= \mathbb{I} \\ \implies \frac{\Sigma^{(1)}}{\sqrt{\det \Sigma^{(1)}}}\left(\frac{\Sigma^{(1)}}{\sqrt{\det \Sigma^{(1)}}}\right)^\dagger &= \frac{1}{\det \Sigma^{(1)}}\mathbb{I} \\ &= \frac{1}{\det U_1}\mathbb{I} \\ &= \mathbb{I}, \end{aligned}$$

and so $\frac{\Sigma^{(n)}}{\sqrt{\det \Sigma^{(n)}}}$ is unitary for $n = 1$.

Now consider $n = k + 1$, assuming $\frac{\Sigma^{(k)}}{\sqrt{\det \Sigma^{(k)}}}$ is unitary.

$$\begin{aligned} \Sigma^{(k+1)}\Sigma^{(k+1)\dagger} &= \left(\sum_{i=1}^{k+1} U_i\right)\left(\sum_{j=1}^{k+1} U_j\right)^\dagger \\ &= \left(U_{k+1} + \sum_{i=1}^{k} U_i\right)\left(U_{k+1}^\dagger + \sum_{j=1}^{k} U_j^\dagger\right) \\ &= U_{k+1}U_{k+1}^\dagger + U_{k+1}\sum_{i=1}^{k} U_j^\dagger + \sum_{i=1}^{k} U_i U_{k+1}^\dagger + \det\left(\Sigma^{(k)}\right)\mathbb{I} \end{aligned} \tag{51}$$

$$U_{k+1}U_{k+1}^\dagger = \sum_{j=0}^{3} {u_{k+1}^{(j)}}^2 \mathbb{I} \tag{52}$$

$$
\begin{aligned}
U_{k+1}\sum_{i=1}^{k}U_i^\dagger &= \left(u_{k+1}^{(0)}\mathbb{I}+i\sum_{j=1}^{3}u_{k+1}^{(j)}\sigma_j\right)\sum_{i=1}^{k}\left(u_i^{(0)}\mathbb{I}-i\sum_{l=1}^{3}u_i^{(l)}\sigma_l\right)\\
&= \sum_{i=1}^{k}\left[u_{k+1}^{(0)}u_i^{(0)}\mathbb{I}+i\sum_{j=1}^{3}\left(u_{k+1}^{(j)}u_i^{(0)}-u_{k+1}^{(0)}u_i^{(j)}\right)\sigma_j+\sum_{j,l=1}^{3}u_{k+1}^{(j)}u_i^{(l)}\sigma_j\sigma_l\right]\\
&= \sum_{i=1}^{k}\left[u_{k+1}^{(0)}u_i^{(0)}\mathbb{I}+i\sum_{j=1}^{3}\left(u_{k+1}^{(j)}u_i^{(0)}-u_{k+1}^{(0)}u_i^{(j)}\right)\sigma_j\right.\\
&\qquad\left.+\sum_{j,l=1}^{3}u_{k+1}^{(j)}u_i^{(l)}\left(\delta_{jl}\mathbb{I}+\sum_{m=1}^{3}i\epsilon_{jlm}\sigma_m\right)\right]\\
&= \sum_{i=1}^{k}\left[\sum_{j=0}^{3}u_{k+1}^{(j)}u_i^{(j)}\mathbb{I}+i\sum_{j=1}^{3}\left(u_{k+1}^{(j)}u_i^{(0)}-u_{k+1}^{(0)}u_i^{(j)}\right)\sigma_j+i\sum_{\substack{j,l,m=1\\ j\neq l}}^{3}u_{k+1}^{(l)}u_i^{(l)}\epsilon_{jlm}\sigma_m\right]
\end{aligned}
\tag{53}
$$

Similarly,

$$
\sum_{i=1}^{k}U_iU_{k+1}^\dagger=\sum_{i=1}^{k}\left[\sum_{j=0}^{3}u_{k+1}^{(j)}u_i^{(j)}\mathbb{I}-i\sum_{j=1}^{3}\left(u_{k+1}^{(j)}u_i^{(0)}-u_{k+1}^{(0)}u_i^{(j)}\right)\sigma_j-i\sum_{\substack{j,l,m=1\\ j\neq l}}^{3}u_{k+1}^{(l)}u_i^{(l)}\epsilon_{jlm}\sigma_m\right]
\tag{54}
$$

$$
(53),(54)\implies U_{k+1}\sum_{i=1}^{k}U_j^\dagger+\sum_{i=1}^{k}U_iU_{k+1}^\dagger=2\sum_{j=0}^{3}\sum_{i=1}^{k}u_{k+1}^{(j)}u_i^{(j)}\mathbb{I}
\tag{55}
$$

$$
\det\left(\Sigma^{(k)}\right)=\sum_{j=0}^{3}\left(\sum_{i=1}^{k}u_i^{(j)}\right)^2
\tag{56}
$$

Combining (51), (52), (55) and (56) gives

$$
\begin{aligned}
\Sigma^{(k+1)}\Sigma^{(k+1)\dagger} &= \sum_{j=0}^{3}\left[{u_{k+1}^{(j)}}^2+2\sum_{i=1}^{k}u_{k+1}^{(j)}u_i^{(j)}+\left(\sum_{i=1}^{k}u_i^{(j)}\right)^2\right]\mathbb{I}\\
&= \sum_{j=0}^{3}\left({u_{k+1}^{(j)}}^2+2\sum_{i=1}^{k}u_{k+1}^{(j)}u_i^{(j)}+\sum_{i=1}^{k}u_i^{(j)2}+\sum_{\substack{i,l=1\\ i\neq l}}^{k}u_i^{(j)}u_l^{(j)}\right)\mathbb{I}\\
&= \sum_{j=0}^{3}\left(\sum_{i=1}^{k+1}u_i^{(j)2}+\sum_{\substack{i,l=1\\ i\neq l}}^{k+1}u_i^{(j)}u_l^{(j)}\right)\mathbb{I}\\
&= \sum_{j=0}^{3}\left(\sum_{i=1}^{k+1}u_i^{(j)}\right)^2\mathbb{I}\\
&= \det\left(\Sigma^{(k+1)}\right)\mathbb{I}\implies\frac{\Sigma^{(k+1)}}{\sqrt{\det\Sigma^{(k+1)}}}\left(\frac{\Sigma^{(k+1)}}{\sqrt{\det\Sigma^{(k+1)}}}\right)^\dagger=\mathbb{I}.
\end{aligned}
$$

Thus by induction, $\frac{\Sigma}{\sqrt{\det\Sigma}}\in SU(2)$. □

F.3 Relating $f(u)$ to $\lambda(\boldsymbol{\phi})$

As is done in E, the non-linear sigma model action (Equation 9) can be split up into $S_{\mathbf{x}}(\boldsymbol{\phi})$ and $T_{\mathbf{x}} \equiv S(\boldsymbol{\phi}) - S_{\mathbf{x}}(\boldsymbol{\phi})$ with

$$S_{\mathbf{x}}(\boldsymbol{\phi}) \equiv -\frac{\beta}{2} \operatorname{Tr}\Big(\boldsymbol{\phi}(\mathbf{x})\boldsymbol{\Sigma}^{\dagger}(\mathbf{x})\Big),$$

giving

$$\begin{aligned} \lambda(\boldsymbol{\phi}) &\propto e^{-S(\boldsymbol{\phi})} \\ &= \exp\left[\frac{\beta}{2} \operatorname{Tr}\Big(\boldsymbol{\phi}(\mathbf{x})\boldsymbol{\Sigma}^{\dagger}(\mathbf{x})\Big)\right] e^{-T_{\mathbf{x}}(\boldsymbol{\phi})}. \end{aligned} \tag{26}$$

Again, only the distribution

$$\begin{aligned} \lambda_{\mathbf{x}}(\boldsymbol{\phi}) &\propto \exp\left[\frac{\beta}{2} \operatorname{Tr}\Big(\boldsymbol{\phi}(\mathbf{x})\boldsymbol{\Sigma}^{\dagger}(\mathbf{x})\Big)\right] \\ &= \exp\left[\frac{\beta\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})}}{2} \operatorname{Tr}\left(\boldsymbol{\phi}(\mathbf{x})\frac{\boldsymbol{\Sigma}^{\dagger}(\mathbf{x})}{\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})}}\right)\right] \end{aligned} \tag{57}$$

needs to be considered for an algorithm updating each site in turn.

Comparing Equation 48 and Equation 57, generating $\boldsymbol{\phi}'(\mathbf{x})$ from $\lambda_{\mathbf{x}}(\boldsymbol{\phi}')$ using the Gibbs sampler is equivalent to generating u from $f(u)$ with $\rho = \beta\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})} \in \mathbb{R}$ and $\Pi = \frac{\boldsymbol{\Sigma}(\mathbf{x})}{\sqrt{\det \boldsymbol{\Sigma}(\mathbf{x})}} \in SU(2)$.