

A study of thermal NRQCD with machine learning methods

SAMUEL PETER OFFLER

Department of Physics
Swansea University

*Submitted to Swansea University in fulfilment of the requirements for the Degree of
Doctor of Philosophy*



**Prifysgol Abertawe
Swansea University**

Declarations

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed..........

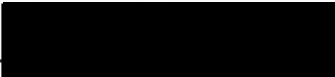
Date.....22/06/2022.....

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed........

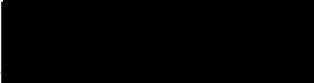
Date.....22/06/2022.....

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed..........

Date.....22/06/2022.....

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed..........

Date.....22/06/2022.....

Abstract

The aim of this thesis is to develop a machine learning model capable of the spectral reconstruction of Euclidean lattice correlators at finite temperature. The early part of this thesis is dedicated to a review of the QCD phase diagram and correlation functions to establish the relationship between the Euclidean correlator and spectral function. An analysis of FASTSUM ensembles of Euclidean correlators is performed to determine effective masses and thermal modification for bottomonium states. An initial model using Kernel Ridge Regression is examined and implemented for the Υ state. The latter part of this thesis focuses on improving the generation of training data for the machine learning method and the machine learning method itself. This work concludes with the implementation of the Kernel Ridge Regression for a variety of bottomonium states.

Acknowledgements

There are numerous people I would like to offer my thanks to, without whom this work would not have been possible:

To my supervisor Gert Aarts for all his support and guidance over the last 4 years. His knowledge and advice has been invaluable for work presented in this thesis. To other members of the FASTSUM collaboration, especially Chris Allton, Simon Hands, and Seyong Kim, whose help has been indispensable when my own understanding of the subject has been lacking.

To my family, and in particular my parents, Teri and Dave, I would like to thank them for their ongoing support and attempts at understanding what I have done. And to my Granddad Peter, who always wanted a Doctor in the family.

To Jonas Gleasen, whose code has formed the basis for much of my own, and Ed Bennett for his expertise.

And lastly to the many friends from the office, especially Ben Heath who has always kept the office interesting with his various surprises.

Acronyms

QCD - Quantum Chromodynamics

QGP - Quark-Gluon Plasma

NRQCD - Non-Relativistic Quantum Chromodynamics

QED - Quantum Electrodynamics

LLS - Linear-Least-Squares

KRR - Kernel Ridge Regression

MAE - Mean Absolute Error

MSE - Mean-Squared-Error

SVD - Singular-Value-Decomposition

Conventions

$\hbar = c = 1$.

$\log(x)$ refers to the natural logarithm. The base will shown explicitly if it differs.

Contents

Introduction	1
1 QCD at nonzero temperature	3
1.1 QCD phase diagram	3
1.2 Phase Transitions	4
1.2.1 Chiral symmetry	4
1.2.2 Center symmetry	6
1.3 Probing the Quark-Gluon Plasma	9
1.4 FASTSUM Gen2L ensembles	10
2 Spectral Reconstruction: An Ill-posed Problem	12
2.1 Motivation	12
2.2 Derivation of Green's functions	16
2.2.1 Kubo-Martin-Schwinger condition	18
2.2.2 Euclidean Correlator	19
3 NRQCD	22
3.1 Lattice NRQCD	22
3.2 Properties of the system from Correlators	25
3.2.1 Effective mass	26
3.2.2 Higher order masses	27
3.2.3 Thermal Modification	34
4 Introduction to Kernel Ridge Regression	36
4.1 Regression methods	36
4.1.1 Linear regression	36
4.1.2 Ridge Regression	38
4.1.3 Kernel Ridge Regression	40
4.2 Data Generation	44
4.2.1 Spectral functions	44
4.2.2 Representations of spectral functions	46
4.3 A first look at KRR	49
4.3.1 Regularization constant	52

4.3.2	Scale Constant	53
4.3.3	Grid Search	61
4.4	First prediction of NRQCD data	62
4.5	Alternative methods	64
4.5.1	Ansatz	64
4.5.2	Backus-Gilbert	64
4.5.3	Maximum Entropy Method	66
5	Data generation revisited	68
5.1	Examination of current data	68
5.2	Data generation with NRQCD influence	74
5.3	Alternative datasets	77
5.3.1	Alternative parameter distributions	77
5.3.2	SVD coefficients	77
5.4	NRQCD correlator noise	85
6	Improvements to the KRR Model	87
6.1	Initial Testing	87
6.2	Attempts for improvements	90
6.2.1	Scaling of Θ	90
6.2.2	Scaling of correlators	92
6.2.3	Training Set Size	94
6.2.4	Cross-validation	94
6.3	Application to NRQCD	101
	Conclusion	110
A	Lattice formulation	112
A.1	Path integral of a 1D system	112
A.2	Relations to statistical mechanics	114
B		115
B.1	Bootstrap sampling	115
	Bibliography	122

Introduction

The theory of Quantum Chromodynamics (QCD) focuses solely on the strongly interacting matter that is quarks and gluons. Since its advent in the 1970s [1], there have been numerous methods developed to investigate it. In particular the work of Wilson helped with the development of Lattice QCD [2]. This is a non-perturbative approach in which spacetime has been discretized into a lattice, so that the path integral [3] of the theory can be evaluated by methods such as Monte Carlo. There is an ongoing struggle to study QCD theoretically due to issues such as the ill-posed problem of spectral reconstruction or the sign problem. The latter of these restricts our ability to investigate the theory at finite baryon chemical potential. However we are able to study the theory at non-zero temperature which can be compared to experimental studies involving heavy-ion collisions. In this thesis we aim to develop a machine learning model to reconstruct spectral functions at finite temperature for further study. This thesis is organised as follows:

In Chapter 1 we will examine the phase diagram of QCD, focusing on the different states of QCD for zero density. These can be explored experimentally using heavy-ion collisions. In the latter half of this chapter we go into further detail of the two states, those being hadrons and quark-gluon plasma (QGP). From this, we can begin to focus on the bound states of $c\bar{c}$ and $b\bar{b}$ mesons, or charmonium and bottomonium respectively. These are of particular interest as they can be used as a thermometer for the QGP. To conclude the chapter, the details of the ensembles generated by the FASTSUM collaboration will be given. From previous work, we are able to determine the transition temperature between hadrons and QGP.

In Chapter 2 we give an introduction to thermal field theory. By making use of the thermal expectation value and time evolution, we demonstrate how a variety of correlation functions can be developed. Whilst two-point correlations contain the detailed information of a system, it is often difficult to access. The spectral function is far more accessible. Hence we will make a point of showing how some of these correlation functions are related to the spectral function. Finally we will discuss the ill-posed problem of spectral reconstruction where the correlation function is in a discrete form, as it is for Lattice QCD.

In Chapter 3 we return to bottomonium in the context of Non-Relativistic QCD (NRQCD). We begin with an explanation of NRQCD as an effective field theory at non-zero temperature. Initially this will be done in the continuum limit before moving to the lattice equivalent. Following this, we obtain our first results of this work. Before any attempts of spectral reconstruction, we use the correlation functions to determine effective masses. After comparisons to experimental values, we perform multi-exponential fits. This is done using the CORFITTER package for the purpose of estimating the masses of first excited states. The chapter then concludes with an examination of the thermal dependence of S and P-waves states.

In Chapter 4 we develop an initial Kernel Ridge Regression (KRR) model that will be used as the basis for later chapters. Starting from simple linear regression, we build up our KRR model making use of analogies to the initial example. Using the training data, a systematic investigation of the KRR is done to observe the effects of altering the hyperparameters. We conclude this chapter with an initial attempt at reconstructing the spectral function for the Υ at different temperatures.

In Chapter 5 we reconsider the generation of the training data. Making use of the improved dispersion relation, we reduce the range of energy for the spectral functions. Next, we alter the form of the training data and compare to the Gen 2L ensembles to find similar correlation functions. Using these similar functions from the mock data, a new dataset is generated using the corresponding mock spectral functions as a foundation. Finally we use the ensembles to estimate noise so that we may incorporate this into the KRR model.

In Chapter 6 we revisit the KRR model, armed with our new data, and propose methods to further improve the model. This includes rescaling of both input and target data, and the effect of varying training sample size. We also introduce cross-validation for determining optimal hyperparameters. The thesis concludes with a prediction of groundstate masses and widths, and spectral functions for various bottomonium states.

The work presented in this thesis focuses developing a method for the spectral reconstruction of FASTSUM Gen 2L bottomonium correlators. For those unfamiliar with lattice field theory, see Appendix A for a brief introduction to the subject.

Chapter 1

QCD at nonzero temperature

1.1 QCD phase diagram

Since the advent of Quantum Chromodynamics (QCD) in the 1970s [1], there have been numerous methods developed to study it. This can be experimentally by using heavy-ion collisions, or theoretically with perturbation theory or lattice gauge theory. Each of these methods offers its own insight into this theory, and by combining these we may develop the theory as a whole.

Despite these efforts, there is still much of the theory that is unknown or simply conjecture. The current proposed phase diagram is shown in Figure 1.1. There is no single theoretical method can be used to explore the phase diagram, and modern colliders lack the capabilities to provide insight for all regions. The QCD phase diagram shown in Figure 1.1 is a function of the temperature T and baryon chemical potential μ_B . Although Figure 1.1 contains several interesting features, we will only be investigating the vertical axis in this work. For vanishing μ_B quarks and gluons can exist in two phases: hadrons, and quark-gluon plasma (QGP). The properties of the QGP are the subject of ongoing investigations by heavy-ion collisions at RHIC at Brookhaven National Laboratory and ALICE at the LHC. In particular the RHIC has been used to search for the critical point shown in Figure 1.1 [4, 5].

It is difficult to study finite chemical potential using Lattice QCD due to the *sign problem*. This problem arises when the action for QCD becomes complex as is the case for non-zero μ_B . Although it is difficult to probe this region with lattice QCD, various methods have been developed to do so. These include the imaginary μ_B approach and complex Langevin [7, 8]. However it is due to studies of finite μ_B that a first-order phase transition is thought to exist [9].

For physical quark mass (m_q), lattice simulations have shown that this transition is an analytical crossover [10–12]. Therefore we can say that the phases of Hadrons

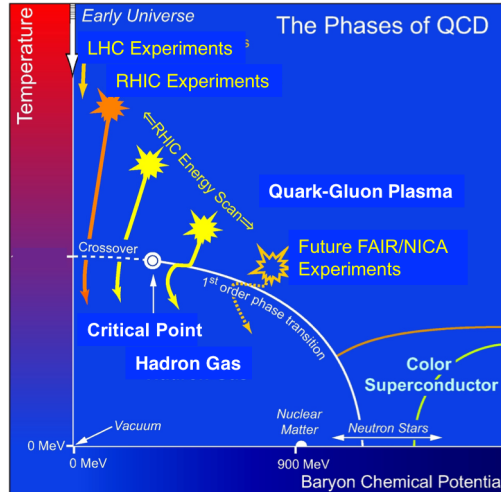


Figure 1.1: Sketch of the Phase diagram of QCD. [6]

and QGP are analytically connected, and hence there is no true order parameter. Instead, we use observables such as the chiral condensate and Polyakov loop to estimate a pseudocritical temperature, T_{pc} [13–15]. Due to this transition being an analytical crossover, these observables vary continuously between the two phases. In the case of a first-order phase transition there is a sudden change in the value of the order parameter. This result leads to the expectation of the first-order transition line that must end with a critical point before the temperature axis.

1.2 Phase Transitions

Although there is no first or second-order transition for the physical m_q at zero μ_B , studies have shown that transitions could exist for alternative values of m_q . In the following, we will show how two symmetries are explicitly broken by physical quarks and how a choice of non-physical quark masses can restore these.

1.2.1 Chiral symmetry

The first symmetry to be considered is chiral symmetry. In order to demonstrate how chiral symmetry is broken by physical quarks, we introduce the continuum Euclidean QCD action,

$$S_{QCD} = \int d^4x \bar{\psi} (\not{D} + m_q) \psi + \frac{1}{2g^2} \text{Tr} [F_{\mu\nu} F^{\mu\nu}], \quad (1.2.1)$$

where $\psi/\bar{\psi}$ are the quark and antiquark fields respectively, \not{D} is the covariant derivative, g is the coupling strength, and $F_{\mu\nu}$ is the field strength tensor. The covariant derivative is given by,

$$\not{D} = \gamma_\mu (\partial_\mu + iA_\mu), \quad (1.2.2)$$

where A_μ is the gluon field. The quark and gluon fields carry several indices although only the Lorentz index μ has been shown here. Both quark and gluon field have a space-time argument, typically denoted by x . The quark fields carry a single colour index, whilst the gluon field carries two. The colour indices of the quark fields are summed over with the corresponding indices of A_μ , thus coupling the quarks with the gluon field. Note that the trace in the second term of Eq.(1.2.1) is over the colour indices. The gluon field for a given x and μ , $A_\mu(x)$, is represented by a traceless, hermitian, 3×3 matrix. The quark fields can also have a flavour index. Finally, the quark fields and γ_μ have Dirac indices, with each of the quark fields having a single Dirac index that corresponds to one of the two indices carried by γ_μ .

Although the work presented in this thesis uses lattice QCD, we used the continuum Lagrangian here for simplicity. Furthermore, we would expect lattice results to agree with the continuum as the lattice spacing vanishes. We focus on the first term of Eq. (1.2.1), as $F_{\mu\nu}$ remain unchanged by a chiral rotation. Performing a chiral rotation transforms the quark and antiquark fields, ψ and $\bar{\psi}$ respectively, as follows,

$$\begin{aligned} \psi &\rightarrow \psi' = e^{i\alpha\gamma_5}\psi \\ \bar{\psi} &\rightarrow \bar{\psi}' = \bar{\psi}e^{i\alpha\gamma_5} \end{aligned} \quad (1.2.3)$$

where α is a constant, real parameter and γ_5 is the chirality matrix acting in Dirac space. In Euclidean space $\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$, where $\gamma_{1,2,3,4}$ are the Euclidean Dirac matrices. The covariant derivative transforms as follows,

$$\begin{aligned} \bar{\psi}'\not{D}\psi &\rightarrow \bar{\psi}'\gamma_\mu (\partial_\mu + iA_\mu) \psi' \\ &= \bar{\psi}e^{i\alpha\gamma_5}\gamma_\mu (\partial_\mu + iA_\mu) e^{i\alpha\gamma_5}\psi \\ &= \bar{\psi}e^{i\alpha\gamma_5}e^{-i\alpha\gamma_5}\gamma_\mu (\partial_\mu + iA_\mu) \psi \\ &= \bar{\psi}\gamma_\mu (\partial_\mu + iA_\mu) \psi, \end{aligned} \quad (1.2.4)$$

where we have made use of the $\gamma_\mu\gamma_5 = -\gamma_5\gamma_\mu$. Hence the covariant derivative is invariant under chiral transformations. Now we consider the mass term which transforms as follows,

$$\begin{aligned} m_q\bar{\psi}\psi &\rightarrow m_q\bar{\psi}'\psi' \\ &= m_q\bar{\psi}'\psi' \\ &= m_q\bar{\psi}e^{i2\alpha\gamma_5}\psi. \end{aligned} \tag{1.2.5}$$

Clearly this is not invariant under chiral transformations. However if massless quarks are used instead, $m_q = 0$, then chiral symmetry is restored. The order parameter for this symmetry is the chiral condensate which is given by,

$$\langle\bar{\psi}\psi\rangle = \frac{T}{V} \frac{\partial\ln Z}{\partial m_q} \tag{1.2.6}$$

where T is still temperature, V is the physical volume of the system, and Z is the partition function. The corresponding susceptibility is given by,

$$\chi_{\bar{\psi}\psi} = \frac{T}{V} \frac{\partial^2\ln Z}{\partial m_q^2} \tag{1.2.7}$$

1.2.2 Center symmetry

The second symmetry we will consider is the center symmetry for SU(3). We define the center group, Z(3), as a set of elements in SU(3) that commute with all other elements of SU(3). The elements of Z(3) have the form $z_i\mathbf{I}_3$, where $z_i = \{1, e^{2\pi i/3}, e^{-2\pi i/3}\}$ such that $z_i^3 = 1$ and \mathbf{I}_3 is the 3×3 identity matrix.

Unlike chiral symmetry, it is the coupling of the quarks to the gluon field that breaks the Z(3) symmetry. To show this, we will introduce the naive lattice fermion action as it is easiest to demonstrate center symmetry breaking on the lattice. For details on the derivation of this action, see [16]. The naive lattice fermion action for a single flavour is given by,

$$S_F = a^4 \sum_{n \in \Lambda} \bar{\psi}(n) \left(\sum_{\mu=1}^4 \gamma_\mu \frac{U_\mu(n)\psi(n + \hat{\mu}) - U_{-\mu}(n)\psi(n - \hat{\mu})}{2a} + m\psi(n) \right), \tag{1.2.8}$$

where a is the lattice spacing, n is the 4-coordinate that is related to the space-time coordinate by $x = an$, Λ is the 4D lattice that is the set of all n , and $U_\mu(n)$ is the link variable that connects the site n to $n + \hat{\mu}$. These link variables must elements

of the gauge group $SU(3)$. Note that the quark fields still carry the same indices as before. The quantity $\hat{\mu}$ is a unit vector. The first term is analogous to the covariant derivative Eq.(1.2.2).

Under a center transformation, the time-like link, $U_4(n)$, transforms as follows,

$$U_4(n) \rightarrow z \mathbf{I}_3 U_4(n) \quad (1.2.9)$$

where z is an element of z_i discussed above. As it is only the time-like links that are affected, we only consider $\mu = 4$ terms of Eq.(1.2.8). Therefore the naive lattice fermion action transforms accordingly,

$$S_F \rightarrow S'_F = a^4 \sum_{n \in \Lambda} \bar{\psi}(n) \left(\gamma_4 \frac{z \mathbf{I}_3 U_4(n) \psi(n + \hat{\mathbf{e}}_4) - z^* \mathbf{I}_3 U_{-4}(n) \psi(n - \hat{\mathbf{e}}_4)}{2a} + m \psi(n) \right), \quad (1.2.10)$$

where $\hat{\mathbf{e}}_4$ is the time-like unit vector. The action is not invariant as there is now a factor of z and its conjugate z^* in the summation, and hence the center symmetry is broken.

To restore this symmetry, we use quarks of infinite mass in our theoretical model so that they decouple and we left with pure $SU(3)$ gauge theory. In order to explicitly show that this obeys center symmetry, we introduce the Wilson gauge action,

$$S_G = \frac{2}{g^2} \sum_{n \in \Lambda} \sum_{\mu < \nu} \text{Re} \{ \text{Tr} [\mathbf{I} - U_{\mu\nu}(n)] \} \quad (1.2.11)$$

where $U_{\mu\nu}(n)$ is a plaquette variable. This is a product of four link variables that form a closed loop,

$$\begin{aligned} U_{\mu\nu}(n) &= U_\mu(n) U_\nu(n + \hat{\mu}) U(n + \hat{\mu} + \hat{\nu}) U_{-\nu}(n + \hat{\nu}) \\ &= U_\mu(n) U_\nu(n + \hat{\mu}) U(n + \hat{\nu})^\dagger U_\nu(n)^\dagger. \end{aligned} \quad (1.2.12)$$

All that remains to be checked is that the plaquette variable is invariant under center transformation. Given that center transformation only affects time-like link variables, any plaquette that does not include $U_4(n)$ is unaffected by the transformation. Otherwise a plaquette will transform according to,

$$\begin{aligned} U_{\mu\nu}(n) \rightarrow U'_{\mu\nu}(n) &= U_\mu(n) z \mathbf{I}_3 U_4(n + \hat{\mu}) U(n + \hat{\mathbf{e}}_4)^\dagger z^* \mathbf{I}_3 U_4(n)^\dagger \\ &= z z^* \mathbf{I}_3 U_\mu(n) U_4(n + \hat{\mu}) U(n + \hat{\mathbf{e}}_4)^\dagger U_4(n)^\dagger \\ &= U_{\mu\nu}(n). \end{aligned} \quad (1.2.13)$$

In the first step, we made use of fact that the elements of the center group commute with all elements of the $SU(3)$ group, and recall that the link variables must be elements of $SU(3)$. The second step is trivial as the identity matrix leaves $U_\mu(n)$ unchanged and $zz^* = 1$.

As the quarks are infinitely heavy, they are only able to move along the temporal axis of the lattice, which corresponds to the Polyakov loop [17]. The key results of such simulations is the discovery of a first-order phase transition [18,19]. Below the corresponding T_c , $\langle P \rangle = 0$, and the phase is confined, whilst above T_c , $\langle P \rangle \neq 0$, and the phase is deconfined.

The results of lattice simulations for both $m_q = 0$ and $m_q \rightarrow \infty$ are summarized in the Columbia plot shown in Figure 1.2. The Columbia plot is schematic of the possible thermal phase transitions in $N_f = 2 + 1$ QCD at vanishing baryon chemical potential, where N_f is the number of flavours. Note that $m_u = m_d$. In full QCD

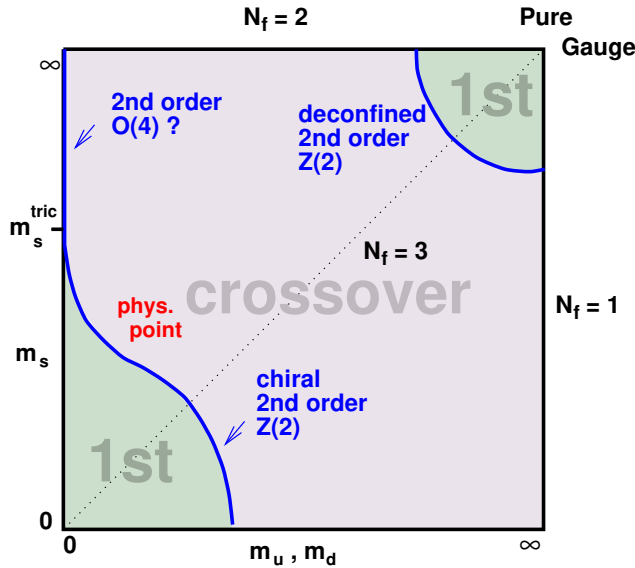


Figure 1.2: Columbia plot: schematic of thermal phase transition behaviour of $N_f = 2 + 1$ QCD at $\mu_B = 0$ for a range of u , d , and s quark masses ($m_u = m_d$) [20].

though neither of these cases are true but despite this it is possible to use $\langle \bar{\psi}\psi \rangle$, $\chi_{\bar{\psi}\psi}$, and $\langle P \rangle$ to estimate a pseudocritical transition temperature, T_{pc} . For this reason they are referred to as remnant order parameters. However it is more common to use chiral susceptibility, as the physical quarks are closer to $m_q = 0$ than $m_q = \infty$. This is reflected in Figure 1.2, although this is only a schematic.

1.3 Probing the Quark-Gluon Plasma

We focus now on the deconfined region of the QCD phase diagram. As has been stated before, for $\mu_B = 0$ the thermal phase transition is an analytical crossover. Whilst the chiral susceptibility can be used to estimate the temperature required for the formation of the QGP, it tells us little about the phase itself. Instead we consider now an alternative signal for QGP formation: the sequential suppression of quarkonia. Quarkonium is a heavy quark-antiquark state that can be used to probe the QGP.

First we consider J/ψ suppression which was proposed as a signal for the formation of QGP. To understand this we need to clarify what deconfinement refers to. The confinement of quarks in hadrons is ultimately determined by the Debye screening of the quark colour charge. This screening radius r_D is dependent on the temperature via the string tension. Therefore at sufficiently high temperatures, $r_D < r_B$, where r_B is the binding radius of J/ψ . At this point the confining force cannot hold the quarks together anymore and the J/ψ melts, or dissociates. These terms are used interchangeably. It was determined that the minimum screening radius that allowed a “Coulombic” bound state is given by,

$$r_D^{\min} = (0.84m\alpha_{\text{eff}}(T))^{-1} \quad (1.3.1)$$

where m is the quark rest mass and $\alpha_{\text{eff}}(T)$ is the “Coulombic” interaction coupling. Note that we use the term “Coulombic” because it is analogous to a true Coulombic interaction which involves electric charge, only in this case it involves the colour charge. Thus it was shown that for SU(3) with dynamical quarks J/ψ is completely suppressed for temperatures only just above T_{pc} [21]. Instead any c quarks produced will bind with light quarks. Hence we can indeed use this dissociation as a signal for the formation of QGP. As such, charmonium has been studied extensively over the years [22–24]. However now that b quarks are produced abundantly at the LHC, the focus of both experimental and theoretical studies has shifted to bottomonium.

This has been motivated in part by phenomenological studies that propose that bottomonium is affected by statistical recombination to a lesser extent [25]. Statistical recombination refers to the case in which a quark and antiquark from dissociated mesons in the QGP come into contact and form a meson again. To understand why

statistical recombination affects bottomonium to a lesser extent, first note that there is a lower density of bottom quarks compared to charm quarks in the QGP. This is partly due to $M_b \gg T$ where T is the temperature reached in the heavy-ion collision experiments, meaning that new pairs of $b\bar{b}$ pairs cannot be produced in the QGP but new pairs of $c\bar{c}$ can. In the case of charmonium, statistical recombination is suppressed due to charm quarks binding with light (u, d) quarks. However in the case of the bottomonium quark, the density of bottom quarks in the QGP is much smaller than of charm quarks, the probability of bottom quarks and anti-bottom quarks to recombine is much smaller.

With the inclusion of bottomonium studies it is now possible to discuss sequential suppression. This refers to the different temperatures at which various quarkonium states are completely suppressed. The inverse proportionality of r_D^{\min} to the quark mass in Eq.(1.3.1) means that bottomonium survives longer before dissociating due to colour screening. Within bottomonium itself there is a hierarchy of suppression temperatures as excited states of the Υ have larger binding radii as does the χ_b (1P). This hierarchy can be used as a thermometer for the QGP [26]. Sequential suppression of the Υ system has been observed experimentally by CMS [27], which has led to phenomenological work in order to explain it [28, 29].

1.4 FASTSUM Gen2L ensembles

In this section we discuss the details of the ensembles generated by the FASTSUM collaboration [30], specifically the Gen2L ensembles [31]. For these ensembles, an anisotropic lattice was used, by which we mean that the spacing of lattice sites along the temporal axis varies from that of the spatial lattice sites. We define the spatial lattice spacing as a_s and the temporal spacing as a_τ . For the Gen2L ensembles, this anisotropy was determined to be $\xi = a_s/a_\tau = 3.453(6)$. The lattice spacing were as follows $a_s = 0.1136(6)\text{fm}$ and $a_\tau = 0.0330(2)\text{fm}$ which equates to $a_\tau^{-1} = 5.997(34)\text{GeV}$. The number of dynamical quarks flavours $N_f = 2 + 1$, meaning only light quarks and the strange quarks were considered. Only the strange quark mass m_s was set to its physical value. The light quarks (the up and down quarks) were heavier than physical, which is represented by π^0 mass used.

N_τ	128	64	56	48	40	36	32	28	24	20
T [MeV]	47	94	107	125	150	167	187	214	250	300
N_{cfg}	1024	1041	1042	1123	1102	1119	1090	1031	1016	1030

Table 1.1: The temporal extent, corresponding temperatures in MeV, and number of configurations for the Generation 2L ensembles of the FASTSUM collaboration

Observable	$\langle \bar{\psi}\psi \rangle_R$	$\chi_{\bar{\psi}\psi}$
T_{pc} [MeV]	164(2)	165(2)(2)

Table 1.2: The estimated pseudocritical temperatures from the renormalised chiral condensate and chiral susceptibility. The second error for the chiral susceptibility is an estimate of the systematic uncertainty.

For the Gen2L ensembles $m_{\pi^0} = 236(2)\text{MeV}$, which is heavier than the physical value. Table 1.1 displays the value of N_τ used for the ensembles along with the corresponding temperature and number of configurations, N_{cfg} . N_τ is the total number of temporal lattice sites, and the temperature is given by $T = 1/(a_\tau N_\tau)$. For all configurations the number of spatial lattice sites was $N_s^3 = 32^3$.

Using these ensembles it has been possible to estimate a pseudocritical temperature, T_{pc} . Table 1.2 displayed the results produced by the FASTSUM collaboration to estimate this pseudocritical temperature [14]. Note that the differing values of T_{pc} is only an indication of the crossover region. If there was a first-order phase transition, the critical temperatures identified by each order parameter would need to match. However as the transition is an analytical crossover, the differing values of T_{pc} are acceptable. Further work was done to estimate a T_{pc} for physical light quark mass using the chiral condensate. This yielded $T_{pc} = 159(6)\text{MeV}$ which is consistent with [13].

Chapter 2

Spectral Reconstruction: An Ill-posed Problem

As we have already mentioned, the main purpose of this work has been to develop a method to extract the spectral function from a Euclidean correlator. The focus of this chapter is deriving the relationship between these two quantities in order to demonstrate what is meant when referring to spectral reconstruction as an ill-posed problem.

2.1 Motivation

Before this though, we need to consider what Euclidean correlator and spectral function represent, and why spectral reconstruction is even necessary. For context we need to present the system that is being modelled in the lattice simulations for the FASTSUM Gen2L ensembles. These correlators represent one of five bottomonium states propagating through a medium of at non-zero temperature. The bottomonium states are not in thermal equilibrium with the medium, and hence are used as a probe. This medium consists of dynamical light and strange quarks. With this in mind, we can discuss Euclidean correlator and spectral function.

We begin with the Euclidean correlator which is a two point function involving the quark and antiquark fields. For this example, we will use the meson interpolator, $O_M(x)$, given by,

$$O_M(\tau) = \bar{\psi}(x)\Gamma\psi(x), \quad (2.1.1)$$

where $\Gamma = \{\gamma_4, \gamma_4\gamma_5, \gamma_4\gamma_i, \gamma_i\gamma_5, \gamma_i\gamma_j\}$ with $i, j = 1, 2, 3$, and x is the space-time coordinate. For this work we will only be considering bottomonium particles with zero momentum. Given this, $\psi/\bar{\psi}$ can be replaced for b/\bar{b} respectively. Depending on our choice of Γ , the interpolator represents a different bottomonium state, which

is shown in Table 2.1. With this, we can write the Euclidean correlator as,

$$\langle O_M(\tau)O_M(0) \rangle = \langle \bar{b}(\tau)\Gamma b(\tau)\bar{b}(0)\Gamma b(0) \rangle, \quad (2.1.2)$$

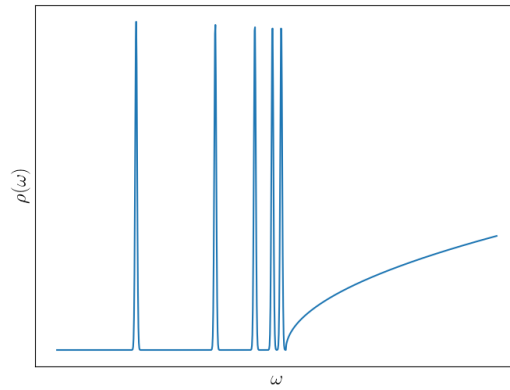
where τ is Euclidean time. In this way, all the quantum numbers are encoded in the Euclidean correlator.

State	Name	Γ	$J^{(PC)}$	$n^{(2S+1)}L_J$
Scalar	η_b	γ_4	0^{-+}	1^1S_0
Pseudoscalar	Υ	$\gamma_4\gamma_5$	1^{--}	1^3S_1
Vector	χ_{b0}	$\gamma_4\gamma_i$	0^{++}	1^3P_0
Axial vector	χ_{b1}	$\gamma_i\gamma_5$	1^{++}	1^3P_1
Tensor	h_b	$\gamma_i\gamma_j$	1^{+-}	1^1P_1

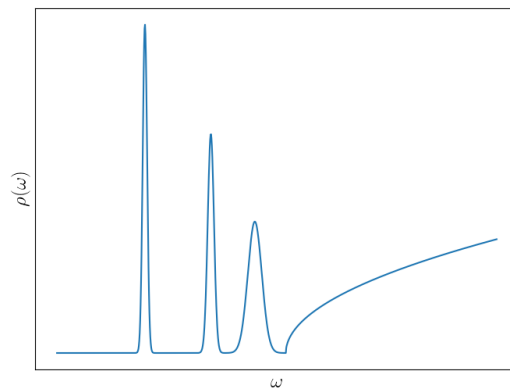
Table 2.1: Bottomonium interpolators with associated quantum numbers and relevant gamma matrices.

The spectral function represents the allowed spectrum of states for the system, in this case bottomonium in the medium, as a function of the energy of the system, ω . As we have set $\hbar = 1$, ω can also represent frequency. In Fig. 2.1 we have attempted to display the effects of temperature on the appearance of the spectral function. These sketches are conjecture to a degree; as we have already mentioned, there is difficulty in obtaining spectral functions from lattice data. However, the main point of Fig. 2.1 is that with just the visual representation of the spectral functions, we are able to make qualitative observations on the thermal dependence of the system.

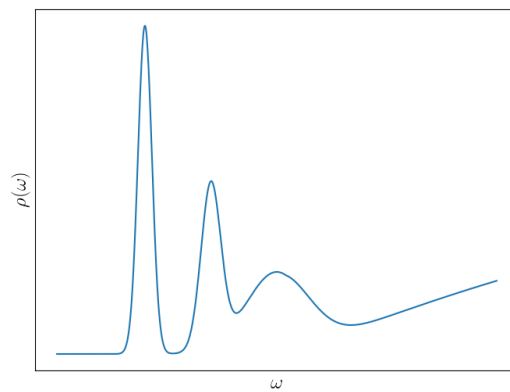
An example spectral function of the system at zero temperature is shown in Fig. 2.1a. We reiterate that this is only a sketch for the purpose of providing some insight. The spectral function for the system at zero temperature appears as a series of delta functions with a continuum. The differences in energy between states decreases with higher order excited states, until ultimately it appears that they overlap. In reality there would be more than the five states shown in Fig. 2.1a before this happened, but we have not included more for the sake of clarity. The continuum we see in Fig. 2.1a occurs when there is sufficient energy to produce multi-particle states. Multi-particle states refers to the case in which the system has enough energy to produce



(a) Zero Temperature



(b) Low Temperature



(c) High Temperature

Figure 2.1: Sketch of example spectral functions for increasing temperature: (a) Zero temperature, (b) Low temperature which produces the first signs of thermal broadening, (c) High temperature that results in dissociation of particle leading to overlap of different energy states and the continuum.

a quark and anti-quark pair such a u and \bar{u} . The b and \bar{b} quarks of the bottomonium state may then bind with these new quarks to produce $u\bar{b}$ and $b\bar{u}$ states.

Fig. 2.1b represent the system at low temperatures in which some effects of the thermal medium are visible via thermal broadening. This is seen by the change from delta functions to Gaussians function as the widths of the resonances increase. The fourth and fifth resonances were removed for clarity. As the model used for FASTSUM Gen2L ensembles does not include Quantum Electrodynamics (QED) or the weak force, the widths of resonances are not related to the decays via either of these. Instead it relates to the motion on the bottomonium within the thermal medium. Although we treat the bottomonium has having zero momentum upon creation, interactions with the thermal medium lead to the bottomonium state undergoing Brownian motion. The continuum in this sketch represents the same as before.

The final sketch in 2.1c represents the system at high temperatures. In this case the thermal broadening has led to excited states overlapping. This is a graphical representation of dissociation referred to in Sec. 1.3. Note that the higher order excited states also overlap with the continuum.

From these sketches, we hope to show that the spectral function would be useful for spectroscopy, as the energies of the ground and excited states are immediately obvious. However, the spectral function can be of further use. Although it is not possible in this work as we have not included QED in our model, other studies have studies have looked at the differential thermal cross section for the production of dilepton pairs which is dependent on the spectral function [32–34]. It is also possible to determine transport coefficients. For example, the electrical conductivity σ which can be obtained from the Kubo relation

$$\frac{\sigma}{T} = \frac{1}{6T} \lim_{\omega \rightarrow \infty} \frac{\rho^{\text{em}}(\omega)}{\omega}, \quad (2.1.3)$$

where $\rho^{\text{em}}(\omega)$ is the spectral functions for the electromagnetic current [35, 36]. For work regarding determining the diffusion coefficient, see [34, 37, 38]. There has also been work to determine thermodynamic quantities such as the pressure and energy density for a system using the spectral function [39]. Although normally such quantities can be determined from the partition function [40, 41], an alternative method would provide a check for the reliability of the reconstructed spectral functions.

2.2 Derivation of Green's functions

The following has been provided to demonstrate why spectral reconstruction is an ill-posed problem. For supporting material see [42, 43]. We begin with the simple case of a single mode of a free scalar bosonic field. The field operator is given by,

$$\phi_{\mathbf{k}}(t) = \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \left(a_{\mathbf{k}} e^{-i\omega_{\mathbf{k}}t} + a_{\mathbf{k}}^{\dagger} e^{i\omega_{\mathbf{k}}t} \right), \quad (2.2.1)$$

where $a_{\mathbf{k}}$, $a_{\mathbf{k}}^{\dagger}$ are the bosonic annihilation and creation operators, t is Minkowski time or real time, and $\omega_{\mathbf{k}}$, which is given by $\sqrt{\mathbf{k}^2 + m^2}$, is the energy of the mode (\mathbf{k} and m are the corresponding 3-momentum and mass). The creation and annihilation operators obey the following,

$$\left[a_{\mathbf{k}}, a_{\mathbf{k}'}^{\dagger} \right] = \delta_{\mathbf{k}\mathbf{k}'} \quad \text{and} \quad \left\langle a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}'} \right\rangle = n_B(\omega_{\mathbf{k}}) \delta_{\mathbf{k}\mathbf{k}'}, \quad (2.2.2)$$

where $n_B(\omega_{\mathbf{k}})$ is the Bose distribution. Hence all expectations values are proportional to $\delta_{\mathbf{k}\mathbf{k}'}$ but we are considering one mode at this stage.

With the field operator defined, we begin our discussion of the various correlation functions. The first of these are the Wightman functions that are not time ordered,

$$G_{\mathbf{k}}^>(t-t') = \langle \phi_{\mathbf{k}}(t) \phi_{\mathbf{k}}(t') \rangle \quad \text{and} \quad G_{\mathbf{k}}^<(t-t') = \langle \phi_{\mathbf{k}}(t') \phi_{\mathbf{k}}(t) \rangle. \quad (2.2.3)$$

By substituting in the field operator given in Eq. (2.2.1) we find the following relation,

$$\begin{aligned} G_{\mathbf{k}}^>(t-t') &= \frac{1}{2\omega_{\mathbf{k}}} \left\langle \left(a_{\mathbf{k}} e^{-i\omega_{\mathbf{k}}t} + a_{\mathbf{k}}^{\dagger} e^{i\omega_{\mathbf{k}}t} \right) \left(a_{\mathbf{k}} e^{-i\omega_{\mathbf{k}}t'} + a_{\mathbf{k}}^{\dagger} e^{i\omega_{\mathbf{k}}t'} \right) \right\rangle \\ &= \frac{1}{2\omega_{\mathbf{k}}} \left(\left\langle a_{\mathbf{k}} a_{\mathbf{k}}^{\dagger} \right\rangle e^{-i\omega_{\mathbf{k}}(t-t')} + \left\langle a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} \right\rangle e^{i\omega_{\mathbf{k}}(t-t')} \right) \\ &= \frac{1}{2\omega_{\mathbf{k}}} \left(e^{-i\omega_{\mathbf{k}}(t-t')} + \left\langle a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} \right\rangle \left[e^{-i\omega_{\mathbf{k}}(t-t')} + e^{-i\omega_{\mathbf{k}}(t-t')} \right] \right) \\ &= \frac{1}{2\omega_{\mathbf{k}}} \left(e^{-i\omega_{\mathbf{k}}(t-t')} + 2n_B(\omega_{\mathbf{k}}) \cos[\omega_{\mathbf{k}}(t-t')] \right) \\ &= G_{\mathbf{k}}^<(t'-t). \end{aligned} \quad (2.2.4)$$

In the first step we have expanded the brackets, making use of $\langle a_{\mathbf{k}}^2 \rangle = \langle a_{\mathbf{k}}^{\dagger 2} \rangle = 0$ to eliminate the remaining terms. In the following step, we used the commutation relation in Eq. (2.2.2), $a_{\mathbf{k}} a_{\mathbf{k}'}^{\dagger} = \delta_{\mathbf{k}\mathbf{k}'} + a_{\mathbf{k}'}^{\dagger} a_{\mathbf{k}}$. Finally we collect the terms in the

square brackets and use the expectation value from Eq. (2.2.2). If we repeat this for $G_{\mathbf{k}}^<(t - t')$, we find the final result of Eq. (2.2.4) holds. We can now use the Wightman functions to define more correlators.

The first of these is the Feynman propagator, which introduces time ordering by using the Heaviside step function, $\theta(t)$,

$$G_{\mathbf{k}}^F(t - t') = \theta(t - t')G_{\mathbf{k}}^>(t - t') + \theta(t' - t)G_{\mathbf{k}}^<(t - t'). \quad (2.2.5)$$

Next is the spectral function $\tilde{\rho}(t - t')$, which in real-space, can be defined as the expectation value of the commutator

$$\begin{aligned} \tilde{\rho}(t - t') &= \langle [\phi_{\mathbf{k}}(t), \phi_{\mathbf{k}}(t')] \rangle = G_{\mathbf{k}}^>(t - t') - G_{\mathbf{k}}^<(t - t') \\ &= \frac{1}{i\omega_{\mathbf{k}}} \sin(\omega_{\mathbf{k}}(t - t')). \end{aligned} \quad (2.2.6)$$

For completeness, we also consider anticommutator which gives the statistical two-point function,

$$\begin{aligned} F(t - t') &= \frac{1}{2} \langle \{\phi_{\mathbf{k}}(t), \phi_{\mathbf{k}}(t')\} \rangle = \frac{1}{2} (G_{\mathbf{k}}^>(t - t') + G_{\mathbf{k}}^<(t' - t)) \\ &= \frac{1}{2\omega_{\mathbf{k}}} (1 + 2n_B(\omega_{\mathbf{k}})) \cos(\omega_{\mathbf{k}}(t - t')). \end{aligned} \quad (2.2.7)$$

The final two correlators are the retarded and advanced Green's functions, which can be defined in terms of the spectral function,

$$G_{\mathbf{k}}^R(t - t') = -i\theta(t - t') \langle [\phi_{\mathbf{k}}(t), \phi_{\mathbf{k}}(t')] \rangle = -i\theta(t - t')\tilde{\rho}(t - t') \quad (2.2.8)$$

$$G_{\mathbf{k}}^A(t - t') = +i\theta(t' - t) \langle [\phi_{\mathbf{k}}(t), \phi_{\mathbf{k}}(t')] \rangle = +i\theta(t' - t)\tilde{\rho}(t - t'). \quad (2.2.9)$$

Using the standard Fourier transform we can define the spectral function in frequency space as¹

$$\begin{aligned} \rho(\omega) &= -\frac{i}{\omega_{\mathbf{k}}} \int_{-\infty}^{\infty} dt e^{-i\omega t} \sin(\omega_{\mathbf{k}}t) \\ &= \frac{2\pi}{2\omega_{\mathbf{k}}} [\delta(\omega - \omega_{\mathbf{k}}) - \delta(\omega + \omega_{\mathbf{k}})] \end{aligned} \quad (2.2.10)$$

for $t' = 0$. Hence we expect two resonances; one at $\omega = \omega_{\mathbf{k}}$ and the other at $\omega = -\omega_{\mathbf{k}}$.

Recall that we are considering a single mode for the free scalar field. Repeating this

¹Note we use $\rho(\omega)$ and $\tilde{\rho}(t)$ for the spectral function. For all other Green's functions we will use $G(t)$ and $\tilde{G}(\omega)$. This is because we will focus on ρ as a function frequency/energy, ω , and the Euclidean correlator as a function of Euclidean time, τ .

for the Wightman functions and the statistical two-point function, it follows they are all proportional to $\rho(\omega) = \tilde{G}^>(\omega) - \tilde{G}^<(\omega)$. As for the retarded/advanced Green's functions we need to take $\theta(t)$ into account. Using the identity

$$\theta(t) = i \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{e^{-i\omega t}}{\omega + i\epsilon}, \quad (2.2.11)$$

and the inverse Fourier transform for the spectral function we define the retarded Green's function in frequency space as,

$$\begin{aligned} \tilde{G}_{\mathbf{k}}^R(\omega) &= -i \int_{-\infty}^{\infty} dt \theta(t) \tilde{\rho}_{\mathbf{k}}(t) \\ &= \int_{-\infty}^{\infty} dt e^{i\omega t} \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{e^{-i\omega' t}}{\omega' + i\epsilon} \tilde{\rho}_{\mathbf{k}}(t) \\ &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{\tilde{\rho}_{\mathbf{k}}(t)}{\omega' + i\epsilon} e^{i(\omega - \omega')t} \\ &= \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{\rho_{\mathbf{k}}(\omega')}{\omega' - (\omega + i\epsilon)} \\ &= \frac{1}{2\omega_{\mathbf{k}}} \left[\frac{1}{\omega_{\mathbf{k}} - (\omega + i\epsilon)} + \frac{1}{\omega_{\mathbf{k}} + (\omega + i\epsilon)} \right] \\ &= \frac{1}{\omega_{\mathbf{k}}^2 - (\omega + i\epsilon)^2}. \end{aligned} \quad (2.2.12)$$

To go from the third to the fourth line, we used the following transformation $\omega' \rightarrow \omega - \omega'$. As a consequence of this $d\omega' \rightarrow -d\omega'$. For completeness,

$$\tilde{G}^A(\omega) = \frac{1}{\omega_{\mathbf{k}}^2 - (\omega - i\epsilon)^2}. \quad (2.2.13)$$

so that $\tilde{G}_{\mathbf{k}}^R(\omega) = \tilde{G}_{\mathbf{k}}^{A*}(\omega) = \tilde{G}_{\mathbf{k}}^A(-\omega)$. Using the Sokhotski–Plemelj theorem,

$$\frac{1}{x + i\epsilon} = \mathcal{P} \frac{1}{x} - i\pi\delta(x), \quad (2.2.14)$$

where \mathcal{P} is the Cauchy principal value, it follows directly that $\tilde{G}_{\mathbf{k}}^R(\omega) - \tilde{G}_{\mathbf{k}}^A(\omega) = i\rho_{\mathbf{k}}(\omega)$, so we can define the spectral function as

$$\rho(\omega) = -i \left[\tilde{G}_{\mathbf{k}}^R(\omega) - \tilde{G}_{\mathbf{k}}^A(\omega) \right] = 2\text{Im}\tilde{G}_{\mathbf{k}}^R(\omega). \quad (2.2.15)$$

2.2.1 Kubo-Martin-Schwinger condition

Thus far we have only been considering free scalar fields. As we introduce interactions, we need to consider whether these relations still hold non-perturbatively. Fortunately we can make use of the Kubo-Martin-Schwinger (KMS) condition [44, 45].

The only requirement for this condition is that the system is in thermal equilibrium. To begin, we replace the free scalar bosonic fields that have been used until this point with a pair of operators $A(t)$ and $B(t)$ that evolve according to,

$$A(t) = e^{iHt}A(0)e^{-iHt}, \quad (2.2.16)$$

such that the Wightman functions are now written as $G_{\mathbf{k}}^>(t-t') = \langle A(t)B(t') \rangle$ and $G_{\mathbf{k}}^<(t-t') = \langle B(t')A(t) \rangle$. The thermal expectation value of an operator O is given by $\langle O \rangle = Z^{-1}\text{Tr} [e^{-\beta H}O]$, where Z is the partition function, given by $Z = \text{Tr} [e^{-\beta H}]$. The Hamiltonian in the time evolution and in the thermal expectation value, are identical which signifies that the system is in thermal equilibrium. Now by making use of the cyclicity of trace, it follows in the bosonic case,

$$G^>(t_1 - t_2) = G^<(t_1 - t_2 + i\beta). \quad (2.2.17)$$

This is the KMS condition and any state that satisfies it is referred to as a KMS state. As the Hamiltonian may contain interactions, this holds non-perturbatively as desired. By performing a Fourier transform, it follows that $\rho(\omega) = \tilde{G}^>(\omega) - \tilde{G}^<(\omega)$ still holds.

2.2.2 Euclidean Correlator

In order to obtain the Euclidean correlator, we perform a Wick rotation $t \rightarrow \tau = it$, where τ is Euclidean time. In this case we can treat $e^{-\beta H}$ as the evolution operator. The Euclidean correlator is thus defined as

$$G^E(\tau) = \langle O(\tau)O(0) \rangle. \quad (2.2.18)$$

In this case τ runs from 0 to $1/T = \beta$ where T and β are temperature and inverse temperature respectively. This provides insight into how the Euclidean correlator can be used to studied a system at finite temperature. Given that the Euclidean correlator is defined for imaginary time it can be written as

$$G^E(\tau) = T \sum_n e^{-i\omega_n \tau} \tilde{G}^E(\omega_n), \quad (2.2.19)$$

where ω_n is the Matsubara frequency. For bosons, $\omega_n = 2n\pi T$ in order to satisfy the periodic boundary condition $G^E(\tau) = G^E(\tau + \beta)$. With this definition of the

Euclidean correlator we now need to relate it to $\rho(\omega)$. The first step towards this is to perform a Fourier transform to determine the correlator in frequency space

$$\tilde{G}^E(\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} G^E(\tau). \quad (2.2.20)$$

For simplicity we revert back to considering a free scalar field. It should be clear that by setting $O(\tau) = \phi_{\mathbf{k}}(\tau)$ that Eq. (2.2.18) is equivalent to Eq. (2.2.4), with $t \rightarrow \tau = it$ and $t' = 0$. Making use of the penultimate line of Eq. (2.2.4) and these substitutions for t and t' , the Euclidean correlator can be defined as

$$G_{\mathbf{k}}^E(\tau) = \frac{\cosh(\omega_{\mathbf{k}}[\tau - 1/2T])}{2\omega_{\mathbf{k}}\sinh(\omega_{\mathbf{k}}/2T)}. \quad (2.2.21)$$

In frequency space, this becomes

$$\tilde{G}_{\mathbf{k}}^E(\omega_n) = \frac{1}{\omega_n^2 + \omega_{\mathbf{k}}^2}. \quad (2.2.22)$$

In this form we can relate it to another of the previous Green's functions by coordinate transformation. Setting $\omega_n \rightarrow i\omega - \epsilon$ we find $\tilde{G}_{\mathbf{k}}^E(\omega_n \rightarrow i\omega - \epsilon) = \tilde{G}_{\mathbf{k}}^R(\omega)$, and conversely $\tilde{G}_{\mathbf{k}}^R(\omega + i\epsilon) = \tilde{G}_{\mathbf{k}}^E(i\omega_n)$. With this we can relate the Euclidean correlator to the spectral function in frequency space using the fourth line of Eq. (2.2.12) and performing the coordinate transformation, $\omega + i\epsilon \rightarrow i\omega_n$,

$$\tilde{G}_{\mathbf{k}}^E(\omega_n) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\rho_{\mathbf{k}}(\omega)}{\omega - i\omega_n}. \quad (2.2.23)$$

Although we have produced a relation between the Euclidean correlator and the spectral function there is still one more step. Lattice simulations are done in position space, not frequency. This is easily resolved by using Eq. (2.2.19),

$$\begin{aligned} G_{\mathbf{k}}^E(\tau) &= T \sum_n e^{-i\omega_n \tau} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\rho_{\mathbf{k}}(\omega)}{\omega - i\omega_n} \\ &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} K(\tau, \omega) \rho_{\mathbf{k}}(\omega) \end{aligned} \quad (2.2.24)$$

where $K(\tau, \omega)$ is the kernel function, or thermal kernel to avoid later confusion. With this we have direct relation between the Euclidean correlator in Euclidean time to the spectral function. In this form it may seem that it is trivial to obtain the spectral function. Unfortunately that is not the case, as the correlator obtained by our lattice simulations is discrete and thus $G_{\mathbf{k}}^E(\tau)$ is only known numerically at N_τ points. In Table 1.1 we see at most $N_\tau = 128$. Furthermore these correlators will have

associated errors. The spectral function on the other hand can take various forms as we have attempted to show in Fig. 2.1. At zero temperature we expect a series of delta functions, whereas at very high temperatures we expect be a continuous function. Note that in both cases, we expect there to be a continuum.

As there is no direct inversion, we refer to this as an ill-posed problem. Numerous techniques have been developed to tackle this problem; more recently these have included machine learning methods. One such method will be the focus of later work.

Chapter 3

NRQCD

Before we delve into spectral reconstruction we will examine the Euclidean correlators themselves. By using the correlators themselves it is possible to estimate the mass of the ground and first few excited states of the bottomonium states. We will also use the correlator to determine the thermal dependence of the bottomonium particles.

3.1 Lattice NRQCD

We begin by discussing how the FASTSUM Gen2L ensembles were generated using Lattice Non-relativistic QCD (NRQCD). This is one of many effective field theories (EFTs) developed for the purpose of investigating quarkonia. These EFTs differ depending on the hierarchy of the scales. The scales of interest are the heavy quark mass M , the QGP temperature T , and the strong coupling α_s which is used to determine the inverse system size $M\alpha_s$, the binding energy $M\alpha_s^2$, and inverse Debye screening length $\sqrt{\alpha_s}T \sim m_D$. It is the ordering of these that has led to the formulation of various EFTs [46–50]. For example, the inverse Debye screening length and system size relate to the Debye radius and binding radius discussed in Sec. 1.3. Therefore the ordering of these two in an EFT determines whether the quarkonium has dissociated or not. Another example would be whether we want to consider a theory at high temperature, in which case

$$M \gg T > M\alpha_s > m_D \gg M\alpha_s^2,$$

whilst at low temperatures

$$M \gg M\alpha_s \gg T \gg M\alpha_s^2 \gg m_D.$$

Even for the EFTs dedicated to high temperatures [48], $M \gg T$ holds for bottomonium which should be clear from Table 1.1. In this work the highest temperature

considered is 300 MeV whilst $M_b \sim 5$ GeV. NRQCD refers to the EFTs where the bottom quark mass has been integrated out [51, 52]. A thorough derivation and explanation of NRQCD can be found in [53], but here we will give a brief overview. Firstly, we need to introduce two more energy scales: the heavy quark 3-momentum $M_b v$ in the bottomonium rest frame, and the kinetic energy $M_b v^2$. The heavy quark velocity is then clearly $v = |\mathbf{p}|/M_b$, in the bottomonium rest frame. It turns out that $M_b v^2 \sim \Lambda_{QCD}$ where Λ_{QCD} is the scale of nonperturbative effects of light quarks and gluons.

To acquire NRQCD, we start with full QCD and introduce an ultraviolet cutoff Λ , that explicitly removes relativistic heavy quarks, gluons, and light quarks with momenta of order M_b . Next a Foldy-Wouthuysen-Tani [54] transformation is implemented so that quark and antiquark degrees of freedom are decoupled from the theory. Finally we expand the Lagrangian to $\mathcal{O}(v^4)$. Although this expansion would need infinite terms to exactly reproduce QCD, only a small number of terms are necessary for sufficient accuracy as $v^2 \sim 0.1c^2$ for the bottom quark [55].

$$\mathcal{L} = \mathcal{L}_0 + \delta\mathcal{L}. \quad (3.1.1)$$

The first term, \mathcal{L}_0 is the leading term of the expansion which describes the heavy quarks and antiquarks,

$$\mathcal{L}_0 = \psi^\dagger \left(D_\tau - \frac{\mathbf{D}^2}{2M} \right) \psi + \phi^\dagger \left(D_\tau + \frac{\mathbf{D}^2}{2M} \right) \phi, \quad (3.1.2)$$

where D_τ and \mathbf{D} are the gauge covariant temporal and spatial derivatives respectively, and ψ/ϕ are the quark/antiquark fields. Note that the covariant temporal derivative is in Euclidean time. The correction term $\delta\mathcal{L}$ is introduced to restore relativistic effects, in order to reduce the systematic error of the leading order,

$$\begin{aligned} \delta\mathcal{L} = & -\frac{c_1}{8M_b^3} [\psi^\dagger (\mathbf{D}^2)^2 \psi - \phi^\dagger (\mathbf{D}^2)^2 \phi] \\ & + c_2 \frac{ig}{8M_b^2} [\psi^\dagger (\mathbf{D} \cdot \mathbf{E} - \mathbf{E} \cdot \mathbf{D}) \psi + \phi^\dagger (\mathbf{D} \cdot \mathbf{E} - \mathbf{E} \cdot \mathbf{D}) \phi] \\ & - c_3 \frac{ig}{8M_b^2} [\psi^\dagger \boldsymbol{\sigma} \cdot (\mathbf{D} \times \mathbf{E} - \mathbf{E} \times \mathbf{D}) \psi + \phi^\dagger \boldsymbol{\sigma} \cdot (\mathbf{D} \times \mathbf{E} - \mathbf{E} \times \mathbf{D}) \phi] \\ & - \frac{c_4}{2M_b} [\psi^\dagger \boldsymbol{\sigma} \cdot \mathbf{B} \psi - \phi^\dagger \boldsymbol{\sigma} \cdot \mathbf{B} \phi], \end{aligned} \quad (3.1.3)$$

where \mathbf{E} and \mathbf{B} are the chromoelectric and chromomagnetic fields, and $\boldsymbol{\sigma}$ are the Pauli spin matrices. Although there are several corrections that could be included,

this is sufficient for significantly reducing the systematic errors for $\mathcal{O}(v^4)$ expansion. Further expansion of the Lagrangian would require tuning more coefficients.

Thus far we have focused on continuum NRQCD. In order to reformulate this on a space-time lattice we introduce the forward, backward and centered lattice covariant derivatives,

$$\begin{aligned} a\Delta_\mu^{(+)}\psi(x) &= U_{x,\mu}\psi(x + a\hat{\mu}) - \psi(x) \\ a\Delta_\mu^{(-)}\psi(x) &= \psi(x) - U_{x-a\hat{\mu},\mu}^\dagger\psi(x - a\hat{\mu}) \\ \Delta_\mu^{(\pm)} &= \frac{1}{2}(\Delta_\mu^{(+)} + \Delta_\mu^{(-)}). \end{aligned} \quad (3.1.4)$$

The quark fields are defined on the node x_μ of the lattice (for antiquark fields, replace $\psi(x)$ with $\phi(x)$). The unitary matrices $U_{x,\mu}$ are elements of SU(3) and are defined on the links joining nodes. Lastly a could either be a_s or a_τ . With these, we can determine a lattice NRQCD Lagrangian and action. For the leading order term of the continuum lagrangian, \mathcal{L}_0 , the quark propagators satisfy the evolution equation,

$$G(\mathbf{x}, \tau + a_\tau) = \left(1 - \frac{a_\tau H_0|_{\tau+a_\tau}}{2k}\right)^k U_4^\dagger(\mathbf{x}, \tau) \left(1 - \frac{a_\tau H_0|_\tau}{2k}\right)^k G(\mathbf{x}, \tau), \quad (3.1.5)$$

where the so-called Lepage's parameter k is included to counter instabilities at large momenta. However for the b quark $k = 1$ is sufficient [56]. $U_4^\dagger(\mathbf{x}, \tau)$ is the temporal gauge link at node x and H_0 is the leading kinetic energy operator,

$$H_0 = -\frac{\Delta^{(2)}}{2M_b}. \quad (3.1.6)$$

where $\Delta^{(2n)} = \sum_{i=1}^3 (\nabla_i^+ \nabla_i^-)^n$, the lattice Laplace operator. After defining the relativistic corrections in terms of lattice variables the evolution equation becomes,

$$G(\mathbf{x}, \tau + a_\tau) = \left(1 - \frac{a_\tau H_0|_{\tau+a_\tau}}{2k}\right)^k U_4^\dagger(\mathbf{x}, \tau) \left(1 - \frac{a_\tau H_0|_\tau}{2k}\right)^k (1 - a_\tau \delta H) G(\mathbf{x}, \tau), \quad (3.1.7)$$

where δH is the corrections. Explicitly these corrections are given by

$$\begin{aligned} \delta H &= -\frac{(\Delta^{(2)})^2}{8M_b^3} + \frac{ig_0}{8M_b^2}(\nabla^\pm \cdot \mathbf{E} - \mathbf{E} \cdot \nabla^\pm) \\ &\quad - \frac{g_0}{8M_b^2} \boldsymbol{\sigma} \cdot (\nabla^\pm \times \mathbf{E} - \mathbf{E} \times \nabla^\pm) - \frac{g_0}{2M_b} \boldsymbol{\sigma} \cdot \mathbf{B} \\ &\quad + \frac{a_s^2 \Delta^{(4)}}{24M_b} - \frac{a_\tau (\Delta^{(2)})^2}{16M_b^2}. \end{aligned} \quad (3.1.8)$$

The evolution equation has the condition that $G(\mathbf{x}, \tau) = 0$ for $\tau < 0$. For $\tau = 0$, $G(\mathbf{x}, 0) = S(\mathbf{x})$, where $S(\mathbf{x})$ is the source. For the work presented in this thesis point sources were used. Note the similarities to \mathcal{L} , however in this case \mathbf{E} and \mathbf{B} are the lattice definitions of the chromoelectric and chromomagnetic fields [57, 58]. The coefficients in Eq. (3.1.3), c_i where $i = 1, 2, 3, 4$, are all set to one. We implement tadpole improvement [59],

$$U_i(x) \rightarrow \frac{U_i(x)}{u_s}, \quad U_4(x) \rightarrow \frac{U_4(x)}{u_\tau}, \quad (3.1.9)$$

where $u_{s,\tau}$ are the mean space (s) and time-like (τ) links. The mean space-like link, u_s is determined as the fourth root of the plaquette expectation value, whilst u_τ is usually set equal to 1 [60]. The final two terms of Eq. (3.1.8) are corrections to the kinetic energy in order to remove the $\mathcal{O}(a_s^2)$ errors in H_0 and $\mathcal{O}(a_\tau)$ errors in the evolution equation due to finite lattice spacing.

3.2 Properties of the system from Correlators

The evolution equation, Eq. (3.1.7), has been used to generate the correlators given in Table 1.1 with the lattice setup described in Sec. 1.4. Recall that the correlation functions contain information about the system, albeit in a format that makes it difficult to extract said information. That is not to say however that the correlations cannot be used to determine some properties of the system. In the following we will focus on the thermal modifications of the NRQCD correlators and effective masses.

However before we can do this, it is necessary to tune the bare heavy quark mass M_b for a full investigation of bottomonium spectroscopy. This was done by fitting the non-relativistic hadronic dispersion relation for the Υ channel at zero temperature,

$$a_\tau E(\hat{P}^2) = a_\tau E(0) + \frac{\hat{P}^2}{2\xi a_s M_2} \quad (3.2.1)$$

$$\hat{P}^2 = 4 \sum_{i=1}^3 \sin^2 \left(\frac{\pi n_i}{N_s} \right)$$

where $a_s^2 \hat{P}^2$ is the lattice momenta and $n_i = -N_s/2+1, \dots, N_s/2$, and $N_s = 32$. Using this method, the bare lattice bottom quark mass was determined to be $a_s M_b = 2.06$.

This corresponds to $M_b = 3.578 \text{ GeV}$, smaller than the approximate physical value of $\sim 4.5 \text{ GeV}$. Now we can begin analysis of the NRQCD ensembles, starting with the effective masses m_{eff} of the bottomonium states.

3.2.1 Effective mass

Previously we have stated that the bottomonium is not in thermal equilibrium with the medium of dynamic quarks. Instead the thermal effects enter the theory by the in-medium effects such as interactions between the bottomonium and light quarks. However, thus far we have not shown this explicitly. To do this, we need to show that the kernel, $K(\tau, \omega)$ is independent of temperature, T , and that any thermal modification to the correlators is due to changes in the associated $\rho(\omega)$ induced via the interactions with the medium of light quarks and gluons. For clarity, ω is energy and τ is Euclidean time,

To do this, first note that as bottomonia are boson states, only the odd terms of Eq. (2.2.21) are included in $K(\tau, \omega)$ to give

$$K(\tau, \omega) = \frac{\cosh(\omega[\tau - 1/2T])}{\sinh(\omega/2T)}, \quad (3.2.2)$$

if periodic temporal boundary conditions were obeyed. If periodic temporal boundary conditions are not obeyed, then the b quarks are no longer in thermal equilibrium. The next step to remove T dependence from Eq. (3.2.2) is to set $\omega = \omega' + 2M$. Finally recall that for NRQCD $M \gg T$, and so we take $M_b/T \rightarrow \infty$. This leads to

$$K(\tau, \omega) = \frac{\cosh(\omega[\tau - 1/2T])}{\sinh(\omega/2T)} \rightarrow e^{-\omega\tau}. \quad (3.2.3)$$

Hence we have removed the temperature dependence from $K(\tau, \omega)$. Given that this does not obey periodic temporal boundary conditions, our bottomonium cannot be in thermal equilibrium. For continuum NRQCD we can rewrite Eq. (2.2.21) as,

$$G(\tau) = \int_0^\infty \frac{d\omega}{2\pi} e^{-\omega\tau} \rho(\omega) = \int_{-2M_b}^\infty \frac{d\omega'}{2\pi} e^{-\omega'\tau} \rho(\omega'). \quad (3.2.4)$$

For lattice NRQCD, the expression is the same with the exception of the limits used. At zero temperature it is assumed that the spectral function of any of the bottomonium states can be expressed as a series of delta functions as there is no thermal broadening. For now we will only be considering a single term that corresponds to

m_{eff} . Explicitly, we define the $\rho(\omega)$ in this case as,

$$\rho(\omega) = 2\pi Z \delta(\omega - m_{\text{eff}}), \quad (3.2.5)$$

where Z relates to the amplitude of the state. Hence write Eq. (3.2.4) as,

$$G(\tau) = Z \int d\omega e^{-\omega\tau} \delta(\omega - m_{\text{eff}}) = Z e^{-\tau m_{\text{eff}}}. \quad (3.2.6)$$

For the anisotropic lattice $\tau = a_\tau n_\tau$, where $n_\tau = 0, 1, \dots, N_\tau - 1$. We can then define the effective mass as follows,

$$m_{\text{eff}} = -a_\tau^{-1} \ln \left(\frac{G(\tau + a_\tau)}{G(\tau)} \right). \quad (3.2.7)$$

We do not have zero temperature data and so instead we have had to use $N_\tau = 128$. Rather than using the central value and variance for these configurations, plots and errors were generated using bootstrap analysis (see Appendix B).

Although Fig. 3.1 clearly shows that the assumption $\rho(\omega) = \delta(\omega - m_{\text{eff}})$ does not hold, there is an evident constant behaviour at later n_τ for all channels, though we note the P-waves states become noisy after $n_\tau \approx 100$. A single value for the effective mass of each channel was determined using the mean of the final 60 points, which can be seen in Table 3.1. The Υ is of particular interest, as we will now use it to define a necessary energy shift E_0 . Given that $a_\tau m_{\text{eff}} = 0.33273(1)$, $m_{\text{eff}} = 0.33273 \times 5.997 = 1.9954(6)$ GeV, significantly lower than the experimental value of 9.460 GeV. This is because only energy differences are relevant in NRQCD rather than the rest-mass of the states directly, as this energy can be removed from the heavy quark dispersion relation by performing a field transformation. Hence our need for E_0 , which was found to be

$$E_0 = M_{\text{expt}} - m_{\text{eff}} = 9.460 - 1.9954 = 7.46461(4) \text{ GeV}.$$

Table 3.1 displays the effective masses for the four states, in both lattice units and GeV, along with the experimental values [61].

3.2.2 Higher order masses

Fig. 3.1 demonstrates the need for an alternative spectral function, whether that be higher order terms or a different function altogether. For now we will continue with

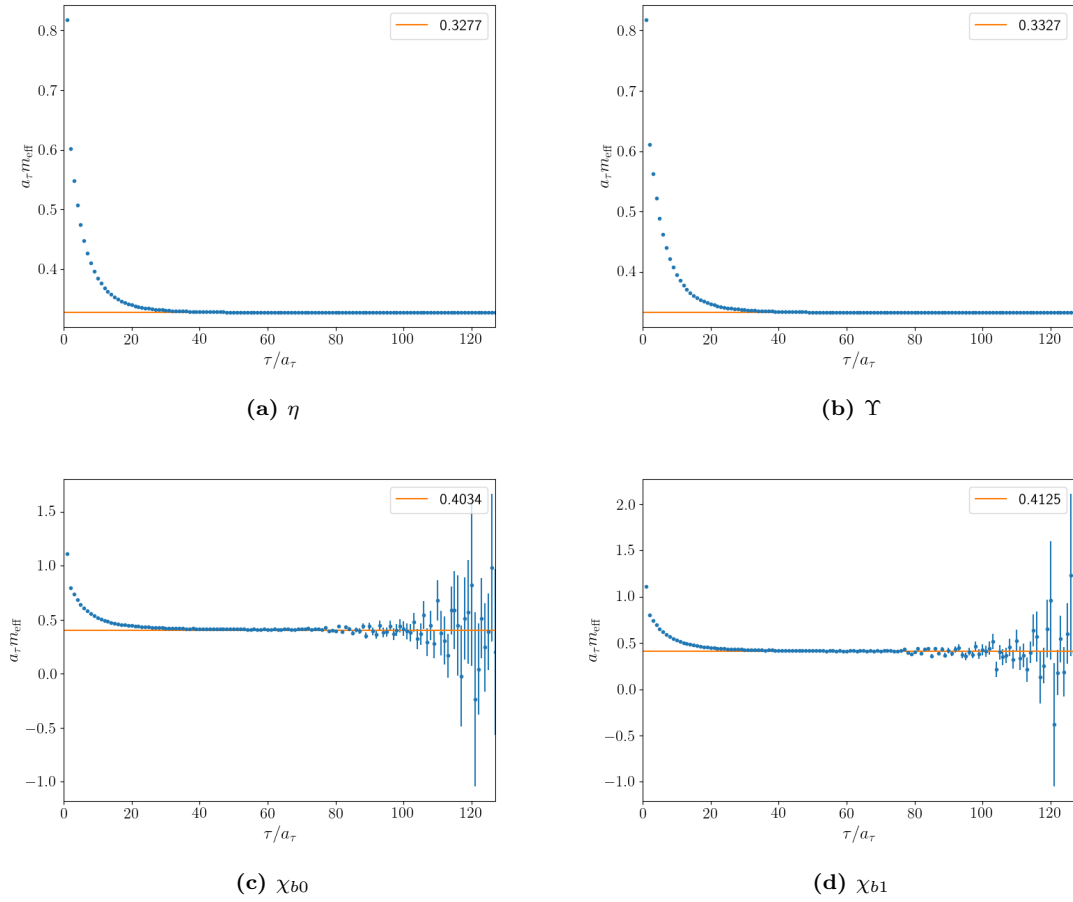


Figure 3.1: Effective mass according to Eq. (3.2.7) for following bottomonium states: scalar η_b (a), pseudoscalar Υ (b), vector χ_{b0} (c), and axial vector χ_{b1} (d) channels for $N_\tau = 128$. The orange line represents the mean effective mass of the last 60 points.

Name	$J^{(PC)}$	$n^{(2S+1)}L_J$	$a_\tau m_{\text{eff}}$	$E_0 + m_{\text{eff}}$	M_{exp} [GeV]
η_b	0^{-+}	1^1S_0	0.32774(1)	9.43045(6)	9.399
Υ	1^{--}	1^3S_1	0.33273(1)	9.46039(4)	9.4603
h_b	1^{+-}	1^1P_1	0.415(17)	9.95(10)	9.8993
χ_{b0}	0^{++}	1^3P_0	0.403(14)	9.884(83)	9.8594
χ_{b1}	1^{++}	1^3P_1	0.412(16)	9.939(94)	9.8928

Table 3.1: Bottomonium spectrum with effective masses in lattice units and MeV. Experimental values were taken from the Particle Data Group [61]

the delta functions but include higher order terms, such that the correlators can be

defined as,

$$G(\tau) = \sum_j^{N_p} Z_j \int d\omega e^{-\omega\tau} \delta(\omega - m_j) = \sum_j^{N_p} Z_j e^{-\tau m_j} \quad (3.2.8)$$

where Z_j is the amplitude associated with each term and N_p is the number of terms used which would correspond to the number of resonances or peaks in the spectral function. For convenience, Z_j and m_j were determined using the CORFITTER package [62]. For further details of this fitting procedure, see [63–65].

As with the effective mass, only $N_\tau = 128$ was used. However, unlike the effective mass which used only the last 60 timeslices to determine a value, we have used all but the first 4 timeslices. The decision to ignore the timeslices is that they are heavily affected by lattice artifacts. The reason for only ignoring 4 timeslices is that we need to choose the equivalent of at least 1 spatial lattice spacing. Recall that $\xi = 3.45$, so that $a_s = 3.45a_\tau$, and since we need to neglect at least 1 spatial lattice spacing, we round up to 4. The results of this fitting procedure for 1-5 terms is displayed in Table 3.2, after being converted to GeV as the output initially corresponds to $a_\tau m_j$. In Figure 3.2 we have plotted the groundstate masses alongside the experimental values and the effective masses of Table 3.1.

We see that it takes 4 terms for the groundstate mass to lie within the uncertainty of the effective mass. Only in the case of the Υ do all results agree which is to be expected since this state was used for tuning. Looking at Table 3.2, we see that increasing the number of terms to 5 leads to significant uncertainty on the masses for the highest excited state. This is due to the increased difficulty in fitting parameters as the number of terms increase. It is more pronounced for P-waves than S-waves due to the variance of these configurations, which can clearly be seen in Fig. 3.1c and Fig. 3.1d.

The results of the fitting procedure have been used to reconstruct the correlators using Eq.(3.2.8). The percentage error for each state and number of terms is displayed in Table 3.3, and the reconstructed correlators G_{recon} for Υ and χ_{b1} states for 1-4 terms are plotted in Fig. 3.3 and 3.4 respectively.

Given all the result so far, we decided that for both states, $N_p = 4$ is the most appropriate number of terms. For the S-waves this is because there is little

n = 1					
j	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	10.906(50)	10.979(31)	11.370(13)	11.226(18)	11.508(27)
n = 2					
j	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	9.5304(55)	9.5838(46)	10.2372(40)	10.1781(46)	10.2412(33)
2	11.424(15)	11.4973(89)	12.2135(80)	12.104(13)	12.2155(68)
n = 3					
j	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	9.4401(14)	9.47163(78)	10.01473(93)	9.9782(13)	10.01165(93)
2	10.4267(88)	10.4563(45)	11.0472(51)	10.9550(61)	11.0398(42)
3	12.284(17)	12.3054(96)	13.0200(63)	12.9501(86)	13.0187(64)
n = 4					
j	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	9.43086(12)	9.460860(75)	9.9619(17)	9.9373(15)	9.96018(99)
2	10.1444(31)	10.1620(16)	10.624(16)	10.575(11)	10.6261(74)
3	11.2477(91)	11.2669(46)	11.886(36)	11.797(27)	11.891(18)
4	13.265(16)	13.2185(75)	13.862(53)	13.789(36)	13.872(26)
n = 5					
j	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	9.430416(71)	9.460341(48)	9.9501(62)	9.923(19)	9.9486(47)
2	10.0549(65)	10.0698(36)	10.485(93)	10.34(21)	10.476(69)
3	10.759(31)	10.771(19)	11.42(26)	11.03(25)	11.38(20)
4	11.995(55)	11.981(33)	12.95(41)	12.36(26)	12.89(32)
5	14.28(10)	14.110(58)	15.9(1.6)	14.39(41)	15.5(1.2)

Table 3.2: Masses in GeV for each bottomonium state using 1, 2, 3, 4, and 5 terms. The index j refers the term in the summation in Eq. (3.2.8). All values were determined using the CORFITTER package.

improvement of the error for $N_p = 4$ and $N_p = 5$ and there is little change between Fig. 3.3c and Fig. 3.3d. For the P-waves states there was another reason to select

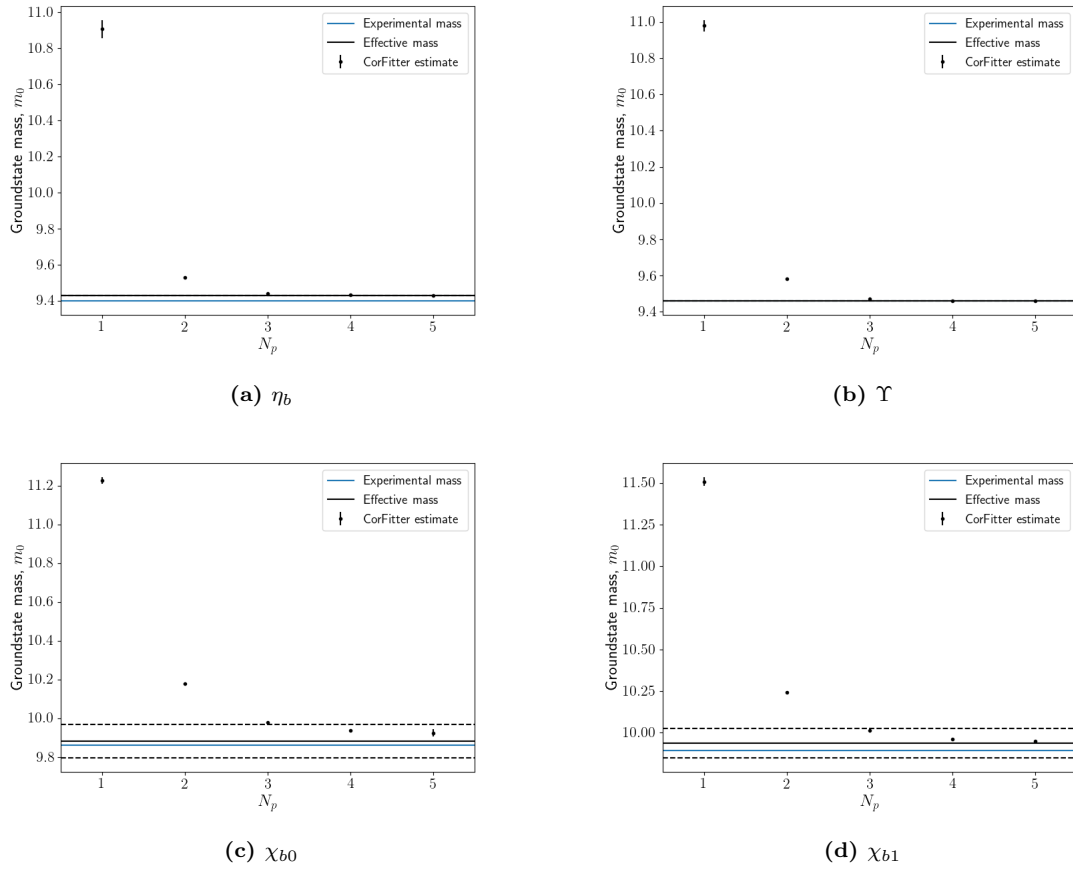


Figure 3.2: Groundstate masses for the bottomonium states using a multi-exponential fit with 1-5 terms. Experimental masses are represented by blue lines. Effective masses are represented by black solid lines and the uncertainty on these values by black dashed lines.

$N_p = 4$ as the most appropriate number of peaks. Although the error continues to reduce by notable amount, we noticed from Table 3.2 that uncertainty on m_5 for h_b and χ_{b1} was significant.

In summary, the CORFITTER fitting procedure has done an excellent job at determining groundstate masses of the bottomonium states if sufficient terms are used. However its estimate of higher order states begins to deteriorate if the number of terms is increased too far. The χ_{b0} and χ_{b1} still have percentage errors of 15.3% and 13.5% respectively. This is expected to some extent though, as although we have treated these configurations as zero temperature, that is not strictly true. Recall that for $N_\tau = 128$, $T = 47$ MeV and so there may be some thermal effects. Recall from Fig. 2.1b that at low temperatures we do not expect the spectral function for bottomonium at finite temperature to just be a set of delta functions. Instead we would expect resonances to begin broadening due to thermal effects. Furthermore,

N_p	η_b	Υ	h_b	χ_{b0}	χ_{b1}
1	32.1%	30.6%	36.9%	79.3%	69.7%
2	16.0%	15.0%	22.3%	49.2%	44.8%
3	14.1%	12.9%	20.1%	42.5%	40.3%
4	5.5%	5.0%	8.2%	16.8%	16.4%
5	4.8%	4.4%	6.2%	15.3%	13.5%

Table 3.3: Percentage error between reconstructed correlators and FASTSUM Gen2L correlators for $N_p = 1 - 5$

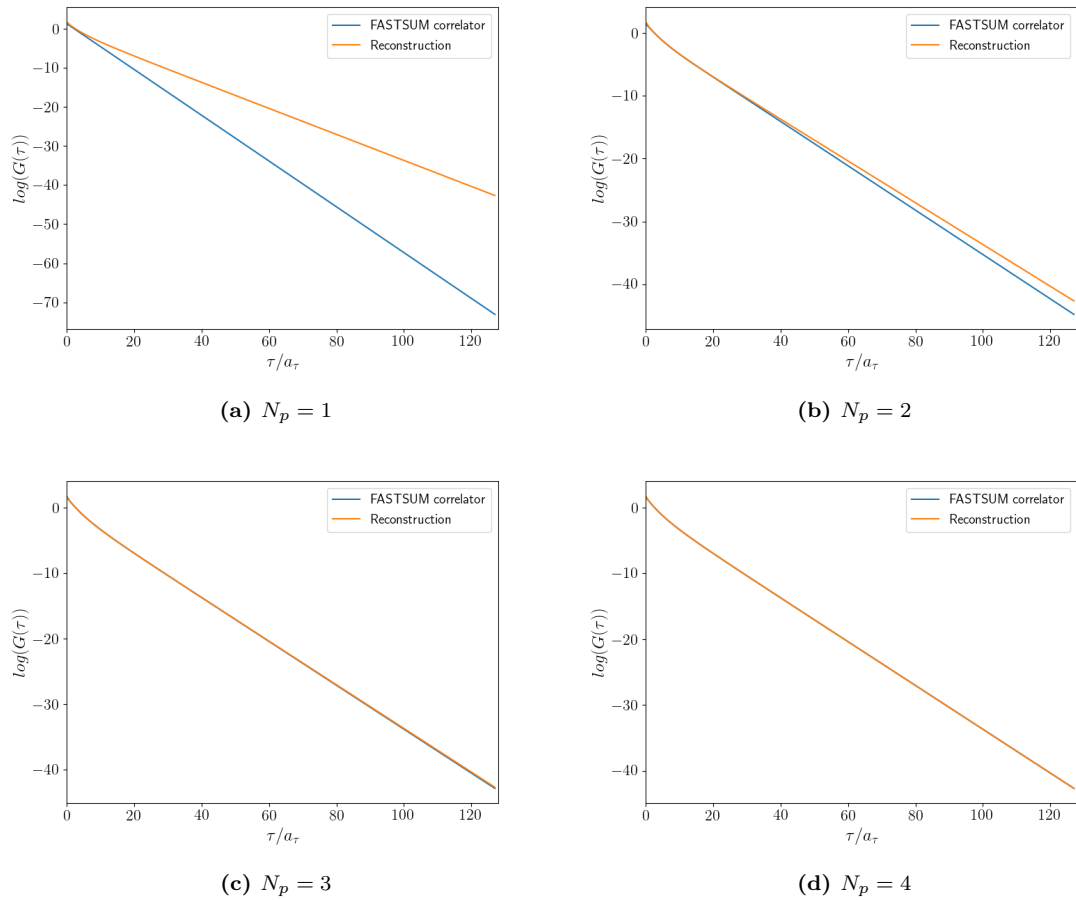


Figure 3.3: Comparison of reconstructed correlator for $N_p = 1 - 4$ and FASTSUM Gen2L correlator for the Υ .

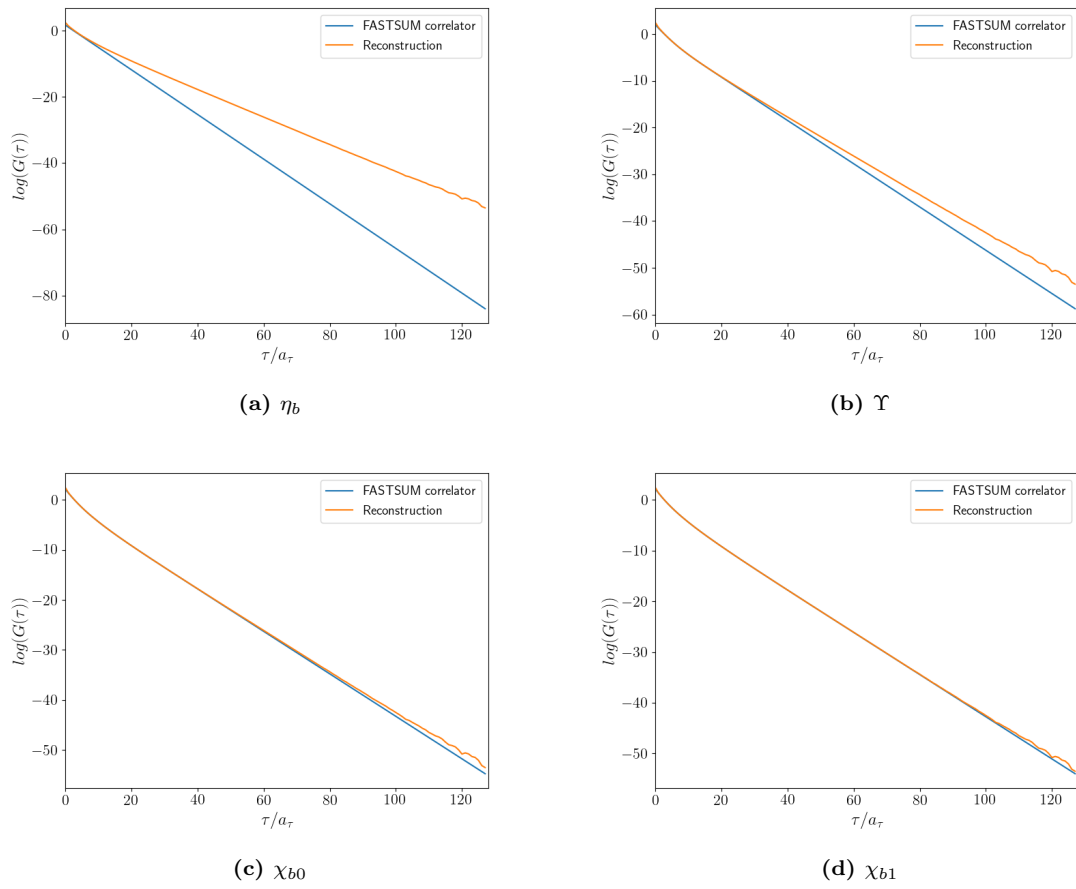


Figure 3.4: Comparison of reconstructed correlator for $N_p = 1 - 4$ and FASTSUM Gen2L correlator for the χ_{b1} .

we have not even considered including a continuum. Nevertheless, these results provide a reasonable starting point for the following work.

3.2.3 Thermal Modification

We conclude this chapter with the examination of thermal dependence of the correlators for η_b , Υ , χ_{b0} , and χ_{b1} channels. This was done using the ratio $G(T)/G(T_0)$, where $G(T_0) \approx G(N_\tau = 128)$ which corresponds to $T = 47$ MeV. This was done using bootstrap analysis for all the other temperatures from Table 1.1 with the central value for $G(T_0)$ and the configurations for $G(T)$ (see Fig. 3.5).

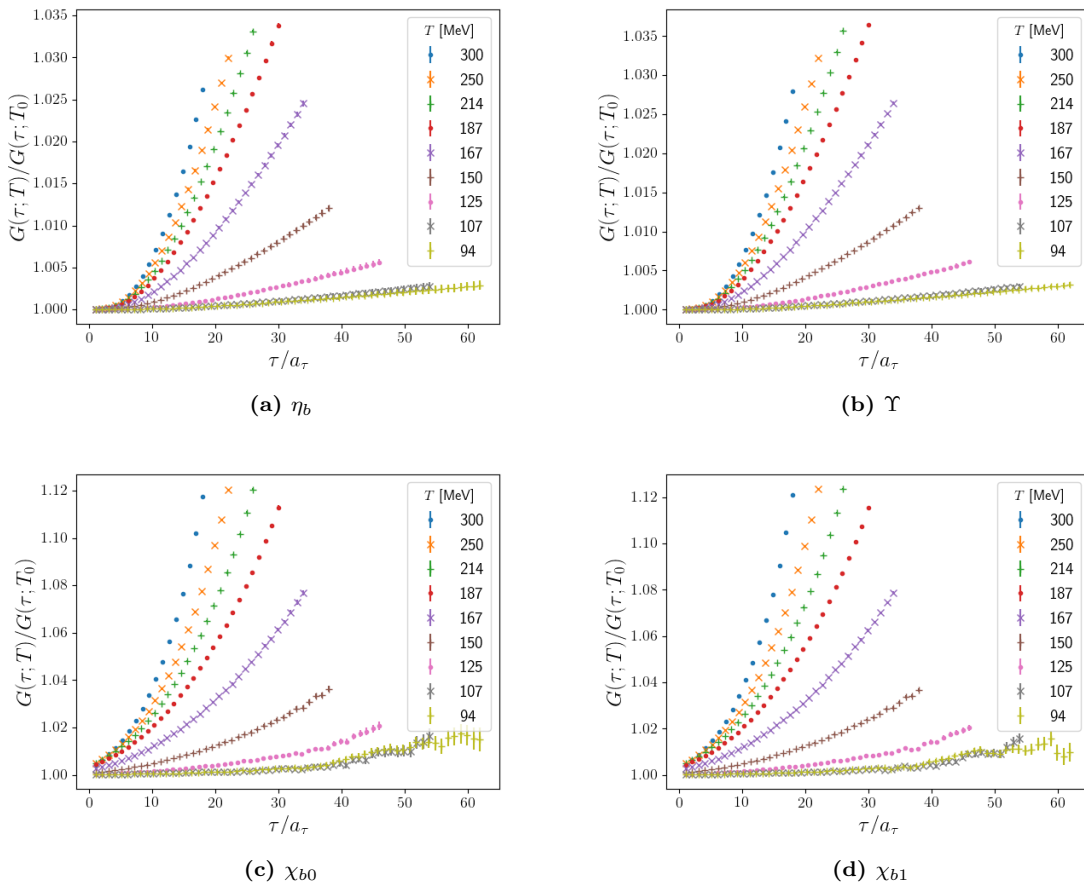


Figure 3.5: Thermal modification, $G(\tau; T)/G(\tau; T_0)$ with $T_0 = 47$ MeV, of the correlation functions.

There are two observations to be made regarding these plots. The first is common to all four states and that is that the largest increases occurs between 125 MeV and 187 MeV. Recall from Chapter 1 that the crossover region is estimated to lay within this range of temperatures. Thus we observe that the largest increase corresponds

to a change of phase. The second observation is the difference in vertical scales between the S and P-wave states. For both the η_b and Υ this is an increase of $\sim 4\%$ whereas for the χ_{b0} and χ_{b1} there is an increase $\sim 12\%$. Again this is expected from the discussion of sequential suppression in Chapter 1, as P-wave states dissociate at a lower temperatures than S-wave states, and therefore have a more pronounced thermal dependence. These observations, will be used in the following section for the generation of training data.

Chapter 4

Introduction to Kernel Ridge Regression

In recent years there has been a growing interest in machine learning and the techniques that fall under this term. If the ill-posed problem of spectral reconstruction is treated as 1D image reconstruction, a common problem in machine learning, then it should be possible to apply these methods. For this work we decided to develop a Kernel Ridge Regression (KRR) model using the work of [66] as inspiration.

4.1 Regression methods

4.1.1 Linear regression

The method of linear regression is simple, and one that most will be familiar with. Despite this familiarity, a brief description will still be provided so that the explanation of KRR will be easier to follow, by making use of analogies to linear regression. An example has also been provided to help with this explanation. For further insight, see [67, 68].

For any form of regression the aim is to develop a mapping between input data x and target data y . This requires a dataset of both input and target data, represented by the column vectors \mathbf{x} , and \mathbf{y} , respectively. The length of these vectors is equal to the number of datapoints, or observations we have, which we will refer to as N_{train} . In the simplest case we assume y to only be dependent on a single variable,

$$y = g(x) + \eta(\mu, \sigma), \quad (4.1.1)$$

where $g(x)$ is the true function that maps x to y and η is noise and is introduced to represent the uncertainty in the data. Usually this noise is chosen for each x from a normal distribution with mean $\mu = 0$ and suitable standard deviation σ . To

determine this mapping we assume

$$\mathbf{y} = \mathbf{f}(\mathbf{x})\mathbf{w}, \quad (4.1.2)$$

where \mathbf{w} is a vector of parameters to be determined, and $\mathbf{f}(\mathbf{x})$ is a matrix where each row corresponds to a different x and each column is a different function of x , $f(x)$. The form of $f(x)$ is arbitrary but for this example we will use powers of x so that

$$\mathbf{f}(\mathbf{x})\mathbf{w} = \sum_n w_n x^n. \quad (4.1.3)$$

In this case $f_n(\mathbf{x}) = \mathbf{x}^n$. Note that polynomials are a poor choice for $f(x)$, especially if there is an error associated with the data, as it is numerically unstable. The problem worsens with the number of terms used. However for the purpose of this example it will be adequate. We can replace $\mathbf{f}(\mathbf{x})$ with the matrix \mathbf{X} . If the number of terms of the polynomial is defined as N , then we see that \mathbf{w} is a column vector of length N and \mathbf{X} is a matrix of size $N_{\text{train}} \times N$ and the j -th column now corresponds \mathbf{x}^j .

The first step of regression is to determine the optimal values of \mathbf{w} , $\hat{\mathbf{w}}$, which is usually done by minimizing a cost function or maximizing likelihood. We refer to this step as the training the model. For Linear Least-Squares (LLS) regression we choose the cost function E ,

$$E(\mathbf{w}) = |\mathbf{y} - \mathbf{X}\mathbf{w}|^2. \quad (4.1.4)$$

Differentiating the above with respect to w_n and setting this gradient to 0 gives [69],

$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (4.1.5)$$

It becomes possible to predict new values of y using

$$y_{\text{new}} = \mathbf{f}(x_{\text{new}})\hat{\mathbf{w}}. \quad (4.1.6)$$

We have reverted to using $\mathbf{f}(x)$, as the method is the same regardless of definition of $\mathbf{f}(x)$. For the following example we have chosen $y_{\text{true}} = \sin(x)$ for the true form of the target data, though ordinarily we would not know the true form of y which is why we resort to regression. For each observation of x the corresponding y is

determined with the addition of noise, $\eta(0, 0.1)$. Therefore, for a single element of \mathbf{y} ,

$$y = \sin(x) + \eta(0, 0.1) = \sum_n w_n x^n. \quad (4.1.7)$$

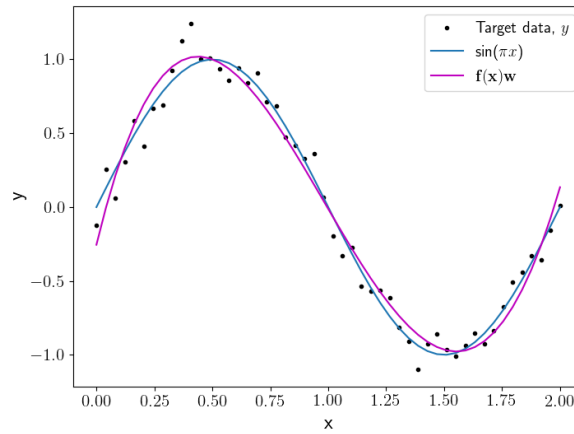
In Fig. 4.1a, b, and c we have plotted three separate models with $n = 4, 10,$ and 16 for $\sum_n w_n x^n$ respectively. The input data consisted of 50 values of x , hence $N_{train} = 50$. We see that even with only 4 terms it is possible to obtain a curve that closely resembles the true function. This is quite the positive result, especially when we consider that $\hat{\mathbf{w}}$ is generated using the noisy data. As we increase the number of terms to 10, the curve generated by our model more closely resembled the true function. However, note the final curve that was generated using 16 terms and how it diverges from the true function to pass through the noisy data instead. Recall that we are using noisy data to train our model, and in reality we would not know what the true function. One issue with this is that to the model, the path of the noisy data is the true function. Therefore it will try to best match that. As we make the model more complex, for example by adding more terms, it will be better able to match the noisy data. This behaviour is referred to as overfitting. The main issue with this is that when we introduce new data, a new x , the model will be unlikely to make a reasonable prediction of the corresponding y . In order to resolve this we need to regulate the model.

4.1.2 Ridge Regression

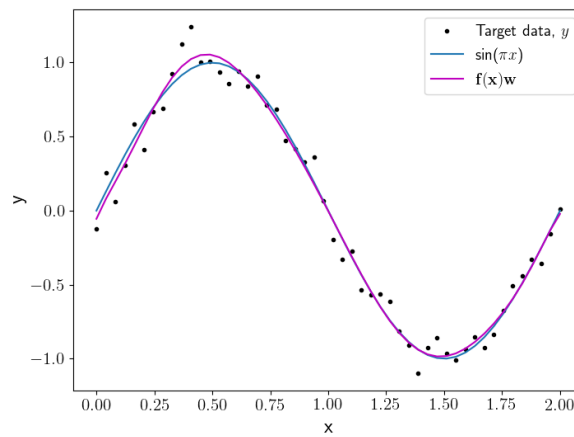
Despite the success of LLS in the example above, there are two flaws with LLS. The first is overfitting which we discussed above. The second is that when \mathbf{X} is singular it becomes impossible to calculate $\hat{\mathbf{w}}$ as $\mathbf{X}^T \mathbf{X}$ is no longer invertible. To resolve both these issues, we implement Ridge Regression (RR), which is identical to the above apart from an additional term in the cost function,

$$E(\mathbf{w}) = |\mathbf{y} - \mathbf{X}\mathbf{w}|^2 + \lambda |\mathbf{w}^T \mathbf{I}\mathbf{w}|^2, \quad (4.1.8)$$

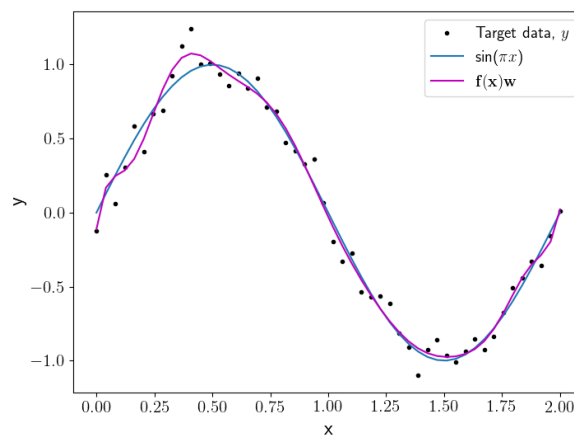
where \mathbf{I} is the identity matrix of size $N \times N$, and λ is the regularization constant. Ridge Regression is a special case of the more general Tikhonov regularization [70, 71]. See [68] for a discussion on LASSO regression, an alternative to RR.



(a) $n = 4$



(b) $n = 10$



(c) $n = 16$

Figure 4.1: Example of Linear Least-Squares regression using polynomials with 4(a), 10(b), and 16(c) terms.

Determining $\hat{\mathbf{w}}$ is done in the same manner as above and hence it follows,

$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{x}^T \mathbf{y} \quad (4.1.9)$$

We have plotted the models, see Fig. 4.2(a), (b), and (c), again using $n = 4, 10,$ and $16,$ generated using the new definition of $\hat{\mathbf{w}}$ with $\lambda = 0.1.$

For both $n = 4$ and $n = 10$ we see little change between the LLS models and the RR models. However for $n = 16,$ the curve produced by our model has smoothed out to resemble Fig. 4.2b. Although this is certainly an improvement compared to Fig. 4.1c, we need to consider whether generating a model with more terms is necessary as there is little difference between the $n = 10$ and $n = 16$ cases. We have already commented that the numerical instability that comes from trying to fit polynomials increases with the number of terms.

4.1.3 Kernel Ridge Regression

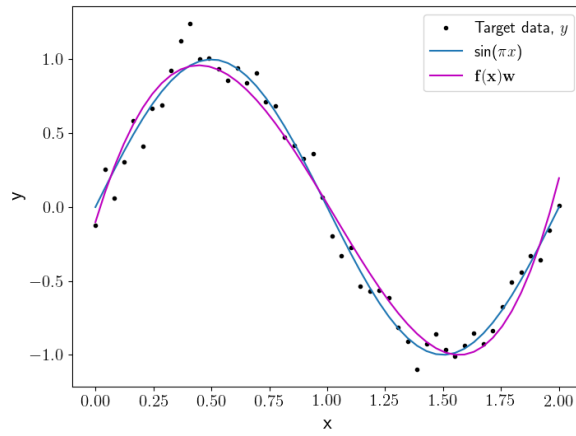
Unfortunately both the above suffer from the same problem; how many basis functions to use in $\mathbf{f}(x).$ Rather than attempting to determine this we introduce Kernel Ridge Regression (KRR). In principle, the method is identical to the above; minimizing a cost function to determine parameters and using these to make predictions. The main difference comes from replacing the input data \mathbf{x} with a kernel function $C(x_i, x_j),$ from which the method gets its name. There is no standard kernel, but the function must be a distance between elements x_i and $x_j.$ For this reason common kernel functions involve L-1 or L-2 norms. For this example we used the following kernel function,

$$C_{ij} = C(x_i, x_j) = \exp(-\gamma [x_i - x_j]^2) \quad (4.1.10)$$

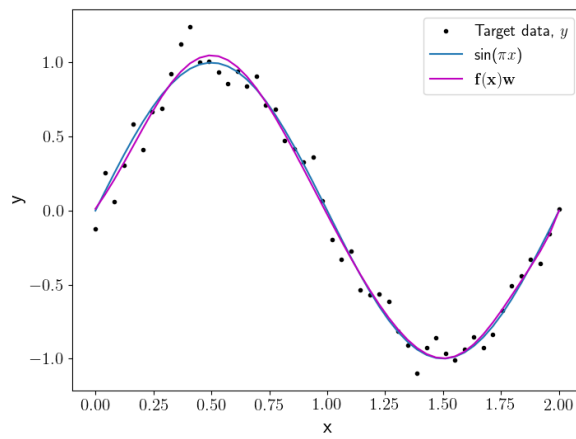
where γ is the scale length and is used control the magnitude of correlations between x_i and $x_j.$ The parameters λ and γ are the model's hyperparameters. Now, rather than testing models with different $\mathbf{f}(x),$ we only need to determine a suitable value for $\gamma.$ By using all elements \mathbf{x} for x_i and x_j in Eq.(4.1.10), we generate a matrix \mathbf{C} of size $N_{\text{train}} \times N_{\text{train}}.$

The cost function is redefined as,

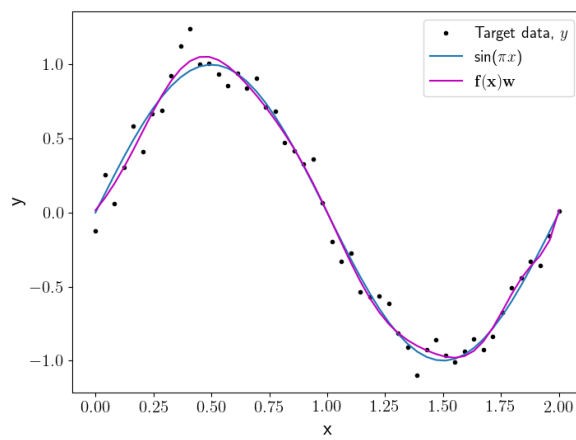
$$E(\boldsymbol{\alpha}) = |\mathbf{y} - \mathbf{C}\boldsymbol{\alpha}|^2 + \lambda |\boldsymbol{\alpha}^T \mathbf{C}\boldsymbol{\alpha}|^2, \quad (4.1.11)$$



(a) $n=4$



(b) $n=10$



(c) $n=16$

Figure 4.2: Example of Ridge regression with $\lambda = 0.05$ using polynomials with 4(a), 10(b), and 16(c) terms.

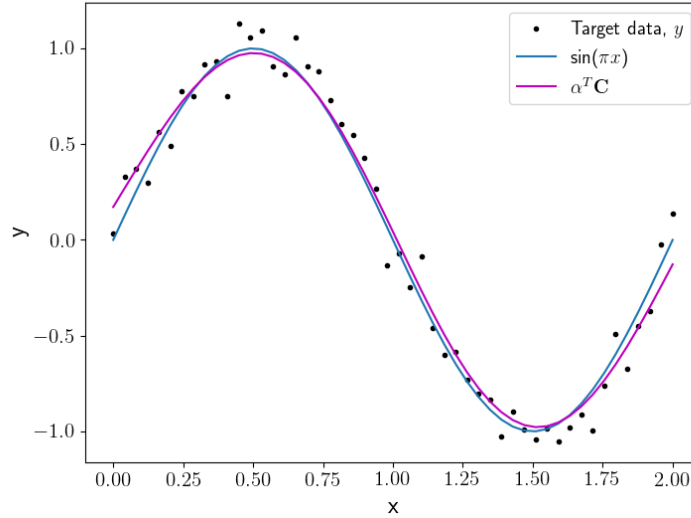


Figure 4.3: Example of Kernel Ridge regression using a radial basis function (RBF) kernel with $\gamma = 0.1$, and $\lambda = 0.1$.

where $\boldsymbol{\alpha}$ is a matrix of parameters that replaces \mathbf{w} , with dimensions $N_{\text{train}} \times 1$ for this example. Regardless of the form of $C(x_i, x_j)$, it follows that the optimal matrix of parameters, $\hat{\boldsymbol{\alpha}}$, is defined as

$$\hat{\boldsymbol{\alpha}} = (\mathbf{C} + \lambda \mathbf{I})^{-1} \mathbf{y} \quad (4.1.12)$$

Finally we define a prediction using new data as

$$y_{\text{new}} = \mathbf{C}_{\text{new}}^T \hat{\boldsymbol{\alpha}} \quad (4.1.13)$$

where elements of \mathbf{C}_{new} are given by Eq.(4.1.10) but with $x_j \rightarrow x_{\text{new}}$ so that \mathbf{C}_{new} is a column vector of length N_{train} . The example in Fig. 4.3 was done using $\gamma = 0.1$. In Table 4.1 we include the mean absolute error (MAE) of each model for y and y_{true} , denoted by $\text{MAE}_{\text{target}}$ and MAE_{true} respectively, for comparison.

With this table of errors it becomes easier to determine how the various models compare. Firstly, the error from the target data decreases as n increases for both LLS and Ridge regression models. This is expected, since there will exist a model with more term capable of producing a curve that goes through all data points with $\text{MAE}_{\text{target}} = 0$. This however would be an extremely overfitted model. Regarding MAE_{true} , all errors are lower than the corresponding $\text{MAE}_{\text{target}}$ so even with the noise, all models are capable of producing good estimates of the true function. We

Model	MAE _{target}	MAE _{true}
LLS ₄	0.0932	0.0624
LLS ₁₀	0.0738	0.0218
LLS ₁₆	0.0636	0.0370
RR ₄	0.1001	0.0652
RR ₁₀	0.0759	0.0217
RR ₁₆	0.0695	0.0280
KRR	0.0848	0.0388

Table 4.1: Mean Absolute Error on each model, determined using the noisy target data, y , and true values, $\sin(x)$.

also remark that for both LLS and Ridge regression, the $n = 10$ model is the best. The addition of the regulator has worsened the $n = 4$ and $n = 10$ models, though this is to be expected due to its effect on the cost function. The improvement for the $n = 16$ is negligible. Finally, although the KRR model does not have the lowest errors of the models, the convenience of not having to decide the number of terms and not having to be concerned about the numerical instability of the polynomial make it the most practical of the models.

In order to relate KRR to spectral reconstruction, it is useful to substitute x and y with the functions $G(\tau)$ and $\rho(\omega)$ respectively. With this substitution we redefine the kernel, $C(x_i, x_j) \rightarrow C(G_i(\tau), G_j(\tau))$ with the form

$$C_{ij} = \exp \left(-\gamma \sum_{n_\tau=4}^{N_\tau} (G_i(a_\tau n_\tau) - G_j(a_\tau n_\tau))^2 \right). \quad (4.1.14)$$

For the time being it is assumed that

$$\rho(\omega) = \mathbf{C}^T \boldsymbol{\alpha}. \quad (4.1.15)$$

Note that there is no addition of noise for the time being. The cost function becomes,

$$E = |\rho - \mathbf{C}^T \boldsymbol{\alpha}|^2 + \lambda |\boldsymbol{\alpha}^T \mathbf{C} \boldsymbol{\alpha}|^2, \quad (4.1.16)$$

Hence, the optimal matrix is given by,

$$\hat{\boldsymbol{\alpha}} = (\mathbf{C} + \lambda \mathbf{I})^{-1} \rho \quad (4.1.17)$$

Here $\boldsymbol{\alpha}$ has dimensions $N_{train} \times N_\rho$, where N_ρ is the number of points used to generate ρ , typically $\mathcal{O}(1000)$.

4.2 Data Generation

4.2.1 Spectral functions

Before we begin to implement KRR though, we require training data in order to train the model and generate α . The generation of the training data is a key step, not just for KRR, but for many machine learning methods that require training. Disregarding the obvious reason that any such algorithm is incapable of making predictions without training, and hence training data, there is a more subtle importance to the training data. It is only at this stage that prior knowledge of the system can be incorporated to provide restrictions on the later predictions.

For now the training data was generated as follows. Each spectral function is a superposition of 4 Gaussian functions,

$$\rho(\omega) = \sum_r^{N_p} A_r \exp\left(-\frac{(\omega - m_r)^2}{\Gamma_r}\right), \quad (4.2.1)$$

where m , Γ , and a represent mass, width and amplitude respectively. The upper limit N_p is the number of Gaussian functions, or number of peaks. The reuse of N_p from Chapter 3 is intentional, as these Gaussians will substitute for $\delta(\omega - m_j)$ in Eq. (3.2.8). In fact, the reason Gaussian functions were chosen was due to their resemblance to a δ function in the limit $\Gamma \rightarrow 0$. In Sec. 3.2.2 we have already discussed why $N_p = 4$ is a more appropriate choice, particularly for S-wave states which we will be focusing on for this initial testing of a KRR model. For simplicity we have not included a continuum or any low frequency behaviour to represent a transport peak. Instead our concern for the time being will be to generate mock functions that mimic the bottomonium groundstate and excited states. A dataset of 20000 functions was generated.

The results of Chapter 3 have been used as inspiration for selecting the values of the parameters themselves. Given the results in Table 3.2 the masses have an approximate range of 9 - 16 GeV. Rather than using a uniform distribution, an exponential distribution was chosen instead. This was to allow for peaks to appear beyond 16 GeV, as well as favouring lower energy for the masses. To generate the

masses, values were chosen using the following distribution,

$$p(\omega) = \frac{1}{\kappa} \exp\left(-\frac{\omega}{\kappa}\right), \quad (4.2.2)$$

where κ is the scale parameter of the distribution. In this case $\kappa = 1$. An additive constant of 8.5 was added to the selected values in order for them to represent masses in GeV, M_{GeV} . Before the spectral functions can be generated, they need converting to lattice units which was done using,

$$m_{\text{lat}} = \frac{M_{\text{GeV}} - E_0}{a_\tau^{-1}}. \quad (4.2.3)$$

Widths, Γ , were chosen by first taking a random number, x , from a uniform distribution such that $-4 \leq x \leq 0$ which is related to Γ by $\Gamma = 10^x$. By selecting Γ this way, it is hoped that we will generate a balanced selection of narrow peaks, $\Gamma = \mathcal{O}(10^{-4})$, and broad peaks, $\Gamma = \mathcal{O}(1)$. Selecting a number from 10^{-4} to 1, directly would have led to broader peaks dominating. Amplitudes were initially chosen from a uniform distribution such that $0.1 \leq A' \leq 10$, then scaled accordingly to get A . To understand why it needs scaling, consider Eq. (3.2.4) for $\tau = 0$,

$$G(0) = \int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{2\pi} \rho(\omega). \quad (4.2.4)$$

This sets the condition that $\rho(\omega)$ must be normalised to $G(0)$. From the Gen2L ensembles we find that $G(0) \approx 6$ for η_b, Υ , and h_b , whilst $G(0) \approx 12$ for χ_{b0} and χ_{b1} . In order to satisfy this condition, we need to decide on appropriate values for ω_{\max} and ω_{\min} . As we are working on the lattice, the range for ω cannot be infinite and must instead be chosen judiciously. For now, we determine the range using the unimproved NRQCD lattice dispersion relation,

$$a_\tau E(\mathbf{n}) = -\log\left(1 - \frac{\hat{p}(\mathbf{n})^2}{2\xi a_s M_b}\right), \quad (4.2.5)$$

where $\hat{p}(\mathbf{n})^2$ is the lattice momenta at site \mathbf{n} in the first Brillouin zone, defined as,

$$\hat{p}(\mathbf{n})^2 = 4 \sum_{i=1}^3 \sin^2\left(\frac{\pi n_i}{N_s}\right), \quad n_i = -\frac{N_s}{2} + 1, \dots, \frac{N_s}{2}. \quad (4.2.6)$$

To determine a range for ω , we decided to initially use a range that corresponded to the range of energy $\Delta E(\mathbf{n})$. This depends on the maximum and minimum momenta which are 12 ($\mathbf{n} = (N_s/2, N_s/2, N_s/2)$) and 0 ($\mathbf{n} = (0, 0, 0)$) respectively. This leads

to $a_\tau \Delta E(\mathbf{n}) = 1.86$. In order to use this to set range for ω , we introduce the free lattice spectral functions for S-waves states

$$a_s \rho_S(\omega) = \frac{4\pi N_c}{\xi N_s^3} \sum_{\mathbf{n}} \delta(a_\tau \omega - 2a_\tau E(\mathbf{n})), \quad (4.2.7)$$

where N_c is the number of colour charges, usually 3. In this case it does not matter what N_c is as we are only concerned about the δ function which is only non-zero for $\omega = 2E(\mathbf{n})$. Hence $a_\tau (\omega_{\max} - \omega_{\min}) = 3.72$. Rather than setting $\omega_{\min} = 0$ and $\omega_{\max} = 3.72$, there was an offset of 0.0005, so $\omega_{\min} = 0.0005$ and $\omega_{\max} = 3.7205$. The offset from 0 is for later computational convenience. Using these values, ρ is normalised by first selecting the parameters from the distributions above and substituted into Eq. (4.2.4). We define a normalization constant z , such that,

$$G(0) = z \int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{2\pi} \rho(\omega). \quad (4.2.8)$$

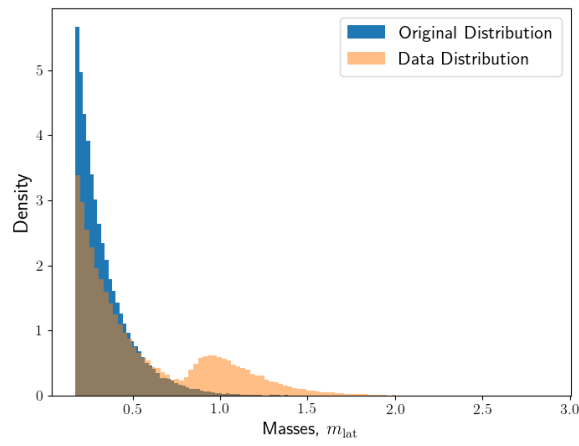
All original amplitudes scaled accordingly, $A'_r \rightarrow A_r = zA'_r$.

There are several constraints imposed. The first of these is that the smallest masses corresponds to the smallest widths and largest amplitudes. For the largest masses, the inverse is true. To justify this, recall the sketches from Chapter 2 in Fig. 2.1 where higher order states are more affected by thermal effects. Next, $\rho(\omega_{\min}) = \rho(\omega_{\max}) \approx 0$ and $m_j < \omega_{\max}$ had to be obeyed. These together ensure that the spectral function can only be non-zero between our chosen range for ω . The final condition set was that $\rho > 10^{-100}$ between the range $\omega_{\min} \leq \omega \leq \omega_{\max}$, the justification for this will follow shortly but it is mentioned here for transparency.

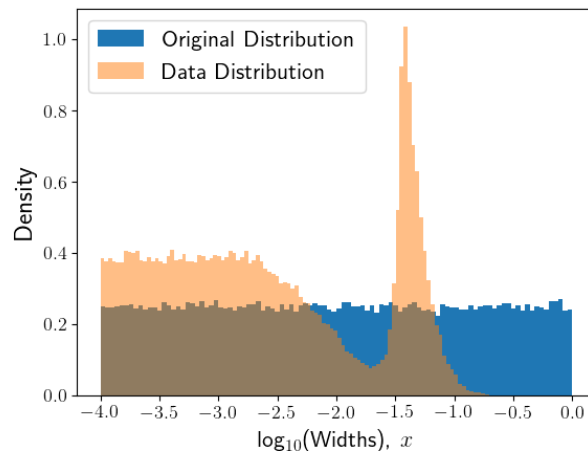
In Fig. 4.4 the distributions of the selected parameters are displayed, alongside a representation of the original distributions in blue. Note that none of the distributions of selected parameters follow the original. It was expected that the amplitude would vary in terms of range due to the scaling, but that the shape of the distribution is closer to an exponential distribution is unexpected. The exact reason for these deviations is currently unknown but it is in part due to the imposed constraints.

4.2.2 Representations of spectral functions

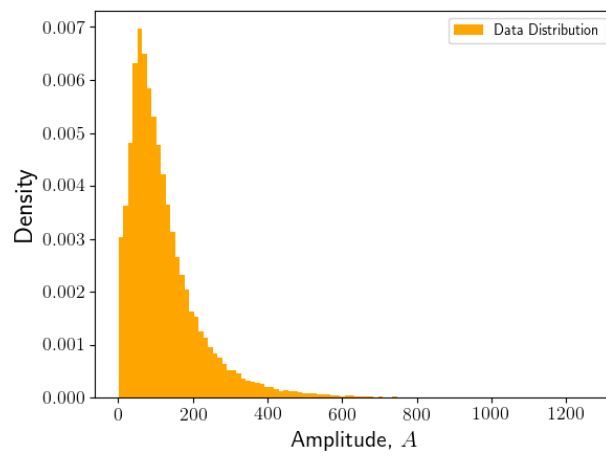
Clearly any predictions of ρ using one of the above regression methods will produce a discrete depiction of the spectral function. Hence we need to decide how many



(a)



(b)



(c)

Figure 4.4: Distribution of unconstrained parameters (blue) and distribution of selected parameters (orange)

points are needed to produce a seemingly continuous spectral function. For this work, we decided $\mathcal{O}(1000)$ points would be sufficient. Given the correlators have N_τ points with ranges from 8 to 128, we see that this inversion is ill-posed as we stated at the end of Chapter 2. To help with this ρ can be represented by a set of parameters instead which we will refer to as $\boldsymbol{\theta}$, with length N_θ . This alters Eq. (4.1.17),

$$\hat{\boldsymbol{\alpha}} = (\mathbf{C} + \lambda \mathbf{I})^{-1} \boldsymbol{\Theta} \quad (4.2.9)$$

where $\boldsymbol{\Theta}$ is a matrix of parameters for all ρ , with size $N_{\text{train}} \times N_\theta$. The j -th row of $\boldsymbol{\Theta}$ would correspond to $\boldsymbol{\theta}_j$, the parameters for the j -th training ρ . The matrix $\hat{\boldsymbol{\alpha}}$ now has dimensions $N_{\text{train}} \times N_\theta$. We have actually already introduced the first of our representations above where $\boldsymbol{\theta} = (m_r, \Gamma_r, A_r)$.

An alternative representation uses Bryan's method [72], which involves taking the Singular Value Decomposition (SVD) of a thermal kernel $K(\tau', \omega,)$ to generate a set of orthogonal basis functions, $\mathbf{f}(\omega)$. This representation will henceforth be referred to as the SVD representation. We use τ' in the kernel because this kernel may correspond to a different N_τ than that of the correlators used for training or predictions. In this work $N_{\tau'} = N_\theta = 80$. We then define the spectral function,

$$\rho_{\text{SVD}}(\omega) = m(\omega) \exp \left(\sum_{j=1}^{80} f_j(\omega) c_j \right), \quad (4.2.10)$$

where $m(\omega)$ is the default model. In this case $\boldsymbol{\theta} = (c_1, \dots, c_{80})$ is a set of coefficients determined by,

$$\boldsymbol{\theta} = \mathbf{f}(\omega) \cdot \log \left(\frac{\rho_{\text{SVD}}(\omega)}{m(\omega)} \right), \quad (4.2.11)$$

which makes use of the orthogonality of the basis functions. To see this, consider the contribution to $\rho(\omega)$ from the j -th basis function and coefficient,

$$\begin{aligned} \rho_{\text{SVD}}^{(j)}(\omega) &= m(\omega) \exp(f_j(\omega) c_j) \\ c_j f_j(\omega) &= \log \left(\frac{\rho_{\text{SVD}}^{(j)}(\omega)}{m(\omega)} \right) \end{aligned} \quad (4.2.12)$$

$$c_j f_i(\omega) \cdot f_j(\omega) = c_j \delta_{ij} = f_j(\omega) \cdot \log \left(\frac{\rho_{\text{SVD}}^{(j)}(\omega)}{m(\omega)} \right)$$

In the final line, we make use of the fact that $\mathbf{f}(\omega)$ is a set of orthogonal basis functions. Note that the logarithm in Eq. (4.2.10) is the reason for our requirement

that $\rho(\omega) > 10^{-100}$, otherwise $\log(\rho(\omega))$ cannot be computed numerically. The default model is set by to normalizing $\rho(\omega)$ according to

$$m(\omega) = G(0) / \left[\Delta\omega \sum_{\omega_{\min}}^{\omega_{\max}} \exp \left(\sum_{j=1}^{80} f_j(\omega)\theta_j \right) \right], \quad (4.2.13)$$

where $\Delta\omega = \omega_{n+1} - \omega_n$. Therefore $m(\omega)$ in this work is just a normalisation constant, B .

The final part of the data generation is calculating the corresponding correlators for each of the mock spectral functions. Rather than using Eq. (3.2.4), the discrete definition was used since ρ_{SVD} is discrete,

$$G_{\text{SVD}}(\tau) = \Delta\omega \sum_{\omega_{\min}}^{\omega_{\max}} K(\tau, \omega) \rho_{\text{SVD}}(\omega) = B \Delta\omega \sum_{\omega_{\min}}^{\omega_{\max}} e^{-\tau\omega} \exp \left(\sum_{j=1}^{80} f_j(\omega)\theta_j \right). \quad (4.2.14)$$

With the mock data generated, it is now time to implement KRR.

4.3 A first look a KRR

Since ρ is now defined terms of the parameters $\boldsymbol{\theta}$, the cost function and optimal matrix are as follows

$$E = |\boldsymbol{\Theta} - \mathbf{C}^T \boldsymbol{\alpha}|^2 + \lambda |\boldsymbol{\alpha}^T \mathbf{C} \boldsymbol{\alpha}|^2, \quad (4.3.1)$$

$$\hat{\boldsymbol{\alpha}} = (\mathbf{C} + \lambda \mathbf{I})^{-1} \boldsymbol{\Theta}, \quad (4.3.2)$$

The kernel, \mathbf{C} , is still as written in Eq. (4.1.14).

In order to test the predictability of a KRR, a second dataset is required that shares no functions with the original training set. This set is referred as the testing set with size N_{test} . The predicted $\boldsymbol{\Theta}$ are given by,

$$\boldsymbol{\Theta}_{\text{pred}} = \mathbf{C}_{\text{test}}^T \hat{\boldsymbol{\alpha}}, \quad (4.3.3)$$

where an element of \mathbf{C}_{test} is defined as,

$$C_{ij}^{(\text{test})} = \exp \left(-\gamma \sum_{\tau} (G_i(\tau) - G_j(\tau))^2 \right). \quad (4.3.4)$$

The index $i = 1, \dots, N_{\text{train}}$ and the index $j = 1, \dots, N_{\text{test}}$

In the following example, $N_{\text{train}} = 10000$, $N_{\text{test}} = 1000$, the SVD representation was used with $N_{\theta} = 80$, and $N_{\tau} = 40$. The normalization constant B is not included

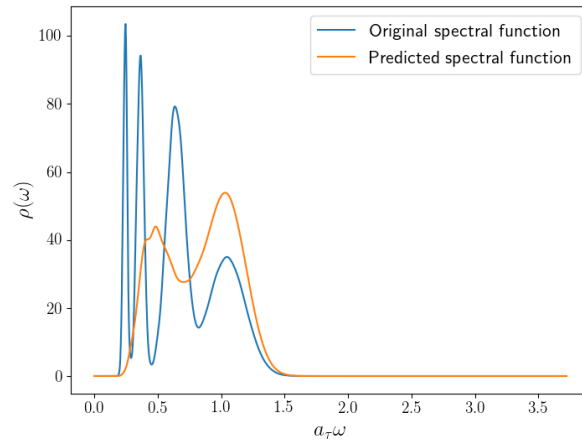
in $\boldsymbol{\theta}$ as it can be determined afterwards. The first 4 timeslices are neglected for the same reason given in Sec. 3.2.2; they are affected by lattice artifacts. The hyperparameters (σ, λ) were both set to 1. In order to gain insight into the predictive ability of the KRR model, the mean-squared-error (MSE) was used. For now it will be determined using the actual values of the coefficients and the predicted values $\boldsymbol{\theta}_{\text{test}}$, rather than the corresponding ρ ,

$$\text{MSE}_{\text{test}} = \frac{1}{N_{\text{test}}} \frac{1}{N_{\theta}} \sum_{n=0}^{N_{\text{test}}} \sum_{j=0}^{N_{\theta}} \left(\theta_{j,n}^{(\text{test})} - \theta_{j,n}^{(\text{pred})} \right)^2. \quad (4.3.5)$$

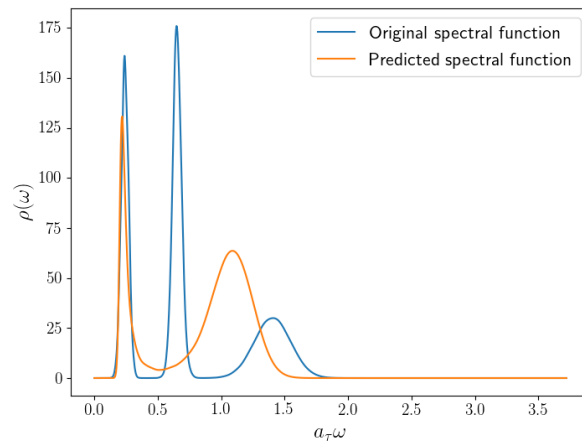
The value of the MSE alone will not mean much but as changes to the current KRR model are made, the error should decrease. The definition of the training MSE is the same after the substitution $N_{\text{test}} \rightarrow N_{\text{train}}$ and $\boldsymbol{\theta}_{\text{test}} \rightarrow \boldsymbol{\theta}_{\text{train}}$. For the MSE of the coefficients for a single spectral function, drop the factor of $1/N_{\text{test}}$, or $1/N_{\text{train}}$, and the appropriate summation.

The MSEs for the entire training and testing sets, alongside the lowest, median, and highest MSEs in the testing set, MSE_{low} , MSE_{med} , and MSE_{high} respectively, are shown in Table 4.2. These numbers alone mean very little so the spectral functions for three example MSEs have been plotted in Fig. 4.5. These plots provide a visual interpretation of the KRR model's current predictive ability. The SVD coefficients were used instead of the parameters (m_r, Γ_r, A_r) in recreating the original spectral functions, which is especially obvious in Fig. 4.5c as there are more than 4 peaks. This is because at this stage it is more important to demonstrate the KRR model's ability to predict the correct coefficients rather than its ability to recreate the correct ρ . By low and high MSE, this is relative to the MSEs of the test set and not a universal standard. Going forward, these three sets of coefficients will be used to determine if the changes to be made are improving the KRR model or not.

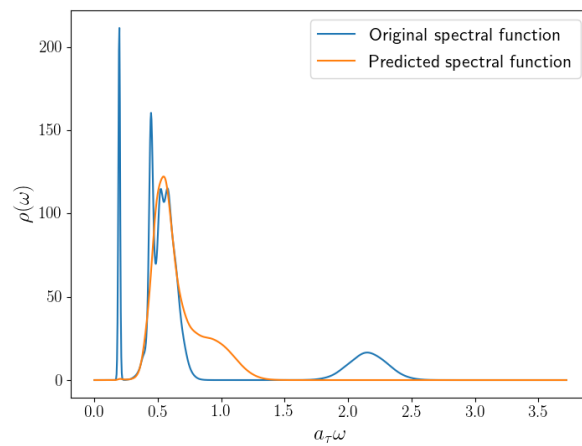
Before moving onto improving the KRR model, there are observations to be made even from this simple example. The model struggles as the separation of the peaks increases, though when comparing the spectral functions directly, it would be difficult to distinguish which of these has the largest error. Observe as well that the median MSE is much less than the mean MSE for the test set. Therefore it would seem that the mean MSE is skewed by a minority of the dataset. This means when it comes to minimizing the MSE, it may be that any decrease to the mean MSE



(a)



(b)



(c)

Figure 4.5: Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (1, 1)$.

(λ, σ)	(1, 1)
$\text{MSE}_{\text{train}}$	4425.97
MSE_{test}	4547.88
MSE_{low}	15.63
MSE_{med}	2847.49
MSE_{high}	40179.18

Table 4.2: MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed.

come about by changes to the sets of coefficients with highest MSE.

There is no way to check how this choice of hyperparameters has affected every function in the test set without looking at each. Given that $N_{\text{test}} = 1000$, this would be inefficient. To observe any effects of changing the KRR model, these three examples will be used. This leaves the question of how to improve the KRR model. As of yet no consideration has been given to what values γ , and λ should take [73]. The decision to use 1 for the initial value of both hyperparameters was just to provide an example of the KRR model, albeit a bad example.

4.3.1 Regularization constant

As discussed above, the purpose of regularization is to prevent singular matrices so that $\hat{\boldsymbol{\alpha}}$ can be calculated and also reduces overfitting. In using functions for our input and target data, the concept of overfitting becomes more abstract but is still a concern as $\lambda \rightarrow 0$. For the first attempt at improving the KRR model we vary λ using a range of $-10 \leq \log_{10}(\lambda) \leq 0$. Smaller values of λ have a risk of $(\mathbf{C} + \mathbf{I}\lambda)$ becoming singular. The training and testing MSEs have been plotted in Fig. 4.6 against λ .

Clearly the minimum testing MSE corresponds to $\lambda = 10^{-10}$. A new KRR model was determined using this value of λ and the new MSEs for the training and testing sets, as well as MSE_{low} , MSE_{med} , and MSE_{high} are displayed in Table 4.3. There is a clear improvement across the five MSE with the exception of MSE_{low} .

However when we consider the new reconstructed spectral functions in Fig. 4.7, it appears to suggest a different behaviour. In Fig. 4.7a and 4.7b the predicted

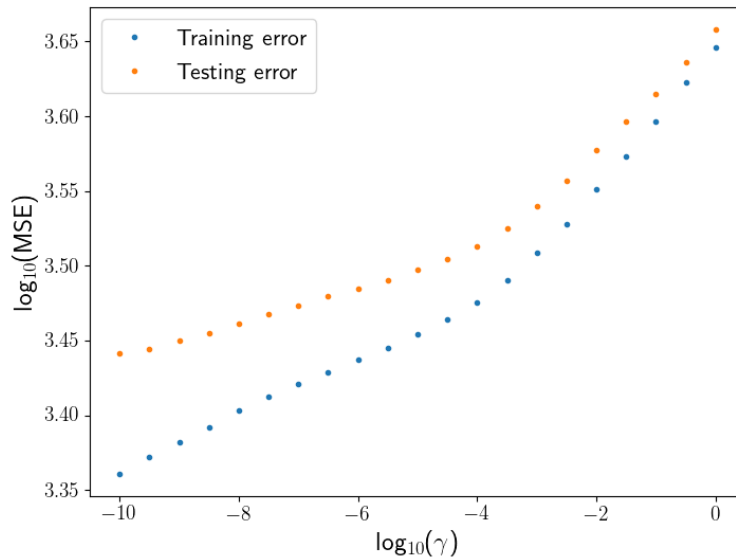


Figure 4.6: Training and testing MSEs as a function λ . Note the values on both axes are logarithmic.

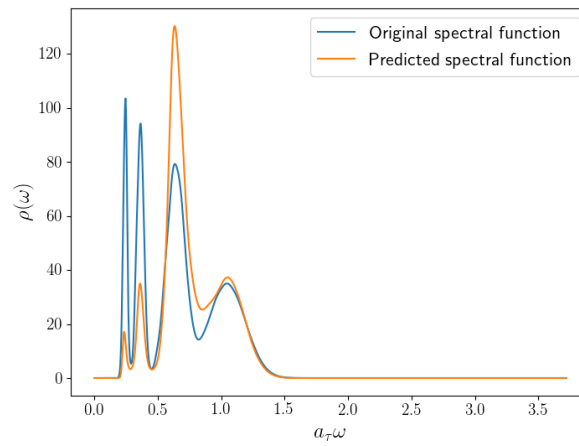
(λ, γ)	(1,1)	$(10^{-10}, 1)$
$\text{MSE}_{\text{train}}$	4425.97	2295.77
MSE_{test}	4547.88	2762.75
MSE_{low}	15.63	200.05
MSE_{med}	2847.49	919.19
MSE_{high}	40179.18	389.84

Table 4.3: MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed.

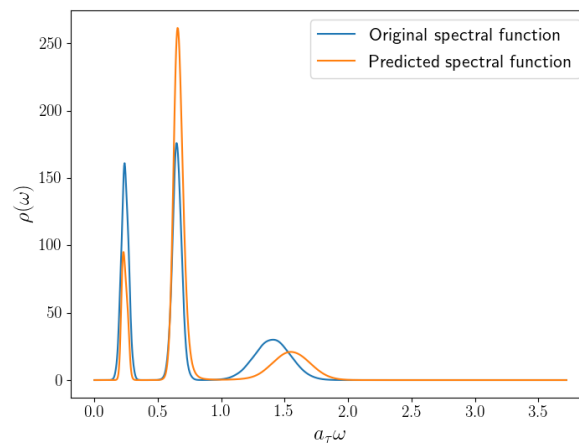
spectral functions have the correct number of peaks and in the correct positions. Meanwhile in Fig. 4.7c any similarity with the low frequency region of the original spectral function has been lost. Now only the position of the final peak matches, contrary to the initial case. For now, our focus is on how the hyperparameters affect the test set as a whole, so we will not examine the SVD in detail at this time.

4.3.2 Scale Constant

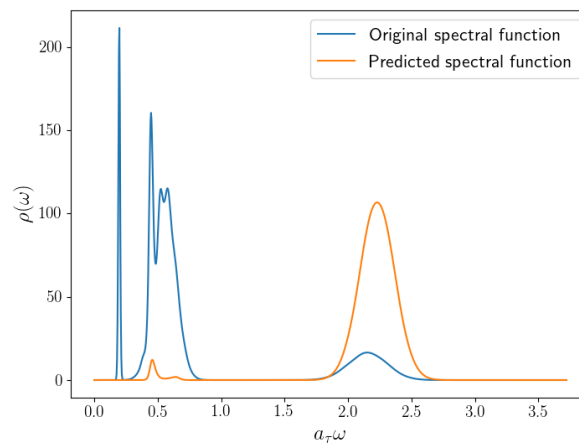
Next we will investigate the effect of varying γ and how the MSEs change. Before this though, we will provide further explanation of γ . It was previously referred to



(a)



(b)



(c)

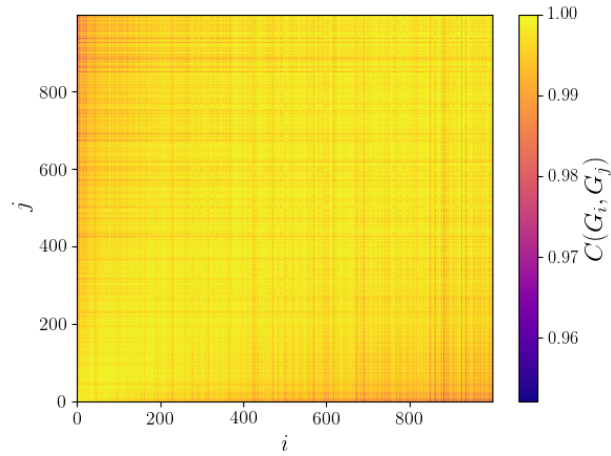
Figure 4.7: Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (10^{-10}, 1)$.

$(\lambda 1, \gamma)$	$(1, 1)$	$(10^{-10}, 1)$	$(1, 10^{1.5})$
$\text{MSE}_{\text{train}}$	4425.97	2295.77	3985.66
MSE_{test}	4547.88	2762.75	4268.37
MSE_{low}	15.63	200.05	333.47
MSE_{med}	2847.49	919.19	4141.96
MSE_{high}	40179.18	389.84	40239.93

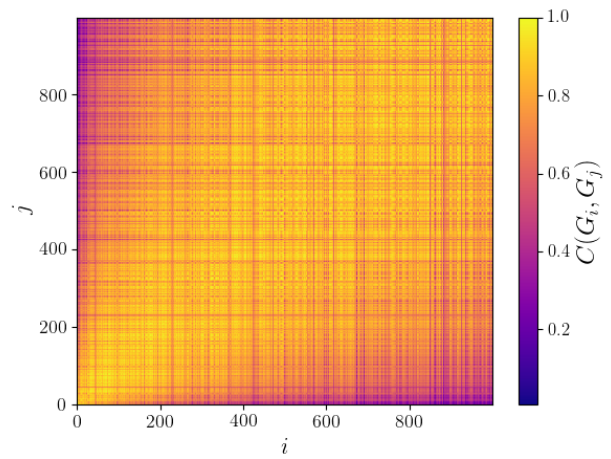
Table 4.4: MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed.

as the scale length. In Fig. 4.8 three examples of the kernel from Eq. (4.1.14) have been displayed as surface plots. 1000 correlators from the training set were used and the correlators for these plots were sorted by their second timeslice for clarity. In Fig. 4.8a the entire kernel ~ 1 , meaning that predictions will be influenced by the entire training set. In Fig. 4.8c the entire kernel ~ 0 with the exception of the diagonal and near off-diagonal elements. This means that predictions will only be influenced by training correlators that are equal or extremely similar to the new data. Fig. 4.8b represents a kernel between these two extremes where predictions will be influenced by the entire training set but to varying degrees.

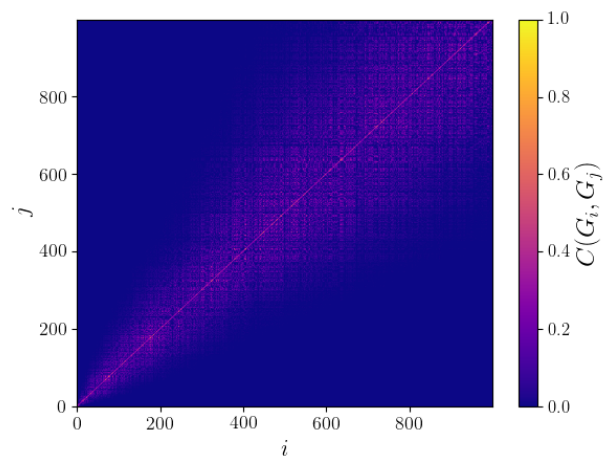
We return to examining the effect of varying γ . This was done using a range of $-5 \leq \log_{10}(\gamma) \leq 5$ for $\lambda = 10^{-10}$ and $\lambda = 1$. By using two values for the regularization constant, it is possible to confirm whether the hyperparameters can be optimized independent of one another when minimizing the cost function. To begin we look at the case where $\lambda = 1$, the training and testing MSEs of which are shown in Fig. 4.9. Unlike the regularization constant where both MSEs decreased with λ , there is a minimum. Although difficult to distinguish, the minimum corresponds to $\gamma = 10^{1.5}$. We know this corresponds to a kernel like Fig. 4.8c which could be due to the relatively large value of λ . Both the training and test MSE display similar behaviour with varying γ , as as the case when varying λ . The usual MSEs are displayed in Table 4.4, where we see an improvement in the MSEs for the entire training and testing sets but the three examples have worsened. Again these examples are plotted in Fig. 4.10. There is little difference to Fig. 4.5 apart from the appearance on a small peak in the correct position in Fig. 4.10a and 4.10c.



(a)



(b)



(c)

Figure 4.8: Examples of an ordered kernel calculated using Eq.(4.1.11) for $\gamma = 10^{-2}$ (a), 1 (b), and 10^2 (c).

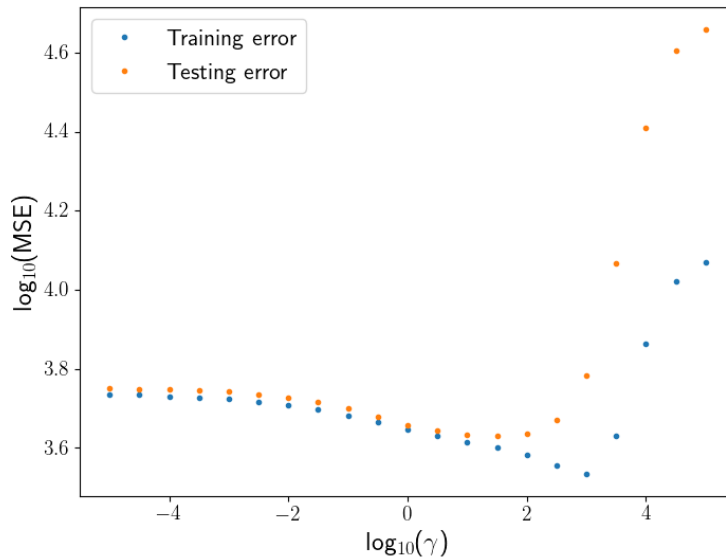
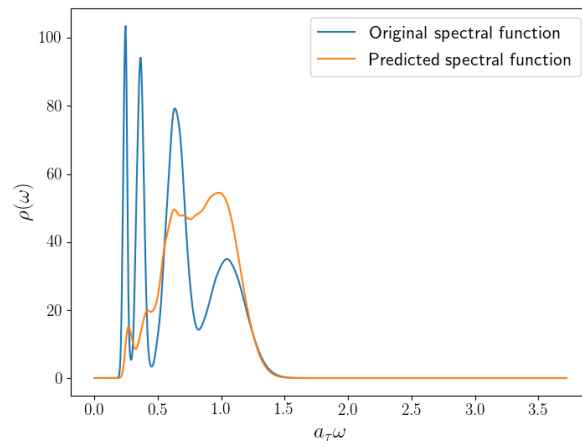


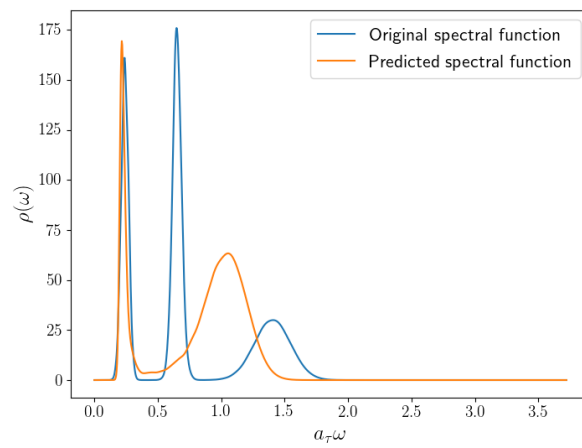
Figure 4.9: Training and testing MSEs as a function γ for $\lambda = 1$. Note the values on both axes are logarithmic.

For the case where $\lambda = 10^{-10}$, there is a clear difference in behaviour between the training and testing MSEs. For $\gamma \geq 10$, the errors begin to diverge from one another and whilst this did occur for the regularization constant, both errors in that case continued to decrease. This is the first instance in which the testing MSE suffers as the training MSE improves. To understand this behaviour it is useful to consider the effect of varying γ on the kernel functions from a qualitative perspective. Consider the examples in Fig. 4.8, in particular 4.8c. With high γ , the only contributions to the predictions will correspond to $\sum_{\tau} (G_i(\tau) - G_j(\tau))^2 \rightarrow 0$, where $i = 1, \dots, N_{\text{train}}$ and $j = 1, \dots, N_{\text{train}}$ or $j = 1, \dots, N_{\text{test}}$. Hence the training MSE $\rightarrow 0$, as all the diagonal elements $\mathbf{C} = 1$. Predictions of the test set suffer as a consequence as there should not be any shared functions between the training and test set.

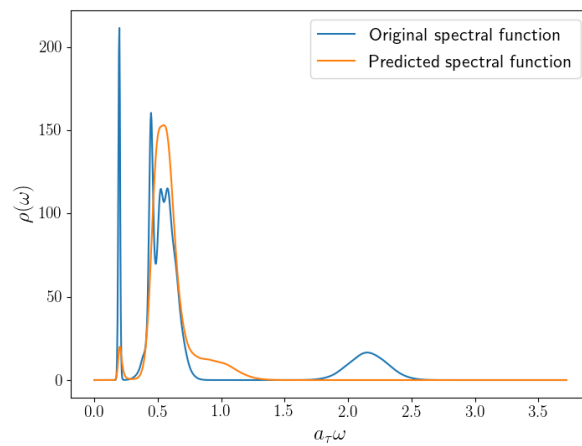
The minimum in this case occurs at $\gamma = 10^{-0.5}$. As before, the usual MSEs are displayed in Table 4.5 and the corresponding predicted spectral functions for the usual examples are plotted in Fig. 4.12. Rather than discussing the changes, we focus on the main observation: (λ, γ) cannot be optimized individually.



(a)



(b)



(c)

Figure 4.10: Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (1, 10^{1.5})$.

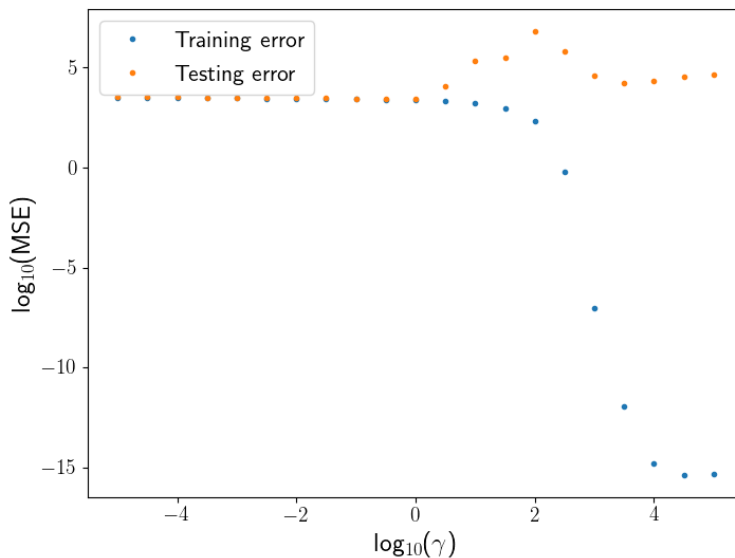
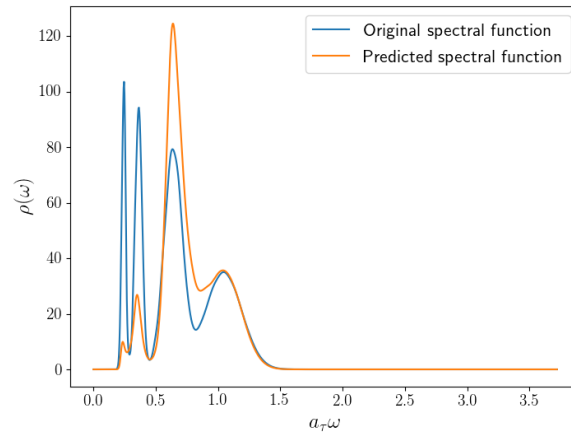


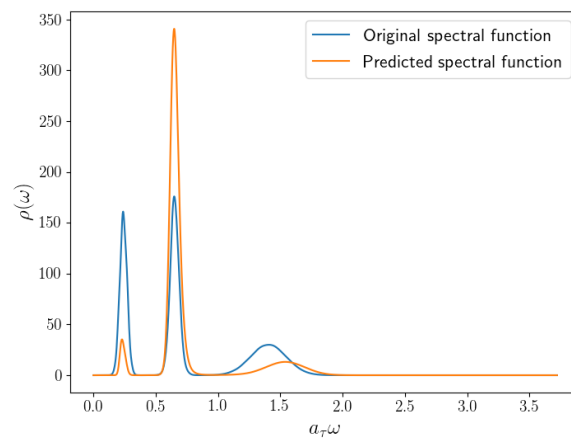
Figure 4.11: Training and testing MSEs as a function γ for $\lambda = 10^{-10}$. Note the values on both axes are logarithmic.

(λ, γ)	$(1, 1)$	$(10^{-10}, 1)$	$(1, 10^{1.5})$	$(10^{-10}, 10^{-0.5})$
$\text{MSE}_{\text{train}}$	4425.97	2295.77	3985.66	2446.35
MSE_{test}	4547.88	2762.75	4268.37	2757.06
MSE_{low}	15.63	200.05	333.47	212.59
MSE_{med}	2847.49	919.19	4141.96	1520.80
MSE_{high}	40179.18	389.84	40239.93	514.11

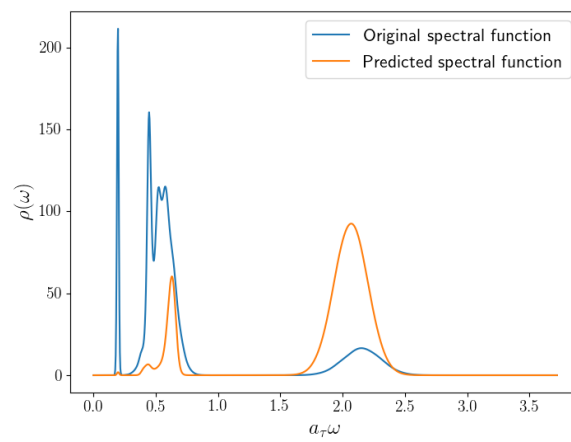
Table 4.5: MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed.



(a)



(b)



(c)

Figure 4.12: Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (10^{-10}, 10^{1.5})$.

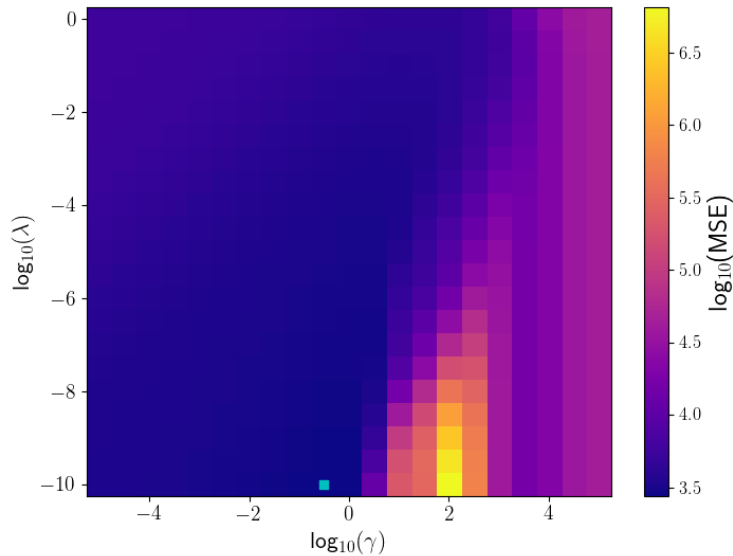


Figure 4.13: Surface plot of MSE_{test} for hyperparameters grid search. The green square marks the optimal hyperparameters.

4.3.3 Grid Search

In order to determine an optimal pair of hyperparameters, it is necessary to find a global minimum for the cost function Eq. (4.3.1). This could be achieved using a gradient descent method. However this would required including $\hat{\alpha}$ as an additional variable to be minimized. Since we know that $\hat{\alpha}$ can be determined for a particular (λ, γ) pair using Eq. (4.3.2), it was decided that a grid search method would be used instead. This method is exactly as it sounds, a grid of hyperparameters pairs is generated. In this case the same ranges from the previous test for λ and γ was used. From each pair of hyperparameters, the cost function is determined and then used to calculate $\hat{\alpha}$. A surface plot is then generated using all these values for the cost function, as seen in Fig. 4.13.

Ideally there would be a clear minimum which is not the case here. Unfortunately, the value of λ cannot be decreased much further without the calculation of $\hat{\alpha}$ suffering due to the inversion involved. Furthermore, it is unlikely we would want to decrease it further since it is questionable already whether $\lambda = 10^{-10}$ is large enough to prevent overfitting should it be necessary in future. By coincidence, the optimal pair appears to be $\lambda = 10^{-10}$ and $\sigma = 10^{-0.5}$. Hence the errors and examples plots can be seen in Table 4.5 and Fig. 4.10 respectively.

4.4 First prediction of NRQCD data

Whilst testing is certainly important, it is ultimately redundant if the model cannot make good predictions when faced with the actual NRQCD data. For that reason it is appropriate to observe the predictive ability of the current model using an unknown dataset. For this we use the Υ as this was the channel used to set E_0 . From effective mass calculations, we expect a peak around 9.460 GeV. From the ratio $G(\tau; T)/G(\tau; T_0)$ we don't expect to see significant variation as N_τ changes.

To reiterate, the setup of the model is by no means the optimal choice. 80 coefficients were used solely as an example as these are sufficient to reproduce the mock spectral functions in most cases. 128 coefficients could have been used instead, but this approaches $N_\theta \gg N_\tau$ for the smaller values of N_τ . This is not currently the focus though, as the point of this chapter was simply to introduce algorithm, data generation and one method of selecting the hyperparameters.

This provides an opportunity to test whether the hyperparameters have any dependence on N_τ , or thermal dependence. For each N_τ the setup previously described was used, $G(\tau)$ was calculated using Eq. (4.2.14) and the training kernel is determined using Eq. (4.1.14), ignoring the first 4 timeslices. The hyperparameters are then identified using the grid search method and used to define the optimal $\hat{\alpha}$. The parameters were calculated using,

$$\theta_\Upsilon = \mathbf{C}_\Upsilon^T \hat{\alpha} \quad (4.4.1)$$

where \mathbf{C}_Υ is a new kernel and i -th element of \mathbf{C}_Υ is defined as,

$$C_i^\Upsilon = \exp \left(-\gamma \sum_\tau (G_i(\tau) - G_\Upsilon(\tau))^2 \right). \quad (4.4.2)$$

The correlator in this case is the central value of the appropriate configurations. We found that for $N_\tau = 20, 24, 28,$ and 32 , the optimal scale length was $\gamma = 1$, whilst for $N_\tau = 36, 40, 48, 56, 64,$ and 128 , the optimal scale length was $\gamma = 10^{-0.5}$. In both case the optimal regularization constant was $\lambda = 10^{-10}$. Though the optimal value of γ appears to have some thermal dependence, it may be less significant than it currently appears. Increasing the resolution of γ would be the only way to check how much it actually varies with temperature.

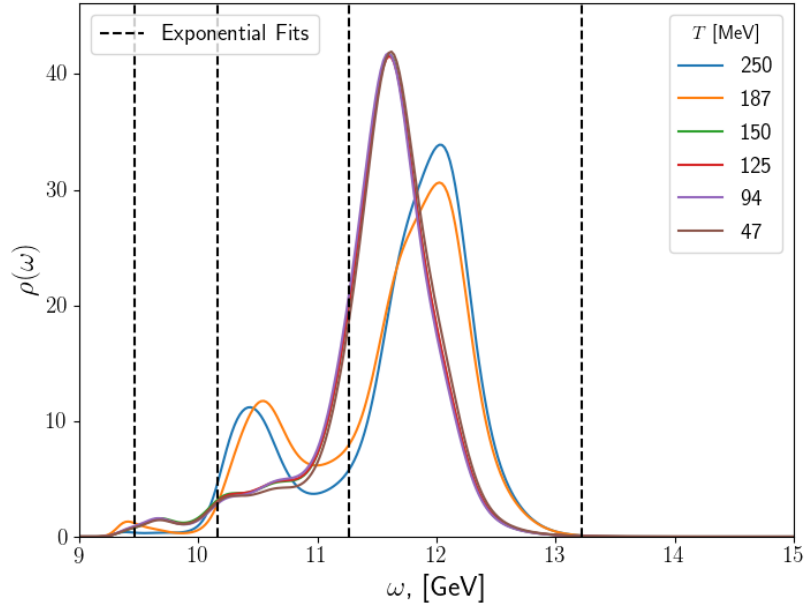


Figure 4.14: First attempt at reconstructed spectral function for Υ using KRR for various temperatures. Black dashed lines represent the masses shown in Table 3.2 that were determined using the CORFITTER package with the Υ correlator for $N_\tau = 128$, which corresponds to $T = 47$ MeV.

The first prediction for the Υ using the KRR model for a selection of temperatures can be seen in Fig. 4.14. It is immediately obvious that prediction is not what we would expect. The only temperature for which a peak coincides with the ground-state mass is $T = 187$ MeV. The dominant peak occurs at higher energies, which is true for all the temperatures displayed. This may be due to the L-2 distance in Eq. (4.1.14) which dominated by the earlier timeslices. At early timeslices the correlator is affected by any excited states, whilst the groundstate influence grows with n_τ . For $N_\tau = 128$ it reaches a point where the groundstate is the sole contribution to the correlator. It would be generous to claim that the higher order peak are close to the expected values, especially $T = 187$ MeV and 250 MeV. Rather than this prediction for the spectral function being correct, it is more probable that the KRR needs improving. This can be done by examining the training data to determine if it is appropriate or requires improvement, or altering how the input and target data is fed into the KRR model.

4.5 Alternative methods

We conclude this chapter with a discussion of alternative methods that have been implemented for either mock data or data from lattice simulations. The purpose of this is to demonstrate that due to spectral reconstruction being an ill-posed problem, it is only when multiple methods agree that we should trust the predictions.

This discussion of other spectral reconstruction methods is by no means a complete list. Work has been done to reconstruct spectral functions using other machine learning methods such as Neural Networks [74,75] and Gaussian processes [76]. The purpose of introducing these methods was to demonstrate the range of alternatives that can be used for comparison, as was done in recent work by the FASTSUM collaboration [77].

4.5.1 Ansatz

This method is perhaps the simplest of the alternative to be discussed. An ansatz that can be parameterized is proposed for ρ . The number of fit parameters should be less than the number of data points which is determined by N_τ . The ansatz can be as simple as a single Gaussian and delta function [78], but many studies use an ansatz consisting of a single bound state peak, a transport peak, and a continuum to represent perturbative contributions [78–81]. A corresponding form for the correlator can be determined by using Eq. (3.2.4). The parameters used to define ρ are then fitted using a minimization of χ^2 , maximum likelihood, or an alternative method of parameter fitting.

The issue with this method is that any proposed ansatz introduces bias that is difficult to overcome. Care needs to be taken to choose a form of ρ that will exhibit temperature dependence, should we expect ρ to behave differently at high and low temperatures, as we might for the P-waves states.

4.5.2 Backus-Gilbert

Another alternative is a method proposed in 1968 by Backus and Gilbert [82]. For work studies of spectral reconstruction using Backus-Gilbert, see [83–86] An estimate

solution to Eq. (3.2.4), $\hat{\rho}$, is generated from the true spectral function via,

$$\hat{\rho}(\omega_0) = \int_{\omega_{\min}}^{\omega_{\max}} d\omega A(\omega, \omega_0) \rho(\omega), \quad (4.5.1)$$

where $A(\omega, \omega_0)$ is the set of resolution functions that would ideally approximate to $\delta(\omega - \omega_0)$. These resolution functions are given by a linear combination of the thermal kernel, $K(\tau, \omega)$,¹

$$A(\omega, \omega_0) = \sum_{\tau} b_{\tau}(\omega_0) K(\tau, \omega) \quad (4.5.2)$$

where $b_{\tau}(\omega_0)$ are a set of coefficients to be determined, the Backus-Gilbert coefficients. Substituting Eq. (4.5.2) into Eq. (4.5.1) gives,

$$\hat{\rho}(\omega_0) = \sum_{\tau} b_{\tau}(\omega_0) G(\tau). \quad (4.5.3)$$

The Backus-Gilbert coefficients are then determined by minimizing

$$\begin{aligned} J(\omega_0) &= \int_{\omega_{\min}}^{\omega_{\max}} d\omega [A(\omega, \omega_0) - \delta(\omega - \omega_0)]^2. \\ &= \int_{\omega_{\min}}^{\omega_{\max}} d\omega \left[\sum_{\tau} b_{\tau}(\omega_0) K(\tau, \omega) - \delta(\omega - \omega_0) \right]^2, \end{aligned} \quad (4.5.4)$$

which corresponds to solving,

$$\mathcal{K}_{\tau, \tau'} \cdot b_{\tau'}(\omega_0) = K(\tau, \omega_0), \quad (4.5.5)$$

where

$$\mathcal{K}_{\tau, \tau'} = \int_{\omega_{\min}}^{\omega_{\max}} d\omega K(\tau, \omega) K(\tau', \omega), \quad (4.5.6)$$

is the kernel width matrix. The issue with this is that \mathcal{K} is close to singular. In Sec. 4.1.2 regularisation was introduced to ensure inversion required for calculating $\hat{\mathbf{w}}$ is possible. In order to invert Eq. (4.5.6) we make use of regularisation again and the covariance matrix Σ of $G(\tau)$. The kernel width matrix is transformed as such,

$$\mathcal{K} \rightarrow \mathcal{K}'(\lambda) = \lambda \mathcal{K} + (1 - \lambda) \Sigma, \quad (4.5.7)$$

¹Note the form of the kernel depends on the application and is not needed here. For NRQCD it is given by $K(\tau, \omega) = \exp(-\omega\tau)$.

where λ again is the regularisation constant, $\lambda \in (0, 1]$. As a consequence of this transformation, there is a loss of resolving power. As the regularisation becomes greater so does this loss. This is one of the main issues with the Backus-Gilbert method, as poor resolving power makes it difficult to reconstruct the narrow resonance we would expect for the groundstate. Another issue with the method is that the number of coefficients is temperature dependent; at higher temperatures it becomes harder to estimate ρ as there are fewer resolution functions.

4.5.3 Maximum Entropy Method

The next method to be discussed is the popular Maximum Entropy Method (MEM) [72, 87]. For studies involving MEM see, [88–91]. Before we go into the details of MEM, we note that we can rescale the thermal kernel and spectral function,

$$K(\tau, \omega) \rightarrow \omega K(\tau, \omega), \quad \rho(\omega) \rightarrow \rho(\omega)/\omega, \quad (4.5.8)$$

to stabilise the inversion [92]. This leaves the correlators given by Eq. (2.2.24) or Eq. (3.2.4) unchanged. The spectral function still needs parameterizing, but typically this is done by,

$$\frac{\rho(\omega)}{\omega} = \frac{m(\omega)}{\omega} \sum_{i=1}^{N_{\text{coeffs}}} b_i f_i(\omega) \quad (4.5.9)$$

where b_i are the coefficients to be determined. The functions $m(\omega)$ and $f(\omega)$ are the default model and basis functions from Bryan’s method described in Sec. 4.2.2. For simplicity we will set $m(\omega) = 1$. Unlike KRR where the number of basis functions is arbitrary, MEM is restricted by N_τ , as it is dependent on $K(\tau, \omega)$. This leads to poorer resolution as temperature increases.

The coefficients, b_i , are determined using Bayes’ theorem,

$$P(\rho|DH) = \frac{P(D|\rho H)P(\rho|H)}{P(D|H)}, \quad (4.5.10)$$

where D is the input data, H represents any prior knowledge of the system, and $P(A|B)$ is the conditional probability of A given B . Let $D = G(\tau)$, where $G(\tau)$ correlator defined as the central value of the ensemble. In the above, $P(D|\rho H)$ is the likelihood function, $P(\rho|H)$ is the prior probability, and $P(D|H)$ is the normalisation. The likelihood function is typically given by $P(D|\rho H) = Z_L^{-1} \exp(-L)$,

where Z_L is a normalisation constant with,

$$L = \frac{1}{2} \sum_{j,k} (G(\tau_j) - G^{(b)}(\tau_j)) c_{jk}^{-1} (G(\tau_k) - G^{(b)}(\tau_k)), \quad (4.5.11)$$

where $G^{(b)}(\tau_j)$ is the correlator from Eq.(2.2.24) using Eq.(4.5.9) for the definition of ρ which is dictated by coefficients b_i , and c_{jk}^{-1} is the covariance matrix given by,

$$c_{jk} = \frac{1}{N_{\text{conf}}(N_{\text{conf}} - 1)} \sum_{m=0}^{N_{\text{conf}}} (G^{(m)}(\tau_j) - G(\tau_j)) (G^{(m)}(\tau_k) - G(\tau_k)), \quad (4.5.12)$$

where N_{conf} is the number of configurations for that temperature, see Table 1.1, and $G^{(m)}(\tau_j)$ is the m -th configuration of the ensemble. For MEM, the prior probability is given by $P(\rho|H) = Z_S^{-1} \exp(\alpha S)$, where Z_S^{-1} is a normalisation constant and S is the Shannon-Jaynes entropy,

$$S = \int_0^\infty d\omega \left[\rho(\omega) - m(\omega) - \rho(\omega) \log \left(\frac{\rho(\omega)}{m(\omega)} \right) \right]. \quad (4.5.13)$$

The coefficients, b_i , and therefore the MEM estimate for the spectral function, $\hat{\rho}$, are determined by maximizing $\alpha S - L$.

There are alternatives to MEM that also make use of Bayesian methods. The work of Ding et al. [81] uses stochastic analytical inference, which is proposed as a generalization of MEM, whilst in [38], a new set of axioms is postulated for the prior probability.

Chapter 5

Data generation revisited

In the previous chapter we tested the current KRR model with the FASTSUM Gen2L lattice correlators. The results were not what we had expected, suggesting that the KRR model needs improving or at least further examination. However, we propose at this time that further work is needed to ensure the training data involved is appropriate. There is the advantage of having access to the lattice correlators which can be compared with our mock data.

5.1 Examination of current data

We begin with an inspection of the current mock dataset compared with the correlators for each of the bottomonium states with the exception of the h_b . In order to do this for the χ_{b0} and χ_{b1} , for which $G(0) \approx 12$, A_r needs scaling appropriately rather than generating an entirely new dataset. This was naively done by letting $A_r^{(P)} = 2A_r^{(S)}$, where $A_r^{(P)}$ and $A_r^{(S)}$ are the amplitudes for the P-wave and S-wave mock data respectively.

Whilst there is little thermal modification for the S-wave states, this is not the case for the P-waves states as can be seen in Fig. 3.5. For this reason the mock and lattice correlators for $N_\tau = 32$ and 128 were used. In Fig. 5.1 the distributions for the mock correlators for $N_\tau = 32$ and 128 have been represented as heatmaps. Although the lattice correlators lie within the bounds of the mock correlators, they diverge from the majority which can be seen by the band of red in Fig. 5.1. The P-waves states appear to be separate from the majority for all timeslices, and even the S-waves states are only close to majority for the first several timeslices. Due to the size of the mock dataset, which is 20000, the density quickly drops below 0.1 away from this majority.

In order to see there are any correlators that match with the lattice correlators we look at the distribution for individual timeslices. Fig. 5.2 shows the distribution

of the mock correlators at specific timeslices, $n_\tau = 7, 31$, and 127, for both S-wave and P-wave states for $N_\tau = 128$. These plots further demonstrate how poorly the current mock data emulates the lattice data. There is noticeable agreement between the S-wave states and mock data up to at least $n_\tau = 31$. For $n_\tau = 127$ for the S-wave states and all timeslices for P-wave states there is little to no agreement between the mock and lattice data. Therefore we have confirmed that the mock data is inadequate. Reviewing the heatmaps for inspiration on how to resolve this, we observed that in all cases the majority of the mock correlators are much greater than the lattice correlators across the timeslices.

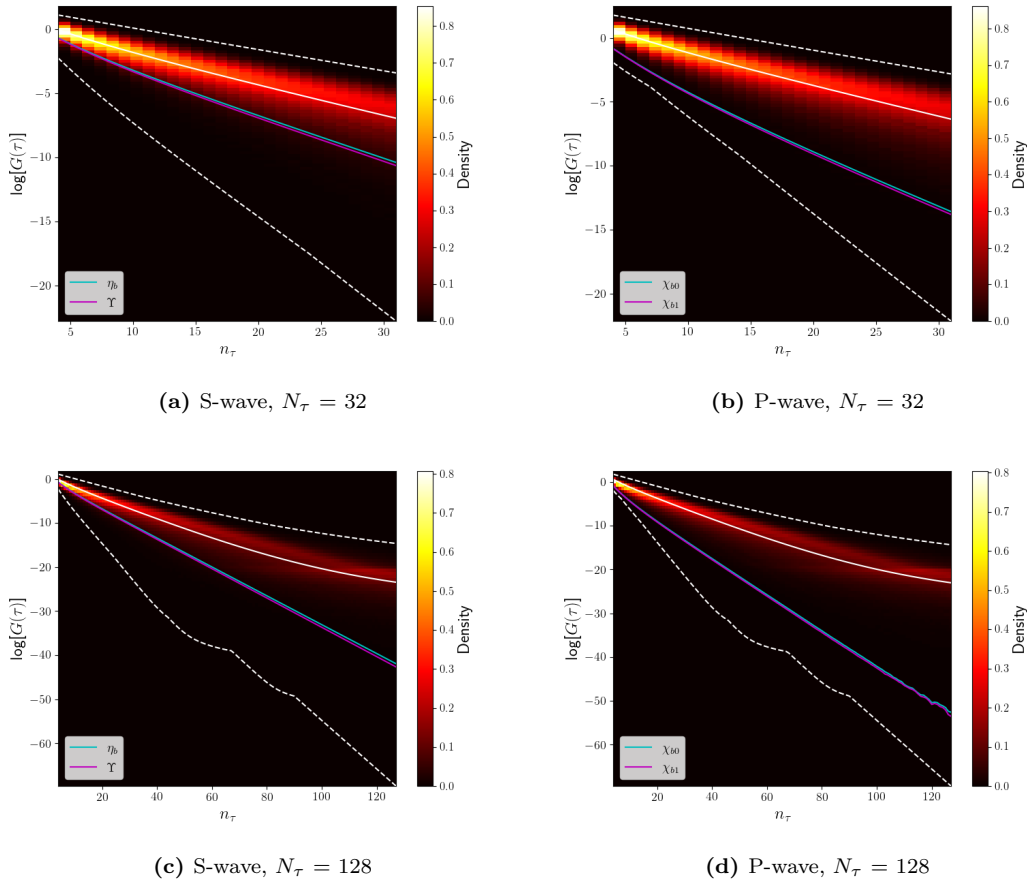


Figure 5.1: Distribution of the natural logarithm of correlators in the mock dataset at each timeslice for $N_\tau = 32$ and 128 for S-wave (a, c) and P-wave (b, d) states. Dashed lines represent the upper and lower limits of the correlators in the mock dataset for each value of n_τ , and white solid line represent mean of the mock correlators. The other solid lines represent the Euclidean correlators obtained in simulations. Note that the two channels in each figure – η_b, Υ (a, c) and χ_{b0}, χ_{b1} (b, d) – are hardly distinguishable.

However before we attempt to resolve this, there is another concern with the current data. See Fig. 4.4a and 4.5 to observe there is a significant range for which $\rho(\omega) = 0$ between the final resonance and ω_{\max} . The range of ω was thus changed using the improved lattice NRQCD dispersion relation [56],

$$a_\tau E(\mathbf{n}) = -2 \log \left(1 - \frac{1}{2} \frac{\hat{p}^2}{2\xi a_s M_b} \right) - \log \left(1 - \frac{\hat{p}^4}{24\xi a_s M_b} + \left(1 + \frac{a_s M_b}{2\xi} \right) \frac{(\hat{p}^2)^2}{8\xi (a_s M_b)^3} \right) \quad (5.1.1)$$

$$\hat{p}^2 = 4 \sum_{i=1}^3 \sin^2 \left(\frac{\pi n_i}{N_s} \right), \quad \hat{p}^4 = 16 \sum_{i=1}^3 \sin^4 \left(\frac{\pi n_i}{N_s} \right), \quad n_i = -\frac{N_s}{2} + 1, \dots, \frac{N_s}{2}, \quad (5.1.2)$$

along with Eq. (4.2.7). The new range was determined to $\omega_{\max} - \omega_{\min} = 1.4$.

Rather than selecting the subset of the original mock functions that obey the new constraint that ρ is non-zero only between $\omega_{\max} = 1.4005$ and $\omega_{\min} = 0.0005$, a new set of mock functions was generated. This is only to ensure that there is an equal number of functions across the two sets for later comparison. A quick check of Fig. 4.4a shows there are masses, and hence functions, from the original set that violate the constraint $m_j < \omega_{\max}$ for $\omega_{\max} = 1.4005$. The method for generating these new functions is identical to that described in Chap 4.2. The distributions for the parameters were kept the same but the additive constant used for generating the masses was changed from 8.5 to 9.

The heatmaps for the updated mock dataset are displayed in Fig. 5.3 and the distribution of mock correlators at specific timeslices are displayed in Fig. 5.4. Even with this change of range for ω , there does not appear to be any significant changes from the original dataset. Certainly there are some noticeable differences, in particular Fig. 5.4e and 5.4f. There is better agreement between the later timeslices for both S and P-wave states, whilst agreement for the early timeslices has worsened.

In principle there is a possible way of making reasonable predictions with the KRR model. If we could ensure that these particular mock correlators, and their corresponding parameters, are included in the training data, then it would be a simple matter of using a large scale length $\gamma > 1$.

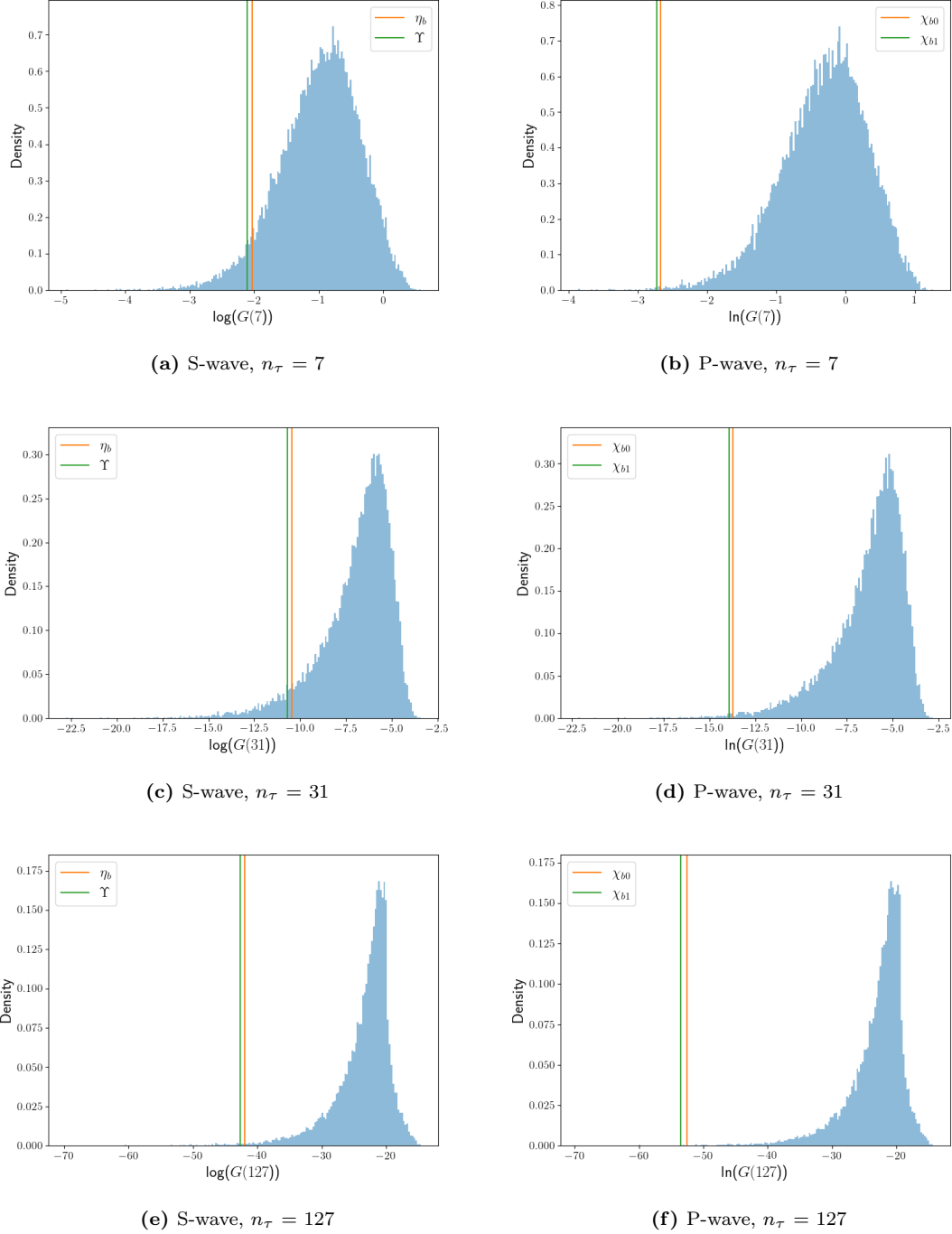


Figure 5.2: Distribution of the natural logarithm of correlators in the mock dataset at $n_\tau = 7, 31$, and 127 for S-wave (a, c, e) and P-wave (b, d, f) states. Solid lines represent the Euclidean correlators obtained in simulations.

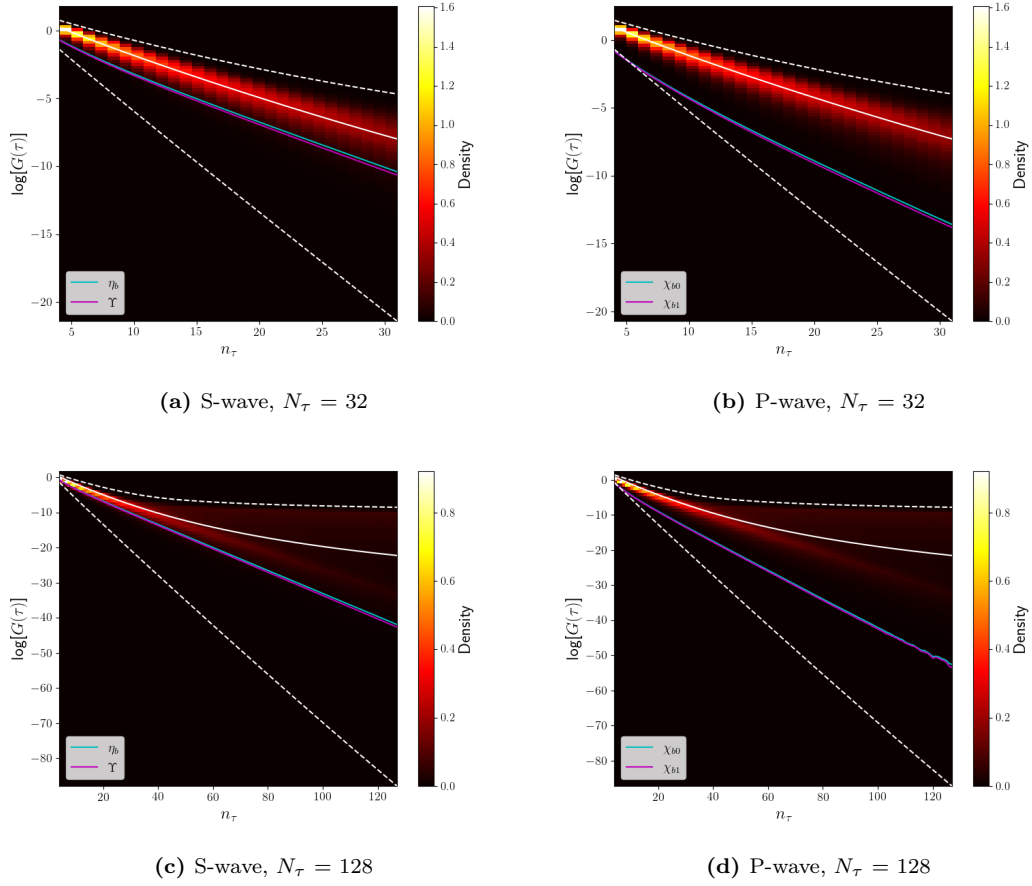


Figure 5.3: Distribution of the natural logarithm of correlators in the updated ($\omega_{\max} = 1.4005$) mock dataset at each timeslice for $N_\tau = 32$ and 128 for S-wave (a, c) and P-wave (b, d) states. Solid and dashed lines represent the same as Fig. 5.1.

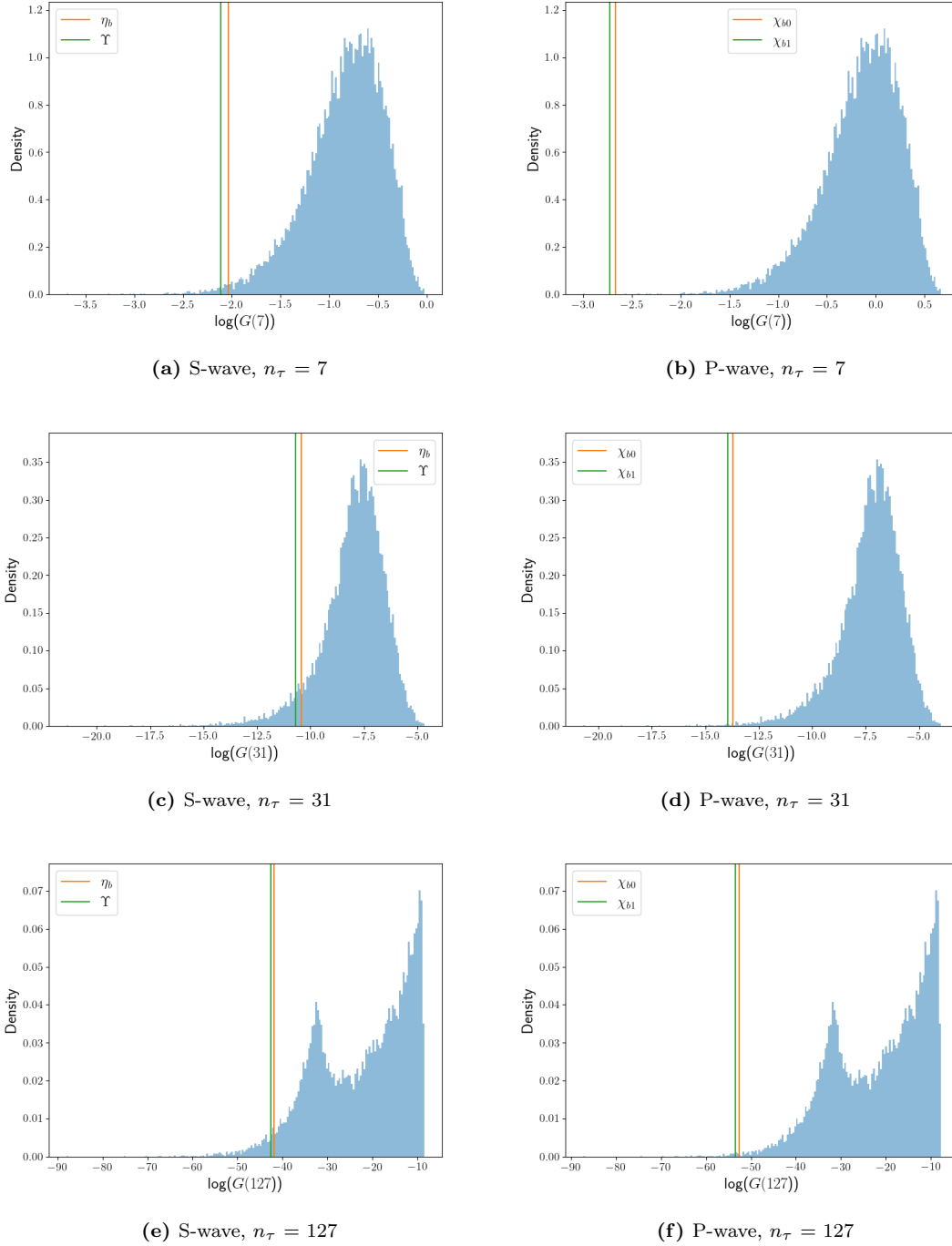


Figure 5.4: Distribution of the natural logarithm of correlators in the updated ($\omega_{\max} = 1.4005$) mock dataset at $n_\tau = 7, 31$, and 127 for S-wave (a, c, e) and P-wave (b, d, f) states. Solid lines represent the same as Fig. 5.2

5.2 Data generation with NRQCD influence

With that in mind, we set about developing a method for producing a better dataset.

To begin, we selected mock correlators according to

$$i = \operatorname{argmin} \sum_{n_\tau=4}^{N_\tau-1} \left| \log \left[\frac{G_i(n_\tau)}{G_{\text{NRQCD}}(n_\tau)} \right] \right|. \quad (5.2.1)$$

In the above, the ratio of the i -th mock correlator and the NRQCD correlator at each timeslice from $n_t = 4$ to $n_t = N_\tau - 1$ is determined. The natural logarithm of these ratios is taken and summed. After this has been calculated for each of the mock correlators have been calculated, we select the index of the mock correlator corresponding to the minimum sum. The correlator itself is unimportant. Instead, it is the parameters, (m_r, Γ_r, A_r) , that generated the corresponding spectral function that are retained. This was done for all N_τ in Table 1.1 for the all states except h_b , which will be disregarded from now on. In total we select 20 sets of parameters for S and P-wave states using this method.

For each of these subsets of parameters, the mean and variance were calculated. A multivariate Gaussian distribution can then be generated using these means and variances. With this multivariate distribution, we follow the same method described in Sec. 4.2, only that the parameters are selected from this multivariate distribution rather than the distributions described previously. However, rather than using the means and variances for Γ and A directly, the following was used instead,¹

$$x_\Gamma = \log_{10}(\Gamma), \quad x_A = \log_{10}(A) \quad (5.2.2)$$

$$\sigma_x = \frac{1}{\ln(10)} \frac{\sigma_\Gamma}{\Gamma}, \quad \sigma_a = \frac{1}{\ln(10)} \frac{\sigma_A}{A}. \quad (5.2.3)$$

This is to prevent negative values for the widths and amplitudes being selected. In fact since this prevents negative values for the widths and amplitudes we will use these parameters instead in future predictions using the KRR model. 20000 spectral functions, each with four peaks, were generated along with the corresponding correlators. The exact form of ρ , Eq. (4.2.1) was used for calculating the correlators,

¹We write \ln rather than the usual \log to avoid confusion.

which were calculated using,

$$G(\tau) = \Delta\omega \sum_{\omega_{\min}}^{\omega_{\max}} K(\tau, \omega) \rho(\omega) = \Delta\omega \sum_{\omega_{\min}}^{\omega_{\max}} e^{-\tau\omega} \sum_{r=1}^4 A_r \exp\left(-\frac{(\omega - m_r)^2}{\Gamma_r}\right). \quad (5.2.4)$$

where $\Gamma_r = 10^{x_r}$ and $A_r = 10^{a_r}$. The heatmaps for this new dataset are displayed in Fig. 5.5 and the distributions for the usual timeslices are displayed in Fig. 5.6.

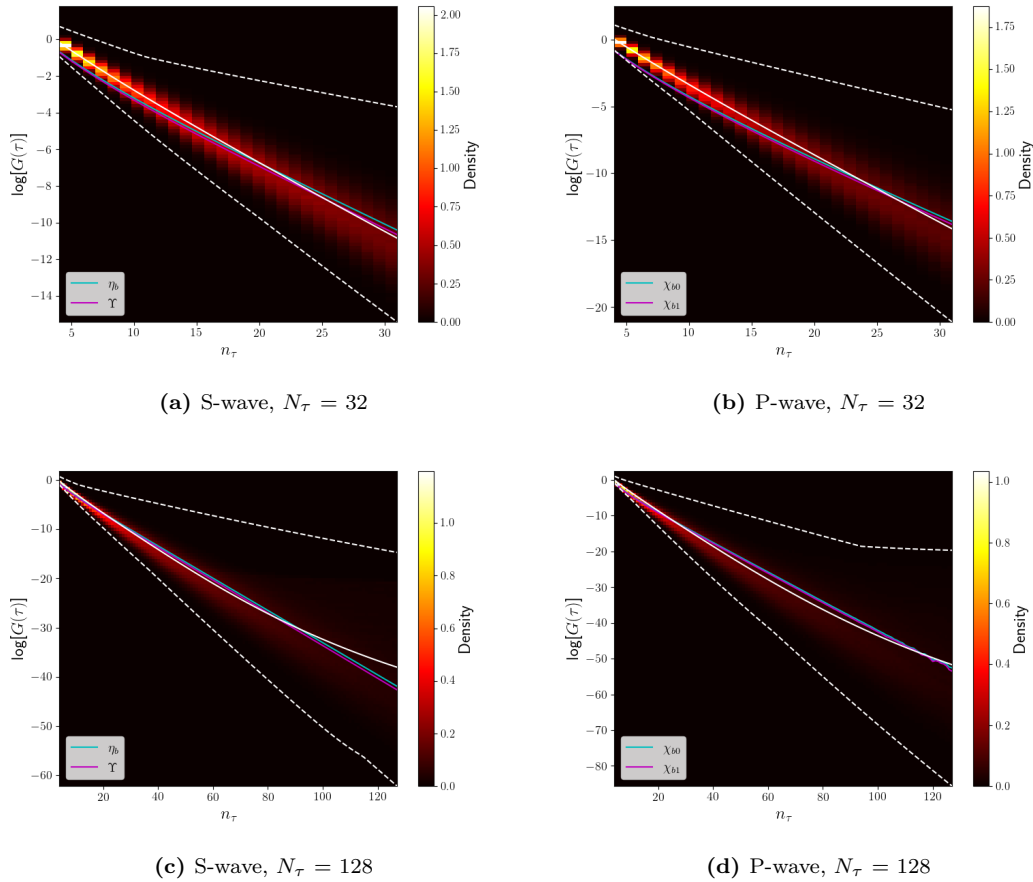


Figure 5.5: Same as Fig. 5.3 but for improved dataset.

Whilst there are still some flaws, this new dataset is certainly a better imitation of the lattice correlators. We can see that the majority follows a similar path to that of the lattice correlators, though there is a greater spread than the original dataset in Fig. 5.1. The similarity between the mock data and lattice correlators for early timeslices could be better but there are now a reasonable number of mock correlators that match with the P-wave correlators for $n_\tau = 7$, which did not occur in either of the previous cases. The spread of the data is not an issue as the KRR model requires correlators that bear little resemblance to the lattice correlators. Rather

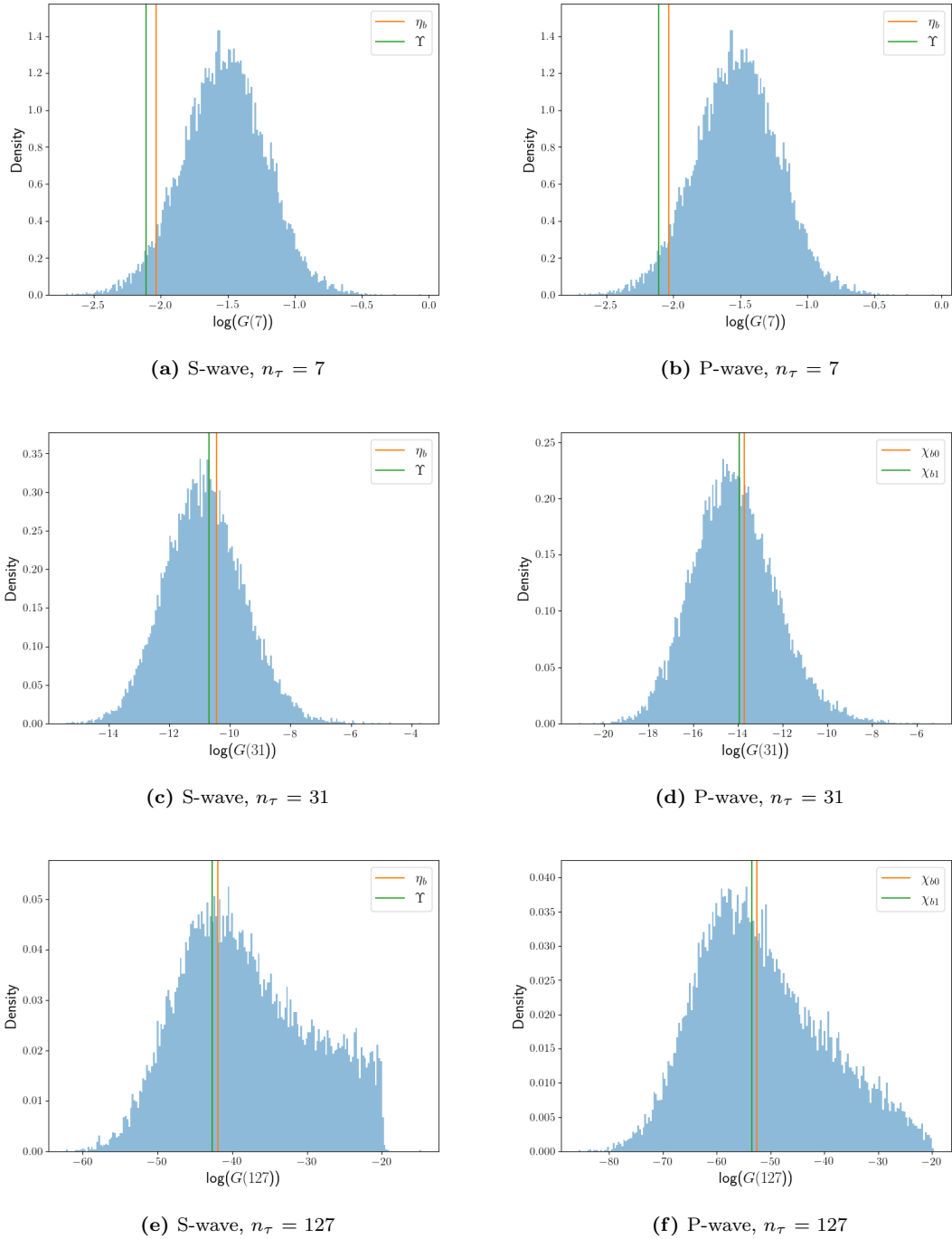


Figure 5.6: Same as Fig. 5.4 but for improved dataset.

than simply taking this new dataset to continue the investigation of KRR, there is an opportunity to examine alternative spectral functions.

5.3 Alternative datasets

We used the above method to generate new datasets for 2 alternative cases. The first of these is similar to the above but with a change to distributions for Γ and A and we introduce an additional peak. The second dataset makes use of the same starting dataset as the first case but uses the corresponding SVD coefficients in an attempt to make an improved dataset.

5.3.1 Alternative parameter distributions

As we stated above for the first alternative dataset we are considering, the distributions for Γ and A were changed. For Γ we still used a uniform distribution but with the range $-6 \leq x \leq 0$, where $\Gamma = 10^x$. For A , a uniform distribution with the range $0.1 \leq A \leq 100$ was used, then rescaled as described in Sec. 4.2.1. Until now, four Gaussian have been used due to the results of Chapter 3. However these results were dependent of the spectral function being a series of δ functions, not a series of Gaussian as we are currently using. By using these larger ranges for parameter and an additional peak, it is hoped that there will be a greater range of mock spectral functions. The method described above for generating an improved dataset was repeated and heatmaps for this new dataset are displayed in Fig. 5.7 and the distributions for the usual timeslices are displayed in Fig. 5.8.

Other than Fig. 5.6b and Fig. 5.8b, there does not appears to be any change. By comparing these two plots, we observe that the new dataset would appear to be somewhat worse as there are fewer matching correlators than in Fig. 5.6b. This could be due to the additional peak; recall from the results of Chapter 3 that $N_p = 4$ was chosen as the appropriate number of peaks.

5.3.2 SVD coefficients

The final dataset to be considered is one generated using the SVD representation. Before discussing this further though, it is worth examining the SVD representation

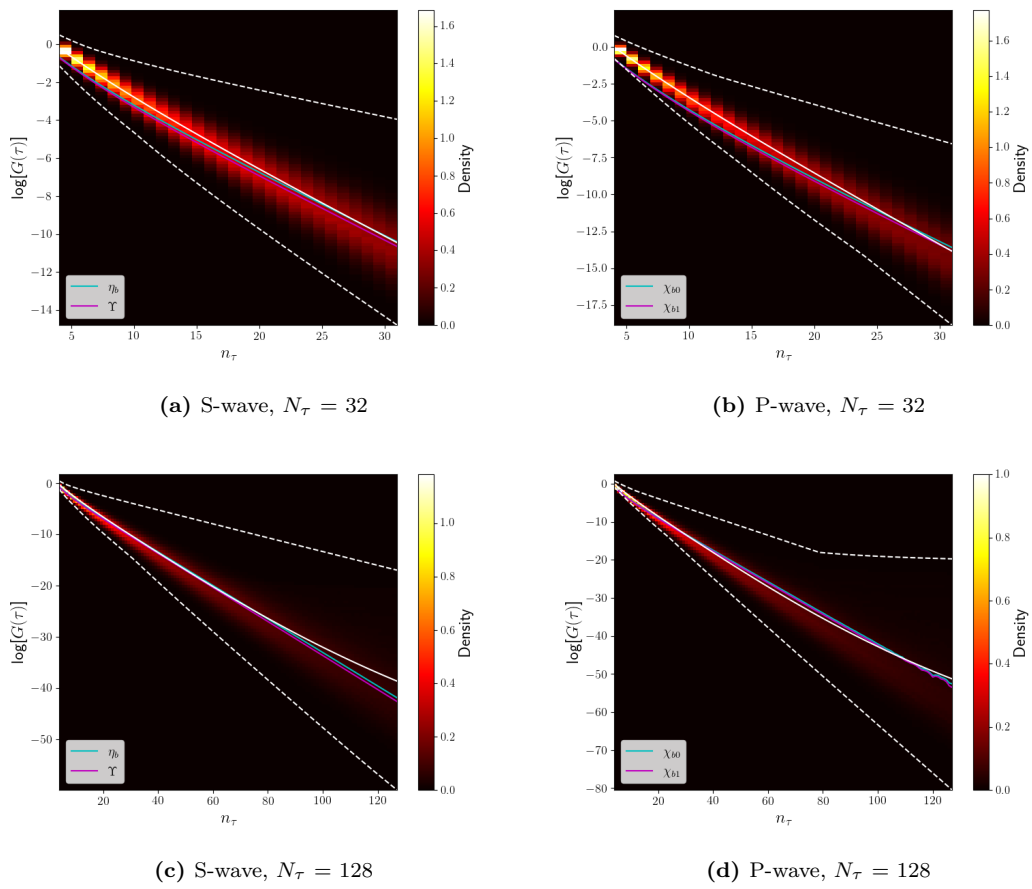


Figure 5.7: Same as Fig. 5.4 but for dataset generated from alternative parameter dataset.

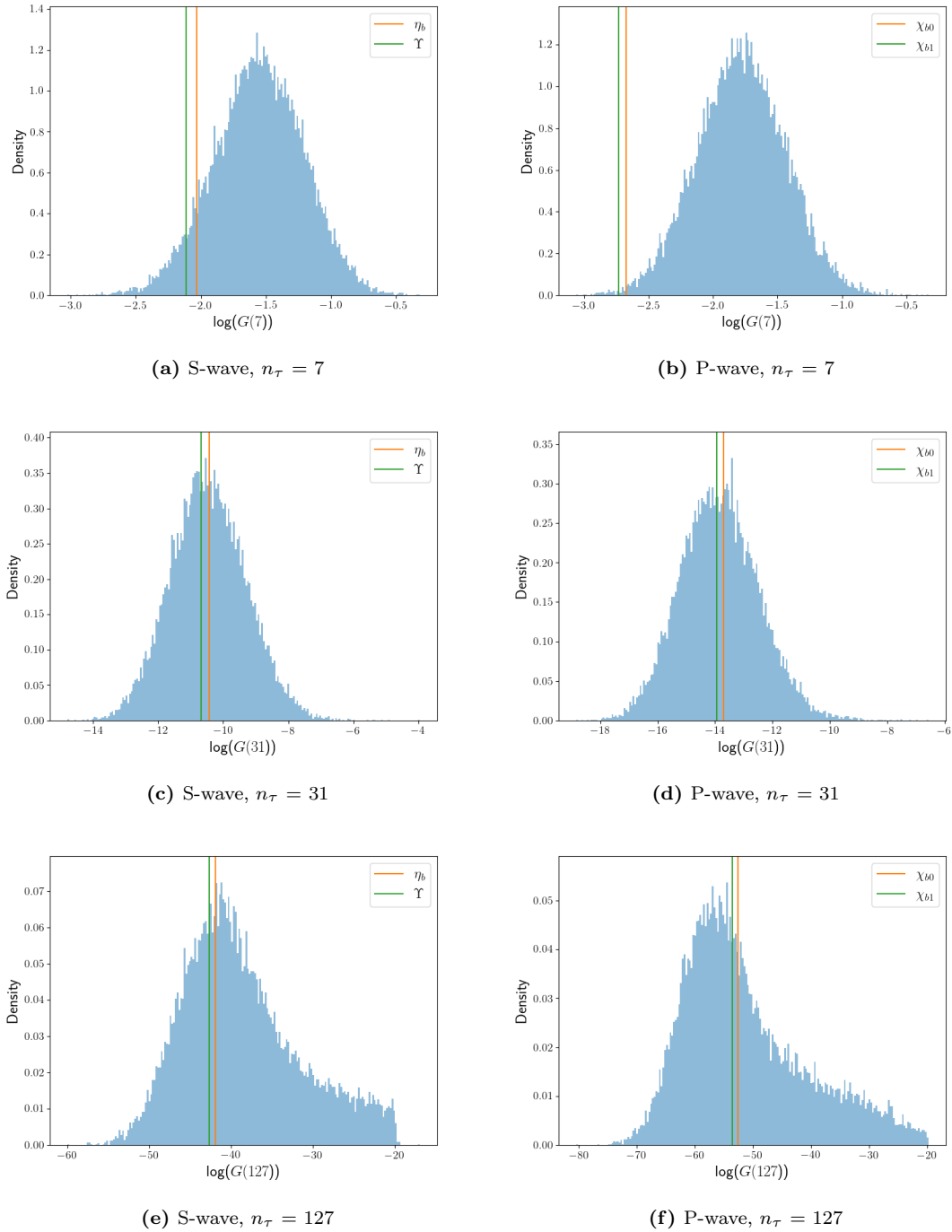


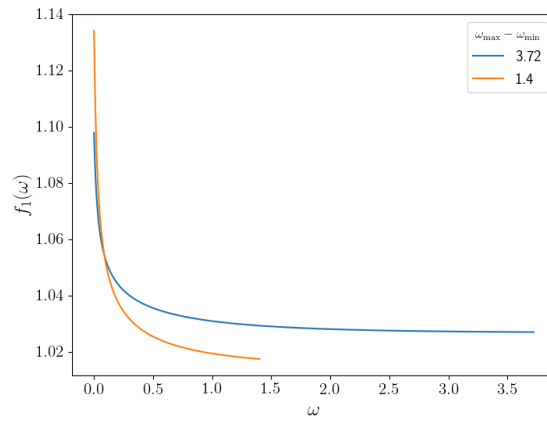
Figure 5.8: Same as Fig. 5.4 but for dataset generated from alternative parameter dataset.

in more detail than we have thus far. A selection of the basis functions, for both ranges of ω , are shown in Fig. 5.9. Reproducing narrow peaks in the spectral functions requires the later basis functions, though more basis functions become relevant with this shorter range for ω . Even so, it is unlikely a spectral function containing a Gaussian with width $\Gamma \sim \mathcal{O}(10^{-6})$ can be reconstructed using the SVD representation. Rather than trying to, we decided to generate a new dataset directly from the basis functions.

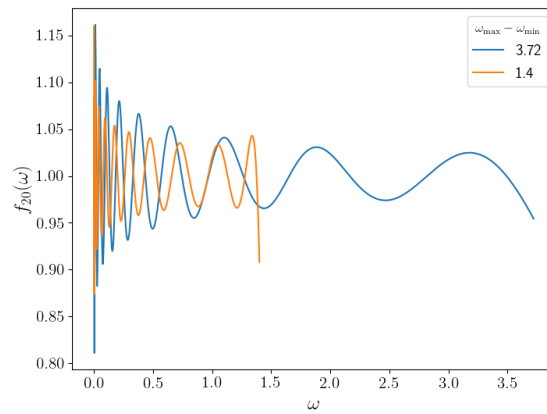
In Fig. 5.10 are box and whisker plots to show the mean and extent of the coefficients for the dataset used in Chapter 4 and the original dataset generated in Sec. 5.3.1. This helps explain the large errors we saw in Chapter 4 which we expect is mainly due to the first several coefficients. However since these correspond to simple functions such as the one seen in Fig. 5.9a, which is unfortunate as the current KRR model will prioritise reducing this error. In future we must determine a way in which all the coefficients are of approximately equal magnitude. For now though, we return our focus to generating a dataset using the SVD representation.

The first step of this was to determine coefficients from the spectral functions using Eq. (4.2.11), which has already been done to produce Fig. 5.10b. Corresponding correlators are calculated using Eq. (4.2.14). Using these correlators, a subset of the SVD coefficients excluding B are selected using the method described in Sec. 5.2 and Eq. (5.2.1). These coefficients are used to generate the multivariate distribution as opposed to the parameters (m_r, Γ_r, A_r) . 20000 sets of coefficients, θ , are selected from this distribution and the corresponding spectral functions are calculated using Eq. (4.2.10). During this step B is determined such that Eq. (4.2.4) holds. Any spectral function with $B > 5$ is rejected and replaced. This is to prevent spectral functions consisting of a single narrow peak.

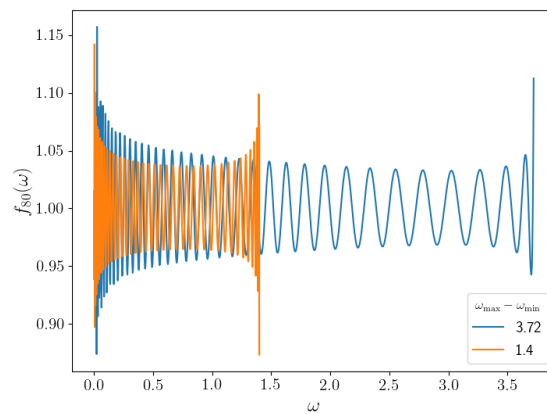
In previous datasets there is an assumption that the spectral functions for the lattice correlators would be a series of Gaussian functions. The reason for using the SVD representation to generate a dataset is to avoid this assumption. Admittedly the original spectral functions are generated from a series of Gaussian functions, so the assumption still has some influence on the final version of this dataset. The heatmaps for this dataset are displayed in Fig. 5.11 and the distributions for the usual timeslices are displayed in Fig. 5.12.



(a)

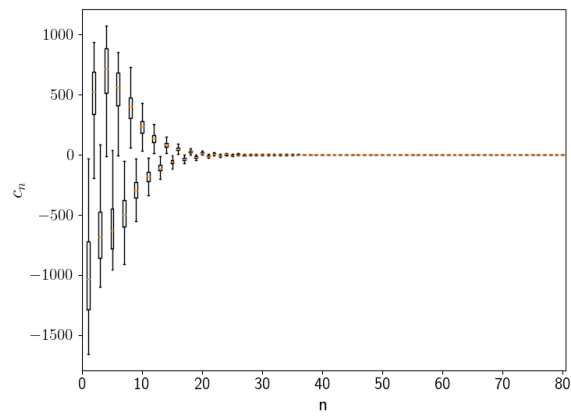


(b)

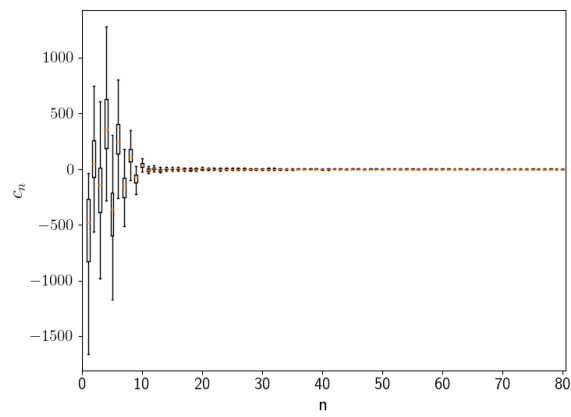


(c)

Figure 5.9: The first (a), twentieth (b), and eightieth (c) basis functions for the SVD representation for ω range $\omega_{\max} - \omega_{\min} = 3.72$ and $\omega_{\max} - \omega_{\min} = 1.4$.



(a)



(b)

Figure 5.10: Box and whisker plots for the coefficients of the original dataset with $\omega_{\max} - \omega_{\min} = 3.72$ (a) and the dataset described in Sec. 5.3.1 before improvement.

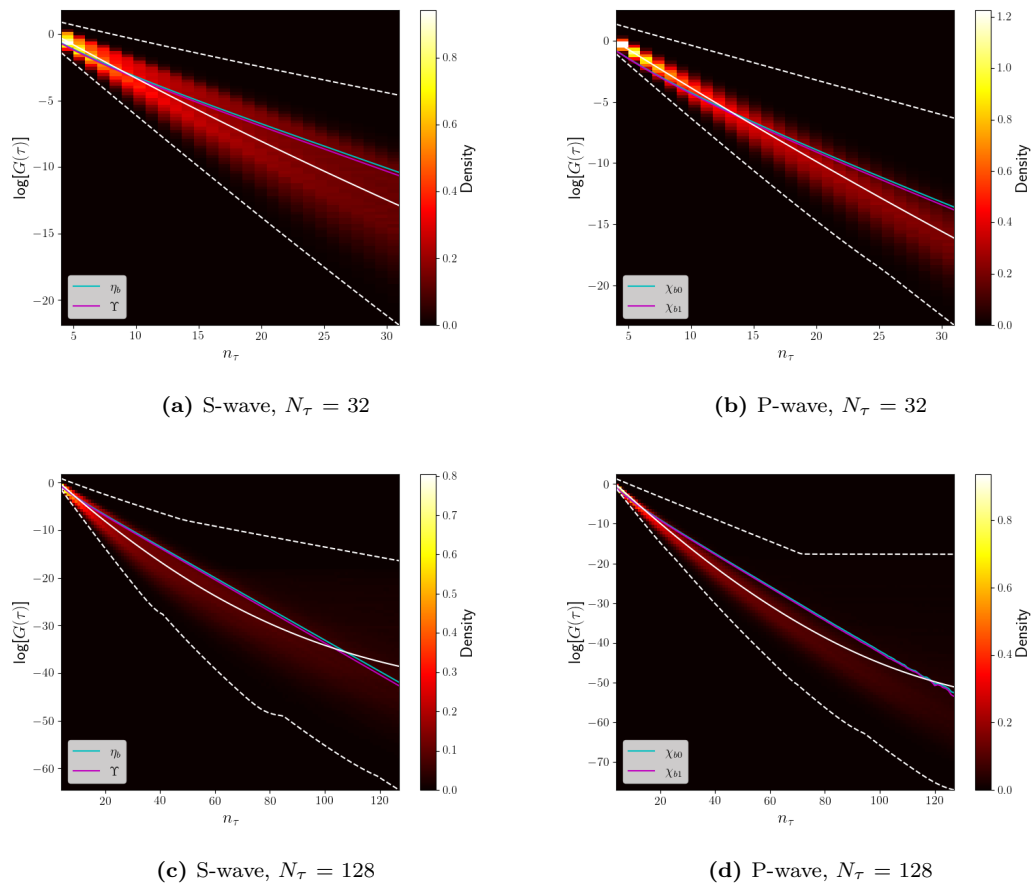


Figure 5.11: Same as Fig. 5.4 but for dataset generated from coefficients.

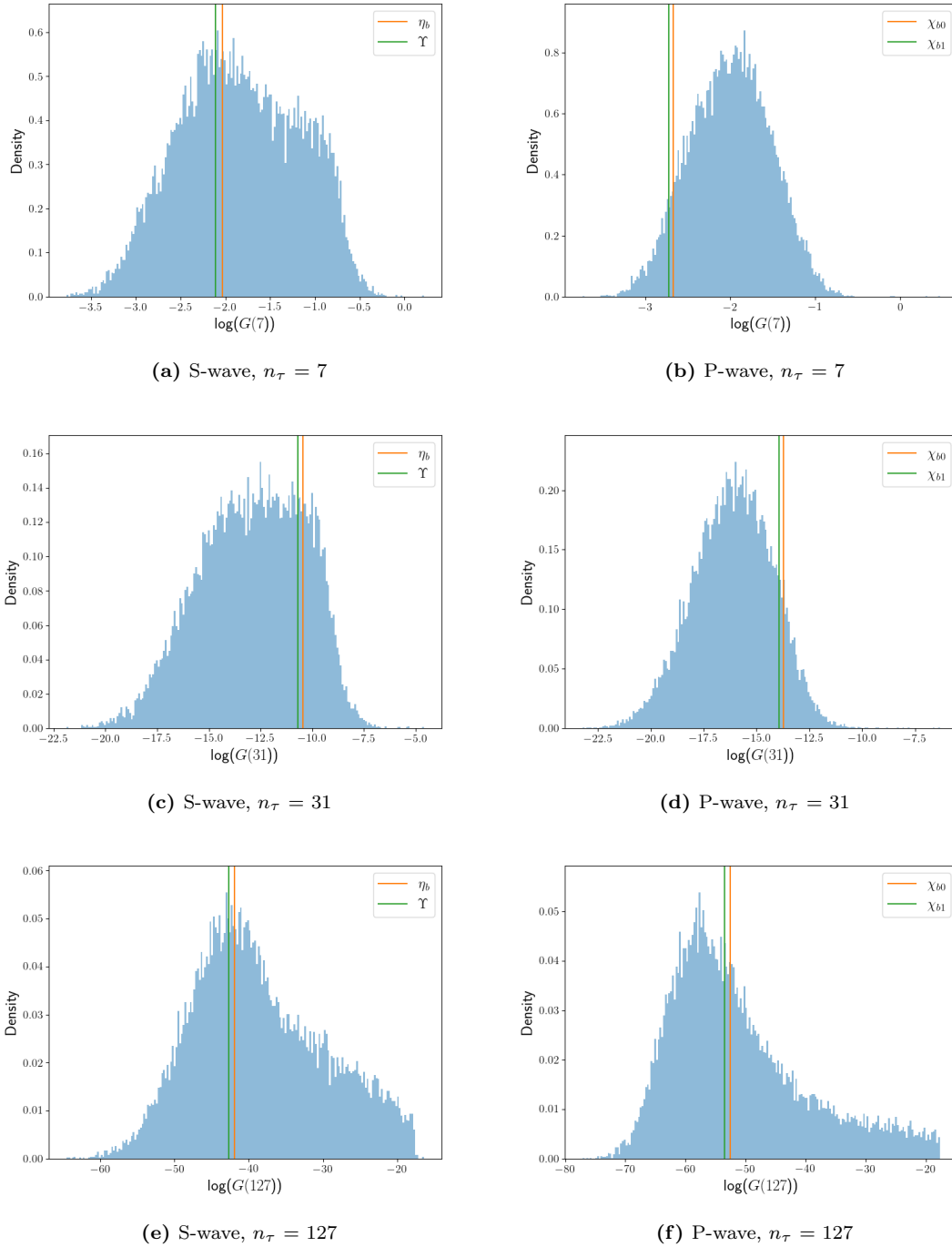


Figure 5.12: Same as Fig. 5.4 but for dataset generated from coefficients.

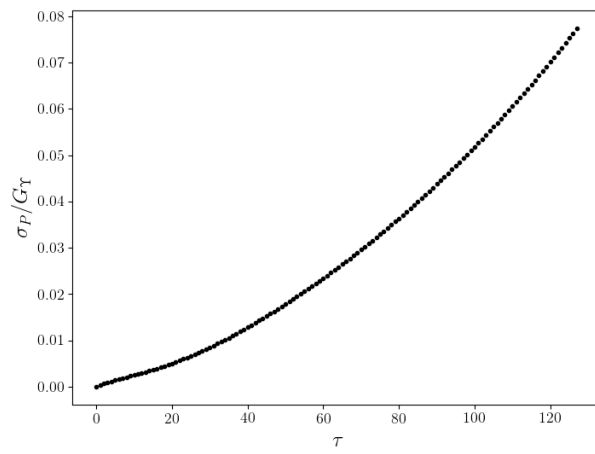
Whilst the dataset does not imitate the lattice correlators as well as the previous two cases, that is not to say it is inadequate. The majority of the dataset is close to the lattice correlator, though it is worse for the P-wave data, than it was in the previous cases. The spread of the data is also similar to that of the alternative dataset of Sec. 5.3.1. Despite the previous datasets appearing to more appropriate training data, the greater range of possible spectral functions provided by this current dataset makes it worth trialing.

5.4 NRQCD correlator noise

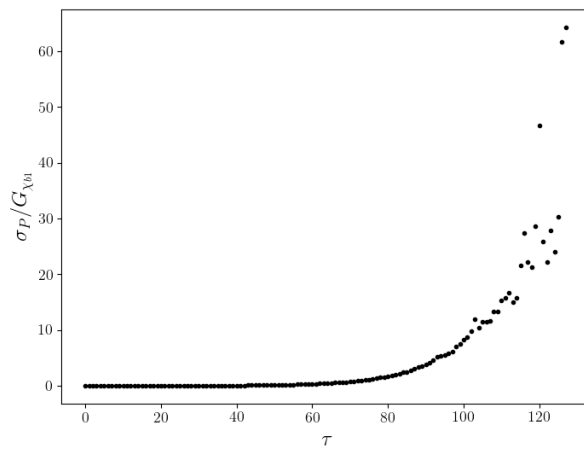
Until now we have ignored an important component of regression; noise or uncertainty on the training data. This is introduced in order to better simulate the lattice data with its own level of uncertainty, which was observed in Chapter 3 when calculating the effective masses. Given that we have attempted to generate data that resembles these ensembles, it follows that we would use the lattice correlators again to estimate suitable noise. Noise will be added to the training data by adding a value chosen from a multivariate Gaussian distribution for each timeslice, $\eta(0, \sigma_{S,P})$, where $\sigma_{S,P}$ is the standard deviation of the distribution. The subscript refers to whether the training data is for the S-wave or P-wave states. As there is a separate dataset for the S and P-wave states, excluding h_b , it follows that the noise to be added is dependent on which state the KRR model is being trained for. The method for estimating the appropriate noise for the training data is as follows: first determine the standard deviation for each state and N_τ at each value of τ , then add in quadrature for S-wave and P-wave states.

$$\sigma_S = \sqrt{\sigma_{\eta_b}^2 + \sigma_{\Upsilon}^2}, \quad \sigma_P = \sqrt{\sigma_{\chi_{b0}}^2 + \sigma_{\chi_{b1}}^2} \quad (5.4.1)$$

Given the results of Chapter 3, it is expected that the noise will be more significant in the P-wave states than S-wave states, see Fig. 3.1. to confirm this, the relative errors for $N_\tau = 128$ at each timeslice, σ_S/G_Υ and $\sigma_P/G_{\chi_{b1}}$, are plotted in Fig. 5.13 using the central values for the Υ and χ_{b1} . Indeed we see a significant difference between the S-wave and P-wave states. For S-wave we see an increase of $\sim 8\%$ whilst for the P-wave, the increase is $\sim 6000\%$. If we decide to add noise to the training data, we may need an alternative for the P-wave states.



(a)



(b)

Figure 5.13: Relative uncertainty for Υ (a) and χ_{b1} (b) for $N_\tau = 128$.

Chapter 6

Improvements to the KRR Model

In the previous chapter several datasets were generated and compared to the lattice correlators to determine if they are appropriate for training data. Now we look to improve KRR model using the improved datasets from Sec. 5.3.1 and 5.3.2, despite the improved dataset of Sec. 5.2 appearing to be better suited. This is because the chosen datasets potentially correspond to a greater variety of spectral functions. We state potentially since this cannot be confirmed without checking all 20000 functions for each dataset. For the proceeding work we will refer to the improved datasets from Sec. 5.3.1 and 5.3.2. as the Parameter and SVD datasets, as these were the representations used to generate each one.

In the following we will consider several methods that may potentially improve the KRR model. These will be the scaling the target data, scaling the input data, and increasing the size of the training set. We will also introduce a new method of selecting the optimal hyperparameters.

Throughout this chapter we will focus of four KRR models, in particularly the latter two. These are: model 1 which is the model from Chapter 4, model 2 which uses the Parameter dataset and SVD representation, model 3 which uses Parameter dataset and $\boldsymbol{\theta} = (m_r, x_r, a_r)$ representation, and model 4 which uses the SVD dataset and SVD representation.

6.1 Initial Testing

Before any attempts are made at improving the KRR, we will compare the four models. However, there is an issue that requires that an alternative measure of performance be used. The magnitude of the parameters $\boldsymbol{\theta} = (m_r, x_r, a_r)$ differs significantly from the coefficients $\boldsymbol{\theta} = (c_1, \dots, c_{80})$, making the comparison of MSEs difficult. Hence, the mean coefficient of determination, \bar{R}^2 will be used instead. This

is given by,

$$\bar{R}_{\text{test}}^2 = \frac{1}{N_\theta} \sum_{j=1}^{N_\theta} R_j^2, \quad (6.1.1)$$

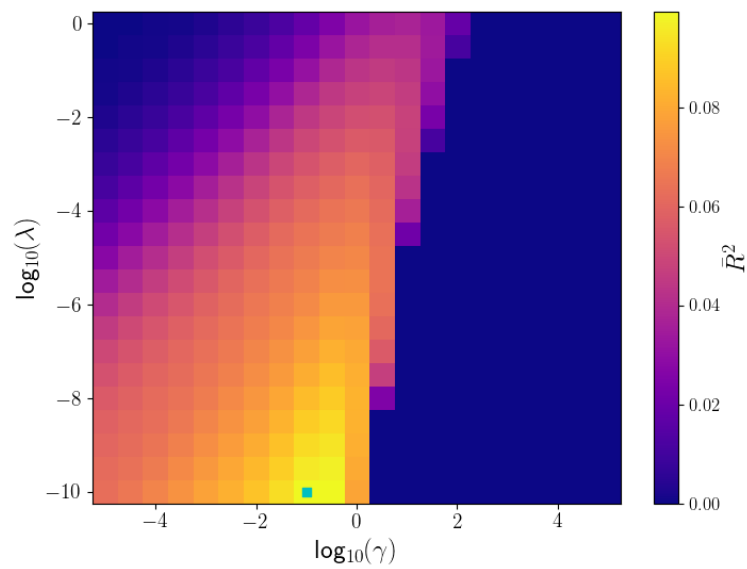
where,

$$R_j^2 = 1 - \frac{\left(\boldsymbol{\theta}_j^{\text{pred}} - \boldsymbol{\theta}_j^{\text{true}}\right)^2}{\left(\boldsymbol{\theta}_j^{\text{true}} - \bar{\boldsymbol{\theta}}_j^{\text{true}}\right)^2}, \quad (6.1.2)$$

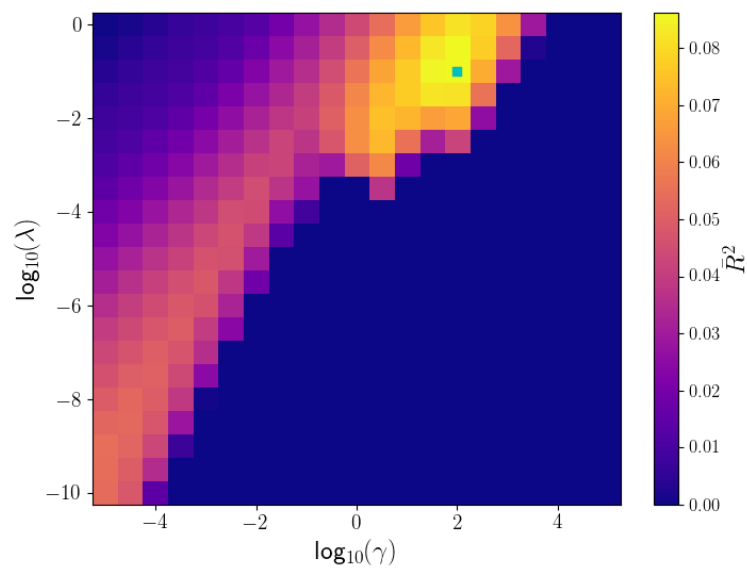
where $\bar{\boldsymbol{\theta}}^{\text{true}}$ is the mean of the j -th parameter of the test set. Currently the optimal hyperparameters are determined using only 1000 test functions without any consideration for the larger training set. To counter this, we now use a larger test set of $N_{\text{test}} = 5000$. The quantity \bar{R}^2 can vary between 1 and $-\infty$. A value of 1 would mean our model is able to perfectly predict the true values of all the parameters. A value of 0 for \bar{R}^2 could either mean that this is the case for all parameters or that the sum in Eq. (6.1.1) is equal to 0 due to the presence of negative and positive R_j^2 . Finally, a negative value of \bar{R}^2 would mean that any predictions made by the model are worse than simply taking the mean of the true parameters. For this reason, any negative values of \bar{R}^2 have been manually set to 0.

To begin, we compare models 1 and 2, using the similar setups for the KRR models. Recall that in Chapter 4 we used the SVD representation with 80 basis functions for $N_\tau = 40$. For consistency, we generated an additional set of mock correlators using Eq. (4.2.14) as was done in Chapter 4. For a fair comparison, we performed a grid search for each of the datasets to determine optimal hyperparameters. Although a grid search has already been done for the original dataset of Chapter 4, this was done using the testing MSE and for a smaller test set so there is no guarantee that this will select the same optimal hyperparameters. As before, we used $N_\tau = 40$. In Fig. 6.1, we have plotted the colourmaps of \bar{R}^2 for the two KRR models.

First of all, we find that redoing the grid search for the original data was the correct decision, as the optimal γ has shifted from $10^{-0.5}$ to 10^{-1} . Note how the regions of better hyperparameters - the areas of yellow in the surface plots - have shifted with the change in training data. For the new model the optimal hyperparameters are $(\lambda, \gamma) = (10^{-1}, 10^2)$. Apart from this though there is little change between the two models in terms of performance. In fact, it would appear the new model is marginally worse.



(a) Model 3



(b) Model 4

Figure 6.1: Colourmaps of \bar{R}^2 for models 1 (a) and 2 (b). The green square marks the position of optimal hyperparameters.

Model	\bar{R}_{train}^2	\bar{R}_{test}^2	$\text{MSE}_{\text{train}}$	MSE_{test}
1	0.1274	0.0994	2570.95	2802.38
2	0.1724	0.0862	4130.29	4404.49
3	0.2743	0.2554	0.0495	0.0494
4	0.0386	0.0291	498.11	502.52

Table 6.1: \bar{R}^2 and MSEs for training and testing sets for models 1, 2, 3, and 4.

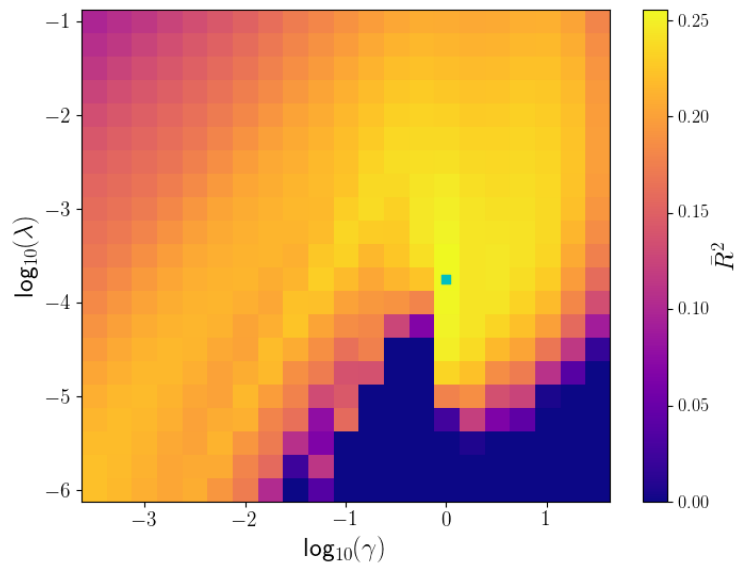
From this point on, we focus on models 3 and 4. Again, we have done a grid search to determine the optimal hyperparameters so that a comparison with the previous two models can be made. The colourmaps representing these grid searches are shown in Fig. 6.2. The training and testing MSEs and \bar{R}^2 are displayed in Table 6.1. The optimal hyperparameters for models 3 and 4 are $(\lambda, \gamma) = (10^{-3.75}, 1)$ and $(10^{-2}, 10)$ respectively.

The MSEs have been shown to justify our change in metric; clearly the different orders of magnitude for the different models make it difficult to compare the models with the MSEs. With these alone, it would appear model 4 is better than model 1 or 2. Yet with \bar{R}^2 , we see that model 4 is the worst. Model 2 is worse than model 1, yet not as significant as the MSEs would suggest. Certainly though the superior model at this time is model 3. Even so, a value of ~ 0.25 for \bar{R}^2 is a poor result, and provides motivation to find methods for improving the KRR models.

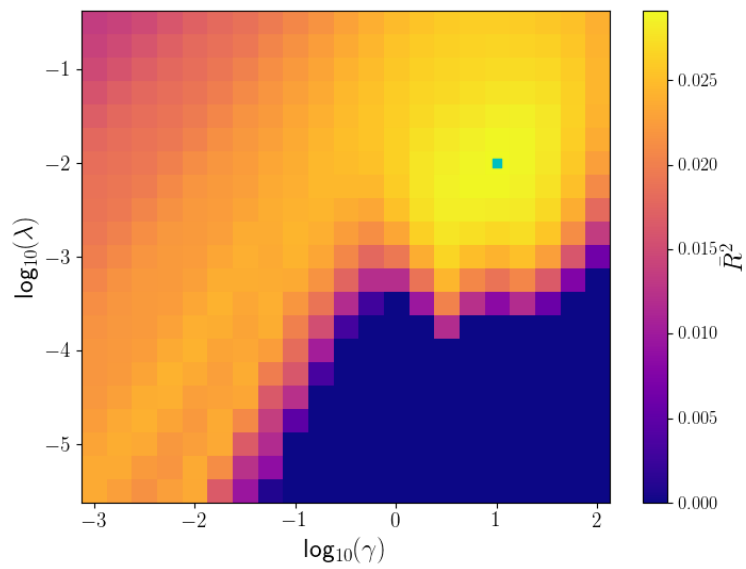
6.2 Attempts for improvements

6.2.1 Scaling of Θ

The first proposed method of improving the models is to scale the target data, Θ , such that all elements are of $\mathcal{O}(1)$. If the metric for determining the fit of the model were still the MSE, then this scaling may be more beneficial. As we have already seen in Fig. 5.10, the coefficients obtained by Bryan's method differ by several orders of magnitude with the first contributing significant to the MSE. Whilst to a lesser extent, this is true in the case where $\theta_j = (m_r, x_r, a_r)$ as well. However, as \bar{R}^2 is now the chosen metric, any improvement to the model is likely to be less significant.



(a) Model 3



(b) Model 4

Figure 6.2: Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b). The green square marks the position of optimal hyperparameters.

Even so, we will consider the technique in the hope of some improvement.

The target data is scaled as follows,

$$\boldsymbol{\theta}_j \rightarrow \boldsymbol{\theta}'_j = \frac{\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}}}{\boldsymbol{\sigma}_\theta}, \quad (6.2.1)$$

where $\bar{\boldsymbol{\theta}}$ is the mean set of parameters for the training data, and $\boldsymbol{\sigma}_\theta$ is the standard deviation of the parameters in the training data. Note that the testing set is also transformed this way using the mean and standard deviation of the training set. Again a grid search was performed for models 3 and 4, but with scaled target data.

The results of these grid search are displayed in Fig. 6.3. The optimal hyperparameters for both models have not changed with the scaling of target data. The \bar{R}^2 for the training and testing set are given in Table 6.2. For model 3 there is noticeably improvement but the \bar{R}^2 for training and testing sets are still lower than we would like. The improvement for model 4 it is negligible. Thus we continue with our attempts to improve the models.

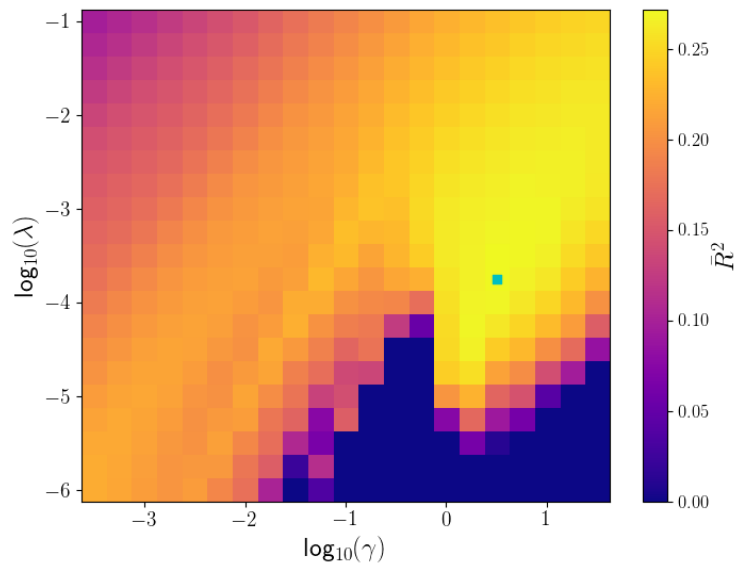
Model	\bar{R}^2_{train}	\bar{R}^2_{test}
3	0.2848	0.2717
4	0.0387	0.0295

Table 6.2: \bar{R}^2 and MSEs for training and testing sets for models 3 and 4 with scaled target data.

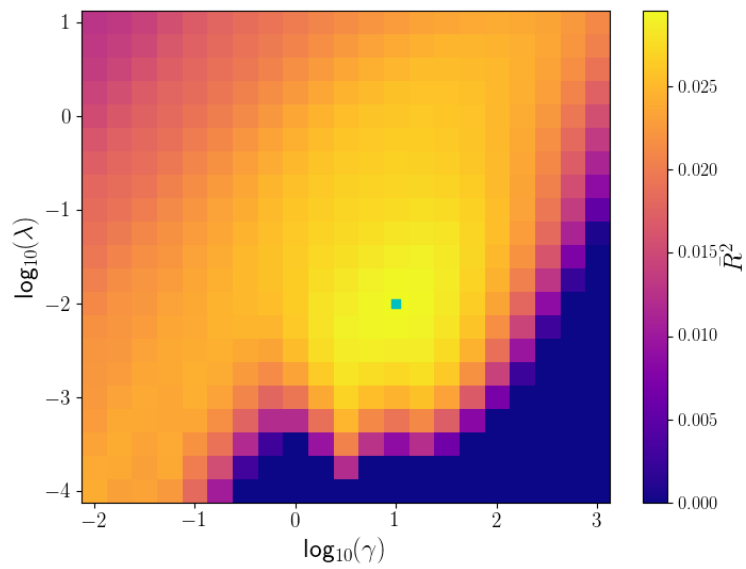
6.2.2 Scaling of correlators

Next we consider scaling the correlators in the hopes of further improving the KRR models. The motivation for scaling the correlators is so that all timeslices of a correlator are of equal magnitude. As we discussed in Sec. 4.4, early timeslices are likely be the largest contributors to the L2 distance in all the kernels. However, we would rather the value of $G(\tau)$ at each timeslices contributes relatively equally. The scaling of the correlators was done using $G_j(\tau) \rightarrow G'_j(\tau) = G_j(\tau)/\bar{G}(\tau)$, where $\bar{G}(\tau)$ is the mean of the training correlators. This was also done for the testing set though the denominator of this ratio is still the mean of the training set.

The results of this scaling are plotted in Fig. 6.4. The optimal hyperparameters have shifted for both models with this scaling. The new optimal pairs are $(\lambda, \gamma) =$



(a) Model 3



(b) Model 4

Figure 6.3: Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b) with scaled target data. The green square marks the position of optimal hyperparameters.

$(10^{-4.25}, 10^{-3})$ and $(10^{-1.2}, 10^{-1.8})$ for model 3 and model 4 respectively. The \bar{R}^2 for the training and testing set are given in Table 6.3. Again we see for model 3 there is noticeable improvement but to a lesser extent than the previous case. The \bar{R}^2 for training and testing sets are still lower than we would like. For model 4 only the training \bar{R}^2 has improved and to a greater extent than the previous case. As for the testing \bar{R}^2 it is negligibly worse.

Model	\bar{R}_{train}^2	\bar{R}_{test}^2
3	0.2967	0.2768
4	0.0453	0.0289

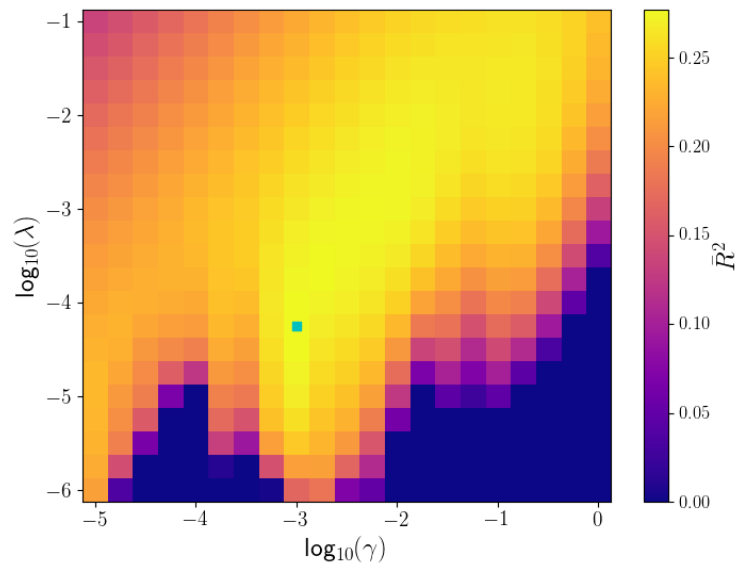
Table 6.3: \bar{R}^2 and MSEs for training and testing sets for models 3 and 4 for scaled target data and scaled correlators which are input data.

6.2.3 Training Set Size

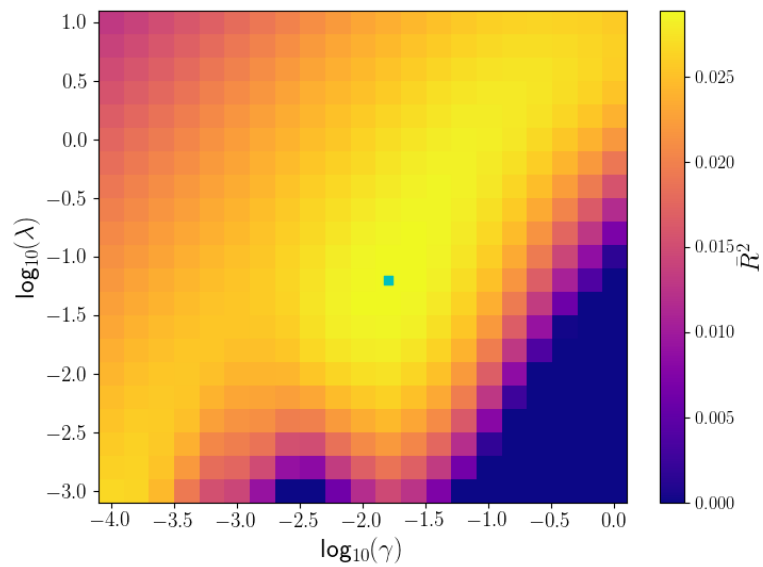
Until now we have used $N_{\text{train}} = 10000$. This is partly due to computing resources and to ensure the grid searches could be completed in reasonable time. However with access to better computing resources such as Swansea’s supercomputer, Sunbird, we can consider using a larger training set. For this test N_{train} ranges from 1000 to 19000 in increments of 1000, whilst $N_{\text{test}} = 1000$ again to prevent any overlap of the two datasets. The models 3 and 4 were used with scaled target and input data, with their respective optimal hyperparameters given in Sec. 6.2.2. The training and testing \bar{R}^2 for the training sizes are plotted in Fig. 6.5. We see that \bar{R}_{train}^2 acts as an upper limit for \bar{R}_{test}^2 , and that there is little improvement in either model if we increase the training size beyond 10000. Increasing the training size beyond 20000 may result in further improvement but we are not able to perform grid searches for this size of training set with the current computing resources.

6.2.4 Cross-validation

Despite there only being a minor increase in \bar{R}_{test}^2 beyond 10000, we chose to increase the training size to 18000. This is because it will provide a greater range of functions to learn from when it comes to finally applying the KRR models to the NRQCD

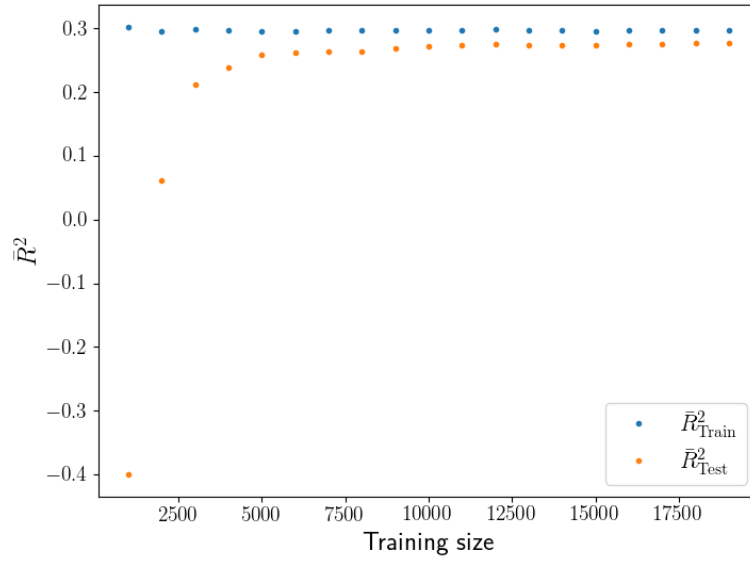


(a) Model 3

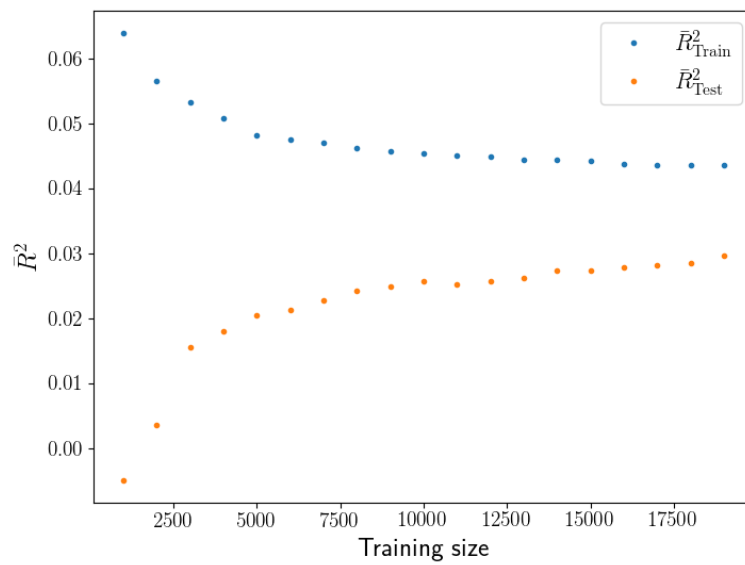


(b) Model 4

Figure 6.4: Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b) with scaled correlators and target data. The green square marks the position of optimal hyperparameters.



(a) Model 3



(b) Model 4

Figure 6.5: Training and testing \bar{R}^2 as a function training data size for Model 3 (a) and Model 4 (b).

lattice correlators. We keep the final 2000 functions for testing purposes.

It has already been discussed that the issue with selecting optimal hyperparameters is dependent solely on optimizing \bar{R}_{test}^2 . Consider Fig. 6.5, where any increase in \bar{R}_{test}^2 comes at the expense of \bar{R}_{train}^2 . In Chapter 4, Fig. 4.11, we see the training MSE increase as the testing MSE approaches the minimum for $1 \leq \gamma \leq 10^4$. In Fig. 6.6 we have plotted \bar{R}_{train}^2 for models 3 and 4 using the scaled input and target data. Again we see that \bar{R}_{train}^2 is reduced in order to improve \bar{R}_{test}^2 .

The issue with this is that the testing set is the minority, especially now it only consists of 2000 functions. A new method for selecting optimal hyperparameters is required that takes both the training and testing set into consideration. This can be done using k -fold cross-validation. This is similar to the grid search method, except the KRR models are trained k times at each pair of hyperparameters. The full training set is split into a smaller training set and a testing set, where the size of testing set is $N_{\text{train}}/k = 18000/k$ and smaller training set consists of the rest of the functions. For this following work $k = 5$. Each iteration of training for a particular pair of hyperparameters using a different set of functions for training and testing, such that no function appears in two or more of the testing sets. We have shown schematically in Fig. 6.7 for a single pair of hyperparameters. For each iteration \bar{R}_{train}^2 and \bar{R}_{test}^2 are determined, and the mean of these are used for the final estimates of \bar{R}_{train}^2 and \bar{R}_{test}^2 . Using 5-fold CV, the optimal hyperparameters were found to be $(\lambda, \gamma) = (10^{-2.75}, 10^{-1.5})$ and $(\lambda, \gamma) = (10^{-1}, 10^{-1.6})$ for models 3 and 4 respectively. The \bar{R}^2 for the training and testing set for both models are given in Table 6.4. Similar to the previous cases there are only minor increases in \bar{R}^2 .

Model	\bar{R}_{train}^2	\bar{R}_{test}^2
3	0.3257	0.2790
4	0.0455	0.0309

Table 6.4: \bar{R}^2 and MSEs for training and testing sets for models 3 and 4 after 5-fold cross-validation.

Unfortunately this has exhausted our ideas for improving the KRR models. Despite this, both KRR models will be applied using the NRQCD lattice data. For the model 3, this is to obtain estimates for the groundstate mass and width for the Υ

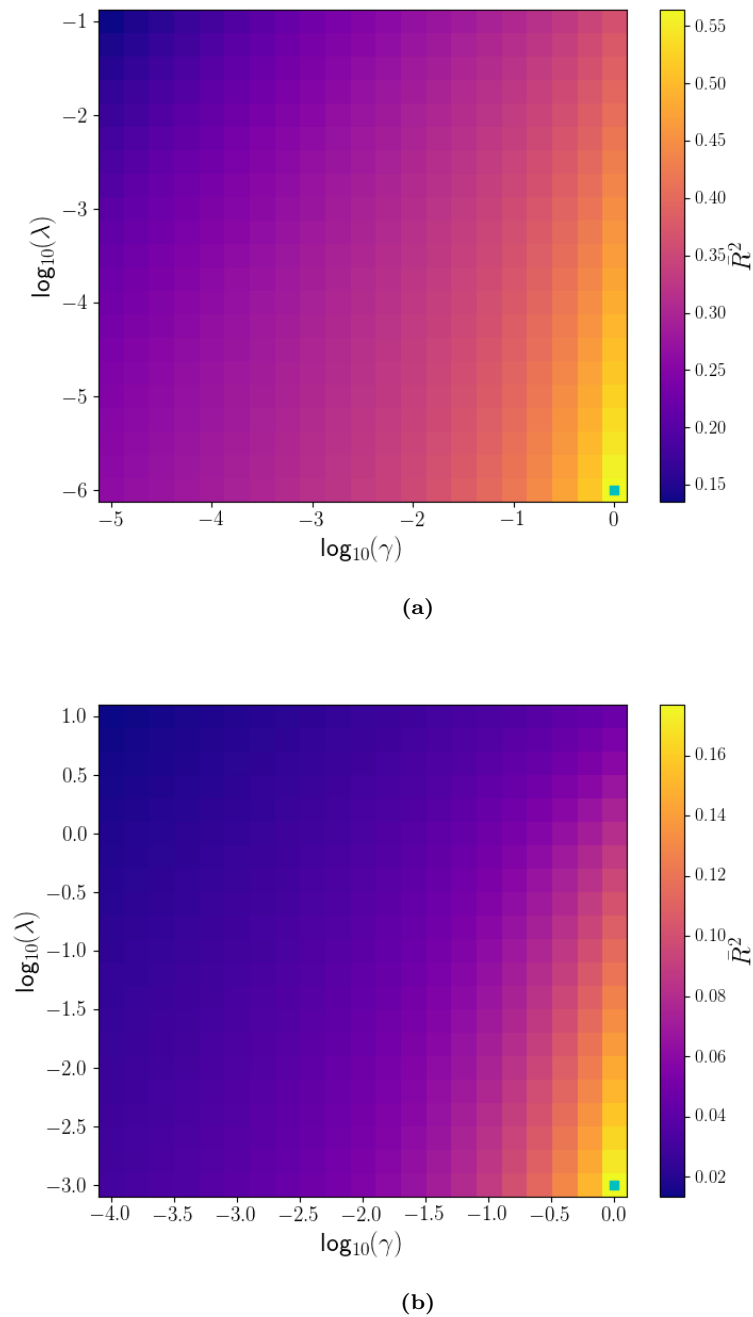


Figure 6.6: Colourmaps of \bar{R}_{train}^2 for models 3 (a) and 4 (b) with scaled correlators and target data. The green square marks the position of optimal hyperparameters.

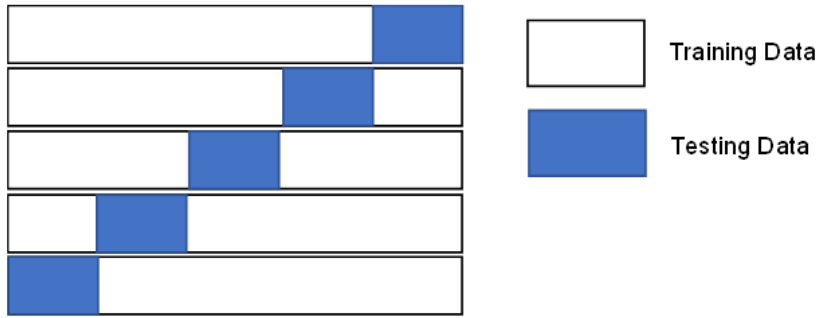
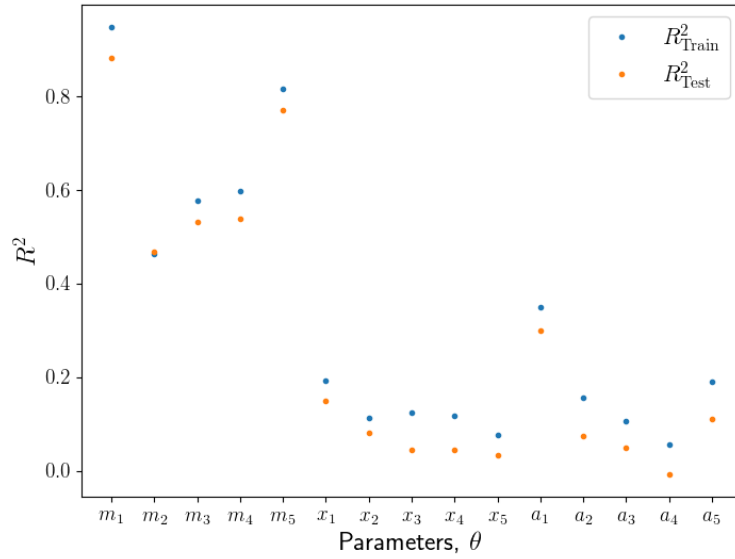


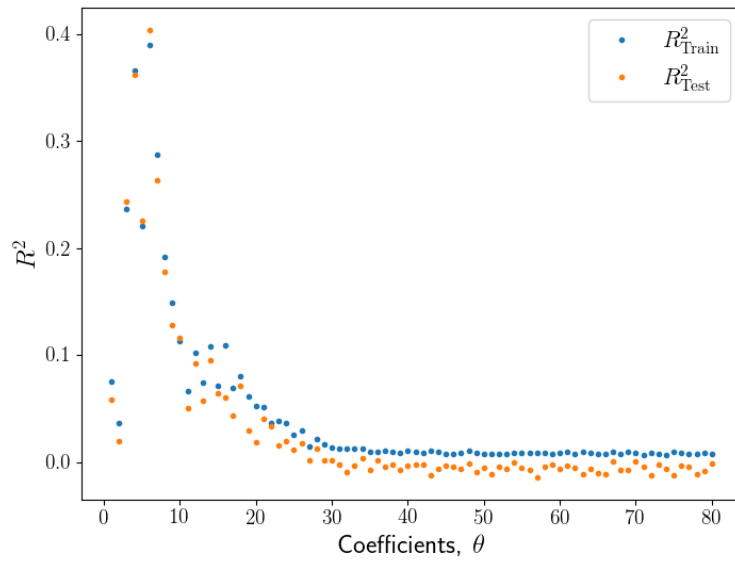
Figure 6.7: Schematic of 5-fold cross-validation. Each row represent a different iteration. Blue blocks represent the testing set and white space of each row represent the training data.

and χ_{b1} . For model 4, this is simply to investigate the potential of a SVD dataset. Before we implement this let us consider the R^2 for the individual parameters in θ for each model. To do this, each model was trained using 18000 functions and the optimal hyperparameters determined by the 5-fold cross-validation. Testing was then done with the final 2000 functions of the mock dataset. The individual R^2 for the parameters are plotted in Fig. 6.8.

We find that whilst model 3 is reasonably good at predicting the masses of the resonances, it suffers with the other parameters, especially the widths. Whilst it is comforting that when we apply it to the NRQCD correlators the prediction for the groundstate masses should be reliable, that the prediction for the widths will be not is disappointing. The amplitude is of less concern. The R^2 of the coefficients of model 4 make it clear why the \bar{R}_{test}^2 has always been so low. It is interesting that the largest R^2 corresponds to the 5th coefficient; not because it is the 5th specifically but that there is a specific coefficient that the model has favoured. It is unclear why the R^2 for model 4 has this pattern but if the underlying reason can be determined it should be possible to improve the model. This could also true for model 3 with regards to the widths and amplitudes. The issue with model 4 is that we need the later coefficients for the finer details of the spectral functions such as the narrow peaks. Even so, we will attempt to implement both models now using the NRQCD lattice correlators.



(a)



(b)

Figure 6.8: Individual R^2 for the training and testing data models 3 (a) and 4 (b) using 18000 training functions and the hyperparameters determined using 5-fold cross-validation.

6.3 Application to NRQCD

We will be using model 3 to predict the groundstate masses and widths for the $\eta_b, \Upsilon, \chi_{b0}$ and χ_{b1} , along with the corresponding spectral functions. For model 4 we will only be attempting to produce reasonable spectral functions to determine whether the SVD dataset is a viable mock dataset. In order to do a full analysis, we needed to do the cross-validation for all temperatures. For transparency, the optimal hyperparameters for both models and for both S and P-wave datasets are shown in Table 6.5.

N_τ	20	24	28	32	36	40	48	56	64	128
Model 3, S-wave										
$\log_{10}(\lambda_{\text{opt}})$	-3.25	-3.5	-3.0	-3.5	-2.75	-2.75	-3.0	-3.0	-3.0	-3.0
$\log_{10}(\gamma_{\text{opt}})$	-1.5	-1.75	-1.75	-2.25	-1.5	-1.5	-1.75	-1.75	-1.75	-1.75
Model 3, P-wave										
$\log_{10}(\lambda_{\text{opt}})$	-3.5	-3.75	-3.5	-3.75	-4.0	-3.75	-3.0	-2.75	-2.75	-2.75
$\log_{10}(\gamma_{\text{opt}})$	-1.5	-2.0	-1.5	-1.75	-2.0	-2.0	-1.25	-1.25	-1.25	-1.25
Model 4, S-wave										
$\log_{10}(\lambda_{\text{opt}})$	-1.8	-1.8	-2.2	-1.6	-1.4	-1.0	-1.2	-1.4	-1.0	0.6
$\log_{10}(\gamma_{\text{opt}})$	-1.0	-1.2	-1.8	-1.6	-1.6	-1.6	-1.8	-2.0	-1.8	-0.8
Model 4, P-wave										
$\log_{10}(\lambda_{\text{opt}})$	-2.4	-1.8	-1.8	-1.2	-1.0	-1.0	-0.6	0.4	0.4	0.2
$\log_{10}(\gamma_{\text{opt}})$	-1.2	-1.2	-1.4	-1.2	-1.2	-1.2	-1.2	-0.4	-0.4	-0.6

Table 6.5: Optimal hyperparameters for models 3 and 4 for both S-wave and P-wave datasets. Provided for reproducibility purposes.

The predictions for the groundstate masses and widths for the four states mentioned above are shown in Table 6.6 - 6.9 done using these hyperparameters with the appropriate training data for model 3. There are two uncertainties associated with each of these predictions. The first is the statistical δ_{stat} uncertainty that is generated from bootstrapping using the configurations rather than the central value. The second is the systematic that is an estimate of the error of the KRR model. It

is determined using the spare 2000 functions from the mock dataset. The model is trained, then tested using these spare functions. The mean-absolute-error (MAE) for each parameter θ_j which could be one of $(m_1, m_2, \dots, x_1, x_2, \dots, a_4, a_5)$ is determined using,

$$\text{MAE}_j = \frac{1}{2000} \sum_{n=1}^{2000} |\theta_n^{(\text{test})} - \theta_n^{(\text{pred})}|, \quad (6.3.1)$$

and taken as the estimate of uncertainty of the model itself. In Table 6.6 and 6.7 there is no $\delta\Gamma_{\text{stat}}$ stated as it is several orders of magnitude smaller than $\delta\Gamma_{\text{sys}}$.

N_τ	128	64	56	48	40	36	32	28	24	20
T [MeV]	47	94	107	125	150	167	187	214	250	300
m_0 [MeV]	9381	9382	9382	9381	9382	9380	9369	9374	9382	9385
δm_{stat} [MeV]	1	2	2	2	2	2	2	2	3	3
δm_{sys} [MeV]	36	36	36	36	34	35	41	40	42	48
Γ_0 [keV]	19.2	19.2	19.2	19.2	19.1	19.3	19.6	20.3	20.8	22.2
$\delta\Gamma_{\text{sys}}$ [keV]	11.2	11.2	11.2	11.2	11.2	11.3	11.5	12.0	12.3	13.2

Table 6.6: Predicted groundstate mass and width for η_b for the range of temperatures using model 3.

N_τ	128	64	56	48	40	36	32	28	24	20
T [MeV]	47	94	107	125	150	167	187	214	250	300
m_0 [MeV]	9404	9404	9404	9403	9404	9402	9390	9395	9400	9399
δm_{stat} [MeV]	1	1	1	1	1	1	1	1	1	1
δm_{sys} [MeV]	36	36	36	36	34	35	41	40	42	48
Γ_0 [keV]	18.3	18.3	18.3	18.3	18.3	18.5	18.2	19.2	19.5	20.6
$\delta\Gamma_{\text{sys}}$ [keV]	10.7	10.7	10.7	10.7	10.7	10.8	10.7	11.3	11.6	12.3

Table 6.7: Predicted groundstate mass and width for Υ for the range of temperatures using model 3.

N_τ	128	64	56	48	40	36	32	28	24	20
T [MeV]	47	94	107	125	150	167	187	214	250	300
m_0 [MeV]	9917	9911	9912	9904	9898	9869	9832	9828	9918	9941
δm_{stat} [MeV]	56	52	56	64	57	66	70	51	43	47
δm_{sys} [MeV]	16	16	16	16	18	18	18	20	26	30
Γ_0 [keV]	7.7	7.5	7.6	7.5	7.5	7.2	6.6	6.9	8.9	10.0
$\delta\Gamma_{\text{stat}}$ [keV]	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.1	0.1	0.1
$\delta\Gamma_{\text{sys}}$ [keV]	2.1	2.0	2.0	2.0	2.0	1.9	1.8	1.9	2.4	2.8

Table 6.8: Predicted groundstate mass and width for χ_{b0} for the range of temperatures using model 3.

N_τ	128	64	56	48	40	36	32	28	24	20
T [MeV]	47	94	107	125	150	167	187	214	250	300
m_0 [MeV]	9945	9941	9941	9934	9934	9907	9869	9864	9952	9969
δm_{stat} [MeV]	33	32	34	41	35	41	44	31	26	30
δm_{sys} [MeV]	36	36	36	36	34	355	41	40	42	48
Γ_0 [keV]	7.48	7.38	7.38	7.27	7.40	7.07	6.55	6.81	8.68	9.64
$\delta\Gamma_{\text{stat}}$ [keV]	0.14	0.13	0.14	0.16	0.14	0.15	0.13	0.08	0.05	0.04
$\delta\Gamma_{\text{sys}}$ [keV]	2.00	1.98	1.98	1.95	1.98	1.89	1.77	1.84	2.37	2.65

Table 6.9: Predicted groundstate mass and width for χ_{b1} for the range of temperatures using model 3.

These results are summarised in Fig. 6.9, with the experimental masses for comparison. The errorbars represent the error associated with the quantities added in quadrature. It is encouraging to see that with the exception of the Υ , the predictions for masses largely agree with these experimental values. It is surprising that the Υ does not agree given that it was the state used to tune the Gen2L ensembles. It is unclear at this time if this is coincidence or not. The masses for S-wave states appear to be consistent which we expect from our work on the thermal modification in Chapter 3, in that there is little change with temperature. We would however expect to see some increase in P-wave masses as temperature increases. The behaviour we see instead is a possible indicator that our training data is still inadequate as

there should not be the dip between 175 - 225 MeV.

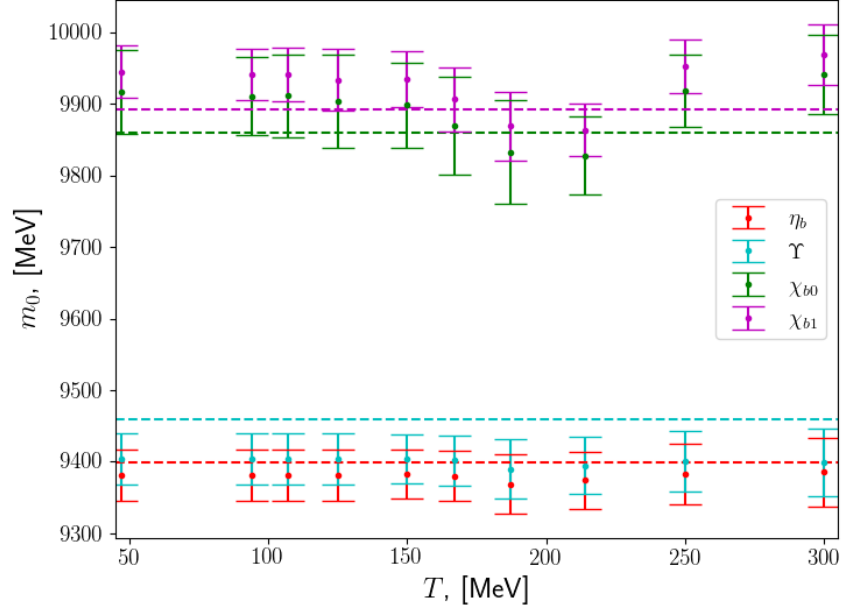
Regarding the width predictions, recall from Chapter 2 that the widths of these resonances are related to the thermal effects of medium and not on any decays as we have not included QED or the weak force in the theoretical model for the lattice simulations. Therefore we have no experimental results to compare these widths. We also note that these predictions are limited by the training data, making it impossible to predict a width smaller than any Γ used in the mock spectral functions. The purpose of these width prediction was to test the model, so that we may be able to compare with other methods in the future. There is one interesting observation though, which is the significant difference in uncertainty between the S-wave and P-wave states. We see that the S-wave states suffer from a large uncertainty, though this is not surprising given Fig. 6.8. It is surprising that the P-waves do not suffer from a similar level of uncertainty. At this time, it is unclear why there is this discrepancy between the S and P-wave states. Finally we note that for the S-wave states, there is a slight increase in groundstate width as $T > 200$ MeV, but given the large uncertainties this cannot be confirmed.

The final inquiry of this work will be to reconstruct the spectral functions of the correlators using both models. The spectral functions produced from using model 3 for the S-wave and P-wave states shown in Fig. 6.10 and Fig. 6.11 respectively, for a selection of temperature. We have omitted error bands for clarity, as the motivation of this test is to determine whether this method has further potential.

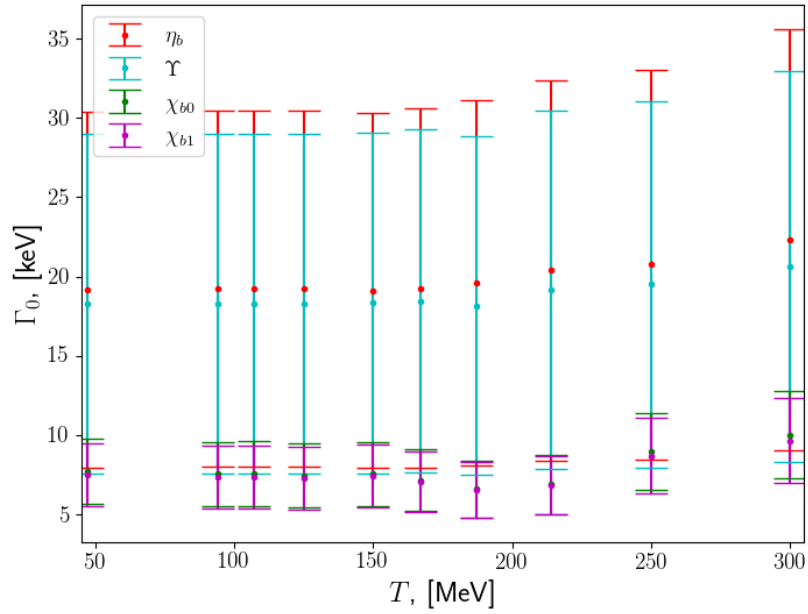
For both S-wave states, the predictions are fairly consistent though the resonances fail to match with any of the exponential fits. The first resonance appears close, but we have already shown in Fig. 6.9a that this is not the case. For the P-wave state there is reasonable agreement across the prediction for the position of the groundstate, a central resonance, and the final one. However, none of the temperatures are capable of matching the exponential fits beyond the groundstate. However the interesting feature of these predictions is the final peak. Though we have not included a continuum feature in the mock data, it could be that the model is making use of the final resonance as a substitute for the continuum. It could simply be a coincidence though.

In Fig. 6.12 we have plotted the predictions using model 4 for the S-wave states

for a selection of temperatures. There is little agreement with the exception of the peak at ~ 12 GeV in Fig. 6.12a. Any agreement between the peaks of the spectral functions and the exponential fits could just as likely be by chance as a successful reconstruction of a resonance. However, this dataset and model was developed as a trial of using basis functions to generate a mock dataset. With refinement, it may be possible generate a suitable dataset with greater freedom and range of possible spectral functions than a series of Gaussian functions.



(a)



(b)

Figure 6.9: Prediction of bottomonium groundstate masses (a) and widths (b), with uncertainties. Dashed lines of the same colour in (a) represent the experimental masses for the bottomonium states.

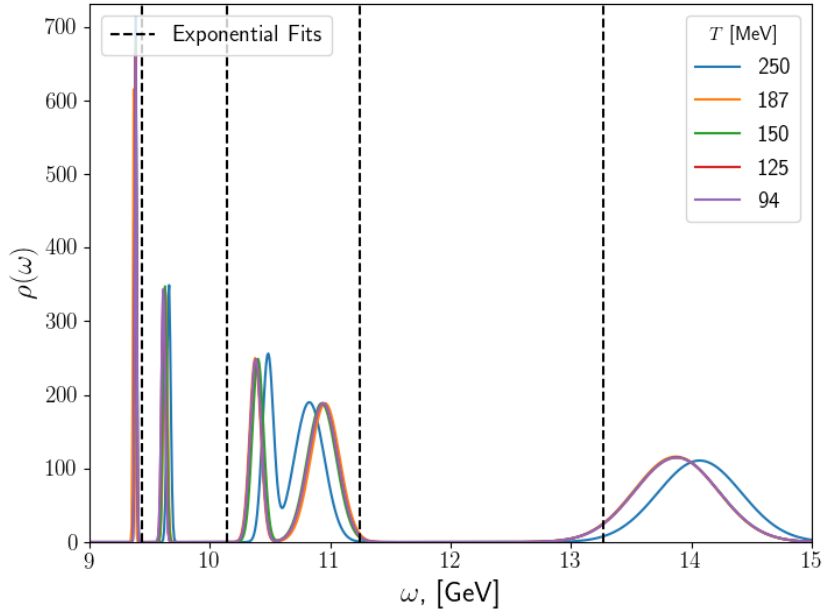
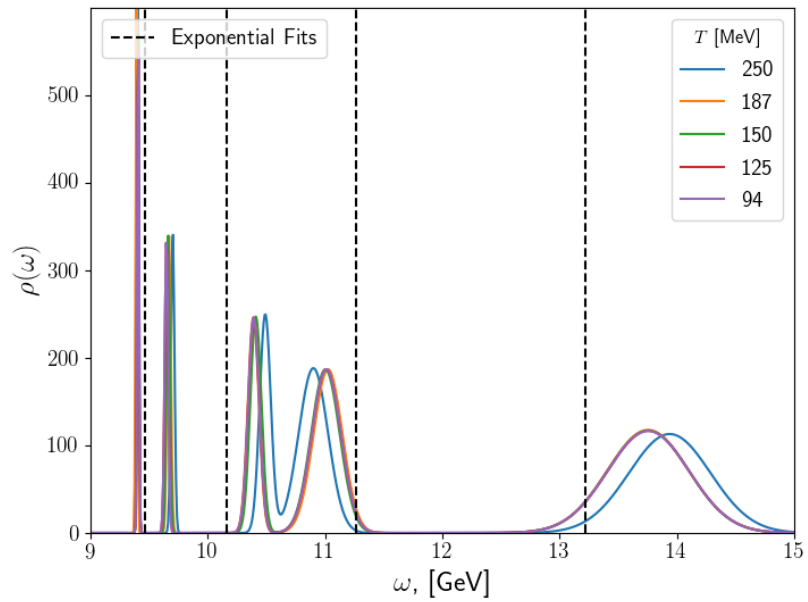

 (a) η_b

 (b) Υ

Figure 6.10: Predicted spectral functions using model 3 for the bottomonium S-wave states: η_b (a), Υ (b).

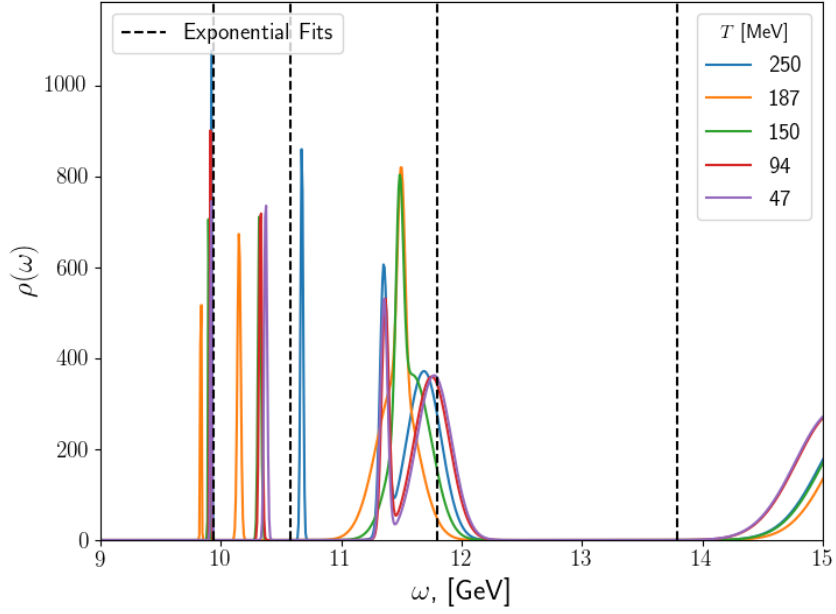
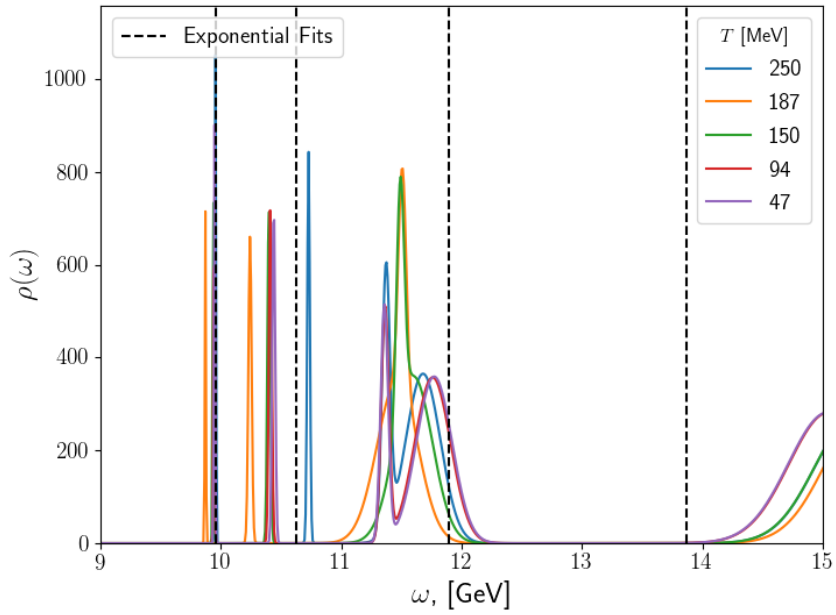
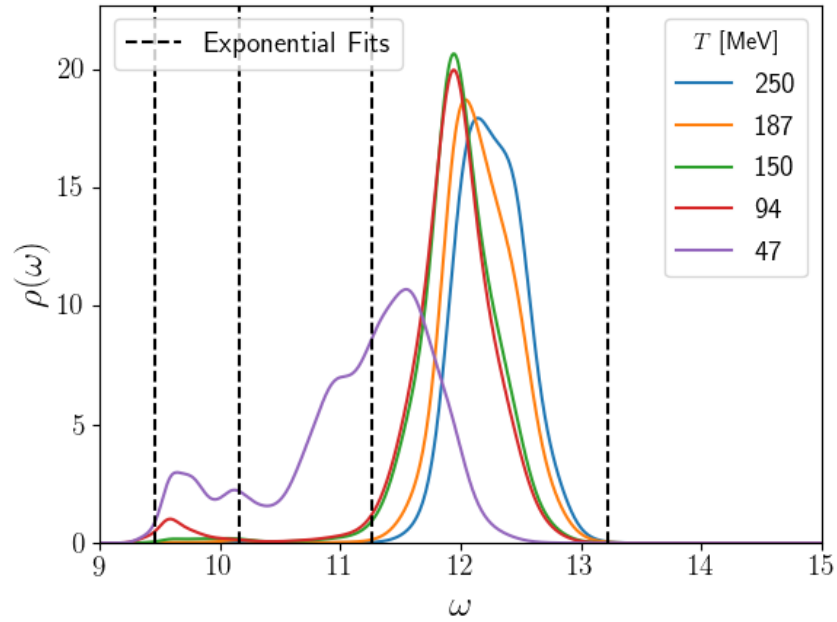
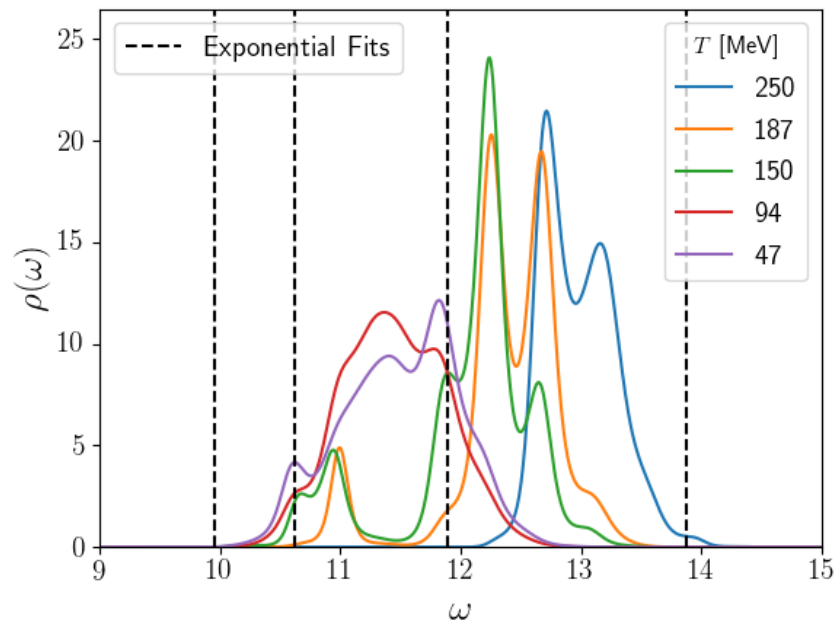

 (a) χ_{b0}

 (b) χ_{b1}

Figure 6.11: Predicted spectral functions using model 3 for the bottomonium P-wave states: χ_{b0} (a), and χ_{b1} (b).



(a) Υ



(b) χ_{b1}

Figure 6.12: Predicted spectral functions using model 4 for Υ (a) and χ_{b1} (b).

Conclusion

The work of this thesis has been to develop a machine learning method capable of investigating thermal NRQCD.

In Chapter 3 we focused on analysis of the NRQCD lattice correlators. Using a simple ansatz of a single δ function for a spectral function, effective masses for the bottomonium states were determined. The effective mass for the Υ was used to set the energy shift $E_0 = 7.465 \text{ GeV}$. The CORFITTER package was used to determine the masses of the groundstate and excited states for the bottomonium states. We investigated the thermal modification of the states and determined that whilst the S-waves are relatively unaffected for the temperatures used in this work, there is a noticeable enhancement for P-wave states of $\sim 12\%$.

In Chapter 4 we developed an initial KRR model. A detailed study of determining optimal hyperparameters was given. It was shown explicitly that they cannot be determined separately and thus a grid search method was implemented. The KRR was tested using the FASTSUM correlators for the Υ for a range of temperatures. The spectral functions predicted by KRR were unexpected and indicated this KRR model was inadequate.

Before trialing methods to improve the KRR model, we considered whether the training data we had was suitable, which we determined it was not. Chapter 5 was dedicated to producing more appropriate data for training the KRR model. The importance of appropriate training data is not restricted to KRR but any regression machine learning method. It is hoped that the ideas put forward in this chapter could help other studies of spectral reconstruction. A method was developed by making use of an initial dataset and selecting appropriate functions by comparing mock correlators with the NRQCD lattice correlators. Two datasets were produced using a series of Gaussian functions for one and basis functions from Bryan's method for the other. This chapter ended with a consideration of how to estimate suitable noise using the NRQCD correlators. This noise could be added to the training data to simulate uncertainty in the system.

The final chapter of the thesis we considered methods with the aim of improving two new KRR models using the new datasets. These methods produced some

improvement but not as much as hoped. Cross-validation was introduced as an alternative to the the grid search method for determining the optimal hyperparameters. The groundstate masses and widths for bottomonium states were predicted. The masses were found to be consistent with experimental values with the exception of the Υ .

There are other studies that this work could be applied to. For one, despite our consideration of estimating appropriate noise, we did not produce a KRR model trained with noisy data in this work. There are also alternative configurations that have used a smeared source. Despite the work done to improve the generation of appropriate data, it could be taken further by including a continuum and/or transport peak.

Despite the promising result of prediction of the groundstate masses and its consistency with experimental mass for three of the bottomonium states, there are still clear issues with the KRR models presented in this work. However, it is hoped that the work and ideas presented in this thesis will provided a step towards a successful method of spectral reconstruction using KRR.

Appendix A

Lattice formulation

In this appendix we outline some of the steps for deriving the Euclidean path integral. The relationship between the path integral and statistical mechanics is discussed afterwards. For a more in depth derivation see [16, 93].

A.1 Path integral of a 1D system

For simplicity we will consider a particle of mass m restricted to motion along the x -axis only, a potential $V(x)$. The Hamiltonian of the system is given by,

$$\hat{H} = \frac{\hat{p}^2}{2m} + \hat{V}(x) \quad (\text{A.1.1})$$

where \hat{p} is the momentum operator,

$$\hat{p} = i \frac{d}{dx}, \quad (\text{A.1.2})$$

with eigenstates $|p\rangle$ and eigenvalues p . Note that we use natural units ($c = \hbar = 1$). The position operator, \hat{x} , with eigenstates $|x\rangle$ and eigenvalues x , and \hat{p} obey the commutation relation,

$$[\hat{x}, \hat{p}] = i. \quad (\text{A.1.3})$$

Consider the system starts at time t_i in an initial position, $x_i(t_i)$ and evolves to with time until some later time t_f in a final position $x_f(t_f)$. This evolution would be given by following transfer probability

$$T_{i,f} = \langle x_i | e^{-i\hat{H}(t_f-t_i)} | x_f \rangle. \quad (\text{A.1.4})$$

It is possible to decompose this into two evolutions by introducing the identity matrix

$$I = \int dx' |x'\rangle \langle x'| = \int dp' |p'\rangle \langle p'| \quad (\text{A.1.5})$$

into the the original definition of the transfer probability,

$$T_{i,f} = \int dx' \langle x_i | e^{-i\hat{H}(t'-t_i)} | x' \rangle \langle x' | e^{-i\hat{H}(t_f-t')} | x_f \rangle. \quad (\text{A.1.6})$$

The integrand represents the evolution of the system from the state $|x_i(t_i)\rangle$ to $|x'(t')\rangle$ then from the intermediate state $|x'(t')\rangle$ to $|x_f(t_f)\rangle$. we obtain the original probability by integrating over all possible x' . We can repeat this process N times so that the time evolution of each iteration becomes infinitesimal, $(t_f - t'_N = t'_N - t'_{N-1} \rightarrow dt)$.

$$T_{i,f} = \int dx'_1 \dots dx'_N \langle x_i | e^{-\hat{H}dt} | x'_1 \rangle \langle x'_1 | e^{-Hdt} | x'_2 \rangle \dots \langle x'_N | e^{-i\hat{H}dt} | x_f \rangle. \quad (\text{A.1.7})$$

Again we obtain the original transfer probability by integrating over each x' . The next step in obtaining the path integral of the system is to introduce the identity matrix again, only this time in terms of the momentum eigenstates before each $|x'\rangle$,

$$\begin{aligned} T_{i,f} &= \int dx'_1 \dots dx'_N dp'_1 \dots dp'_N \langle x_i | e^{-i\hat{H}dt} | p'_1 \rangle \langle p'_1 | x'_1 \rangle \\ &\times \langle x'_1 | e^{-i\hat{H}dt} | p'_2 \rangle \langle p'_2 | x'_2 \rangle \dots \langle x'_N | e^{-i\hat{H}dt} | p'_N \rangle \langle p'_N | x_f \rangle. \end{aligned} \quad (\text{A.1.8})$$

In order to simplify the above, we use the following result

$$\langle x | p \rangle = \frac{1}{\sqrt{2\pi}} e^{ipx}. \quad (\text{A.1.9})$$

For example, if we consider a single evolution and $\hat{V}(x) = 0$.

$$\begin{aligned} \langle x | e^{-t\hat{p}^2/2m} | x' \rangle &= \int dp \langle x | e^{-t\hat{p}^2/2m} | p \rangle \langle p | x' \rangle \\ &= \int dp \langle x | p \rangle \langle p | x' \rangle e^{-tp^2/2m} \\ &= \frac{1}{2\pi} \int dp e^{-i(x-x')p} e^{-tp^2/2m} \\ &= C e^{-(x-x')^2 m/2t}. \end{aligned} \quad (\text{A.1.10})$$

The final step comes from using the Gaussian integral. Note that we had to set $\hat{V}(x) = 0$ for this to work. As $V(x)$ does not commute with \hat{p} it is not a simple matter to reintroduce a non-zero potential. To resolve this, we split the potential symmetrically and for a infinitesimal timestep, dt we write [16],

$$e^{-i\hat{H}dt} = e^{-i\hat{V}dt} e^{-i\hat{p}^2 dt/2m} e^{-i\hat{V}dt} (1 + \mathcal{O}(dt)). \quad (\text{A.1.11})$$

With the leading order of this formula we find,

$$\begin{aligned} \langle x | e^{-i\hat{H}dt} | x' \rangle &= e^{-iV(x)dt} \langle x | e^{-i\hat{p}^2 dt/2m} | x' \rangle e^{-iV(x')dt} \\ &= C e^{-iV(x)dt} e^{-iV(x')dt} e^{-(x-x')^2 m/2dt}. \end{aligned} \quad (\text{A.1.12})$$

We can define $\dot{x}_j(t) = (x'_{j+1} - x_j)/dt + \mathcal{O}(dt)$. With this results, we can redefine Eq. (A.1.7),

$$\begin{aligned} T_{i,f} &= \int dx'_1 \dots dx'_N dp'_1 \dots dp'_N \langle x_i | e^{-i\hat{H}dt} | p'_1 \rangle \langle p'_1 | x'_1 \rangle \\ &\quad \times \langle x'_1 | e^{-i\hat{H}dt} | p'_2 \rangle \langle p'_2 | x'_2 \rangle \dots \langle x'_N | e^{-i\hat{H}dt} | p'_N \rangle \langle p'_N | x_f \rangle \\ &= C^N \int dx'_1 \dots dx'_N \exp \left(-dt \sum_{j=1}^N \left[\frac{m}{2} \dot{x}_j(t)^2 + V(x_j) \right] \right). \end{aligned} \quad (\text{A.1.13})$$

As we increase the number of timesteps $N \rightarrow \infty$, $dt \sum_{j=0}^N \rightarrow \int_{t_i}^{t_f} dt$. Hence,

$$dt \sum_{j=1}^N \left[\frac{m}{2} \dot{x}_j(t)^2 + V(x_j) \right] = \int_{t_i}^{t_f} dt \left(\frac{m}{2} \dot{x}(t)^2 + V(x) \right) = \int_{t_i}^{t_f} dt \mathcal{L}(x, \dot{x}) = S_E[x, \dot{x}], \quad (\text{A.1.14})$$

where $S_E[x, \dot{x}]$ is the Euclidean action.

A.2 Relations to statistical mechanics

To see how we can write the above in the language of statistical mechanics, first consider the partition function Z ,

$$Z = \text{tr} \left[e^{-T\hat{H}} \right] = \int dx \langle x | e^{-T\hat{H}} | x \rangle. \quad (\text{A.2.1})$$

Note that $T = t_f - t_i$; it is not the transfer matrix given above. It follows from Eq. (A.1.13),

$$\begin{aligned} Z &= \int dx_0 T_{i,f} = C^N \int dx_0 dx_1 \dots dx_N \exp \left(-dt \sum_{j=1}^N \left[\frac{m}{2} \dot{x}_j(t)^2 + V(x_j) \right] \right) \\ &= C^N \int \mathcal{D}x e^{-S_E[x, \dot{x}]} \end{aligned} \quad (\text{A.2.2})$$

where $\mathcal{D}x = \prod_j dx_j$. Hence we find we can replace the Boltzman factor $e^{-\beta H}$ from statistical mechanics with the weight factor $e^{-S_E[x, \dot{x}]}$. Given this, we can define the expectation value of an observable O as,

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}x e^{-S_E[x, \dot{x}]} O. \quad (\text{A.2.3})$$

We omit the C^N as it is a constant that can be incorporated into Z .

Appendix B

B.1 Bootstrap sampling

Bootstrap sampling is used in cases where simple error propagation cannot be used to calculate the uncertainty of a result, for example if $f(\mathbf{x})$ is logarithmic. It involves taking a random selection from the dataset allowing individual correlators to be selected multiple times new dataset must be the same size as the original dataset. For example, consider a dataset $\mathbf{x} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$, then the new dataset could look like this $\mathbf{x}' = \{x_2, x_6, x_4, x_6, x_1, x_7, x_6, x_4\}$. In this example the 4th and 6th entries of the original dataset appear multiple times in this new dataset, whilst the 3rd, 5th and 8th entries do not appear at all. This is repeated to generate multiple of these sampled datasets; typically 1000 such datasets are made. From these sampled datasets, estimates for the mean and variance, \bar{x} and σ_x , can be determined from the mean average of each sampled dataset,

$$\bar{x}' = \frac{1}{N} \sum_i \bar{x}'_i, \quad \sigma_{x'} = \frac{1}{N-1} \sum_i (f(x'_i) - \bar{f}(x')) \quad (\text{B.1.1})$$

The variance is determined as would be expected. However this two quantities can be determined directly from the original dataset and so bootstrap sampling is redundant at this stage. Instead it becomes a useful tool when looking to estimate a mean and variance of function that uses the dataset \mathbf{x} , where error propagation is difficult. The mean and variance for this function is then estimated by calculating $f(\mathbf{x}')$ for each sampled dataset. The mean and variance are then estimated using

$$\bar{f}(x) \approx \frac{1}{N} \sum_i f(x'_i). \quad (\text{B.1.2})$$

and

$$\text{Var}(f(x)) \approx \frac{1}{N-1} \sum_i (f(x'_i) - \bar{f}(x))^2. \quad (\text{B.1.3})$$

List of Figures

1.1	Sketch of the Phase diagram of QCD. [6]	4
1.2	Columbia plot: schematic of thermal phase transition behaviour of $N_f = 2 + 1$ QCD at $\mu_B = 0$ for a range of u , d , and s quark masses ($m_u = m_d$) [20].	8
2.1	Sketch of example spectral functions for increasing temperature: (a) Zero temperature, (b) Low temperature which produces the first signs of thermal broadening, (c) High temperature that results in dissociation of particle leading to overlap of different energy states and the continuum.	14
3.1	Effective mass according to Eq. (3.2.7) for following bottomonium states: scalar η_b (a), pseudoscalar Υ (b), vector χ_{b0} (c), and axial vector χ_{b1} (d) channels for $N_\tau = 128$. The orange line represents the mean effective mass of the last 60 points.	28
3.2	Groundstate masses for the bottomonium states using a multi-exponential fit with 1-5 terms. Experimental masses are represented by blue lines. Effective masses are represented by black solid lines and the uncertainty on these values by black dashed lines.	31
3.3	Comparison of reconstructed correlator for $N_p = 1-4$ and FASTSUM Gen2L correlator for the Υ	32
3.4	Comparison of reconstructed correlator for $N_p = 1-4$ and FASTSUM Gen2L correlator for the χ_{b1}	33
3.5	Thermal modification, $G(\tau; T)/G(\tau; T_0)$ with $T_0 = 47$ MeV, of the correlation functions.	34
4.1	Example of Linear Least-Squares regression using polynomials with 4(a), 10(b), and 16(c) terms.	39
4.2	Example of Ridge regression with $\lambda = 0.05$ using polynomials with 4(a), 10(b), and 16(c) terms.	41

4.3	Example of Kernel Ridge regression using a radial basis function (RBF) kernel with $\gamma = 0.1$, and $\lambda = 0.1$	42
4.4	Distribution of unconstrained parameters (blue) and distribution of selected parameters (orange)	47
4.5	Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (1, 1)$	51
4.6	Training and testing MSEs as a function λ . Note the values on both axes are logarithmic.	53
4.7	Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (10^{-10}, 1)$	54
4.8	Examples of an ordered kernel calculated using Eq.(4.1.11) for $\gamma = 10^{-2}$ (a), 1 (b), and 10^2 (c).	56
4.9	Training and testing MSEs as a function γ for $\lambda = 1$. Note the values on both axes are logarithmic.	57
4.10	Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (1, 10^{1.5})$	58
4.11	Training and testing MSEs as a function γ for $\lambda = 10^{-10}$. Note the values on both axes are logarithmic.	59
4.12	Comparison of spectral functions from original and predicted coefficients for MSE_{low} (a), MSE_{med} (b), and MSE_{high} (c), for $(\lambda, \gamma) = (10^{-10}, 10^{1.5})$	60
4.13	Surface plot of MSE_{test} for hyperparameters grid search. The green square marks the optimal hyperparameters.	61
4.14	First attempt at reconstructed spectral function for Υ using KRR for various temperatures. Black dashed lines represent the masses shown in Table 3.2 that were determined using the CORFITTER package with the Υ correlator for $N_\tau = 128$, which corresponds to $T = 47 \text{ MeV}$	63

5.1 Distribution of the natural logarithm of correlators in the mock dataset at each timeslice for $N_\tau = 32$ and 128 for S-wave (a, c) and P-wave (b, d) states. Dashed lines represent the upper and lower limits of the correlators in the mock dataset for each value of n_τ , and white solid line represent mean of the mock correlators. The other solid lines represent the Euclidean correlators obtained in simulations. Note that the two channels in each figure – η_b, Υ (a, c) and χ_{b_0}, χ_{b_1} (b, d) – are hardly distinguishable. 69

5.2 Distribution of the natural logarithm of correlators in the mock dataset at $n_\tau = 7, 31$, and 127 for S-wave (a, c, e) and P-wave (b, d, f) states. Solid lines represent the Euclidean correlators obtained in simulations. 71

5.3 Distribution of the natural logarithm of correlators in the updated ($\omega_{\max} = 1.4005$) mock dataset at each timeslice for $N_\tau = 32$ and 128 for S-wave (a, c) and P-wave (b, d) states. Solid and dashed lines represent the same as Fig. 5.1. 72

5.4 Distribution of the natural logarithm of correlators in the updated ($\omega_{\max} = 1.4005$) mock dataset at $n_\tau = 7, 31$, and 127 for S-wave (a, c, e) and P-wave (b, d, f) states. Solid lines represent the same as Fig. 5.2 73

5.5 Same as Fig. 5.3 but for improved dataset. 75

5.6 Same as Fig. 5.4 but for improved dataset. 76

5.7 Same as Fig. 5.4 but for dataset generated from alternative parameter dataset. 78

5.8 Same as Fig. 5.4 but for dataset generated from alternative parameter dataset. 79

5.9 The first (a), twentieth (b), and eightieth (c) basis functions for the SVD representation for ω range $\omega_{\max} - \omega_{\min} = 3.72$ and $\omega_{\max} - \omega_{\min} = 1.4$ 81

5.10 Box and whisker plots for the coefficients of the original dataset with $\omega_{\max} - \omega_{\min} = 3.72$ (a) and the dataset described in Sec. 5.3.1 before improvement. 82

5.11 Same as Fig. 5.4 but for dataset generated from coefficients. 83

5.12	Same as Fig. 5.4 but for dataset generated from coefficients.	84
5.13	Relative uncertainty for Υ (a) and χ_{b1} (b) for $N_\tau = 128$	86
6.1	Colourmaps of \bar{R}^2 for models 1 (a) and 2 (b). The green square marks the position of optimal hyperparameters.	89
6.2	Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b). The green square marks the position of optimal hyperparameters.	91
6.3	Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b) with scaled target data. The green square marks the position of optimal hyperparameters.	93
6.4	Colourmaps of \bar{R}_{test}^2 for models 3 (a) and 4 (b) with scaled correlators and target data. The green square marks the position of optimal hyperparameters.	95
6.5	Training and testing \bar{R}^2 as a function training data size for Model 3 (a) and Model 4 (b).	96
6.6	Colourmaps of \bar{R}_{train}^2 for models 3 (a) and 4 (b) with scaled correlators and target data. The green square marks the position of optimal hyperparameters.	98
6.7	Schematic of 5-fold cross-validation. Each row represent a different iteration. Blue blocks represent the testing set and white space of each row represent the training data.	99
6.8	Individual R^2 for the training and testing data models 3 (a) and 4 (b) using 18000 training functions and the hyperparameters determined using 5-fold cross-validation.	100
6.9	Prediction of bottomonium groundstate masses (a) and widths (b), with uncertainties. Dashed lines of the same colour in (a) represent the experimental masses for the bottomonium states.	106
6.10	Predicted spectral functions using model 3 for the bottomonium S-wave states: η_b (a), Υ (b).	107
6.11	Predicted spectral functions using model 3 for the bottomonium P-wave states: χ_{b0} (a), and χ_{b1} (b).	108
6.12	Predicted spectral functions using model 4 for Υ (a) and χ_{b1} (b).	109

List of Tables

1.1	The temporal extent, corresponding temperatures in MeV, and number of configurations for the Generation 2L ensembles of the FASTSUM collaboration	11
1.2	The estimated pseudocritical temperatures from the renormalised chiral condensate and chiral susceptibility. The second error for the chiral susceptibility is an estimate of the systematic uncertainty. . . .	11
2.1	Bottomonium interpolators with associated quantum numbers and relevant gamma matrices.	13
3.1	Bottomonium spectrum with effective masses in lattice units and MeV. Experimental values were taken from the Particle Data Group [61]	28
3.2	Masses in GeV for each bottomonium state using 1, 2, 3, 4, and 5 terms. The index j refers the term in the summation in Eq. (3.2.8). All values were determined using the CORFITTER package.	30
3.3	Percentage error between reconstructed correlators and FASTSUM Gen2L correlators for $N_p = 1 - 5$	32
4.1	Mean Absolute Error on each model, determined using the noisy target data, y , and true values, $\sin(x)$	43
4.2	MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed. . .	52
4.3	MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed. . .	53
4.4	MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed. . .	55
4.5	MSEs for training and testing sets, alongside the lowest, median, and highest MSEs of the testing set for the hyperparameters displayed. . .	59
6.1	\bar{R}^2 and MSEs for training and testing sets for models 1, 2, 3, and 4. .	90

6.2	\bar{R}^2 and MSEs for training and testing sets for models 3 and 4 with scaled target data.	92
6.3	\bar{R}^2 and MSEs for training and testing sets for models 3 and 4 for scaled target data and scaled correlators which are input data.	94
6.4	\bar{R}^2 and MSEs for training and testing sets for models 3 and 4 after 5-fold cross-validation.	97
6.5	Optimal hyperparameters for models 3 and 4 for both S-wave and P-wave datasets. Provided for reproducibility purposes.	101
6.6	Predicted groundstate mass and width for η_b for the range of temperatures using model 3.	102
6.7	Predicted groundstate mass and width for Υ for the range of temperatures using model 3.	102
6.8	Predicted groundstate mass and width for χ_{b0} for the range of temperatures using model 3.	103
6.9	Predicted groundstate mass and width for χ_{b1} for the range of temperatures using model 3.	103

Bibliography

- [1] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Physics Letters B* **47**, 365 (1973).
- [2] K. G. Wilson, *Phys. Rev. D* **10**, 2445 (1974).
- [3] R. Feynman, A. Hibbs, and D. Styer, *Quantum Mechanics and Path Integrals* (Dover Publications, 2010).
- [4] A. Bzdak *et al.*, *Phys. Rept.* **853**, 1 (2020), 1906.00936.
- [5] STAR collaboration, M. M. Aggarwal *et al.*, (2010), 1007.2613.
- [6] T. K. Nayak, *J. Phys. Conf. Ser.* **1602**, 012003 (2020), 2008.04643.
- [7] P. de Forcrand and O. Philipsen, *Phys. Rev. Lett.* **105**, 152001 (2010), 1004.3144.
- [8] G. Aarts, F. Attanasio, B. Jäger, and D. Sexty, *JHEP* **09**, 087 (2016), 1606.05561.
- [9] S. Ejiri, *Phys. Rev. D* **78**, 074507 (2008).
- [10] Y. Aoki, Z. Fodor, S. D. Katz, and K. K. Szabo, *JHEP* **01**, 089 (2006), hep-lat/0510084.
- [11] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, *Nature* **443**, 675 (2006), hep-lat/0611014.
- [12] Y. Aoki *et al.*, *JHEP* **06**, 088 (2009), 0903.4155.
- [13] Wuppertal-Budapest, S. Borsanyi *et al.*, *JHEP* **09**, 073 (2010), 1005.3508.
- [14] G. Aarts *et al.*, (2020), 2007.04188.
- [15] RBC, HotQCD, F. Karsch, *J. Phys. G* **35**, 104096 (2008), 0804.4148.
- [16] C. Gattringer and C. B. Lang, *Quantum chromodynamics on the lattice* (Springer, Berlin, Heidelberg, 2010).
- [17] A. M. Polyakov, *Phys. Lett. B* **59**, 79 (1975).

- [18] A. M. Polyakov, Phys. Lett. B **72**, 477 (1978).
- [19] L. Susskind, Phys. Rev. D **20**, 2610 (1979).
- [20] C. Bonati, P. de Forcrand, M. D’Elia, O. Philipsen, and F. Sanfilippo, PoS **LATTICE2011**, 189 (2011), 1201.2769.
- [21] T. Matsui and H. Satz, Phys. Lett. B **178**, 416 (1986).
- [22] NA60, R. Arnaldi, Nucl. Phys. A **830**, 345C (2009), 0907.5004.
- [23] N. Brambilla *et al.*, Eur. Phys. J. C **71**, 1534 (2011), 1010.5827.
- [24] H. Satz, Nucl. Phys. A **783**, 249 (2007), hep-ph/0609197.
- [25] R. Rapp, D. Blaschke, and P. Crochet, Prog. Part. Nucl. Phys. **65**, 209 (2010), 0807.2470.
- [26] A. Mocsy, Eur. Phys. J. C **61**, 705 (2009), 0811.0337.
- [27] CMS, S. Chatrchyan *et al.*, Phys. Rev. Lett. **109**, 222301 (2012), 1208.2826, [Erratum: Phys.Rev.Lett. 120, 199903 (2018)].
- [28] M. Strickland, Phys. Rev. Lett. **107**, 132301 (2011), 1106.2571.
- [29] F. Brezinski and G. Wolschin, Phys. Lett. B **707**, 534 (2012), 1109.0211.
- [30] FASTSUM collaboration, fastsum.gitlab.io/.
- [31] G. Aarts *et al.*, PoS **LATTICE2019**, 075 (2019), 1912.09827.
- [32] L. D. McLerran and T. Toimela, Phys. Rev. D **31**, 545 (1985).
- [33] E. Braaten, R. D. Pisarski, and T. C. Yuan, Phys. Rev. Lett. **64**, 2242 (1990).
- [34] P. Petreczky, J. Phys. G **39**, 093002 (2012), 1203.5320.
- [35] C. Allton *et al.*, J. Phys. Conf. Ser. **509**, 012015 (2014), 1310.5135.
- [36] G. Aarts and A. Nikolaev, Eur. Phys. J. A **57**, 118 (2021), 2008.12326.
- [37] O. Kaczmarek, Nucl. Phys. A **931**, 633 (2014), 1409.3724.
- [38] Y. Burnier and M. Laine, Eur. Phys. J. C **72**, 1902 (2012), 1201.1994.

- [39] S. Ejiri and T. Hatsuda, PoS **LAT2005**, 183 (2006), hep-lat/0509119.
- [40] M. Cheng *et al.*, Phys. Rev. D **77**, 014511 (2008), 0710.0354.
- [41] P. Petreczky, Eur. Phys. J. ST **155**, 123 (2008), 0711.2280.
- [42] M. L. Bellac, *Thermal Field Theory* (Cambridge University Press, 2011).
- [43] M. Laine and A. Vuorinen, **925** (2016), 1701.01554.
- [44] R. Kubo, Journal of the Physical Society of Japan **12**, 570 (1957), <https://doi.org/10.1143/JPSJ.12.570>.
- [45] P. C. Martin and J. Schwinger, Physical Review **115**, 1342 (1959).
- [46] N. Brambilla, A. Pineda, J. Soto, and A. Vairo, Rev. Mod. Phys. **77**, 1423 (2005), hep-ph/0410047.
- [47] N. Brambilla, M. A. Escobedo, J. Ghiglieri, J. Soto, and A. Vairo, JHEP **09**, 038 (2010), 1007.4156.
- [48] M. Laine, O. Philipsen, P. Romatschke, and M. Tassler, JHEP **03**, 054 (2007), hep-ph/0611300.
- [49] A. Beraudo, J. P. Blaizot, and C. Ratti, Nucl. Phys. A **806**, 312 (2008), 0712.4394.
- [50] Y. Burnier, M. Laine, and M. Vepsalainen, JHEP **01**, 043 (2008), 0711.1743.
- [51] G. T. Bodwin, E. Braaten, and G. P. Lepage, Phys. Rev. D **51**, 1125 (1995), hep-ph/9407339, [Erratum: Phys.Rev.D 55, 5853 (1997)].
- [52] W. Caswell and G. Lepage, Physics Letters B **167**, 437 (1986).
- [53] G. P. Lepage, L. Magnea, C. Nakhleh, U. Magnea, and K. Hornbostel, Phys. Rev. D **46**, 4052 (1992), hep-lat/9205007.
- [54] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
- [55] C. Quigg, AIP Conf. Proc. **424**, 173 (1998), hep-ph/9707493.
- [56] G. Aarts *et al.*, JHEP **07**, 097 (2014), 1402.6210.

- [57] M. Cardoso, N. Cardoso, and P. Bicudo, PoS **LAT2009**, 233 (2009), 0910.0133.
- [58] TUMQCD, N. Brambilla *et al.*, PoS **LATTICE2019**, 109 (2019), 1911.03290.
- [59] G. P. Lepage and P. B. Mackenzie, Phys. Rev. D **48**, 2250 (1993), hep-lat/9209022.
- [60] G. Aarts *et al.*, JHEP **11**, 103 (2011), 1109.4496.
- [61] P. D. Group, Phys. Rev. D **86**, 010001 (2012).
- [62] P. Lepage, CORRFITTER 8.1.1, <https://github.com/gplepage/corrfitter/releases/tag/v8.1.1>.
- [63] G. P. Lepage *et al.*, Nucl. Phys. B Proc. Suppl. **106**, 12 (2002), hep-lat/0110175.
- [64] K. Hornbostel *et al.*, Phys. Rev. D **85**, 031504 (2012), 1111.1363.
- [65] C. M. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, Phys. Rev. D **90**, 054506 (2014), 1406.2279.
- [66] L.-F. Arsenault, R. Neuberg, L. A. Hannah, and A. J. Millis, arXiv e-prints (2016), 1612.04895.
- [67] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).
- [68] P. Mehta *et al.*, **810**, 1 (2019), 1803.08823.
- [69] S. D. Hodges and P. G. Moore, Journal of the Royal Statistical Society. Series C (Applied Statistics) **21**, 185 (1972).
- [70] A. Neubauer, Inverse Problems **5**, 541 (1989).
- [71] A. Tikhonov, *Nonlinear Ill-Posed Problems* (Springer Netherlands, 2014).
- [72] R. Bryan, European Biophysics Journal **18**, 165 (1990).
- [73] K. Vu *et al.*, arXiv e-prints , arXiv:1501.03854 (2015), 1501.03854.
- [74] L. Kades *et al.*, Phys. Rev. D **102**, 096001 (2020), 1905.04305.

-
- [75] S. Y. Chen, H. T. Ding, F. Y. Liu, G. Papp, and C. B. Yang, (2021), 2110.13521.
- [76] J. Horak *et al.*, (2021), 2107.13464.
- [77] T. Spriggs *et al.*, (2021), 2112.04201.
- [78] T. Spriggs *et al.*, (2021), 2112.01599.
- [79] A. Francis and O. Kaczmarek, *Prog. Part. Nucl. Phys.* **67**, 212 (2012), 1112.4802.
- [80] B. B. Brandt, A. Francis, H. B. Meyer, and H. Wittig, *JHEP* **03**, 100 (2013), 1212.4200.
- [81] H.-T. Ding, O. Kaczmarek, S. Mukherjee, H. Ohno, and H. T. Shu, *Phys. Rev. D* **97**, 094503 (2018), 1712.03341.
- [82] G. Backus and F. Gilbert, *Geophysical Journal International* **16**, 169 (1968).
- [83] B. B. Brandt, A. Francis, B. Jäger, and H. B. Meyer, *Phys. Rev. D* **93**, 054510 (2016), 1512.07249.
- [84] T. Harris, H. B. Meyer, and D. Robaina, *PoS LATTICE2016*, 339 (2016), 1611.02499.
- [85] M. Hansen, A. Lupo, and N. Tantalo, *Phys. Rev. D* **99**, 094508 (2019), 1903.06476.
- [86] B. Page *et al.*, (2021), 2112.02075.
- [87] Y. Nakahara, M. Asakawa, and T. Hatsuda, *Phys. Rev. D* **60**, 091503 (1999), hep-lat/9905034.
- [88] T. Umeda, K. Nomura, and H. Matsufuru, *Eur. Phys. J. C* **39S1**, 9 (2005), hep-lat/0211003.
- [89] A. Jakovac, P. Petreczky, K. Petrov, and A. Velytsky, *Phys. Rev. D* **75**, 014506 (2007), hep-lat/0611017.
- [90] H. T. Ding *et al.*, *Phys. Rev. D* **83**, 034504 (2011), 1012.4963.

- [91] A. Amato *et al.*, Phys. Rev. Lett. **111**, 172001 (2013), 1307.6763.
- [92] G. Aarts, C. Allton, J. Foley, S. Hands, and S. Kim, Phys. Rev. Lett. **99**, 022002 (2007), hep-lat/0703008.
- [93] G. Roepstorff, *Path Integral Approach to Quantum Physics* (Springer, Berlin, Heidelberg, 1994).