Reconstruction of bottomonium spectral functions in thermal QCD using Kernel Ridge Regression

Sam Offler,[a,*] **Gert Aarts,**[a,b] **Chris Allton,**[a] **Benjamin Jäger,**[c] **Seyong Kim,**[d] **Maria Paola Lombardo,**[e] **Ben Page,**[a] **Sinead M. Ryan,**[f] **Jon-Ivar Skullerud**[c] **and Thomas Spriggs**[a]

[a]*Department of Physics, Swansea University, Swansea SA2 8PP, United Kingdom*

[b]*European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT*) & Fondazione Bruno Kessler Strada delle Tabarelle 286, 38123 Villazzano (TN), Italy*

[c]*CP3-Origins & Danish IAS, Department of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense M, Denmark*

[d]*Department of Physics, Sejong University, Seoul 143-747, Korea*

[e]*INFN, Sezione di Firenze, 50019 Sesto Fiorentino (FI), Italy*

[f]*School of Mathematics, Trinity College, Dublin 2, Ireland*

[g]*Dept. of Theoretical Physics, National University of Ireland Maynooth, County Kildare, Ireland*

E-mail: s.p.offler.967106@swansea.ac.uk

We discuss results for bottomonium at nonzero temperature obtained using NRQCD on Fastsum *Generation 2L* ensembles, as part of the Fastsum collaboration's programme to determine the spectrum of the bottomonium system as a function of temperature using a variety of approaches. Here we give an update on results for spectral functions obtained using Kernel Ridge Regression. We pay in particular attention to the generation of training data and introduce the notion of using lattice QCD ensembles to learn how to improve the generation of training data. A practical implementation is given.

The 38th International Symposium on Lattice Field Theory, LATTICE 2021 26th-30th July, 2021
Zoom/Gather@Massachusetts Institute of Technology

---

*Speaker

1. Introduction

Since the initial proposal that suppression of J/ψ could be a signal for the formation of the quark-gluon plasma (QGP) [1], quarkonia, bound states of heavy quark-antiquark pairs, have been intensively studied. These states serve as important probes for the plasma created in heavy-ion collisions as they do not reach thermal equilibrium with the medium. Initial work focused on the charmonium system, but more recently the focus of both experimental [2] and theoretical (see e.g. Refs. [3, 4]) studies has shifted to bottomonium.

The purpose of this contribution is to continue the investigation of bottomonium at nonzero temperature by the FASTSUM collaboration [5–7], with the aim of determining masses and widths of ground and excited states, provided they exist, as well as the state-dependent melting or dissociation temperatures. In Ref. [8] Kernel Ridge Regression (KRR) was introduced to reconstruct bottomonium spectral functions from Euclidean lattice correlators. KRR is a machine learning method that requires training data and here we revisit intricacies of generating this data. In particular we introduce the notion of using lattice QCD ensembles to learn how to improve the generation of mock data and demonstrate how this may be implemented.

2. Lattice details

Given that the bottom quark mass scale M lies above any other scale in the problem, including the temperature T of the quark-gluon plasma, it can be integrated out to obtain standard non-relativistic QCD (NRQCD) [9, 10], also at nonzero temperature [11]. We use the same formulation as in Refs. [6, 7], i.e. terms up to $O(v^4)$ are included, where v is the heavy quark velocity in the bottomonium rest frame. In the relation between the bottomonium Euclidean correlator and the spectral function, terms involving M/T are exponentially suppressed, leading to a relation with a simplified ("$T = 0$") kernel,

$$G(\tau) = \int_{\omega_{\min}}^{\omega_{\max}} \frac{d\omega}{2\pi} K(\tau, \omega)\rho(\omega), \qquad K(\tau, \omega) = e^{-\omega\tau}, \qquad (1)$$

where it is understood that an additive constant $\sim 2M$ has to be added to the lower limit of the integral, to make contact with phenomenology in full QCD.

For the finite-temperature study, we use the anisotropic FASTSUM *Generation 2L* ensembles, with $N_f = 2 + 1$ flavours of clover-improved Wilson fermions, see Ref. [12]. The light quarks are heavier than in nature, but the strange quark mass is at its physical value. Relevant parameters are given in Table 1. In the fixed-scale approach, the temperature is varied by changing N_τ, using the relation $T = 1/(N_\tau a_\tau)$. The N_τ values and corresponding temperatures are shown in Table 2.

$1/a_\tau$ [GeV]	a_s [fm]	$\xi = a_s/a_\tau$	N_s	m_π [MeV]	$m_\pi L$	$T_{\rm pc}$ [MeV]
5.997(34)	0.01136(6)	3.453(6)	32	236(2)	4.36	164(2)

Table 1: Parameters relevant for the ensembles: a_s (a_τ) is the spatial (temporal) lattice spacing; ξ is the renormalised anisotropy; N_s is the number of points in the spatial direction; m_π is the mass of the pion; $T_{\rm pc}$ is the pseudocritical temperature, obtained from the inflection point of the renormalised chiral condensate [12].

2

N_τ	128	64	56	48	40	36	32	28	24	20	16
T [MeV]	47	94	107	125	150	167	187	214	250	300	375

Table 2: Temporal lattice sizes and temperatures for the FASTSUM Generation 2L ensembles [12].

3. Data generation

As is the case with many machine learning methods, Kernel Ridge Regression (KRR) requires training in order to make predictions. In the current context, this means it is necessary to generate a set of mock spectral functions and calculate the corresponding correlators according to Eq. (1), yielding the training set $\{\rho_i(\omega), G_i(\tau)\}$ ($i = 1, \ldots, N_{\text{train}}$). The importance of generating appropriate training data cannot be understated; it is only at this stage that we can incorporate any constraints/information we know of.

In this study we choose to construct mock spectral functions from the combination of $N_p = 5$ Gaussian peaks, such that a single spectral function is written as

$$\rho(\omega) = \sum_{p=1}^{N_p} Z_p \exp\left(-\frac{(\omega - m_p)^2}{\Gamma_p}\right). \tag{2}$$

Here Z_p is the amplitude, m_p is the position and Γ_p is related to the width of the Gaussian peak. Other mock functions can be constructed; in Ref. [8] the logarithm of the spectral function was expanded in an orthonormal set of incomplete basis functions. Continuing with the Gaussian peaks here, we generate a collection of peaks by sampling values for $\{Z_p, m_p, \Gamma_p\}$ from a set of distributions. Masses are chosen from the following exponential distribution,

$$m_{\text{GeV}} = 9 + \frac{1}{\zeta} e^{-\omega/\zeta}, \tag{3}$$

where ζ determines the decay rate of the distribution. These values are then converted to lattice units using

$$m_{\text{lat}} = a_\tau \left(m_{\text{GeV}} - \Delta M\right), \tag{4}$$

where $\Delta M = 7.465$ GeV is the additive constant mentioned below Eq. (1). As in Refs. [6, 7], it is determined by equating the calculated zero-temperature ground state mass in the Υ channel to its experimental value. Widths are determined by writing $\Gamma = 10^{-x}$ and selecting x from a uniform distribution, $0 < x < 6$. The amplitudes were selected for a uniform distribution, $0 < Z < 10$. We put $\rho(\omega) = 0$ for $\omega < \omega_{\text{min}}$ and $\omega > \omega_{\text{max}}$. Finally, the mock spectral functions (2) are normalised, using the relation at $\tau = 0$,

$$G(0) = \int_{\omega_{\text{min}}}^{\omega_{\text{max}}} \frac{d\omega}{2\pi} K(0, \omega)\rho(\omega) = \int_{\omega_{\text{min}}}^{\omega_{\text{max}}} \frac{d\omega}{2\pi} \rho(\omega), \tag{5}$$

where $G(0)$ is the source in the NRQCD formulation. In total we generated a set of 20 000 mock spectral functions, each consisting of $N_p = 5$ peaks and characterised by $3N_p = 15$ parameters. Only $N_{\text{train}} = 15 000$ functions from this dataset were used for training the KRR model.

To assess whether the training set is potentially relevant for the correlators obtained from the actual lattice QCD simulations, we carry out the following comparison. At each timeslice we
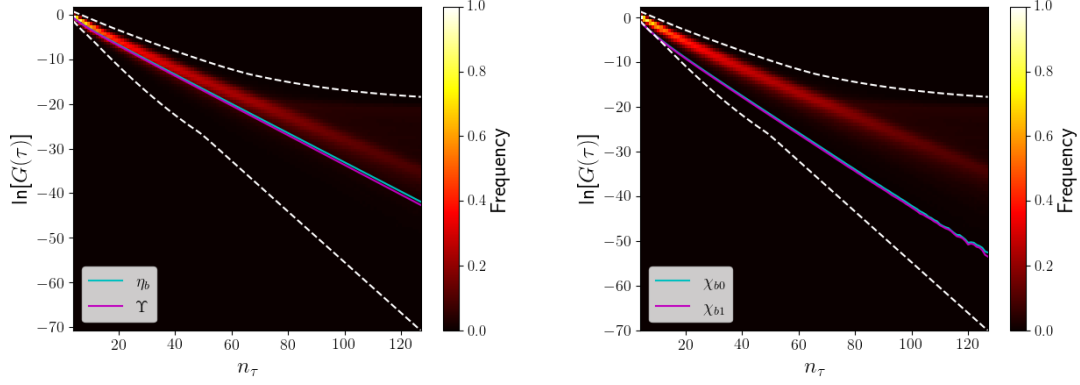
Figure 1: Distribution of the natural logarithm of correlators in the training set at each timeslice for $N_\tau = 128$ for S-wave (left) and P-wave (right) states. Dashed lines represent the upper and lower limits of the correlators in the training set for each value of n_τ. Solid lines represent the Euclidean correlators obtained in simulations. Note that the two channels in each figure – η_b, Υ (left) and χ_{b_0}, χ_{b_1} (right) – are hardly distinguishable.

determine the distribution of (the logarithm of) the correlators in the training set, as obtained from the mock spectral functions. Given that the mock set contains 20 000 spectral functions, the density quickly falls below 10%. This distribution is shown in Fig. 1 via a heat map. The dashed lines represent the upper and lower limits of the correlators in the training set for each value of n_τ. In addition we show the actual correlators obtained in the simulations. Both the left and the right panel in Fig. 1 show two correlators – η_b, Υ (left) and χ_{b_0}, χ_{b_1} (right) –, but note that these are hardly distinguishable.

The key observation is that the training data compares poorly to the actual correlators, i.e. there is little overlap between the most common values seen in the training set and the real data. This effect is worse in the P-wave channels. Since machine learning algorithms assume that training and validation data sets are drawn from the same ensembles as (or at least are representative of) the real data sets, this is not a desired feature. We remark that the spread of the training data itself is not a problem; it is beneficial for KRR to have access to outliers in principle.

In order to generate a better set of training data, we use the actual correlators to learn which members of the training set are the most representative. This is implemented as follows: from the original training set, a subset of correlators are chosen by determining which correlator $G_i(\tau)$ lies closest to the actual correlator $G_{\mathrm{NRQCD}}(\tau)$, at each value of the temperature and for S-waves (η_b, Υ) and P-waves (χ_{b_0}, χ_{b_1}) separately. Closeness is here defined via

$$i = \operatorname{argmin} \sum_{n_\tau=4}^{N_\tau-1} \left| \ln\left[\frac{G_i(n_\tau)}{G_{\mathrm{NRQCD}}(n_\tau)} \right] \right|. \tag{6}$$

Note that the sum starts at $n_\tau = 4$; the first few timeslices are not included in the summation. Given that there are ten different temperatures, see Table 2, and two S-wave or P-waves channels, this procedure yields a subset of up to 20 representative correlators (note that repeats are possible).

This subset is used to construct a larger, more representative, training set, as follows. For each correlator in the subset, the corresponding spectral function and, more importantly, the set of $3N_p$
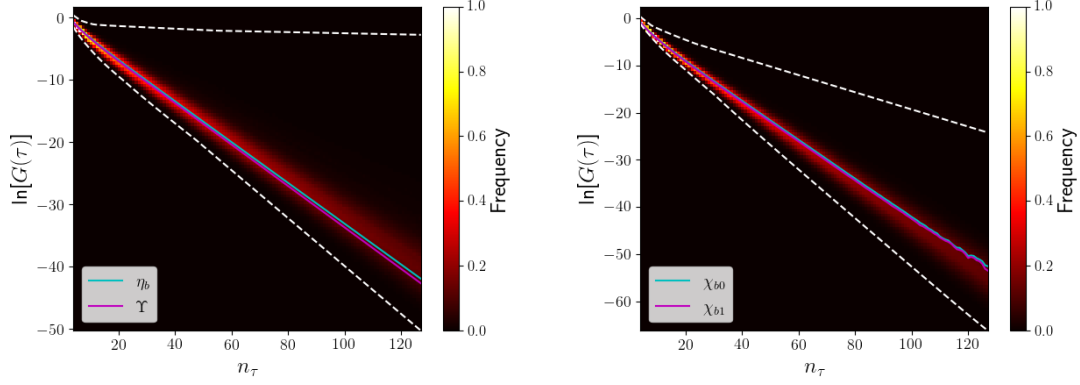
Figure 2: As above, for the improved training set.
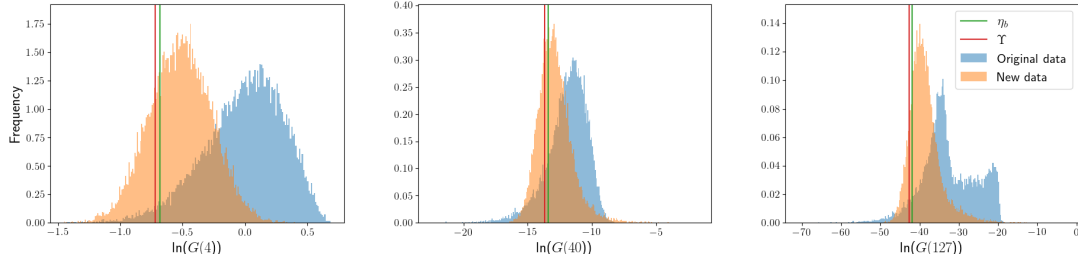
Figure 3: Distribution of the original and the improved training set for the correlators at $n_\tau = 4$ (left), 40 (centre), and 127 (right) for $N_\tau = 128$. Solid vertical lines represent the value of the actual Euclidean correlators at the specified timeslices.

parameters $\{Z_p, m_p, \Gamma_p\}$ are known. From those values the mean and variance is computed and new parameter values are sampled from Gaussian distributions with these means and variances. Note that for the amplitude Z and the parameter Γ this is done indirectly, using $\Gamma = 10^{-x}$ and an equivalent relation for Z, as above. From this point onwards the procedure is the same as earlier, but with different distributions for the parameters characterising the mock spectral functions.

Fig. 2 displays the same comparison as in Fig. 1, but now for the improved training set. As can be seen, the most common correlators in the training set now coincide with the actual correlators, leading potentially to a more appropriate training set. This is further demonstrated in Fig. 3, where the distribution of correlators obtained from the original and the improved training sets are shown at selected timeslices for the $N_\tau = 128$ lattice. The actual correlator values at the selected timeslices are shown with the vertical lines. We indeed observe a better overlap between the actual correlators and the data in the training set. The latter still has a nonnegligible width, allowing for flexibility in the application to spectral reconstruction.

4. Kernel Ridge Regression

Kernel Ridge Regression (see e.g. Ref. [13]) combines two techniques: the kernel method and ridge regression. Consider first linear regression in which $y = w^T \phi(x)$, where y are the target data,

$\phi(x)$ is a vector of functions of the input data x, and w is a vector of parameters to be determined. The cost function to be minimized is $E = \frac{1}{2}[y - w^T \phi(x)]^2$. In our application, the input data are the training set of Euclidean correlators $G_i(\tau)$ ($i = 1, \ldots, N_{\text{train}}$). However, rather than using these directly, they are used to generate an $N_{\text{train}} \times N_{\text{train}}$ matrix $\mathbf{C}$ (kernel function), with matrix elements

$$C_{ij} = \exp\left(-\gamma \sum_{n_\tau=4}^{N_\tau-1} \left[\widetilde{G}_i(n_\tau) - \widetilde{G}_j(n_\tau)\right]^2\right), \qquad \widetilde{G}_i(n_\tau) = \frac{G_i(n_\tau)}{\overline{G}(n_\tau)}. \qquad (7)$$

Note that the form of the kernel function is not unique. Here γ is a hyper-parameter, which sets a correlation length in the space of correlators. The correlator data appearing is normalised with $\overline{G}(n_\tau) = \sum_i G_i(n_\tau)/N_{\text{train}}$, the mean of the training data at each timeslice. This is done to account for the difference in absolute size of the correlators as τ changes.

The target data in our application are the spectral functions. To control the size of the target data, we use here expansion (2) in terms of Gaussian peaks, leading to $N_a \equiv 3N_p = 15$ parameters for spectral function. The target data is then represented by a $N_{\text{train}} \times N_a$ matrix $\mathbf{Y}$, in which each row contains the parameters of a single spectral function. It is now assumed that the input and target data are related as

$$\mathbf{Y} = \mathbf{C}\alpha, \qquad (8)$$

in which the $N_{\text{train}} \times N_a$ matrix α is the analogue of the vector w in linear regression. The aim of the training stage is to determine the matrix α.

To prevent overfitting, it is common to add an additional term in the cost function, proportional to the square of the parameters (ridge regression). Hence the cost function to be minimised reads

$$E(\mathbf{Y}, \mathbf{C}, \alpha) = \frac{1}{2}(\mathbf{Y} - \mathbf{C}\alpha)^2 + \frac{1}{2}\lambda \alpha^T \mathbf{C}\alpha, \qquad (9)$$

where λ is the second hyper-parameter, used to regularise the influence of the additional term. Minimising this cost function with respect to α then determines the optimal parameter matrix, for given σ and λ, as

$$\alpha_{\text{opt}} = (\mathbf{C} + \lambda \mathbf{I})^{-1} \mathbf{Y}. \qquad (10)$$

The hyper-parameters are determined via a cross-validation procedure, see below.

After this training stage, it is possible to make predictions for a spectral function ($\mathbf{Y}'$) given an actual Euclidean correlator using

$$\mathbf{Y}' = \mathbf{C}'\alpha_{\text{opt}}, \qquad (11)$$

where $\mathbf{C}'$ is determined from the squared rescaled difference between the actual correlator and the training correlators, i.e. it is a matrix of size $1 \times N_{\text{train}}$.

5. Optimization

To determine the optimal choice of hyper-parameters, we consider the so-called R^2 score, the coefficient of determination, defined in general as [14]

$$R^2(y) = 1 - \frac{\sum_i (y_i^{\text{true}} - y_i^{\text{pred}})^2}{\sum_i (y_i^{\text{true}} - \overline{y}^{\text{true}})^2}. \qquad (12)$$
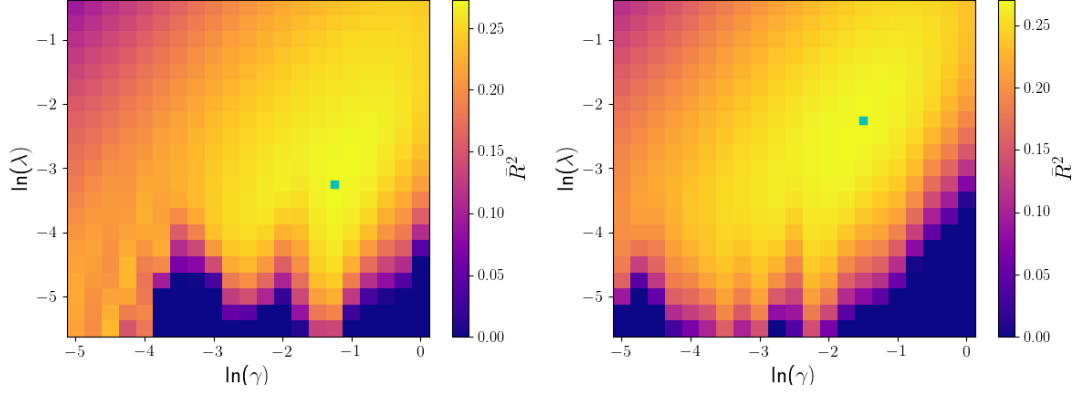
Figure 4: Heatmap of $\overline{R}^2$ used for cross-validation in the plane of hyper-parameters (λ, γ), for $N_\tau = 20$ (left) and $N_\tau = 128$ (right). The maxima, indicated with the little squares, provide the optimal choice of hyper-parameters. In cases where $\overline{R}^2 <0$, it has manually been set equal to 0 for clarity.

In our application, y is one of the $N_a = 15$ parameters parametrising a spectral function and we consider the mean R^2 score, averaged over the parameters,

$$\overline{R}^2 = \frac{1}{N_a} \sum_{a=1}^{N_a} R^2(y_a).$$ (13)

Rather than using a training and a validation set, we use cross-validation [13] in which a single set is used for both training and testing. The optimisation procedure seeks to find the value of $\overline{R}^2$ closest to 1. Results for $\overline{R}^2$ are shown in Fig. 4, for all the hyper-parameters pairs tested, and for two ensembles, $N_\tau = 20$ (left) and $N_\tau = 128$ (right). Though the plots are similar, there does appear to be some thermal dependence. In both cases there is a broad region of hyper-parameter pairs that perform to a similar standard. The most optimal choice of hyper-parameters is shown by the little squares, with $(\ln(\gamma), \ln(\lambda)) = $ (-1.25, -3.25) (left) and $(\ln(\gamma), \ln(\lambda)) = $ (-1.5, -2.25) (right). It is noted that the optimal values of $\overline{R}^2$ are rather low, i.e. not close to 1. Looking at the R^2 scores for each of the N_a parameters individually, we note that the R^2 score for the ground state mass is $R^2_{m_0} = 0.869$, whereas the R^2 scores for the groundstate width and amplitude are significantly lower, 0.148 and 0.336 respectively. We may therefore further improve the procedure by treating the various parameters on unequal footing.

As a preliminary result, Fig. 5 finally presents the predicted spectral functions in the Υ channel at four different temperatures. It is noted that the masses of both the ground and first excited state are smaller than the expected values. This may be due to the choice of expressing the spectral functions in terms of $N_p = 5$ Gaussian peaks and requires further investigation. Returning to the representation described in [8] may improve this. For the application of alternative methods to the same data, see Refs. [15, 16], also presented at this conference.
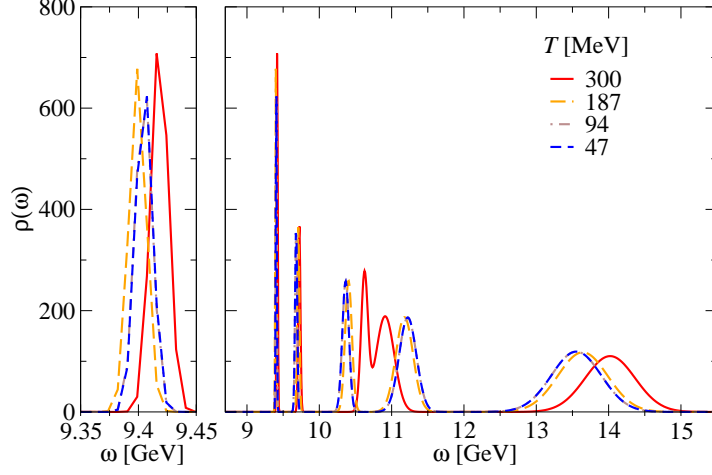
Figure 5: Spectral functions in the Υ channel at 4 different temperatures, obtained using Kernel Ridge Regression (KRR). The left pane zooms in on the ground state peak.

6. Summary

The main message of this contribution is that it is possible to adapt the training set by comparing its predictions to the actual correlator data from lattice QCD simulations, at the level of the correlators. A large discrepancy would suggest that the training sets are not representative of the real data sets, an undesired feature. It is noted that improving the overlap is necessary, but not sufficient. As a general remark, we note that this method can be iterated in principle and is applicable to other parametrisations of training spectral functions as well. The parametrisation used here – spectral functions are written as sums of Gaussian peaks – may well be an important limiting factor.

Acknowledgments

Southern Denmark for use of computing facilities. We are grateful to the Hadron Spectrum Collaboration for the use of their zero temperature ensemble.

References

[1] T. Matsui and H. Satz, Phys. Lett. B **178** (1986), 416-422.

[2] S. Chatrchyan *et al.* [CMS], Phys. Rev. Lett. **109** (2012), 222301 [arXiv:1208.2826 [nucl-ex]].

[3] R. Rapp, D. Blaschke and P. Crochet, Prog. Part. Nucl. Phys. **65** (2010), 209-266 [arXiv:0807.2470 [hep-ph]].

[4] R. Larsen, S. Meinel, S. Mukherjee and P. Petreczky, Phys. Lett. B **800** (2020), 135119 [arXiv:1910.07374 [hep-lat]].

[5] G. Aarts, S. Kim, M. P. Lombardo, M. B. Oktay, S. M. Ryan, D. K. Sinclair and J. I. Skullerud, Phys. Rev. Lett. **106** (2011), 061602 [arXiv:1010.3725 [hep-lat]].

[6] G. Aarts, C. Allton, S. Kim, M. P. Lombardo, M. B. Oktay, S. M. Ryan, D. K. Sinclair and J. I. Skullerud, JHEP **11** (2011), 103 [arXiv:1109.4496 [hep-lat]].

[7] G. Aarts, C. Allton, T. Harris, S. Kim, M. P. Lombardo, S. M. Ryan and J. I. Skullerud, JHEP **07** (2014), 097 [arXiv:1402.6210 [hep-lat]].

[8] S. Offler, G. Aarts, C. Allton, J. Glesaaen, B. Jäger, S. Kim, M. P. Lombardo, S. M. Ryan and J. I. Skullerud, PoS **LATTICE2019** (2019), 076 [arXiv:1912.12900 [hep-lat]].

[9] G. P. Lepage, L. Magnea, C. Nakhleh, U. Magnea and K. Hornbostel, Phys. Rev. D **46** (1992), 4052-4067 [arXiv:hep-lat/9205007 [hep-lat]].

[10] G. T. Bodwin, E. Braaten and G. P. Lepage, Phys. Rev. D **51** (1995), 1125-1171 [arXiv:hep-ph/9407339 [hep-ph]].

[11] Y. Burnier, M. Laine and M. Vepsalainen, JHEP **01** (2008), 043 [arXiv:0711.1743 [hep-ph]].

[12] G. Aarts, C. Allton, J. Glesaaen, S. Hands, B. Jäger, S. Kim, M. P. Lombardo, A. A. Nikolaev, S. M. Ryan and J. I. Skullerud, *et al.* [arXiv:2007.04188 [hep-lat]].

[13] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York Inc.

[14] R^2 score, scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html

[15] T. Spriggs *et al.*, PoS **LATTICE2021** 077.

[16] B. Page *et al.*, PoS **LATTICE2021** 134.