Syddansk Universitet

Alternatives to the stochastic "noise vector" approach

Forcrand, Philippe de; Jäger, Benjamin

Alternatives to the stochastic "noise vector" approach

Philippe de Forcrand[1,2,★] and *Benjamin* Jäger[1,3]

[1]*Institute for Theoretical Physics, ETH, CH-8093 Zürich, Switzerland*
[2]*CERN, Theory Division, CH-1211 Geneva, Switzerland*
[3]*CP3-Origins & Danish Institute for Advanced Study, Department of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense M, Denmark*

Abstract. Several important observables, like the quark condensate and the Taylor coefficients of the expansion of the QCD pressure with respect to the chemical potential, are based on the trace of the inverse Dirac operator and of its powers. Such traces are traditionally estimated with "noise vectors" sandwiching the operator. We explore alternative approaches based on polynomial approximations of the inverse Dirac operator.

1 Introduction

Measuring observables on an ensemble of gauge field configurations can require computing resources similar to those required to generate the configurations. However, comparatively little effort has been devoted to optimizing the computing strategy for such measurements. An important category of fermionic observables can be expressed as the trace of some function of the Dirac operator, or some other related operator. For example, the quark condensate $\bar{\psi}\psi$ is obtained from the trace of the inverse of the Dirac operator $\slashed{D}$. Generically, the trace for a given gauge field configuration is estimated by using "noise vectors" $\eta_k, k = 1, .., n$, via

$$\mathrm{Tr}A \approx \frac{1}{n} \sum_{k=1}^{n} \eta_k^\dagger A \eta_k \qquad (1)$$

with $\langle \eta_k^{i\dagger} \eta_k^j \rangle = \delta^{ij}$. Here, i and j label components of a noise vector η, and $\langle .. \rangle$ means averaging over noise vectors (for a given gauge field). Clearly, there are two sources of noise here: the *stochastic noise* coming from the limited number n of noise vectors used on a given gauge field configuration; and the *gauge noise* coming from the fluctuations caused by the variation of the gauge field from configuration to configuration. The balance between these two types of noise depends on the observable, and on the parameters of the Monte Carlo simulation: in particular, fluctuations coming from the gauge field will severely increase in the vicinity of a phase transition.

Here, we consider two situations of relevance at finite temperature:
(*i*) the measurement of the chiral condensate and its first 4 moments near the finite-temperature deconfinement transition. The moments are necessary to construct the Binder cumulant (or kurtosis), whose value at the phase transition is known by universality;

★e-mail: forcrand@phys.ethz.ch

(*ii*) the measurement of the Taylor coefficients of the expansion of the pressure in μ/T.
In the latter situation, it is not unusual to take 1000 noise vectors or more per gauge configuration. Can one be more efficient?

2 Related recent progress

Applied mathematicians have proposed a novel approach to determine the spectrum of a Hermitian matrix A [1]. Instead of obtaining the eigenvalues directly by a Krylov-based method, the idea is to determine the "eigenvalue count", namely the *number* $\mu(a, b)$ of eigenvalues in some interval $[a, b]$. Upon varying a and b, the positions of the jumps in the eigenvalue count give the desired eigenvalues. Thus, one needs a sliding bandpass filter $f(x, a, b)$ which is $= 1$ if $x \in [a, b]$, 0 otherwise. In practice:
- The function f is approximated by Chebyshev polynomials and applied to the matrix A. Then, the "eigenvalue count" $\mu(a, b)$ is Tr $f(A, a, b)$.
- The trace is approximated by using a small number of noise vectors following eq.(1).

One great advantage is that, by varying the coefficients of the polynomial, a and b can be varied at will without any additional matrix-vector multiplication, and thereby the *complete* spectrum of A can be determined at low cost.

Two lattice groups have tested this approach so far:
• In [2], a degree-8000 Chebyshev approximation was constructed and applied to $\slashed{D}^\dagger \slashed{D}$, using 20 noise vectors for the trace. This required 80000 Dirac-matrix-vector multiplications. These are very moderate resources, considering that the lattice size was $64^3 \times 96$, and that the whole spectrum could be extracted. After averaging over 10 gauge configurations, a precise determination of $\langle \bar{\psi}\psi \rangle$ could be obtained.
• In [3], an overlap Dirac operator was used, on the same lattice size, with a Chebyshev approximation of the same degree 8000, and 1 noise vector per configuration only. The gauge ensemble consisted of 50 configurations. Again, a precise determination of $\langle \bar{\psi}\psi \rangle$ could be obtained.

These promising first results were obtained at zero temperature. We are interested in applying a similar approach at $T \sim T_c$, the deconfinement temperature, where the "gauge noise" is expected to be much greater. And we want, in addition to $\bar{\psi}\psi = \sum_i \lambda_i^{-1}$, higher negative moments of the eigenvalues, i.e. $\sum_i \lambda_i^{-m}, m = 1, 2, 3, 4, ...$

3 Application I: Binder cumulant for the finite-temperature transition

A suitable order parameter for the finite-T transition is the quark condensate $\bar{\psi}\psi$. To probe the order of the phase transition, it is customary to measure its Binder cumulant (or equivalently its kurtosis)

$$B_4(\bar{\psi}\psi) \equiv \frac{\langle (\delta\bar{\psi}\psi)^4 \rangle}{\langle (\delta\bar{\psi}\psi)^2 \rangle^2}, \quad \delta\bar{\psi}\psi = \bar{\psi}\psi - \langle \bar{\psi}\psi \rangle \tag{2}$$

Thus, the first 4 moments $(\bar{\psi}\psi)^k, k = 1, .., 4$ must be measured, or estimated in an unbiased way. The latter approach is normally chosen, using noise vectors. One must then decide how many noise vectors per configuration to use: using too few will increase the "stochastic noise" on B_4 and the associated statistical error; using too many is wasteful of computing resources.
Here, we chose the most economical approach: 4 noise vectors $\eta_1, .., \eta_4$, forming unbiased estimators
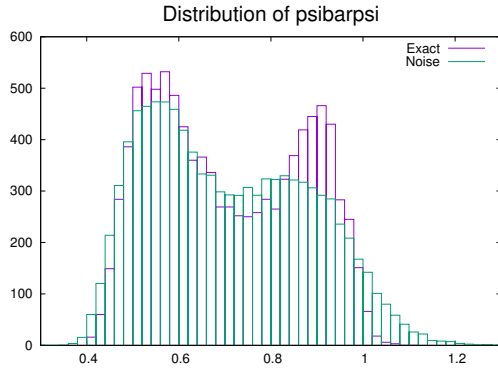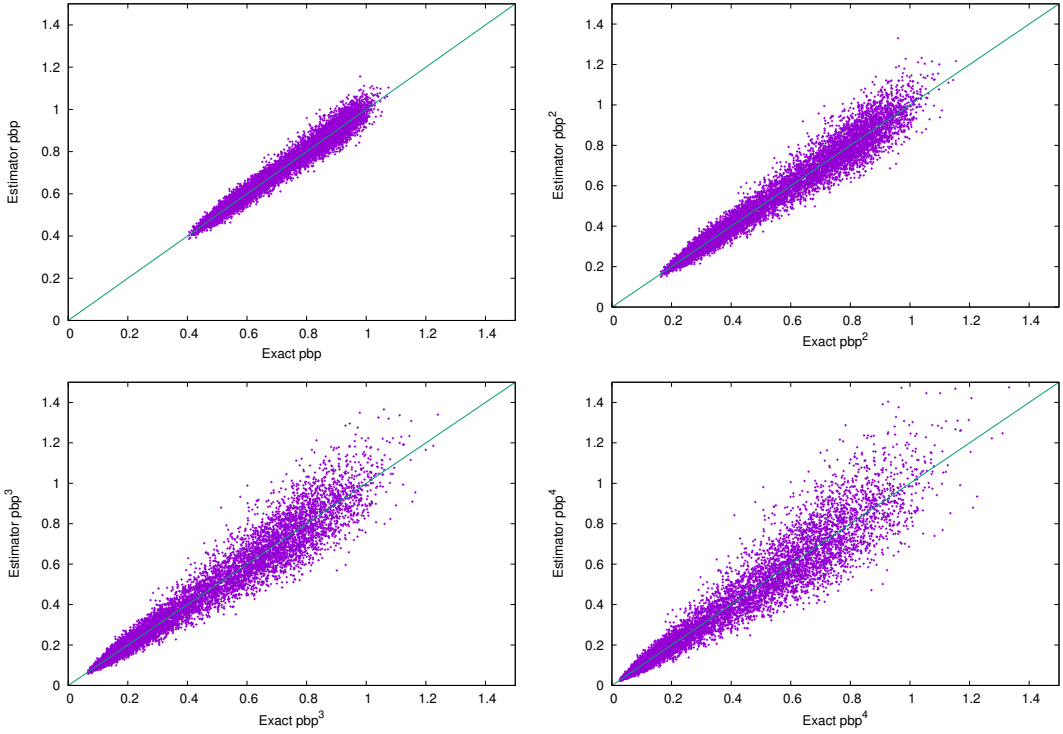
Figure 1. Comparison of the first 4 powers of $\bar{\psi}\psi$ with their stochastic estimators eq.(3), on an ensemble of $6^3 \times 4$ SU(3), $N_f = 4$ configurations with staggered fermions, near the finite-temperature deconfinement transition. The bottom figure compares the resulting distribution of the exactly measured $\bar{\psi}\psi$ with that of its estimator: the two-bump distribution signals the vicinity of the finite-T transition; the "stochastic noise" barely increases the variance of the distribution.

as

$$\bar{\psi}\psi \quad \longleftarrow \quad \frac{1}{4}\left(\eta_1^\dagger D\!\!\!/^{-1}\eta_1 + \eta_2^\dagger D\!\!\!/^{-1}\eta_2 + \dots\right)$$

$$(\bar{\psi}\psi)^2 \quad \longleftarrow \quad \frac{1}{6}\left(\eta_1^\dagger D\!\!\!/^{-1}\eta_1 \cdot \eta_2^\dagger D\!\!\!/^{-1}\eta_2 + \dots\right)$$

$$(\bar{\psi}\psi)^3 \quad \longleftarrow \quad \frac{1}{4}\left(\eta_1^\dagger D\!\!\!/^{-1}\eta_1 \cdot \eta_2^\dagger D\!\!\!/^{-1}\eta_2 \cdot \eta_3^\dagger D\!\!\!/^{-1}\eta_3 + \dots\right)$$

$$(\bar{\psi}\psi)^4 \quad \longleftarrow \quad \left(\eta_1^\dagger D\!\!\!/^{-1}\eta_1 \cdot \eta_2^\dagger D\!\!\!/^{-1}\eta_2 + \eta_3^\dagger D\!\!\!/^{-1}\eta_3 \cdot \eta_4^\dagger D\!\!\!/^{-1}\eta_4\right) \tag{3}$$

In Fig. 1, we compare the estimators above with the exact values obtained by diagonalization of $D\!\!\!/$, for an ensemble of $6^3 \times 4$ SU(3) configurations, where the parameters ($N_f = 4$, $am = 0.07$) have been chosen to be in the vicinity of a sharp temperature-driven crossover. Indeed, the exact distribution of $\bar{\psi}\psi$ in the last panel shows a clear two-peak distribution. What is striking is that the distribution of the stochastic estimator of $\bar{\psi}\psi$, in the same figure, is hardly broader, even though the number of noise vectors has been kept to its minimum value 4. Thus, in this example, the "stochastic noise" is negligible compared to the "gauge noise". Whether or not this persists on larger lattices needs to be checked.
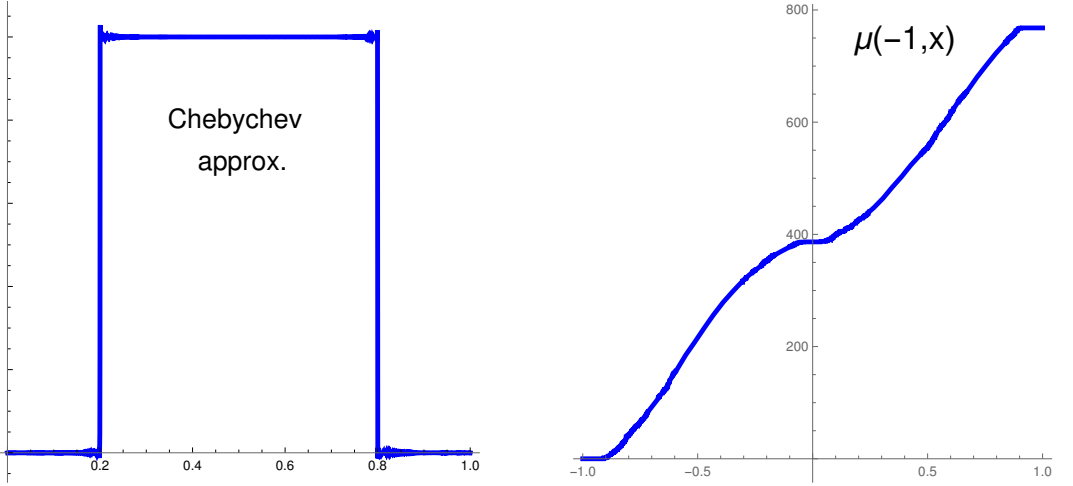
Figure 2. Chebyshev bandpass filter (*left*) and eigenvalue count $\mu(-1, x)$ as a function of x, for the rescaled, staggered operator $(iD\!\!\!/)$ on a 4^4 lattice.

The novel method outlined in Sec. 2 can give us nearly exact results at a potentially reduced cost, compared with exact diagonalization. A Chebyshev bandpass filter is applied to $(iD\!\!\!/)$, so that the number μ of eigenvalues in $[a, b]$ can be approximated as

$$\mu(a, b) = \frac{1}{n}\sum_{j=1}^{n}\sum_{p=0}^{p_{\max}} g_p(a, b)\,\langle\langle\eta_j^\dagger T_p(iD\!\!\!/)\,\eta_j\rangle\rangle \tag{4}$$

where T_p is the Chebyshev polynomial of degree p, g_p are the coefficients of the filter approximation, and $\langle\langle\cdots\rangle\rangle$ represents the averaging over the gauge field. Fig. 2 illustrates the bandpass filter (*left*) and

the eigenvalue count $\mu(-1, x)$ versus x (*right*) [1], for one 4^4 configuration ($\not{D}$ has been rescaled so that its spectrum is contained in $[-1, +1]$). $\mu(-1, x)$ is a piecewise constant function, increasing by one as x passes each eigenvalue. We approximate the eigenvalues as solutions of $\mu(-1, x)$ being half-integer. Since all eigenvalues are obtained in this way, we can form arbitrary moments $\text{Tr}\left(\not{D}^{-k}\right), k = 1, 2, \ldots$. This is illustrated in Table 1 below. Note the remarkable accuracy, even for high moments.

k	exact	approx.	approx./exact
2	-453.767	-451.148	0.994229
4	3871.66	3844.25	0.99292
10	$-8.136 \cdot 10^7$	$-8.091 \cdot 10^7$	0.994547

Table 1. $\text{Tr}\left(\not{D}^{-k}\right)$ using 4 noise vectors and a degree-16k polynomial on a 4^4 lattice.

4 Application II: Taylor expansion of free energy

QCD at non-zero quark chemical potential μ represents a long-standing challenge for lattice simulations: the Dirac determinant becomes complex when $\mu \neq 0$, which prevents its interpretation as a sampling probability. Several strategies have been proposed to circumvent this "sign" (or phase) problem. One of the first is that of Taylor-expanding the QCD free energy, or the pressure P, in powers of (μ/T) around $\mu = 0$:

$$P(\mu, T) = P(\mu = 0, T) + \sum_{k=1} c_k(T) \left(\frac{\mu}{T}\right)^k. \tag{5}$$

Charge conjugation symmetry implies that odd-order coefficients c_k vanish. However, products of odd derivatives do contribute in the presence of multiple quark flavors, each with its chemical potential. Now, c_k is simply the kth derivative of the free energy evaluated at $\mu = 0$, so it can be measured in conventional, sign-problem free simulations. The μ-dependence appears only in the Dirac determinant, so that the generic expression for c_k is the trace of a polynomial expression in $\not{D}^{-1}$ and in derivatives $\partial^m \not{D}/\partial\mu^m$. The number of terms grows exponentially with the order k, with positive or negative signs. This makes the calculation of high-order coefficients c_k extremely challenging: no calculation has ever yielded information at 10th order. Improvements in efficiency are needed.

One such improvement has been proposed recently by Gavai and Sharma [5, 6]. The suggestion is to change the way the chemical potential is introduced on the lattice, and adopt a naive, linear prescription rather than the exponential prescription of Hasenfratz and Karsch [7]. This is rather counter-intuitive, since it re-introduces a UV divergence in the pressure. Gavai and Sharma argue that this divergence can be removed by a numerical subtraction. The gain is that derivatives $\partial^m \not{D}/\partial\mu^m$ now vanish for $m > 1$. This greatly simplifies the structure of the c_k's, which become the traces of homogeneous polynomials of degree k in A^{-1}, with $A = \not{D}\left(\frac{\partial \not{D}}{\partial\mu}\right)^{-1}$ (note that A is neither hermitian, nor anti-hermitian – see Fig. 3).

At this stage, our approach may become advantageous. Suppose we can determine the [complex] spectrum of A. Then, c_k will be obtained as a sum of products of factors $\text{Tr}(A^{-m})$, which can *all* be determined in one go!

---

[1] Note the similarity with Lüscher's "mode number" [4].
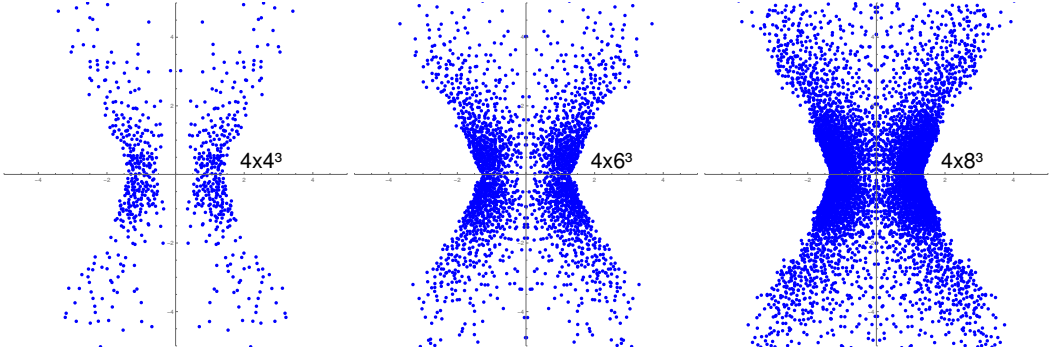
Figure 3. Complex spectrum of $\left(\not{D} \left(\frac{\partial \not{D}}{\partial \mu} \right)^{-1} \right)$, for lattices of increasing spatial size near the finite-T transition.

Here, we take first steps in that promising direction. First, we generalize the notion of eigenvalue count to the complex plane. The number of eigenvalues $\mu(\Gamma)$ inside a complex contour Γ can be obtained by a Cauchy integral [8]:

$$\mu(\Gamma) = \frac{1}{2\pi i} \oint_{\Gamma} dz \, \mathrm{Tr}\left[(z\mathbb{1} - A)^{-1} \right] = \frac{1}{2\pi i} \oint_{\Gamma} dz \sum_j \frac{1}{z - \lambda_j} \tag{6}$$

For Γ, we choose a circle of center c and radius r, and discretize the integral using the trapezoidal rule:

$$\mu(r, c) \sim \frac{1}{N} \sum_{k=0}^{N-1} r \, e^{\frac{2\pi i}{N} k} \, \mathrm{Tr}\left[(z_k \mathbb{1} - A)^{-1} \right] \text{ with } z_k = c + r \, e^{\frac{2\pi i}{N} k} \tag{7}$$

where the trace is estimated using noise vectors. The so-obtained value of $\mu(r, c)$ will be nearly an integer equal to the number of eigenvalues inside the circle. One can now discard the circles which contain zero eigenvalue, and divide by 2 the radius of non-empty circles. This procedure is repeated recursively, until each circle contains one and only one eigenvalue, and has a radius smaller than some prescribed tolerance ϵ. In this fashion, the complete spectrum of A is determined. Arbitrary moments $\mathrm{Tr}A^{-m}$ can be evaluated.

The crucial point is that the estimation of the trace in eq.(7) proceeds by solving linear systems $(z_k \mathbb{1} - A) x = \eta$, using a multi-shift linear solver. Then the cost of solving many such systems is not much greater than that of solving just one, because these systems differ only by the shift z_k, and the Krylov space built by the linear solver is shift-invariant. In addition, each iteration of the solver involves matrix-vector multiplication by $A = \not{D} \left(\frac{\partial \not{D}}{\partial \mu} \right)^{-1}$. At first sight, the inversion of $\frac{\partial \not{D}}{\partial \mu}$ looks daunting. But this operator is completely local in space, so that the corresponding matrix is made of small diagonal blocks, each of which is easy to invert.

We illustrate our approach in Figs. 3 and 4. Fig. 3 shows the complex spectrum of A, for 3 configurations obtained with the same action, near the finite-T transition, on lattices of increasing spatial sizes.

Fig. 4 shows 3 successive stages of our recursive procedure: at each stage, circles empty of eigenvalues are discarded, and non-empty circles are replaced by 4 circles with half the radius. The final output is an estimate of all eigenvalues of A, from which successive moments $\mathrm{Tr}\left(\not{D} \frac{\partial \not{D}}{\partial \mu}^{-1} \right)^{-m}$ can be estimated as well. Illustrative results are provided in Table 2 below.
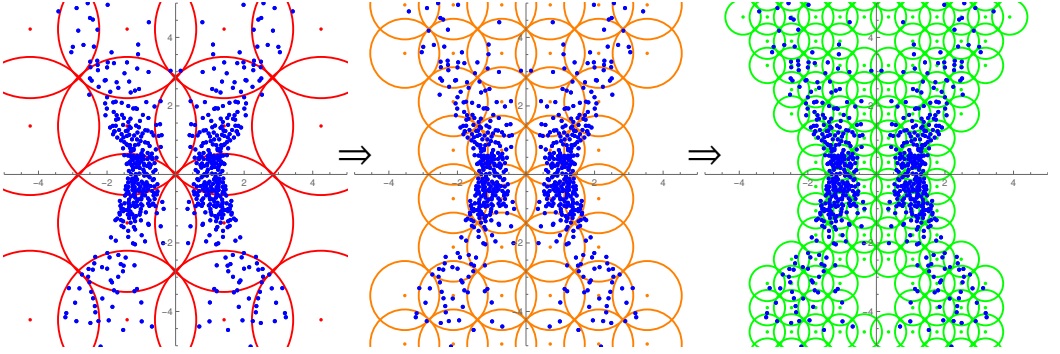
Figure 4. Circular contours are recursively refined until each contour encloses one eigenvalue (or a few).

m	exact	approx.	ratio
1	4.10306i	0.0352058 + 4.10474i	1.00041
2	142.497	142.607 − 0.00748385i	1.00077
3	10.8641i	0.04117 + 10.8765i	1.00114
4	−14.4443	−14.3171 − 0.0751784i	0.991195
10	−133.716	−135.113 − 3.93177i	1.01045
20	−79320.5	−81511.3 + 979.464i	1.02762

Table 2. $\mathrm{Tr}\,(A^{-m})$ for 100 quadrature points and 100 noise vectors for $\epsilon = 10^{-3}$, on a 4^4 gauge configuration.

5 Outlook

We have described a novel approach to obtain estimates of moments $\mathrm{Tr}\,(A^{-m})$ of a sparse matrix A, in the Hermitian and in the non-Hermitian case. Two applications were tested: measuring the Binder cumulant of the quark condensate ($A = i\slashed{D}$), and the coefficients of the Taylor expansion of the pressure in μ/T ($A = \slashed{D}\left(\frac{\partial \slashed{D}}{\partial \mu}\right)^{-1}$). In both cases, the regime of interest is near the finite-temperature deconfinement transition. In this regime, fluctuations of the gauge field are large, so that the "gauge noise" may well exceed the "stochastic noise" caused by the limited number of noise vectors used to estimate $\mathrm{Tr}\,(A^{-m})$. Indeed, this is what happens in the first application: it is then sufficient to use 4 noise vectors per configuration, at least for the small lattice sizes considered in this test.

In the second application, we have generalized the eigenvalue count to the complex plane. The viability of our approach rests on a massive usage of non-Hermitian multi-shift linear solvers. Great simplifications in the contour integrals are possible and are under study. However, while we obtain nearly exact results for the first 20 moments $\mathrm{Tr}\,(A^{-m})$, we have not measured the "gauge noise", which presumably increases quickly with m. This may represent the ultimate, irreducible origin of the computing cost of the Taylor expansion of the pressure, as foreseen in [9].

6 Acknowledgements

References

[1] E. Di Napoli, E. Polizzi, Y. Saad, Numerical Linear Algebra with Applications **23**, 674 (2016)

[2] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, C.H. Wong, PoS **LATTICE2015**, 310 (2016), `1605.08091`

[3] G. Cossu, H. Fukaya, S. Hashimoto, T. Kaneko, J.I. Noaki, PTEP **2016**, 093B06 (2016), `1607.01099`

[4] L. Giusti, M. Lüscher, JHEP **03**, 013 (2009), `0812.3638`

[5] R.V. Gavai, S. Sharma, Phys. Lett. **B749**, 8 (2015), `1406.0474`

[6] R.V. Gavai, S. Sharma, PoS **CPOD2014**, 015 (2015), `1503.03575`

[7] P. Hasenfratz, F. Karsch, Physics Letters B **125**, 308 (1983)

[8] Y. Futamura, H. Tadano, T. Sakurai, JSIAM Letters **2**, 127 (2010)

[9] P. de Forcrand, PoS **LAT2009**, 010 (2009), `1005.0539`