Machine-learning approaches to accelerating lattice simulations

Scott Lawrence[*]

Los Alamos National Laboratory Theoretical Division T-2,
Los Alamos, NM 87545, USA

E-mail: srlawrence@lanl.gov

The last decade has seen an explosive growth of interest in exploiting developments in machine learning to accelerate lattice QCD calculations. On the sampling side, generative models are a promising approach to mitigating critical slowing down and topological freezing. Meanwhile, signal-to-noise problems have been shown to be improvable by the use of optimized improved observables. Both techniques can be made free of bias, resulting in trustworthy but reduced statistical errors. This talk reviews recent developments in this field.

---

[*]Speaker

Contents

1. Introduction

Lattice Monte Carlo methods (particularly lattice QCD) allow nonperturbative, first-principles access to a wide range of observables in strongly coupled systems, but at considerable computational cost. Algorithmic improvements that reduce this cost, while preserving the merits of these Monte Carlo methods—particularly controllable systematic errors, predictable polynomial scaling, and access to nonperturbative physics—are valuable. Recent searches for such improvements have often been inspired by techniques from machine learning (ML) and particularly deep learning. To slightly over-generalize, one may describe an algorithmic improvement that depends on the existence of some function $f(x)$ such that a functional $L[f(x)] \approx 0$ is small. The function $f(x)$ is then parameterized and sought via ML techniques (typically an appropriate equivalent of stochastic gradient descent on the loss functional $L[\cdot]$). This approach has the virtue that the last decade's substantial progress in deep learning techniques can be brought to bear on lattice QCD with relatively little modification.

This talk is focused on those ML and ML-inspired methods which accelerate lattice simulations in a way that provably introduces no systematic bias, without significantly altering the overall framework of lattice QCD. The ML-inspired methods discussed in this talk can be divided into two rough families, depending on whether the method is intended to assist with sampling or with measurement. On the sampling side, there have been an array of proposals for learning optimized proposal distributions. The greatest successes in gauge theories have been seen by flow-based samplers, discussed in Section 2. On the measurement side, signal-to-noise problems (and, closely related, sign problems) can be alleviated by performing contour deformations as discussed

in Section 3, or by explicitly constructing observables with equal mean but reduced variance as discussed in Section 4. Finally, measurement can be made cheaper without great sacrifices to precision by learning approximate observables which may be quickly evaluated, while using a small number of configurations to remove the bias (see Section 5).

Of necessity a wide variety of techniques have been excluded from this talk, not because of a lack of merit but because they do not fit with the narrow framework described above. To conclude this section let us highlight a portion of this work. Various machine learning methods have been applied to regularize the ill-posed problem of determining spectral functions [1, 2, 3, 4, 5], relevant to determining real-time evolution, decay rates, and cross sections from Euclidean lattice data. In something of a call-back to the origins of deep learning of a field, phases or action parameters can be learned from samples (or a single sample) of field configurations [6, 7, 8, 9, 10, 11, 12]. Following the introduction of neural-network quantum states [13], neural networks have been used to represent wavefunctions of lattice systems (in the Hamiltonian formalism) [14, 15]. Finally, neural networks have recently been used to parameterize a many-parameter lattice action [16] with an eye towards minimizing lattice discretization effects.

2. Flow-based sampling

A substantial fraction of the computational effort of lattice QCD comes from the cost of obtaining a sequences of N samples U_n from the Boltzmann distribution $p(U) \propto e^{-S_{\text{eff}}(U)}$. The statistical errors on a Monte Carlo measurement scale as $\sigma \sim N^{-1/2}$, so assuming negligible systematics and cheap measurement[1], a method that accelerates sampling by a factor A will reduce error bars by a factor $\sqrt{A}$.

Thus we are motivated to consider the vast family of machine-learning methods for efficiently sampling from various probability distributions. The focus here will be on one particular class of such methods: normalizing flows (also highlighted in a plenary talk at the previous lattice conference [17]). Several other ML methods for accelerating sampling have been explored in recent years, including restricted Boltzmann machines [18, 19], diffusion models [20, 21], and generative adversarial networks [22, 23]. Our focus is on normalizing flows because normalizing flows have been tested directly on gauge theories—the performance of other methods on gauge models of size comparable to QCD lattice is generally unexplored.

Let us say we have a probability distribution on $\mathbb{R}$, $p(x)$, from which we wish to sample. A normalizing flow for this distribution is a function $x(z)$ that obeys:

$$p(x)dx = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz. \tag{1}$$

A normally distributed random variable z can be turned into a sample from p simply by applying the function $x(z)$. Normalizing flows exist for any distribution p, and can be obtained in one dimension by inverting the cumulative distribution function.

---

[1]Neither is a universally valid assumption: many calculations have errors dominated by systematics, and measurement is typically at least comparably expensive to sampling due to the need to invert the Dirac matrix.
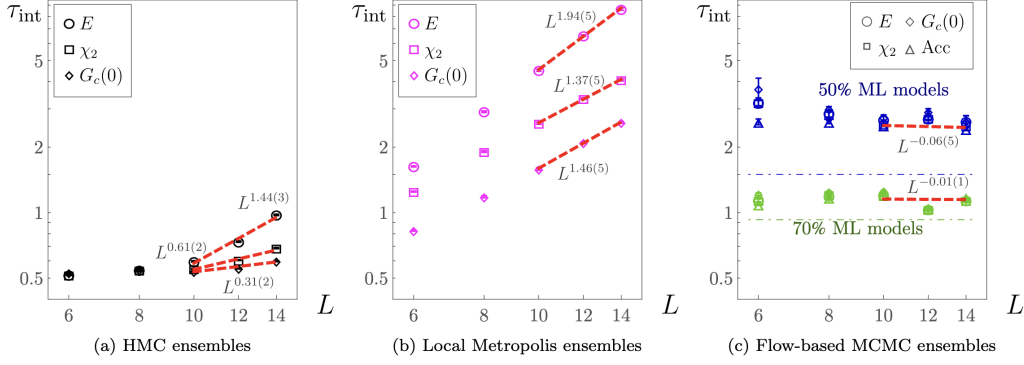
Figure 1: Comparison of the autocorrelation time as a function of lattice size when using HMC updates, local Metropolis updates, and normalizing flows. Reproduced from [25].

Normalizing flows naturally generalize to multivariate distributions, where the defining equation is

$$p[\phi(z)] \det \frac{\partial \phi}{\partial z} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{2}$$

An early published example of a normalizing flow is the Box-Müller transform [24], which converts between a uniform sample on $[0, 1]^2$ and two normally distributed random variables. In physics, normalizing flows first appeared under the name of *trivializing maps* as a method for solving field theories. Nicolai presented one as a solution to $\mathcal{N} = 4$ supersymmetric Yang-Mills [26], and three decades later Luscher discussed trivializing maps in the context of lattice gauge theory [27].

So far all of these normalizing flows have been *exact* normalizing flows, meaning that (2) holds as written. When attempting to learn a normalizing flow, one does not typically expect to be able to satisfy such an equality exactly. We instead search for flows $\phi(z)$ such that (2) is satisfied only approximately. Expectation values over the original distribution $p(\cdot)$ are then estimated either by using the normalizing flow to generate proposals in a Markov-chain Monte Carlo, or by reweighting. For good approximations, this is often still much more efficient than sampling from $p(\cdot)$ by other means.

Training and using a numerical normalizing flow involves repeated evaluations of the Jacobian determinant. For an arbitrary function ϕ this calculation scales nearly as the cube of the number of degrees of freedom, and is by far the most computationally intensive step. Therefore, practical normalizing flows use a restricted class of functions for which both ϕ and its Jacobian determinant can be evaluated efficiently—typically in time linear in the number of degrees of freedom. These families include NICE [28], RealNVP [29], and neural ordinary differential equations [30].

Continuous normalizing flows have been used to sample the Nambu-Goto string [31] and to construct normalizing flows the exactly respect the symmetries of the field theory [32]. However, most large-scale numerical works have used RealNVP.

A trained normalizing flow allows the partition function to be directly estimated [33, 34]. In the context of lattice quantum field theory, the partition function is typically not of physical interest, and so this property is not much used.

The use of RealNVP (and various extensions) has led to a series of remarkable results in accelerating lattice sampling. The first demonstration of this was in scalar field theory [25]. In

Reference	N_{params}	ESS at $\beta = 6$
[27]	8	< 1%
[41]	420	70%
[35]	$\sim 10^6$	48%

Table 1: Comparison of the performance of three different normalizing flows, as measured by effective sample size (ESS). Larger ESS corresponds to a higher-precision normalizing flow and a more efficient algorithm. Table reproduced from [41]

that work, the flow is used to generate proposals for a Markov-chain Monte Carlo ensemble, and so the success of the normalizing flow manifests as a short autocorrelation time. This is shown in Figure 1.

Applications to gauge theories followed from the design of gauge-equivariant flows [35, 36, 37], which are only capable of representing gauge-invariant probability distributions [2]. This has since been exted to theories with fermions [38], applied to the lattice Schwinger model [39], and demonstrated at arbitrary space-time dimension [40].

Most works in this line have adopted the usual perspective of deep learning, that more parameters leads to better results. At least in the case of $1 + 1$-dimensional gauge theory (a special case, being an exactly solvable model), this is not necessarily true. In [41], it was demonstrated that a few-parameter ansatz (inspired by Lüscher's flows) is capable of substantially out-performing numerically trained flows with $\sim 10^6$ parameters. The key table from this work is reproduced as Table 1.

3. Contour deformations

We now turn from methods designed to accelerate sampling, to those designed to modify the signal-to-noise ratio after sampling has been completed. The method of *contour deformations* is based on Cauchy's integral theorem. In the case of a single complex dimension, for any contour γ obtained by continuously deforming $\mathbb{R} \subset \mathbb{C}$, we have the equality

$$\int_{\mathbb{R}} f(z)dz = \int_{\gamma} f(z)dz \tag{3}$$

for any function $f(z)$ which is holomorphic. The complex integration measure is defined as $dz = dx + idy$. The two integrals are equal, but the values the integrand takes along $\mathbb{R}$ are not the same as those the integrand takes along the contour γ. As a result, one integral may be numerically easier than the other.

In the N-dimensional case, Cauchy's integral theorem states[3]

$$\int_{\mathbb{R}^N} f(z)dz = \int_{\mathbb{R}^N} f(\phi(x)) \det \frac{\partial \phi}{\partial x} dz \tag{4}$$

---

[2]Note that this is not because gauge-equivariance is required for correctness, but rather because it is believed to be easier to train a flow which already "knows about" all relevant symmetries.

[3]As pointed out in [42], this construction excludes those contours homologous but not homotopic to the real plane. No such contour has yet been used to mitigate a sign problem.
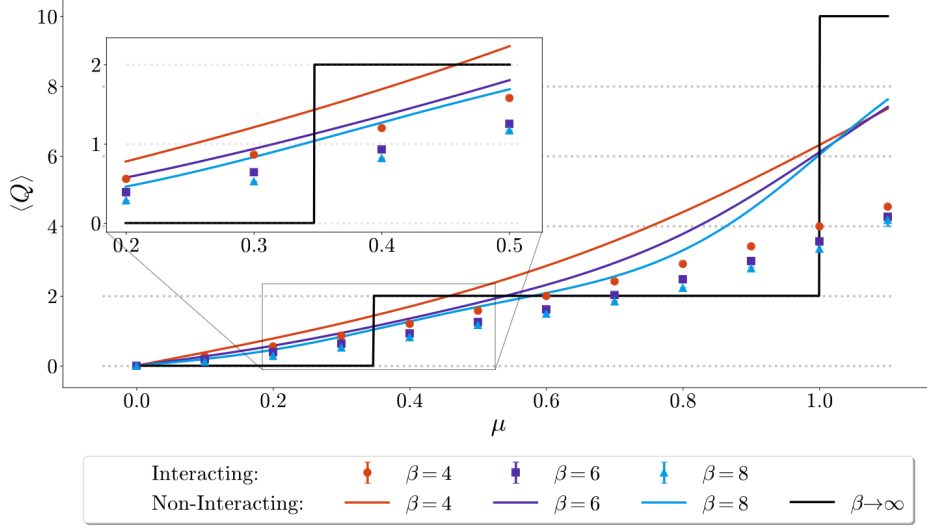
Figure 2: Reproduced from [43], the charge of Perylene as a function of chemical potential. Computed with a contour deformation, defined by one parameter.

for any continuous (and piecewise differentiable) $\phi : \mathbb{R}^N \to \mathbb{C}^N$ parameterizing an N-dimensional integration contour. Suitable generalizations to functions defined on manifolds that only locally look like $\mathbb{R}^N$ are readily available. This includes the integrals over $SU(N)$ which are relevant to lattice QCD.

This fact has been applied to both sign problems (where the Boltzmann factor itself posesses strong fluctuations) and signal-to-noise problems, in slightly different ways.

3.1 Sign problems

Historically, the motivating applications of contour deformations were theories that exhibit *sign problems*. In short[4], we have a theory for which the action is in general complex, and therefore the Boltzmann factor e^{-S} no longer defines a probability distribution. In this context expectation values are commonly computed by sampling with respect to the "quenched ensemble", defined by a probability distribution proportional to $|e^{-S}|$. This comes at a severe cost. Defining the *average phase*

$$\langle\sigma\rangle \equiv \frac{\int e^{-S(U)}dU}{\int |e^{-S(U)}|dU} \equiv \frac{Z}{Z_Q}, \tag{5}$$

the number of samples required to overcome the sign problem scales as $\langle\sigma\rangle^{-2}$. As the average phase typically decays exponentially with the system volume, this imposes an exponential cost on lattice calculations.

Now consider performing a contour deformation—instead of integrating over $\mathbb{S}U(3)^{\otimes V}$, we integrate over some contour $\gamma \subset SL(3;\mathbb{C})^{\otimes V}$. Expectation values are expressible as a ratio of two

---

[4]See [44] for a review.

holomorphic integrals, and are therefore not affected by contour deformation:

$$\langle O \rangle = \frac{\int_{SU(3)^{\otimes V}} O(U) e^{-S(U)} d^V U}{\int_{SU(3)^{\otimes V}} e^{-S(U)} d^V U} = \frac{\int_\gamma O(M) e^{-S(M)} d^V M}{\int_\gamma e^{-S(M)} d^V M}. \tag{6}$$

The figure of merit above, however—the average phase—cannot be written in this way. The quenched partition function Z_Q is dependent on the choice of contour. We therefore conclude that it is possible to improve the sign problem by judicious choice of integration contour.

It remains to find a useful integration contour: one on which $\langle \sigma \rangle$ is as large as possible. Inspired by early work on analytic continuation of field theories [45], the first proposed contours were Lefschetz thimbles [46]: a particular family of contours for which the imaginary part of the action happens to be constant[5]. The Lefschetz thimbles are computationally expensive to find, and so it was proposed to use supervised learning to obtain an efficient approximation [47].

Shortly after, it was realized that instead of attempting to approximate the Lefschetz thimbles, one could directly maximize the average phase $\langle \sigma \rangle$ over a family of contour deformations [48, 49]. This unsupervised learning algorithm has been applied to a variety of systems, including fermions at finite density in $2 + 1$ dimensions [50].

One lesson learned from the many works searching for integration contours with a mild sign problem, is that very simple contour deformations are often sufficient in practice. The simplest nontrivial contour deformation is a "constant shift", where each degree of freedom has the same imaginary part added:

$$\phi_x \rightarrow \phi_x + iC.$$

This sort of contour deformation has proven useful in the study of small systems of strongly correlated electrons [51]. Note that when performing a Monte Carlo on a deformed contour, there is a Jacobian determinant that must be evaluated, coming from the curvature of the contour. A constant shift has no curvature, and so there is no cost associated to this determinant. This line of work culminated in the calculation of the behavior and spectrum of doped Perylene [43]. From that work, the charge as a function of chemical potential is shown in Figure 2.

More complicated contours, defined by many-parameter neural networks, were demonstrated to be valuable for the calculation of scalar $\lambda \phi^4$ field theory at complex coupling [42]. This model is of possible interest as a $\mathcal{PT}$-symmetric field theory [52, 53, 54, 55]. Key figures from this work are reproduced in Figure 3.

These sorts of machine-learned contour deformations have also been used to study the *XY* model in $2 + 1$ dimensions [56], $U(1)$ gauge theory with complex parameters [57], one-dimensional QCD [58], QCD in the heavy-dense limit [59].

As with normalizing flows (discussed above), it is generally believed that an ansatz will perform better if the symmetries of the problem are "built-in", so that they do not have to be learned during training. This has provided motivation for the study of the performance of neural networks that take as input only gauge-invariant quantities [60].

---

[5]To prevent a common misconception, note that the Lefschetz thimbles do not in general maximize $\langle \sigma \rangle$. This is both because the Jacobian determinant introduces additional phase fluctuations, and because of cancellations between the thimbles.
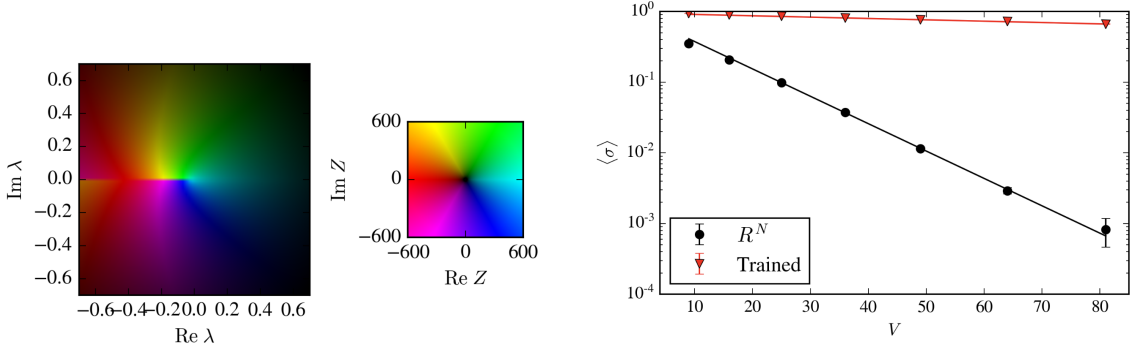
Figure 3: From [42], investigations of lattice scalar field theory at complex coupling. At left is shown the partition function of $0 + 1$-dimensional scalar ($\lambda\phi^4$) field theory as a function of complex anharmonicity λ; the branch cut on $\mathbb{R}_-$ can clearly be seen. At right, the average phase of $1 + 1$-dimensional lattice scalar field theory at $\lambda = i$, as a function of volume. Both the defining contour $\mathbb{R}^V$ and the trained contour exhibit exponential decays, but the trained contour has a slope more than an order of magnitude smaller.

Not all successes of the contour deformation method have stemmed from machine learning. A popular alternative is the *holomorphic gradient flow*, which defines a family of contours that approximate the Lefschetz thimbles. This family has been used to study the chiral random matrix model [61] and QED in $1 + 1$ dimensions [62].

The contour deformation method, on its face, is only able to handle systems with continuous parameters (which can then be complexified). This can be extended to spin systems by introducing auxiliary degrees of freedom [63, 64, 65].

Finally, most of the works above have focused on systems at a finite chemical potential (which introduces the sign problem). Via the lattice Schwinger-Keldysh formalism, real-time dynamics can also be studied, albeit with a severe sign problem. This sign problem has been the target of various contour deformation-based methods [66, 67, 64], including machine learning [68]. However, in the case of field variables on compact spaces (particularly relevant to gauge theory), there are substantial technical obstacles to preserving unitarity in time-evolution [69].

3.2 Signal-to-noise problems

Sign problems and signal-to-noise problems are closely related. In fact, a sign problem can be seen as a signal-to-noise problem in evaluating the ratio of partition functions Z/Z_Q. Just as a sign problem can be improved by choice of appropriate contour (because Z_Q is not the integral of a holomorphic quantity), so can the noise in an observable be reduced by choice of an appropriate integration contour. This comes from the fact that the variance associated to $\langle O \rangle$ is $\langle O^\dagger O \rangle$—the operator in the expectation value is not holomorphic.

Thus motivated, simple contour deformations have been explored in the context of $U(1)$ gauge theory and complex scalar field theory [71], and two-dimensional $SU(2)$ Yang-Mills [70]. A demonstration from the latter work is reproduced as Figure 4. The improvement in the signal-to-noise ratio is shown to be exponential in the size of the Wilson loop.
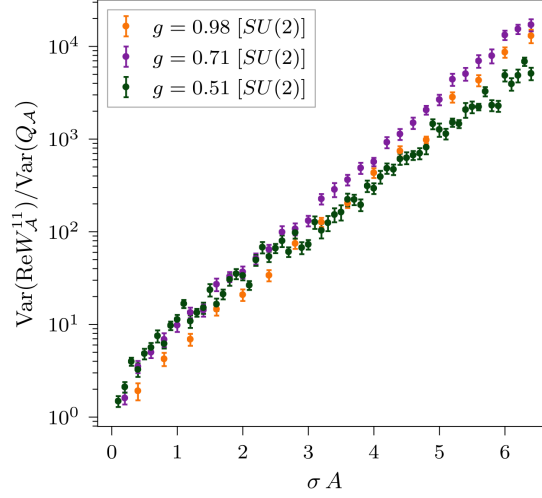
Figure 4: From [70], evidence of exponential improvement in the signal-to-noise problem of two-dimensional $SU(2)$ Yang-Mills. Shown is the ratio of the variance of the estimator of the Wilson loop, to the the variance of a contour-improved estimator of the same quantity.

3.3 Theoretical aspects

Beyond the machine-learning aspects of this method, a great deal is now known about the space of possible contour deformations and their sign problems. Here we briefly summarize a few important results and major open questions.

If, in the definition (2) of a normalizing flow, the function $x : \mathbb{R}^N \to \mathbb{C}^N$ is permitted to be complex-valued, then $x(z)$ can be seen to parameterize a complex integration contour. This observation leads to the notion of a *complex normalizing flow* [68], which can be used to work around the difficulty of efficiently sampling on an arbitrary contour deformation. In the context of machine learning, by learning a complex normalizing flow, one obtains many of the benefits of a contour deformation and a normalizing flow simultaneously.

The discussion of contour deformations above hinged on the Boltzmann factor e^{-S} being holomorphic. For some applications (e.g. nuclear forces) this assumption does not hold. It has been shown that this approximation can be relaxed, either by approximating the action by a nearby holomorphic one [72], or by constructing a Riemann surface on which the Boltzmann factor is holomorphic [73]. In both cases, a key insight is that the calculation we want to perform is defined only on the real plane. The behavior of the action (and other functions) away from the real plane is a choice we make for algorithmic convenience.

It has been shown that a necessary and sufficient condition for a contour γ to be a *local maximum* of $\langle \sigma \rangle$, is that the integrand $e^{-S}dz$ has locally constant phase as long as that phase is defined. In other words, the phase of $e^{-S}dz$ is piecewise constant, with regions of different phase being separated by zeros of that integrand. This is reminiscent of a similar property of Lefschetz thimbles, where it is the phase of e^{-S} that is constant except where that Boltzmann factor vanishes. It has been conjectured [51, 74] that there exists an expansion in $\hbar$, of which the Lefschetz thimbles are the leading-order term and the perfect contours are the sum. Rigorous results to this effect are
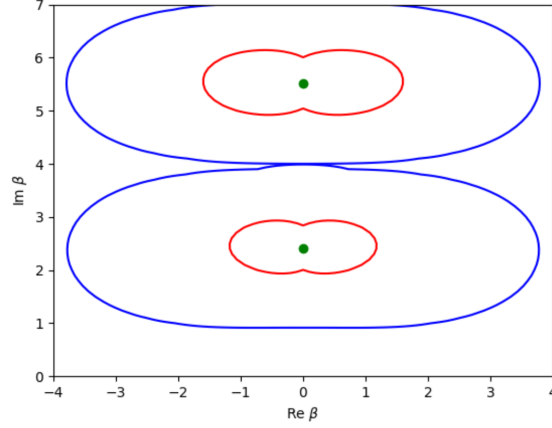
Figure 5: No-go regions for $1 + 1$-dimensional lattice Yang-Mills at complex coupling β, from [74]. Outside of the blue lines, perfect integration contours (on which $\langle \sigma \rangle = 1$) were found numerically; within the red lines, proofs that no such contours existed were found. The black points indicate zeros of the partition function.

not yet available.

Empirically, most studies that use contour deformations have failed to find a perfect contour. That is, there is usually some residual average phase, which decays exponentially with the volume of the system. When this happens, it is *a priori* unclear whether no such contour exists, or the learning algorithm has simply failed to find the perfect contour (in which case applying more computational power may be advisable). In [74], a method of proving that no perfect contour exists was devised. Writing the quenched partition function over a contour γ as $\int_\gamma |\omega|$ for some differential form ω, one finds another differential form α obeying both $d\alpha = 0$ and $|\alpha| \leq |\omega|$. For any such form α one has the following sequence of inequalities:

$$\left| \int_\mathbb{R} \alpha \right| = \left| \int_\gamma \alpha \right| \leq \int_\gamma |\alpha| \leq \int_\gamma |\omega|, \tag{7}$$

establishing that $\int_\mathbb{R} \alpha$ is a lower bound on the quenched partition function, no matter the choice of contour. By this method an exponentially decaying upper bound on $\langle \sigma \rangle$ was found for $1 + 1$-dimensional lattice Yang-Mills for a range of complex couplings (see Figure 5). No such bounds have yet been found for theories that are not exactly solvable; nor have any such bounds been derived for improvements in signal-to-noise problems.

4. Control variates

For the remainder of this talk we continue with the theme of variance reduction. Of course this can often be accomplished via normalizing flows [75]: the variance is reduced if the number of samples can be efficiently increased. However it may be that no high-quality normalizing flow can be constructed, or that the cost of measurement outweighs the cost of sampling. In such cases we desired to modify the observable in such a way that the variance changes while the expectation value remains the same.

Suppose we know of a function f such that $\langle f \rangle = 0$ vanishes exactly. Then instead of computing $\langle O \rangle$, we might computed $\langle O - f \rangle$. The two expectation values are exactly equal, but the statistical noise might be improved. The function f is termed a *control variate*. Control variates are extraordinarily general: for every observable there exists a "perfect" control variate that entirely removes the noise [76]. Moreover, other variance reduction methods (including contour deformations as discussed above) can often be re-written as control variates [76]. This generality is both a blessing and a curse: we know that there exist desirable control variates, but we have little guidance in how to find them.

Control variates are in principle applicable to both sign problems and signal-to-noise problems. The most striking successes of control variates have all been in their application to signal-to-noise problems, so this section will focus on those cases. Small but significant improvements have been obtained in the sign problems for fermionic systems [77, 76] and Ising models at complex coupling [78].

To make this approach practical, we first need a large family of functions f whose expectation values can be proven[6] to vanish exactly. Such a family of functions is provided by noticing that the integral of a total derivative (on a compact space of field configurations) must vanish. As a result we have

$$\langle \partial g \rangle = \langle g \partial S \rangle, \tag{8}$$

where S is the action and g is any function of field configurations. Equalities of this form, which hold for any first-order derivative ∂, are termed *Schwinger-Dyson relations*.

The immediate problem with applying this method to lattices of realistic sizes is that there are simply too many Schwinger-Dyson relations, and in a sense we do not know which ones to use. To be concrete, consider lattice scalar field theory with V sites. At leading order there are V^2 Schwinger-Dyson relations available, of the form

$$\langle \frac{\partial}{\partial \phi_x} \phi_y \rangle = \langle \phi_y \frac{\partial}{\partial \phi_x} S \rangle \tag{9}$$

for (not necessarily distinct) sites (x, y). Using translational invariance this is reduced to a set of V relations. A general control variate may be written as a linear combination of these. This control variate has V free coefficients, which must be determined by a fit. On a $2 + 1$-dimensional lattice of reasonable ($\sim 100^3$) size, this fit requires $\gtrsim 10^6$ configurations, which are likely not available.

One approach, proposed in [79], is to use L_1 regularization. The assumption is made that a good control variate exists for which the vast majority of coefficients are exactly 0, and such a "sparse" control variate is optimized numerically. The performance of this method is shown in the left panel of Figure 6: an order-of-magnitude improvement in the signal-to-noise ratio, corresponding to a speed-up of a factor of $\sim 10^2$, is obtained.

Another proposed approach to this high-dimensional fit has been to parameterize the function g by a neural network [80]. Recall that any function g defines a control variate. The high-dimensional fitting problem can then be approached in any number of ways standard in machine learning, including regularization by limiting training time. Tested in scalar field theory, this method

---

[6]It is critical that $\langle f \rangle = 0$ be a theorem, rather than an empirical statement from Monte Carlo data. By construction, the observable f will be noisy.
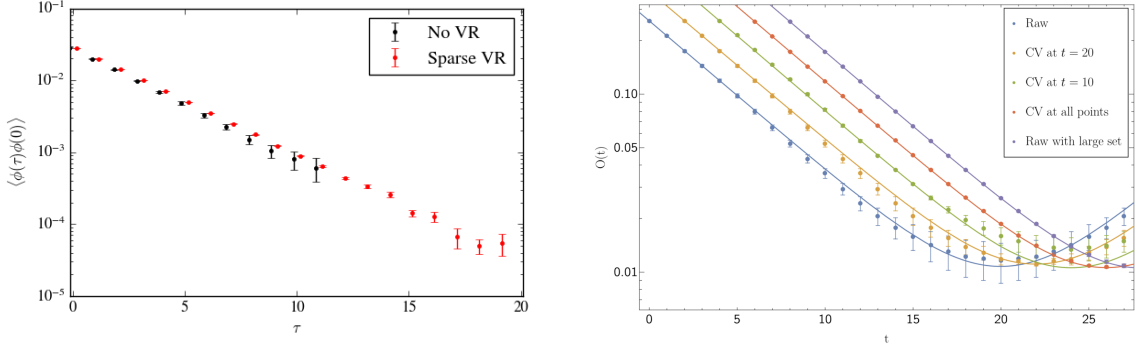
Figure 6: Two approaches to reducing the signal-to-noise problem associated with the correlator in scalar field theory, both based on Schwinger-Dyson control variates. At left [79], a complete basis of leading-order Schwinger-Dyson relations is used, with L_1 regularization to mitigate overfitting. At right [80], a neural network parameterizes the function $g(\phi)$ which defines the Schwinger-Dyson control variate.

also yields an improvement in the signal-to-noise ratio of more than an order of magnitude (see again Figure 6).

Recent work has also proposed constructing control variates in perturbation theory [81], although with less success than the previously mentioned approaches.

5. Surrogate observables

The previous section focused on reducing the noise associated with measuring an expectation value $\langle O \rangle$. This is valuable when, for any reason, we cannot perform a sufficiently large number of calculations of $O(U)$ on independent configurations. This may be because collecting the configurations is expensive (the cost of HMC), or because given U, the task of evaluating $O(U)$ is expensive. In either case, reducing the noise will reduce the number of configurations on which $O(U)$ is needed.

When $O(U)$ is expensive and collecting configurations U is (relatively) cheap, another approach is available. By one method or another, we construct an approximation $\tilde{O}(U) \approx O(U)$, which is cheaper to evaluate. We can evaluate $\tilde{O}$ on a larger number of configurations; however its use introduces a systematic bias from the difference $\langle O \rangle - \langle \tilde{O} \rangle \neq 0$. This bias can be corrected by evaluating $\langle O - \tilde{O} \rangle$. Per-sample, this is as expensive as evaluating O; however, because this last expectation value has smaller variance, it does not need to be evaluated on as many samples.

This approach is summarized by the equation

$$\langle O \rangle = \langle \tilde{O} \rangle + \langle O - \tilde{O} \rangle \approx \langle \tilde{O} \rangle_{\tilde{N}} + \langle O - \tilde{O} \rangle_N, \qquad (10)$$

where $\langle \cdot \rangle$ denotes the true lattice expectation value, and $\langle \cdot \rangle_n$ an expectation value with respect to n samples. The ratio $\frac{N}{\tilde{N}} < 1$ controls which term in the estimator dominates the variance.

It remains only to construct an approximation to the observable O, and at this point the connection to machine learning is obvious. The true observable O is computed on some set of configurations, which are used to train a neural network or other device to represent the approximation $\tilde{O}$. This approximation is by construction cheap to compute, enabling the above procedure.
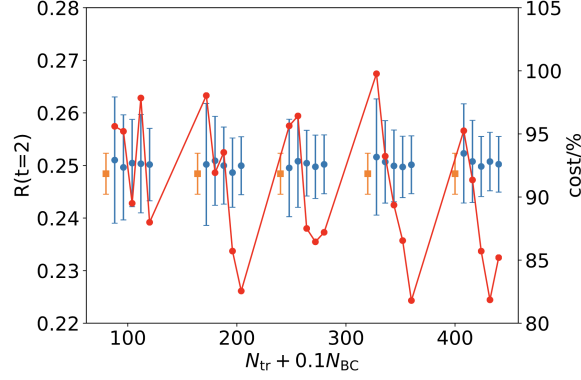
Figure 7: Predicting $R(t = 2)$ (ratio of 3-pt to 2-pt) of kaon quasi-PDF correlators at $(p_{pred}, t_{pred}, t_{sep}) = (4, 4, 5)$ from measurements at $(3, 4, 5)$. From [82]. The red curve shows the cost savings from the use of surrogate observables, with a maximum of around 20%.

This method has been applied to lattice QCD at scale, with mild success. In [83], boosted decision trees were used to construct approximations to, among other things, nucleon three-point functions given already-computed two-point functions. That work reported reduction in computational cost of 7%-38%. In [82], both the boosted decision trees and a linear regression model were used to extract kaon quasi-PDF correlators. One figure from this latter study is reproduced in Figure 7. The amount of improvement is similar to that in the previous study. A similar approach, using machine learning to extrapolate observables to longer flow times or lighter quark masses, is explored in [84].

6. Discussion

We have seen four broad families of machine learning algorithms for accelerating lattice simulations: normalizing flows for improving sampling, contour deformations chiefly for improving sign problems, control variates for improving the signal-to-noise ratio, and surrogate observables for accelerating measurement. All can be used without introducing any additional systematic bias into the Monte Carlo results. Normalizing flows and surrogate observables have already been proven somewhat useful "at scale"—at least, in 4-dimensional gauge theories. Contour deformations and control variates have proven useful in smaller systems, but not yet in lattice QCD itself.

A recurring theme in this field is that methods that involve large numbers of parameters (usually some form of "deep learning") do not systematically out-perform methods with few, or even no, parameters. This was highlighted in Table 1 above with respect to normalizing flows, but the phenomenon has also repeatedly been reported in the context of contour deformations (see for example [47, 50]).

A second pattern, noticeable throughout this talk, is that relatively few of these methods have been successfully applied to systems of physical interest. The most remarkable exception is the study of $C_{20}H_{12}$-Perylene via contour deformations [43], although this work is also not quite sufficient to make direct contact with experiment.

Those interested in using these techniques—or just gaining practical experience for self-edification—may find useful a normalizing flow tutorial published several years ago [85], and two software packages targetting normalizing flows for lattice field theory: [86, 87].

Acknowledgments

References

[1] S.Y. Chen, H.T. Ding, F.Y. Liu, G. Papp and C.B. Yang, *Machine learning spectral functions in lattice QCD*, 2110.13521.

[2] S. Offler, *A study of thermal NRQCD with machine learning methods*, Ph.D. thesis, Swansea U., 2022.

[3] R. Fournier, L. Wang, O.V. Yazyev and Q. Wu, *Artificial neural network approach to the analytic continuation problem*, *Phys. Rev. Lett.* **124** (2020) 056401.

[4] L. Kades, J.M. Pawlowski, A. Rothkopf, M. Scherzer, J.M. Urban, S.J. Wetzel et al., *Spectral reconstruction with deep neural networks*, *Phys. Rev. D* **102** (2020) 096001.

[5] L. Wang, S. Shi and K. Zhou, *Reconstructing spectral functions via automatic differentiation*, *Phys. Rev. D* **106** (2022) L051502 [2111.14760].

[6] J. Carrasquilla and R.G. Melko, *Machine learning phases of matter*, *Nature Physics* **13** (2017) 431.

[7] J.-H. Peng, Y.-H. Tseng and F.-J. Jiang, *Machine learning phases of an Abelian gauge theory*, *PTEP* **2023** (2023) 073A03 [2212.14655].

[8] A. Tanaka and A. Tomiya, *Detection of phase transition via convolutional neural network*, *J. Phys. Soc. Jap.* **86** (2017) 063001 [1609.09087].

[9] E.P.L. van Nieuwenburg, Y.-H. Liu and S.D. Huber, *Learning phase transitions by confusion*, *Nature Phys.* **13** (2017) 435 [1610.02048].

[10] J.F. Rodriguez-Nieva and M.S. Scheurer, *Identifying topological order through unsupervised machine learning*, *Nature Phys.* **15** (2019) 790 [1805.05961].

[11] P. Broecker, J. Carrasquilla, R.G. Melko and S. Trebst, *Machine learning quantum phases of matter beyond the fermion sign problem*, *Sci. Rep.* **7** (2017) 8823.

[12] P.E. Shanahan, A. Trewartha and W. Detmold, *Machine learning action parameters in lattice quantum chromodynamics*, *Phys. Rev. D* **97** (2018) 094506 [1801.05784].

[13] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, *Science* **355** (2017) 602.

[14] D.-L. Deng, X. Li and S.D. Sarma, *Machine learning topological states*, *Phys. Rev. B* **96** (2017) 195145.

[15] D. Luo, G. Carleo, B.K. Clark and J. Stokes, *Gauge Equivariant Neural Networks for Quantum Lattice Gauge Theories*, *Phys. Rev. Lett.* **127** (2021) 276402 [2012.05232].

[16] K. Holland, A. Ipp, D.I. Müller and U. Wenger, *Machine learning a fixed point action for SU(3) gauge theory with a gauge equivariant convolutional neural network*, 2401.06481.

[17] G. Kanwar, *Flow-based sampling for lattice field theories*, PoS **LATTICE2023** (2024) 114 [2401.01297].

[18] L. Huang and L. Wang, *Accelerated Monte Carlo simulations with restricted Boltzmann machines*, *Phys. Rev. B* **95** (2017) 035105 [1610.02746].

[19] A. Tanaka and A. Tomiya, *Towards reduction of autocorrelation in HMC by machine learning*, 1712.03893.

[20] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, *JHEP* **05** (2024) 060 [2309.17082].

[21] L. Wang, G. Aarts and K. Zhou, *Generative Diffusion Models for Lattice Field Theory*, in *37th Conference on Neural Information Processing Systems*, 11, 2023 2311.03578.

[22] K. Zhou, G. Endrődi, L.-G. Pang and H. Stöcker, *Regressive and generative neural networks for scalar field theory*, *Phys. Rev. D* **100** (2019) 011501 [1810.12879].

[23] J.M. Pawlowski and J.M. Urban, *Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks*, *Mach. Learn. Sci. Tech.* **1** (2020) 045011 [1811.03533].

[24] G.E. Box and M.E. Muller, *A note on the generation of random normal deviates*, *The annals of mathematical statistics* **29** (1958) 610.

[25] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515 [1904.12072].

[26] H. Nicolai, *On a New Characterization of Scalar Supersymmetric Theories*, *Phys. Lett. B* **89** (1980) 341.

[27] M. Luscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, *Commun. Math. Phys.* **293** (2010) 899 [0907.5491].

[28] L. Dinh, D. Krueger and Y. Bengio, *NICE: Non-linear Independent Components Estimation*, 1410.8516.

[29] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using Real NVP*, 1605.08803.

[30] D.P. Kingma and P. Dhariwal, *Glow: Generative Flow with Invertible 1x1 Convolutions*, 1807.03039.

[31] M. Caselle, E. Cellini and A. Nada, *Sampling the lattice Nambu-Goto string using Continuous Normalizing Flows*, *JHEP* **02** (2024) 048 [2307.01107].

[32] P. de Haan, C. Rainone, M.C.N. Cheng and R. Bondesan, *Scaling Up Machine Learning For Quantum Field Theory with Equivariant Continuous Flows*, 2110.02673.

[33] K.A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller and P. Kessel, *Asymptotically unbiased estimation of physical observables with neural samplers*, *Phys. Rev. E* **101** (2020) 023304 [1910.13496].

[34] K.A. Nicoli, C.J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel et al., *Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models*, *Phys. Rev. Lett.* **126** (2021) 032001 [2007.07115].

[35] D. Boyda, G. Kanwar, S. Racanière, D.J. Rezende, M.S. Albergo, K. Cranmer et al., *Sampling using $SU(N)$ gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504 [2008.05456].

[36] G. Kanwar, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, S. Racanière et al., *Equivariant flow-based sampling for lattice gauge theory*, *Phys. Rev. Lett.* **125** (2020) 121601 [2003.06413].

[37] G. Kanwar, *Machine Learning and Variational Algorithms for Lattice Field Theory*, Ph.D. thesis, MIT, 2021. 2106.01975.

[38] R. Abbott et al., *Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions*, *Phys. Rev. D* **106** (2022) 074506 [2207.08945].

[39] M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, G. Kanwar, S. Racanière et al., *Flow-based sampling in the lattice Schwinger model at criticality*, *Phys. Rev. D* **106** (2022) 014514 [2202.11712].

[40] R. Abbott et al., *Normalizing flows for lattice gauge theory in arbitrary space-time dimension*, 2305.02402.

[41] S. Bacchio, P. Kessel, S. Schaefer and L. Vaitl, *Learning trivializing gradient flows for lattice gauge theories*, *Phys. Rev. D* **107** (2023) L051504 [2212.08469].

[42] S. Lawrence, H. Oh and Y. Yamauchi, *Lattice scalar field theory at complex coupling*, *Phys. Rev. D* **106** (2022) 114503 [2205.12303].

[43] M. Rodekamp, E. Berkowitz, C. Gäntgen, S. Krieg, T. Luu, J. Ostmeyer et al., *Single Particle Spectrum of Doped* $C_{20}H_{12}$*-Perylene*, 2406.06711.

[44] A. Alexandru, G. Basar, P.F. Bedaque and N.C. Warrington, *Complex paths around the sign problem*, *Rev. Mod. Phys.* **94** (2022) 015006 [2007.05436].

[45] E. Witten, *Analytic Continuation Of Chern-Simons Theory*, *AMS/IP Stud. Adv. Math.* **50** (2011) 347 [1001.2933].

[46] AuroraScience collaboration, *New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble*, *Phys. Rev. D* **86** (2012) 074506 [1205.3996].

[47] A. Alexandru, P.F. Bedaque, H. Lamm and S. Lawrence, *Deep Learning Beyond Lefschetz Thimbles*, *Phys. Rev. D* **96** (2017) 094505 [1709.01971].

[48] Y. Mori, K. Kashiwa and A. Ohnishi, *Toward solving the sign problem with path optimization method*, *Phys. Rev. D* **96** (2017) 111501 [1705.05605].

[49] A. Alexandru, P.F. Bedaque, H. Lamm and S. Lawrence, *Finite-Density Monte Carlo Calculations on Sign-Optimized Manifolds*, *Phys. Rev. D* **97** (2018) 094510 [1804.00697].

[50] A. Alexandru, P.F. Bedaque, H. Lamm, S. Lawrence and N.C. Warrington, *Fermions at Finite Density in 2+1 Dimensions with Sign-Optimized Manifolds*, *Phys. Rev. Lett.* **121** (2018) 191602 [1808.09799].

[51] C. Gäntgen, E. Berkowitz, T. Luu, J. Ostmeyer and M. Rodekamp, *Fermionic sign problem minimization by constant path integral contour shifts*, *Phys. Rev. B* **109** (2024) 195158 [2307.06785].

[52] C.M. Bender, N. Hassanpour, S.P. Klevansky and S. Sarkar, *PT-symmetric quantum field theory in D dimensions*, *Phys. Rev. D* **98** (2018) 125003 [1810.12479].

[53] P. Romatschke, *On the negative coupling O(N) model in 2d at high temperature*, 2412.10496.

[54] S. Lawrence, R. Weller, C. Peterson and P. Romatschke, *Instantons, analytic continuation, and PT-symmetric field theory*, *Phys. Rev. D* **108** (2023) 085013 [2303.01470].

[55] R.D. Weller, *Can negative bare couplings make sense? The $\vec{\phi}^4$ theory at large N*, 2310.02516.

[56] Z. Tulipant, M. Giordano, K. Kapas, S.D. Katz and A. Pasztor, *Exponential improvement of the sign problem via contour deformations in the 2+1D XY model at non-zero density*, *Phys. Rev. D* **106** (2022) 054512 [2202.07561].

[57] K. Kashiwa and Y. Mori, *Path optimization for U(1) gauge theory with complexified parameters*, *Phys. Rev. D* **102** (2020) 054519 [2007.04167].

[58] G. Basar and J. Marincel, *Sign optimization and complex saddle points in one-dimensional QCD*, *Phys. Rev. D* **106** (2022) L091503 [2208.02072].

[59] G. Basar and J. Marincel, *Heavy-dense QCD, sign optimization, and Lefschetz thimbles*, *Phys. Rev. C* **109** (2024) 045208 [`2311.06343`].

[60] Y. Namekawa, K. Kashiwa, A. Ohnishi and H. Takase, *Gauge invariant input to neural network for path optimization method*, *Phys. Rev. D* **105** (2022) 034502 [`2109.11710`].

[61] M. Giordano, A. Pasztor, D. Pesznyak and Z. Tulipant, *Alleviating the sign problem in a chiral random matrix model with contour deformations*, *Phys. Rev. D* **108** (2023) 094507 [`2301.12947`].

[62] A. Alexandru, G. Başar, P.F. Bedaque, H. Lamm and S. Lawrence, *Finite Density QED$_{1+1}$ Near Lefschetz Thimbles*, *Phys. Rev. D* **98** (2018) 034506 [`1807.02027`].

[63] K. Kashiwa, Y. Namekawa, A. Ohnishi and H. Takase, *Application of the path optimization method to a discrete spin system*, *Phys. Rev. D* **108** (2023) 094504 [`2309.06018`].

[64] N.C. Warrington, *Real-time spin systems from lattice field theory*, *JHEP* **12** (2023) 156 [`2310.19761`].

[65] T.C. Mooney, J. Bringewatt, N.C. Warrington and L.T. Brady, *Lefschetz thimble quantum Monte Carlo for spin systems*, *Phys. Rev. B* **106** (2022) 214416 [`2110.10699`].

[66] A. Alexandru, G. Basar, P.F. Bedaque and G.W. Ridgway, *Schwinger-Keldysh formalism on the lattice: A faster algorithm and its application to field theory*, *Phys. Rev. D* **95** (2017) 114501 [`1704.06404`].

[67] A. Alexandru, G. Basar, P.F. Bedaque, S. Vartak and N.C. Warrington, *Monte Carlo Study of Real Time Dynamics on the Lattice*, *Phys. Rev. Lett.* **117** (2016) 081602 [`1605.08040`].

[68] S. Lawrence and Y. Yamauchi, *Normalizing Flows and the Real-Time Sign Problem*, *Phys. Rev. D* **103** (2021) 114509 [`2101.05755`].

[69] G. Kanwar and M.L. Wagman, *Real-time lattice gauge theory actions: Unitarity, convergence, and path integral contour deformations*, *Phys. Rev. D* **104** (2021) 014513 [`2103.02602`].

[70] W. Detmold, G. Kanwar, H. Lamm, M.L. Wagman and N.C. Warrington, *Path integral contour deformations for observables in $SU(N)$ gauge theory*, *Phys. Rev. D* **103** (2021) 094517 [`2101.12668`].

[71] W. Detmold, G. Kanwar, M.L. Wagman and N.C. Warrington, *Path integral contour deformations for noisy observables*, *Phys. Rev. D* **102** (2020) 014514 [`2003.05914`].

[72] G. Kanwar, A. Lovato, N. Rocco and M. Wagman, *Mitigating Green's function Monte Carlo signal-to-noise problems using contour deformations*, *Phys. Rev. C* **109** (2024) 034317 [`2304.03229`].

[73] S. Lawrence, S. Valgushev, J. Xiao and Y. Yamauchi, *Contour deformations for non-holomorphic actions*, `2401.16733`.

[74] S. Lawrence and Y. Yamauchi, *Convex optimization of contour deformations*, *Phys. Rev. D* **110** (2024) 014508 [2311.13002].

[75] R. Abbott, A. Botev, D. Boyda, D.C. Hackett, G. Kanwar, S. Racanière et al., *Applications of flow models to the generation of correlated lattice QCD ensembles*, *Phys. Rev. D* **109** (2024) 094514 [2401.10874].

[76] S. Lawrence and Y. Yamauchi, *Deep learning of fermion sign fluctuations*, *Phys. Rev. D* **107** (2023) 114505 [2212.14606].

[77] S. Lawrence, *Perturbative Removal of a Sign Problem*, *Phys. Rev. D* **102** (2020) 094504 [2009.10901].

[78] S. Lawrence and Y. Yamauchi, *Mitigating a discrete sign problem with extreme learning machines*, 2312.12636.

[79] T. Bhattacharya, S. Lawrence and J.-S. Yoo, *Control variates for lattice field theory*, *Phys. Rev. D* **109** (2024) L031505 [2307.14950].

[80] P.F. Bedaque and H. Oh, *Leveraging neural control variates for enhanced precision in lattice field theory*, *Phys. Rev. D* **109** (2024) 094519 [2312.08228].

[81] S. Lawrence, *Schwinger-Dyson control variates for lattice fermions*, 2404.10707.

[82] R. Zhang, Z. Fan, R. Li, H.-W. Lin and B. Yoon, *Machine-learning prediction for quasiparton distribution function matrix elements*, *Phys. Rev. D* **101** (2020) 034516 [1909.10990].

[83] B. Yoon, T. Bhattacharya and R. Gupta, *Machine Learning Estimators for Lattice QCD Observables*, *Phys. Rev. D* **100** (2019) 014504 [1807.05971].

[84] J. Kim, G. Pederiva and A. Shindler, *Machine learning mapping of lattice correlated data*, *Phys. Lett. B* **856** (2024) 138894 [2402.07450].

[85] M.S. Albergo, D. Boyda, D.C. Hackett, G. Kanwar, K. Cranmer, S. Racanière et al., *Introduction to Normalizing Flows for Lattice Field Theory*, 2101.08176.

[86] A. Tomiya and S. Terasaki, *GomalizingFlow.jl: A Julia package for Flow-based sampling algorithm for lattice field theory*, 2208.08903.

[87] K.A. Nicoli, C.J. Anders, L. Funcke, K. Jansen, S. Nakajima and P. Kessel, *NeuLat: a toolbox for neural samplers in lattice field theories*, *PoS* **LATTICE2023** (2024) 286.