PROCEEDINGS OF SCIENCE

Fixed point actions from convolutional neural networks

Kieran Holland,[a] **Andreas Ipp,**[b] **David I. Müller**[b] **and Urs Wenger**[c,*]

[a] *University of the Pacific, 3601 Pacific Ave., Stockton, CA 95211, USA*

[b] *Institute for Theoretical Physics, TU Wien, Wiedner Hauptstraße 8-10/136, A-1040 Vienna, Austria*

[c] *Albert Einstein Center for Fundamental Physics, Institute for Theoretical Physics, University of Bern, Sidlerstraße 5, 3012 Bern, Switzerland*

E-mail: kholland@pacific.edu, ipp@hep.itp.tuwien.ac.at, dmueller@hep.itp.tuwien.ac.at, wenger@itp.unibe.ch

Lattice gauge-equivariant convolutional neural networks (L-CNNs) can be used to form arbitrarily shaped Wilson loops and can approximate any gauge-covariant or gauge-invariant function on the lattice. Here we use L-CNNs to describe fixed point (FP) actions which are based on renormalization group transformations. FP actions are classically perfect, i.e., they have no lattice artifacts on classical gauge-field configurations satisfying the equations of motion, and therefore possess scale invariant instanton solutions. FP actions are tree–level Symanzik–improved to all orders in the lattice spacing and can produce physical predictions with very small lattice artifacts even on coarse lattices. We find that L-CNNs are much more accurate at parametrizing the FP action compared to older approaches. They may therefore provide a way to circumvent critical slowing down and topological freezing towards the continuum limit.

The 40th International Symposium on Lattice Field Theory (Lattice 2023)
July 31st - August 4th, 2023
Fermi National Accelerator Laboratory

*Speaker

PoS(LATTICE2023)038

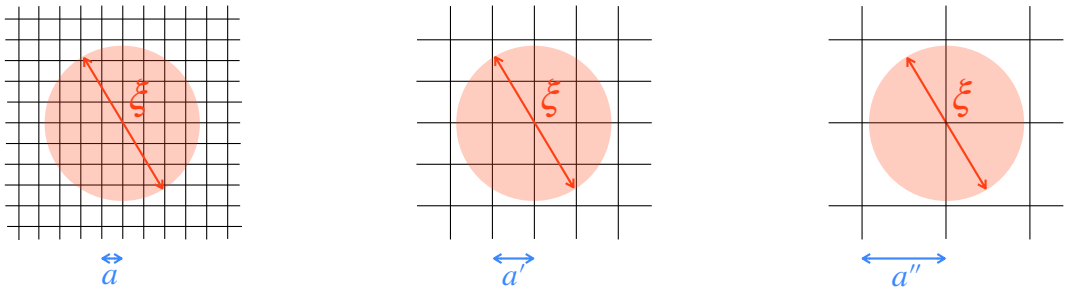


Figure 1: Sketch of the procedure of taking the continuum limit: as the gauge coupling is decreased from right to left, $g < g' < g''$, or equivalently $\beta > \beta' > \beta''$ is increased, the lattice spacing decreases, $a < a' < a''$. In the limit $\beta \to \infty$ the lattice spacing $a \to 0$, i.e., $\xi/a \to \infty$, and the continuum limit is reached. Renormalization group transformations map the system from the left to the right side.

1. Introduction

Consider an asymptotically free gauge field theory on the lattice, e.g., SU(N_c) lattice gauge theory, which is described by the partition function

$$Z(\beta) = \int \mathcal{D}U \exp\{-\beta A[U]\}$$

with gauge coupling $\beta = 2N_c/g^2$ and the gauge action $A[U]$ which is a function of the gauge links U. The integration over the gauge links is defined via the Haar measure $\mathcal{D}U$ of the gauge group. Expectation values for observables $O_\xi[U]$ with a characteristic length scale ξ are defined as

$$\langle O_\xi(\beta)\rangle = \frac{1}{Z}\int \mathcal{D}U \exp\{-\beta A[U]\}\, O_\xi[U].$$

A typical observable could be, for example, a correlation function of operators whose exponential decay at asymptotically large time separations is governed by ξ. When the physical scale is expressed in units of the lattice spacing a, the result ξ/a is dimensionless. The lattice spacing itself is determined by the gauge coupling, i.e., $a = a(\beta)$. Specifically, the continuum limit of the theory is reached by taking $\beta \to \infty$, such that $a \to 0$ and $\xi/a \to \infty$. This situation is sketched in Figure 1. From the point of view of a statistical lattice system the continuum limit of the lattice field theory corresponds to approaching the critical point of a continuous (second order) phase transition where the correlation length ξ diverges. The universality at the critical point guarantees the independence of the so-obtained physical observables from the details of the microscopic lattice definition of the theory, i.e., different discretizations lead to the same universal results. This has long been understood in the context of renormalization group transformations (RGT). Performing a (real space) RGT by blocking the degrees of freedom increases the lattice spacing $a \to a' \to a''$ and maps the system from left to right in Figure 1, while keeping the physical length scales unchanged. It is therefore in principle possible to extract continuum physics, i.e., values of the observables in the continuum, from systems defined at finite lattice spacings as long as the corresponding correlation length is still well defined, i.e., $\xi/a \gtrsim 1$.

It is well known that taking the continuum limit in practice is computationally very demanding due to the problem of *critical slowing down* when approaching a critical point. Moreover, for SU(N_c) gauge theories *topological freezing* poses an additional problem [1]. It is therefore prohibitive for lattice simulations to reach very fine lattice spacings. On the other hand, simulations at coarse lattice spacings are computationally cheap, but (with the usual discretizations) not very helpful as the lattice artifacts are large and not well controlled. This is where the RGT comes to the rescue by providing *discretizations without lattice artifacts* which allow cheap simulations at coarse lattice spacings, thereby avoiding critical slowing down and topological freezing, as well as uncontrollable lattice artifacts at the same time (see Ref. [2] for a pedagogical introduction to the topic).

2. Renormalization group transformations

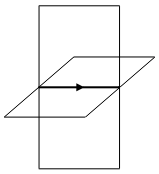
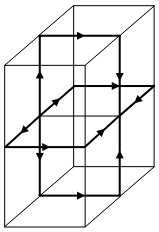
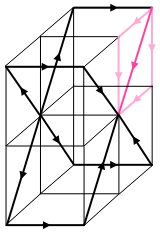
A real space RGT can be defined by averaging (blocking) the degrees of freedom on the fine lattice before integrating them out. More specifically, one has

$$\exp\{-\beta' A'[V]\} = \int \mathcal{D}U \exp\{-\beta\,(A[U] + T[U,V])\}\,, \tag{1}$$

where $T[U,V]$ is a blocking kernel relating the fine gauge links U to the coarse gauge links V. For gauge theories, the blocking kernel can be defined as

$$T[U,V] = -\kappa \sum_{x_B,\mu} \left\{ \mathrm{ReTr}\left(V_\mu(x_B)\cdot Q^\dagger_\mu(x_B)\right) - \mathcal{N}^\beta_\mu \right\},$$

where the sum is over the lattice sites x_B of the coarse lattice and the normalization factor $\mathcal{N}^\beta_\mu$ guarantees $Z(\beta') = Z(\beta)$, i.e., unchanged long distance physics. The blocked link $Q_\mu(x_B)$ is obtained by first smearing the fine links using a linear combination of the original link with planar, spatially diagonal and hyperdiagonal staples according to

$S^{\mathrm{smeared}}_\mu = s_0 \cdot$  $+\, s_{\mathrm{pl}} \cdot$  $+\, s_{\mathrm{d}} \cdot$  $+\, s_{\mathrm{hd}} \cdot \ldots,$

with $s_0, s_{\mathrm{pl}}, s_{\mathrm{d}}, s_{\mathrm{hd}}$ being arbitrary constants [3]. Note that the smeared links S^{smeared}_μ are no longer an element of the gauge group. These smeared links are now multiplied together such that they connect lattice sites corresponding to the ones on the coarse lattice,

$$Q_\mu(x_B) = S^{\mathrm{smeared}}_\mu(x)\cdot S^{\mathrm{smeared}}_\mu(x+\hat{\mu})\,.$$

It is easy to see that this procedure produces a linear combination of a plethora of gauge link paths connecting x and $x + 2\hat{\mu}$ and taking all links within the attached hypercubes into account.

The effective action $\beta' A'[V]$ on the LHS of Eq. (1) is in general described by infinitely many couplings $\{c'_\alpha\}$. Repeating the RGT yields a sequence of sets of couplings which maps out a flow

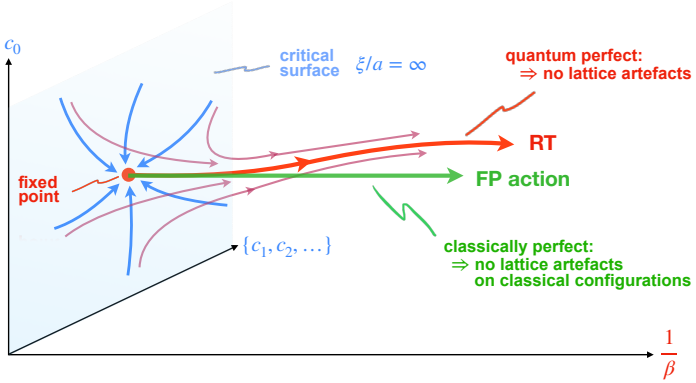


Figure 2: Illustration of the RGT flow in the infinite-dimensional space of couplings. Under repeated RGTs the couplings approach the renormalized trajectory (RT), unless one starts on the critical surface in which case the couplings flow into the FP of the RGT. Figure taken from [4].

in the infinite dimensional coupling space as illustrated in Figure 2. Under repeated RGTs the couplings approach the renormalized trajectory (RT), unless the RGT procedure starts from a set of couplings on the critical surface where $\xi/a = \infty$. In that case the couplings flow into the fixed point (FP) of the RGT defined by $\{c_\alpha^{\rm FP}\} \stackrel{\rm RGT}{\longrightarrow} \{c_\alpha^{\rm FP}\}$. Note that the infrared physics described by the actions on the RT is unchanged since the RGT only integrates out ultraviolet modes. Hence, actions defined along the RT reproduce continuum physics without any lattice artifacts and are therefore *quantum perfect*.

Finding such quantum perfect actions faces two practical challenges, namely *(a)* how to parametrize actions on the RT, i.e., how to choose a necessarily finite set of couplings $\{c_\alpha\}$, and *(b)* how to determine $\{c_\alpha^{\rm RT}\}$ or $\{c_\alpha^{\rm FP}\}$. For asymptotically free theories, the latter problem has been solved by Hasenfratz and Niedermayer in Ref. [5]. They observed that in the limit $\beta \to \infty$ (on the critical surface) the integration on the RHS of Eq. (1) is dominated by the minimizing configuration and hence becomes a classical saddle point problem independent of β,

$$A^{\rm FP}[V] = \min_{\{U\}} \left\{ A^{\rm FP}[U] + T[U,V] \right\} \ . \tag{2}$$

They subsequently showed that the corresponding action employed at finite values of β along the straight line emanating from the FP on the critical surface is *classically perfect*. This means that there are no discretization effects when the action is evaluated on configurations fulfilling the classical equations of motions. As a consequence, lattice artifacts of $O(a^{2n})$ are absent for all n. While the FP action is not quantum perfect, i.e., lattice artifacts of $O(g^2 a^{2n})$ are present, it is expected that these effects are suppressed sufficiently close to the critical surface where the couplings of the FP action closely follow the RT. That this is indeed the case has been demonstrated in Ref. [6] where a rich parametrization of the FP action was investigated in Monte Carlo simulations based on the RGT defined in Ref. [3]. The results for the deconfinement phase transition, the static quark-antiquark potential, and the glueball mass spectrum showed only very small lattice artifacts, if any, up to lattice spacings as coarse as $a \sim 0.33$ fm.

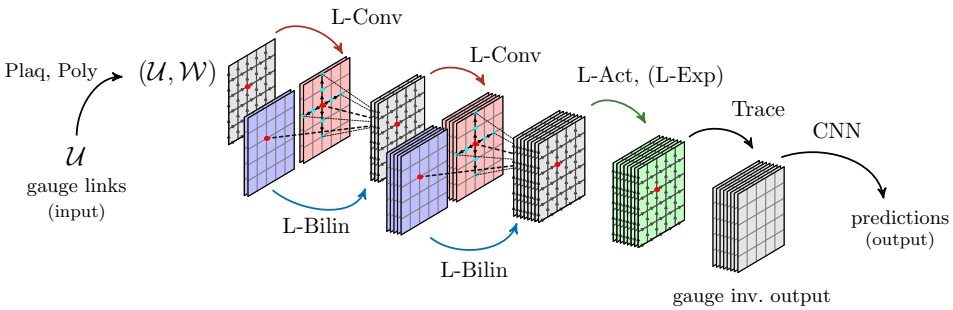


Figure 3: Illustration of a particular lattice gauge-equivariant convolutional neural network (L-CNN) architecture, cf. text and Ref. [7] for further details.

3. Machine learning the FP action

The second challenge in the construction of FP actions is to find a suitable parametrization. Here we make use of the recent developments in machine learning (ML) architectures. In Refs. [7, 8] a lattice gauge-equivariant convolutional neural network (L-CNN) was constructed which is capable of learning any gauge-covariant or gauge-invariant function of gauge fields on a lattice. As such, the architecture is predestined to accurately describe FP actions. The key elements of the L-CNN architecture are the convolutional (L-Conv) and the bilinear (L-Bilin) layers. The L-Conv layer parallel-transports gauge-covariant objects W from a finite region around the lattice site x (the *receptive field*) according to

$$W_i^{\text{L-Conv}}(x) = \sum_{j,\mu,k} \omega_{i,j,\mu,k} U_{k\cdot\mu}(x) W_j(x + k\cdot\hat{\mu}) U^\dagger_{k\cdot\mu}(x),$$

where $U_{k\cdot\mu}(x)$ is the product of gauge links along a path connecting x to $x + k\cdot\hat{\mu}$. The receptive field is determined by the kernel size K with $|k| < K$. The indices i, j label the channels of the data and ω are trainable parameters of the layer. The L-Bilin layer produces new gauge-covariant objects by forming bilinear combinations of two gauge-covariant objects W and W' at lattice site x,

$$W_i^{\text{L-Bilin}}(x) = \sum_{j,j'} \alpha_{i,j,j'} W_j(x) W'_{j'}(x)\,,$$

where α are trainable parameters. Additional layers contain activation functions (L-Act), exponentiation (L-Exp), or tracing (Trace) of the gauge-covariant objects at each lattice site. The input of the L-CNN is a gauge link configuration and the plaquettes at every lattice site. An example of the full architecture is sketched in Figure 3. With this architecture, it is possible to recursively generate combinations of arbitrarily complicated closed loops of gauge links of any shape. Hence, any contribution to the FP action can in principle be generated and described by the L-CNN.

The data for learning the FP action is generated as follows. For a given coarse gauge field configuration V the FP action value is determined by the sequence of minimizing configurations $U, U', \ldots$ according to an inception procedure defined by iterating the FP Eq. (2),

$$A^{\text{FP}}[V] = \min_{\{U\}} \left\{A^{\text{FP}}[U] + T[U,V]\right\} = \min_{\{U',U\}} \left\{A^{\text{FP}}[U'] + T[U',U] + T[U,V]\right\} = \ldots\,.$$

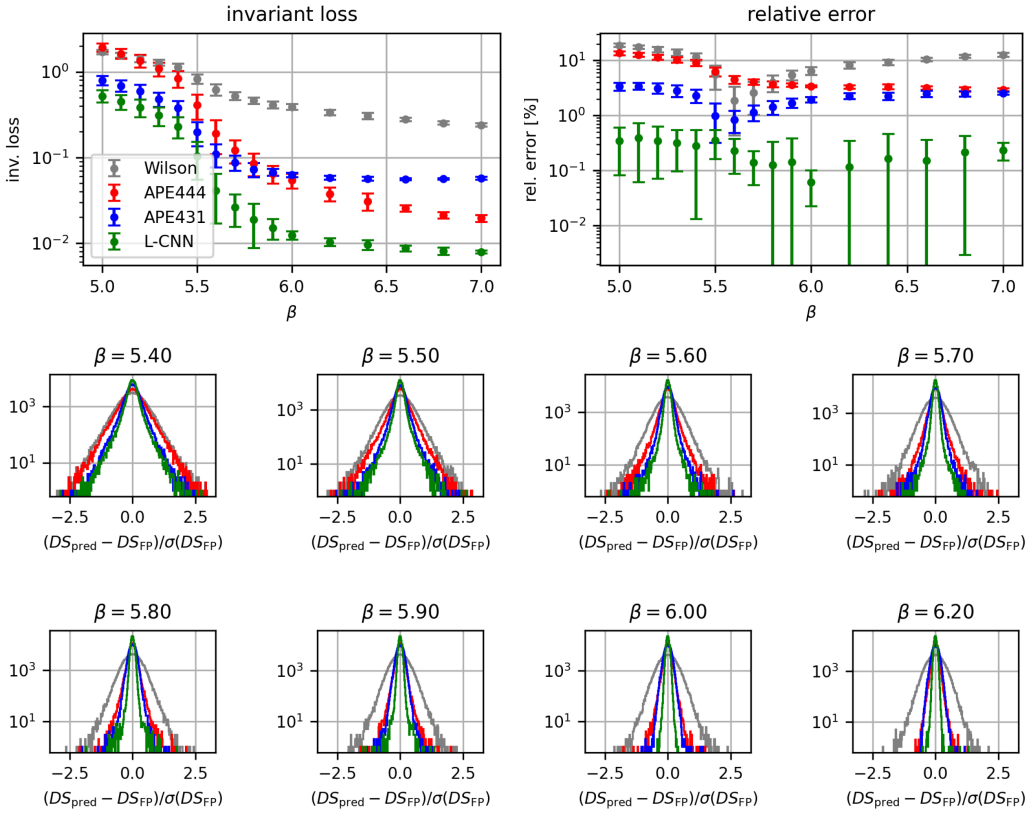


Figure 4: Results for the FP action parametrized by a particular L-CNN model as described in the text. We show the invariant loss (*top left plot*) and the relative error on the action values (*top right plot*) evaluated on the ensembles generated with various values of β using the Wilson action. The plots in the lower two rows show the distributions of the difference between true and predicted derivatives. For comparison, we also show the results for the Wilson gauge action and the APE444 and APE431 parametrizations of the FP action.

The so-obtained exact FP action values $A^{\mathrm{FP}}[V]$ are used for training, testing, and validation. In practice, we only perform one iteration of the procedure and use an existing, sufficiently good parametrization of the FP action on the RHS of Eq. (2) such that the error on $A^{\mathrm{FP}}[V]$ is well controlled.

In addition, we make use of the derivatives of the FP action with respect to the gauge links. They are determined through the FP equation and are given by the derivatives of the blocking kernel,

$$D^{\mathrm{FP}}_{x,\mu,a}[V] \equiv \frac{\delta A^{\mathrm{FP}}[V]}{\delta V^a_{x,\mu}} = \frac{\delta T[U,V]}{\delta V^a_{x,\mu}} = -\kappa\,\mathrm{Re\,Tr}(it^a\,V_{x,\mu}Q^\dagger_{x,\mu})\,,$$

i.e., they implicitly depend on the minimizing configuration through $Q^\dagger_{x,\mu} = Q^\dagger_{x,\mu}[U]$. The derivatives yield $4\times(N_c^2-1)\times L^4$ data per configuration, one data point for each link and color index. Apart from providing an immensely larger amount of FP data for training the L-CNN, the derivatives are particularly suitable in the ML procedure, because they are automatically accessible through the backpropagation process as they are the derivatives of part of the loss function w.r.t. the input. The ML loss function, which is minimized during training, is defined as a weighted sum of

the following two contributions:

$$\mathcal{L}_1 = \frac{1}{L^4}\frac{1}{N_{\text{cfg}}}\sum_i \left|A^{\text{FP}}[V_i] - A^{\text{L-CNN}}[V_i]\right|,$$
$$\mathcal{L}_2 = \frac{1}{8(N_c^2-1)L^4}\frac{1}{N_{\text{cfg}}}\sum_{i,x,\mu,a}(D^{\text{FP}}_{x,\mu,a}[V_i] - D^{\text{L-CNN}}_{x,\mu,a}[V_i])^2.$$

The first expression measures the absolute error of the action density. The second expression measures the error of the derivatives in a gauge invariant way, which we term *invariant loss*. Here, N_{cfg} is the number of configurations while $A^{\text{L-CNN}}$ and $D^{\text{L-CNN}}_{x,\mu,a}$ are the action and derivative values predicted by the L-CNN. The data set we use for the supervised learning is produced by first generating ensembles of coarse gauge field configurations for a large range of fluctuations using the Wilson gauge action with the corresponding sets labeled by the gauge coupling β_{wil}. For each configuration V_i, we then find the configuration U_i minimizing the RHS of Eq. (2) and producing the training data.

4. Results

We are currently in the process of evaluating different L-CNN models and learning strategies. In Fig. 4 we show the results for the FP action parametrized by a particular L-CNN model with three L-Conv and L-Bilin layers containing 12 channels each, parallel transport with $k = \pm 1$ in the first two L-Conv layers, and no transport in the third L-Conv layer. Neither activation layers nor a traditional CNN is used after the final Trace layer. As baselines for comparison, we use the Wilson gauge action, and two older, but rather expressive parametrizations of the FP action denoted by APE444 and APE431. The latter has been optimized to satisfy the FP equation specifically on coarse lattices and has been extensively used and tested in MC simulations [6]. We find that the L-CNN describes the FP action values better than the best old parametrizations over a large range of gauge field fluctuations corresponding to $5.0 \leq \beta_{\text{wil}} \leq 7.0$, and with a relative error by about one order of magnitude smaller. We see a similar improvement in the description of the FP action derivatives.

A crucial point in the practical implementation of the FP program is how local the FP action is, i.e., whether the generated couplings are sufficiently short-ranged. We probe the locality by calculating $\delta^2 A^{\text{L-CNN}}[V]/\delta V^a_{x,\mu}\delta V^b_{y,\nu}$ and forming a suitable, gauge-invariant norm $D^2S(x-y)$. The result on a coarse configuration at $\beta_{\text{wil}} = 5.0$ is shown in Fig. 5. We find that the couplings fall off exponentially, as desired, and take very small values at the distances where the L-CNN parametrization is truncated (by the choice of the number of layers and kernel sizes).

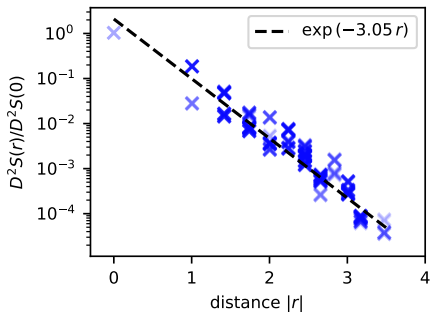


Figure 5: Measure of the gauge link couplings as a function of separation in lattice units for one specific L-CNN architecture including three layers.

5. Summary and conclusions

Using highly improved gauge actions for generating gauge field ensembles holds the promise to overcome both the problems of *critical slowing down* and *topological freezing* when approaching the continuum limit in a gauge field theory. This is achieved by simulating the improved actions on coarse lattices, where both problems are absent, while keeping the lattice artifacts under sufficient control to allow a reliable and solid continuum limit. A radical way to implement this approach is to use RGTs in order to construct *quantum perfect actions* which have no lattice artifacts at all. In a way, this program is similar in spirit to those attempting to construct normalizing flows [9–12] or diffusion models [13, 14] to overcome critical slowing down and topological freezing. In those approaches, the invertible flows generate maps from a trivial or simple distribution of gauge field configurations to a desired target distribution without including any physical information apart from the target distribution. In contrast, the RG approach makes use of the RGT flow in order to inform the map, however, the RG flow is of course not invertible.

While quantum perfect actions have so far been elusive, *classically perfect FP actions* have been constructed and put to use in the past. As such they can immediately be employed in simulations of four-dimensional SU(3) gauge theories in order to overcome the above-mentioned problems. In this work, we revisit the construction of the FP actions and propose to make use of the latest developments in designing L-CNNs and ML techniques. In this context, two crucial questions arise. Firstly, can the FP action be parametrized sufficiently well, or even better than before, with the new L-CNN architectures? Secondly, is the FP action sufficiently local such that any necessary truncation in the couplings is negligible? Both questions are addressed in these proceedings and answered in the affirmative. In fact, it turns out that the L-CNNs are capable of describing the FP actions to a higher accuracy than before and over a much larger range of gauge field fluctuations.

The next task in our program is to investigate how well the L-CNN parametrization of the FP action behaves in actual Monte Carlo simulations and what its scaling properties are. The availability of gauge-link derivatives of the FP action is the stepping stone for these further developments, since both the HMC and the Langevin algorithms, as well as observables based on the gradient flow, make use of derivatives. The ultimate goal would of course be to apply exact RGT steps. The results presented in these proceedings provide a promising basis for further steps in that direction.

Acknowledgments: This work is supported by the US National Science Foundation under Grant No. 2014150, the Austrian Science Fund (FWF) projects P 32446, P 34455 and P 34764, and the AEC and ITP at the University of Bern. The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC) and computing resources at the University of Bern.

References

[1] ALPHA collaboration, S. Schaefer, R. Sommer and F. Virotta, *Critical slowing down and error analysis in lattice QCD simulations*, *Nucl. Phys. B* **845** (2011) 93–119, [1009.5228].

[2] P. Hasenfratz, *The Theoretical background and properties of perfect actions*, in *Advanced Summer School on Nonperturbative Quantum Field Physics*, pp. 137–199, 3, 1998. hep-lat/9803027.

[3] M. Blatter and F. Niedermayer, *New fixed point action for SU(3) lattice gauge theory*, *Nucl. Phys. B* **482** (1996) 286–304, [hep-lat/9605017].

[4] K. Holland, A. Ipp, D. I. Müller and U. Wenger, *(in preparation)*, .

[5] P. Hasenfratz and F. Niedermayer, *Perfect lattice action for asymptotically free theories*, *Nucl. Phys. B* **414** (1994) 785–814, [hep-lat/9308004].

[6] F. Niedermayer, P. Rüfenacht and U. Wenger, *Fixed point gauge actions with fat links: Scaling and glueballs*, *Nucl. Phys. B* **597** (2001) 413–450, [hep-lat/0007007].

[7] M. Favoni, A. Ipp, D. I. Müller and D. Schuh, *Lattice Gauge Equivariant Convolutional Neural Networks*, *Phys. Rev. Lett.* **128** (2022) 032003, [2012.12901].

[8] J. Aronsson, D. I. Müller and D. Schuh, *Geometrical aspects of lattice gauge equivariant convolutional neural networks*, 2303.11448.

[9] 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].

[10] 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].

[11] M. Gerdes, P. de Haan, C. Rainone, R. Bondesan and M. C. N. Cheng, *Learning Lattice Quantum Field Theories with Equivariant Continuous Flows*, 2207.00283.

[12] 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].

[13] L. Wang, G. Aarts and K. Zhou, *Diffusion Models as Stochastic Quantization in Lattice Field Theory*, 2309.17082.

[14] 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.

PoS(LATTICE2023)038