Neural Markov Chain Monte Carlo

Piotr Korcyl

1 Motivation and significance of the project

Computational sciences constitute nowadays one of the driving force towards scientific discovery. As the numerical simulations and computations become more and more realistic and sophisticated the required computer resources increase. Today's simulations are possible thanks to the simultaneous joint progress in efficient computer architectures, algorithms and supercomputer facilities. Nevertheless, further breakthroughs need new approaches and new ideas to increase the algorithmic performance.

In order to be more specific we take the example of *ab initio* elementary particle physics simulations which are based on Markov Chain Monte Carlo type of algorithms. Often calculating expectation values of physical observables corresponds to evaluating highly dimensional integrals. The latter in the spirit of Monte Carlo integration requires sampling from some very complicated probability distribution. This can be realized exactly using Markov Chains with a carefully chosen transition probabilities. Such approach turned out to be extremely successful as it guarantees exactness based on results in the theory of stochastic processes, reproducibility and robust handling of uncertainties. Monte Carlo simulations in different branches of physics [4, 5]: statistical physics, solid state physics, astroparticle physics [3], elementary particle physics and Beyond the Standard Model physics [6] but also in chemistry, economy have yielded a great amount of new results.

The computational cost of such simulations is most often proportional to the cost of generating new, statistically independent samples. As we approach the continuum limit in our discretized simulation the cost of the new independent sample increases significantly which is a well known phenomenon called *critical slowing down*. Also, when the simulated system exhibits different phases with different symmetry properties or has non-local excitations, for instance topological kind of objects, the sampling becomes ineffective and the simulations tend to be trapped in only one of the phases. Autocorrelations in such situation may increase and the computational cost rises by factors of 100 or more.

In this project we tackle this fundamental problem with Artificial Intelligence inspired algorithms. We propose to investigate the effectiveness of typical sampling algorithms where the proposals for new samples are generated by neural networks. In the last two years it was shown in a couple of pioneering papers that the incorporation of a Variational Autoregressive Neural (VAN) network into the Hastings-Metropolis algorithm in the case of the 2D Ising model has a large impact on the effective autocorrelation time Ref. [9, 10]. This approach was named Neural Markov Chain Monte Carlo (NMCMC) and was further generalized to more complex systems, such as a U(1) gauge theory in two dimensions in Ref. [7]. The importance of this research direction can be demonstrated by the fact that the leading American universities, MIT, Harvard, Northeastern and Tufts have recently launched a dedicated research institute: Institute for Artificial Intelligence and Fundamental Interactions (IAIFI) [8]. One of the three research topics developed at IAIFI is the described above *Ab Initio AI for Theory Calculations: Accelerating Lattice Field Theory with AI.*

This research direction is exactly aligned with the main topics of DigiWorld. Specifically, it touches the domains **1. Advanced computational methods and Artificial Intelligence** and **4. Artificial Intelligence in exact and natural sciences**.

2 Expected results

In this project we will consider the Ising model in two-dimensions on a square lattice of size $N \times N$. A particular configuration of spins $\mathbf{s}$ occurs with the Boltzmann probability

$$p(\mathbf{s}) = \frac{1}{Z} e^{-\beta H(\mathbf{s})} \tag{1}$$

We construct a neural network which approximates the distribution p by a distribution q_θ. The probability $q_\theta(\mathbf{s})$ of a given spin configuration depends on the weights (parameters) of the neural network θ. To model the distribution probability p we use a VAN neural network. We label all spins and we express the conditional probability of spin s_i in terms of probabilities of previous spins $s_{<i}$ [9] as,

$$q_\theta(\mathbf{s}) = \prod_{i=1}^{N} q_\theta(s_i|s_1, \ldots, s_{i-1}) \tag{2}$$

In practice, the neural network is given on input a spin configuration $\mathbf{s}$ for which it evaluates $\log q_\theta(\mathbf{s})$ [9].

This construction has a great advantage that it can be used to generate a spin configuration $\mathbf{s}$ which follows the probability distribution q_θ. This can be easily done by reverting the conditional probabilities. Having the configuration, we estimate its probability $\log q_\theta(\mathbf{s})$ and compare it with the exact probability $p(\mathbf{s})$. This comparison can be used to correct the weights θ.

Following the ideas of Ref. [10], the spin configuration proposed by the neural network, which follows probability distribution q_θ is subjected to an accept/reject step according to the following rule

$$\min\left(1, \frac{q_\theta(\mathbf{s}')e^{-\beta H(\mathbf{s})}}{q_\theta(\mathbf{s})e^{-\beta H(\mathbf{s}')}}\right) \tag{3}$$

where the spin configuration $\mathbf{s}'$ is the previous analyzed configuration.

We have implemented this method and performed initial tests in the course of two Bachelor thesis [1, 2] performed in 2020 at the Jagiellonian University. Figure 1 demonstrates the data for the mean magnetization obtained in three different ways:

1. Hastings-Metropolis algorithm
2. algorithm proposed in Ref. [9]
3. NMCMC algorithm proposed in Ref. [10]

We demonstrated that the differences between the traditional approach 1 and 3 are all compatible with 0 showing that the sampling is correct in the entire range of temperatures investigated. In figure 2 we present the autocorrelation functions. We see a sharp drop of these functions indicating a short leading autocorrelation time. However, we also discovered a very large, subleading autocorrelation time. It was not noticed in the cited references [9, 10] and its origin is for now unknown. It is the aim of this minigrant to deepen these studies and clarify and quantify the autocorrelations induced by the VAN neural network and the accept/reject step.

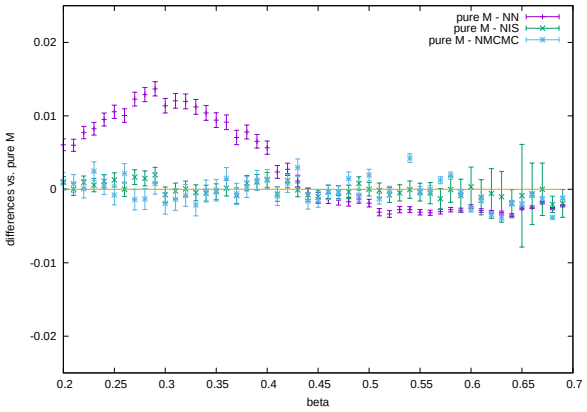

Figure 1: Differences of magnetization obtained for different values of β for a 4^2 lattices. Figure taken from K. Szota Bachelor Thesis, Ref. [1].

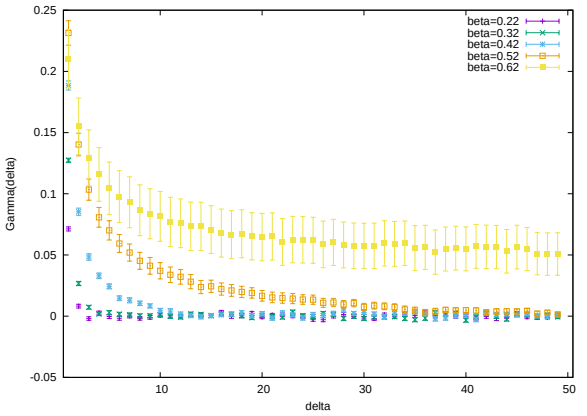

Figure 2: Autocorrelation functions for different values of β for a 4^2 lattice. Figure taken from S. Sauer Bachelor Thesis, Ref. [2].

In the course of this short grant we propose to continue these studies and thoroughly investigate the autocorrelation times of consecutive samples generated by the NMCMC approach for the case of two-dimensional Ising model. We propose to extend the results obtained in Ref. [10] by studying this problem

analytically by analyzing semi-numerically the transfer matrix of the associated Markov Chain. Analytic results will be confirmed by numerical simulations. The expected results will allow to understand the impact of the accept/reject step on the autocorrelation time. Furthermore, we propose to consider a toy-model taken from the studies of topological quantum field theory where states have different topological properties and therefore for which the traditional Markov Chain Monte Carlo would not be applicable as the algorithm is not able to introduce global changes of topology during the local changes in the sampling procedure. A neural network enhanced algorithm may be able to overcome this limitation, thus opening a completely new branch of models suitable to numerical simulations

The results of this project (published in academic journals and in form of internal notes) will be used as a base for a NCN grant application.

3 Innovation of the proposed project

This project is innovative as it leads to a definition of an improved simulation algorithm allowing to generate samples from a given distribution probability with small autocorrelation times.

4 Interdisciplinary nature of the project

Computational sciences are at the crossroads of natural and exact sciences. Advances in simulation algorithms have immediate consequences in different branches of science: chemistry, physics, economy, and others, everywhere where stochastic sampling is used. In the project we will use as a test-bed for our investigations simple models from statistical physics and topological quantum field theory, however results will have deep consequences for application to solid state physics, atomic physics or elementary particle physics. Our team is composed of physicists and computer scientists and spans over the Institute of Theoretical Physics WFAIS and Institute of Applied Computer Sciences WFAIS.

5 International collaboration

The work will be performed in collaboration with Salvatore Cali, a postdoc at Massachussetts Institute of Technology, Cambridge, USA (MIT) from January 2021 in the Lattice Quantum Chromodynamics group of Phiala Shanahan. Currently Salvatore Cali is a postdoc at the Jagiellonian University. Our preprint [11] was submitted to Phys. Rev. Lett. and is in the second stage of the refereeing process. Phiala Shanahan was one of the authors of the generalization of the AI inspired Markov Chain Monte Carlo proposals for the U(1) gauge theory published in Phys. Rev. Lett. [7]. She is also employed at the IAIFI.

References

[1] K. Szota, , Bachelor Thesis 2020, Jagiellonian University,

[2] S. Sauer, , Bachelor Thesis 2020, Jagiellonian University,

[3] S. Sharma, *Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy*, Annual Review of Astronomy and Astrophysics 55 (2017) 213,

[4] S. Brooks, A. Gelman, G. Jones, X. Meng, *Handbook of Markov Chain Monte Carlo*, Chapman&Hall, 2011

[5] W. Gilks, S. Richardson, D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman&Hall, 1998

[6] A. Joseph, *Markov Chain Monte Carlo Methods in Quantum Field Theories*, SpringerBriefs in Physics, 2020

[7] G. Kanwar, M. Albergo, D. Boyda, K. Cranmer, D. Hackett, S. Racanière, D. Rezende, P. Shanahan, *Equivariant Flow-Based Sampling for Lattice Gauge Theory*, PhysRevLett. 125 (2020) 121601,

[8] https://iaifi.org/

[9] D. Wu, L. Wang, P. Zhang, *Solving Statistical Mechanics Using Variational Autoregressive Networks*, arXiv:1809.10606v2 [cond-mat.stat-mech]

[10] K. Nicoli, S. Nakajima, N. Strodthoff, K.-R. Mueller, P. Kessel, *Asymptotically Unbiased Estimation of Physical Observables with Neural Samplers*, arXiv:1910.13496v2 [cond-mat.stat-mech]

[11] S. Cali, K. Cichy, P. Korcyl, J. Simeth, *Running coupling constant from position-space current-current correlation functions in three-flavor lattice QCD*, arxiv:2003.05781