Generative models for sampling of lattice field theories

Matija Medvidović
Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA
Department of Physics, Columbia University, New York 10027, USA
matija.medvidovic@columbia.edu

Juan Carrasquilla
Vector Institute for Artificial Intelligence, MaRS Centre, Toronto, Ontario, Canada
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
carrasqu@vectorinstitute.ai

Lauren E. Hayward
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5, Canada
lhayward@perimeterinstitute.ca

Bohdan Kulchytskyy
Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, N2L 2Y5, Canada
bkulchytskyy@uwaterloo.ca

Abstract

We explore a self-learning Markov chain Monte Carlo method based on the Adversarial Non-linear Independent Components Estimation Monte Carlo, which utilizes generative models and artificial neural networks. We apply this method to the scalar φ^4 lattice field theory in the weak-coupling regime and, in doing so, greatly increase the system sizes explored to date with this self-learning technique. Our approach does not rely on a pre-existing training set of samples, as the agent systematically improves its performance by bootstrapping samples collected by the model itself. We evaluate the performance of the trained model by examining its mixing time and study the ergodicity of generated samples. When compared to methods such as Hamiltonian Monte Carlo, this approach provides unique advantages such as the speed of inference and a compressed representation of Monte Carlo proposals for potential use in downstream tasks.

1 Introduction

Field theory is the theoretical bedrock for unified descriptions of critical phenomena. Such a framework provides us with a microscopic understanding of universality [1]. Its predictions for a variety of strongly correlated systems have been confirmed experimentally to high precision in widely diverse physical systems [2], including thin films of superfluids [3], superconductors [4], ferromangets [5], quantum simulations of frustrated magnets [6], and ultracold atoms [7]. In practice, however, field theories are often intractable and their analytical treatment involves approximations with varying degrees of reliability. Luckily, some of these field theories lend themselves to non-perturbative numerical simulations. These simulations often utilize Markov Chain Monte Carlo (MCMC) algorithms

based on a Feynman path-integral formulation of the theory [8], tensor networks [9–11], or even quantum computing [12].

A key practical concern in MCMC simulations is the autocorrelation that exists between Monte Carlo samples [13]. Reducing the autocorrelation time enables a Markov chain to become shorter while maintaining the same statistical predictive capacity. Such optimization can be achieved through a tailored design of the proposal distribution in an MCMC update. The best proposals dramatically reduce the autocorrelation and thus the computation time required to reach a desired accuracy.

Recently, there has been progress toward an automatic optimization of the proposal distribution in MCMC simulations applied to physical systems. Specifically, the proposal distribution is parameterized as a generative model designed for an inexpensive generation of statistically independent samples, such as a generative adversarial network (GAN) [14, 15] or a flow model [16, 17]. An important ingredient in such optimization is the choice of a loss function that ultimately leads to an efficient sampler. For example, Ref. [15] takes a data-driven approach relying on a GAN's loss-function evaluated over a pre-existing data-set of samples. Contrarily, Ref. [17] takes a variational approach that aims to minimize the action for the samples generated from the model.

Here we examine a general approach for a self-training MCMC known as Adversarial Nonlinear Independent Components Estimation Monte Carlo (A-NICE MC) [18], where a neural network is optimized to minimize the autocorrelation in a Markov chain. The network is trained on samples generated by the model itself and, like most MCMC schemes, requires only the analytical lattice action as input. By extending the A-NICE MC method to sample high-dimensional lattice field theories, we find that this method scales well beyond previously tested target space dimensionalities. Using A-NICE MC on the scalar field φ^4 theory [19] proves that the model can be systematically optimized towards producing decorrelated samples using standard gradient-based optimizers. The φ^4 theory was chosen because of its simplicity. In addition, we study the performance of A-NICE MC by examining its ability to discriminate between random Gaussian noise and actual samples from the weakly coupled theory. We find that the model learns to distinguish between the two and that we can observe the distinction in the training procedure. While the training procedure for A-NICE MC is considerably more expensive than Hamiltonian Monte Carlo (HMC), the strategy has the advantage that a trained A-NICE MC model can simply be stored, shared, and sampled when needed, as opposed to HMC where the result of the calculations is a collection of samples.

2   Methods

We focus on obtaining lattice field samples $\{\varphi^{(i)}\}$ from a Boltzmann-type distribution of the form $P(\varphi) \propto \exp(-S(\varphi))$. The distribution is defined on a square $L \times L$ lattice $\Lambda = \mathbb{R}^{L^2}$, where each sample $\varphi \in \Lambda$. Metropolis-Hastings (MH) [20, 21] MCMC algorithms propose each new sample $\varphi^{(i+1)}$ from $\varphi^{(i)}$ through a predefined proposal function f. Each $\varphi^{(i+1)}$ is either accepted and appended to the Markov Chain, or rejected such that $\varphi^{(i)}$ is appended to the chain instead. Both detailed balance and ergodicity are required of the mapping $\varphi^{(i)} \mapsto \varphi^{(i+1)}$. The details of the MH algorithms can be found in Ref. [13]. Ideally, this probabilistic mapping is chosen such that the statistical correlation between adjacent samples is minimized.

An important and widely used algorithm in the study of lattice quantum field theory is the HMC algorithm [22]. This method supplements the original degrees of freedom φ with fictitious conjugate momentum variables π and introduces a combined Hamiltonian on the same lattice Λ defined as $H_\Lambda(\varphi, \pi) = \frac{1}{2} \sum_{\mathbf{x} \in \Lambda} \pi_{\mathbf{x}}^2 + S_\Lambda(\varphi)$. A state proposal $(\varphi, \pi)^{(i+1)}$ is then obtained by integrating Hamilton's equations for a finite time. Upon integrating out the momentum variables in the distribution $P_{H_\Lambda}(\varphi, \pi) \propto \exp(-H_\Lambda)$, one recovers samples from the desired $P(\varphi)$. Since the Hamiltonian dynamics generate a volume-preserving and time-reversible flow of states, the proposal distribution is symmetric, implying that the rate of rejection of proposed samples is determined by the accuracy of the numerical integrator [22].

A-NICE MC takes inspiration from HMC and embeds its proposal distribution in the augmented space $f : (\varphi, \pi)^{(i)} \to (\varphi, \pi)^{(i+1)}$ where the supplementary degrees of freedom π are Gaussian distributed. Furthermore, A-NICE exploits the invertible and volume-preserving properties of the NICE [23] architecture to parametrize f. As a result, the proposal distribution is symmetrical and does not rely on expensive integration as in HMC.

The NICE proposal f_θ, which is parametrized by θ, is trained alongside an adversary to systematically minimize autocorrelation times. This setup closely resembles a more traditional GAN [14, 24, 25] but retains a few key differences. Most notably, a pairwise discriminator network is used in place of a cost function. Instead of ranking each sample individually as for a more conventional discriminator, the pairwise discriminator ranks pairs of samples together, acting as a proxy for the auto-correlation function between samples. The discriminator is parametrized as a deep neural network and represents a mapping $D_\alpha : \mathbb{R}^{2n} \to \mathbb{R}$ with parameters α. Intuitively, the discriminator takes a pair of samples and is trained to score the degree of their correlation. The objective is to produce a NICE proposal that generates a Markov chain with two desirable properties. Firstly, starting from noise, the proposal should generate a highly-probable sample in a small number of steps such that equilibrium is reached quickly. Secondly, starting from a highly-probable sample, the proposal should generate another decorrelated and probable sample in as few steps as possible.

We employ a bootstrapping procedure to generate a training dataset. The untrained network is initially sampled to obtain a starting dataset $\mathcal{D}_0$, which is expected to exhibit large autocorrelation times. Still, the MH algorithm biases the sampling towards the correct distribution regardless of the proposal [13]. We collect samples from multiple chains and shuffle them to further reduce hidden correlations in the training set. After $N_{\text{bootstrap}}$ training loop iterations using $\mathcal{D}_0$, some fraction $r \in [0, 1]$ of samples $\mathcal{D}_0$ are replaced by new data, yielding $\mathcal{D}_1$. This new dataset is less correlated since it comes from the partially trained model. Iterating this procedure, we obtain samples of increasing quality in each subsequent dataset $\mathcal{D}_i$. In principle, we reach the fixed point of this training procedure when the samples generated by f_θ become as decorrelated as the shuffled samples from the bootstrapped dataset $\mathcal{D}_i$.

The scalar φ^4 theory on a lattice was chosen as the simplest non-Gaussian theory to test the effectiveness of the A-NICE MC approach. Despite its simplicity, this theory has a nontrivial phase diagram and a variety of use cases across particle and many-body physics. As noted earlier, in this case, samples consist of only one real number per lattice site. For the φ^4 lattice field theory, the probability distribution $P_\Lambda(\varphi) \propto e^{-S_\Lambda(\varphi)}$ is defined by the action

$$S_\Lambda(\varphi) = \sum_{\mathbf{x} \in \Lambda} \left[-2\kappa \sum_{\mu=1}^{D} \varphi_{\mathbf{x}} \, \varphi_{\mathbf{x}+\hat{\mathbf{e}}_\mu} + (1 - 2\lambda) \, \varphi_{\mathbf{x}}^2 + \lambda \, \varphi_{\mathbf{x}}^4 \right], \tag{1}$$

where the coupling constants $\kappa, \gamma \in \mathbb{R}$ define different distributions $P_\Lambda(\varphi)$ and determine the phase diagram of the model. We choose $\kappa = 0.21$ and $\lambda = 0.022$, for which the model is in the weak-coupling regime and displays behavior similar to the paramagnetic Ising model [26]. The phase diagram of the theory associated with Eq. 1 was explored in Ref. [27] using standard MCMC algorithms. Our choice of parameters enables us to tackle our goal of exploring novel MCMC frameworks for a non-Gaussian model, while eliminating any strong-coupling effects. Remaining in the weakly-coupled regime allows us to study A-NICE MC performance in the region of parameter space easily reachable by conventional methods which allows for a straightforward comparison of results.

3 Results

It is difficult a priori to verify whether the A-NICE framework will display ergodic behavior since it is stochastically optimized, highly nonlinear, and contains a large number of free parameters. We thus compare our A-NICE algorithm against well-established HMC simulations. First, we take the generated HMC chain and randomly pair the samples (to decorrelate them artificially) in order to examine the output of the fully-trained A-NICE pairwise discriminator. The results are shown in Fig. 1. We observe that the A-NICE and HMC output distributions overlap, which means that the A-NICE distribution is consistent with the behavior of the HMC distribution in that the produced samples are not penalized differently by the discriminator and pushed outside the bulk of the distribution. The fact that all HMC samples being scored similarly suggests that the discriminator had seen all such samples from the NICE proposal during training. We supplement the plot with a reference point provided by samples randomly chosen from a unit Gaussian distribution. Since we are studying our lattice model well within the weakly coupled regime, we expect our field samples to be almost Gaussian and thus penalized in a similar way. We plot the discriminator scores of HMC, A-NICE and the Gaussian random noise against the training iteration in the right panel of Fig. 1. Differences
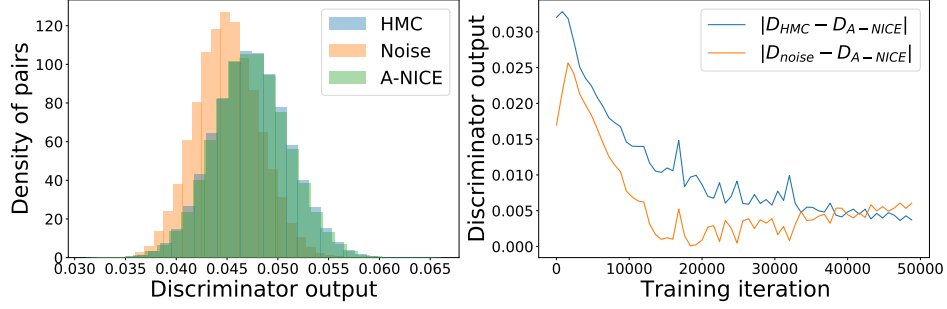
Figure 1: (left) The fully-trained discriminator score distribution of randomly shuffled HMC, A-NICE and (unit) Gaussian random noise for an 8×8 lattice. All three histograms have been normalized to unity. The noise was chosen with the mean and the variance estimated from HMC samples. (right) The mean value of the same three scores with respect to the training step and averaged over 32 independent chains.
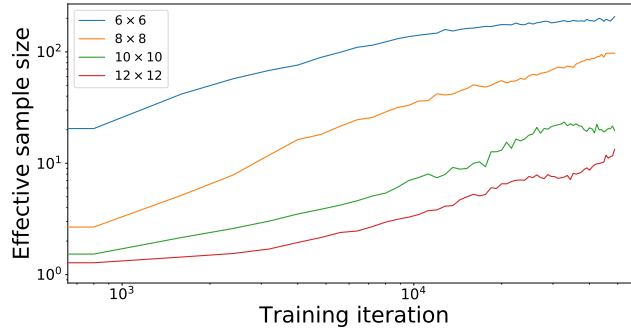

Figure 2: Effective sample size (ESS) as a function of training iteration. The increasing ESS demonstrates A-NICE MC's ability to systematically improve the statistical quality metrics of the Markov chain produced by the model.

between discriminator scores of A-NICE samples and HMC/random samples are shown, where the HMC samples are randomly shuffled to artificially eliminate autocorrelations and provide reliable reference values. We see that, as the discriminator optimizes [24], the difference $|D_{\text{HMC}} - D_{\text{A-NICE}}|$ decreases and thus the A-NICE score approaches the decorrelated HMC score. We perform error analysis on the fully-trained distributions in the left panel of Fig. 1 and find that $|D_{\text{HMC}} - D_{\text{A-NICE}}| = (416.7 \pm 1.4) \times 10^{-5}$, which is less than $|D_{\text{noise}} - D_{\text{A-NICE}}| = (433.0 \pm 1.5) \times 10^{-5}$. Therefore, we conclude that A-NICE MC learns to distinguish between random noise and a weakly-coupled φ^4 theory, since the A-NICE samples become less similar to random noise and become more similar to decorrelated HMC samples by the end of the training.

In Fig. 2, we study the systematic improvement of sample quality during the training process. Specifically, we plot the effective sample size (ESS), which is inversely proportional to the integrated autocorrelation time τ [13] of the chain (ESS $= N/2\tau$, where N is the number of samples) and approximately represents the effective number of independent samples in a given chain. We expect that additional time and computational resources would continue to improve the model.

Next, we study the behaviour of some observables, including the magnetization $M \equiv \langle \frac{1}{L^D} \sum_{\mathbf{x} \in \Lambda} \varphi_{\mathbf{x}} \rangle$, the two point-susceptibility $\chi_2 \equiv \sum_{\mathbf{x} \in \Lambda} G_2(\mathbf{x})$ (where $G_2(\mathbf{x}) \equiv \frac{1}{L^D} \sum_{\mathbf{y} \in \Lambda} [\langle \varphi_{\mathbf{y}} \varphi_{\mathbf{x}+\mathbf{y}} \rangle - \langle \varphi_{\mathbf{y}} \rangle \langle \varphi_{\mathbf{x}+\mathbf{y}} \rangle]$), and the Ising energy density $E_I \equiv \lim_{\lambda \to \infty} \frac{1}{D} \sum_{\mu} G_2(\hat{\mathbf{e}}_{\mu})$. As illustrated in Fig. 3, A-NICE outperforms HMC for most observables of interest since the A-NICE autocorrelation curves decay faster, except for the case of the average magnetization. Note that HMC can be tuned to improve performance, but usually at the cost of increased computational complexity. Since HMC constructs proposals by integrating differential equations, more steps would be required each time. Thus, the proposal would be further away from
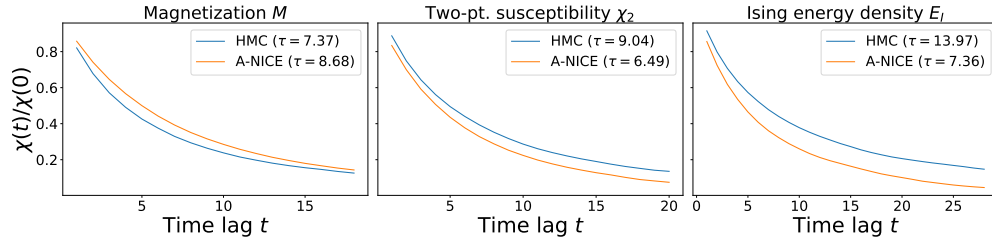
Figure 3: Comparisons of common physical observables as a function of the time difference (lag) t between two MCMC samples for $L = 10$. Each y-axis shows the normalized statistical autocorrelation $\chi(t) = \int_0^\infty dt' \left[\mathcal{O}(t')\mathcal{O}(t + t') - \langle \mathcal{O} \rangle^2 \right]$ for a different observable $\mathcal{O}$. Integrated autocorrelation times [13] for both methods and each observable are reported in the figure legends.

the current state, lowering autocorrelations and decreasing acceptance rates. A-NICE doesn't force such a trade-off but offers no control over acceptance rates in return. For HMC, we take 10 leapfrog steps for each proposal and the step size is tuned so that the acceptance rate matches A-NICE MC.

4 Conclusion

We have applied the A-NICE MC [18] algorithm to the lattice φ^4 field theory in the weakly-coupled regime, increasing the number of dimensions that have been successfully sampled using this method to date. The model can be systematically optimized and produces samples which are consistent with an ergodic exploration of the state-space. The bulk of the computational cost for employing this technique is associated with the training phase. Deployment of the trained model is relatively inexpensive compared to the HMC method tuned to the same performance level. This sampling efficiency of the trained model effectively eliminates the need to store the samples, which cuts down on the storage space requirements and allows us to extract physical observables with low statistical uncertainties.

In future work, it will be interesting to extend our experiments with A-NICE MC to the strongly-interacting regime of the φ^4 field theory as well as to apply this framework to more complicated theories with additional internal degrees of freedom or gauge symmetries. We conjecture that A-NICE MC can be extended to deal with these models with possible modifications required to efficiently explore the state space in a systematic domain-specific way.

Impact statement

Our Adversarial Nonlinear Independent Components Estimation Monte Carlo (A-NICE MC) approach to sampling focuses on training a neural network to be embedded into Markov chain Monte Carlo (MCMC) sampling algorithms. A-NICE MC moves most of the computational cost of MCMC sampling to network training, making sampling itself relatively less expensive compared to established MCMC approaches.

We expect that further improvements to the adversarial network architecture are possible such that the sampling can continue to become even more efficient in the future. Improved efficiency in MCMC sampling schemes has the potential to greatly reduce the overall carbon footprint of high-performance computational (HPC) science. Lower autocorrelation times mean fewer sampling steps which, in turn, mean more effective samples per watt of consumed power. Therefore, we expect that, given further optimizations, machine-learning-based sampling methods such as A-NICE MC have the potential to significantly reduce the scientific carbon footprint.

References

[1] J. Cardy, *Scaling and Renormalization in Statistical Physics*. Cambridge Lecture Notes in Physics. Cambridge University Press, 1996.

[2] S. Sachdev, *Quantum Phase Transitions*. Cambridge University Press, Apr., 2001.

[3] F. M. Gasparini, M. O. Kimball, K. P. Mooney, and M. Diaz-Avila, "Finite-size scaling of ^{4}He at the superfluid transition," *Rev. Mod. Phys.* **80** (Sep, 2008) 1009–1059. https://link.aps.org/doi/10.1103/RevModPhys.80.1009.

[4] B. I. Halperin and D. R. Nelson, "Resistive transition in superconducting films," *Journal of Low Temperature Physics* **36** no. 5, (Sep, 1979) 599–616. https://doi.org/10.1007/BF00116988.

[5] R. Coldea, D. A. Tennant, E. M. Wheeler, E. Wawrzynska, D. Prabhakaran, M. Telling, K. Habicht, P. Smeibidl, and K. Kiefer, "Quantum Criticality in an Ising Chain: Experimental Evidence for Emergent E $_8$ Symmetry," *Science* **327** no. 5962, (Jan., 2010) 177–180.

[6] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Fréchette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato, A. Y. Smirnov, L. J. Swenson, M. H. Volkmann, J. Whittaker, J. Yao, E. Ladizinsky, M. W. Johnson, J. Hilton, and M. H. Amin, "Observation of topological phenomena in a programmable lattice of 1,800 qubits," *Nature* **560** no. 7719, (Aug., 2018) 456.

[7] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch, "Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms," *Nature* **415** no. 6867, (Jan., 2002) 39–44.

[8] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics," *Rev. Mod. Phys.* **20** (Apr, 1948) 367–387. https://link.aps.org/doi/10.1103/RevModPhys.20.367.

[9] J. Haegeman, T. J. Osborne, H. Verschelde, and F. Verstraete, "Entanglement renormalization for quantum fields in real space," *Phys. Rev. Lett.* **110** (Mar, 2013) 100402. https://link.aps.org/doi/10.1103/PhysRevLett.110.100402.

[10] M. Ganahl, J. Rincón, and G. Vidal, "Continuous matrix product states for quantum fields: An energy minimization algorithm," *Phys. Rev. Lett.* **118** (Jun, 2017) 220402. https://link.aps.org/doi/10.1103/PhysRevLett.118.220402.

[11] F. Verstraete and J. I. Cirac, "Continuous matrix product states for quantum fields," *Phys. Rev. Lett.* **104** (May, 2010) 190405. https://link.aps.org/doi/10.1103/PhysRevLett.104.190405.

[12] J. Preskill, "Simulating quantum field theory with a quantum computer," *arXiv:1811.10085 [hep-lat, physics:hep-th, physics:quant-ph]* (Nov., 2018) , arXiv:1811.10085 [hep-lat, physics:hep-th, physics:quant-ph].

[13] M. E. J. Newman and G. T. Barkema, "Monte Carlo Methods in Statistical Physics," *Oxford University Press* (1999) .

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks,". http://arxiv.org/abs/1406.2661.

[15] J. M. Urban and J. M. Pawlowski, "Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks,". http://arxiv.org/abs/1811.03533.

[16] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP,". http://arxiv.org/abs/1605.08803.

[17] 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** (Aug, 2019) 034515. https://link.aps.org/doi/10.1103/PhysRevD.100.034515.

[18] J. Song, S. Zhao, and S. Ermon, "A-NICE-MC: Adversarial Training for MCMC,". http://arxiv.org/abs/1706.07561.

[19] S. Weinberg, *The Quantum Theory of Fields*, vol. 1. Cambridge University Press, 1995.

[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics* **21** no. 6, (1953) 1087–1092.

[21] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika* **57** no. 1, (1970) 97–109.

[22] R. M. Neal, "MCMC using Hamiltonian dynamics,". `http://arxiv.org/abs/1206.1901`.

[23] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear Independent Components Estimation,". `http://arxiv.org/abs/1410.8516`.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in *Proceedings of the 34th International Conference on Machine Learning*. 2017.

[25] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Advances in Neural Information Processing Systems*, pp. 2180–2188. 6, 2016. `http://arxiv.org/abs/1606.03657`.

[26] M. Kardar, *Statistical Physics of Fields*. Cambridge University Press, 1 ed., 2007.

[27] A. K. De, A. Harindranath, J. Maiti, and T. Sinha, "Investigations in 1+1 dimensional lattice phi ^4 theory," *Physical Review D - Particles, Fields, Gravitation and Cosmology* **72** no. 9, (6, 2005) s. `http://arxiv.org/abs/hep-lat/0506002`.