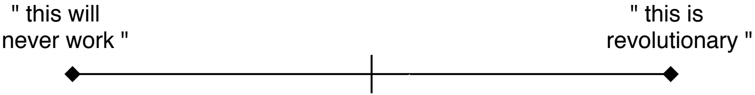
Prospects of Lattice Field Theory Simulations powered by Deep Neural Networks

Julian Urban
ITP Heidelberg
2019/11/06

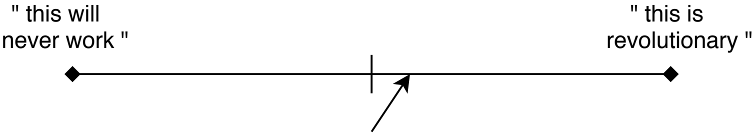
Prospects of Lattice Field Theory Simulations powered by Deep Neural Networks

Julian Urban
ITP Heidelberg
2019/11/06



Prospects of Lattice Field Theory Simulations powered by Deep Neural Networks

Julian Urban
ITP Heidelberg
2019/11/06



Overview

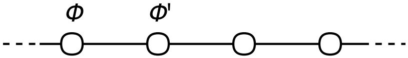
- Stochastic estimation of Euclidean path integrals
- Overrelaxation with Generative Adversarial Networks (GAN)*
- Ergodic sampling with Invertible Neural Networks (INN)†
- Some results for real, scalar ϕ^4-theory in $d = 2$

* Urban, Pawlowski (2018) — "Reducing Autocorrelation Times in Lattice Simulations with Generative Adversarial Networks" — arXiv: 1811.03533

† Albergo, Kanwar, Shanahan (2019) — "Flow-based generative models for Markov chain Monte Carlo in lattice field theory" — arXiv: 1904.12072

Markov Chain Monte Carlo

$$\langle \mathcal{O}(\phi) \rangle_{\phi \sim e^{-S(\phi)}} = \frac{\int \mathcal{D}\phi \; e^{-S(\phi)} \, \mathcal{O}(\phi)}{\int \mathcal{D}\phi \; e^{-S(\phi)}} \cong \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(\phi_i)$$



- accept ϕ' with probability:

$$T_A(\phi'|\phi) = \min \left(1, \; e^{-\Delta S}\right)$$

- autocorrelation function:

$$C_{\mathcal{O}}(t) = \langle \mathcal{O}_i \mathcal{O}_{i+t} \rangle - \langle \mathcal{O}_i \rangle \langle \mathcal{O}_{i+t} \rangle$$

Real, Scalar ϕ^4-Theory on the Lattice

- $\phi(x) \in \mathbb{R}$ discretized on d-cubic Euclidean lattice with volume $V = L^d$ and periodic boundary conditions

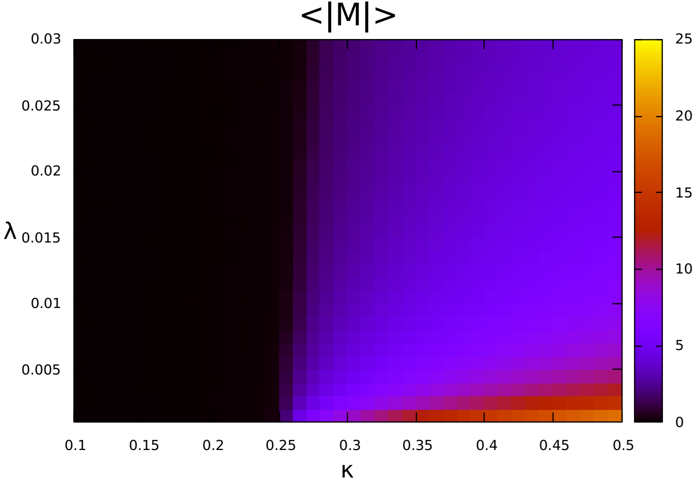
$$S = \sum_x \left[-2\kappa \sum_{\mu=1}^{d} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\,\phi(x)^2 + \lambda\,\phi(x)^4 \right]$$

- magnetization $M = \frac{1}{V} \sum_x \phi(x)$
- connected susceptibility $\chi_2 = V \left(\langle M^2 \rangle - \langle M \rangle^2 \right)$
- connected two-point correlation function

$$G(x,y) = \langle \phi(x)\phi(y) \rangle - \langle \phi(x) \rangle \langle \phi(y) \rangle$$

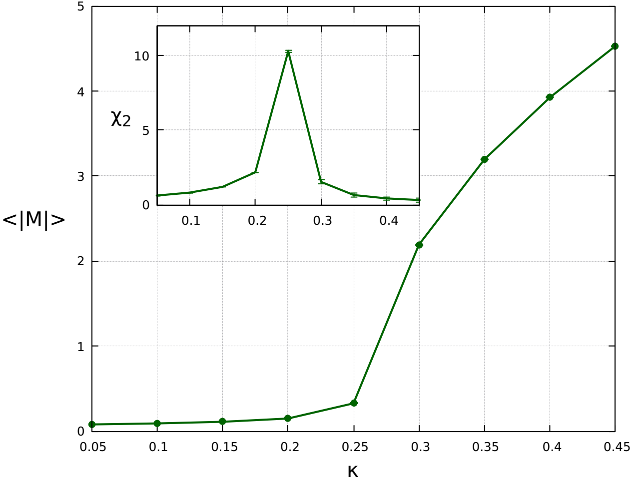
Real, Scalar ϕ^4-Theory on the Lattice

$d = 2$



Real, Scalar ϕ^4-Theory on the Lattice

$$d = 2,\ V = 8^2,\ \lambda = 0.02$$



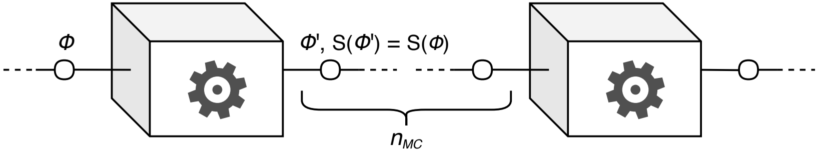
Independent (Black-Box) Sampling

Replace $p(\phi)$ by an approximate distribution $q(\phi)$ generated from a function $g : \mathbb{R}^V \longrightarrow \mathbb{R}^V, \chi \longmapsto \phi$, where the components of χ are i.i.d. random variables (commonly $\mathcal{N}(0, \mathbb{1})$).

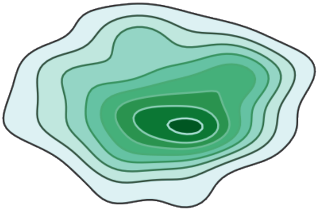
Theoretical / computational requirements:

- ergodic in $p(\phi)$
  - $p(\phi) \neq 0 \Rightarrow q(\phi) \neq 0$
  - sufficient overlap between q and p for practical use on human timescales
- balanced and asymptotically exact
  - statistical selection or weighting procedure for asymptotically unbiased estimation similar to accept/reject correction

Overrelaxation

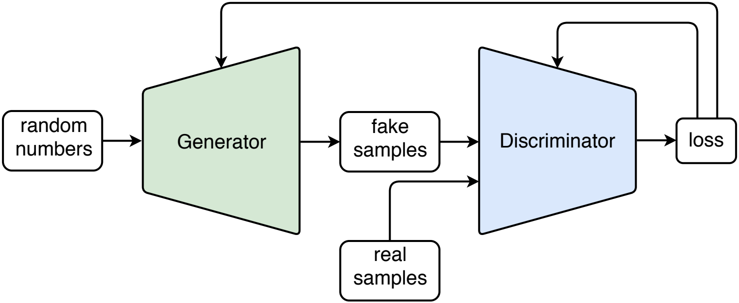


$$T_A(\phi'|\phi) = 1 \quad \text{for} \quad \Delta S = 0$$



- sampling on hypersurfaces of constant S
- ergodicity through normal MC steps
- requirements
  - ability to reproduce all possible S
  - symmetric a priori selection probability

Generative Adversarial Networks

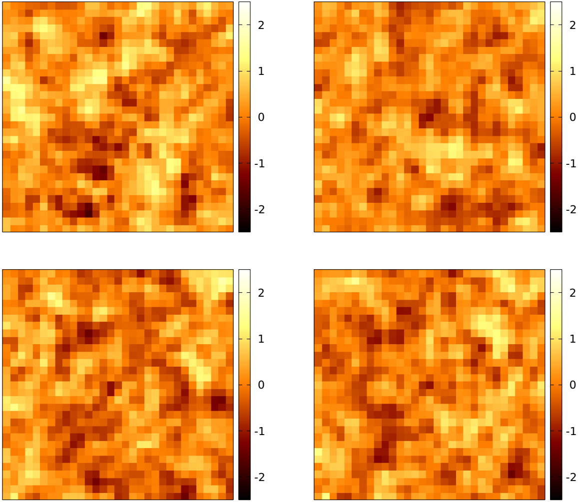


- overrelaxation step: find χ s.t. $S[g(\chi)] = S[\phi]$
- iterative gradient descent solution of

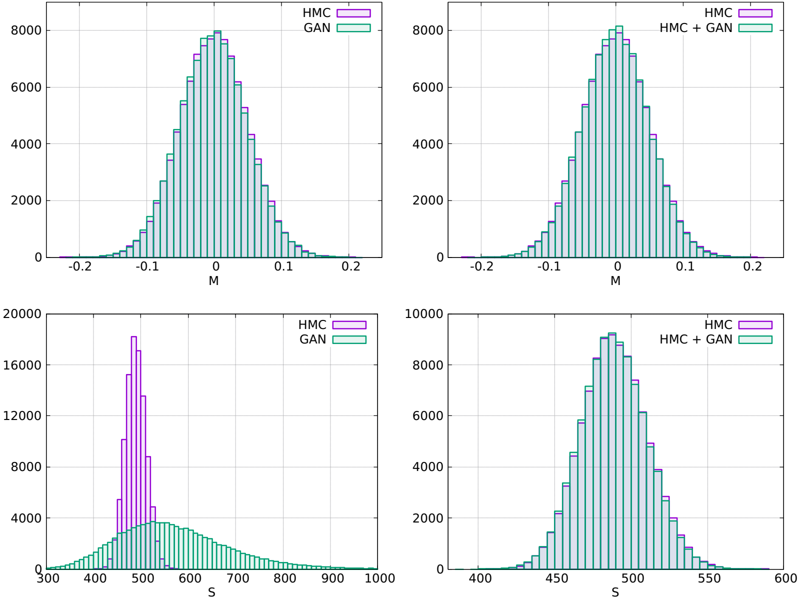
$$\chi' = \arg\min_{\chi} \parallel S[g(\chi)] - S[\phi] \parallel$$

Sample Examples

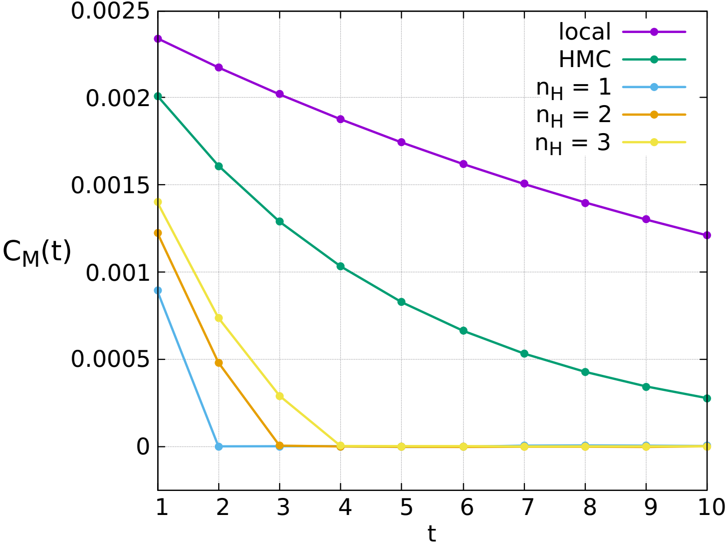
$$d = 2,\ V = 32^2,\ \kappa = 0.21,\ \lambda = 0.022$$



Magnetization & Action Distributions



Reduced Autocorrelations



Problems with this Approach

- GAN
  - relies on the existence of an exhaustive dataset
  - no direct access to sample probability
  - adversarial learning complicates quantitative error assessment
  - convergence/stability issues such as mode collapse
- Overrelaxation
  - still relies on traditional MC algorithms
  - symmetry of the selection probability
  - little effect on autocorrelations of observables coupled to S
  - latent space search is computationally rather demanding

Proper Reweighting to Model Distribution

$$\langle \mathcal{O} \rangle_{\phi \sim p(\phi)} = \int \mathcal{D}\phi \, p(\phi) \; \mathcal{O}(\phi)$$
$$= \int \mathcal{D}\phi \, q(\phi) \; \frac{p(\phi)}{q(\phi)} \; \mathcal{O}(\phi) = \left\langle \frac{p(\phi)}{q(\phi)} \; \mathcal{O}(\phi) \right\rangle_{\phi \sim q(\phi)}$$

Generate $q(\phi)$ through parametrizable, invertible function $g(\chi|\omega)$ with tractable Jacobian determinant:

$$q(\phi) = r(\chi(\phi)) \left| \det \frac{\partial g^{-1}(\phi)}{\partial \phi} \right|$$

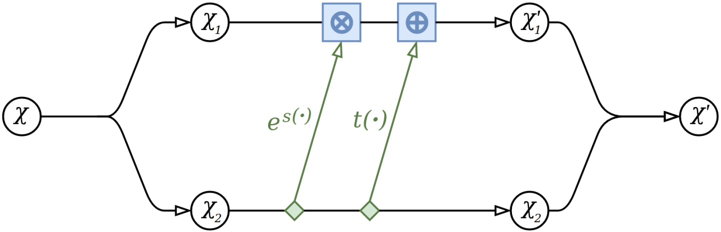
Optimal choice for $q(\phi)$ $\longleftrightarrow$ Minimal relative entropy / Kullback-Leibler divergence

$$D_{\mathrm{KL}}(q \,\|\, p) = -\int \mathcal{D}\phi \, q(\phi) \log \frac{p(\phi)}{q(\phi)} = -\left\langle \log \frac{p(\phi)}{q(\phi)} \right\rangle_{\phi \sim q(\phi)}$$

INN / Real NVP Flow

Ardizzone, Klessen, Köthe, Kruse, Maier-Hein, Pellegrini, Rahner, Rother, Wirkert (2018) — “Analyzing Inverse Problems with Invertible Neural Networks” — arXiv: 1808.04730

Ardizzone, Köthe, Kruse, Lüth, Rother, Wirkert (2019) — “Guided Image Generation with Conditional Invertible Neural Networks” — arXiv: 1907.02392



Advantages of this Approach

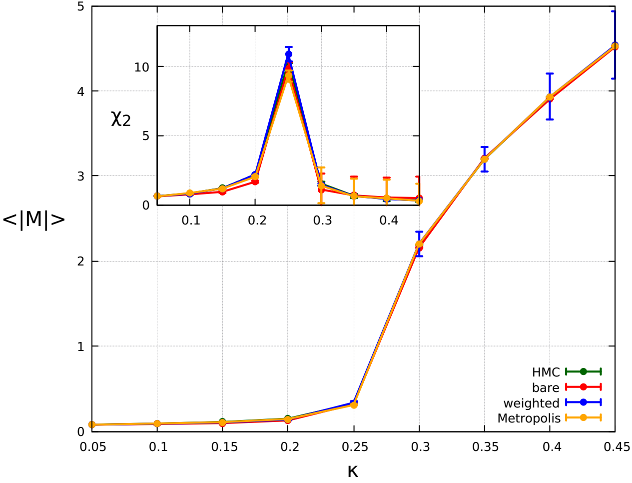
- learning is completely data-independent
- improved error metrics
  - Metropolis-Hastings acceptance rate
  - convergence properties of D_{KL}
- ergodicity & balance + asymptotic exactness satisfied a priori
- no latent space deformation required

Objective: maximization of overlap between $q(\phi)$ and $p(\phi)$.

Comparison with HMC Results

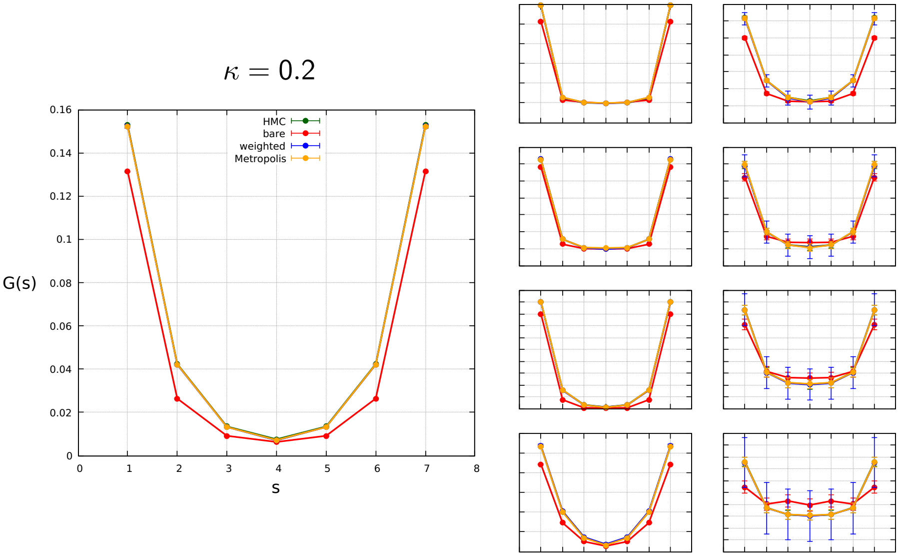
$$d = 2,\ V = 8^2,\ \lambda = 0.02$$

INN: 8 layers, 4 hidden layers, 512 neurons / layer



Comparison with HMC Results

$\kappa = 0.2$



Potential Applications & Future Work

- accelerated simulations of physically interesting theories (QCD, Yukawa, Gauge-Higgs, Condensed Matter)
- additional conditioning (cINN) to encode arbitrary couplings κ, λ
- tackling sign problems with generalized thimble / path optimization approaches by latent space disentanglement
- efficient minimization of D_{KL} i.t.o. the ground state energy of an interacting hybrid classical-quantum system

Challenges & Problems

- scalability to higher dimensions / larger volumes / more d.o.f. (e.g. QCD: $\sim 10^9$ floats per configuration)
  - multi-GPU parallelization
  - progressive growing to successively larger volumes
- architectures that intrinsically respect symmetries and topological properties of the theory
  - gauge symmetry / equivariance
- critical slowing down

Thank you!