LPENS
LABORATOIRE DE PHYSIQUE
DE L'ÉCOLE NORMALE SUPÉRIEURE

SORBONNE UNIVERSITÉ

Machine-learning-assisted Monte Carlo fails at sampling computationally hard problems

Simone Ciarella, **Jeanne Trinquier**, Martin Weigt, Francesco Zamponi

arXiv:2210.11145

Motivations

$$P_B(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z}$$

- Goal: generate equilibrium samples from the Boltzmann distribution
- Universal strategy: local MCMC
- But there are hard-to-sample problems
- There exists system-specific solutions
- Recent line of research: machine learning assisted MCMC moves

An universal solution?

- Suppose we can find an ‘auxiliary’ distribution $P_{AR}(\sigma)$ such that
- It approximates well the target Boltzmann distribution

$$D_{KL}(P_B || P_{AR}) = \left\langle \log \frac{P_B(\sigma)}{P_{AR}(\sigma)} \right\rangle_{P_B} \ll N$$

Merchan, Nemenman (2016)

- It can be sampled easily e.g via autoregressive structure

$$P_{AR}(\sigma) = P_{AR}^1(\sigma_1)\, P_{AR}^2(\sigma_2 | \sigma_1) \cdots P_{AR}^N(\sigma_N | \sigma_{N-1}, \cdots, \sigma_1)$$

Wu, Wang, Zhang (2019)

Global MCMC scheme

1. Propose a new configuration σ_{new} sampled from $P_{AR}(\sigma)$
2. Accept the move with Metropolis-Hastings rate

$$\text{Acc}\left[\sigma_{old} \to \sigma_{new}\right] = \min\left[1, \frac{P_B(\sigma_{new}) \times P_{AR}(\sigma_{old})}{P_B(\sigma_{old}) \times P_{AR}(\sigma_{new})}\right]$$

- The change is non-local (all the spins are changed in one move)
- If the acceptance rate is high ($D_{KL}(P_B || P_{AR})$ is small), very fast decorrelation

McNaughton, Milosevic, Perali, Pilati (2020)
Gabrié, Rotswdkoff, Vanden-Ejiden (2021)

Learning the autoregressive distribution

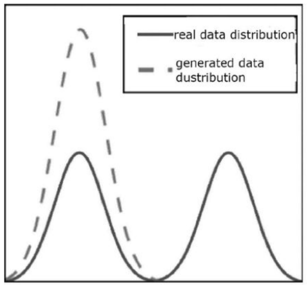
Computing D_{KL} requires sampling from P_B! $D_{KL}(P_B||P_{AR}) = \left\langle \log \frac{P_B(\sigma)}{P_{AR}(\sigma)} \right\rangle_{P_B}$

Proposed solutions:

1. Maximum likelihood: Sample from P_B, learn P_{AR} by minimizing $D_{KL}(P_B||P_{AR})$
2. Variational approach: Minimize $D_{KL}(P_{AR}||P_B) = \left\langle \log \frac{P_{AR}(\sigma)}{P_B(\sigma)} \right\rangle_{P_{AR}} = \beta(F[P_{AR}] - F[P_B])$

   Main problem: Mode collapse

   Wu, Wang, Zhang (2019)



3. Simulated tempering
   a. Generate a sample from P_B at high T via local MCMC
   b. Use this sample to learn P_{AR}
   c. Use P_{AR} to generate a new sample via global (+local) MCMC at T-δT

McNaughton, Milosevic, Perali, Pilati (2020)
Gabrié, Rotswdkoff, Vanden-Ejiden (2021)

Autoregressive architectures

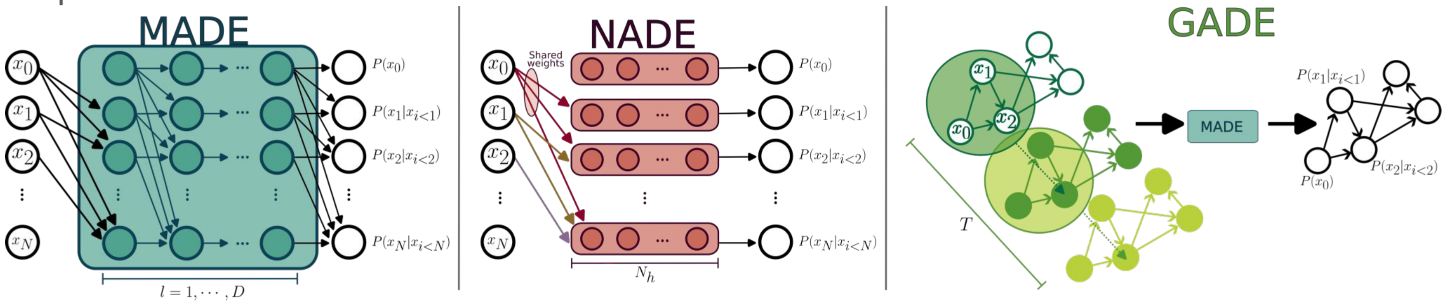
$$P_{AR}(\sigma) = P^1_{AR}(\sigma_1)\, P^2_{AR}(\sigma_2 \,|\, \sigma_1)\ \cdots\ P^N_{AR}(\sigma_N \,|\, \sigma_{N-1}, \cdots, \sigma_1)$$

Shallow MADE

$$P^1_{AR}(\sigma_1) = \frac{e^{h_1 \sigma_1}}{2\cosh(h_1)}$$

$$P^i_{AR}(\sigma_i \,|\, \sigma_{<i}) = \frac{e^{\sum_{j(<i)} J_{ij}\sigma_i\sigma_j + h_i\sigma_i}}{2\cosh\left(\sum_{j(<i)} J_{ij}\sigma_j + h_i\right)}$$

More complicated:

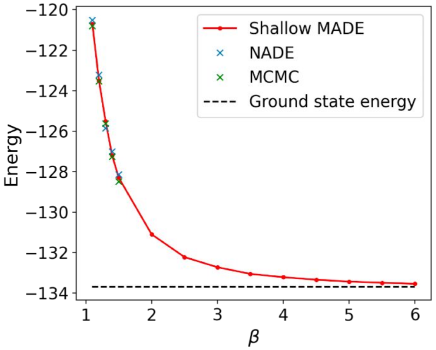


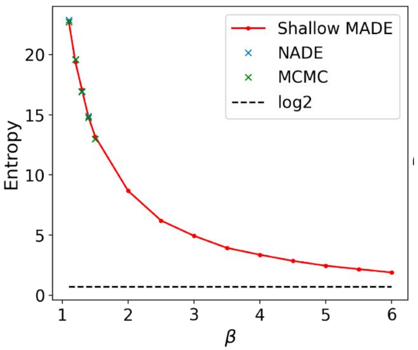
2d Edwards-Anderson spin glass

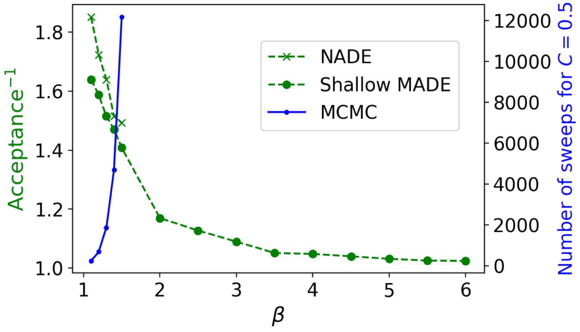
$$H(\sigma) = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j \qquad \sigma_i \in \{-1,1\}$$

- Relaxation time of local MCMC grows fast when T is decreased
- Autoregressive models are able to recapitulate the energy and entropy
- High acceptance rate: decorrelation in a few steps

Local MCMC time /sweep	10^{-5} s
Global MCMC time/sweep	$3\ 10^{-4}$ s
Training time at fixed β	5 mn

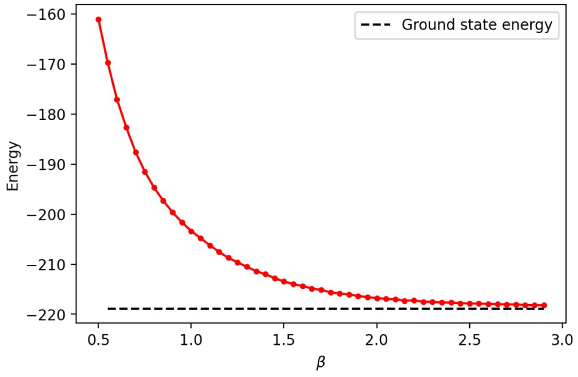


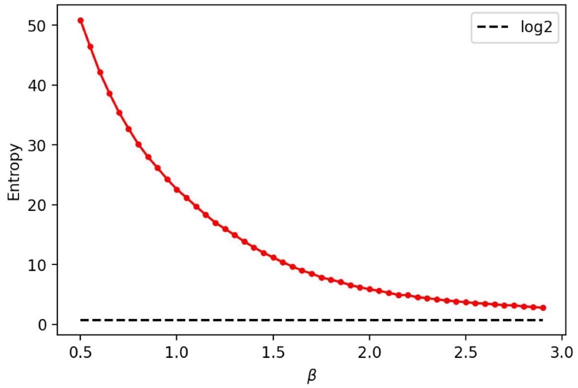


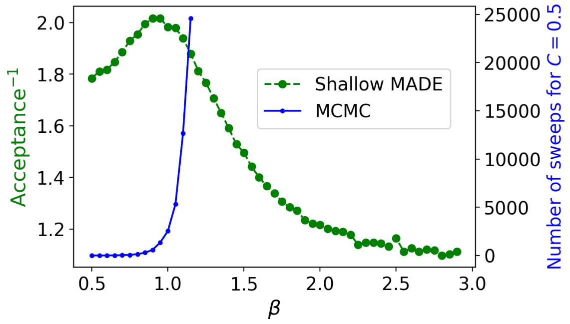


3d Edwards-Anderson spin glass

- Similar results despite the presence of a phase transition
- Needs a more systematic study on the dependance with the system size

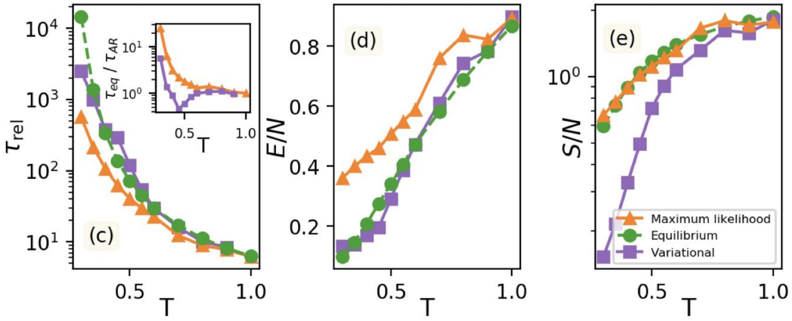






Failure for more complicated models $H(\sigma) = \sum_{\langle i,j\rangle \in \mathcal{G}} \delta_{\sigma_i,\sigma_j}$ $\sigma_i \in \{1,2,\cdots,q\}$

- The most difficult sampling problems are those in the Random First Order Transition class
- The sampling time of local MCMC is $(T\text{-}T_d)^{-\gamma}$ above T_d and *exp(N)* below
- Failure of all models at T < 1 (where local MCMC is still fast)
- No clear perspective for improvement: increasing the model expressivity leads to overfitting



Conclusions and perspectives

- Machine-learning assisted MCMC enables to speed-up by orders of magnitude the sampling time in a broad range of problems
- However, still a challenge to use it for problems with a too complex landscape

- Perspectives
  - More recent architectures (Transformers)
  - Reduce the problem difficulty by learning to sample a subset of the spins given the neighboring spins