A new approach to simulating lattice field theories

Alessandro Nada

Università degli Studi di Torino
INFN, Sezione di Torino

12th November 2023

INFN Theory Retreat in Santo Stefano Belbo

Past and present

PhD student in **Torino** (with M. Caselle and M. Panero)

- lattice gauge theory at non-zero temperature: SU(2) and SU(3) YM equation of state
- effective string theory: testing Nambu-Goto predictions
- algorithms: Jarzynski's equality

Post-doc in **DESY Zeuthen** (with S. Schaefer and R. Sommer)

- algorithms: multilevel for quark correlators
- phenomenology: QCD running coupling
- algorithms: finite-size scaling of gradient flow

Now: **post-doc** of the **Simons Collaboration on Confinement and QCD Strings** (local PI M. Caselle)

- effective string theory: beyond Nambu-Goto picture
- algorithms: Stochastic Normalizing Flows for gauge theories (with M. Caselle, M. Panero, A. Bulgarelli and E. Cellini)

How to compute the path integral in a lattice simulation?

Computing v.e.v. on the lattice

How do we compute expectation values in a lattice simulation?

After discretizing the theory of interest (in Euclidean spacetime), the path integral becomes

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod_i \mathrm{d}\phi_i \, \mathcal{O}(\phi) \exp(-S(\phi))$$

How do we compute expectation values in a lattice simulation?

After discretizing the theory of interest (in Euclidean spacetime), the path integral becomes

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod_i \mathrm{d}\phi_i \; \underbrace{\mathcal{O}(\phi)}_{\text{compute}} \; \underbrace{\exp(-S(\phi))}_{\text{sample}}$$

- use the Boltzmann distribution as a weight to sample numerically configurations of fields ϕ
- simply "measure" the observable by computing it on the configurations

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{conf}}} \sum_{n=1}^{N_{\text{conf}}} \mathcal{O}(\phi^{(n)}), \qquad \phi^{(n)} \sim \exp(-S(\phi^{(n)}))$$

provided that the $\phi^{(n)}$ are effectively sampled according to Boltzmann distribution!

How do we sample from $\exp(-S)/Z$?

Markov Chain Monte Carlo

An elegant and reliable numerical recipe is provided by Markov Chain Monte Carlo **[Creutz; 1979]**

- Generate a (thermalized) Markov chain

$$\underbrace{\phi^{(0)} \stackrel{P_p}{\to} \phi^{(1)} \stackrel{P_p}{\to} \dots \stackrel{P_p}{\to}}_{\text{thermalization}} \underbrace{\phi^{(t)} \stackrel{P_p}{\to} \phi^{(t+1)} \stackrel{P_p}{\to} \dots \to \phi^{(t+N_{\text{conf}})}}_{\text{equilibrium}}$$

(setting $\exp(-S)$ as the equilibrium distribution)

- Measure $\mathcal{O}$ on sampled equilibrium configurations

MCMC algorithm of choice (Metropolis, heatbath, HMC) defined by the transition probability P_p

(details are unimportant for the sake of this presentation)

Critical slowing down

The configurations sampled sequentially in a Markov Chain are **autocorrelated**

$$\cdots \rightarrow \underbrace{\phi^{(t)} \rightarrow \phi^{(t+1)} \rightarrow \ldots}_{2\tau_{\text{int}}} \rightarrow \phi^{(t+n)}$$

Number of effectively independent configurations $= n/2\tau_{\text{int}}$

Critical slowing down: when a critical point is approached τ_{int} diverges

E.g. in the continuum limit $a \rightarrow 0$

$$\tau_{\text{int}} \simeq a^{-z}$$

where z depends on the algorithm and on the observable under study

Topological freezing: topological charge in YM has $z \simeq 5$ **[Schaefer, Sommer, Virotta; 2011]**

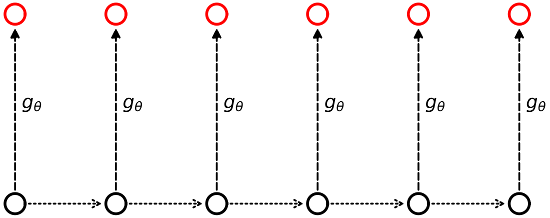
Trivializing and Normalizing flows

A different kind of sampling

Idea

Sample each new configuration (almost) independently from the previous one by construction

Trivializing map g_θ between a trivial base distribution and the "difficult" target



Wilson flow **[Lüscher; 2009]**: approximate trivializing map in Yang-Mills

Normalizing flows for lattice field theory

Normalizing Flows are a promising deep generative architecture that can provide a mapping between the target $p(\phi)$ with some tractable distribution $q_0(z)$

NFs are a diffeomorphism g_θ

$$g_\theta(\phi_0) = (g_{N_{cl}} \circ \cdots \circ g_1)(\phi_0) \qquad \phi_0 \sim q_0$$

composed of N_{cl} transformations $\rightarrow$ the **coupling layers** g_n

Normalizing Flows are a promising deep generative architecture that can provide a mapping between the target $p(\phi)$ with some tractable distribution $q_0(z)$

NFs are a diffeomorphism g_θ

$$g_\theta(\phi_0) = (g_{N_{cl}} \circ \cdots \circ g_1)(\phi_0) \qquad\qquad \phi_0 \sim q_0$$

composed of N_{cl} transformations $\rightarrow$ the **coupling layers** g_n

The generated distribution is

$$q_\theta(\phi) = q_0(g_\theta^{-1}(\phi)) \prod_n |\det J_n(\phi_n)|^{-1}$$

Normalizing Flows are a promising deep generative architecture that can provide a mapping between the target $p(\phi)$ with some tractable distribution $q_0(z)$

NFs are a diffeomorphism g_θ

$$g_\theta(\phi_0) = (g_{N_{cl}} \circ \cdots \circ g_1)(\phi_0) \qquad \phi_0 \sim q_0$$

composed of N_{cl} transformations $\rightarrow$ the **coupling layers** g_n

The generated distribution is

$$q_\theta(\phi) = q_0(g_\theta^{-1}(\phi)) \prod_n |\det J_n(\phi_n)|^{-1}$$

Training: minimize the **Kullback-Leibler** divergence, a measure of the “similarity” between two distributions

$$\tilde{D}_{\mathrm{KL}}(q_\theta \| p) = \int \mathrm{d}\phi \, q_\theta(\phi) \log \frac{q_\theta(\phi)}{p(\phi)}$$

Sampling with Normalizing flows

Sampling

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathrm{d}\phi \, \mathcal{O}(\phi) q_\theta(\phi) \frac{p(\phi)}{q_\theta(\phi)} = \frac{1}{Z} \int \mathrm{d}\phi \, \underbrace{q_\theta(\phi)}_{\text{sample}} \underbrace{\mathcal{O}(\phi) \tilde{w}(\phi)}_{\text{measure}} = \frac{\langle \mathcal{O}(\phi) \tilde{w}(\phi) \rangle_{\phi \sim q_\theta}}{\langle \tilde{w}(\phi) \rangle_{\phi \sim q_\theta}}$$

with a weight

$$\tilde{w}(\phi) = \frac{p}{q_\theta}$$

Get Z directly by sampling from q_θ **[Nicoli et al.; 2020]**

$$Z = \int \mathrm{d}\phi \, p(\phi) = \int \mathrm{d}\phi \, q_\theta(\phi) \tilde{w}(\phi) = \langle \tilde{w}(\phi) \rangle_{\phi \sim q_\theta}$$

Issues with NFs

Incomplete list of recent works:

→ successfully applied in LFTs in 2d: ϕ^4 scalar field theory **[Albergo et al.; 2019]**, **[Kanwar et al.; 2020]**, **[Nicoli et al.; 2020]**, **[Del Debbio et al.; 2021]**, $\mathrm{SU}(N)$ **[Boyda et al.; 2020]**, fermionic theories **[Albergo et al.; 2021]**, $\mathrm{U}(1)$ and $\mathrm{SU}(N)$ with fermions **[Abbott et al.; 2022]**, Schwinger model **[Finkenrath et al.; 2022]**, **[Albergo et al.; 2022]**, first attempts in 4d **[Abbott et al.; 2023]**

→ strongly related to the idea of trivializing maps **[Lüscher; 2009]**, **[Bacchio et al.; 2022]**, **[Albandea et al.; 2023]**

- NFs effectively remove critical slowing in low dimensional theories
- training generally scales badly with volume and in the continuum limit
- quite complicated architectures for gauge theories

Non-equilibrium stochastic evolutions

The Second Law of Thermodynamics

Clausius inequality for an (isothermal) transformation from state A to state B

$$\frac{Q}{T} \leq \Delta S$$

If we use

$$\begin{cases} Q = & \Delta E - W \quad \text{(First Law)} \\ F \stackrel{\text{def}}{=} & E - ST \end{cases}$$

the Second Law becomes

$$W \geq \Delta F = F_B - F_A$$

where the equality holds for reversible processes.

Moving from thermodynamics to statistical mechanics we know that actually

$$\langle W \rangle_f \geq \Delta F = F_B - F_A$$

Jarzynski's equality and the Second Law

Jarzynski's equality [Jarzynski; 1997] is a beautiful result from non-equilibrium statistical mechanics

$$\left\langle \exp\left(-\frac{W}{T}\right)\right\rangle_f = \exp\left(-\frac{\Delta F}{T}\right)$$

valid for a given process f

Using Jensen's inequality $\langle \exp x \rangle \geq \exp\langle x \rangle$ we get

$$\langle W \rangle_f \geq \Delta F$$

Apart from the real world, it can be proved for several processes

→ most relevantly for us: **Markov Chain Monte Carlo** for lattice field theory

Jarzynski's equality in MCMC

Consider two distributions q_0 and p

$$q_0 = \exp(-S_0)/Z_0 \rightarrow \cdots \rightarrow p = \exp(-S)/Z$$

The ratio of partition functions is computed directly with an average over **non-equilibrium evolutions**

$$\frac{Z}{Z_0} = \langle \exp\left(-W\right) \rangle_f$$

Jarzynski's equality in MCMC

Consider two distributions q_0 and p

$$q_0 = \exp(-S_0)/Z_0 \to \cdots \to p = \exp(-S)/Z$$

The ratio of partition functions is computed directly with an average over **non-equilibrium evolutions**

$$\frac{Z}{Z_0} = \langle \exp\left(-W\right) \rangle_f$$

Over a single evolution:

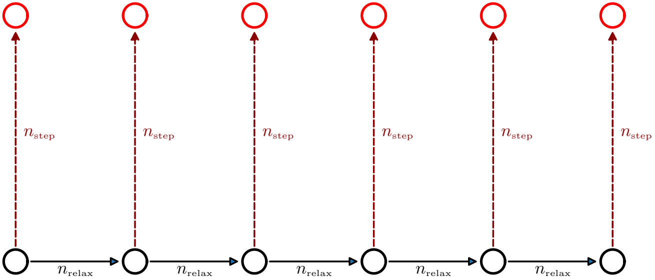
- update the system using regular MC updates over n_{step}
- but the transition probability changes along the evolution according to a **protocol** η_n that interpolates q_0 and p

$$q_0 \simeq e^{-S_{\eta_0}} \to e^{-S_{\eta_1}} \to \cdots \to p \simeq e^{-S}$$

- along the process we compute the **work**

$$W = \sum_{n=0}^{n_{\text{step}}-1} \left\{ S_{\eta_{n+1}}\left[\phi_n\right] - S_{\eta_n}\left[\phi_n\right] \right\}$$

A new paradigm: out-of-equilibrium stochastic evolutions



compute v.e.v. with

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O}(\phi) \exp(-W(\phi_0 \to \phi)) \rangle_{\mathrm{f}}}{\langle \exp(-W(\phi_0 \to \phi)) \rangle_{\mathrm{f}}}$$

Jarzynski's equality for lattice field theory

Several applications, spearheaded by the Torino group over the last few years

Compute ratios of partition functions:

- interface free-energy in the Z_2 gauge theory **[Caselle et al.; 2016]**
- $\mathrm{SU}(3)$ equation of state **[Caselle, Nada and Panero; 2018]**
- running coupling for $\mathrm{SU}(N)$ **[Francesconi, Panero and Preti; 2020]**
- entanglement entropy for Ising model in $2d$ and $3d$ **[Bulgarelli and Panero; 2023]**

But also just to sample difficult distributions:

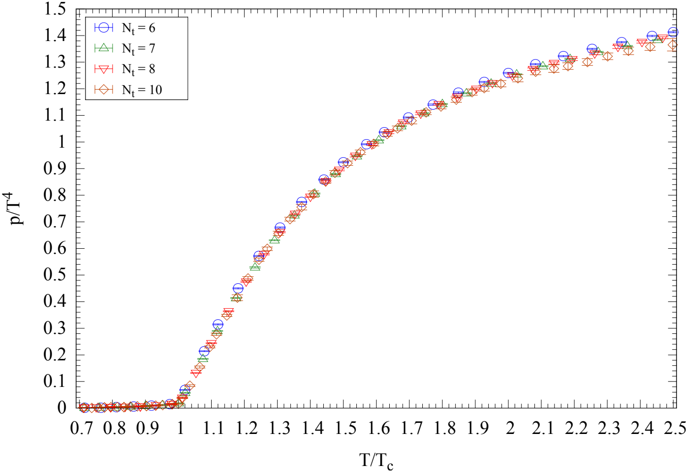
- sampling topological observables on periodic boundary conditions **[Bonanno, Nada and Vadacchino; 2023]**

SU(3) e.o.s. with Jarzynski's equality

Large-scale application: computation of the SU(3) equation of state **[Caselle et al.; 2018]**

Extract the pressure with Jarzynski's equality

$$\frac{p(T)}{T^4} - \frac{p(T_0)}{T_0^4} = \left(\frac{N_t}{N_s}\right)^3 \log\langle e^{-W_{\mathrm{SU}(N_c)}}\rangle_{\mathrm{f}}$$



A common framework: Stochastic Normalizing Flows

Jarzynski's equality is the same formula used to extract Z in NFs

$$\frac{Z}{Z_0} = \langle \tilde{w}(\phi) \rangle_{\phi \sim q} = \langle \exp(-W) \rangle_{\mathrm{f}}$$

Normalizing Flows

$$\phi_0 \to \phi_1 = g_1(\phi_0) \to \cdots \to \phi$$

$$-\log \tilde{w} = S(\phi) - S_0(\phi_0) - \log |\det J|$$

stochastic non-equilibrium evolutions

$$\phi_0 \overset{P_{\eta_1}}{\to} \phi_1 \overset{P_{\eta_2}}{\to} \ldots \overset{P_{\eta_{n_{\mathrm{step}}}}}{\to} \phi$$

$$W = S(\phi) - S_0(\phi_0) - Q$$

A common framework: Stochastic Normalizing Flows

Jarzynski's equality is the same formula used to extract Z in NFs

$$\frac{Z}{Z_0} = \langle \tilde{w}(\phi) \rangle_{\phi \sim q} = \langle \exp(-W) \rangle_{\mathrm{f}}$$

Normalizing Flows

$$\phi_0 \to \phi_1 = g_1(\phi_0) \to \dots \to \phi$$

$$-\log \tilde{w} = S(\phi) - S_0(\phi_0) - \log |\det J|$$

stochastic non-equilibrium evolutions

$$\phi_0 \stackrel{P_{\eta_1}}{\to} \phi_1 \stackrel{P_{\eta_2}}{\to} \dots \stackrel{P_{\eta_{n_{\mathrm{step}}}}}{\to} \phi$$

$$W = S(\phi) - S_0(\phi_0) - Q$$

Stochastic Normalizing Flows (introduced in **[Wu et al.; 2020]**)

$$\phi_0 \to g_1(\phi_0) \stackrel{P_{\eta_1}}{\to} \phi_1 \to g_2(\phi_1) \stackrel{P_{\eta_2}}{\to} \dots \stackrel{P_{\eta_{n_{\mathrm{step}}}}}{\to} \phi$$

$$W = S(\phi) - S_0(\phi_0) - Q - \log |\det J|$$

The proper loss we use for training is

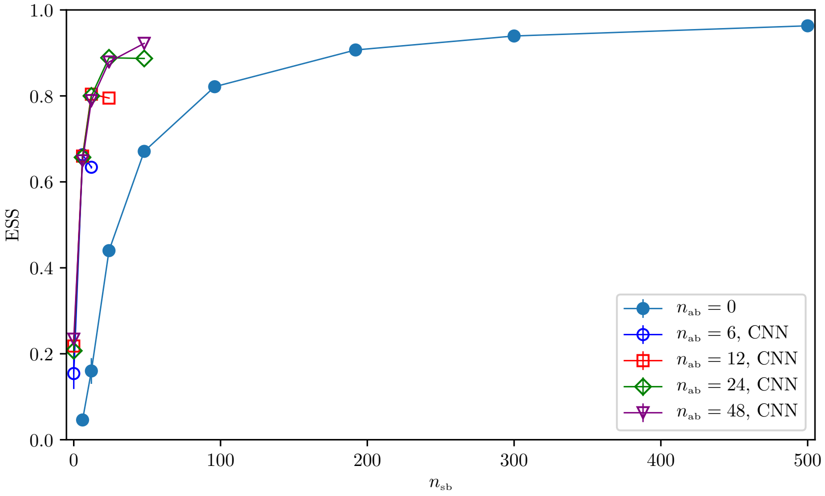
$$\tilde{D}_{\mathrm{KL}}(q_0 P_{\mathrm{f}} \| p P_{\mathrm{r}}) = \int \prod_{i=0}^{n_{\mathrm{step}}} \mathrm{d}\phi_i \, q_0(\phi_0) P_{\mathrm{f}}[\phi_0, \phi_1, \ldots, \phi] \log \frac{q_0(\phi_0) P_{\mathrm{f}}[\phi_0, \phi_1, \ldots, \phi]}{p(\phi) P_{\mathrm{r}}[\phi, \phi_{N-1}, \ldots, \phi_0]}$$

$\rightarrow$ measure of how reversible the process is!

Clear "thermodynamic" interpretation

$$\tilde{D}_{\mathrm{KL}}(q_0 P_{\mathrm{f}} \| p P_{\mathrm{r}}) = \underbrace{\langle W \rangle_{\mathrm{f}} + \log \frac{Z}{Z_0} \geq 0}_{\text{Second Law}}$$

Efficiency test on the $2d$ ϕ^4 scalar field theory



stochastic evolutions vs SNFs on a $N_s \times N_t = 16 \times 8$ lattice

SNFs for lattice field theory

- Theoretical common framework between Jarzynski's equality and NFs is now explicit
- SNFs can be trained!
- improvement over non-equilibrium evolutions: less MCMC updates used
- improvement over normalizing flows: limited training times, good scaling with volume
- perhaps non-equilibrium statistical mechanics can give us even more insights

What next?

SNFs already applied to sample **directly** the Nambu-Goto string on a lattice **[Caselle, Cellini and Nada; 2023]**

- main project of **E. Cellini**
- unfeasible to do numerically with a MCMC, SNFs best approach (and still challenging)
- allows to test NG and Effective String Theories directly, when analytic results are not available (string width)

New approach to sample lattice gauge theories

- SNFs for $\mathrm{SU}(3)$ gauge theory under development (with **A. Bulgarelli** and **E. Cellini**)
- final aim: mitigating (if not removing) topological freezing from lattice gauge theory simulations
- proof of concept in the CP^{N-1} model **[Bonanno, Nada and Vadacchino; 2023]**
- much less risky program! instead of building a trivializing map from the ground up, we **systematically improve** something that already works (non-equilibrium MC)

Thank you for your attention!

Out-of-equilibrium stochastic evolutions

Closer look at the average on the processes in the equality:

$$\frac{Z}{Z_0} = \langle \exp(-W) \rangle_f = \int \mathrm{d}\phi_0 \, \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_N \, q_0(\phi_0) \, P_\mathrm{f}[\phi_0, \phi_1, \ldots, \phi_N] \, \exp(-W)$$

with

$$P_\mathrm{f}[\phi_0, \phi_1, \ldots, \phi_N] = \prod_{n=0}^{N-1} P_{\eta_n}(\phi_n \to \phi_{n+1})$$

- the *actual* probability distribution at each step is NOT the equilibrium distribution $\sim \exp(-S_{\eta_n})$: it's a non-equilibrium process!
- the $\langle \ldots \rangle_f$ average is taken over as many evolutions as possible (all independent from each other!)

Typical toy model for tests: ϕ^4 field theory in 2 dimensions in the unbroken symmetry phase

$$S(\phi) = \sum_{x \in \Lambda} -2\kappa \sum_{\mu=0,1} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\phi(x)^2 + \lambda\phi(x)^4$$

Protocol

η_n interpolates between the prior (normal distribution is recovered with $\kappa = \lambda = 0$) and target parameters $\kappa = 0.2$ and $\lambda = 0.022$

- linear protocol η_n
- heatbath algorithm for the stochastic updates
- n_{sb} = # of stochastic updates

Coupling layer

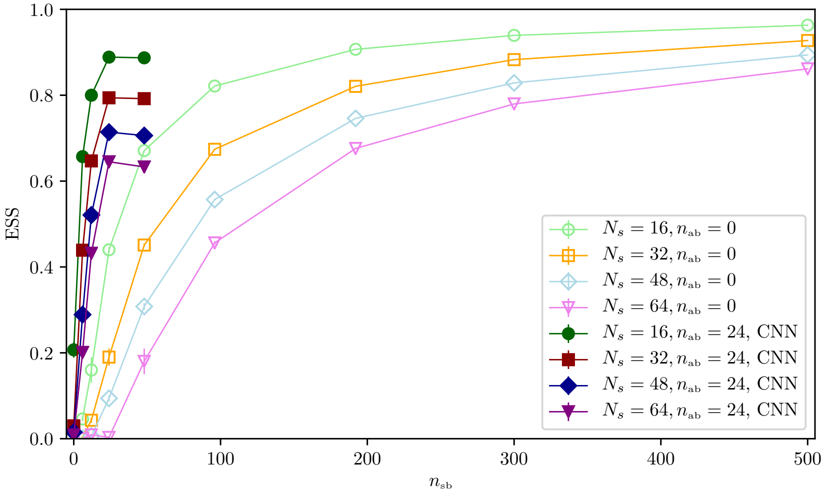
- using affine layers with CNNs
- n_{ab} = # of affine blocks

Metric

Effective Sample Size as metric to evaluate architectures

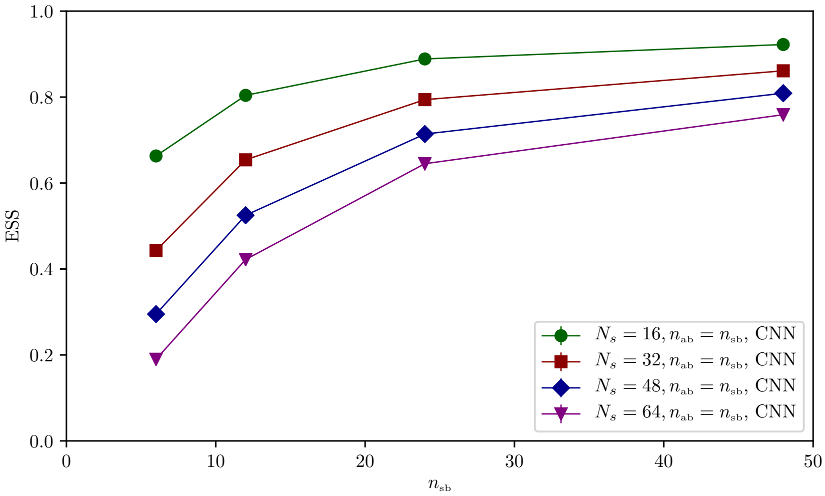
$$\text{ESS} = \frac{\langle \exp(-W) \rangle_f^2}{\langle \exp(-2W) \rangle_f}$$

Training length: 10^4 epochs for all volumes. Slowly-improving regime reached fast

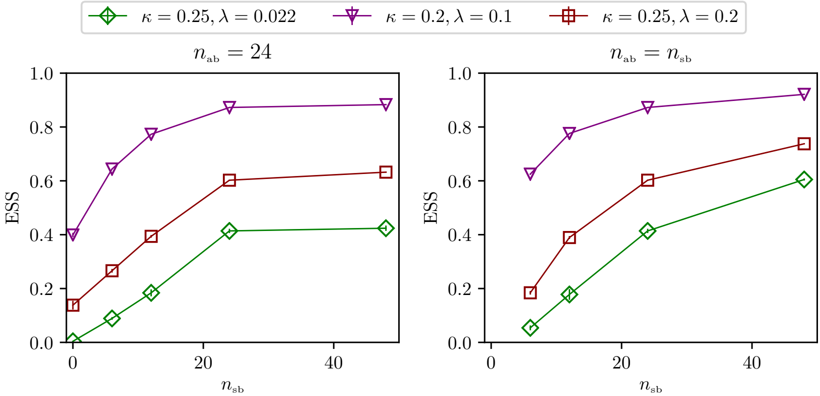


Interesting behaviour for all volumes: a peak for $n_{sb} = n_{ab}$?

SNFs with $n_{sb} = n_{ab}$ as a possible recipe for efficient scaling



More results in the unbroken phase



Stochastic evolutions/SNF/NF wrap-up

	normalizing flows	stochastic evolutions	SNFs
preparation	training	setting the protocol η_n	both
forward prob. P_{f}	$P_{\mathrm{f}} = \prod_n P_n(\phi_n \to \phi_{n+1})$		
transition prob. P_n	$\delta(\phi_{n+1} - g_n(\phi_n))$	$P_{\eta_n}(\phi_n \to \phi_{n+1})$	uses both
KL divergence	$\tilde{D}_{\mathrm{KL}}(q_\theta \| p)$	$\tilde{D}_{\mathrm{KL}}(q_0 P_{\mathrm{f}} \| p P_{\mathrm{r}})$	
"work"	$W = S - S_0 - Q = -\log \tilde{w}$		
"heat" Q	$\sum_{n=0}^{N-1} \log \lvert \det J_n(\phi_n) \rvert$	$\sum_{n=0}^{N-1} S_{\eta_{n+1}}(\phi_{n+1}) - S_{\eta_{n+1}}(\phi_n)$	both
e.v. $\langle \mathcal{O} \rangle$	$\frac{\langle \mathcal{O}(\phi_N) \tilde{w}(\phi_N) \rangle_{\phi_N \sim q_\theta}}{\langle \tilde{w}(\phi_N) \rangle_{\phi_N \sim q_\theta}}$	$\frac{\langle \mathcal{O}(\phi_N) \exp(-W(\phi_0 \to \phi_N)) \rangle_{\mathrm{f}}}{\langle \exp(-W(\phi_0 \to \phi_N)) \rangle_{\mathrm{f}}}$	

Crooks fluctuation theorem

Crooks theorem **[Crooks; 1998]**: another relation deeply connected with Jarzynski's equality

$$\frac{\mathcal{P}_F(W)}{\mathcal{P}_R(-W)} = e^{W-\Delta F}$$

The $\mathcal{P}_{F,R}$ indicate the probability distribution of the work performed in the forward and reverse realizations of the transformation.

JE is easily recovered by moving the $\exp(-W)$ and $\mathcal{P}_R$ factors and integrating in W on both sides.

$W_d = W - \Delta F$ is the **dissipated** work.

Transformations g_n must be invertible + the Jacobian has to be efficiently computable

A class of coupling layers called **affine layers** meets this criteria

- The variables ϕ are divided into two partitions A and B
- For each layer, one is kept "frozen" while the other is transformed following

$$g_n : \begin{cases} \phi_{\mathrm{A}}^{n+1} = \phi_{\mathrm{A}}^{n} \\ \phi_{\mathrm{B}}^{n+1} = e^{-s(\phi_{\mathrm{A}}^{n})} \phi_{\mathrm{B}}^{n} + t(\phi_{\mathrm{A}}^{n}) \end{cases}$$

- s and t are the neural networks where the trainable parameters θ are
- **RealNVP** architecture **[Dinh et al.; 2016]**

Natural choice for lattice variables: checkerboard (i.e. even-odd) partitioning

Affine block = even c. layer + odd c. layer

Some possible issues with NFs

- **multi-modal distributions**
  in the presence of multiple vacua the training procedure "picks" only one

  "mode-collapse": only one mode of the distribution is sampled by the flow

  see **[Hackett et al.; 2021]**

- **scalability**
  measurements of v.e.v. are statistically independent (no autocorrelation)

  not clear however how the training times scale when approaching the continuum limit

  comprehensive discussion in **[Del Debbio et al.; 2021]** and **[Abbott et al.; 2022]**