Normalising Flows as Trivialising Maps for Lattice ϕ^4 Theory

Joe Marsh Rossney

Path integrals → correlations

$$\big\langle \mathcal{O}_0[\Psi]\mathcal{O}_1[\Psi]\dots\big\rangle = \frac{1}{\mathcal{Z}}\int \mathcal{D}\Psi e^{-iS_{\text{mink}}[\Psi]}\mathcal{O}_0(\Psi)\mathcal{O}_1(\Psi)\dots$$

Non-perturbative problems:

- Confinement mechanism
- Origin of mass
  (Reasonably important)

Lattice Field Theory

- Wick rotate $t \to -i\tau$, $S_{\text{mink}} \to iS_{\text{eucl}}$
- Discretise space-time onto lattice Λ

$$\big\langle \mathcal{O}_0[\Psi]\mathcal{O}_1[\Psi]\dots\big\rangle = \frac{1}{\mathcal{Z}}\int \prod_{x\in\Lambda} d\Psi_x e^{-S_{\text{eucl}}(\Psi)}\mathcal{O}_0(\Psi)\mathcal{O}_1(\Psi)\dots$$

Valid probability distribution if $e^{-S_{\text{eucl}}(\Psi)}$ real & positive

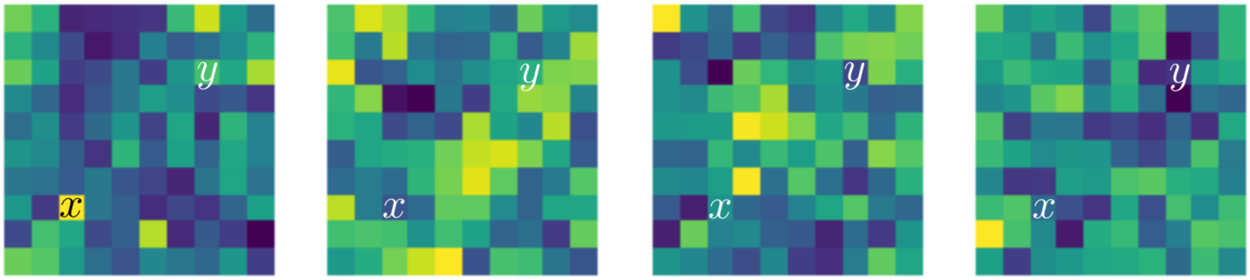
---

Example: ϕ^4 theory in $d=2$ $(a := 1)$

$$S(\phi) = \sum_{x\in\Lambda}\Big[-\beta\sum_{\mu=1}^{2}\phi_{x+e_\mu}\phi_x + \phi_x^2 + \lambda(\phi_x^2-1)^2\Big]$$

$\lambda \to \infty$ is the Ising model

Example: two point correlation



$$\phi_x^{(1)}\phi_y^{(1)} + \phi_x^{(2)}\phi_y^{(2)} + \phi_x^{(3)}\phi_y^{(3)} + \phi_x^{(4)}\phi_y^{(4)} + \cdots$$

$$\langle \phi_x \phi_y \rangle = \frac{1}{N}\sum_{n=1}^{N} \phi_x^{(n)}\phi_y^{(n)} + O(N^{-\frac{1}{2}})$$

How to generate samples?

Stochastic process whose limiting distribution is e^{-S}

$$A(\phi^{(t)} \to \phi^{(t+1)}) = \frac{q(\phi^{(t)} \mid \phi^{(t+1)})}{q(\phi^{(t+1)} \mid \phi^{(t)})}\,\frac{e^{-S(\phi^{(t+1)})}}{e^{-S(\phi^{(t)})}}$$

Get (wrong kind of) correlations $\phi^{(t)} \sim \phi^{(t+1)}$

Errors ~ effective sample size

$$N_{\text{eff}} = \frac{N}{2\tau_{\text{int}}}$$

$$\tau_{\text{int}} = \frac{1}{2}\sum_{t=-\infty}^{\infty}\frac{\Gamma(t)}{\Gamma(0)}$$

$$\Gamma(t) = \langle (\phi_x^{(t)}\phi_y^{(t)})(\phi_x^{(0)}\phi_y^{(0)}) \rangle - \langle \phi_x \phi_y \rangle$$

It gets worse...

$$\tau_{\text{int}} \sim \xi^{z}$$

"Critical slowing down" towards continuum limit 🤦

Generative models

https://thispersondoesnotexist.com/ (GAN)

Joint (correlated) probability distributions which we can model (fit training examples)

Generate new "synthetic" samples $\sim \tilde{p}_\theta(\phi)$

Other uses of generative models

- As classifiers
- Data augmentation (without storage issues)
- Data interpolation
- Procedural generation (e.g. graphics)

Recent idea (Albergo et al. 1904.12072)

$$A(\phi^{(t)} \to \phi^{(t+1)}) = \frac{\tilde{p}_\theta(\phi^{(t)})}{\tilde{p}_\theta(\phi^{(t+1)})} \frac{e^{-S(\phi^{(t+1)})}}{e^{-S(\phi^{(t)})}}$$

Independent proposals → rejections *only* source of autocorrelation

Corrects for "mistakes" in model (underfitting)

Candidate models

- Autoregresive models (expensive)

$$\tilde{p}_\theta(\phi) = \prod_i q_\theta^{(i)}(i \mid j < i)$$

- Variational autoencoders (intractable)

$$\tilde{p}_\theta(\phi) = \int dz \ell_\theta(z) r(z) dz$$

- GANs $\tilde{p}_\theta(\phi) = ???$ (implicit)
- Normalising flows....

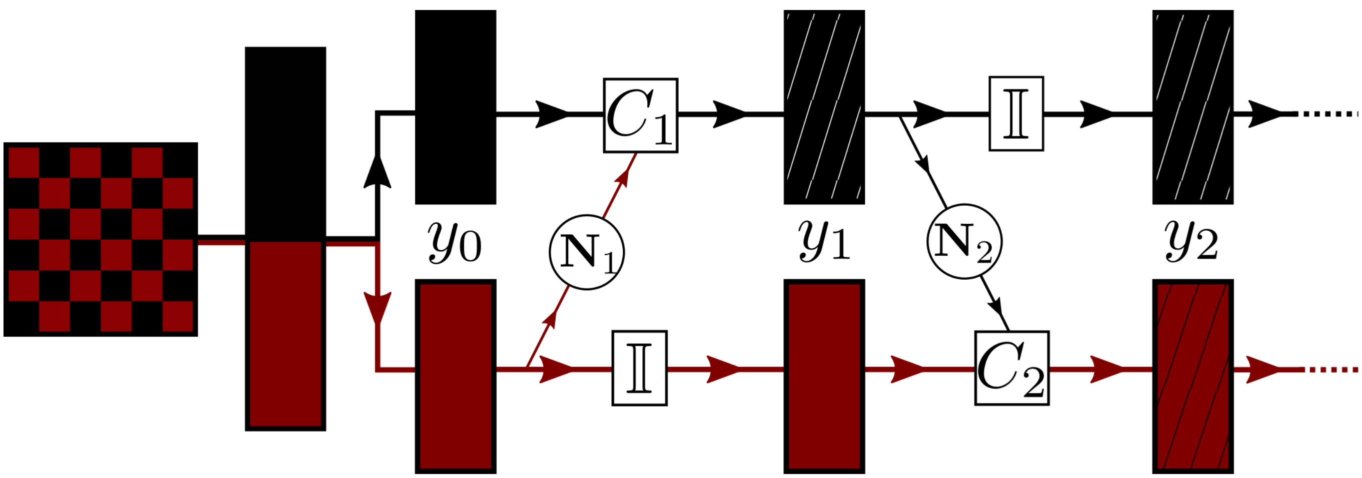
Normalising flows

Model theory as bijection $\phi = f_\theta(z)$ from uncorrelated latent variables

$$r(z) = \prod_{x \in \Lambda} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z_x^2/(2\sigma^2)}$$

$$\tilde{p}_\theta(f_\theta(z)) = r(z) \left| \det \frac{\partial f_\theta(z)}{\partial z} \right|$$

f_θ^{-1} is an (approximate) *trivialising map*



Correlations → weights and biases (θ) of neural networks ($\mathbf{N}$)

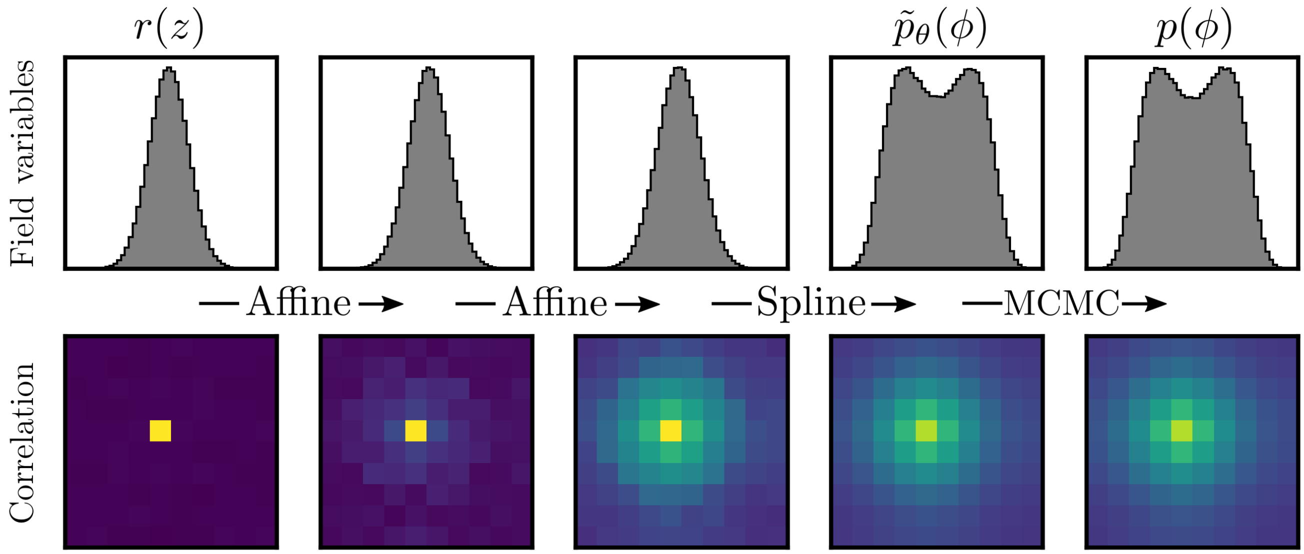
Training (fit parameteric model $\tilde{S}_\theta$)

$$D_{\mathrm{KL}}(\tilde{p}_\theta \,||\, p) = \int \prod_{x \in \Lambda} d\phi_x \, \tilde{p}_\theta(\phi) \log \frac{\tilde{p}_\theta(\phi)}{p(\phi)}$$

$$\hat{D}_{\mathrm{KL}}(\tilde{p}_\theta \,||\, p) = \mathbb{E}_{z \sim r(z)} \left[S\big(f_\theta(z)\big) - \log \left| \det \frac{\partial f_\theta(z)}{\partial z} \right| \right] + \dots$$

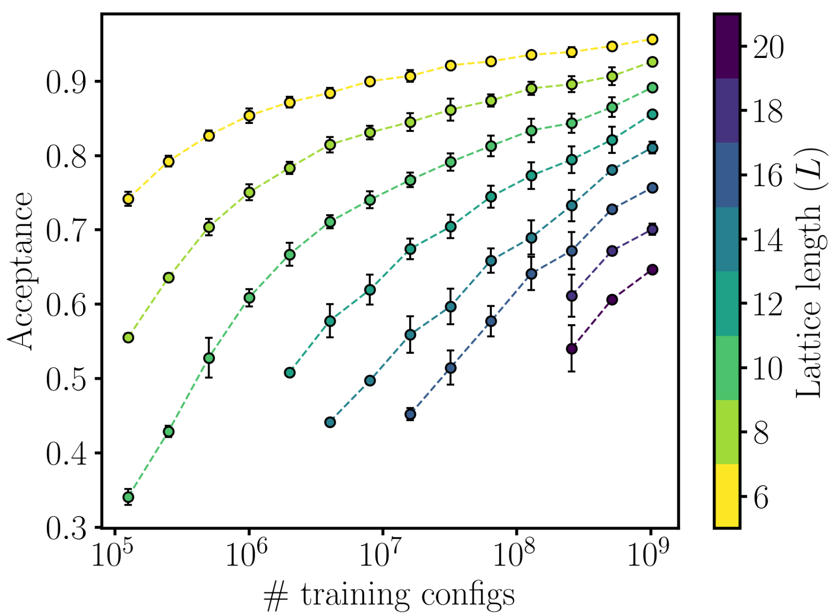
Doesn't require existing training examples!

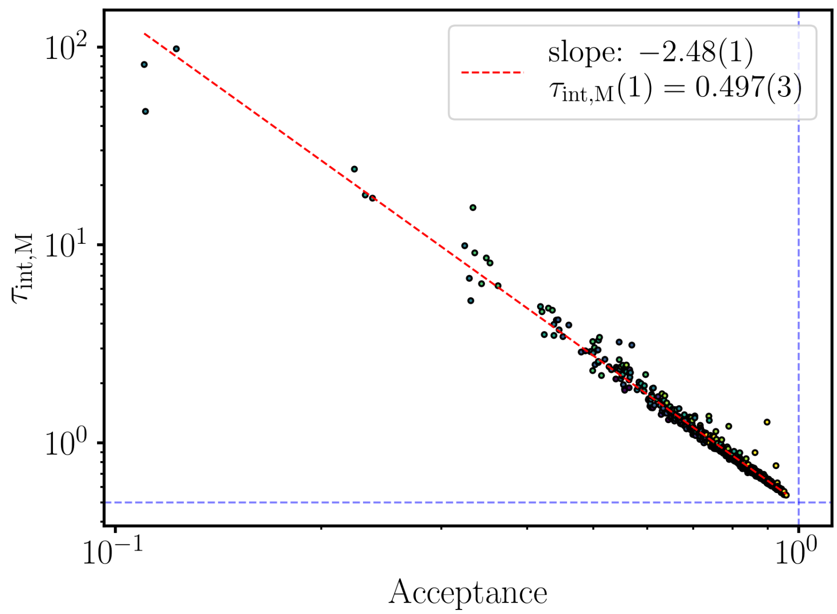
Example of trained model



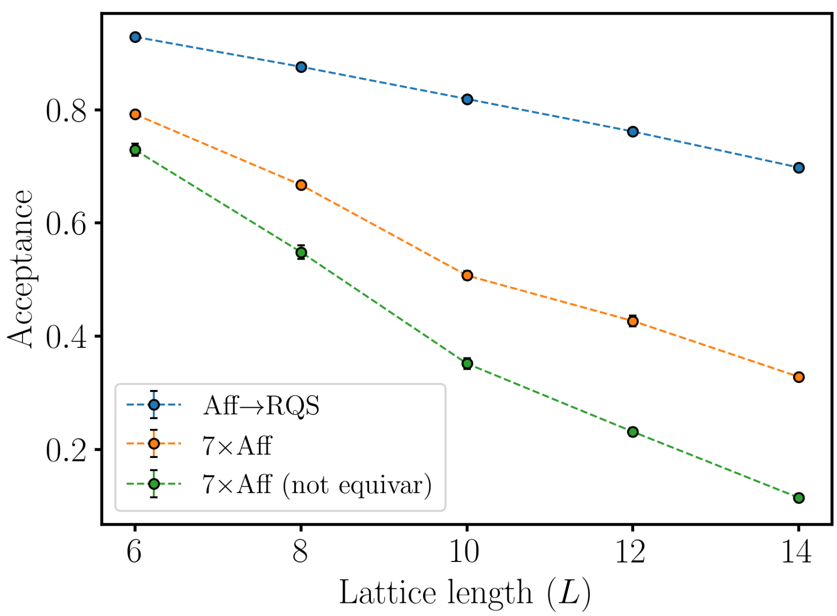
Does it work?

- Want high acceptance rate $\mathbb{E}_{\phi,\phi'\sim\tilde{p}_\theta}[A(\phi \to \phi')]$
- Integrated autocorrelation time $\tau_{\text{int}} \not\propto \xi$





Our models learn much more efficient representations than original...



...but still poor scaling out-of-the-box.

Reasons to get excited

- Large improvements from finding efficient representation
- Highly parallelisable approach (large batch training)

- Low memory requirements (shallow neural networks)

Reasons to go back to sleep

- Scaling sucks
- ϕ^4 is 'easy'

Next steps

- Scaling: exploit symmetries to reduce redundancy
- Scaling: hierarchical / multigrid approaches
- Physics: non-trivial topology e.g. $\mathbb{CP}^{N-1}$ models
- Physics: incorporate ideas from machine-learned RG
- Physics: test recently developed algorithms for gauge fields

---

The bigger picture

Lattice field theories	Real data
Local interactions	☑
Exact symmetries	~
Renormalisable	~
Correlations over multiple scales	☑
Emergent phenomena	☑

Can we use LFT as a test bed where we learn how to *efficiently* encode these properties into machine learning models?

Thanks

Questions? 👀

Joe.Marsh-Rossney@ed.ac.uk