Learning ground states of gauge theories with neural networks

Anthony Ashmore

Skidmore College

Collaborators

Based on `2405.14830` and `260X.XXXX` with

Anuj Apte

(JPMorganChase)

Clay Córdova

(UChicago)

Tzu-Chen Huang

(UChicago)

Many-body quantum systems

Many-body quantum mechanics describes systems with many interacting quantum degrees of freedom

- Spin chains (exactly solvable / integrable models)
- Topological phases of matter (quantum Hall / topological order)
- Fundamental physics (EM / QCD)

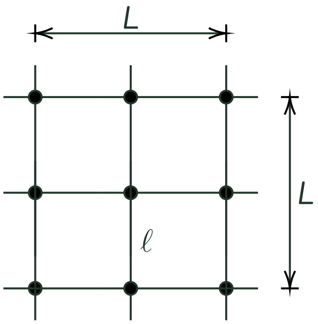
Wavefunction Ψ gives complete description of quantum system

- How do we find the ground-state wavefunction Ψ_0?
- What if system has gauge symmetry?

$\mathbb{Z}_N$ lattice gauge theory in $(2+1)$d

QM description of QFT

- Space discretised as $L \times L$ lattice
- Gauge field degrees of freedom live on links between lattice sites



Clock Q and shift P operators satisfy a $\mathbb{Z}_N$ algebra

$$P^N = Q^N = \mathbf{1}, \qquad P^\dagger Q P = e^{2\pi i/N} Q.$$

States on link ℓ labelled by

$$|q\rangle_\ell, \qquad q \in \{0, \dots, N-1\}$$

Clock associates a $\mathbb{Z}_N$ phase to a link

$$Q|q\rangle = e^{2\pi i q/N}|q\rangle, \qquad P|q\rangle = |q-1\rangle$$

$\mathbb{Z}_N$ lattice gauge theory in $(2+1)$d

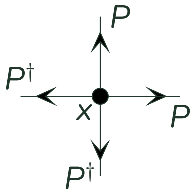
Physics determined by Hamiltonian H with

$$H = \frac{g^2}{2} \sum_{\ell} \left[1 - P_\ell \right] + \frac{1}{2g^2} \sum_{\ell_i \in \Box} \left[1 - Q^\dagger_{\ell_1} Q^\dagger_{\ell_2} Q_{\ell_3} Q_{\ell_4} \right] + \text{h.c.}$$

Coupling g encodes interaction strength

Gauge transformations generated by product of P operators around a lattice site

$$\Theta_x = P_{x,\hat{x}} P_{x,\hat{y}} {P^\dagger}_{x-\hat{x},\hat{x}} {P^\dagger}_{x-\hat{y},\hat{y}}.$$

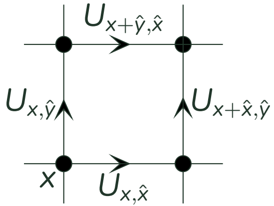


Local symmetry: $[\Theta_x, H] = 0$

- Physical states must be gauge invariant

Gauge-equivariant data

A configuration of the gauge field is given by a set of link phases



$$\mathcal{U} = \{U_{x,\mu}\} \in \mathbb{C}^{2L^2}$$

A wavefunction Ψ maps a configuration $\mathcal{U}$ to a complex number

$$\Psi : \mathbb{C}^{2L^2} \to \mathbb{C}, \qquad \mathcal{U} \mapsto \Psi(\mathcal{U}),$$

where $\Psi(T_\Theta \mathcal{U}) = \Psi(\mathcal{U})$ is a gauge-invariant function

$U_{x,\mu}$ transform non-locally – equivariant data: products of phases around closed loops of all shapes and sizes

An optimisation problem

We seek the ground-state wavefunction Ψ_0 – minimises energy $E[\Psi] = \langle \Psi | H | \Psi \rangle$

$$\Psi_0 = \operatorname{argmin} E[\Psi]$$

High-dimensional optimisation problem for Ψ_0

- Minimum energy $E_0 = E[\Psi_0]$ unknown
- Exponentially large state space: $\dim \mathcal{H} = N^{2L^2}$
- Gauge invariance – highly non-trivial constraint

Want large class of gauge-invariant trial functions for Ψ

Neural quantum states

Neural quantum states

[Carleo, Troyer '16] introduced a new kind of variational ansatz for Ψ: neural quantum state (NQS)

- Wavefunction Ψ_θ represented by neural network with hidden layers
- Parameters θ of network adjusted by stochastic gradient descent to minimise energy $E[\Psi_\theta]$

At end of training, neural network gives approximate ground-state wavefunction

We need a gauge-invariant neural network!

LGE-CNNs

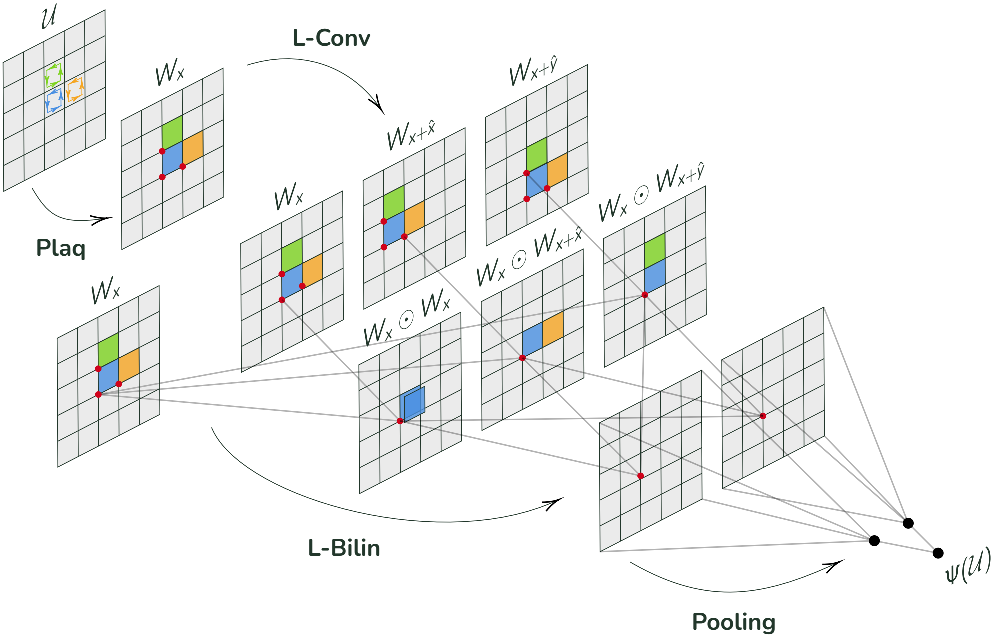
Lattice gauge-equivariant convolutional neural networks (LGE-CNNs) introduced by [Favoni et al. '20] for supervised learning

- Can approximate arbitrary gauge invariant or equivariant functions of a lattice system
- Automatically translation and rotation invariant

Implemented using NetKet 3 [Vicentini et al. '21] and JAX/Flax

(See also [Luo et al. '22] for $\mathbb{Z}_N$, [Luo et al. '22] for U(1), [Spriggs et al. '25] for SU(2))

LGE-CNN architecture



Gauge-invariant trial functions

Deep LGE-CNNs can construct all possible closed loops [Favoni et al. '20] – network constructs a trial function

$$\Psi_\theta(\mathcal{U}) \sim \exp\left[\sum_{\text{lattice}} f_\theta(\text{closed loops})\right]$$

where f_θ is a non-linear gauge-invariant function of closed loops of all sizes and shapes

- Finite network width $\rightarrow$ network keeps only those loops that contribute significantly to ground state

Training minimises $E[\Psi_\theta]$ over θ via stochastic gradient descent

Test bed: $\mathbb{Z}_2$ lattice gauge theory

Ground-state energy

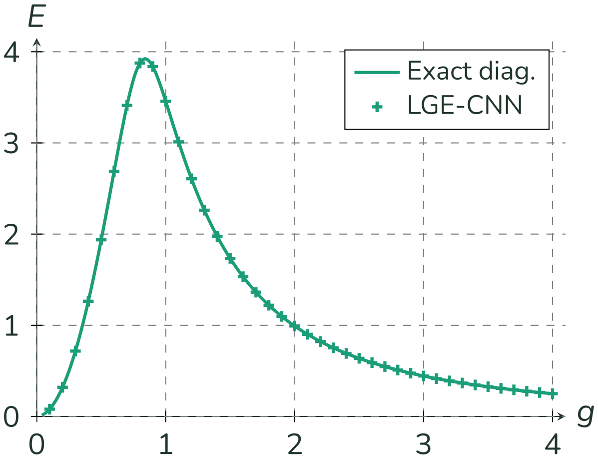


Figure 1: Ground-state energy on a 2×2 lattice as a function of interaction strength g compared with results from exact diagonalisation.

Ground-state energy

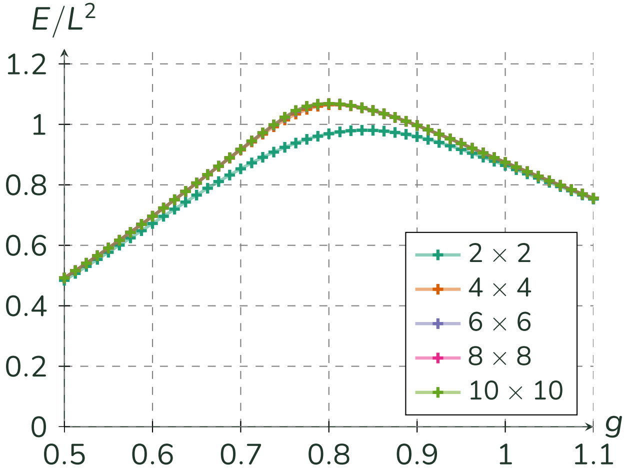


Figure 2: Ground-state energy for varying lattice sizes as a function of interaction strength g.

Phase transition

Energy alone is not that interesting – wavefunction encodes much more!

For $\mathbb{Z}_2$, clock / shift are usual Pauli operators: $Q \equiv \sigma_z$ and $P \equiv \sigma_x$

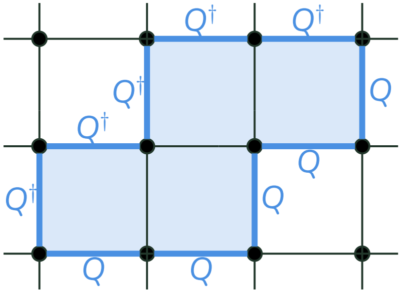
- Gauge theory dual to quantum Ising spin model [Wegner '71]
- In universality class of Ising CFT3 [Wilson, Fisher '72]

As g varies, expect phase transition from deconfined to confined phase (ordered to disordered) – can we see this?

Detecting a phase transition

Wilson loop of Q operators along directed closed path Γ on lattice

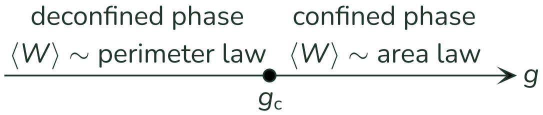
$$W_\Gamma = \prod_\Gamma Q$$



Phase detected by decay of $\langle W_\Gamma \rangle$ [Wegner '71; Kogut '79]

Deconfined phase: decays with perimeter P_Γ of loop

Confined phase: decays with area A_Γ of loop



Order parameter

Expectation value takes the form

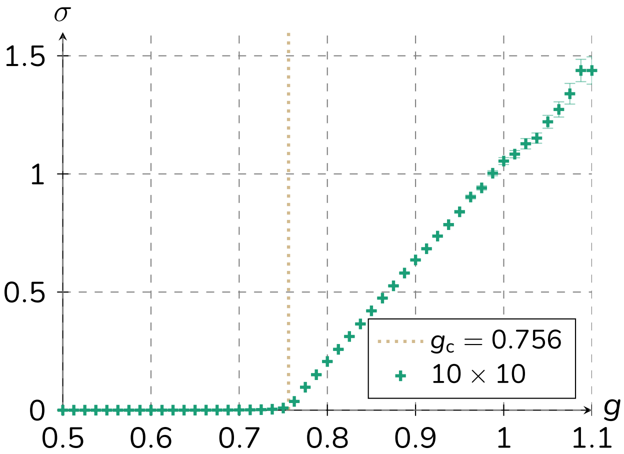
$$\langle W_\Gamma \rangle \sim \exp(-\kappa\, P_\Gamma - \sigma\, A_\Gamma),$$

where string tension σ acts as order parameter

- Use trained NQSs to extract σ as function of g

Signals phase transition at

$$g_c = 0.756$$



Critical exponents

Finite-size scaling and curve collapse determine critical coupling g_c, and correlation length and magnetisation critical exponents ν and β

NQS for $\mathbb{Z}_2$ gauge theory: [AA et al. '24]

$$g_c = 0.7546(8), \qquad \nu = 0.630(3), \qquad \beta = 0.326(4)$$

Previous results from spin models / CFT: [Blöte, Deng '02; Kos et al. '16; Simmons-Duffin '17]

$$g_c = 0.757051, \qquad \nu = 0.629971, \qquad \beta = 0.326419$$

Conclusions

Summary

Neural networks are powerful tools for solving high-dimensional optimisation problems

- Gauge-equivariant networks extend this to local gauge symmetries
- Salient gauge-equivariant features constructed "on the fly"

Provide accurate approximation of ground state for lattice gauge theories – can be used to probe phase transitions, critical exponents, and more

Future directions

Extension to finite groups A_N / D_N [Kitaev ‘03], $SU(N)$ Yang–Mills [Spriggs et al. ‘25]

- Network architecture remains the same

Inspiration for mathematics?

- Optimisation over functions on quotient spaces by local symmetry groups commonly needs invariant basis functions or explicit gauge choice – often impractical
- Gauge-equivariant neural networks might provide a new approach for constrained variational problems in geometry / mathematical physics

Extra

Curve collapse

Near phase transition, finite-size scaling predicts a specific scaling with lattice size L

- Observables are functions of renormalised coupling $\tilde{g} = L^{1/\nu}(g - g_c)/g_c$

e.g. ‘t Hooft string disorder parameter should scale as

$$\langle T \rangle = L^{-2\beta/\nu} t(\tilde{g})$$

where ν and β are correlation length and magnetisation critical exponents

“Curve collapse”: plot of $L^{2\beta/\nu}\langle T \rangle$ vs $\tilde{g}$ should be independent of L

- Extract g_c, β and ν by minimising distance between curves

Curve collapse for $\langle T \rangle$

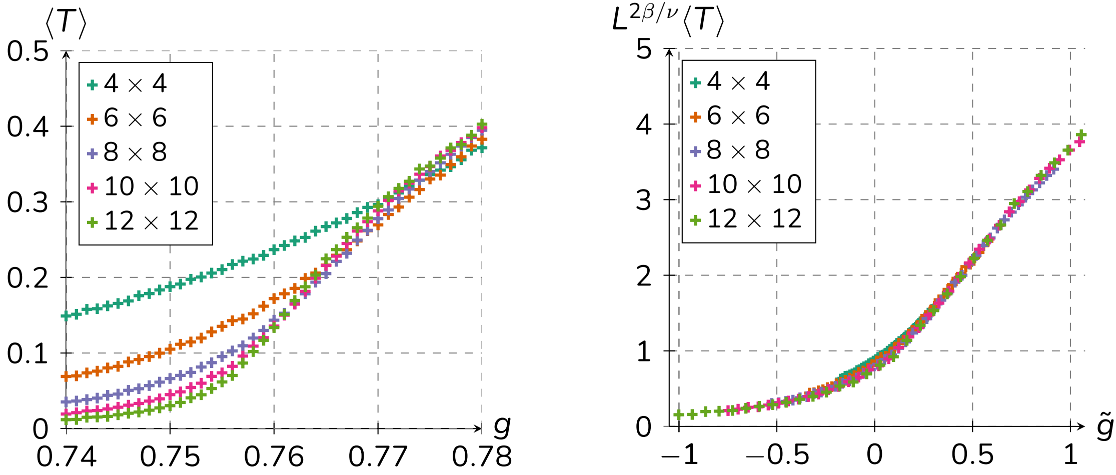


Figure 3: Curve collapse for ‘t Hooft string disorder parameter $T = \prod_{\tilde{\Gamma}} P$.

What do we learn?

What can we learn from the trained network?

As $g \to \infty$, network parameters show loop phases not used at all

- $\Psi(\mathcal{U}) =$ constant

As $g \to 0$, 1×1 loops used, but larger loops not constructed

- $\Psi(\mathcal{U})$ non-zero on trivial flux configurations, $W_x = 1$

Near phase transition, loops of all shapes and sizes constructed

- Scale invariance $\to$ conformal field theory

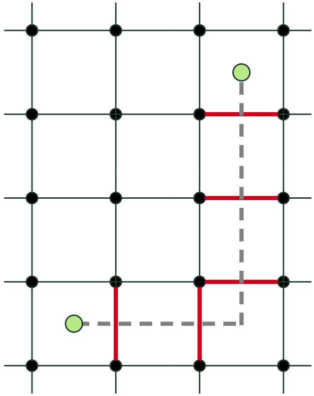
Identifying the critical coupling

Ordered phase	Disordered phase
$g < g_\mathrm{c}$	$g > g_\mathrm{c}$
$\langle W \rangle \sim$ perimeter law	$\langle W \rangle \sim$ area law
$\langle T \rangle \sim$ exp. decay with distance	$\langle T \rangle \sim$ constant
electric flux lines condensed	mag. monopoles condensed
$\mathbb{Z}_2^{(1)}$ broken	$\mathbb{Z}_2^{(1)}$ preserved
"deconfined"	"confined"

[Rayhaun, Williamson '23]

Cleaner to look at decay of 't Hooft string

Disorder parameter



't Hooft string of P operators, $T = \prod_{\tilde{\Gamma}} P$, along open path $\tilde{\Gamma}$ between two points on *dual* lattice

- $\langle T \rangle$ independent of path due to Gauss' law
- In deconfined phase, creates a pair of quasi-particles (magnetic monopoles) – $\langle T \rangle$ decays exponentially with distance
- In confined phase, $\langle T \rangle$ independent of distance – "monopoles condensed"

Disorder parameter

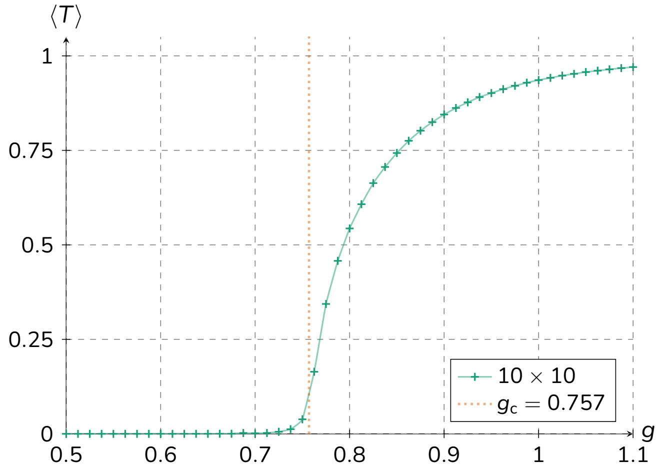
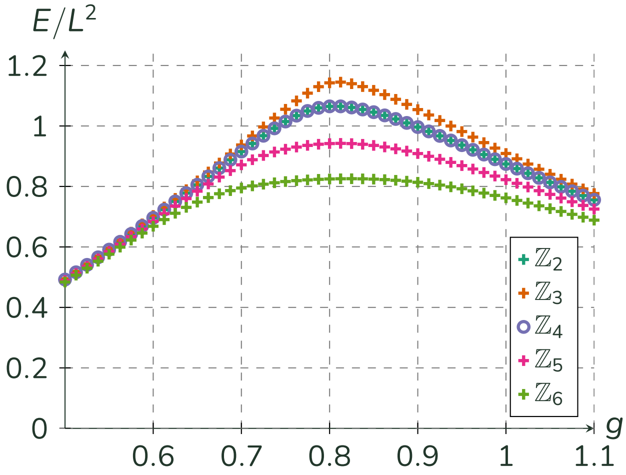


Figure 4: Lattice average of the 't Hooft string operator near the critical point for $L = 10$. Distance $L/2 = 5$ between ends of 't Hooft string.

$\mathbb{Z}_N$ energies

Ground-state energy for $\mathbb{Z}_N$ theories on a 4×4 lattice

The $\mathbb{Z}_4$ theory splits into two decoupled $\mathbb{Z}_2$ theories for every value of the coupling [Grosse et al. '81]



Stochastic reconfiguration

Stochastic gradient descent with a non-trivial weight-space metric, c.f. imaginary time evolution [Sorella et al. ‘07]

$$(1-\epsilon H)|\Psi\rangle \approx e^{-\epsilon H}|\Psi\rangle = e^{-\epsilon H}\sum_{i=0} c_i|i\rangle = e^{-\epsilon E_0}c_0|0\rangle + e^{-\epsilon E_1}c_1|1\rangle + \ldots$$

- Iterating this will project a trial wavefunction onto the ground state

Construct a weight update rule so that the updated state is $(1-\epsilon H)|\Psi\rangle$

$$\theta_\alpha \mapsto \theta_\alpha - \eta(S^{-1})_{\alpha\beta}R_\beta$$

where

$$O_\alpha|\psi\rangle = \partial_\alpha|\psi\rangle, \qquad R_\alpha = \langle O_\alpha^\dagger H\rangle - \langle O_\alpha^\dagger\rangle\langle H\rangle, \qquad S_{\alpha\beta} = \langle O_\alpha^\dagger O_\beta\rangle - \langle O_\alpha^\dagger\rangle\langle O_\beta\rangle$$

Calculating expectation values

Given a state $|\Psi\rangle$, one can estimate $E[\Psi] = \langle H \rangle$ via

$$\langle H \rangle = \langle \Psi | H | \Psi \rangle = \mathbb{E}_{\mathcal{U} \sim \mathrm{p}(\mathcal{U})} \left[E_{\mathrm{loc}}(\mathcal{U}) \right],$$

where $\mathbb{E}_{\mathcal{U} \sim \mathrm{p}(\mathcal{U})}$ takes the expectation value over $\mathcal{U}$, with $\mathcal{U}$ chosen according to the PDF $p(\mathcal{U}) = |\Psi(\mathcal{U})|^2$, and the "local energy" is

$$E_{\mathrm{loc}}(\mathcal{U}) = \sum_{\mathcal{U}'} \langle \mathcal{U} | H | \mathcal{U}' \rangle \frac{\Psi(\mathcal{U}')}{\Psi(\mathcal{U})}.$$