Title: Machine learning for lattice field theory and back

Speakers: Gert Aarts

Series: Machine Learning Initiative

Date: March 10, 2023 - 11:00 AM

URL: https://pirsa.org/23030101

Abstract: Recently, machine learning has become a popular tool to use in fundamental science, including lattice field theory. Here I will report on some recent progress, including the Inverse Renormalisation Group and quantum-field theoretical machine learning, combining insights of lattice field theory and machine learning in a hopefully constructive manner.

Zoom link: https://pitp.zoom.us/j/95456375462?pwd=WmtZMloyclAyZzBwVEZHQ3gxVnkrUT09

Machine learning for lattice field theory and back

Gert Aarts

Swansea University
Prifysgol Abertawe

FONDAZIONE BRUNO KESSLER

ECT*
EUROPEAN CENTRE FOR THEORETICAL STUDIES IN NUCLEAR PHYSICS AND RELATED AREAS

Perimeter, March 2023

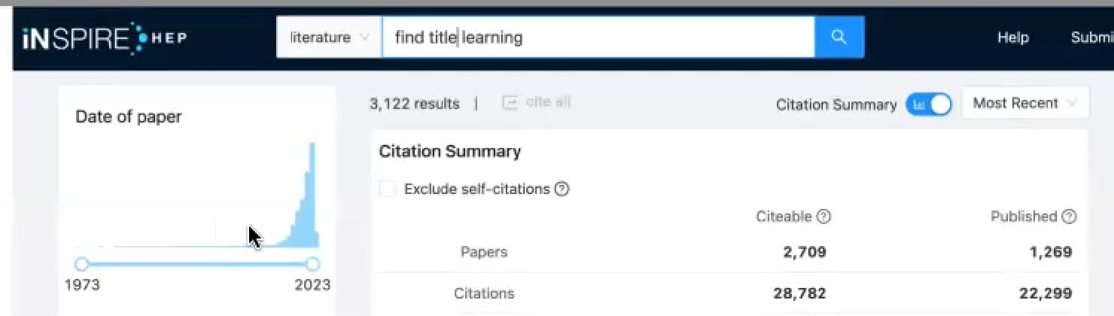
talkperimeter-mar2023.pdf
Page 2 of 47

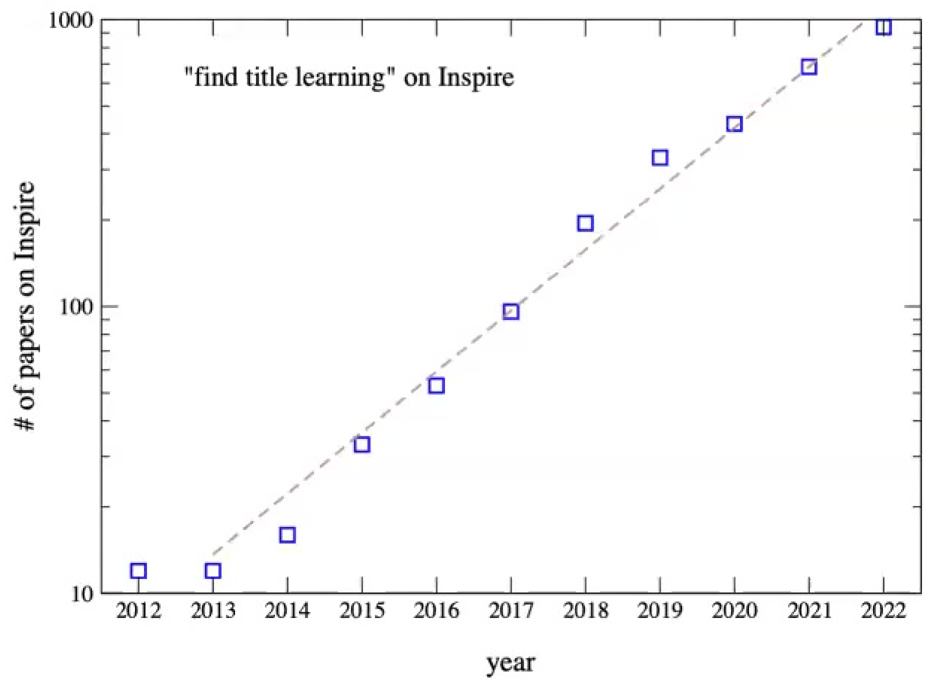
Introduction

- past five years or so has seen a rapid rise of applications of machine learning (ML) in fundamental science, particle physics, theoretical physics
- of course ML has been around for quite some time, especially in experimental particle physics
- nevertheless, there is an **exponential** increase in activity

talkperimeter-mar2023.pdf
Page 3 of 47

Introduction





- find title **learning** on the Inspire data base (high-energy physics)
- exponential growth!

Outline

- classification: order-disorder transition (by now classic application)
- inverse renormalisation group (new application and concepts)
- quantum field-theoretical machine learning (new conceptual ideas to explore)
- outlook

biased towards own work and interests in lattice field theory

talkperimeter-mar2023.pdf
Page 5 of 47

Based on the following papers:

with Dimitrios Bachtis and Biagio Lucini:

- ✓ Extending machine learning classification capabilities with histogram reweighting

  Phys. Rev. E 102 (2020) 033303 [2004.14341 [cond-mat.stat-mech]]
- ✓ Mapping distinct phase transitions to a neural network

  Phys. Rev. E 102 (2020) 053306 [2007.00355 [cond-mat.stat-mech]]
- ✓ Quantum field-theoretic machine learning

  Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]

- ✓ Inverse renormalisation group in quantum field theory

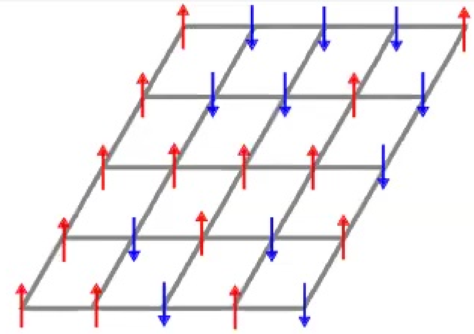
  Phys. Rev. Lett. 128 (2022) 081603 [2107.00466 [hep-lat]], with D Bachtis, B Lucini and Francesco di Renzo

- ✓ Scalar field RBM (in preparation), with Chanju Park and Biagio Lucini

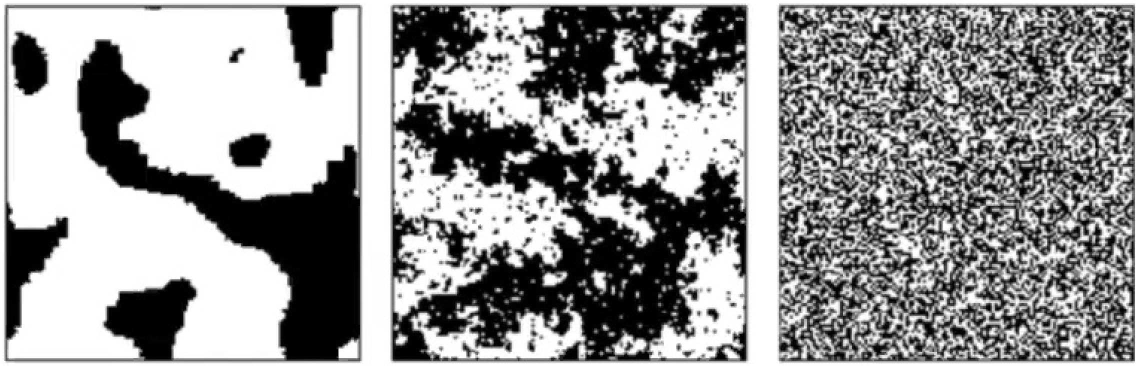
talkperimeter-mar2023.pdf
Page 6 of 47

Search

Classification of phases of matter



- matter can exist in different phases
- prototype: 2d Ising model -> ordered/disordered or cold/hot phases
- task: determine phase a configuration is in, determine critical coupling or temperature



Ordered -- ? -- Disordered

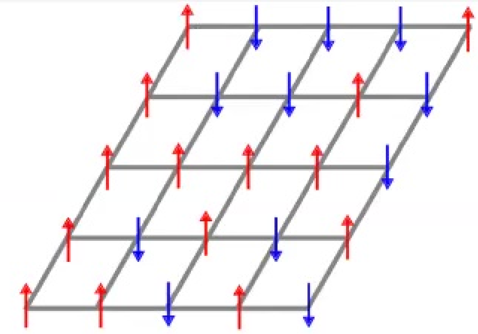
Published: 13 February 2017

Machine learning phases of matter

Juan Carrasquilla & Roger G. Melko

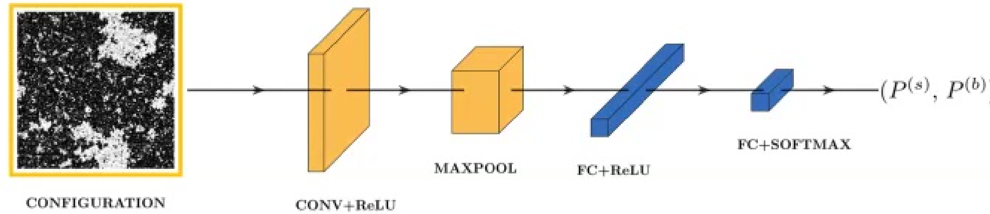
Nature Physics **13**, 431–434(2017) | Cite this article

2d Ising model



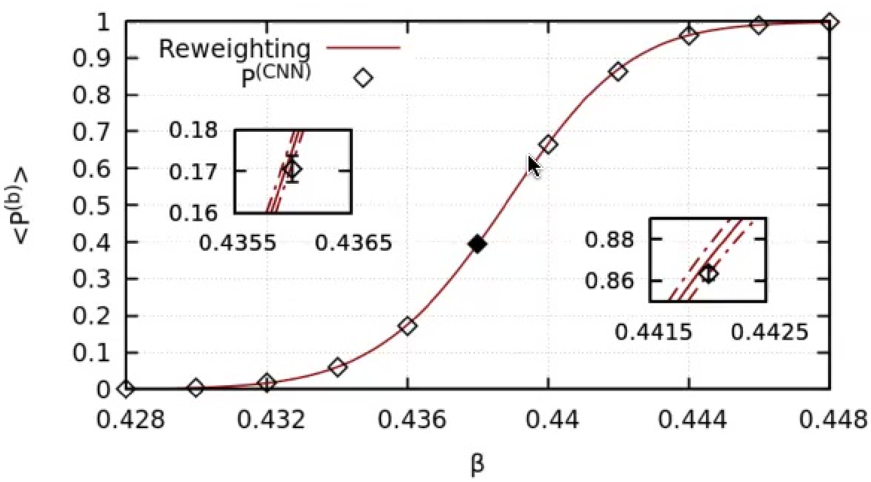
- $Z = \mathrm{Tr}\, e^{-\beta E}$ with $E = -\sum_{<ij>} s_i s_j$ $(s_i = \pm 1)$
- critical coupling or inverse temperature β_c
- correlation length ξ, magnetic susceptibility χ diverge at transition
- critical exponents $\xi \sim |t|^{-\nu}$ $\chi \sim |t|^{-\gamma}$ reduced temperature $t = \dfrac{\beta_c - \beta}{\beta_c}$
- $\nu = 1,$ $\gamma/\nu = 7/4\,,$ $\beta_c = \frac{1}{2}\ln(1+\sqrt{2}) \approx 0.440687$
- finite-size scaling $|t| = \left|\dfrac{\beta_c - \beta_c(L)}{\beta_c}\right| \sim \xi^{-\frac{1}{\nu}} \sim L^{-\frac{1}{\nu}}$ $\chi \sim L^{\gamma/\nu}$

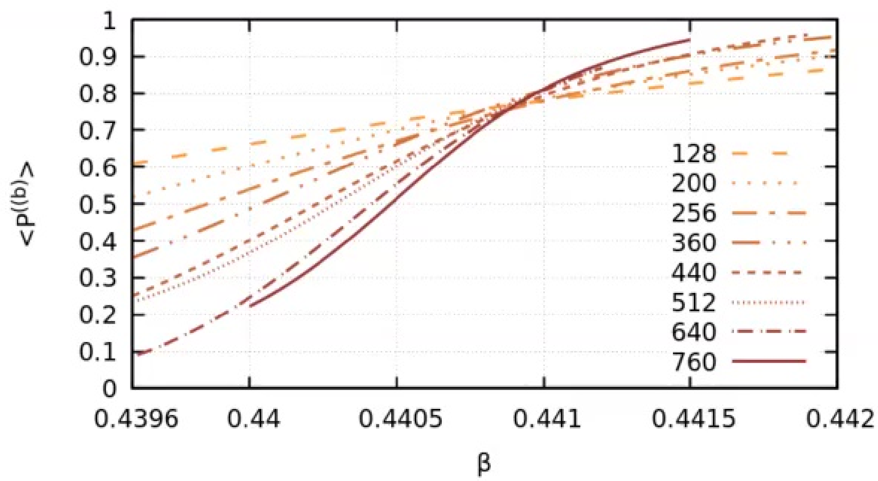
Probability



train NN away from the phase transition: $\beta \leq 0.41$ and $\beta \geq 0.47$, $\beta_c \sim 0.440687$

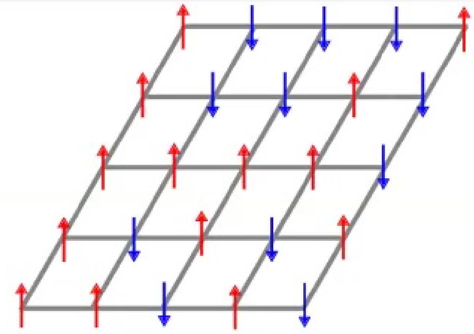
investigate unseen configurations at intermediate β on lattices of different sizes





probability behaves as an approximate order parameter

2d Ising model

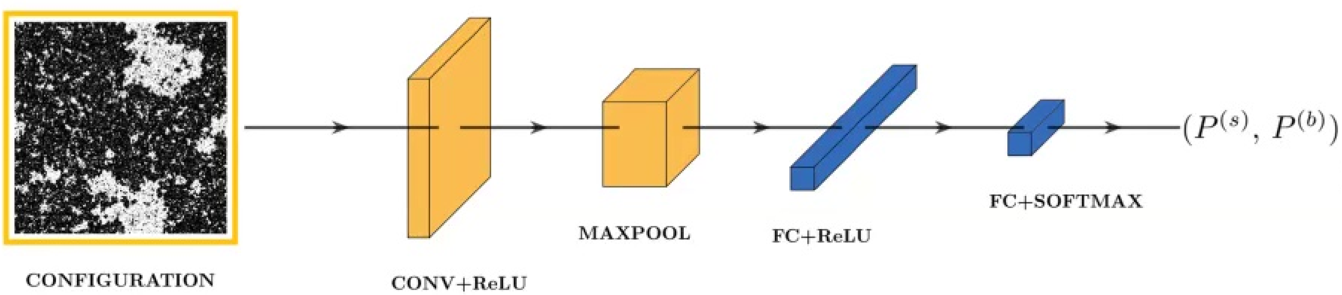


- $Z = \mathrm{Tr}\, e^{-\beta E}$ with $E = -\sum_{<ij>} s_i s_j$ $\quad (s_i = \pm 1)$
- critical coupling or inverse temperature β_c
- correlation length ξ, magnetic susceptibility χ diverge at transition
- critical exponents $\quad \xi \sim |t|^{-\nu} \quad \chi \sim |t|^{-\gamma}$ $\qquad$ reduced temperature $t = \dfrac{\beta_c - \beta}{\beta_c}$
- $\nu = 1, \qquad \gamma/\nu = 7/4\,, \qquad \beta_c = \frac{1}{2}\ln(1+\sqrt{2}) \approx 0.440687$
- finite-size scaling $\quad |t| = \left|\dfrac{\beta_c - \beta_c(L)}{\beta_c}\right| \sim \xi^{-\frac{1}{\nu}} \sim L^{-\frac{1}{\nu}} \qquad \chi \sim L^{\gamma/\nu}$

✓Extending machine learning classification capabilities with histogram reweighting
Phys. Rev. E 102 (2020) 033303 [2004.14341 [cond-mat.stat-mech]]

What’s new?

- by now well-established procedure, what can we add?
- interpret output from a NN as an observable in a statistical system
- input: configurations, distributed according to Boltzmann weight
- output: observable, “order parameter” in statistical system



$$\Rightarrow \quad < P > = \frac{1}{Z} \sum_i P_i \, e^{-\beta E_i}$$

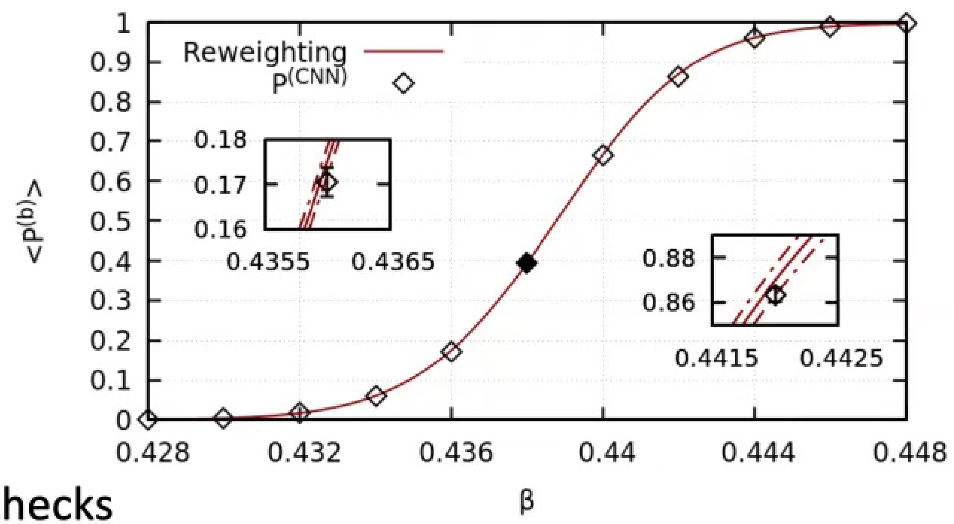
Output of NN as physical observable

- opens up possibility to use “standard” numerical/statistical methods

→ histogram reweighting: extrapolation to other parameter values

- starting from computation at given β_0: extrapolate to other β values

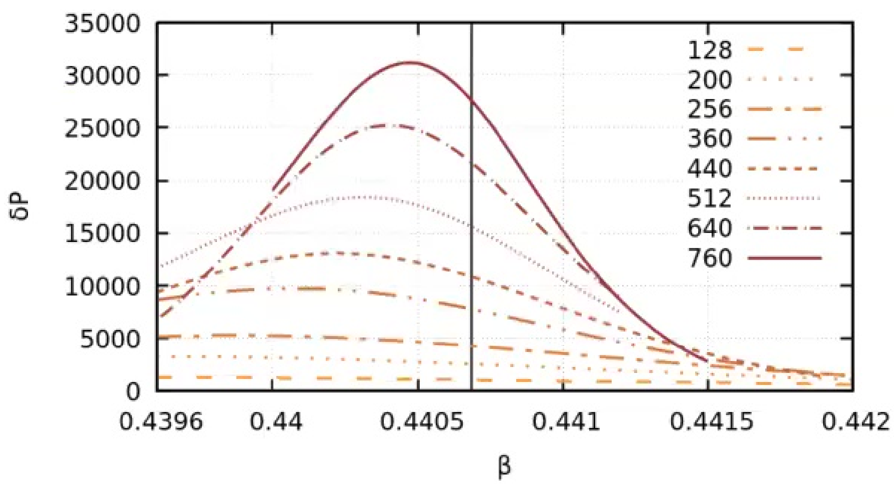
$$< P > (\beta) = \frac{\sum P_i e^{-(\beta-\beta_0)E_i}}{\sum e^{-(\beta-\beta_0)E_i}}$$

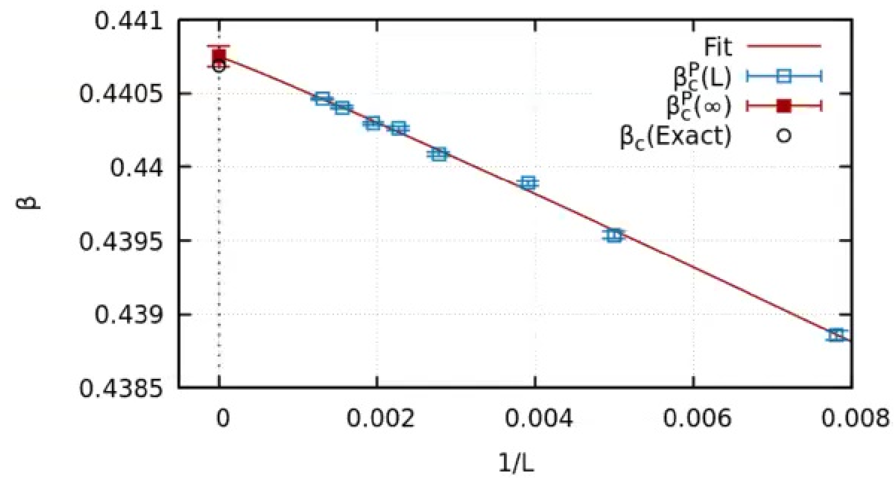


✓ filled diamond at β_0
✓ line obtained by reweighting in β
✓ open diamonds are independent cross checks

Critical behaviour from NN observables

- Determine L dependent susceptibility δP and its maximum at $\beta_c(L)$



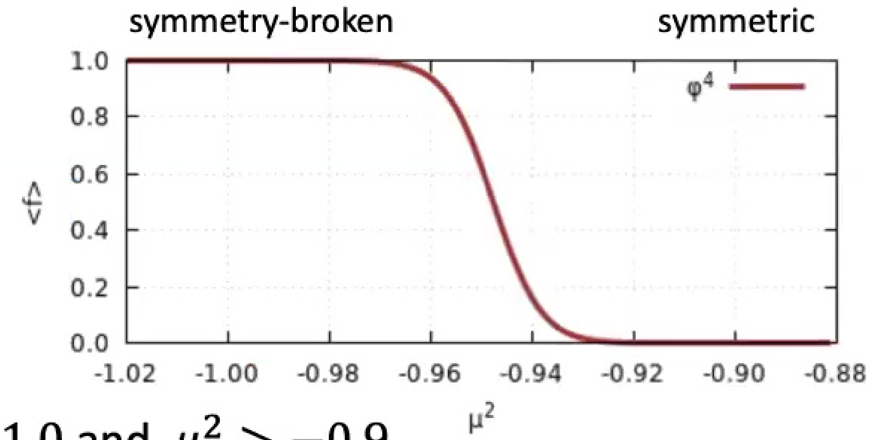


Extract critical properties from NN observables only

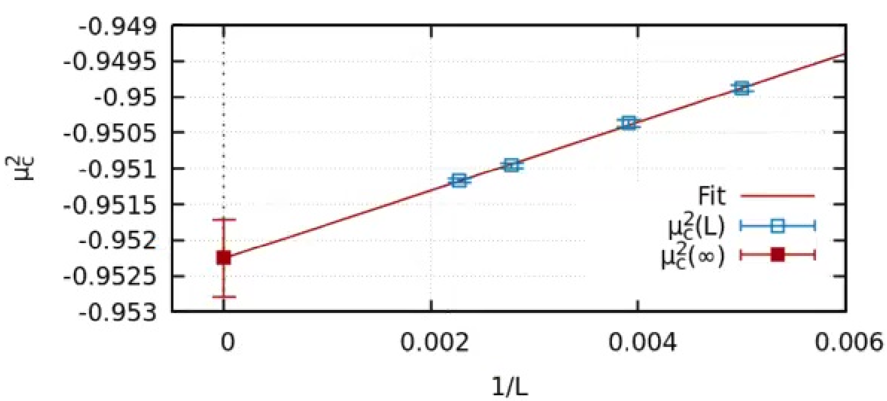
	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	7/4 =1.75

✓ Mapping distinct phase transitions to a neural network
Phys. Rev. E 102 (2020) 053306 [2007.00355 [cond-mat.stat-mech]]

φ^4 scalar field theory



- reweight in mass parameter, μ^2
- identify regions where phase is clear
- transfer learning: retrain NN using $\mu^2 < -1.0$ and $\mu^2 > -0.9$
- repeat finite-size scaling analysis as in 2d Ising model



	μ_c^2	ν	γ/ν
CNN+Reweighting	-0.95225(54)	0.99(34)	1.78(7)

- same universality class as 2d Ising model
- critical mass in agreement with results obtained with standard methods (Binder cumulant, susceptibility)

talkperimeter-mar2023.pdf
Page 13 of 47

Intermediate summary

✓ proposed to identify NN outputs as observables in statistical physics

✓ introduced histogram reweighting to employ in supervised machine learning

✓ critical properties obtained from a finite-size scaling analysis using quantities derived from NN alone

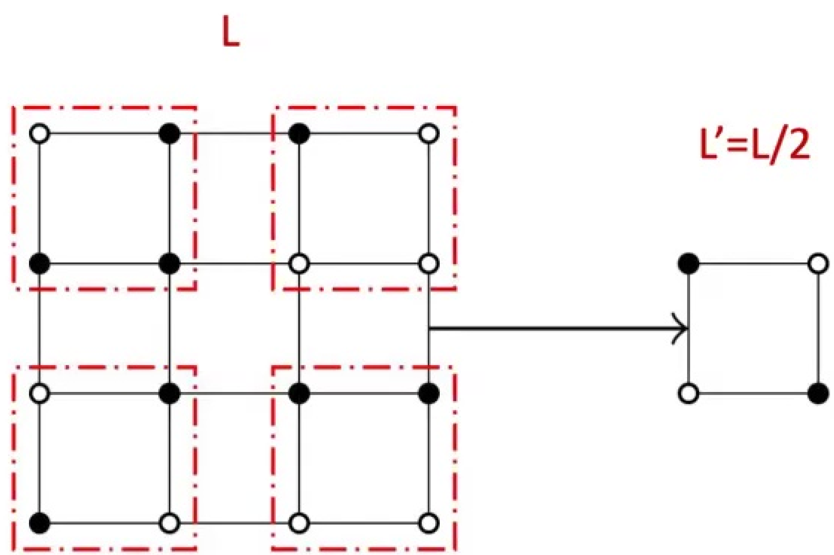
talkperimeter-mar2023.pdf
Page 14 of 47

Renormalisation Group (RG)

- standard renormalisation group: coarse-graining, blocking transformation, integrating out degrees of freedom, …

- Ising model: Kadanoff block spin
- majority rule
- reduction of degrees of freedom
- study critical scaling

- not invertible: semi-group



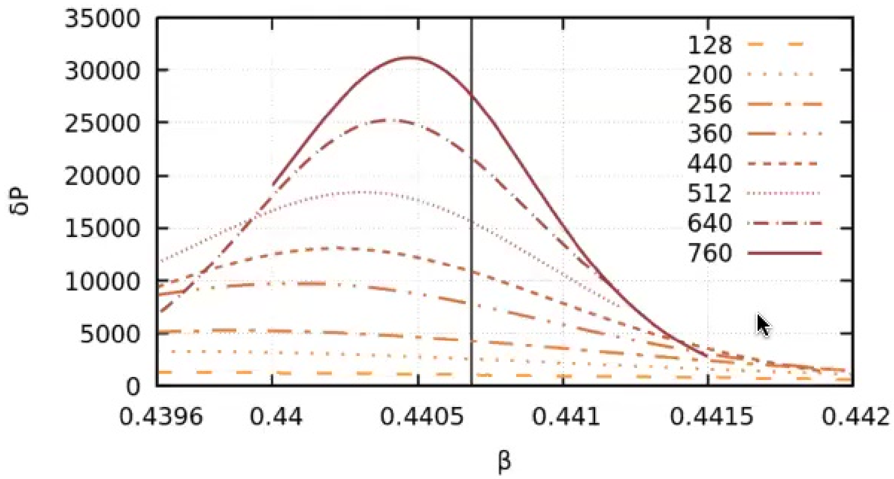
talkperimeter-mar2023.pdf
Page 13 of 47

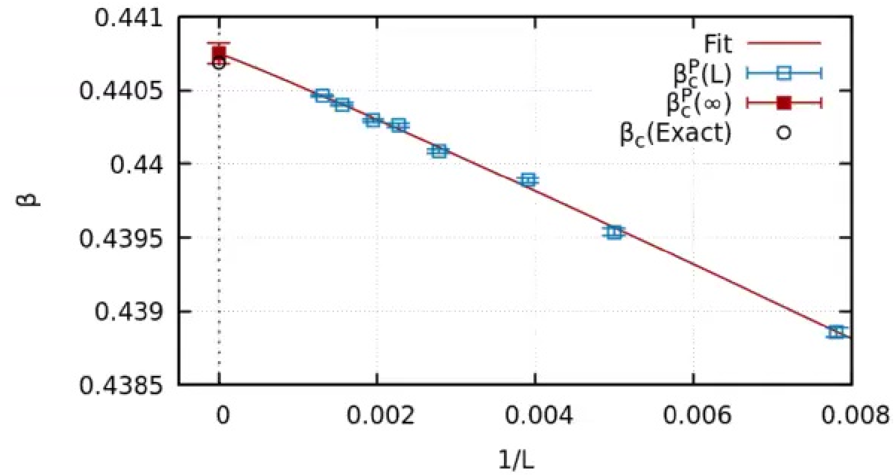
Intermediate summary

- ✓ proposed to identify NN outputs as observables in statistical physics
- ✓ introduced histogram reweighting to employ in supervised machine learning
- ✓ critical properties obtained from a finite-size scaling analysis using quantities derived from NN alone

Critical behaviour from NN observables

- Determine L dependent susceptibility δP and its maximum at $\beta_c(L)$





Extract critical properties from NN observables only

	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	7/4 =1.75

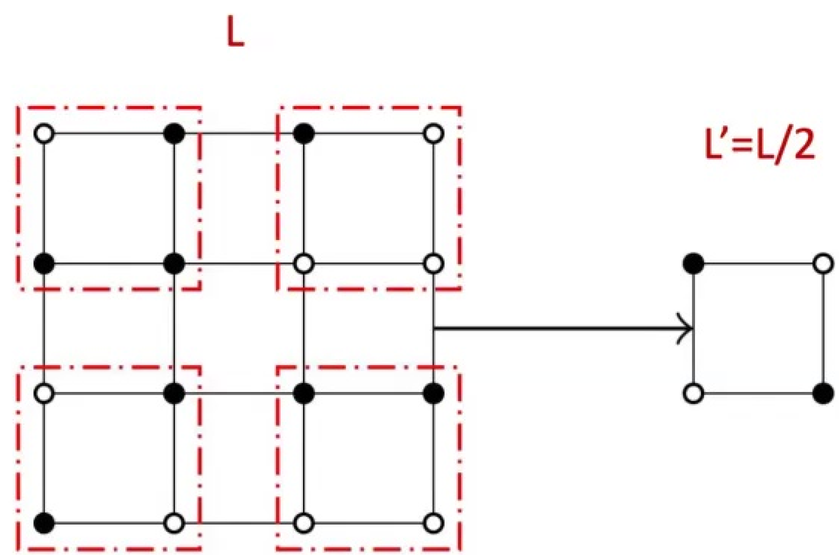
talkperimeter-mar2023.pdf
Page 14 of 47

Renormalisation Group (RG)

- standard renormalisation group: coarse-graining, blocking transformation, integrating out degrees of freedom, ...

- Ising model: Kadanoff block spin
- majority rule
- reduction of degrees of freedom
- study critical scaling

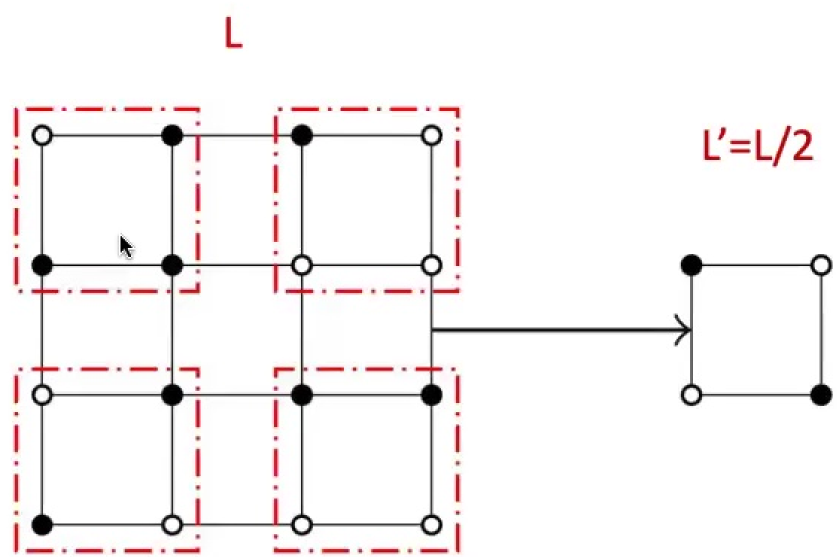
- not invertible: semi-group



Renormalisation group

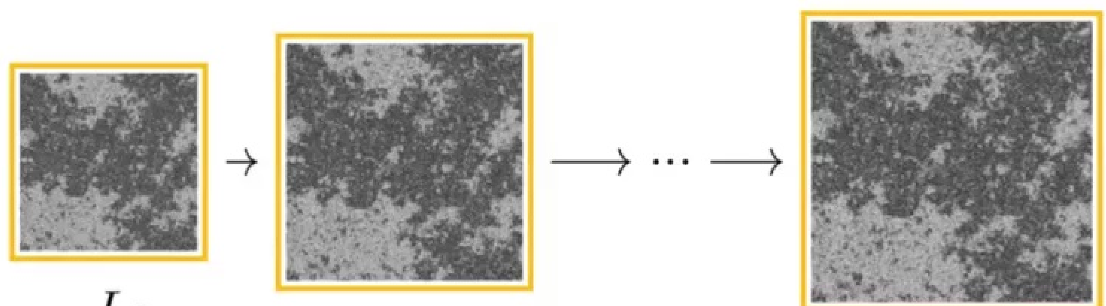
- generates flow in parameter space
- due to repeated blocking: run out of degrees of freedom
- need to start with large system to apply RG step multiple times
- large systems, close to a transition,

  suffer from critical slowing down



Inverse renormalisation group

- what if we could invert the RG?
- add degrees of freedom, fill in the 'details'
- inverse flow in parameter space
- can be applied arbitrary number of steps
- evade critical slowing down



L_0 $\quad$ $L_1 = bL_0$ $\quad$ $L_j = b^{(j-i)}L_i$

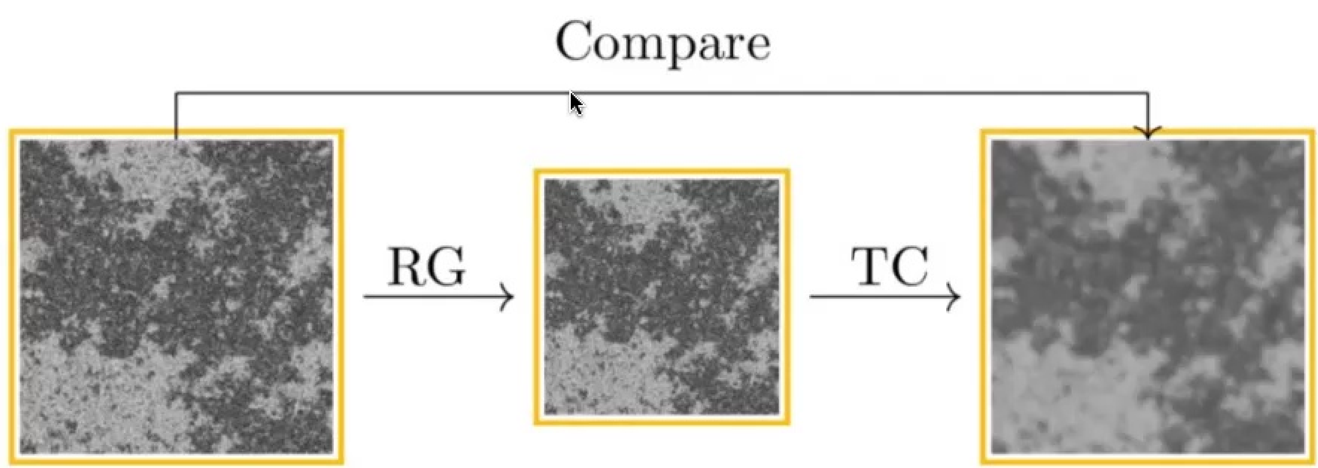
for Ising model: Inverse Monte Carlo Renormalization Group Transformations for Critical Phenomena, D. Ron, R. Swendsen, A. Brandt, Phys. Rev. Lett. 89, 275701 (2002)

Search

✓ Inverse renormalisation group in quantum field theory
Phys. Rev. Lett. 128 (2022) 081603 [2107.00466 [hep-lat]]

How to devise an inverse transformation?

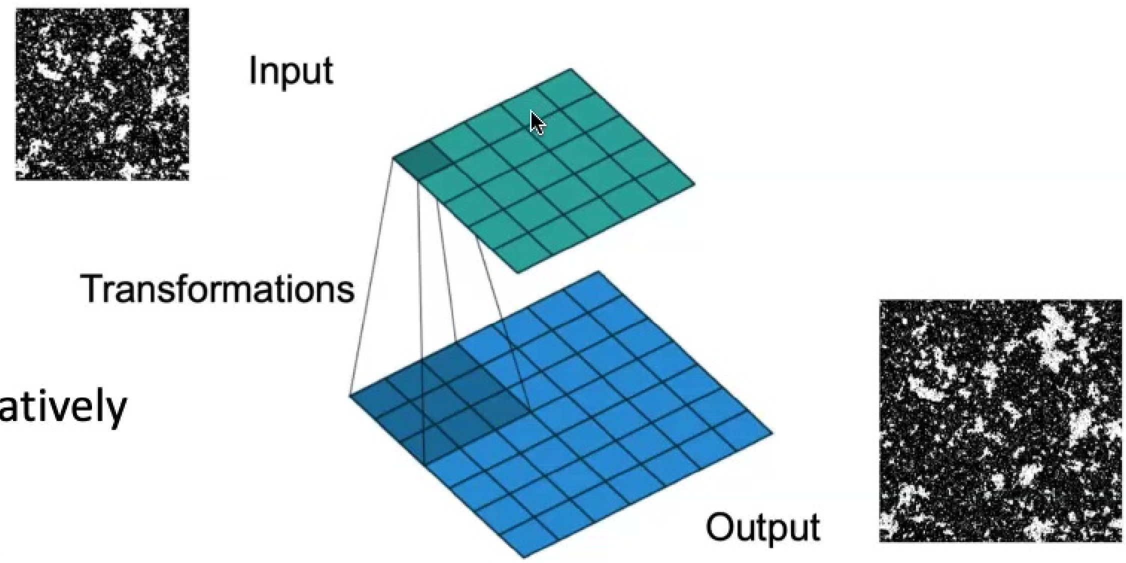
- new degrees of freedom should be introduced
- learn a set of transformations (*transposed convolutions*) to invert a standard RG step
- minimise difference between original and constructed configuration



Search

Inverse renormalisation group

Transposed convolutions



- local transformation
- apply inverse transformations iteratively
- evade critical slowing down
- generate flow in parameter space
- invariance at critical point

Application to φ^4 scalar field theory

- repeated steps
- locking in on critical point

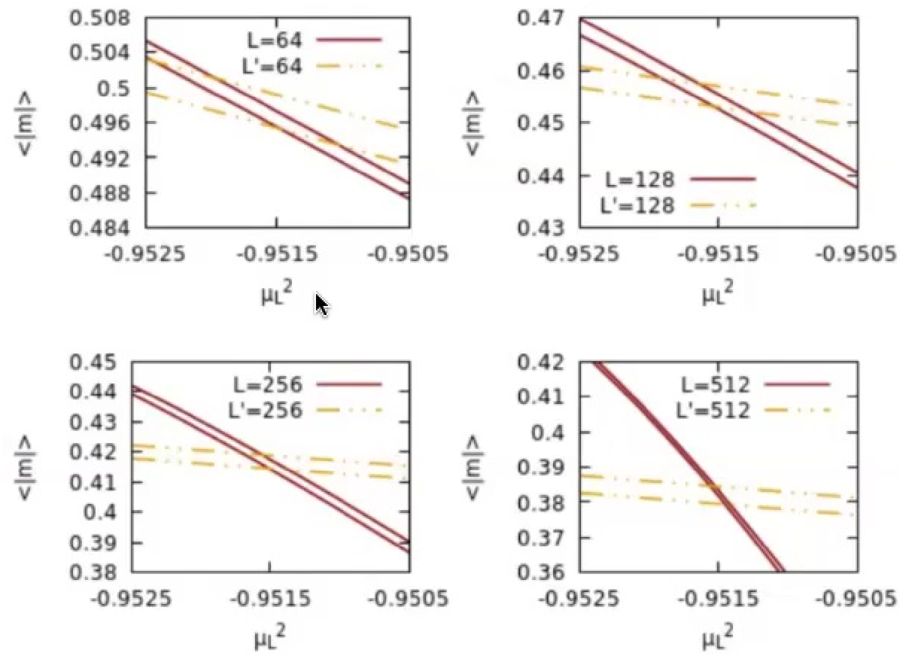
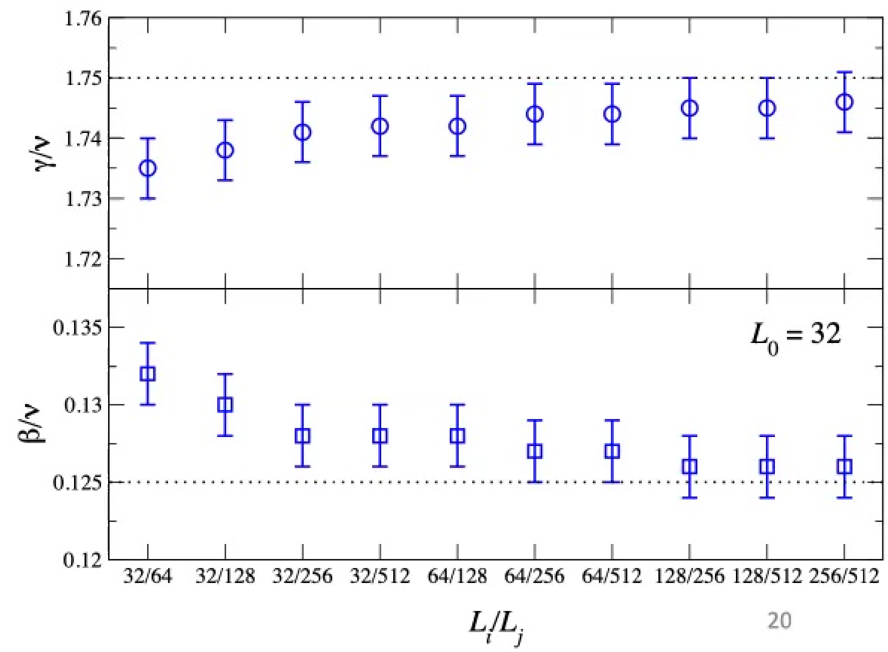


TABLE I. Values of the critical exponents γ/ν and β/ν. The original system has lattice size $L = 32$ in each dimension and its action has coupling constants $\mu_L^2 = -0.9515$, $\lambda_L = 0.7$, and $\kappa_L = 1$. The rescaled systems are obtained through inverse renormalization group transformations.

L_i/L_j	32/64	32/128	32/256	32/512	64/128	64/256	64/512	128/256	128/512	256/512
γ/ν	1.735(5)	1.738(5)	1.741(5)	1.742(5)	1.742(5)	1.744(5)	1.744(5)	1.745(5)	1.745(5)	1.746(5)
β/ν	0.132(2)	0.130(2)	0.128(2)	0.128(2)	0.128(2)	0.127(2)	0.127(2)	0.126(2)	0.126(2)	0.126(2)

Application to φ^4 scalar field theory

- start with lattice of size 32^2 and apply IRG steps repeatedly
- $32^2 \to 64^2 \to 128^2 \to 256^2 \to 512^2$
- IRG flow towards critical point
- extract critical exponents
  γ/ν and β/ν from comparison
  between two volumes
- constructed a large (512^2) lattice
  very close to criticality
  without critical slowing down



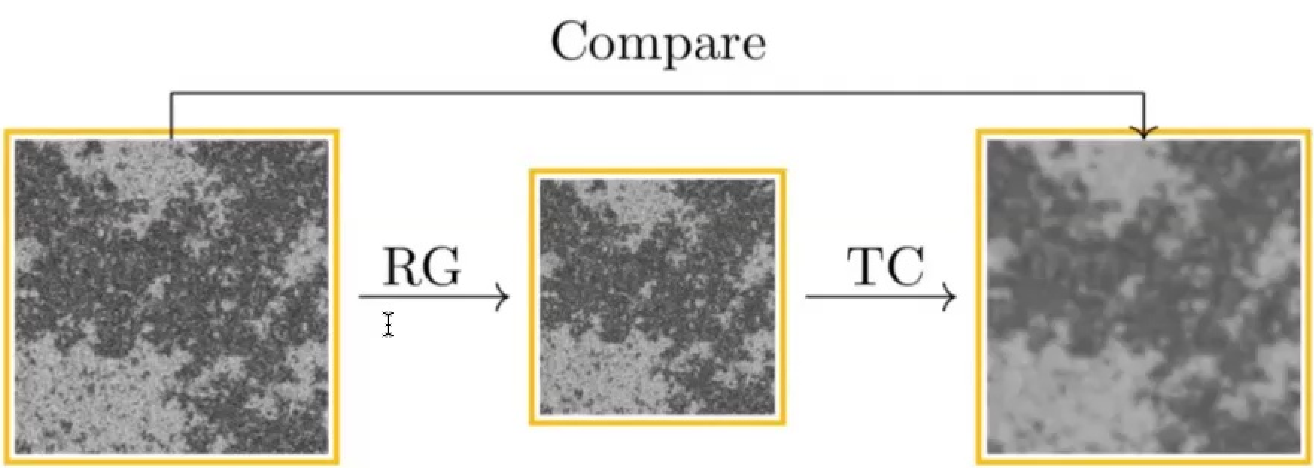
Intermediate summary: inverse RG

- ✓ flow to critical point without critical slowing down
- ✓ reach large lattices from easy-to-simulate lattice sizes
- ✓ relies on 'reliable' blocking step (nontrivial: scalar field majority rule is new)
- ✓ new concept for continuous field theories

✓ Inverse renormalisation group in quantum field theory
Phys. Rev. Lett. 128 (2022) 081603 [2107.00466 [hep-lat]]

How to devise an inverse transformation?

- new degrees of freedom should be introduced
- learn a set of transformations (*transposed convolutions*) to invert a standard RG step
- minimise difference between original and constructed configuration

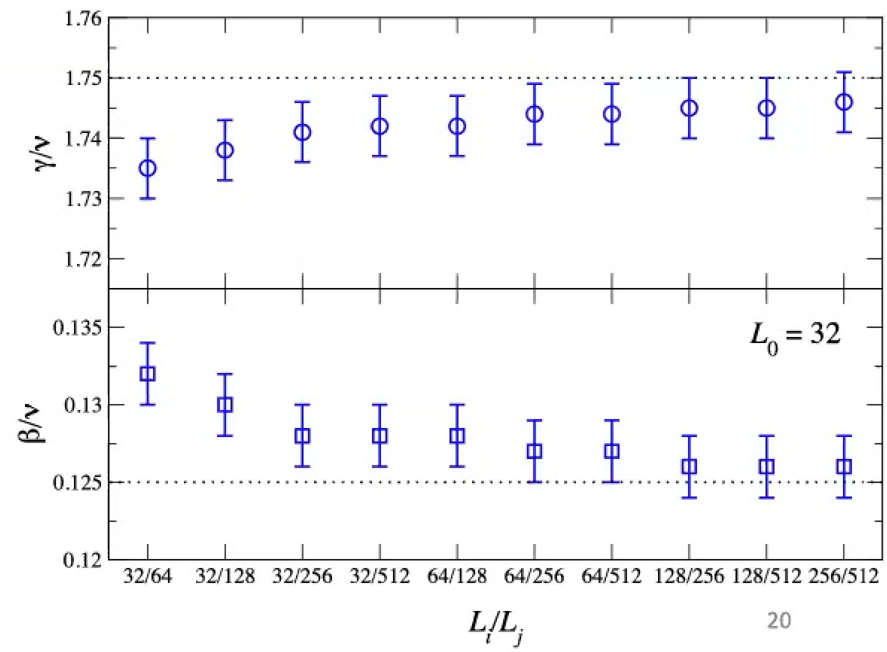


Intermediate summary: inverse RG

- ✓ flow to critical point without critical slowing down
- ✓ reach large lattices from easy-to-simulate lattice sizes
- ✓ relies on ‘reliable’ blocking step (nontrivial: scalar field majority rule is new)
- ✓ new concept for continuous field theories

Application to φ^4 scalar field theory

- start with lattice of size 32^2 and apply IRG steps repeatedly
- $32^2 \rightarrow 64^2 \rightarrow 128^2 \rightarrow 256^2 \rightarrow 512^2$
- IRG flow towards critical point
- extract critical exponents
  γ/υ and β/υ from comparison
  between two volumes
- constructed a large (512^2) lattice
  very close to criticality
  without critical slowing down

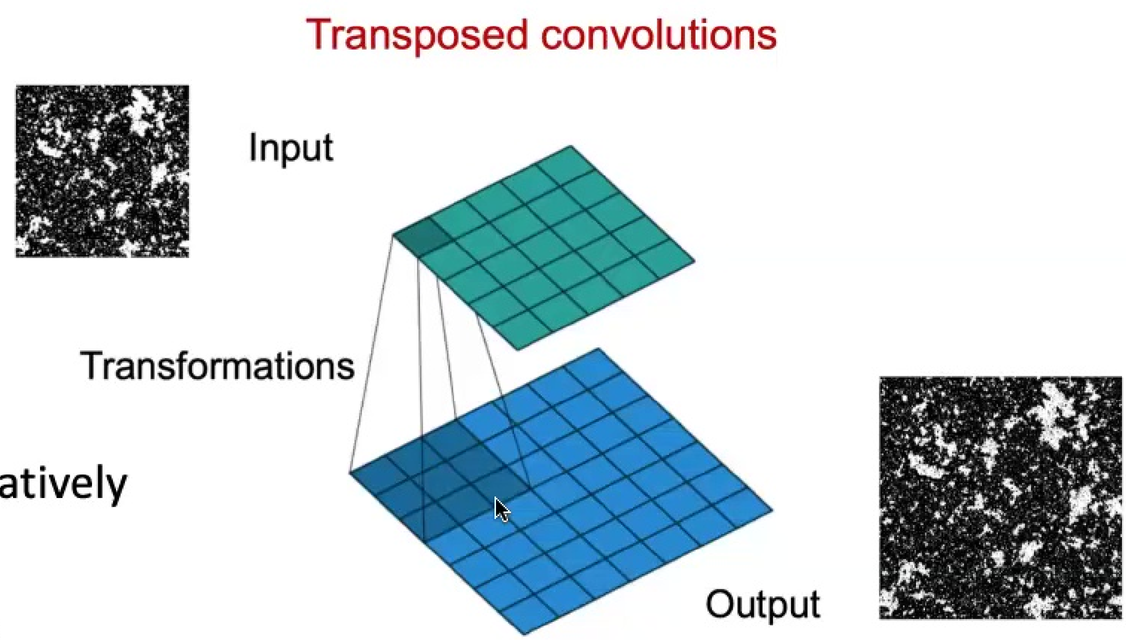


Intermediate summary: inverse RG

- ✓ flow to critical point without critical slowing down
- ✓ reach large lattices from easy-to-simulate lattice sizes
- ✓ relies on 'reliable' blocking step (nontrivial: scalar field majority rule is new)
- ✓ new concept for continuous field theories

talkperimeter-mar2023.pdf
Page 18 of 47

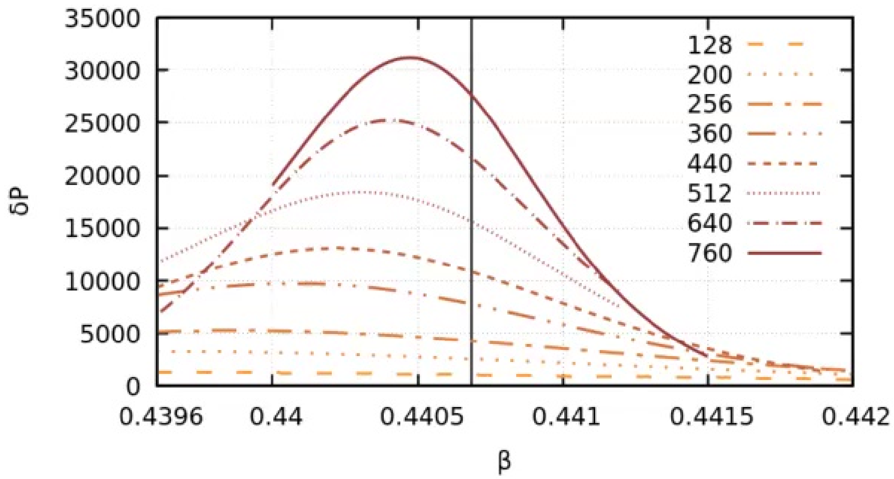
Inverse renormalisation group

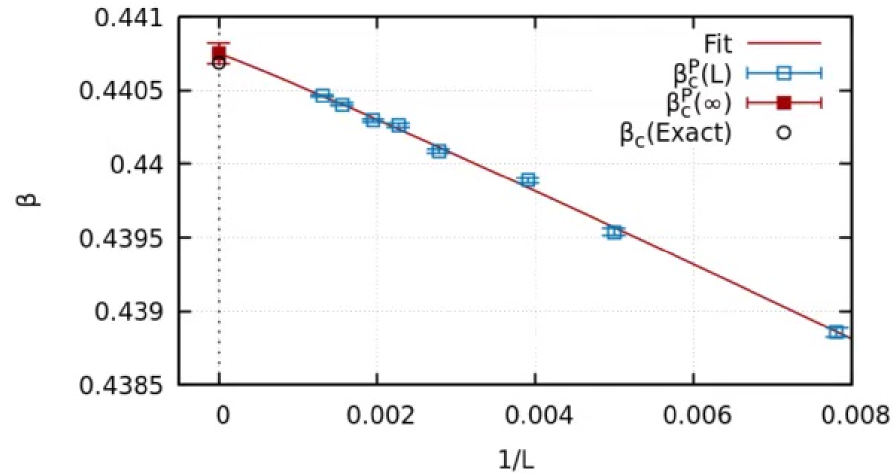


- local transformation
- apply inverse transformations iteratively
- evade critical slowing down
- generate flow in parameter space
- invariance at critical point

Critical behaviour from NN observables

- Determine L dependent susceptibility δP and its maximum at $\beta_c(L)$





Extract critical properties from NN observables only

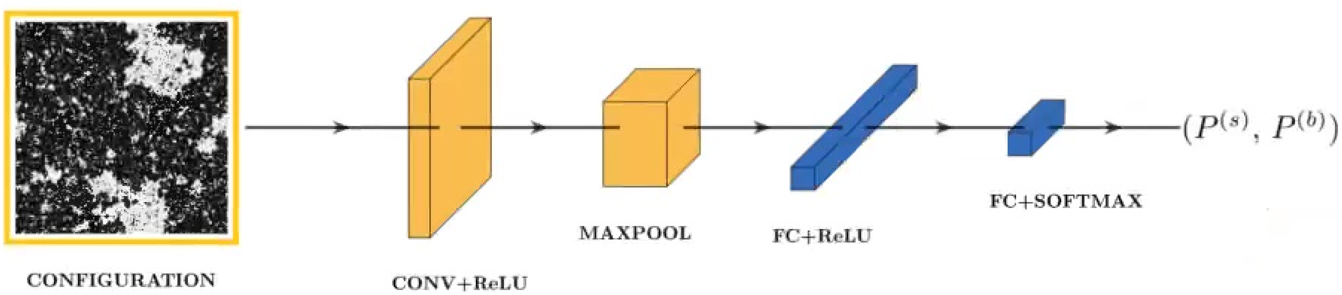
	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	7/4 =1.75

talkperimeter-mar2023.pdf
Page 9 of 47

✓Extending machine learning classification capabilities with histogram reweighting
Phys. Rev. E 102 (2020) 033303 [2004.14341 [cond-mat.stat-mech]]

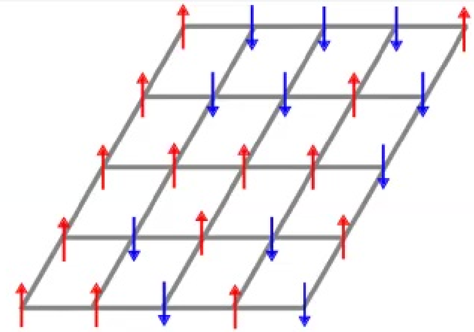
What's new?

- by now well-established procedure, what can we add?
- interpret output from a NN as an observable in a statistical system
- input: configurations, distributed according to Boltzmann weight
- output: observable, "order parameter" in statistical system



$$\Rightarrow \quad < P > = \frac{1}{Z}\sum_i P_i\, e^{-\beta E_i}$$

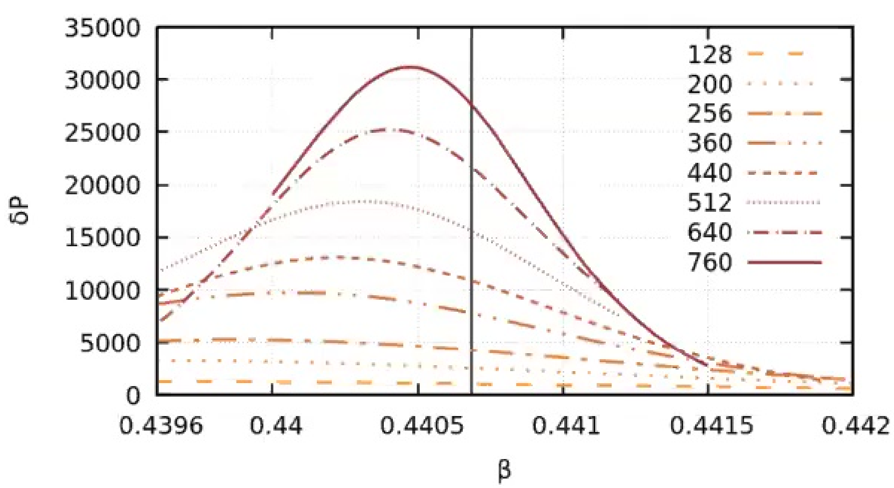
2d Ising model

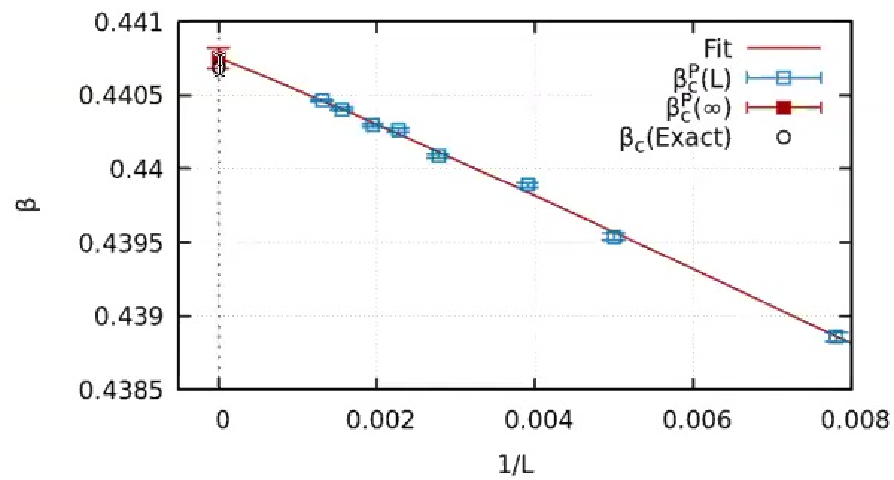


- $Z = \mathrm{Tr}\, e^{-\beta E}$ with $E = -\sum_{<ij>} s_i s_j$ $\quad (s_i = \pm 1)$
- critical coupling or inverse temperature β_c
- correlation length ξ, magnetic susceptibility χ diverge at transition
- critical exponents $\xi \sim |t|^{-\nu}$ $\quad \chi \sim |t|^{-\gamma}$ $\qquad$ reduced temperature $t = \dfrac{\beta_c - \beta}{\beta_c}$
- $\nu = 1,$ $\qquad \gamma/\nu = 7/4\,,$ $\qquad \beta_c = \frac{1}{2}\ln(1+\sqrt{2}) \approx 0.440687$
- finite-size scaling $\quad |t| = \left|\dfrac{\beta_c - \beta_c(L)}{\beta_c}\right| \sim \xi^{-\frac{1}{\nu}} \sim L^{-\frac{1}{\nu}}$ $\qquad \chi \sim L^{\gamma/\nu}$

Critical behaviour from NN observables

- Determine L dependent susceptibility δP and its maximum at $\beta_c(L)$

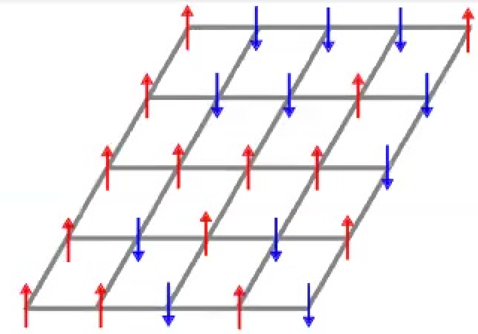




Extract critical properties from NN observables only

	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	7/4 =1.75

2d Ising model



- $Z = \mathrm{Tr}\, e^{-\beta E}$ with $E = -\sum_{<ij>} s_i s_j \quad (s_i = \pm 1)$
- critical coupling or inverse temperature β_c
- correlation length ξ, magnetic susceptibility χ diverge at transition
- critical exponents $\xi \sim |t|^{-\nu} \quad \chi \sim |t|^{-\gamma}$ reduced temperature $t = \dfrac{\beta_c - \beta}{\beta_c}$
- $\nu = 1, \quad \gamma/\nu = 7/4, \quad \beta_c = \frac{1}{2}\ln(1+\sqrt{2}) \approx 0.440687$
- finite-size scaling $|t| = \left|\dfrac{\beta_c - \beta_c(L)}{\beta_c}\right| \sim \xi^{-\frac{1}{\nu}} \sim L^{-\frac{1}{\nu}} \qquad \chi \sim L^{\gamma/\nu}$

Quantum field-theoretical machine learning

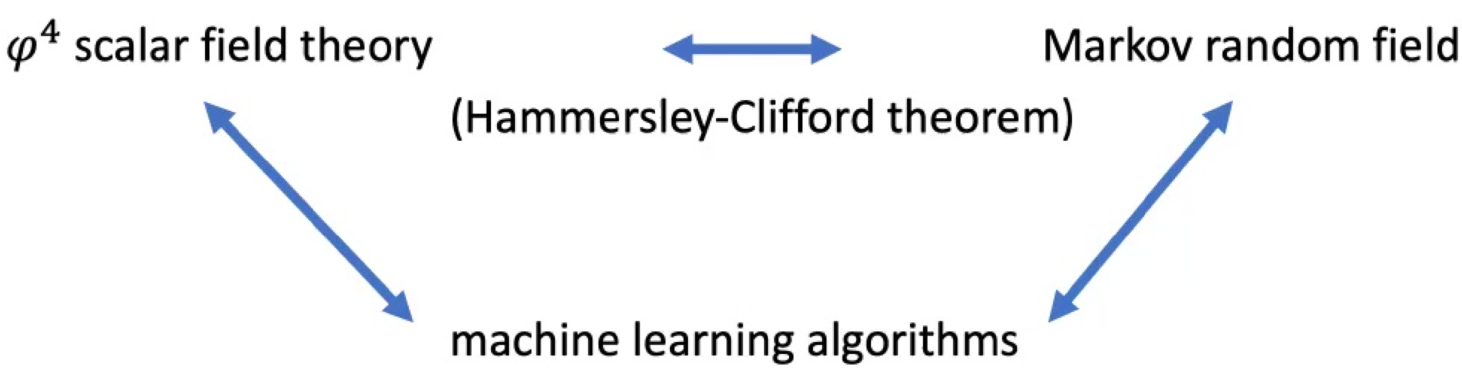
- design ML algorithms
- develop synergies with lattice field theories

✓ Quantum field-theoretic machine learning
Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]
D Bachtis, GA, B Lucini

✓ Chanju Park, BL and GA (in preparation)

Main idea

derive machine learning algorithms from discretized Euclidean field theories



Probability distribution p(φ)

probability distribution p(φ) defined as product of nonnegative functions over maximal cliques:

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi)$$

then p(φ) satisfies local Markov property

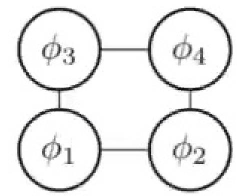
and set of random variables φ define a Markov random field

Theorem 1 (Hammersley-Clifford.) *A strictly positive distribution p satisfies the local Markov property of an undirected graph $\mathcal{G}$, if and only if p can be represented as a product of nonnegative potential functions ψ_c over $\mathcal{G}$, one per maximal clique $c \in C$, i.e.,*

$$p(\phi) = \frac{1}{Z} \prod_{c \in C} \psi_c(\phi), \tag{2}$$

where $Z = \int_{\boldsymbol{\phi}} \prod_{c \in C} \psi_c(\boldsymbol{\phi}) d\boldsymbol{\phi}$ is the partition function and $\boldsymbol{\phi}$ are all possible states of the system.

Partition function of local field theory



$$Z = \int d\varphi \, \exp(-S(\varphi)) \qquad p(\varphi) = \exp(-S(\varphi))/Z$$

- S depends on local (potential) and nearest neighbours (kinetic term)
- explicit example: 2d scalar field $\quad \mathcal{L}_E = \frac{\kappa}{2}(\nabla\phi)^2 + \frac{\mu_0^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4$
- discretise and introduce local couplings: $\quad S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4$
- probability distribution: $\quad p(\phi) = \frac{1}{Z}\prod_{c\in C}\psi_c(\phi)$
- satisfies Hammersley-Clifford theorem $\quad \psi_c = \exp\left[-w_{ij}\phi_i\phi_j + \frac{1}{4}(a_i\phi_i^2 + a_j\phi_j^2 + b_i\phi_i^4 + b_j\phi_j^4)\right]$

➡ discretized φ^4 scalar field is a Markov random field

Lattice φ^4 theory as Markov random field

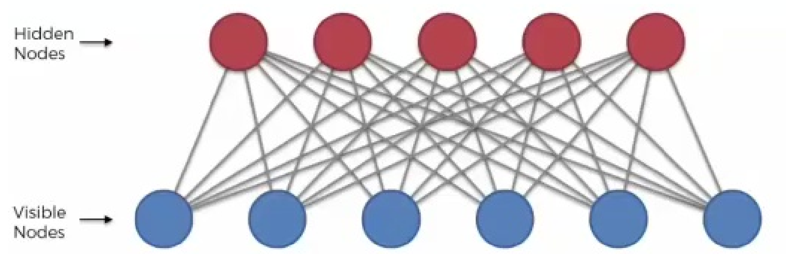
$$S(\phi;\theta) = -\sum_{\langle ij \rangle} w_{ij}\phi_i\phi_j + \sum_i a_i\phi_i^2 + \sum_i b_i\phi_i^4$$

- set of coupling constants/variational parameters: $\theta = \{ w_{ij}, a_i, b_i \}$
- search for an optimal set to complete ML tasks
- allowing them to be local/inhomogeneous increases expressivity
- from Euclidean QFT perspective: slightly strange theory
- but note: QFTs with random couplings or potentials

$$p(\phi;\theta) = \frac{\exp\left[-S(\phi;\theta) \right]}{\int_\phi \exp[-S(\phi,\theta)]d\phi}$$

From scratch: Restricted Boltzmann Machine

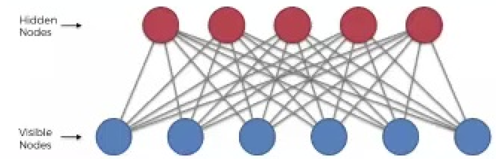
- Restricted Boltzmann Machine (RBM): two-layer generative network
- visible layer: to encode probability distribution
- hidden layer: to encode correlations
- restricted: no connections within a layer
- standard RBM: spin degrees of freedom on each node



$$v_i = \{0,1\}, \qquad i = 1,\dots,N_v \qquad\qquad h_a = \{0,1\}, \qquad a = 1,\dots,N_h$$

https://medium.datadriveninvestor.com/an-intuitive-introduction-of-restricted-boltzmann-machine-rbm-14f4382a0dbb

Restricted Boltzmann machine



- standard RBM: spin degrees of freedom on each node

$$v_i = \{0,1\}, \qquad i = 1,\dots,N_v \qquad\qquad h_a = \{0,1\}, \qquad a = 1,\dots,N_h$$

- energy function $E(v,h) = -\sum_{i,a} v_i w_{ia} h_a - \sum_i b_i v_i - \sum_a \eta_a h_a$

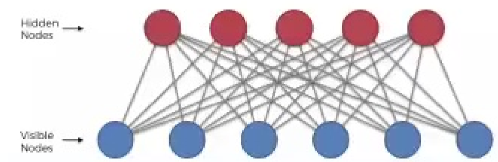
- Gibbs distribution $p(v,h) = \frac{1}{Z} e^{-E(v,h)}$ $\qquad Z = \sum_{\{v_i\},\{h_a\}} e^{-E(v,h)}$

- learn probability distribution $p(v) = \frac{1}{Z} \sum_{\{h_a\}} e^{-E(v,h)}$ from data provided to visible layer

- learn weight matrix and biases, “universal approximator”

M. Carreira-Perpinan and G. Hinton, On contrastive divergence learning, Artificial Intelligence and Statistics (2005)

Scalar field RBM



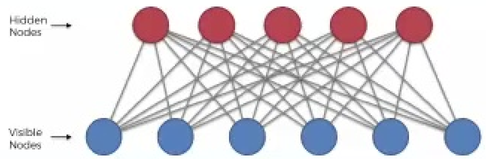
- use field theory approach: start with "free fields": Gaussian-Gaussian RBM
- add interactions later
- distribution: $p(\phi, h) = \frac{1}{Z} e^{-S(\phi,h)}$ $\quad Z = \int D\phi Dh\, e^{-S(\phi,h)}$
- energy (or action): $S(\phi, h) = S_\phi(\phi) + S_h(h) - \sum_{i,a} \phi_i w_{ia} h_a$
- no interaction between nodes in a layer, but potential terms possible, schematically

$$S(\phi, h) = \sum_{i=1}^{N_v} V_\phi(\phi_i) + \sum_{a=1}^{N_h} V_h(h_a) - \sum_{i,a} \phi_i w_{ia} h_a$$

- as stated: start with Gaussian fields on both layers

talkperimeter-mar2023.pdf
Page 32 of 47

Gaussian scalar field RBM



- distribution $p(\phi, h) = \frac{1}{Z} e^{-S(\phi,h)}$ $\qquad Z = \int D\phi Dh\, e^{-S(\phi,h)}$

- action $S(\phi, h) = \sum_i \frac{1}{2}\mu_i^2 \phi_i^2 + \sum_a \frac{1}{2\sigma^2}(h_a - \eta_a)^2 - \sum_{i,a} \phi_i w_{ia} h_a$

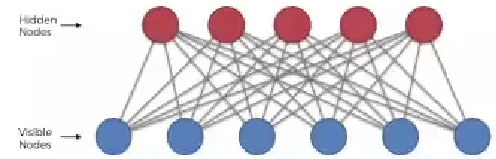
- induced distribution $p(\phi) = \int Dh\, p(\phi, h) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{i,j} \phi_i K_{ij} \phi_j + \sum_i J_i \phi_i \right)$

- kinetic (all-to-all) term $K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$ and source $J_i = \sum_a w_{ia} \eta_a$

talkperimeter-mar2023.pdf
Page 33 of 47

Search

Gaussian scalar field RBM



- kinetic (all-to-all) term $K_{ij} = \mu_i^2 \delta_{ij} - \sigma^2 \sum_a w_{ia} w_{aj}^T$ and source $J_i = \sum_a w_{ia} \eta_a$

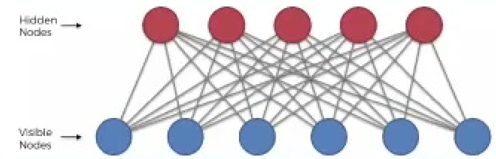
- use matrix notation $p(\phi) = \frac{1}{Z} \exp\left(-\frac{1}{2}\phi^T K \phi + J^T \phi\right)$ $K = \mu^2 \mathbb{1} - \sigma^2 W W^T$

  $J = W\eta$

- convergence: kinetic term should be positive definite

- first constraint on parameters: $\mu^2/\sigma^2 >$ largest eigenvalue of WW^T

Gaussian scalar field RBM



- apply now to concrete problem: free scalar field theory as target theory

- action: $S = \int dx \, \frac{1}{2} \left[(\partial\phi)^2 + m^2\phi^2 \right]$

- standard lattice kinetic term: nearest neighbour terms + diagonal (use 1d for simplicity)

$$K^{\phi}_{ij} = (m^2 + 2)\delta_{ij} - \delta_{i,j+1} - \delta_{i,j-1} \qquad K^{\phi}_{p} = m^2 + \hat{p}^2 \qquad \hat{p}^2 = 2 - 2\cos(p)$$

- directly compare the RBM kinetic term $K = \mu^2 \mathbb{1} - \sigma^2 W W^T$: $K \sim K^{\phi}$

- find explicit representations for the weight matrix W : $W W^T = \frac{1}{\sigma^2}\left(\mu^2 \mathbb{1} - K^{\phi}\right)$

talkperimeter-mar2023.pdf
Page 35 of 47

Gaussian scalar field RBM

$$WW^T = \frac{1}{\sigma^2}\left(\mu^2 \mathbb{1} - K^\phi\right) \equiv \mathcal{K}$$

- find explicit representations for the weight matrix W
- RHS is a $N_v \times N_v$ symmetric real matrix, W is $N_v \times N_h$ matrix

this leads to several conclusions

1. LHS is positive: $\mu^2 > \text{largest eigenvalue of } K^\phi$
2. $\det(\text{RHS}) > 0:$ $\quad \text{rank}(W) \geq N_v$ → $N_h \geq N_v$ result on number of nodes required(!)
3. for Gaussian models, $N_h = N_v$ is in fact necessary and sufficient
4. if $N_h < N_h$: $\det(\text{LHS}) = 0$ → $\text{eigenvalues } K_\phi : \mu^2$

talkperimeter-mar2023.pdf
Page 41 of 47

QFT-RBM

- add interactions as local potentials terms

$$S(\phi, h) = \sum_{i=1}^{N_v} V_\phi(\phi_i) + \sum_{a=1}^{N_h} V_h(h_a) - \sum_{i,a} \phi_i w_{ia} h_a$$

- weight matrix induces non-local interactions, as above (e.g. nearest neighbours)
- dimensionality is encoded here
- bias on hidden nodes induces symmetry breaking, as above $J_i = \sum_a w_{ia} \eta_a$
- potentials, possibly with local coupling parameters, to increase capacity,

  e.g. $V_\phi(\phi_i) = g_i^{(4)} \phi_i^4 + g_i^{(6)} \phi_i^6 + \ldots$ and the same for hidden nodes

to be explored

Scalar-Bernoulli RBM

- add interactions by replacing hidden fields by binary nodes, $h_a = \pm 1$

$$S(\phi, h) = S_\phi(\phi) + \sum_a \eta_a h_a - \sum_{i,a} \phi_i w_{ia} h_a$$

- induced distribution on visible layer: $p(\phi) = \prod_a \sum_{h_a = \pm 1} p(\phi, h)$

$$\begin{aligned} p(\phi) &= \frac{1}{Z} e^{-S_\phi(\phi)} \prod_a \left(e^{\eta_a - \sum_i \phi_i w_{ia}} + e^{-\eta_a + \sum_i \phi_i w_{ia}} \right) \\ &= \frac{1}{Z} \exp\left(-S_\phi(\phi) + \sum_a \ln\left[2\cosh(\psi_a)\right] \right) = \frac{1}{Z} \exp\left(-S_\phi(\phi) + \sum_a \sum_{n=1}^{\infty} c_n \psi_a^{2n} \right) \end{aligned}$$

with $\psi_a = \sum_i \phi_i w_{ia} - \eta_a$

talkperimeter-mar2023.pdf
Page 44 of 47

QFT or Bernoulli RBMs

very different probability distributions:

- local field theories with specific kinetic terms

vs

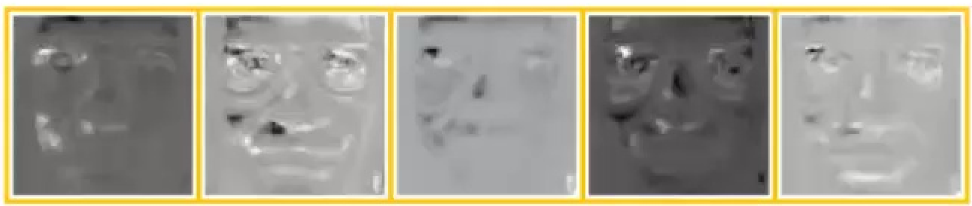
- all-to-all interactions of all powers of ϕ

which one is preferred? depends on application:

- Bernoulli-type RBMs common in ML
- for LFT purposes, local field formulations might be better

Example: Olivetti faces dataset with φ^4 RBM

- 64^2 visible units and 32 hidden units
- are there learned features? coupling constants w_{ij} for a fixed j



- neural network has learned hidden features: abstract face shapes and characteristics
- hidden units can serve as input to a new φ^4 neural network to progressively extract more abstract features in data

✓ Quantum field-theoretic machine learning
Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]

Summary

- ✓ learn ML from LFT and vice versa: formally connected as Markov random fields
- ✓ starting from scratch: Restricted Boltzmann Machine viewed from a LFT perspective
- ✓ found conditions on parameters and explicit representations for weights in “free” (Gaussian) case
- ✓ interactions can be local or highly non-local, depending on hidden nodes

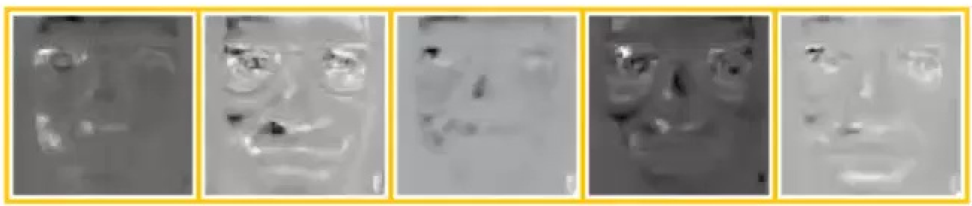
talkperimeter-mar2023.pdf
Page 47 of 47

Outlook

- ✓ inspiring connection between problems in lattice field theory and machine learning
- ✓ new solutions to old problems/old solutions to new problems
- ✓ new insights to both LFT and ML

Example: Olivetti faces dataset with φ^4 RBM

- 64^2 visible units and 32 hidden units
- are there learned features? coupling constants w_{ij} for a fixed j



- neural network has learned hidden features: abstract face shapes and characteristics
- hidden units can serve as input to a new φ^4 neural network to progressively extract more abstract features in data

✓ Quantum field-theoretic machine learning
Phys. Rev. D 103 (2021) 074510 [2102.09449 [hep-lat]]

Gaussian scalar field RBM

$$WW^T = \frac{1}{\sigma^2}\left(\mu^2 \mathbb{1} - K^\phi\right) \equiv \mathcal{K}$$

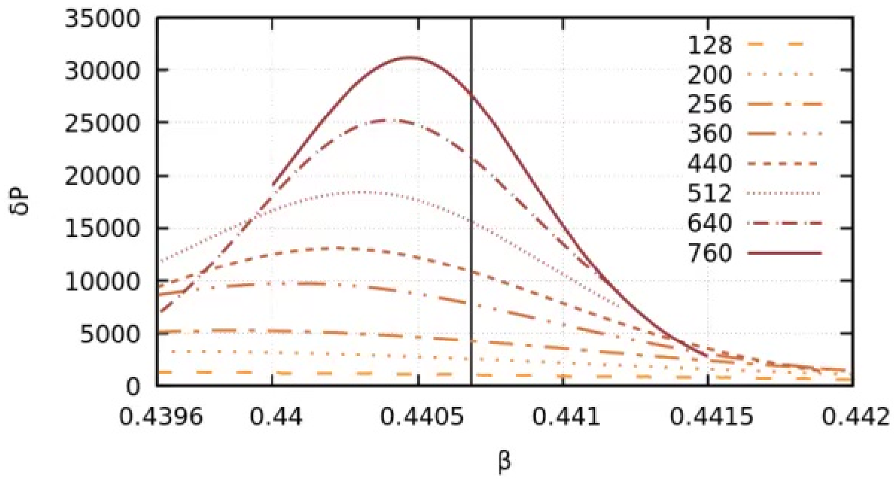
- find explicit representations for the weight matrix W
- RHS is a $N_v \times N_v$ symmetric real matrix, W is $N_v \times N_h$ matrix

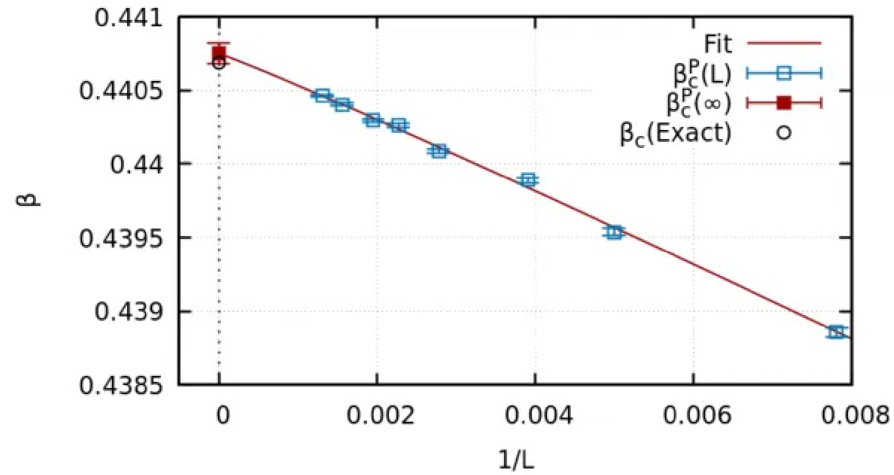
this leads to several conclusions

1. LHS is positive: $\mu^2 > \text{largest eigenvalue of } K^\phi$
2. $\det(\text{RHS}) > 0:$ $\text{rank}(W) \geq N_v$ → $N_h \geq N_v$ result on number of nodes required(!)
3. for Gaussian models, $N_h = N_v$ is in fact necessary and sufficient
4. if $N_h < N_h$: $\det(\text{LHS}) = 0$ → $\text{eigenvalues } K_\phi : \mu^2$

Critical behaviour from NN observables

- Determine L dependent susceptibility δP and its maximum at $\beta_c(L)$





Extract critical properties from NN observables only ➡

	β_c	ν	γ/ν
CNN+Reweighting	0.440749(68)	0.95(9)	1.78(4)
Exact	$\ln(1+\sqrt{2})/2$ ≈ 0.440687	1	7/4 $=1.75$