Constrained curve fitting with Bayesian neural networks

Curtis Taylor Peterson & Anna Hasenfratz

University of Colorado Boulder

curtis.peterson@colorado.edu

Introduction

Common to many analysis pipelines in lattice field theory is the need to fit data to a model that is determined only partially by a finite number of model parameters. Familiar examples include analyses of finite-size scaling (FSS) and ground state spectroscopy. Motivated by promising results in condensed matter physics [1], we conjecture that the expressivity of neural networks makes them good candidates for parameterizing the unknown component of such models. We test this conjecture by performing a curve collapse analysis of the 2nd-order finite-temperature phase transition of the 2D Ising model and the zero-temperature phase transition of the 4D massless $N_f/Nc = 8/3$ gauge-fermion system.

Network architecture & training

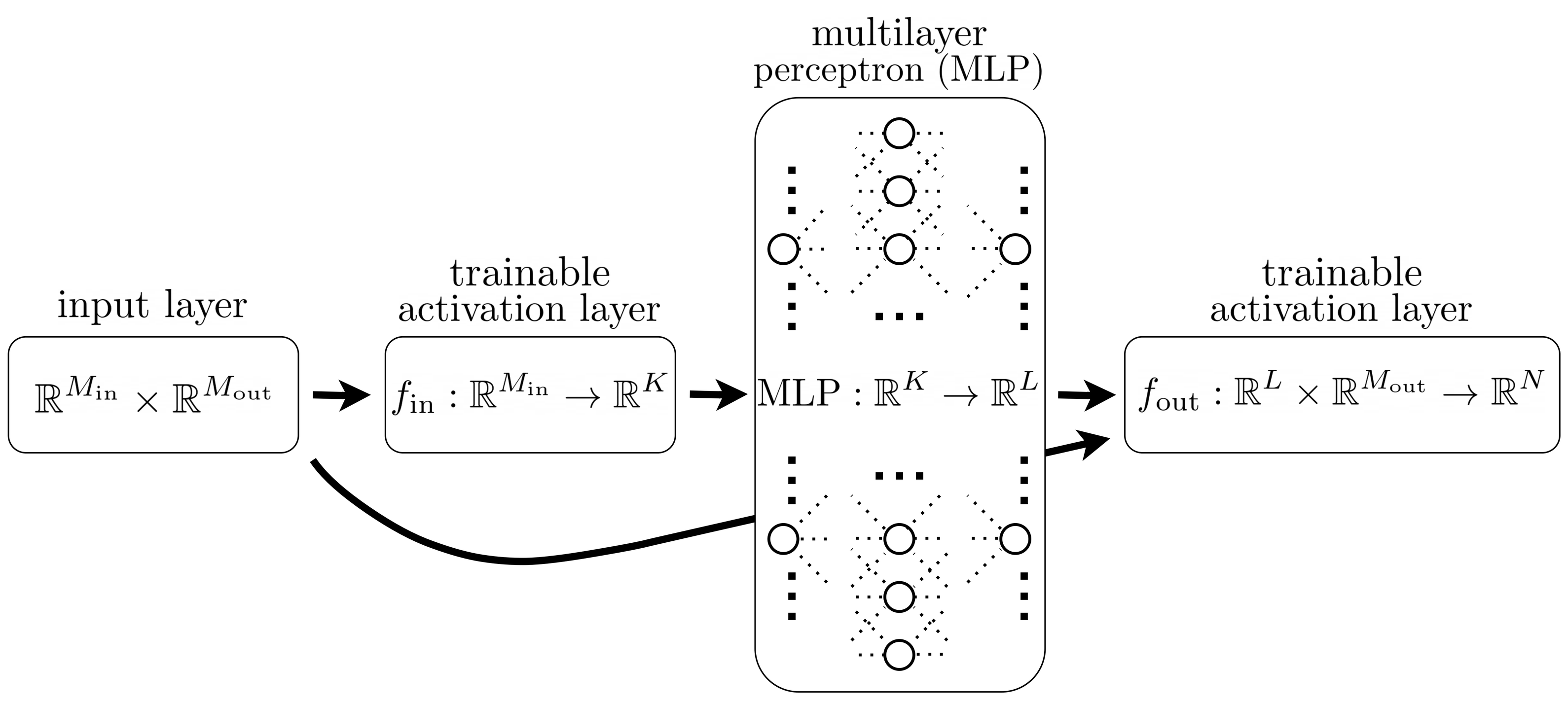

Figure 1: Neural network architecture.

Architecture

- Full model $\mathcal{M}_\Theta$ maps input features in $\mathbb{R}^{M_{\rm in}} \times \mathbb{R}^{M_{\rm out}}$ to output features in $\mathbb{R}^N$ given the model parameters $\{\Theta\} = \{\theta_{\rm in}, \theta_{\rm MLP}, \theta_{\rm out}\}$; see Fig. 1.
  - $\{\theta_{\rm MLP}\}$ denote learnable parameters of a multilayer perception (MLP) that parameterize the a priori unknown component of the model, such as the scaling function in an FSS analysis or the infinite sum of excited-state exponentials in an analysis of hadron spectroscopy.
  - $\{\theta_{\rm in}\}$ and $\{\theta_{\rm out}\}$ denote additional learnable parameters of the model, such as critical exponents or ground-state energies/amplitudes, that enter the model function through explicitly-defined activation functions that act upon a combination of input features of the full network and output features of the MLP.

Training

- Training is performed by least-squares minimization of the augmented χ^2 [2]

$$\chi^2_{\rm aug.} = \chi^2_{\rm data} + \chi^2_{\rm priors},$$

  where $\chi^2_{\rm data}$ and $\chi^2_{\rm priors}$ include information from the covariance of the data & priors, respectively.
  - Priors on the networks weights in $\{\theta_{\rm MLP}\}$ provide a lever arm for controlling overfitting; equivalent to L2 regularization in the ML literature
  - Appropriate priors can be estimated using the *empirical Bayes method*
- We optimize $\chi^2_{\rm aug.}$ using the "basin hopping" algorithm championed in Ref. [3]:
  1. Random perturbation of coordinates (Θ)
  2. Local minimization (trust region reflective)
  3. Metropolis accept/reject

Gradient flow coupling & FSS

The gradient flow coupling at smearing radius $\sqrt{8t} = cL$ defined by

$$g_c^2(\beta, L) \sim t^2 \langle E(t) \rangle \big|_{8t=(cL)^2}$$

can be utilized as a scaling variable for use in finite-size scaling [4].

- At a 1st/2nd-order phase transition, $g_c^2(\beta, L)$ is expected to scale as

$$g_c^2(\beta, L) \sim L^\eta \mathcal{G}_c\big(\tilde{t} L^{1/\nu}\big)$$

  with $\nu = 1/d$ signaling first order transition.
- At a Berezinskii–Kosterlitz–Thouless (BKT) transition the scaling is

$$g_c^2(\beta, L) \sim L^{\gamma/\nu} \mathcal{G}_c\big(L \exp(-\zeta \tilde{t}^{-\nu})\big),$$

  where $\tilde{t} \equiv \beta/\beta_c - 1$.

Results

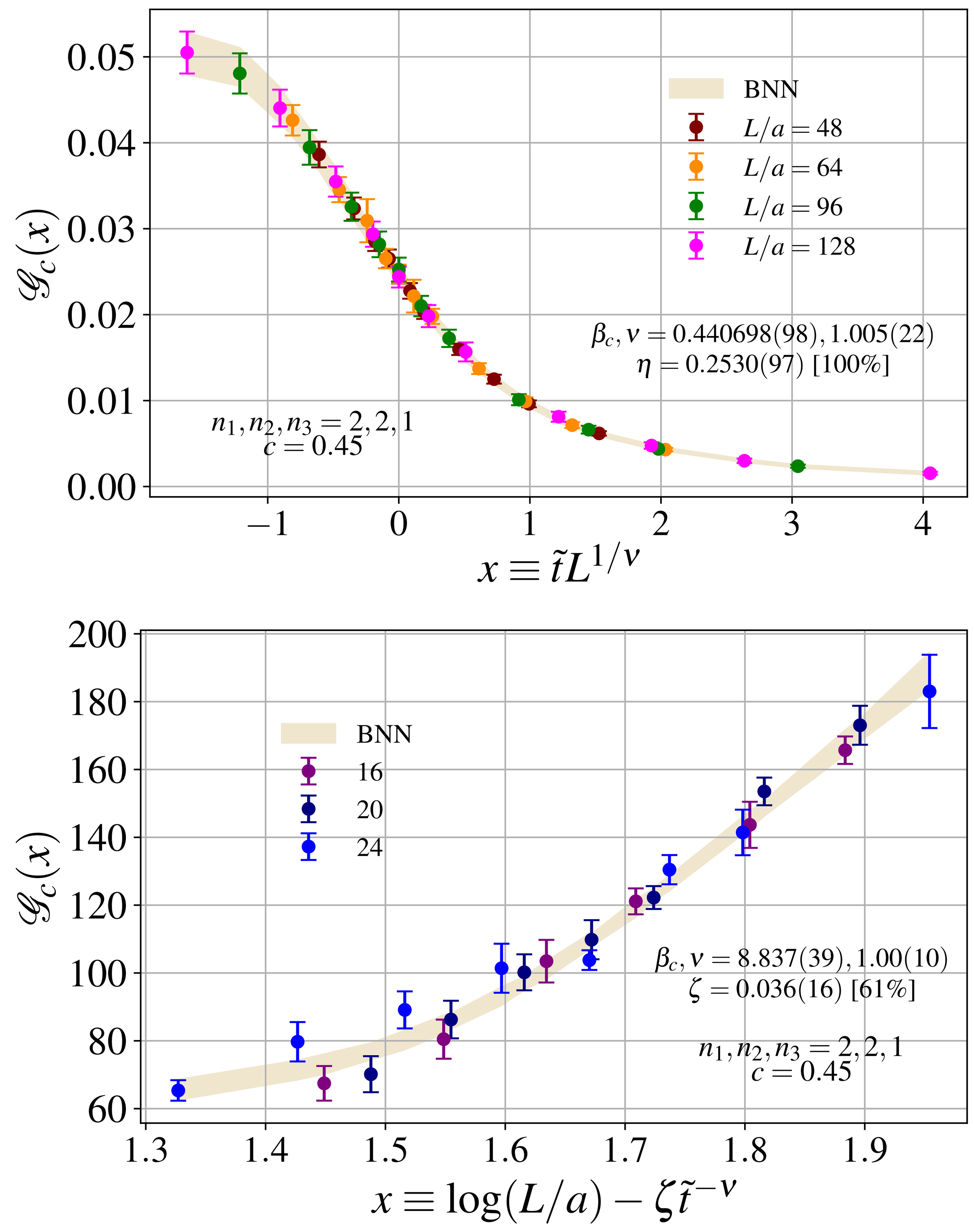

Figure 2: (Top panel) Curve collapse for the 2D Ising model using a 3-layer network with $n_1 = n_2 = 2$ nodes in the first and second layers and $n_3 = 1$ nodes in the third (output) layer. (Bottom panel) Curve collapse for the 4D massless $N_f/N_c = 8/3$ gauge-fermion system with $n_1, n_2, n_3 = 2, 2, 1$ nodes in each layer.

Conclusions & forthcoming research

- Our FSS results are consistent with the published literature
  - Prediction for Ising critical exponents consistent with exact prediction: $\beta_c \approx 0.4407$ and $\nu, \eta = 1, 1/4$.
  - $N_f = 8$ test rules out 1st-order scaling and appears to have a very slight preference for BKT.
- Even a relatively small neural network is sufficiently expressive to accurately represent the scaling function for the systems that we have tested
  - Bayesian priors provide level arm to control overfitting
- We are currently testing this method on the XY model
- We will begin testing the application of this method to ground state spectroscopy

References

[1] Ryosuke Yoneda and Kenji Harada. Neural network approach to scaling analysis of critical phenomena. *Phys. Rev. E*, 107(4):044128, 2023.

[2] G. P. Lepage, B. Clark, C. T. H. Davies, K. Hornbostel, P. B. Mackenzie, C. Morningstar, and H. Trottier. Constrained curve fitting. *Nucl. Phys. B Proc. Suppl.*, 106:12–20, 2002.

[3] David J Wales and Jonathan PK Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[4] Anna Hasenfratz. Emergent strongly coupled ultraviolet fixed point in four dimensions with eight Kähler-Dirac fermions. *Phys. Rev. D*, 106(1):014513, 2022.

[5] Mary Ann Branch, Thomas F Coleman, and Yuying Li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing*, 21(1):1–23, 1999.

Acknowledgements

A.H. acknowledges support by DOE Grant No. DE-SC0010005. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2040434. Computations for this work were carried out in part on facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy, the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards No. ACI-1532235 and No. ACI-1532236), the University of Colorado Boulder, and Colorado State University.