Normalizing flows and control variates for sign problems

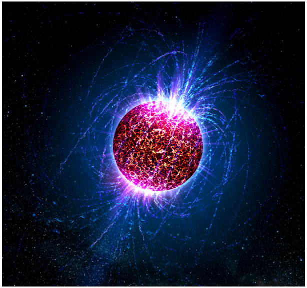
Yukari Yamauchi

Based on: arXiv:2205.12303 with **Scott Lawrence** and **Hyunwoo Oh**
arXiv:2311.13002 with **Scott Lawrence**
arXiv:2312.12636 with **Scott Lawrence**
arXiv:2401.16733 with **Scott Lawrence**, **Semeon Valgushev**, **Jianan Xiao**

December 10th, 2024, CERN

INSTITUTE for NUCLEAR THEORY

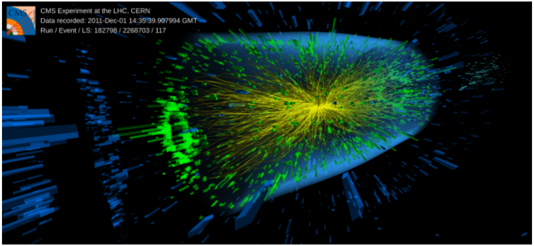
Sign problems in lattice QCD



Casey Reed/Penn State University

Finite density

Real time



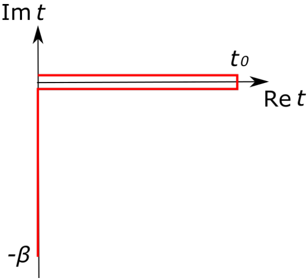
Tom McCauley/CMS/CERN

What is a sign problem?

Let's put both sign problems in once computation...

Schwinger-Keldysh formalism

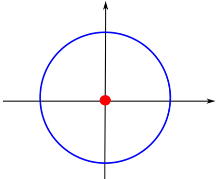
$$\begin{aligned}\langle \mathcal{O}(t_0)\rangle &= \mathrm{Tr}\left[e^{-\beta H}\, e^{iHt_0}\, \mathcal{O}\, e^{-iHt_0}\right] \\ &= \frac{1}{Z}\int \mathcal{D}[\psi, U]\, e^{-S_{\rm SK}}\, \mathcal{O}(t_0) \\ &\approx \frac{1}{N_s}\sum_{i=1}^{N_s} \mathcal{O}(U_i)\end{aligned}$$

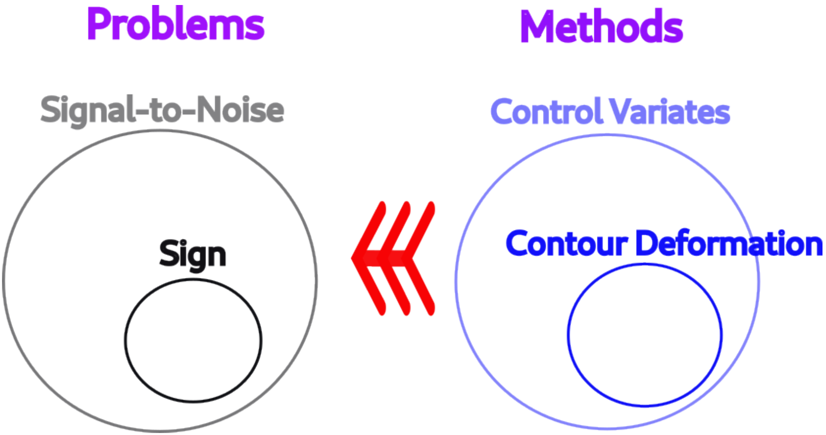


The "distribution" is **NOT** real & positive $\rightarrow$ **Sign problems**

The average phase

$$\langle\sigma\rangle = \frac{\int \mathcal{D}[\psi, U]\, e^{-S_{\rm SK}}}{\int \mathcal{D}[\psi, U]\, |e^{-S_{\rm SK}}|} \propto e^{-V}$$




Problems
Methods
Signal-to-Noise
Control Variates
Sign
Contour Deformation

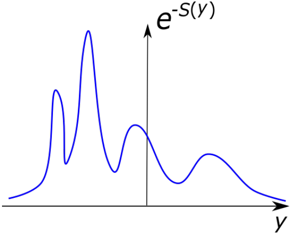
Contour deformation / Complex normalizing flow

Normalizing flows — when no sign problem

Normalizing flow[1] $\vec{y} = f(\vec{x})$:

$$\det\left(\frac{\partial \vec{y}}{\partial \vec{x}}\right) e^{-S(\vec{y})} = \mathcal{N} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \mathcal{N}\, G_N(\vec{x})$$

Some distribution

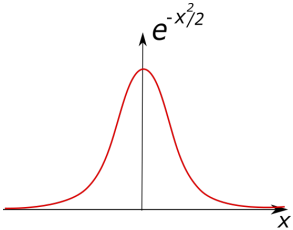


Map

$\leftrightarrow$

$y(x)$

Simple Gaussian distribution



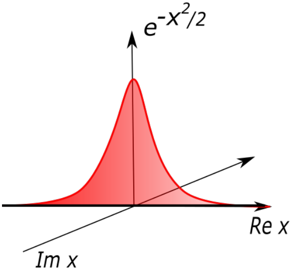
if such a map exists, then

$$\langle \mathcal{O} \rangle = \frac{\int_{\mathbb{R}^N} \mathcal{D}\vec{y}\ e^{-S(\vec{y})} \mathcal{O}(\vec{y})}{\int_{\mathbb{R}^N} \mathcal{D}\vec{y}\ e^{-S(\vec{y})}} = \frac{\int_{\mathbb{R}^N} d\vec{x}\ G_N(\vec{x}) \mathcal{O}(\vec{y}(\vec{x}))}{\int_{\mathbb{R}^N} d\vec{x}\ G_N(\vec{x})}$$

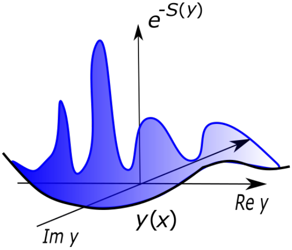
[1] Martin Lüscher, Commun.Math.Phys.293(2010) 899-919
M.Albergo et al. Phys. Rev. D 100, 034515(2019)
K. A. Nicoli, et al. Phys. Rev. E 101, 023304(2020)

Complex normalizing flow[2]/contour deformations

$$\mathcal{N} G_N(\vec{x}) = \det\left(\frac{\partial \vec{y}}{\partial \vec{x}}\right) \; e^{-S(\vec{y}(\vec{x}))}$$



Map
$\leftrightarrow$
$y(x)$



Expectation values:

$$\frac{\int_{\mathbb{R}^N} d\vec{x} \; G_N(\vec{x}) \mathcal{O}(y(x))}{\int_{\mathbb{R}^N} d\vec{x} \; G_N(\vec{x})} = \frac{\int_{\vec{y}(\mathbb{R}^N)} d\vec{y} \; e^{-S(\vec{y})} \mathcal{O}(\vec{y})}{\int_{\vec{y}(\mathbb{R}^N)} d\vec{y} \; e^{-S(\vec{y})}} \stackrel{?}{=} \langle \mathcal{O} \rangle$$

Perfect complex normalizing flow $\rightarrow$ No sign problems!

[2] S. Lawrence and YY, arXiv:2101.05755

Constraints on manifolds[3]

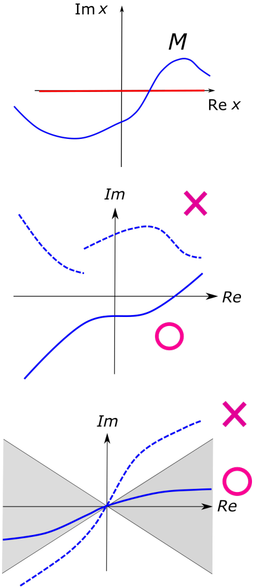
Manifolds give the correct $\langle \mathcal{O} \rangle$

$$\langle \mathcal{O} \rangle = \frac{\int_{\mathbb{R}} dy\ e^{-S(y)} \mathcal{O}(y)}{\int_{\mathbb{R}} dy\ e^{-S(y)}} = \frac{\int_{\mathcal{M}} dz\ e^{-S(z)} \mathcal{O}(z)}{\int_{\mathcal{M}} dz\ e^{-S(z)}}$$

when:

- The manifold (—) is a continuous manifold
- The manifold (—) is in "asymptotically safe" region
- Both e^{-S} and $e^{-S}\mathcal{O}$ are holomorphic functions in the region between (—) and (—)

→ **Cauchy's integral theorem guarantees**



[3] A. Alexandru et al., Phys. Rev. D. 98, 034506(2018)

Machine-learn a map

Contour deformation

Loss function: $L = -\log\langle\sigma\rangle$

The gradient of $\langle\sigma\rangle$ is **sign-free**[4]!

$$\partial_v(-\log\langle\sigma\rangle) = -\frac{\int_{\mathcal{C}} \mathcal{D}\phi\,(\partial_v \operatorname{Re} S)\left|e^{-S}\right|}{\int_{\mathcal{C}} \mathcal{D}\phi\,\left|e^{-S}\right|}$$

- Needs MCMC sampling
- Can find "good enough" contours

Complex normalizing flow

Loss function:

$$L = \left\langle \left| \det\left(\frac{\partial y(x_i)}{\partial x_i}\right) e^{-S(y(x_i))} - \mathcal{N} G_N(x_i) \right| \right\rangle_{G_N}$$

- No MCMC sampling
- Find only very good contours

[4] A. Alexandru, P. Bedaque, H. Lamm, and S. Lawrence, arXiv:1804.00697

Normalizing flows for complex coupling model[5]

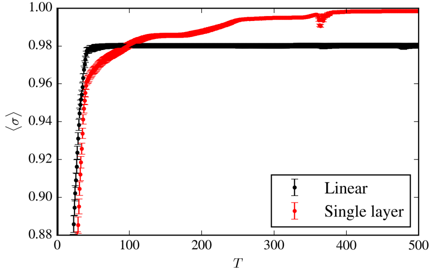
$$S = \sum_{i=1}^{N} \frac{m}{2}\phi_i^2 + \frac{(\phi_i - \phi_{i-1})^2}{2} + \lambda\phi_i^4, \text{ with } \lambda \in \mathbb{C}$$

Neural network $\vec{\phi}(\vec{x}) = f(\vec{x}) + ig(\vec{x})$, σ =sigmoid

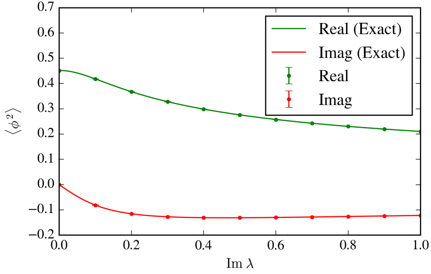
$$f, g(\vec{x}) = \left(L_{N,N} + L_{N,2N} \otimes \sigma \otimes L_{2N,2N} \cdots L_{2N,2N} \otimes \sigma \otimes L_{2N,N}\right)\vec{x}$$

Demonstration: $N = 10, m = 0.5$

$\lambda = 0.1 + i$



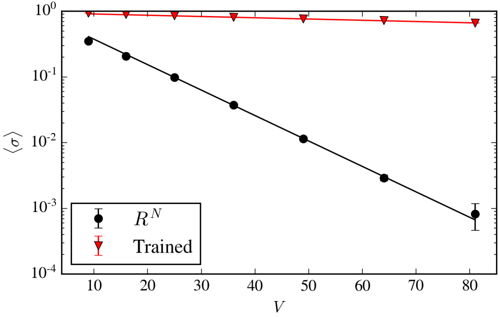
$\text{Re}\,\lambda = 0.1$



[5] S. Lawrence, H. Oh, and YY, arXiv:2205.12303

Contour deformation for complex coupling model

The average sign over the lattice size V



- $m = 0.5, \lambda = i$
- linear contour $\tilde{\phi}(\phi) = \phi + M_R\phi + iM_I\phi$
- square lattice

Yukawa scattering in one spatial dimension[6]

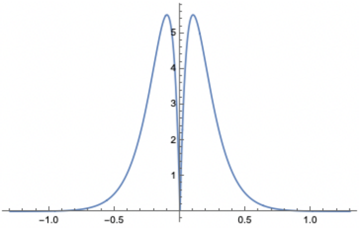
Quantum mechanical path integral

$$\langle\psi(0)|e^{iHT}\ \mathcal{O}\ e^{-iHT}|\psi(0)\rangle$$

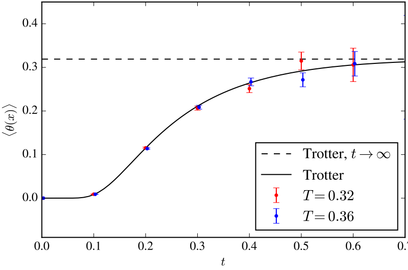
with Yukawa potential

$$V(x) = g\ |x|\ e^{-m|x|}$$

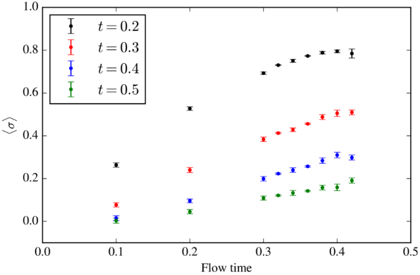
with $m = 10$ and $g = 150$.



Transmission coefficient



Average phase



[6] S. Lawrence, S. Valgushev, J. Xiao, YY, arXiv:2401.16733

Question

Do perfect contours exist for any lattice field theories?

or equivalently

Do complex normalizing flows exist for any lattice field theories?

NO!!

Bounds on the average sign[7]

$$\langle \sigma \rangle = \frac{Z}{Z_Q}$$

An obvious bound is, with $\omega = dz\, e^{-S(z)}$

$$Z_Q \;\geq |Z| \quad \text{because} \quad \int_\gamma |\omega| \;\geq \; \left| \int_\gamma \omega \right| \;, \quad \text{then} \quad \langle \sigma \rangle \leq 1$$

A better bound can be constructed by a new form α

$$|\alpha| < |\omega| \text{ anywhere}$$

Then, for any contours γ

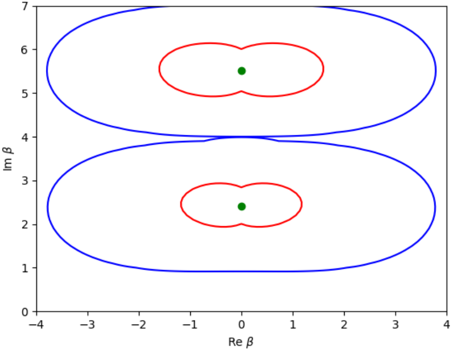
$$Z_Q = \int_\gamma |\omega| \;\geq\; \int_\gamma |\alpha| \;\geq\; \left| \int_\gamma \alpha \right|$$

[7] S. Lawrence and YY, arXiv:2311.13002

(Non)existence in $U(1)$ gauge theory in $1+1$-d

Path integral after gauge tranformation[8]

$$Z = \left[\int_0^{2\pi} d\theta e^{-\beta\cos(\theta)}\right]^N$$



Perfect contours **DO NOT** exist **inside** the red contour
Perfect contours **DO** exist **outside** the blue contour

[8] A.A.Migdal, Sov.Phys.JETP, 42, 413 (1975)

Complex control variants

Control variates for variance reduction

For an observable $\mathcal{O}$ with small signal-to-noide ratio

Subtract a function f from $\mathcal{O}$!!

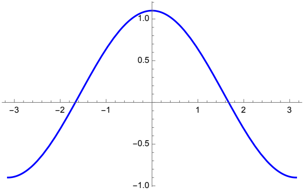
without changing physics, i.e.,

$$\langle \mathcal{O} \rangle = \langle \mathcal{O} - f \rangle \ \text{ but } \ \mathrm{Var}(\mathcal{O}) > \mathrm{Var}(\mathcal{O} - f)$$

So we strictly impose

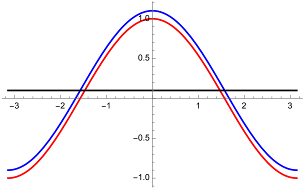
$$\langle f \rangle = \int \mathcal{D}[\phi]\, e^{-S(\phi)}\, f(\phi) = 0$$

$\mathcal{O} = \cos\theta + \epsilon$



$\rightarrow$

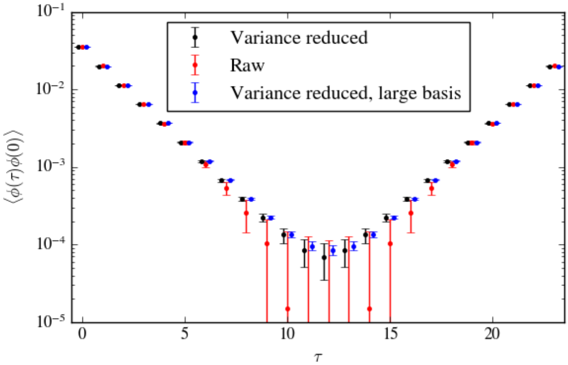
$\mathcal{O} - f = \epsilon$



Variance reduction for signal-to-noise problem[9]

Lattice scalar ϕ^4 theory in Euclidean

$$S = \sum_{\langle r,r'\rangle} \frac{(\phi(r)-\phi(r'))^2}{2} + \sum_r \left[\frac{m^2}{2}\phi^2(r) + \frac{\lambda}{24}\phi^4(r)\right]$$



24×24 lattice, $m^2 = 0.0, \lambda = 2.0$

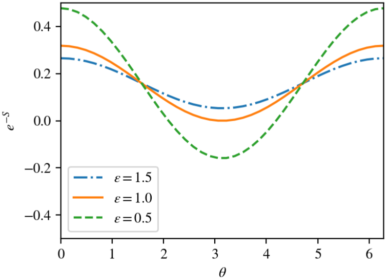
[9] T. Bhattacharya, S. Lawrence, and J. Yoo, arXiv:2307.14950

Perfect control variates always exist!

Example:

$$e^{-S(\theta;\epsilon)} = \cos(\theta) + \epsilon, \quad \theta \in [0, 2\pi)$$

What is the perfect control variates?



More generally, for any e^{-S}

$$f(x) = e^{-S(x)} - \frac{\int \mathcal{D}x \; e^{-S(x)}}{\int \mathcal{D}x \; 1}$$

(Perfect control variates are not unique)

Notes on control variates

Strength of control variates

- Include all contour deformation methods
- No Jacobian
- Can be applied to discrete field space

How do we find good control variates?

1. Analytical approaches
   - Perturbative constructions
   - Via Dyson-Schwinger relations
2. Numerical approaches
   - Start with ansatz and optimize
   - Machine learning

Demonstration: classical Ising model (Lee-Yang zeros[10])

Classical Ising model: $S(\vec{\sigma}) = -J\sum_{\langle i,j\rangle}\sigma_i\sigma_j - h\sum_i \sigma_i$

Goal: Compute $Z = \sum_\sigma e^{-S}$ at **purely imaginary magnetic field**

Measure

$$\frac{Z(h)}{Z(h{=}0)} = \frac{\sum_\sigma \exp\Big(J\sum_{\langle i,j\rangle}\sigma_i\sigma_j\Big)\,\exp\big(h\sum_i\sigma_i\big)}{\sum_\sigma \exp\Big(J\sum_{\langle i,j\rangle}\sigma_i\sigma_j\Big)} = \big\langle e^{h\sum_i \sigma_i}\big\rangle_Q$$

By replacing

$$e^{h\sum_i s_i} \to e^{h\sum_i s_i} - \textbf{CV}$$

and optimizing **CV** to minimize

$$\mathrm{Var}\left(e^{h\sum_i \sigma_i} - \textbf{CV}\right)$$

[10] C.N.Yang and T.D.Lee, Phys.Rev 87 404 (1952)

Extreme learning machine

1. Prepare basis functions (3V+2 of them)

$$\left\{\sigma_i, \sigma_i \cos(\mathrm{Im}\, S), \sigma_i \sin(\mathrm{Im}\, S), \cos(\mathrm{Im}\, S), \sin(\mathrm{Im}\, S)\right\} \times e^{-S(h=0)}$$

2. Input basis functions to ELM

3. Take "divergence"

$$F_i = f(\sigma_i) - f(-\sigma_i)$$

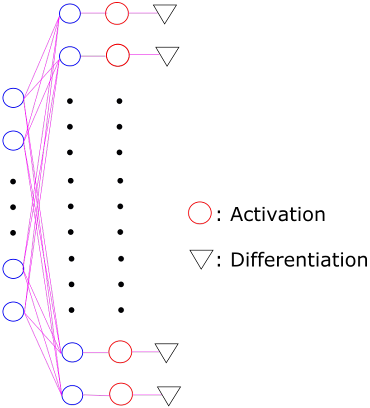
4. The $CV = \sum_i c_i F_i$

The coefficients c_i are optimized by estimating

$$M_{ij} = \langle F_i F_j \rangle, v_j = \langle \mathcal{O} F_i \rangle$$

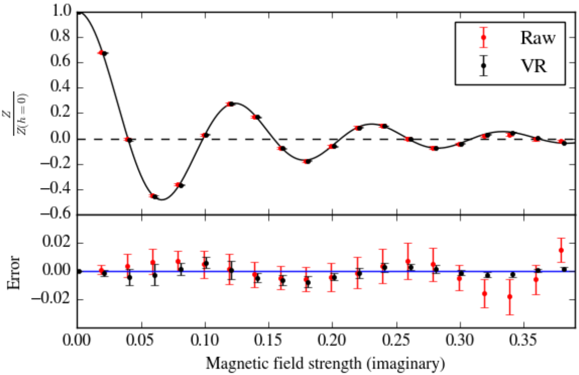
and

$$c = M^{-1}\, v$$



Classical Ising model[11]

At purely imaginary h, $J = 0.4 < J_c \approx 0.441$, 8×8 lattice:

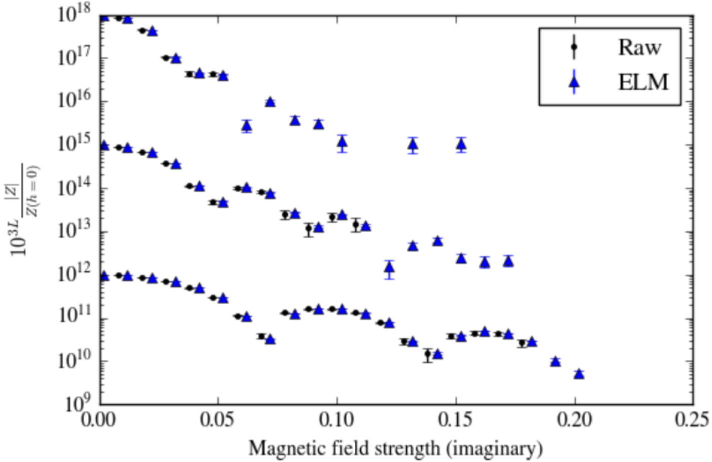


- Raw: 5k samples for Z
- VR: 5k samples to optimize, 5k samples for Z

[11] S. Lawrence and YY, arXiv:2312.12636

On cubic lattices of size $4^3, 5^3, 6^3$

At purely imaginary h, $J = 0.2 < J_c \approx 0.22$



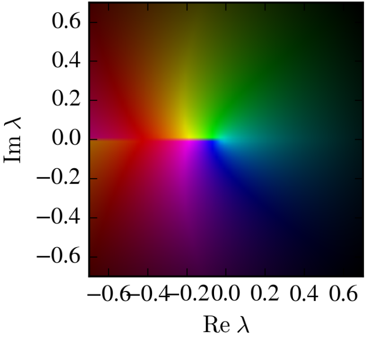
Future focus

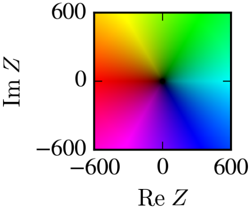
- Light nuclei reactions with contour deformations
- Real-time quantum spins with control variates
- Signal-to-Noise problems

Thank you!

Partition function in $0 + 1$-dimensions

$$S = \sum_{i=1}^{N} \frac{m}{2}\phi_i^2 + \frac{(\phi_i - \phi_{i-1})^2}{2} + \lambda\phi_i^4, \text{ with } \lambda \in \mathbb{C}$$





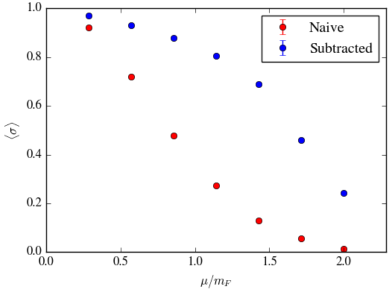
- $N = 10, m = 0.5$
- 1 internal layer
- Adiabatic training

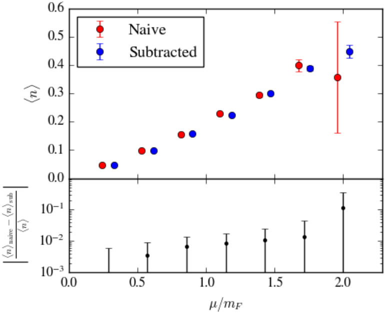
Thirring model in $1 + 1$-dimension[12]

$$S = \sum_{x,\nu} \frac{2}{g^2} (1 - \cos A_\nu(x)) - \log\det K,\ A_\nu \in [0, 2\pi)$$

with the Dirac matrix ($\eta_0 = (-1)^{\delta_{0,x_0}}$ and $\eta_1 = (-1)^{x_0}$)

$$K[A]_{xy} = m\delta_{xy} + \frac{1}{2}\sum_{\nu=0,1} \eta_\nu e^{iA_\nu(x)+\mu\delta_{\nu,0}}\delta_{x+\nu,y} - \eta_\nu e^{-iA_\nu(y)-\mu\delta_{\nu,0}}\delta_{y+\nu,x}$$

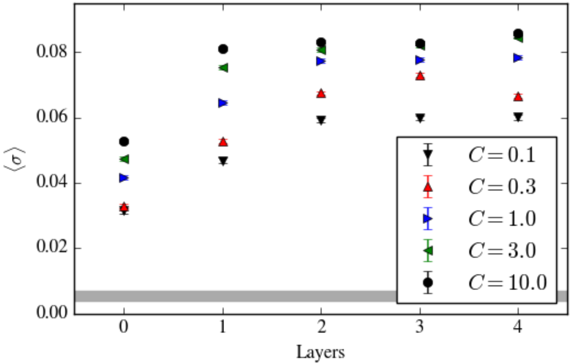




- 4×4 lattice, $m = 0.05, g = 1.0 \rightarrow m_B = 0.33(1), m_F = 0.35(2)$
- MLP with 2 inner layers

[12] S. Lawrence and YY, arXiv:2212.14606

Larger networks give better vector fields



- 6×6 lattice
- $m = 0.05, g = 1.0, \mu = 0.5$

Measurement of observables

Idea 1. No subtraction in the numerator

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{D}x\ e^{-S(x)} \mathcal{O}}{\int \mathcal{D}x\ e^{-S(x)}} = \frac{\int \mathcal{D}x\ (e^{-S(x)} - f(x)) \frac{e^{-S(x)}}{e^{-S(x)} - f(x)} \mathcal{O}}{\int \mathcal{D}x\ e^{-S(x)} - f(x)}$$

(Phase fluctuation moved from denominator to numerator.)

Idea 2. Subtract $\nabla \cdot (\mathcal{O}\vec{v})$ anyway

$$\langle \mathcal{O} \rangle = \frac{\int \mathcal{D}x\ e^{-S(x)} \mathcal{O} - \nabla \cdot (\mathcal{O}\vec{v})}{\int \mathcal{D}x\ e^{-S(x)} - \nabla \cdot \vec{v}}$$

$$= \frac{\int \mathcal{D}x\ (e^{-S(x)} - \nabla \cdot \vec{v}) \left(\mathcal{O} + \frac{\vec{v} \cdot \nabla \mathcal{O}}{e^{-S(x)} - \nabla \cdot \vec{v}} \right)}{\int \mathcal{D}x\ e^{-S(x)} - \nabla \cdot \vec{v}}$$

Hoping that the "extra term" won't cause signal-noise problem.

This seems to work for the density operator.... (why?)