Complex Control Variates

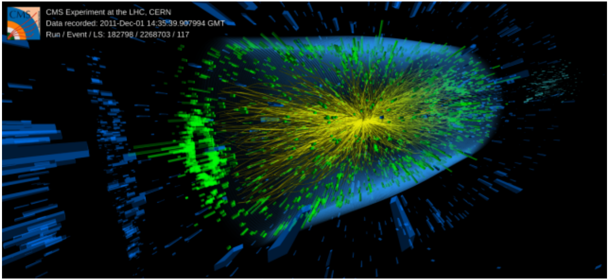
Yukari Yamauchi

Based on arXiv:2212.14606 [hep-lat] with Scott Lawrence

August 1st, 2023, Lattice 2023

INSTITUTE for NUCLEAR THEORY

Sign problems in lattice QCD



Tom McCauley/CMS/CERN

Path integral

$$\langle \mathcal{O}(t)\mathcal{O}(0)\rangle_\beta = \mathrm{Tr}\left[e^{-\beta H}\, e^{iHt}\, \mathcal{O}\, e^{-iHt}\mathcal{O}\right] = \frac{1}{Z}\int \mathcal{D}[\psi, U]\, e^{-S_{\mathrm{SK}}}\, \mathcal{O}(t)\mathcal{O}(0)$$

The average phase

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

Complex control variates

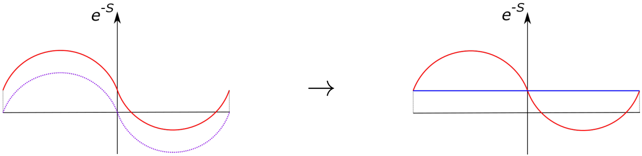
The idea is very simple...

Subtract a function (fluctuation) from e^{-S}!!

without changing physics

$$Z = \int \mathcal{D}x \, \left(e^{-S(x)} - f(x)\right) \text{with} \int \mathcal{D}x \; f(x) = 0$$

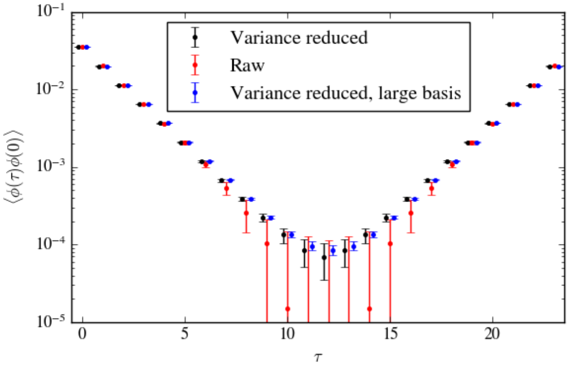
so sign problem is reduced



Variance reduction for signal-to-noise problem[1]

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, $m^2 = 0.0, \lambda = 2.0$

[1] T. Bhattacharya, S. Lawrence, and J. Yoo, arXiv:2307.14950 [hep-lat]

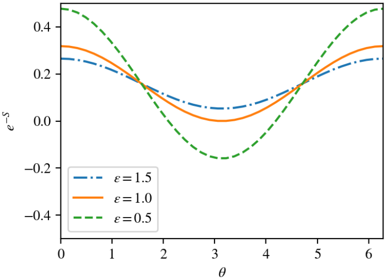
Existence of control variates

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

Other 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 (perturbative) approaches
   - S. Lawrence, arXiv:2009.10901[hep-lat]
   - S. Lawrence and **YY**, arXiv:2212.14606 [hep-lat]

2. Numerical approaches
   - Start with ansatz and optimize
   - Machine learning

Demonstration: Classical Ising model, Thirring model in $1+1$-d

Demonstration: Classical Ising model

Classical Ising model: $S(\vec{s}) = -J\sum_{\langle i,j\rangle} s_i s_j - h\sum_i s_i$

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

Measure

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

By replacing

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

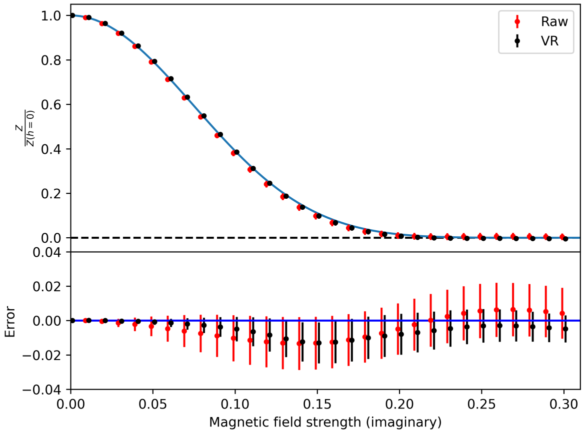
and optimize **CV** to minimize $\mathrm{Var}\left(e^{h\sum_i s_i} - \textbf{CV}\right)$

Constructing **CV** $\quad (\nabla_i f(\vec{s}) = f(s_i) - f(-s_i))$

- $\nabla_i\ e^{-S(h=0)}$
- $\nabla_i\ e^{-S(h)}$
- $\nabla_i\left(S(h=0)\ e^{-S(h)}\right)$
- $\nabla_i\left(\ s_j e^{-S(h=0)}\right)$
- $\nabla_i\left(\ s_i s_j e^{-S(h=0)}\right)$ (**Important!**)

Classical Ising model (preliminary)

At purely imaginary h, $J = 0.2$, 8×8 lattice:



- Raw: 2k samples for Z
- VR: 2k samples to optimize, 2k samples for Z

Numerical optimization of complex control variates

Prepare a family of functions $f_{\rm cv}(x)$ that integate out to zero by

$$f_{\rm cv}(x) = \nabla \cdot \vec{v}(x) \quad \because \int \mathcal{D}x \, \nabla \cdot \vec{v}(x) = 0$$

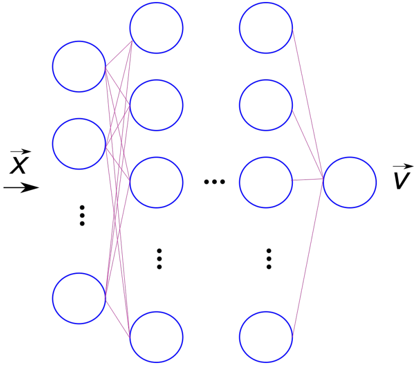
represent $\vec{v}$ via neural network

and minimize

$$C(\vec{v}) = -\log \langle \sigma \rangle_{\vec{v}}$$

where

$$\langle \sigma \rangle_{\vec{v}} = \frac{\int \mathcal{D}x \; e^{-S(x)} - f_{\rm cv}(x)}{\int \mathcal{D}x \; \left| e^{-S(x)} - f_{\rm cv}(x) \right|}$$



The gradient of $\langle \sigma \rangle$ is **sign-free**

$$\partial_v(-\log\langle\sigma\rangle) = -\frac{\int \mathcal{D}x \, (\partial_v \, \mathrm{Re}\, S_v) |e^{-S_v}|}{\int \mathcal{D}\phi \, |e^{-S_v}|}$$

$$S_v = -\log\left(e^{-S(x)} - f_{\rm cv}(x)\right)$$

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.

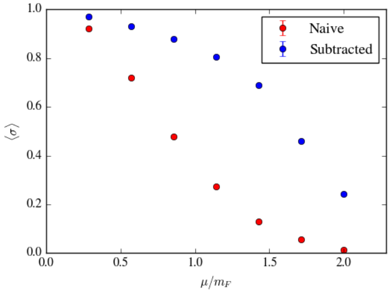
This seems to work for the density operator.... (why?)

Thirring model in 1 + 1-dimension

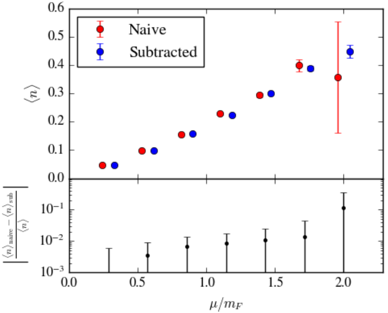
$$S = \sum_{x,\nu} \frac{2}{g^2} (1 - \cos A_\nu(x)) - \log\det K,\ A_\nu \in [0, 2\pi)$$

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

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



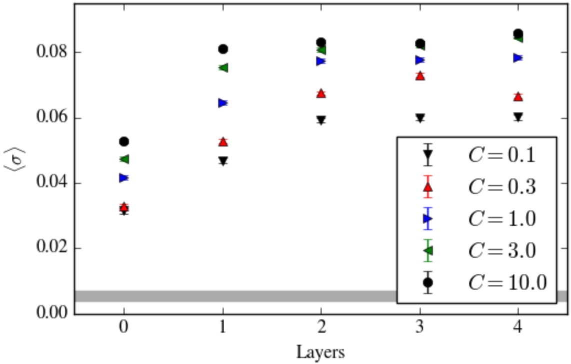
Average sign



Density

- 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

Larger networks give better vector fields



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

Future

- Better ansatz for control variates
- Compare with contour deformation methods
- Scale-up model size and complexity

Thank you!