LA-UR-24-27971

Machine-learning approaches to accelerating lattice simulations

Scott Lawrence

29 July 2024

Lattice 2024; Liverpool, UK

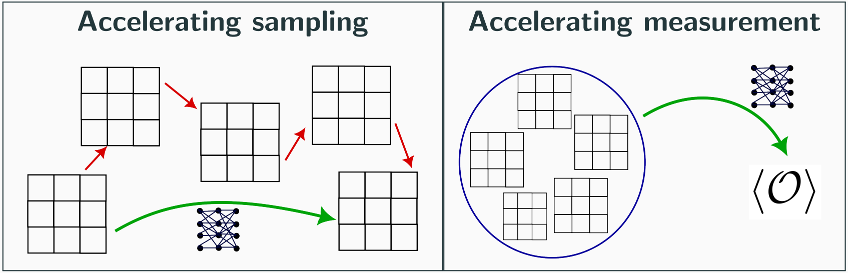
Los Alamos NATIONAL LABORATORY

NNSA National Nuclear Security Administration

Things not in this talk

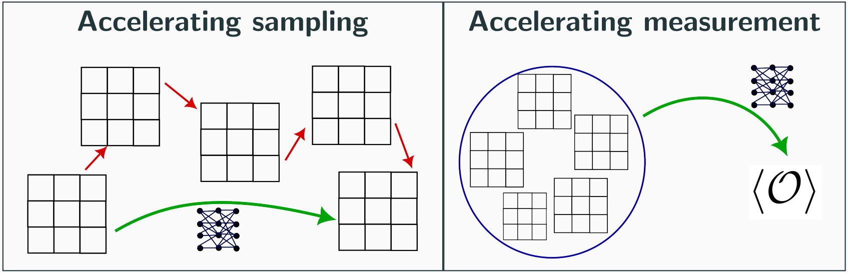
Learning spectral functions	[Chen+ 2110.13521] [Offler 2022] [Fournier+ 2020] [Kades+ 2020] [Wang-Shi-Zhou 2111.14760]
Phases or action parameters	[Carrasquilla-Melko 2017] [Peng-Tseng-Jiang 2212.14655] [Tanaka-Tomiya 1609.09087] [Nieuwenburg-Liu-Huber 1610.02048] [Rodriguez-Nieva-Scheurer 1805.05961] [Broecker+ 2017] [Shanahan-Trewartha-Detmold 1801.05784]
Learning of wave functions	[Carleo-Troyer 2017] [Deng-Li-Sarma 2017] [Luo+ 2012.05232]
Representing many-term actions	[Holland+ 2401.06481]

Design considerations



Requirement: Introduce no bias

Design considerations



Requirement: Introduce no bias

- Some algorithmic modification, specified by a function $f(\cdot)$
- An efficiently differentiable procedure for estimating the speed-up due to the use of f
- An ansatz for $f(\cdot)$ (typically a deep NN)

"Deep learning" is **often optional**.

Problem: sampling is slow

Save time by producing (nearly) statistically independent samples, without repeated querying of the action.

Sample from learned $\pi(U) \approx e^{-S(U)}$. Better approximations result in higher accetance rates.

Restricted boltzmann machines	[Huang-Wang 1610.02746] [Tanaka-Tomiya 1712.03893]
Diffusion models	[Wang-Aarts-Zhou 2309.17082] [Wang-Aarts-Zhou 2311.03578]
GANs	[Zhou+ 1810.12879] [Pawlowski-Urban 1811.03533]

Normalizing flows

An alternate view of probability distributions

Let z be a normal random variable. A general probability distribution is specified by a function $\phi(z)$.

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = p(\phi(z)) d\phi(x)$$

First(?) appeared as a method for sampling from the normal distribution given samples from $[0, 1]$ [Box-Muller 1958].

Appeared in supersymmetric and gauge theories, 40 years apart:

[Nicolai 1980]
[Luscher 0907.5491]

Modern flows, suitable for deep learning:

[Dinh-Krueger-Bengio 1410.8516]
[Dinh-Sohl-Dickstein-Bengio 1605.08803]
[Kingma-Dhariwal 1807.03039]
[Chen+ 1806.07366]

Sampling in lattice theories

Continuous normalizing flows:

[Caselle-Cellini-Nada 2307.01107]
[Haan+ 2110.02673]

Normalizing flows allow direct estimation of partition function:

[Nicoli+ 1910.13496]
[Nicoli+ 2007.07115]

The road to QCD:

[Abbott 2305.02402]
[Abbott 2207.08945]
[Albergo+ 2202.11712]
[Boyda+ 2008.05456]
[Kanwar+ 2003.06413]
[Albergo-Kanwar-Shanahan 1904.12072]

In [Abbott 2305.02402], an 8^4 lattice with $\beta = 1$ is simulated with a 288-layer generative model. ESS is $\sim 75\%$.

Is deep learning necessary?

A few-parameter ansatz (with analytic insight) can often outperform deep learning!

Reference	N_{params}	ESS at $\beta = 6$
[Luscher 0907.5491]	8	$< 1\%$
[Bacchio+ 2212.08469]	420	70%
[Boyda+ 2008.05456]	$\sim 10^6$	48%

Table reproduced from [Bacchio+ 2212.08469]

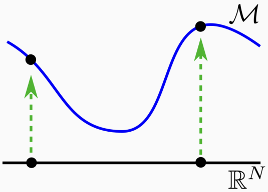
The above is for 2-dimensional Yang-Mills, where an exact normalizing flow is known [Kanwar 2106.01975].

Contour deformations

Problem: either e^{-S} or $\mathcal{O}e^{-S}$ is complex-valued and noisy.

$$\int_{\mathcal{M}} \mathrm{d}z\, e^{-S[z]} = \int_{\mathbb{R}^N} \mathrm{d}x\, e^{-S[z(x)]} \det J$$

$$\langle \sigma \rangle = \frac{\int e^{-S}}{\int |e^{-S}|} \qquad \mathrm{S2N} = \frac{\langle \mathcal{O} \rangle}{\sqrt{\langle |\mathcal{O}|^2 \rangle}}$$



Sign problems and signal-to-noise problems are closely related!

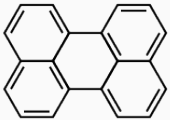
Reviewed in [Alexandru+ 2007.05436]

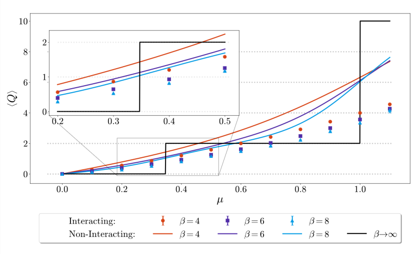
Originally, Lefschetz thimbles: [Witten 1001.2933] [Cristoforetti-Di Renzo-Scorzato 1205.3996]

Connected to analytically continued flows [**SL**-Yamauchi 2101.05755]

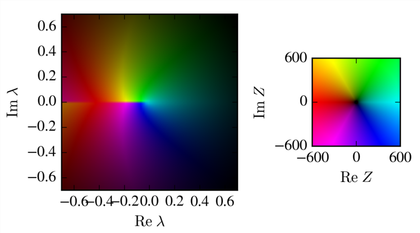
Alleviating sign problems

"Supervised" learning of thimbles:	[Alexandru-**SL**+ 1709.01971]
Although $\langle\sigma\rangle$ is difficult to estimate, $\partial \log\langle\sigma\rangle$ is quite cheap.	[Mori-Kashiwa-Ohnishi 1705.05605] [Alexandru-**SL**+ 1804.00697]
Fermionic systems	[Alexandru-**SL**+ 1808.09799] [Gäntgen+ 2307.06785] [Rodekamp+ 2406.06711]
Gauge theories	[Alexandru-**SL**+ 1807.02027] [Kashiwa-Mori 2007.04167] [Namekawa+ 2109.11710] [Basar-Marincel 2208.02072]
Heavy-dense QCD	[Basar-Marincel 2311.06343]
Real-time dynamics	[**SL**-Yamauchi 2101.05755] [Kanwar-Wagman 2103.02602]
Complex couplings	[**SL**-Oh-Yamauchi 2205.12303]





Charge of perylene ($C_{20}H_{12}$).
From [Rodekamp+ 2406.06711],
with one parameter determined
analytically.



Partition function (analytically continued) of scalar field theory in $0+1$ dimensions, at $m^2 = 0.5$, as a function of complex coupling. Computed with a complex normalizing flow.
From [**SL**-Oh-Yamauchi 2205.12303], with ~ 200 parameters.

Contours for signal-to-noise problems

Re-sampling on a deformed contour is expensive!

[Detmold+ 2003.05914]
[Detmold+ 2101.12668]

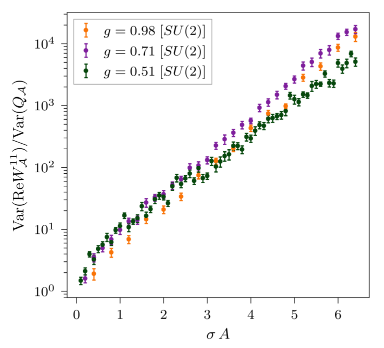
Move the contour-dependence from the domain to the integrand, by parameterizing $\mathbb{R} \to \gamma$:

$$\frac{\int_\gamma dz\, e^{-S}\mathcal{O}}{\int_\gamma dz\, e^{-S}} = \frac{\int_\mathbb{R} dx\, e^{-S_{\text{eff}}(x)}\mathcal{O}(z(x))}{\int_\mathbb{R} dx\, e^{-S_{\text{eff}}(x)}}$$

Rewrite with reweighting:

$$= \left\langle e^{S(x)-S_{\text{eff}}(x)}\mathcal{O}(z(x))\right\rangle \left\langle e^{S-S_{\text{eff}}}\right\rangle^{-1}$$

Expectations w.r.t. *original* ensemble.

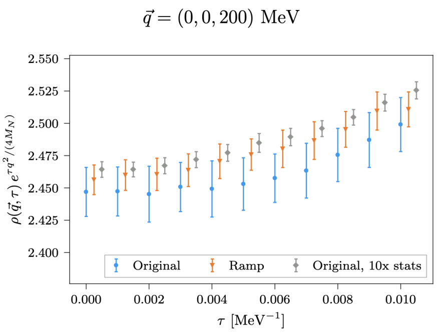


Improvement in variance of estimator of Wilson loops of various areas, in two-dimensional $SU(2)$ Yang-Mills. From [Detmold+ 2101.12668]

Broader applicability of contour deformations

To spin systems: [Kashiwa+ 2309.06018] [Warrington 2310.19761] [Mooney+ 2110.10699]

To non-holomorphic actions: [Kanwar+ 2304.03229] [**SL**+ 2401.16733]



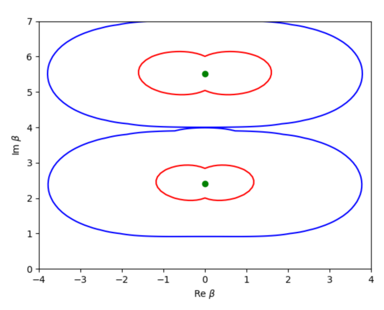
GFMC-measured Euclidean response of density in the deuteron.
From [Kanwar+ 2304.03229], with one parameter.

No-go theorems for contour deformations

Let α be any closed differential form obeying $|\alpha| \leq |e^{-S} dz|$.
Then

$$\left|\int_{\mathbb{R}} \alpha\right| = \left|\int_{\gamma} \alpha\right| \leq \int_{\gamma} |\alpha| \leq \int_{\gamma} |\omega|$$

Searching for such α gives upper bounds on the best-possible average phase.
Bounds of this form are generically exponential in volume.



For 2-d Yang-Mills at complex coupling β: regions for which a perfect contour can be found (outside blue) and can be proven not to exist (inside orange). From [**SL**-Yamauchi 2311.13002]

Not all sign problems can be removed by contour deformations!

Problem: Many observables are expensive.

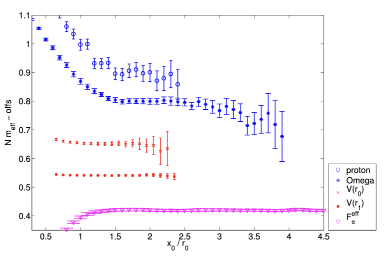
Replace with equivalent, but less expensive, observable.

"Surrogate observables"

Problem: Many observables are noisy.

Replace with equivalent, but less noisy, observable.

"Control variates"



Effective masses, from the FLAG review 2021 [Aoki 2111.09849].

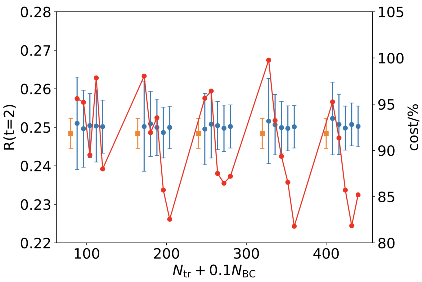
Variance reduction with normalizing flows: [Abbott+ 2401.10874]

Surrogate observables

Approximate $\mathcal{O}$ (physical observable of interest) by $\tilde{\mathcal{O}}$ (efficient approximation).

$$\langle \mathcal{O} \rangle = \langle \tilde{\mathcal{O}} \rangle - \left(\langle \tilde{\mathcal{O}} \rangle_s - \langle \mathcal{O} \rangle_s \right)$$

[Yoon-Bhattacharya-Gupta 1807.05971]
[Zhang+ 1909.10990]

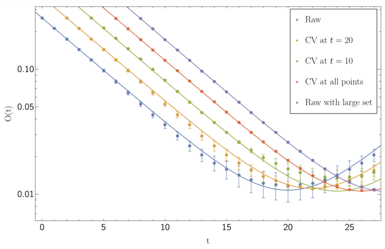


Predicting $R(t = 2)$ (ratio of 3-pt to 2-pt) of kaon quasi-PDF correlators at $(p_{\mathrm{pred}}, t_{\mathrm{pred}}, t_{\mathrm{sep}}) = (4, 4, 5)$ from measurements at $(3, 4, 5)$. From [Zhang+ 1909.10990]; a linear model.

Control variates from Schwinger-Dyson relations

Replace $\mathcal{O}$ by $(\mathcal{O} - f)$, where $\langle f \rangle = 0$. If f is correlated with $\mathcal{O}$, measurements are less noisy.

$$0 = \langle \partial g - g \partial S \rangle$$



Correlator in 2d scalar field theory, improved with deep neural control variates. From [Bedaque-Oh 2312.08228], with $\sim 6 \times 10^3$ parameters.

The object to learn is g.

[Bhattacharya-**SL**-Yoo 2307.14950]
[Bedaque-Oh 2312.08228]
[**SL** 2404.10707]

Not much success for sign problems:

[**SL** 2009.10901]
[**SL**-Yamauchi 2212.14606]
[**SL**-Yamauchi 2312.12636]

On reproducibility

Poor adoption into physics calculations

Many methods not ready for "prime time" (QCD)
—but are also unused on lower-dimensional systems.

Some efforts to diffuse knowledge of techniques, e.g. a normalizing flow tutorial [Albergo+ 2101.08176] and two software packages:

[Tomiya-Terasaki 2208.08903]
[Nicoli+ 2024]

On reproducibility

Poor adoption into physics calculations

Many methods not ready for "prime time" (QCD)
—but are also unused on lower-dimensional systems.

Some efforts to diffuse knowledge of techniques, e.g. a normalizing flow tutorial [Albergo+ 2101.08176] and two software packages:

[Tomiya-Terasaki 2208.08903]
[Nicoli+ 2024]

A good first step: make **code** and **weights** publicly available.

Develop reusable software packages, **specialized** for lattice field theory.