Discussion on machine learning and lattice field theories

Gert Aarts

Swansea University Prifysgol Abertawe

FONDAZIONE BRUNO KESSLER

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

Algorithms, Edinburgh, April 2023

ML and LFT

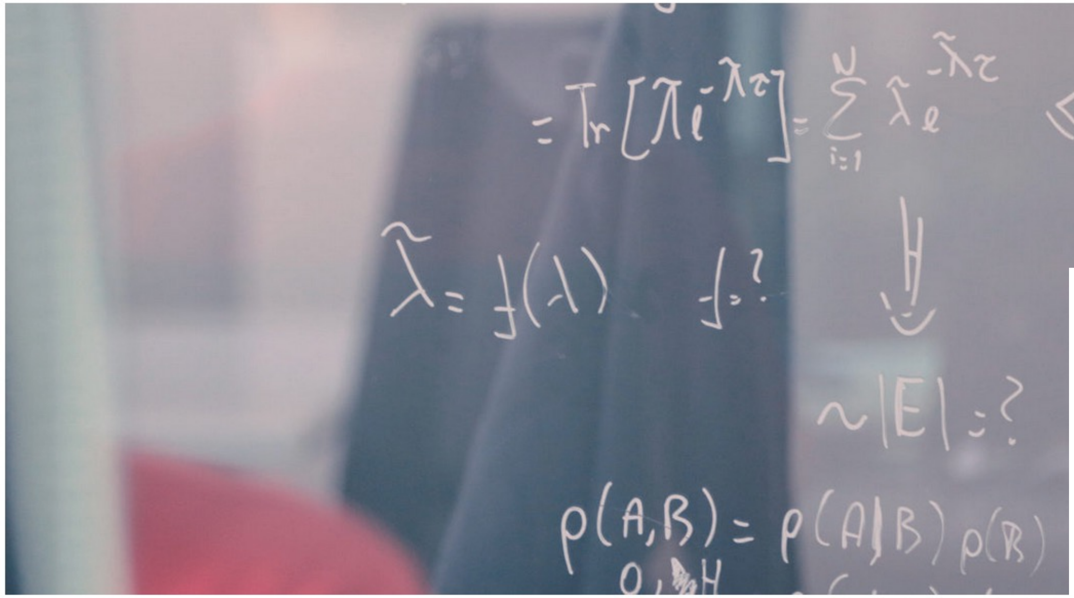
- ensemble generation
- analysis
- sign and noise problems

- understand ML using
  physics insights, stat mech and LFT
- stochastic gradient descent
- quantum ML
- new applications, anything else

FONDAZIONE BRUNO KESSLER

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

MACHINE LEARNING FOR LATTICE FIELD THEORY AND BEYOND



26 June 2023 — 30 June 2023

Organizers

Daniel Hackett (Massachusetts Institute of Technology, United States)
dhackett@mit.edu
Gert Aarts (Swansea University, United Kingdom)
g.aarts@swansea.ac.uk
Dimitrios Bachtis (Swansea University, United Kingdom)
Dimitrios.Bachtis@swansea.ac.uk
Biagio Lucini (Swansea University, United Kingdom)
B.Lucini@swansea.ac.uk
Phiala Shanahan (Massachusetts Institute of Technology, United States)
pshana@mit.edu

registration is open: indico.ectstar.eu/event/171/

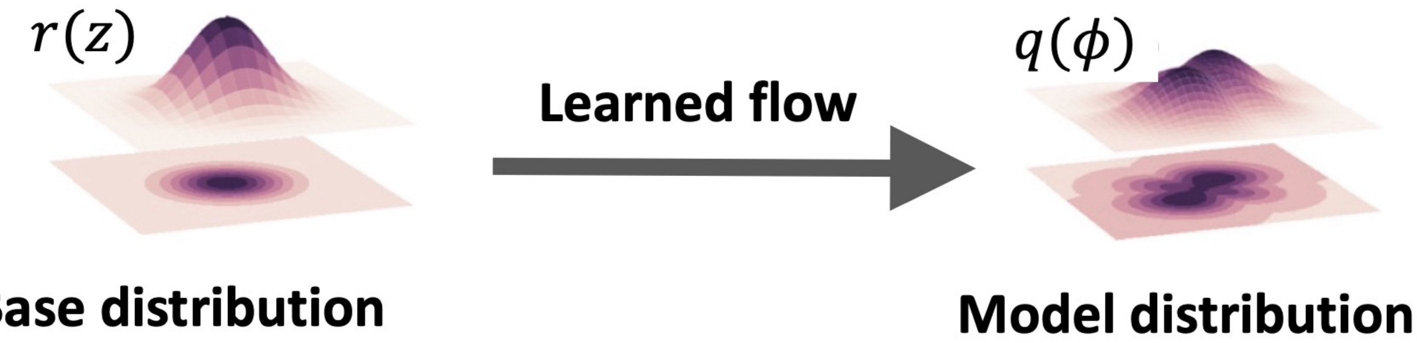
ML applied to many problems in LFT

- ensemble generation
  - normalising flow, trivialising maps
  - learning from data
  - equivariant multi-grid
  - learn parameters in nonstandard actions

- questions to discuss:
  - (provable) exactness, in theory and in practice
  - ergodicity
  - scalability
  - implementation of gauge equivariant formulations
  - from low-dimensional modes to QCD

Abbott et al, 2211.07541

Normalising flow



Exact model density:

$$q(\phi) = \left|\det \frac{\partial \phi}{\partial z}\right|^{-1} r(z)$$

- base distribution: $r(z)$ trivial
- target distribution: $p(\varphi)$
- model distribution: $q(\varphi)$
- aim: $q(\varphi) \sim p(\varphi)$
- exact: accept/reject $q(\varphi)/p(\varphi)$

Ryan Abbott

Disussion – Normalizing Flows

- Current status: can build flows for QCD, but more work is needed particularly on expressivity and scaling to larger models
- Flow layers: spectral[Boyda et al. 2008.05456], residual[Abbott et al. 2304.xxxx], continuous/ODE [Bacchio et al., 2212.08469]
  - Gauge equivariance greatly restricts map (particularly for SU(N)) $\implies$ need more work on gauge-equivariant SU(N) flows
  - Need deeper and more expressive flows for scaling towards practical parameters (current flows are very shallow – only O(1-5) updates/link)
- Fermions – flows work w/exact det, but need more scalable approach
  - Possibilities: [Albergo et al. 2106.05934]
    - Autoregressive modeling: model $p(U)$ and $p(\phi \mid U)$
    - Gibbs sampling: model $p(U \mid \phi)$, sample $p(\phi \mid U)$ exactly
    - Full joint sampling: transform U and ϕ together to $p(U, \phi)$
  - So far focused on autogressive pseudofermion modeling

High-level questions

- deep learning versus multigrid paradigm
- building global and local symmetries into the network versus learning the symmetries

Disussion – Normalizing Flows

- Hybrid algorithms – many possibilities to explore
  - HMC on trivialized distribution [Lüscher 0907.5491]
  - Generalize proposal distribution for HMC [Foreman et al., 2112.01582]
  - Subdomain modeling [Finkenrath, 2201.02216]
  - Stochastic Normalizing Flows [Wu et al. 2002.0670]
  - CRAFT/Annealed Importance Sampling [Matthews et al. 2201.13117]
  - ...
- Beyond sampling
  - Mapping between different actions
  - Contour deformation and density of states approaches to the sign problem [Detmold et al., 2101.12668] [Pawlowski+Urban, 2203.01243] [Lawrence et al., 2205.12303]

ML applied to many problems in LFT

- analysis
  - phase classification, detection of order parameters
  - inverse problems: spectral functions, transport, parton distribution functions
  - error reduction

- questions to discuss:
  - precision
  - reliable error estimates: systematic and statistical
  - comparison with known methods/learn something new
  - parametrised vs non-parametrised approaches

Inverse problem: spectral function reconstruction

- given $\rho(\omega)$: computation of $G(\tau)$ is easy

$$G(\tau) = \int d\omega\, K(\tau,\omega)\rho(\omega)$$

- given $G(\tau)$: computation of $\rho(\omega)$ is hard, ill-posed inversion problem

standard argument:

- $G(\tau)$ known numerically at $\mathrm{O}(16-64)$ points
- $\rho(\omega)$ in principle continuous function, with sharp and broad structures
- integral over known kernel $K(\tau,\omega)$ washes out information

Julian Urban

ML & inverse problems

- Probabilistic algorithms for Fredholm integral inversion
  - Must-have: reliable uncertainty estimates
  - Should-have: incorporating various types of prior knowledge

$\longrightarrow$ Spectral densities of QCD correlators for real-time physics

$\longrightarrow$ Extraction of parton distribution functions

$\longrightarrow$ CMT stuff, e.g. optical conductivity in Hubbard model

- Inverse problems with known solutions as pretext tasks
  - Action parameter regression from raw configurations

  $\longrightarrow$ Identifying order parameters and effective d.o.f.

  - ?

RG & optimal transport

- ERG and OT recently shown to be mathematically equivalent (Polchinski flow = OT gradient flow of relative entropy)
- RG picture of Wilson flow, trivializing maps
- Functional RG interpretation of regularized stochastic quantization (colored-noise Langevin)
- Relation of OT to normalizing flows

$\longrightarrow$ Interesting parallels and connections, but can we exploit these formal insights computationally?

ML applied to many problems in LFT

- sign and noise problems
  - complex actions: finite density, real time
  - signal-to-noise deterioration

- questions to discuss:
  - deformations in complex plane (complex Langevin, thimbles, holomorphic flow)
  - optimise manifold
  - reduce average sign, variance
  - reliable error estimates: systematic and statistical

Tej Kanwar

Exactness in ML for lattice

Sampling algorithms:

- Markov chain guarantees vs practical ergodicity
- Reliability of Effective sample size metric?
- Usage for thermalization?
- Hybrid ML + HMC approaches?

Non-sampling applications:

- Sign/signal-to-noise problems
- Interpolating operators

✅ ML applications **all** designed to be asymptotically exact.

🕐 Need to ask about **practical** exactness in **real** applications.

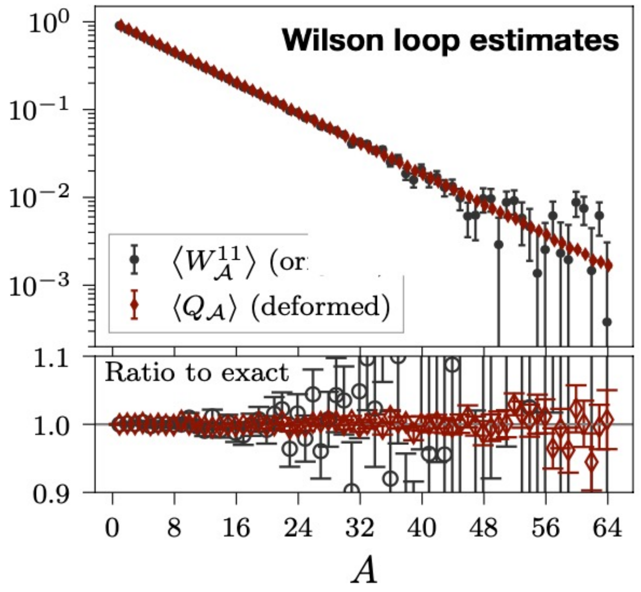
Learned contour deformations for StN problems

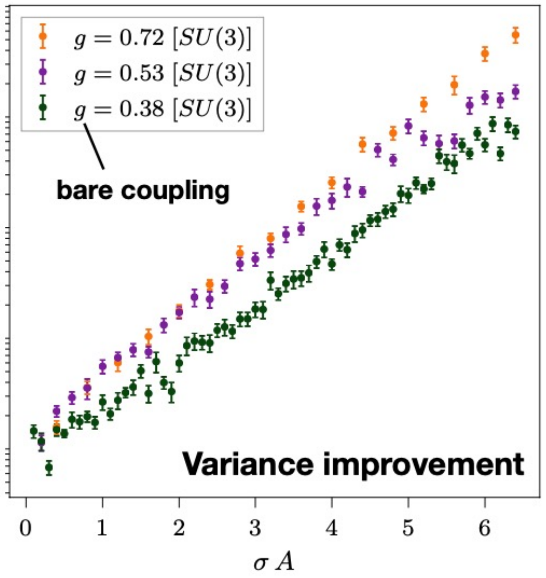
Using complex analysis, one can …

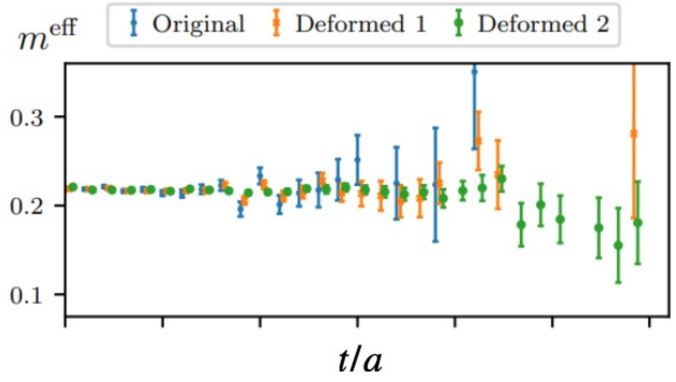
- **Deform observables $\mathcal{O} \to \mathcal{Q}$**, where $\langle \mathcal{O} \rangle = \langle \mathcal{Q} \rangle$ but $\mathrm{Var}[\mathcal{O}] \neq \mathrm{Var}[\mathcal{Q}]$.

- **Minimize variance numerically** (using existing MC samples).

- **Achieve far more precise** measurements in proof-of-principle applications to lattice field theories.







To do: Higher dimensions, fermions, advanced ML techniques, non-lattice methods.

ML applied to many problems in LFT

- understand ML using physics, stat mech and LFT
  - stochastic gradient descent
  - quantum machine learning
  - energy based models, e.g. Restricted Boltzmann Machines
  - phase structure of energy based models [2011.11307]
  - relation to Markov random fields [2102.09449]
  - partition function formulation of deep neural nets [2209.04882]

ML applied to many problems in LFT

- stochastic gradient descent
  - loss function landscape and (discretised) updates
  - implicit gradient regularisation for popular methods, such as Adam
  - batch size dependence and effective noise
  - exactness?

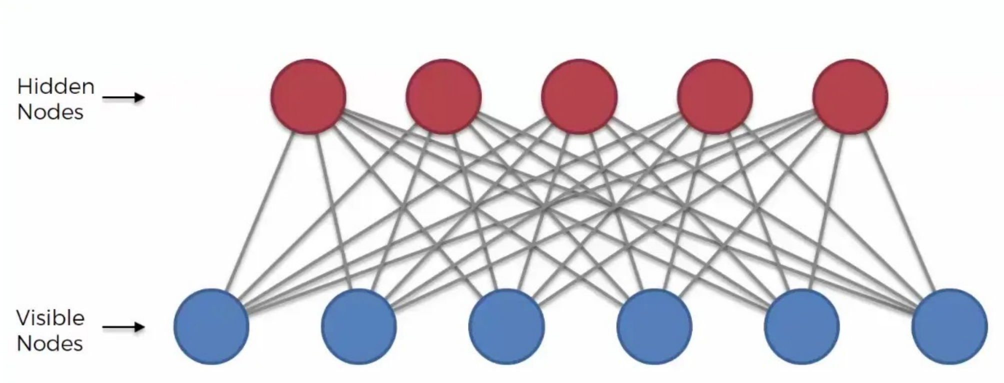
ML applied to many problems in LFT

- quantum machine learning
  - architecture of hardware vs algorithms
  - all-to-all vs nearly linearly connected qubits

  - joint developments?
  - should algorithms/hardware be designed with hardware/algorithms in mind?

  - demonstrate quantum advantage for physically relevant problems?

Quantum Restricted Boltzmann Machine



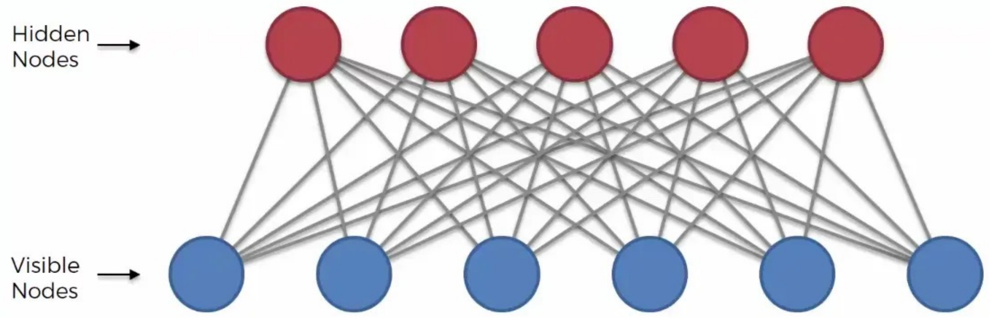
- quantum ground states
- quantum variational approaches
- learn from 100 years of quantum experience
- compare/compete with tensor networks, exact methods (small systems), …
- what is the target? physics, QC, larger systems, …

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

Restricted Boltzmann Machine

- Restricted Boltzmann Machine (RBM): two-layer generative network

- visible layer and hidden layer



- restricted: no connections within a layer

- standard RBM: spin degrees of freedom on each node

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

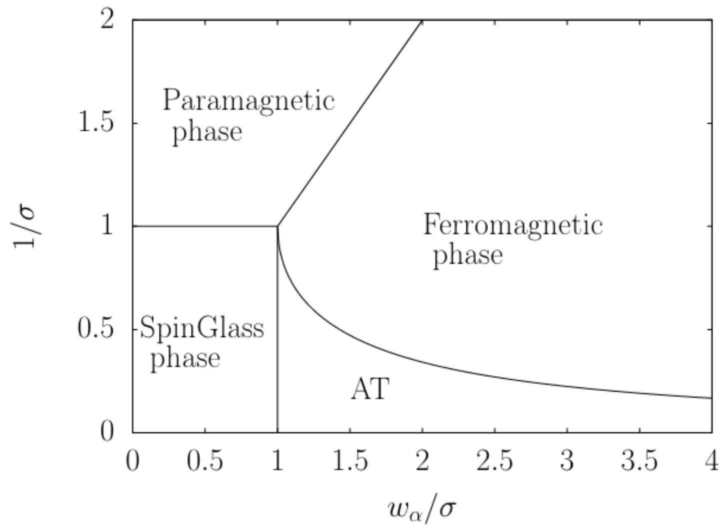
- energy function, distribution

$$E(v,h) = -\sum_{i,a} v_i w_{ia} h_a - \sum_i b_i v_i - \sum_a \eta_a h_a \qquad p(v,h) = \frac{1}{Z} e^{-E(v,h)} \qquad Z = \sum_{\{v_i\},\{h_a\}} e^{-E(v,h)}$$

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

Restricted Boltzmann Machines, recent advances and mean-field theory
A. Decelle and C. Furtlehner,
arXiv:2011.11307

Binary RBM phase diagram



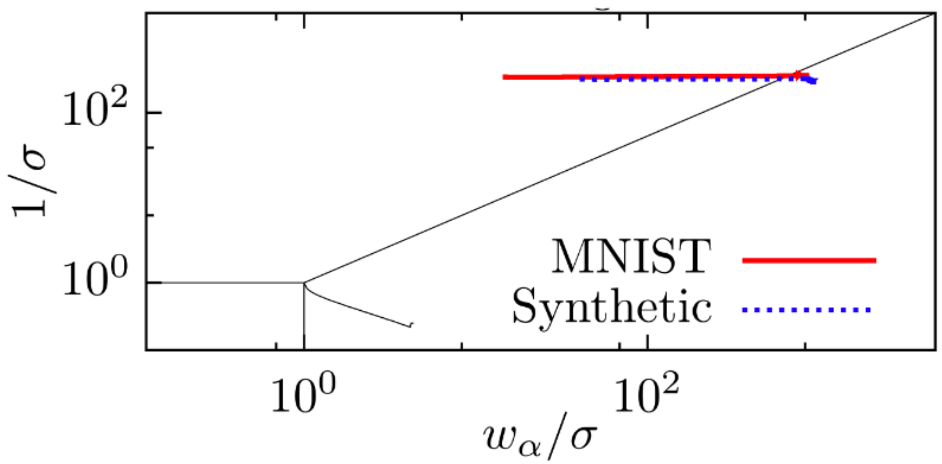


Figure 4: **Left:** the phase diagram of the model. The y-axis corresponds to the variance of the noise matrix, the x-axis to the value of the strongest mode of $\boldsymbol{w}$. We see that the ferromagnetic phase is characterized by having strong mode eigenvalues. In this phase, the system can behave either by recalling one eigenmode of $\boldsymbol{w}$ or by composing many modes together (compositional phase). For the sake of completeness, we indicate the AT region where the replica symmetric solution is unstable, but for practical purpose we are not interested in this phase. **Right:** An example of a learning trajectory on the MNIST dataset (in red) and on a synthetic dataset (in blue). It shows that starting from the paramagnetic phase, the learning dynamics brings the system toward the ferromagnetic phase by learning a few strong modes.

Summary and outlook

- ✓ new solutions to old problems/old solutions to new problems
- ✓ new insights to both LFT and ML