Non-perturbative renormalization for the neural network-QFT correspondence

Harold Erbin

Mit & Iaifi (Usa) & Cea-List (France)

Idp, Tours – 27 January 2022

In collaboration with:

- Vincent Lahoche (Cea-List)
- Dine Ousmane Samary (Université d'Abomey-Calavi)

arXiv: 2108.01403

Massachusetts Institute of Technology

Funded by the European Union (Horizon 2020)

list cea tech

Outline: 1. Motivations

Talk highlights

- neural networks and uses in physics in a nutshell
- NN-QFT correspondence between:
  - statistical ensemble of neural networks
  - (Euclidean) quantum field theory
- describe the correspondence
  - infinite-width neural network = Gaussian process = free QFT (i.e. infinite number of neurons)
  - finite-width = interactions
  - data-space and theory space
  - renormalization group
  - numerical results
- goal: effective theory of learning
  - improve efficiency
  - improve architecture design

Why machine learning?

ML applications in theoretical physics [1903.10563, Carleo et al.]

- cosmology
- lattice theories
- many-body physics
- particle physics
- quantum information
- string theory

Why machine learning?

ML applications in theoretical physics [1903.10563, Carleo et al.]

- cosmology
- lattice theories
- many-body physics
- particle physics
- quantum information
- string theory

Typical problems

- big data: large-dimensional and noisy data to be related to theoretical models
  (particle colliders, gravitational waves, galaxy surveys. . .)
- exploration: landscape of possible models too large or not well understood (BSM phenomenology, NN-QFT. . .)
- computational: interaction structure prevents writing the model explicitly or making analytic computations
  (strong coupling, many-body physics. . .)

Why neural networks?

Universal approximation theorem

Under mild assumptions, a feed-forward network $f(x)$ with a finite number of neurons can approximate any continuous function $F(x)$ on compact subsets of $\mathbb{R}^n$.

[Cybenko '89; Hornik-Stinchcombe-White '89; 1709.02540, Lu et al.]

Why neural networks?

Universal approximation theorem

Under mild assumptions, a feed-forward network $f(x)$ with a finite number of neurons can approximate any continuous function $F(x)$ on compact subsets of $\mathbb{R}^n$.

[Cybenko '89; Hornik-Stinchcombe-White '89; 1709.02540, Lu et al.]

- neural network (NN) $f(x) =$ sequence of:
  - matrix multiplication + translations (learnable parameters)
  - element-wise non-linear functions (fixed)
- supervised learning: given a set of pairs $(x_i, y_i = F(x_i))$, tune parameters with gradient descent such that $\forall i : f(x_i) \approx F(x_i)$
- motivations
  - generically outperform all other machine learning algorithms
  - can outperform human experts
  - transfer learning (train for one task, apply to other tasks)

What is a neural network?

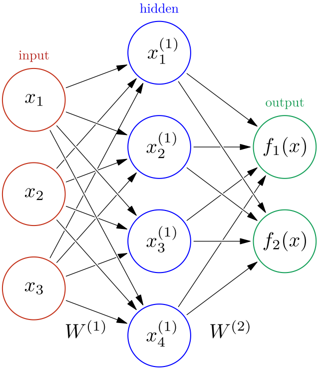
$$x^{(0)}_{i_0} := x_{i_0}$$

$$x^{(1)}_{i_1} = g^{(1)}\left(W^{(1)}_{i_1 i_0} x^{(0)}_{i_0}\right)$$

$$f_{i_2}(x_{i_0}) := x^{(2)}_{i_2} = g^{(2)}\left(W^{(2)}_{i_2 i_1} x^{(1)}_{i_1}\right)$$

$i_0 = 1, 2, 3$; $i_1 = 1, \dots, 4$; $i_2 = 1, 2$

$K = 1$; $d_{\text{in}} = 3$; $d_{\text{out}} = 2$; $N^{(1)} = 4$



- input $x^{(0)} := x \in \mathbb{R}^{d_{\text{in}}}$
- $K \geq 1$ hidden layers, $n \in \{1, \dots, K\}$
  - layer n: $N^{(n)}$ neurons (units) $x^{(n)} \in \mathbb{R}^{N^{(n)}}$
  - learnable weights $W^{(n)} \in \mathbb{R}^{N^{(n)} \times N^{(n-1)}}$
  - learnable biases $b^{(n)} \in \mathbb{R}^{N^{(n)}}$ (not displayed)
  - fixed activation functions $g^{(n)}$ (element-wise)
- output $x^{(K+1)} := f(x) \in \mathbb{R}^{d_{\text{out}}}$

Problems with neural networks

- black box: hard to understand the meaning of computations
- loss landscape: loss function non-convex and very rough, hard to find (global) minimum (related to spin glass)
  [1412.0233, Choromanska et al.; 1712.09913, Li et al.]
- complicated training: expensive computationally, convergence issues...
  [syncedreview.com/cost-of-training-sota-ai-models/]
- hyperparameter tuning: mostly trial and errors or random/Bayesian/bandit optimization
- expressibility: which functions can be approximated, under which conditions?
  [1606.05336, Raghu et al.]

Why physics?

- effective description (no need to know fundamental theory)
- efficient representation of statistical models (path integral, Feynman diagrams)
- collective dynamics of degrees of freedom and organization by scales (renormalization, phase transitions)

$\rightarrow$ develop tools to improve analytical understanding of neural network building and training

[Krauth-Mézard '87; Gardner '88; Gardner-Derrida '88; Krauth-Mézard-Nadal '88; Krauth-Mézard-Nadal '88; Amaldi-Nicolis '89; Krauth-Mézard '89; Mézard-Nadal '89; Nicolis '92; . . . ; 1608.08225, Lin-Tegmark-Rolnick; 1903.10563, Carleo et al.; Zdeborová '21]

Plan

NN-QFT correspondence

For a very general class of architectures, it is possible to associate a quantum field theory (QFT) to a statistical ensemble of neural networks (NN).

[2008.08601, Halverson-Maiti-Stoner (HMS)]

Plan

NN-QFT correspondence

For a very general class of architectures, it is possible to associate a quantum field theory (QFT) to a statistical ensemble of neural networks (NN).

[2008.08601, Halverson-Maiti-Stoner (HMS)]

In this talk [2108.01403, HE-Lahoche-Samary]:

- describe the NN-QFT correspondence
- discuss the theory space
- establish RG flow for the QFT
- provide numerical results

Main "experimental" result

Varying the standard deviation of the weight distribution induces an RG flow in the space of neural networks.

Outline: 2. NN-QFT correspondence

Neural network

- fully connected neural network (one hidden layer)

$$f_{\theta,N} : \mathbb{R}^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$$
$$f_{\theta,N}(x) = W_1\Big(g(W_0 x + b_0)\Big) + b_1$$

- width N, activation function g
- parameters (weights and biases): Gaussian distributions

$$\theta = (W_0, b_0, W_1, b_1)$$
$$W_0 \sim \mathcal{N}(0, \sigma_W^2/d_{\text{in}}), \qquad W_1 \sim \mathcal{N}(0, \sigma_W^2/N)$$
$$b_0, b_1 \sim \mathcal{N}(0, \sigma_b^2)$$

Dual description

- consider statistical ensemble of neural networks defined by distribution in parameter space
- specific NN = sample from distribution

$$f_{\theta,N} \sim P[\theta]$$

Dual description

- consider statistical ensemble of neural networks defined by distribution in parameter space
- specific NN = sample from distribution

$$f_{\theta,N} \sim P[\theta]$$

- dual description: parameter dist. + architecture
  induces distribution in function space

$$f_{\theta,N} \sim p[f]$$

Dual description

- consider statistical ensemble of neural networks defined by distribution in parameter space
- specific NN = sample from distribution

$$f_{\theta,N} \sim P[\theta]$$

- dual description: parameter dist. + architecture
  induces distribution in function space

$$f_{\theta,N} \sim p[f]$$

- training = change parameter dist. = flow in function space

Note: no training in this talk (see [2106.00694, HMS])

Large N limit, Gaussian process and free QFT

Large N limit = infinite layer width:

- NN (function) distribution drawn from Gaussian process (GP) with kernel K (consequence of central limit theorem) [Neal '96]

$$f \sim \mathcal{N}(0, K)$$

- generalize to most architectures [1910.12478, Yang] and training

Large N limit, Gaussian process and free QFT

Large N limit $=$ infinite layer width:

- NN (function) distribution drawn from Gaussian process (GP) with kernel K (consequence of central limit theorem) [Neal '96]

$$f \sim \mathcal{N}(0, K)$$

- generalize to most architectures [1910.12478, Yang] and training
- log probability

$$S_0[f] = \frac{1}{2} \int \mathrm{d}^{d_{\text{in}}} x \mathrm{d}^{d_{\text{in}}} x' \, f(x) \Xi(x, x') f(x'), \qquad \Xi := K^{-1}$$

- n-point correlation (Green) functions (fixed by Wick theorem)

$$G_0^{(n)}(x_1, \ldots, x_n) := \int \mathrm{d}f \, \mathrm{e}^{-S_0[f]} \, f(x_1) \cdots f(x_n)$$

Large N limit, Gaussian process and free QFT

Large N limit = infinite layer width:

- NN (function) distribution drawn from Gaussian process (GP) with kernel K (consequence of central limit theorem) [Neal '96]

$$f \sim \mathcal{N}(0, K)$$

- generalize to most architectures [1910.12478, Yang] and training
- log probability

$$S_0[f] = \frac{1}{2} \int \mathrm{d}^{d_{\text{in}}} x \mathrm{d}^{d_{\text{in}}} x' \, f(x) \Xi(x, x') f(x'), \qquad \Xi := K^{-1}$$

- n-point correlation (Green) functions (fixed by Wick theorem)

$$G_0^{(n)}(x_1, \ldots, x_n) := \int \mathrm{d}f \, \mathrm{e}^{-S_0[f]} \, f(x_1) \cdots f(x_n)$$

"If it looks like a duck, swims like a duck, and quacks like a duck, then it probably is a duck."

Large N limit, Gaussian process and free QFT

Large N limit = infinite layer width:

- NN (function) distribution drawn from Gaussian process (GP) with kernel K (consequence of central limit theorem) [Neal '96]

$$f \sim \mathcal{N}(0, K)$$

- generalize to most architectures [1910.12478, Yang] and training
- log probability

$$S_0[f] = \frac{1}{2} \int \mathrm{d}^{d_{\text{in}}}x \mathrm{d}^{d_{\text{in}}}x' \, f(x) \Xi(x, x') f(x'), \qquad \Xi := K^{-1}$$

- n-point correlation (Green) functions (fixed by Wick theorem)

$$G_0^{(n)}(x_1, \dots, x_n) := \int \mathrm{d}f \, \mathrm{e}^{-S_0[f]} \, f(x_1) \cdots f(x_n)$$

This looks like a free QFT.

Finite N and interactions

- for finite N, non-GP $\Rightarrow$ deviations of Green functions

$$\Delta G^{(n)} := G^{(n)} - G_0^{(n)}$$

Finite N and interactions

- for finite N, non-GP $\Rightarrow$ deviations of Green functions

$$\Delta G^{(n)} := G^{(n)} - G_0^{(n)}$$

- in QFT: non-Gaussian contributions = interactions

$$S[f] = S_0'[f] + S_{\text{int}}[f]$$

note: free action $S_0'[f]$ unknown

Finite N and interactions

- for finite N, non-GP $\Rightarrow$ deviations of Green functions

$$\Delta G^{(n)} := G^{(n)} - G_0^{(n)}$$

- in QFT: non-Gaussian contributions = interactions

$$S[f] = S_0'[f] + S_{\text{int}}[f]$$

note: free action $S_0'[f]$ unknown

- n-point Green functions

$$G^{(n)}(x_1, \ldots, x_n) := \int \mathrm{d}f\, \mathrm{e}^{-S[f]}\, f(x_1) \cdots f(x_n)$$

- effective (IR) 2-point function exactly known ($G^{(2)}$ N-indep.)

$$G^{(2)}(x, y) = K(x, y) = G_0^{(2)}(x, y)$$

- work with 1PI effective action

$$\Gamma[f] = S_0[f] + \Gamma_{\text{int}}[f]$$

Summary of NN-QFT correspondence

	QFT	NN / GP
x	spacetime points	data-space inputs
p	momentum space	dual data-space
f	field	neural network
$K(x, y)$	propagator	Gaussian kernel
S	action	log probability
S_0	free action	Gaussian log probability
S_{int}	interactions	non-Gaussian corrections

Why is it interesting?

- correlation functions between outputs (= Green functions) give measure of learning
- e.g. 1-point function $\langle f(x) \rangle$ = average prediction for input x (relation with symmetry breaking)

Theory space

- data-space $\neq$ spacetime $\rightarrow$ avoid bias from particle physics

Theory space

- data-space $\neq$ spacetime $\rightarrow$ avoid bias from particle physics
- spacetime $= x \in \mathbb{R}^{d_{\text{in}}}$ + group structure + causality
  - dictates how coordinates can be transformed
  - invariances by translation, rotation, coordinate permutation
  - locality (from kinetic operator, prevents causality violation)

Theory space

- data-space $\neq$ spacetime $\rightarrow$ avoid bias from particle physics
- spacetime $= x \in \mathbb{R}^{d_{\text{in}}}$ + group structure + causality
  - dictates how coordinates can be transformed
  - invariances by translation, rotation, coordinate permutation
  - locality (from kinetic operator, prevents causality violation)
- data-space
  - bi-local kernel $\rightarrow$ non-local interactions?
  - symmetries of inputs and outputs?
  - natural UV cutoff: machine precision

Theory space

- data-space $\neq$ spacetime $\rightarrow$ avoid bias from particle physics
- spacetime $= x \in \mathbb{R}^{d_{\text{in}}}$ + group structure + causality
  - dictates how coordinates can be transformed
  - invariances by translation, rotation, coordinate permutation
  - locality (from kinetic operator, prevents causality violation)
- data-space
  - bi-local kernel $\rightarrow$ non-local interactions?
  - symmetries of inputs and outputs?
  - natural UV cutoff: machine precision
- neural network phenomenology
  1. assumptions dictated by numerical evidences
  2. write model to match observations
  3. use model to check theoretical facts (dualities…)

  ex.: local interactions sufficient for simple models [2008.08601, HMS], study of input/output symmetries [2106.00694, HMS]

Examples of interactions

- local interactions

$$S_{\text{int}} = \sum_n g_n \int \mathrm{d}^{d_{\text{in}}} x \, f(x)^n$$

- non-local interactions and coupling functions

$$S_{\text{int}} = \int \mathrm{d}^{d_{\text{in}}} x_1 \cdots \mathrm{d}^{d_{\text{in}}} x_n \, g(x_1, \ldots, x_n) f(x_1) \cdots f(x_n)$$

- delocalized fields [2111.03672, HE-Fırat-Zwiebach]

$$\tilde{f}(x) := \int \mathrm{d}^{d_{\text{in}}} y \, \kappa(x, y) f(y)$$

- tensor models: break permutation invariance

$$S_{\text{int}} = g \int \mathrm{d}^3 x \mathrm{d}^3 y \mathrm{d}^3 z \, \rho(x, y, z) \, f(x_1, y_2, z_3) f(y_1, z_2, x_3) f(z_1, x_2, y_3)$$

too wild, restrict to random tensor field theories [Gurau '16]:
$f(x)$ = 3-tensor field with continuous indices, pairwise contractions

GaussNet

Setup in this talk and [2108.01403, HE-Lahoche-Samary]:

- take $d_{\text{out}} = 1$
- translation-invariant activation function

$$g(W_0 x + b_0) = \frac{\exp(W_0 x + b_0)}{\sqrt{\exp\left[2\left(\sigma_b^2 + \frac{\sigma_W^2}{d_{\text{in}}} x^2\right)\right]}}$$

(stricly speaking, activation func. + normalization)

- GP kernel [2008.08601, HMS]

$$K(x, y) := \sigma_b^2 + K_W(x, y), \quad K_W(x, y) = \sigma_W^2 \, \mathrm{e}^{-\frac{\sigma_W^2}{2 d_{\text{in}}} |x - y|^2}$$

- note: [2008.08601, HMS] also considers ReLU and Erf functions

Numerical setup

[2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

- $d_{\text{in}} = 1, \sigma_b = 1,\ N \in \{2, 3, 4, 5, 10, 20, 50, 100, 500, 1000\}$
- n_{bags} distinct statistical ensembles of n_{nets} networks each
- "experimental" Green functions

$$\bar{G}^{(n)}_{\text{exp}}(x_1, \ldots, x_n) := \frac{1}{n_{\text{bags}}} \sum_{A=1}^{n_{\text{bags}}} G^{(n)}_{\text{exp}}(x_1, \ldots, x_n)\big|_{\text{bag } A}$$

$$G^{(n)}_{\text{exp}}(x_1, \ldots, x_n) := \frac{1}{n_{\text{nets}}} \sum_{\alpha=1}^{n_{\text{nets}}} f_\alpha(x_1) \cdots f_\alpha(x_n)$$

$$\Delta G^{(n)}_{\text{exp}} := \bar{G}^{(n)}_{\text{exp}} - G^{(n)}_0, \qquad m_n := \frac{\Delta G^{(n)}_{\text{exp}}}{G^{(n)}_0}$$

- $x^{(1)}, \ldots, x^{(6)} \in \{-0.01, -0.006, -0.002, 0.002, 0.006, 0.01\}$
  $\rightarrow$ evaluate Green functions for all inequivalent combinations

Effective action

- numerical results

$$\forall N: \quad m_2 \approx 0, \qquad \forall n \geq 2: \quad m_{2n} = O\left(\frac{1}{N}\right)$$

- agreement with N-scaling formulas [2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

$$G_c^{(2n)} = O\left(\frac{1}{N^{n-1}}\right)$$

- extract single number $\langle |m_n| \rangle$: average $|m_n(x_1, \ldots, x_n)|$ over all combinations of points
- compare with background: standard deviation of $G_{\text{exp}}^{(n)}$ over all bags, then average over all combinations of points
  (compare statistical deviation and deviation from free result)

Effective action

- numerical results

$$\forall N: \quad m_2 \approx 0, \qquad \forall n \geq 2: \quad m_{2n} = O\left(\frac{1}{N}\right)$$

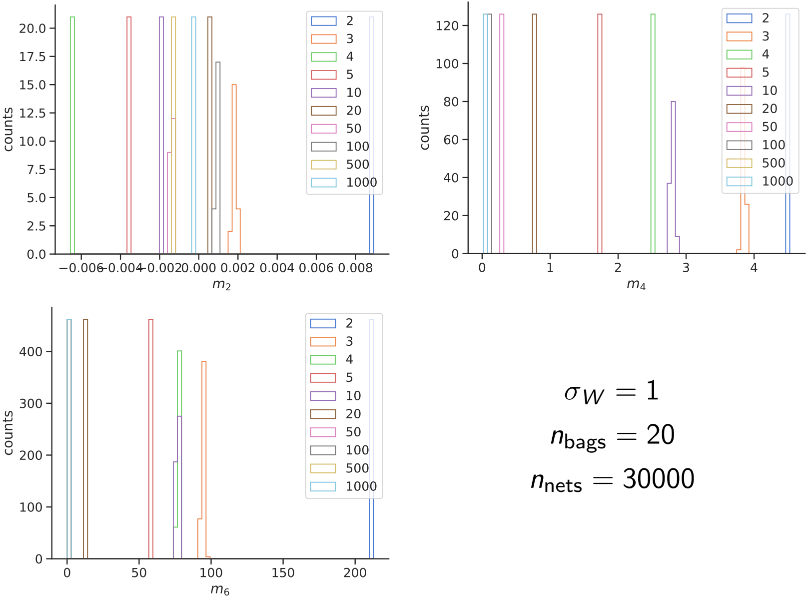
- agreement with N-scaling formulas [2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

$$G_c^{(2n)} = O\left(\frac{1}{N^{n-1}}\right)$$

- extract single number $\langle |m_n| \rangle$: average $|m_n(x_1, \dots, x_n)|$ over all combinations of points
- compare with background: standard deviation of $G_{\text{exp}}^{(n)}$ over all bags, then average over all combinations of points
  (compare statistical deviation and deviation from free result)
- compute 1PI action with quartic and sextic interactions:

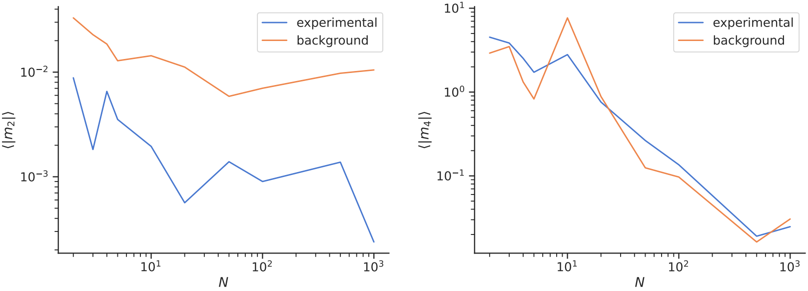
$$\Gamma = S_0 + \frac{u_4}{4!} \int \mathrm{d}^{d_{\text{in}}} x\, f(x)^4 + \frac{u_6}{6!} \int \mathrm{d}^{d_{\text{in}}} x\, f(x)^6$$

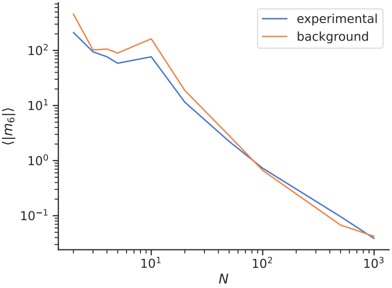
Green function deviations: histogram



$$\sigma_W = 1$$
$$n_{\text{bags}} = 20$$
$$n_{\text{nets}} = 30000$$

Green function deviations: mean values



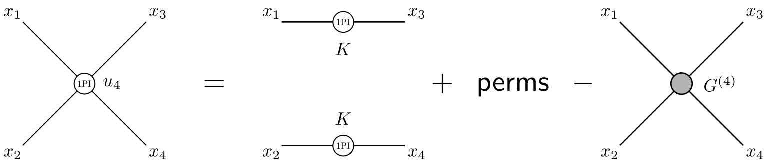


$$\sigma_W = 1$$
$$n_{\text{bags}} = 20$$
$$n_{\text{nets}} = 30000$$

Extract quartic coupling

[2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

- 4-point Feynman diagrams (1PI $\to$ no loops)

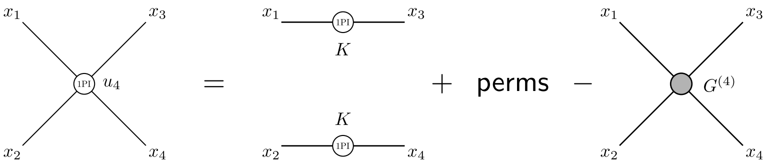


([2008.08601, HMS] works with microscopic action, studies only $|u_4|$)

Extract quartic coupling

[2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

- 4-point Feynman diagrams (1PI → no loops)



- measure u_4 from $G^{(4)}_{\text{exp}}$

$$u_4(x_1, x_2, x_3, x_4) = -\frac{\Delta G^{(4)}_{\text{exp}}(x_1, x_2, x_3, x_4)}{N_K(x_1, x_2, x_3, x_4)}$$

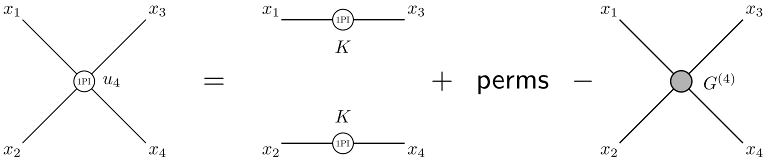
$$N_K := \int \mathrm{d}^{d_{\text{in}}} x \, K_W(x, x_1) K_W(x, x_2) K_W(x, x_3) K_W(x, x_4)$$

([2008.08601, HMS] works with microscopic action, studies only $|u_4|$)

Extract quartic coupling

[2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]

- 4-point Feynman diagrams (1PI $\to$ no loops)



- measure u_4 from $G^{(4)}_{\text{exp}}$

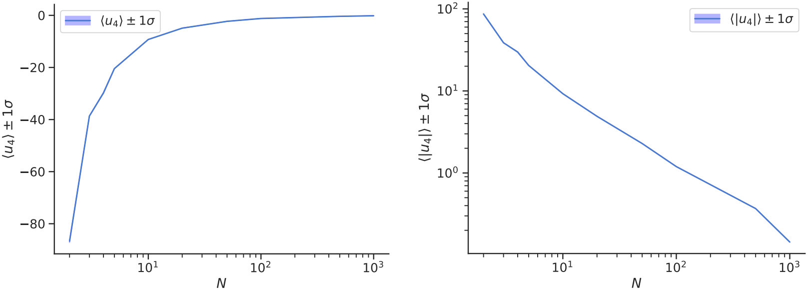
$$u_4(x_1, x_2, x_3, x_4) = -\frac{\Delta G^{(4)}_{\text{exp}}(x_1, x_2, x_3, x_4)}{N_K(x_1, x_2, x_3, x_4)}$$

$$N_K := \int \mathrm{d}^{d_{\text{in}}} x \, K_W(x, x_1) K_W(x, x_2) K_W(x, x_3) K_W(x, x_4)$$

- result: $u_4 \approx$ constant < 0
  $\to$ need $u_6 > 0$ for path integral stability

([2008.08601, HMS] works with microscopic action, studies only $|u_4|$)

Quartic coupling



$$\sigma_W = 1, \qquad n_{\text{bags}} = 30, \qquad n_{\text{nets}} = 30000$$

Outline: 3. Renormalization group in NN-QFT

Non-perturbative RG

- partition function and microscopic action

$$Z[j] := \mathrm{e}^{W[j]} := \int \mathrm{d}\phi \, \mathrm{e}^{-S[\phi] - j\cdot\phi}$$

$S[\phi]$ encodes microscopic (UV) physics

Non-perturbative RG

- partition function and microscopic action

$$Z[j] := \mathrm{e}^{W[j]} := \int \mathrm{d}\phi \, \mathrm{e}^{-S[\phi]-j\cdot\phi}$$

  $S[\phi]$ encodes microscopic (UV) physics

- classical field and 1PI effective action

$$\varphi(x) := \frac{\delta W}{\delta j}, \qquad \Gamma[\varphi] := j\cdot\varphi - W[j]$$

  $\Gamma[\varphi]$ encodes effective (IR) physics

Non-perturbative RG

- partition function and microscopic action

$$Z[j] := \mathrm{e}^{W[j]} := \int \mathrm{d}\phi\, \mathrm{e}^{-S[\phi]-j\cdot\phi}$$

  $S[\phi]$ encodes microscopic (UV) physics

- classical field and 1PI effective action

$$\varphi(x) := \frac{\delta W}{\delta j}, \qquad \Gamma[\varphi] := j\cdot\varphi - W[j]$$

  $\Gamma[\varphi]$ encodes effective (IR) physics

- renormalization group (RG) flow:
  - organize theory according to length scales
  - integrate degrees of freedom (dof) step by step
    $\rightarrow$ flow in the theory space
  - connect UV to IR

- review: [cond-mat/0702365, Delamotte]

Wilson RG: momentum-shell integration

- split field in slow and fast modes with respect to scale k

$$\phi(p) = \phi_<(p) + \phi_>(p), \qquad \begin{cases} \phi_<(p) := \theta(|p| < k)\, \phi(p) \\ \phi_>(p) := \theta(|p| \geq k)\, \phi(p) \end{cases}$$

- kinetic operator decomposes

$$\Xi(p) = \Xi_<(p) + \Xi_>(p), \qquad \begin{cases} \Xi_<(p) := \theta(|p| < k)\, \Xi(p) \\ \Xi_>(p) := \theta(|p| \geq k)\, \Xi(p) \end{cases}$$

Wilson RG: momentum-shell integration

- split field in slow and fast modes with respect to scale k

$$\phi(p) = \phi_<(p) + \phi_>(p), \qquad \begin{cases} \phi_<(p) := \theta(|p| < k)\,\phi(p) \\ \phi_>(p) := \theta(|p| \geq k)\,\phi(p) \end{cases}$$

- kinetic operator decomposes

$$\Xi(p) = \Xi_<(p) + \Xi_>(p), \qquad \begin{cases} \Xi_<(p) := \theta(|p| < k)\,\Xi(p) \\ \Xi_>(p) := \theta(|p| \geq k)\,\Xi(p) \end{cases}$$

- Wilsonian effective action for $\phi_<$

$$S_{\text{eff}}[\phi_<] := \frac{1}{2}\,\phi_< \cdot \Xi_< \cdot \phi_< + S_{\text{eff,int}}[\phi_<]$$
$$\mathrm{e}^{-S_{\text{eff,int}}[\phi_<]} := \int \mathrm{d}\phi_>\,\mathrm{e}^{-\frac{1}{2}\phi_> \cdot \Xi_> \cdot \phi_> - S_{\text{int}}[\phi_< + \phi_>]}$$

$\phi_<$ background, $\phi_>$ fluctuations

Wilson–Polchinski RG

- hard cutoff not convenient, use smooth regulator

$$\Xi_k(p) := R_k(p)\,\Xi(p), \qquad R_k(p) \to \begin{cases} 1 & p \ll k \\ 0 & p \gg k \end{cases}$$

- measure factorization $\Rightarrow$ field decomposition

$$\phi(p) = \chi(p) + \Phi(p)$$

$$\int \mathrm{d}\phi\, \mathrm{e}^{-\frac{1}{2}\phi\cdot\Xi\cdot\phi} = \left(\int \mathrm{d}\chi\, \mathrm{e}^{-\frac{1}{2}\chi\cdot\Xi_k\cdot\chi}\right) \times \left(\int \mathrm{d}\Phi\, \mathrm{e}^{-\frac{1}{2}\Phi\cdot(\Xi-\Xi_k)\cdot\Phi}\right)$$

Wilson–Polchinski RG

- hard cutoff not convenient, use smooth regulator

$$\Xi_k(p) := R_k(p)\,\Xi(p), \qquad R_k(p) \to \begin{cases} 1 & p \ll k \\ 0 & p \gg k \end{cases}$$

- measure factorization $\Rightarrow$ field decomposition

$$\phi(p) = \chi(p) + \Phi(p)$$

$$\int \mathrm{d}\phi\, \mathrm{e}^{-\frac{1}{2}\phi\cdot\Xi\cdot\phi} = \left(\int \mathrm{d}\chi\, \mathrm{e}^{-\frac{1}{2}\chi\cdot\Xi_k\cdot\chi}\right) \times \left(\int \mathrm{d}\Phi\, \mathrm{e}^{-\frac{1}{2}\Phi\cdot(\Xi-\Xi_k)\cdot\Phi}\right)$$

- effective action at scale k (UV cut-off for χ)

$$\mathrm{e}^{-S_{\mathrm{int},k}[\chi]} := \int \mathrm{d}\Phi\, \mathrm{e}^{-\frac{1}{2}\Phi\cdot(\Xi-\Xi_k)\cdot\Phi - S_{\mathrm{int}}[\chi+\Phi]}$$

Wilson–Polchinski RG

- hard cutoff not convenient, use smooth regulator

$$\Xi_k(p) := R_k(p)\,\Xi(p), \qquad R_k(p) \to \begin{cases} 1 & p \ll k \\ 0 & p \gg k \end{cases}$$

- measure factorization ⇒ field decomposition

$$\phi(p) = \chi(p) + \Phi(p)$$

$$\int \mathrm{d}\phi\, \mathrm{e}^{-\frac{1}{2}\phi\cdot\Xi\cdot\phi} = \left(\int \mathrm{d}\chi\, \mathrm{e}^{-\frac{1}{2}\chi\cdot\Xi_k\cdot\chi}\right) \times \left(\int \mathrm{d}\Phi\, \mathrm{e}^{-\frac{1}{2}\Phi\cdot(\Xi-\Xi_k)\cdot\Phi}\right)$$

- effective action at scale k (UV cut-off for χ)

$$\mathrm{e}^{-S_{\text{int},k}[\chi]} := \int \mathrm{d}\Phi\, \mathrm{e}^{-\frac{1}{2}\Phi\cdot(\Xi-\Xi_k)\cdot\Phi - S_{\text{int}}[\chi+\Phi]}$$

- Polchinski equation

$$k\,\frac{\mathrm{d}S_{\text{int},k}}{\mathrm{d}k} = \int \frac{\mathrm{d}^d p}{(2\pi)^d}\; k\,\frac{\mathrm{d}\Xi_k(p)}{\mathrm{d}k} \left[\frac{\delta^2 S_{\text{int},k}}{\delta\chi(p)\delta\chi(-p)} - \frac{\delta S_{\text{int},k}}{\delta\chi(p)}\,\frac{\delta S_{\text{int},k}}{\delta\chi(-p)}\right]$$

Wetterich formalism

- non-perturbative truncation with Polchinski equation difficult
  $\rightarrow$ Wetterich formalism
- regularize path integral

$$Z_k[j] := \mathrm{e}^{W_k[j]} := \int \mathrm{d}\phi \, \mathrm{e}^{-S[\phi] - \frac{1}{2}\phi \cdot R_k \cdot \phi - j \cdot \phi}$$

- R_k cutoff function s.t. $W_{k=\infty} = S$, $W_{k=0} = W$

$$R_{k=\infty}(p) = \infty, \qquad R_{k=0}(p) = 0, \qquad R_k(|p| > k) \approx 0$$

Wetterich formalism

- non-perturbative truncation with Polchinski equation difficult
  $\rightarrow$ Wetterich formalism
- regularize path integral

$$Z_k[j] := \mathrm{e}^{W_k[j]} := \int \mathrm{d}\phi \, \mathrm{e}^{-S[\phi] - \frac{1}{2}\phi \cdot R_k \cdot \phi - j \cdot \phi}$$

- R_k cutoff function s.t. $W_{k=\infty} = S$, $W_{k=0} = W$

$$R_{k=\infty}(p) = \infty, \qquad R_{k=0}(p) = 0, \qquad R_k(|p| > k) \approx 0$$

- effective average action action at scale k (IR cutoff for φ)

$$\varphi(x) := \frac{\delta W_k}{\delta j}, \qquad \Gamma_k[\varphi] := j \cdot \varphi - W_k[j] - \frac{1}{2}\varphi \cdot R_k \cdot \varphi$$

- Legendre transform requires correction to satisfy:

$$\Gamma_{k=0}[\varphi] = \Gamma[\varphi], \qquad \Gamma_{k=\infty}[\varphi] = S[\varphi]$$

Wetterich equation

- Wetterich equation

$$\frac{\mathrm{d}\Gamma_k}{\mathrm{d}k} = \frac{1}{2}\frac{\mathrm{d}R_k}{\mathrm{d}k}\,\mathrm{tr}\left(\Gamma_k'' + R_k\right)^{-1}$$

  Γ_k'' second derivatives of Γ w.r.t. φ

- solving requires approximation
  - restrict theory space to finite-dimensional subspace
  - derivative / local potential expansion

- non-perturbative formalism, finite coupling constants

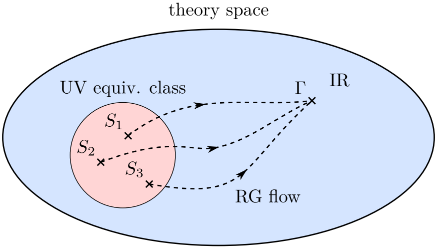
- large N expansion: keeping up to $\phi^{2n} \leftrightarrow O(1/N^{n-1})$ effects

RG for NN-QFT

- machine learning: find patterns in large dataset, ignoring noise
  $\rightarrow$ similar to RG flow

RG for NN-QFT

- machine learning: find patterns in large dataset, ignoring noise
  $\rightarrow$ similar to RG flow
- action: effective (IR) known, microscopic (UV) unknown
  - opposite as usual, need to reverse flow
  - since information is lost, no 1-to-1 map UV / IR
  - but any microscopic theory in IR universality class is fine



(Note: [2008.08601, Halverson-Maiti-Stoner] defines RG flow w.r.t. IR cutoff)

Momentum space 2-point function

- momentum space propagator

$$K(p) = (\sigma_W^2)^{1-\frac{d_{\text{in}}}{2}} \left(\frac{d_{\text{in}}}{2\pi}\right)^{\frac{d_{\text{in}}}{2}} \exp\left[-\frac{d_{\text{in}}}{2\sigma_W^2}\, p^2\right]$$

- momentum expansion (derivatives subleading in IR, $|p| \to 0$)

$$K(p) \approx \frac{Z_0^{-1}}{m_0^2 + p^2 + O(p^2)}, \qquad m_0^2 := \frac{2\sigma_W^2}{d_{\text{in}}}$$

$\rightarrow$ can be used in deep IR

- typical mass scale $\rightarrow$ correlation length $\xi := m_0^{-1}$

Momentum space 2-point function

- momentum space propagator

$$K(p) = (\sigma_W^2)^{1-\frac{d_{\text{in}}}{2}} \left(\frac{d_{\text{in}}}{2\pi}\right)^{\frac{d_{\text{in}}}{2}} \exp\left[-\frac{d_{\text{in}}}{2\sigma_W^2}\, p^2\right]$$

- momentum expansion (derivatives subleading in IR, $|p| \to 0$)

$$K(p) \approx \frac{Z_0^{-1}}{m_0^2 + p^2 + O(p^2)}, \qquad m_0^2 := \frac{2\sigma_W^2}{d_{\text{in}}}$$

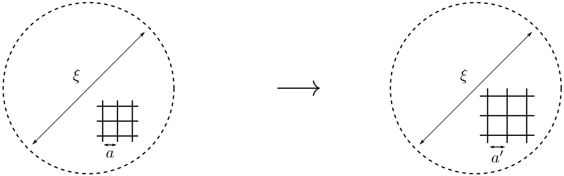
$\to$ can be used in deep IR

- typical mass scale $\to$ correlation length $\xi := m_0^{-1}$
- two possible RG scales: a_0^{-1} (machine precision) and m_0
- effective action: kinetic term + local potential

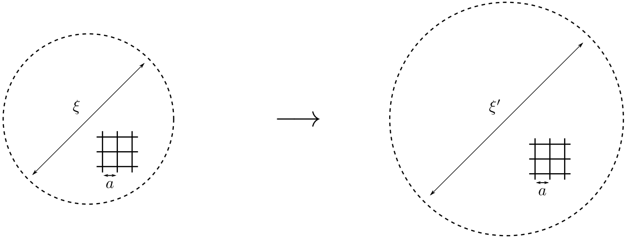
$$\Gamma_k = \Gamma_{k,0} + \frac{u_4(k)}{4!} \int \mathrm{d}^{d_{\text{in}}} x\, \varphi(x)^4 + \frac{u_6(k)}{6!} \int \mathrm{d}^{d_{\text{in}}} x\, \varphi(x)^6$$

Passive / active RG

- passive RG: keep $m_0 = \xi^{-1}$ fixed, vary $k = a^{-1} \leq a_0^{-1}$
  (keep neural network fixed, vary data)



- active RG: keep a_0 fixed, vary $k = m \geq m_0$
  (keep data fixed, vary neural network)



Passive RG: deep IR

- 2-derivative approximation

$$\Gamma_{k,0} = \frac{1}{2}\int \frac{\mathrm{d}^{d_{\text{in}}}p}{(2\pi)^{d_{\text{in}}}}\,\varphi(-p)(p^2 + m(k)^2)\varphi(p)$$

- flow equations

$$k\,\frac{\mathrm{d}\bar{u}_2}{\mathrm{d}k} = -2\,\bar{u}_2 - \frac{K_{d_{\text{in}}}\bar{u}_4}{(1+\bar{u}_2)^2}$$

$$k\,\frac{\mathrm{d}\bar{u}_4}{\mathrm{d}k} = -(4-d_{\text{in}})\,\bar{u}_4 - \frac{K_{d_{\text{in}}}\bar{u}_6}{(1+\bar{u}_2)^2} + \frac{6K_{d_{\text{in}}}\bar{u}_4^2}{(1+\bar{u}_2)^3}$$

$$k\,\frac{\mathrm{d}\bar{u}_6}{\mathrm{d}k} = -(6-2d_{\text{in}})\,\bar{u}_6 + \frac{30K_{d_{\text{in}}}\bar{u}_4\bar{u}_6}{(1+\bar{u}_2)^3} - \frac{90K_{d_{\text{in}}}\bar{u}_4^3}{(1+\bar{u}_2)^4}$$

where

$$\bar{u}_{2n} := k^{(n-1)d_{\text{in}}-2n}u_{2n}, \qquad u_2 := m^2$$

$$K_{d_{\text{in}}} := \frac{1}{(2\pi)^{d_{\text{in}}}}\,\frac{\pi^{d_{\text{in}}/2}}{\Gamma(d_{\text{in}}/2+1)}$$

- can also study deep UV (need momentum-dependent vertices)

Active RG

- propagator looks like zero-momentum propagator with UV regulator with scale k

$$K_k(p) := \frac{\mathrm{e}^{-p^2/k^2}}{k^2}, \qquad k^2 := \frac{2\sigma_W^2}{d_{\text{in}}}$$

- changing $\sigma_W \approx$ changing UV cutoff k
  $\rightarrow$ define running scale

Active RG

- propagator looks like zero-momentum propagator with UV regulator with scale k

$$K_k(p) := \frac{\mathrm{e}^{-p^2/k^2}}{k^2}, \qquad k^2 := \frac{2\sigma_W^2}{d_{\mathrm{in}}}$$

- changing $\sigma_W \approx$ changing UV cutoff k
  $\rightarrow$ define running scale
- classical action with K_k satisfies Polchinski equation
  but should be the effective propagator $\Rightarrow$ define

$$\Gamma_k''(p) + R_k(p) := k^2\,\mathrm{e}^{p^2/k^2}$$

Active RG

- propagator looks like zero-momentum propagator with UV regulator with scale k

$$K_k(p) := \frac{\mathrm{e}^{-p^2/k^2}}{k^2}, \qquad k^2 := \frac{2\sigma_W^2}{d_{\text{in}}}$$

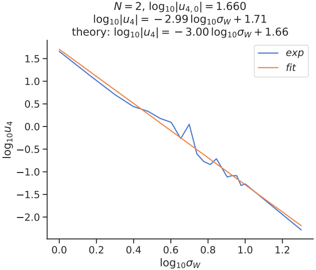
- changing $\sigma_W \approx$ changing UV cutoff k
  $\to$ define running scale
- classical action with K_k satisfies Polchinski equation but should be the effective propagator $\Rightarrow$ define

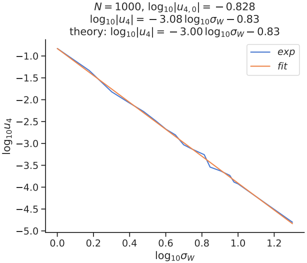
$$\Gamma_k''(p) + R_k(p) := k^2\, \mathrm{e}^{p^2/k^2}$$

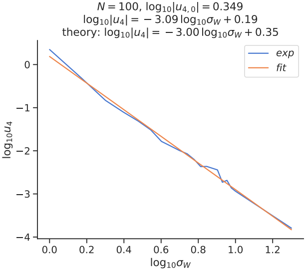
- flow equations

$$\sigma_W \frac{\mathrm{d}u_4}{\mathrm{d}\sigma_W} = (4 - d_{\text{in}})\, u_4, \qquad \sigma_W \frac{\mathrm{d}u_6}{\mathrm{d}\sigma_W} = (6 - 2d_{\text{in}})\, u_6$$

Results: active RG







$\sigma_W \in \{1.0, 1.5, \dots, 10, 20\}$

$n_{\text{bags}} = 30, \quad n_{\text{nets}} = 30000$

Outline: 4. Conclusion

Conclusion and outlook

Achievements:

- additional checks of the NN-QFT correspondence
- map of the possible theory space
- change in standard deviation = RG flow
- numerical tests of the equations

Conclusion and outlook

Achievements:

- additional checks of the NN-QFT correspondence
- map of the possible theory space
- change in standard deviation = RG flow
- numerical tests of the equations

Future directions:

- increase d_{in}, d_{out}, and order in N expansion; large d_{in} limit
- increase number of hidden layers
- extend to non-translation invariant kernels (ReLU. . .)
  - 2PI formalism [2102.13628, Blaizot-Pawlowski-Reinosa]
  - field redefinitions for non-local theories [2111.03672, HE-Fırat-Zwiebach]
- study evolution of QFT under training