Non-perturbative renormalization for the neural network-QFT correspondence

Harold Erbin

MIT & Iaifi (USA) & CEA-List (France)

In collaboration with:

− Vincent Lahoche (CEA-List)
− Dine Ousmane Samary (Université d’Abomey-Calavi)

arXiv: 2108.01403
Motivations

NN-QFT correspondence

Renormalization group in NN-QFT

Conclusion
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)
  - 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?

\[
\begin{align*}
    x_{i_0}^{(0)} & := x_{i_0} \\
    x_{i_1}^{(1)} & = g^{(1)} \left( W_{i_1i_0}^{(1)} x_{i_0}^{(0)} \right) \\
    f_{i_2}(x_{i_0}) & := x_{i_2}^{(2)} = g^{(2)} \left( W_{i_2i_1}^{(2)} x_{i_1}^{(1)} \right)
\end{align*}
\]

\[i_0 = 1, 2, 3; \quad i_1 = 1, \ldots, 4; \quad i_2 = 1, 2\]
\[K = 1; \quad d_{in} = 3; \quad d_{out} = 2; \quad N^{(1)} = 4\]

- input \(x^{(0)} := x \in \mathbb{R}^{d_{in}}\)
- \(K \geq 1\) hidden layers, \(n \in \{1, \ldots, K\}\)
  - 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 \(x^{(K+1)} := f(x) \in \mathbb{R}^{d_{out}}\)
Why a QFT?

Problems with neural networks:

- black box: very hard to understand the meaning of NN computation
- loss landscape problem: loss function non-convex and very rough, hard to find (global) minimum (related to spin glass) \([1412.0233, \text{Choromanska et al.}; 1712.09913, \text{Li et al.}]\)
- training can be complicated (expensive computationally, convergence issues. . .)
- hyperparameter tuning (find architecture / best parameters): mostly trial and errors or random optimization
Why a QFT?

Problems with neural networks:

▶ black box: very hard to understand the meaning of NN computation

▶ loss landscape problem: 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.]

▶ training can be complicated (expensive computationally, convergence issues. . . )

▶ hyperparameter tuning (find architecture / best parameters): mostly trial and errors or random optimization

→ develop tools to improve analytical understanding of neural network building and training
Plan

**NN-QFT correspondence**

For a very general class of architectures, it is possible to associate a Wilsonian effective action (QFT) to a neural network (NN).

[2008.08601, Halverson-Maiti-Stoner (HMS)] (see also [2106.00694, HMS; 2106.10165, Roberts-Yaida-Hanin; 2109.13247, Grosnevor-Jefferson...])

In this talk:

▶ 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.
Plan

**NN-QFT correspondence**

For a very general class of architectures, it is possible to associate a Wilsonian effective action (QFT) to a neural network (NN).

\[2008.08601, \text{Halverson-Maiti-Stoner (HMS)}\] (see also \[2106.00694, \text{HMS}; 2106.10165, \text{Roberts-Yaida-Hanin}; 2109.13247, \text{Grosnevor-Jefferson}…\])

In this talk \[2108.01403, \text{HE-Lahoche-Samary}\]:

- 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

Motivations

NN-QFT correspondence

Renormalization group in NN-QFT

Conclusion
Neural network

- fully connected neural network (one hidden layer)

\[ f_{\theta,N} : \mathbb{R}^{d_{in}} \rightarrow \mathbb{R}^{d_{out}} \]

\[ f_{\theta,N}(x) = W_1 \left( g(W_0 x + b_0) \right) + b_1 \]

width \( N \), activation function \( g \)
parameters (weights and biases) \( \theta = (W_0, b_0, W_1, b_1) \)

\( W_0 \sim \mathcal{N}(0, \sigma_W^2 / d_{in}) \), \( W_1 \sim \mathcal{N}(0, \sigma_W^2 / N) \)

\( b_0, b_1 \sim \mathcal{N}(0, \sigma_b^2) \)
Neural network

- fully connected neural network (one hidden layer)

\[
f_{\theta,N} : \mathbb{R}^{d_{\text{in}}} \rightarrow \mathbb{R}^{d_{\text{out}}}
\]

\[
f_{\theta,N}(x) = W_1 \left( g(W_0 x + b_0) \right) + b_1
\]

width \( N \), activation function \( g \)
parameters (weights and biases) \( \theta = (W_0, b_0, W_1, b_1) \)

\[
W_0 \sim \mathcal{N}(0, \sigma_W^2/d_{\text{in}}), \quad W_1 \sim \mathcal{N}(0, \sigma_W^2/N)
\]

\[
b_0, b_1 \sim \mathcal{N}(0, \sigma_b^2)
\]

- change of perspective (compared to applied ML)
  - consider statistical ensemble of neural networks defined by distribution in parameter space
  - specific NN = sample from distribution
  - dual description: parameter dist. + architecture induces distribution in function space
  - 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) \cite{Neal96}
- generalize to most architectures \cite{1910.12478, Yang} and training

\[
S_0[f] = \frac{1}{2} \int d\mathbf{x} d\mathbf{x}' f(\mathbf{x}) \Xi(\mathbf{x}, \mathbf{x}') f(\mathbf{x}') \\
\Xi := K^{-1}
\]

\[
G_n(0)(x_1, \ldots, x_n) := \int df e^{-S_0[f]} f(x_1) \cdots f(x_n)
\]
Large $N$ limit, Gaussian process and free QFT

Large $N$ limit $\Rightarrow$ infinite layer width:

- NN (function) distribution drawn from Gaussian process (GP) with kernel $K$ (consequence of central limit theorem) [Neal '96]
- generalize to most architectures [1910.12478, Yang] and training
- log likelihood

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

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

$$G_0^{(n)}(x_1, \ldots, x_n) := \int df \, e^{-S_0[f]} f(x_1) \cdots f(x_n)$$
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) \cite{Neal1996}
- generalize to most architectures \cite{1910.12478, Yang} and training
- log likelihood

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

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

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

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

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

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

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

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

- free action $S'[f]$ 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
  \[ S[f] = S'_0[f] + S_{\text{int}}[f] \]

- free action $S'_0[f]$ unknown

- $n$-point Green functions
  \[ G^{(n)}(x_1, \ldots, x_n) := \int df \ e^{-S[f]} \ f(x_1) \cdots f(x_n) \]

- effective (IR) 2-point function exactly known ($G^{(2)}$ $N$-indep.)
  \[ G^{(2)}(x, y) = K(x, y) = G_0^{(2)}(x, y) \]

- $N$-scaling [2008.08601, HMS; 2108.01403, HE-Lahoche-Samary]
  \[ G_c^{(2n)} = O \left( \frac{1}{N^{n-1}} \right) \]
## Summary of NN-QFT correspondence

<table>
<thead>
<tr>
<th>QFT</th>
<th>NN / GP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>spacetime points</td>
</tr>
<tr>
<td>$p$</td>
<td>momentum space</td>
</tr>
<tr>
<td>$f(x)$</td>
<td>field</td>
</tr>
<tr>
<td>$K(x, y)$</td>
<td>propagator</td>
</tr>
<tr>
<td>$S_0$</td>
<td>free action</td>
</tr>
<tr>
<td>interactions</td>
<td></td>
</tr>
</tbody>
</table>
Theory space

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

- data-space ≠ spacetime → 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 ≠ spacetime** → avoid bias from particle physics
- **spacetime = x ∈ ℝ^{d_{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 → non-local interactions?
  - symmetries of inputs and outputs?
  - natural UV cutoff: machine precision
Theory space

- **data-space ≠ spacetime** $\rightarrow$ avoid bias from particle physics
- **spacetime** $= x \in \mathbb{R}^{d_{in}} +$ 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**

\[ S_{\text{int}} = \sum_n g_n \int d^{d_{\text{in}}} x f(x)^n \]

- **non-local interactions and coupling functions**

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

- **delocalized fields**

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

Good UV behavior (e.g. string theory [1604.01783, Pius-Sen])

- **tensor models: break permutation invariance**

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

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 $d_{out} = 1$
▶ translation-invariant activation function

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

(stricly speaking, activation func. + normalization)

▶ GP kernel [2008.08601, HMS]

\[
k(x, y) := \sigma_b^2 + K_W(x, y), \quad K_W(x, y) = \sigma_W^2 e^{-\frac{\sigma_W^2}{2d_{in}} |x-y|^2}
\]

▶ note: [2008.08601, HMS] also considers ReLU and Erf functions
Numerical setup

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

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

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

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

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

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

- local action with quartic and sextic interactions:

\[ S = S_0' + \frac{u_4}{4!} \int d^{d_{in}}x \, \phi(x)^4 + \frac{u_6}{6!} \int d^{d_{in}}x \, \phi(x)^6 \]

reminder: \( S_0' \) unknown
Effective action

- local action with quartic and sextic interactions:

\[ S = S_0' + \frac{u_4}{4!} \int d^{d_{\text{in}}} x \, \phi(x)^4 + \frac{u_6}{6!} \int d^{d_{\text{in}}} x \, \phi(x)^6 \]

reminder: \( S_0' \) unknown

- numerical results

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

- 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
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]

- Feynman diagrams (assume $u_6 \ll 1$)

\[
\begin{align*}
\int d^4x & \exp\left(u^4(x_1, x_2, x_3, x_4)\right) \\
& \approx \int d^4x K W(x_1, x_2) K W(x_2, x_3) K W(x_3, x_4) \\
& + \text{perms} - G^{(4)}(x_1, x_2, x_3, x_4)
\end{align*}
\]
Extract quartic coupling

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

- Feynman diagrams (assume $u_6 \ll 1$)

$$
\begin{align*}
\left( \begin{array}{c}
{x_1} \\
{x_2} \\
{u_4} \\
{x_3} \\
{x_4}
\end{array} \right) & \approx \\
\frac{K}{x_2} & + \text{perms} & - \\
\frac{K}{x_4}
\end{align*}
$$

- measure from $G_{exp}^{(4)}$

$$
\begin{align*}
u_4(x_1, x_2, x_3, x_4) & = - \frac{\Delta G_{exp}^{(4)}(x_1, x_2, x_3, x_4)}{N_K(x_1, x_2, x_3, x_4)} \\
N_K & := \int d^{d_{in}} x \ K_W(x, x_1)K_W(x, x_2)K_W(x, x_3)K_W(x, x_4)
\end{align*}
$$
Extract quartic coupling

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

▶ Feynman diagrams (assume $u_6 \ll 1$)

\[
\begin{align*}
  x_1 & \to x_3 \\
  x_2 & \to x_4 \\
  u_4 & \quad \approx \quad \begin{array}{c}
    x_1 \\
    \otimes \otimes \\
    x_3
  \end{array} \\
  K & \quad \quad \quad \quad \quad \quad + \text{ perms} - \\
  x_2 & \to x_4
\end{align*}
\]

▶ measure from $G^{(4)}_{\exp}$

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

\[
N_K := \int d^{d_{\text{in}}} x \ W(x, x_1) W(x, x_2) W(x, x_3) W(x, x_4)
\]

▶ result: $u_4 \approx \text{constant} < 0$

→ need $u_6 > 0$ (or higher-order even power) for path integral stability ([2008.08601, HMS] studies only $|u_4|$)
Quartic coupling

\[ \sigma_W = 1, \quad n_{\text{bags}} = 30, \quad n_{\text{nets}} = 30000 \]
Outline: 3. Renormalization group in NN-QFT

Motivations

NN-QFT correspondence

Renormalization group in NN-QFT

Conclusion
Non-perturbative RG

- partition function and microscopic action

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

\( S[\phi] \) encodes microscopic (UV) physics
Non-perturbative RG

- partition function and microscopic action

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

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

- classical field and 1PI effective action

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

\( \Gamma[\varphi] \) encodes effective (IR) physics
Non-perturbative RG

- partition function and microscopic action

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

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

- classical field and 1PI effective action

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

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

- renormalization group (RG) flow:
  - organize theory according to length scales
  - integrate degrees of freedom (dof) step by step
    → 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 $k$

\[
\phi(p) = \phi_<(p) + \phi_>(p),
\]

\[
\begin{cases}
\phi_<(p) := \theta(|p| < k) \phi(p) \\
\phi_>(p) := \theta(|p| \geq k) \phi(p)
\end{cases}
\]

- kinetic operator decomposes

\[
\Xi(p) = \Xi_<(p) + \Xi_>(p),
\]

\[
\begin{cases}
\Xi_<(p) := \theta(|p| < k) \Xi(p) \\
\Xi_>(p) := \theta(|p| \geq k) \Xi(p)
\end{cases}
\]
Wilson RG: momentum-shell integration

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

\[
\phi(p) = \phi_<(p) + \phi_>(p),
\]

\[
\begin{align*}
\phi_<(p) &:= \theta(|p| < k) \phi(p) \\
\phi_>(p) &:= \theta(|p| \geq k) \phi(p)
\end{align*}
\]

- kinetic operator decomposes

\[
\Xi(p) = \Xi_<(p) + \Xi_>(p),
\]

\[
\begin{align*}
\Xi_<(p) &:= \theta(|p| < k) \Xi(p) \\
\Xi_>(p) &:= \theta(|p| \geq k) \Xi(p)
\end{align*}
\]

- Wilsonian effective action for $\phi_<$

\[
S_{\text{eff}}[\phi_] := \frac{1}{2} \phi_\cdot \Xi_\cdot \phi_ + S_{\text{eff,int}}[\phi_]
\]

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

$\phi_<$ background, $\phi_>$ fluctuations
Wilson–Polchinski RG

- hard cutoff not convenient, use smooth regulator

\[ \Xi_k(p) := R_k(p) \Xi(p), \quad R_k(p) \rightarrow \begin{cases} 
1 & p \ll k \\
0 & p \gg k 
\end{cases} \]

- measure factorization \( \Rightarrow \) field decomposition

\[ \phi(p) = \chi(p) + \Phi(p) \]

\[ \int d\phi e^{-\frac{1}{2}\phi \cdot \Xi \cdot \phi} = \left( \int d\chi e^{-\frac{1}{2}\chi \cdot \Xi_k \cdot \chi} \right) \times \left( \int d\Phi e^{-\frac{1}{2}\Phi \cdot (\Xi - \Xi_k) \cdot \Phi} \right) \]
Wilson–Polchinski RG

- hard cutoff not convenient, use smooth regulator

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

- measure factorization \( \Rightarrow \) field decomposition

\[ \phi(p) = \chi(p) + \Phi(p) \]

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

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

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

- hard cutoff not convenient, use smooth regulator

\[ \Xi_k(p) := R_k(p) \Xi(p), \quad R_k(p) \rightarrow \begin{cases} 1 & p \ll k \\ 0 & p \gg k \end{cases} \]

- measure factorization \( \Rightarrow \) field decomposition

\[ \phi(p) = \chi(p) + \Phi(p) \]
\[ \int d\phi e^{-\frac{1}{2} \phi \cdot \Xi \cdot \phi} = \left( \int d\chi e^{-\frac{1}{2} \chi \cdot \Xi_k \cdot \chi} \right) \times \left( \int d\Phi e^{-\frac{1}{2} \Phi \cdot (\Xi - \Xi_k) \cdot \Phi} \right) \]

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

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

- Polchinski equation

\[ k \frac{dS_{int,k}}{dk} = \int \frac{d^d p}{(2\pi)^d} k \frac{d\Xi_k(p)}{dk} \left[ \frac{\delta^2 S_{int,k}}{\delta \chi(p) \delta \chi(-p)} - \frac{\delta S_{int,k}}{\delta \chi(p)} \frac{\delta S_{int,k}}{\delta \chi(-p)} \right] \]
Wetterich formalism

- non-perturbative truncation with Polchinski equation difficult
  → Wetterich formalism
- regularize path integral

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

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

\[
R_{k=\infty}(p) = \infty, \quad R_{k=0}(p) = 0, \quad R_k(|p| > k) \approx 0
\]
Wetterich formalism

- non-perturbative truncation with Polchinski equation difficult
  \(\rightarrow\) Wetterich formalism

- regularize path integral

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

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

\[
R_{k=\infty}(p) = \infty, \quad R_{k=0}(p) = 0, \quad R_k(|p| > k) \approx 0
\]

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

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

- Legendre transform requires correction to satisfy:

\[
\Gamma_{k=0}[\varphi] = \Gamma[\varphi], \quad \Gamma_{k=\infty}[\varphi] = S[\varphi]
\]
Wetterich equation

\[ \frac{d\Gamma_k}{dk} = \frac{1}{2} \frac{dR_k}{dk} \text{tr} (\Gamma_k'' + R_k)^{-1} \]

\( \Gamma_k'' \) second derivatives of \( \Gamma \) w.r.t. \( \varphi \)

- 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
  → similar to RG flow
RG for NN-QFT

- machine learning: find patterns in large dataset, ignoring noise → 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

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

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

\[
K(p) \approx \frac{Z_0^{-1}}{m_0^2 + p^2 + O(p^2)}, \quad m_0^2 := \frac{2\sigma_W^2}{d_{in}}
\]

→ can be used in deep IR

- typical mass scale → 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, \(|p| \to 0\))

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

→ can be used in deep IR

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

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

▶ 2-derivative approximation (deep IR)

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

▶ flow equations

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

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

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

where

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

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

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

\[ K_k(p) := \frac{\exp^{-p^2/k^2}}{k^2}, \quad k^2 := \frac{2\sigma_W^2}{d_{\text{in}}} \]

- 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 $k$

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

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

- classical action with $K_k$ satisfies Polchinski equation
  problem: should be the effective propagator ⇒ define

$$\Gamma_k''(p) + R_k(p) := k^2 \varepsilon^{p^2/k^2}$$
Active RG

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

\[ K_k(p) := \frac{e^{-p^2/k^2}}{k^2}, \quad k^2 := \frac{2\sigma_W^2}{d_{in}} \]

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

- classical action with $K_k$ satisfies Polchinski equation

problem: should be the effective propagator $\Rightarrow$ define

\[ \Gamma''_k(p) + R_k(p) := k^2 e^{p^2/k^2} \]

- flow equations

\[ \sigma_W \frac{du_4}{d\sigma_W} = 2(4 - d_{in}) u_4, \quad \sigma_W \frac{du_6}{d\sigma_W} = 2(6 - 2d_{in}) u_6 \]
Results: active RG

\( N = 2, \log_{10}|u_{4,0}| = 1.660 \)
\( \log_{10}|u_4| = -2.99 \log_{10} \sigma_W + 1.71 \)
theory: \( \log_{10}|u_4| = -3.00 \log_{10} \sigma_W + 1.66 \)

\( N = 1000, \log_{10}|u_{4,0}| = -0.828 \)
\( \log_{10}|u_4| = -3.08 \log_{10} \sigma_W - 0.83 \)
theory: \( \log_{10}|u_4| = -3.00 \log_{10} \sigma_W - 0.83 \)

\( \sigma_W \in \{1.0, 1.5, \ldots , 10, 20\} \)
\( n_{\text{bags}} = 30, \; n_{\text{nets}} = 30000 \)
Detour: string field theory

- active RG expected from string field theory
- standard deviation $\sigma_W \sim$ stub parameter $s_0$
- GaussNet QFT $= p$-adic string theory
  
  \[\text{[hep-th/0003278, Ghoshal-Sen; hep-th/0207107, Moeller-Zwiebach]}\]
Detour: string field theory

▶ active RG expected from string field theory
▶ standard deviation $\sigma_W \sim$ stub parameter $s_0$
▶ GaussNet QFT = $p$-adic string theory
  \[\text{[hep-th/0003278, Ghoshal-Sen; hep-th/0207107, Moeller-Zwiebach]}\]
▶ “adding stubs to string vertices / propagator” = vary region of moduli space corresponding to fundamental interactions
▶ changing stub length = Wilsonian RG
  \[\text{[Brustein-De Alwis '91; hep-th/0105272, Nakatsu; 1609.00459, Sen]}\]
▶ stub parameter responsible for good UV behavior in string theory \[1604.01783, \text{Pius-Sen}\]
Detour: string field theory

- active RG expected from string field theory
- standard deviation $\sigma_W \sim$ stub parameter $s_0$
- GaussNet QFT = $p$-adic string theory
  \[ \text{[hep-th/0003278, Ghoshal-Sen; hep-th/0207107, Moeller-Zwiebach]} \]
- “adding stubs to string vertices / propagator” = vary region of moduli space corresponding to fundamental interactions
- changing stub length = Wilsonian RG
  \[ \text{[Brustein-De Alwis '91; hep-th/0105272, Nakatsu; 1609.00459, Sen]} \]
- stub parameter responsible for good UV behavior in string theory \[1604.01783, \text{Pius-Sen}\]
- QFT with stubs \[2108.04312, \text{Chiaffrino-Sachs; HE-Godet, to appear}\]
- non-locality studied in \[\text{HE-Fırat-Zwiebach, to appear}\]
Outline: 4. Conclusion

Motivations

NN-QFT correspondence

Renormalization group in NN-QFT

Conclusion
Conclusion and outlook

Achievements:

- additional checks of the NN-QFT correspondence
- map of the possible theory space
- passive and active RG flow equations for neural networks
- 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
▶ passive and active RG flow equations for neural networks
▶ change in standard deviation = RG flow
▶ numerical tests of the equations

Future directions:
▶ increase $d_{\text{in}}, d_{\text{out}},$ and order in $N$ expansion
▶ investigate non-locality and random tensor models
▶ consider more general architectures
▶ extend to non-translation invariant kernels (ReLU…)
▶ numerical tests for passive RG flow