$\lambda\phi^4$ Scalar Neural Network Field Theory

Anindita Maiti

Machine Learning for Lattice Field Theory and Beyond
ECT* Trento

30-June-2023

Based on 2307.xxxxx w/ Halverson, Schwartz, Demirtas, Stoner

Neural Network Field Theories: Non-Gaussianity, Actions, and Locality

Mehmet Demirtas[1,2,i], James Halverson[1,2,ii], Anindita Maiti[1,2,4,iii], Matthew D. Schwartz[1,3,iv], and Keegan Stoner[1,2,v]

[1] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

[2] *Department of Physics, Northeastern University, Boston, MA 02115 USA*

[3] *Department of Physics, Harvard University, Cambridge, MA 02138 USA*

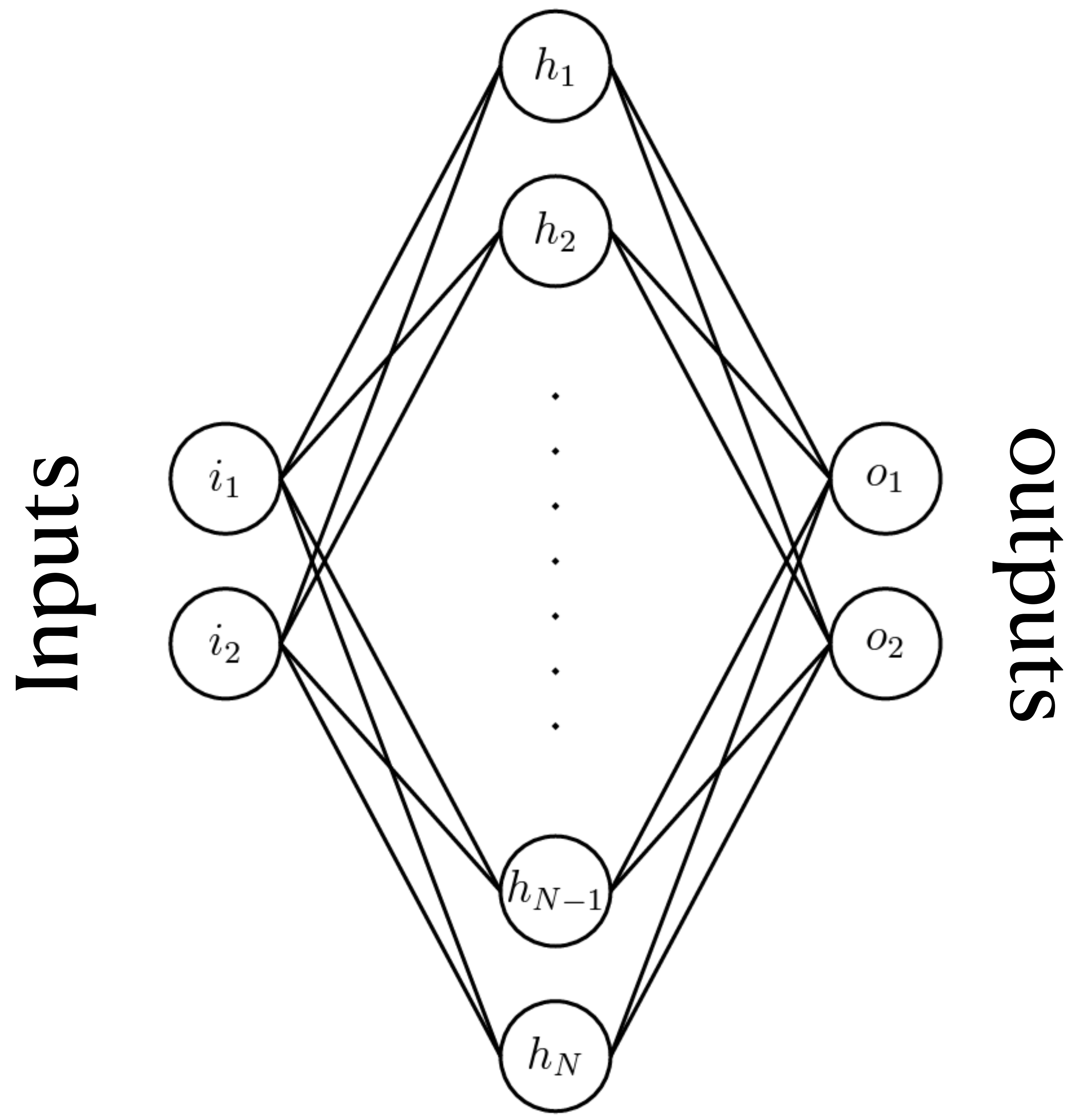
[4] *Department of Applied Mathematics, Harvard University, Cambridge, MA 02138 USA*

We study foundational aspects of field theories that are defined by neural networks. In certain large-N limits these theories are generalized free field theories due to the Central Limit Theorem. Interactions are induced by parameterically violating assumptions of the theorem, including infinite-N and statistical independence. The Edgeworth expansion yields a method for computing actions in terms of connected Feynman diagrams, whose vertices are the usual connected correlators. This general field theory result may be used to derive actions for neural network field theories. Specific interacting theories may be engineered by representing action deformations in the parameter space of the neural network, in which case interactions arise from the breaking of statistical independence of neural network parameters. Using this technique, ϕ^4 theory is realized as an infinite neural network field theory.

June 2023

[i] m.demirtas@northeastern.edu [ii] j.halverson@northeastern.edu [iii] amaiti@seas.harvard.edu [iv] schwartz@g.harvard.edu [v] stoner.ke@northeastern.edu

Introduction



$$\phi(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(x)$$

Notations:

1. Output $\phi(x)$ for input x.
2. Depends on architecture details, width N, parameters $P(h; \vec{\alpha})$.
3. Statistical independence at $\vec{\alpha} = 0$.

Initialize NN many times, do not train. Outputs are drawn from same statistical distribution, fixed by architecture.

Each initialized architecture corresponds to a distribution over fields or functions. Introduce $Z[\phi] = \int D\phi e^{-S[\phi]}$.

What is a Neural Network Field Theory?

It's a field theory defined by NN architecture.

Connections to this Workshop

Q1. Can we design Neural Network (NN) architectures, such that corresponding functional densities are the same as those of known field theories, e.g. scalar, fermionic, gauge?

Q2. Why care, though?

Answer for Q2: An alternative sampling strategy. One can start with such an architecture instead of using Monte Carlo.

Answer for Q1: We have one architecture for NN scalar field theory at initialization. No solutions for fermions, gauge yet.

Related works!

ML for Lattice Field Theory -

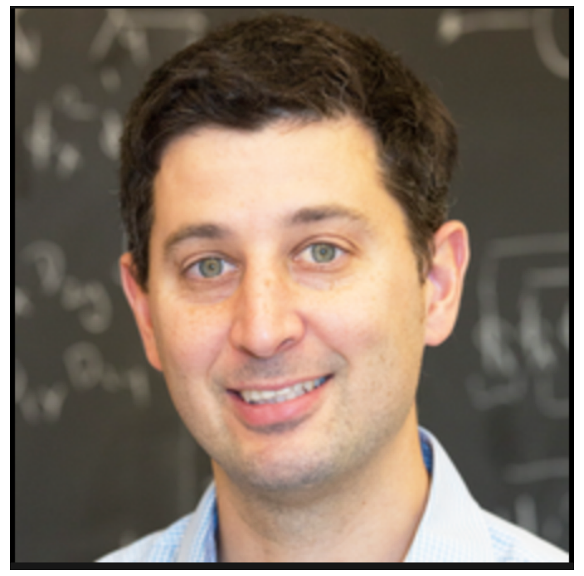
- [Albergo et al. 2019]
- [Hackett et al. 2021]
- [Bachtis et al. 2021]
- [de Haan et al. 2021]
- [Bachtis et al. 2022]
- [Abbott et al. 2022]
- [Gerdes at al. 2022]

Gaussian and non-Gaussian processes in NN -

- [Neal 1995]
- [Yaida 2019]
- [Yang 2019]
- [Halverson et al. 2020]
- [Maiti et al. 2021]
- [Naveh et al. 2021]
- [Jefferson et al. 2021]
- [Halverson 2022]
- [Roberts et.al. 2022]
- [Erbin et al. 2022]
- [Banta et al. 2023]

Collaborators!

Jim Halverson



Matt Schwartz

Mehmet Demirtas

Keegan Stoner

Q. How to engineer NN architectures for scalar field theory at initialization?

Step 1: Given some NN architecture, find principles to predict corresponding field action $S[\phi]$ of output ensembles.

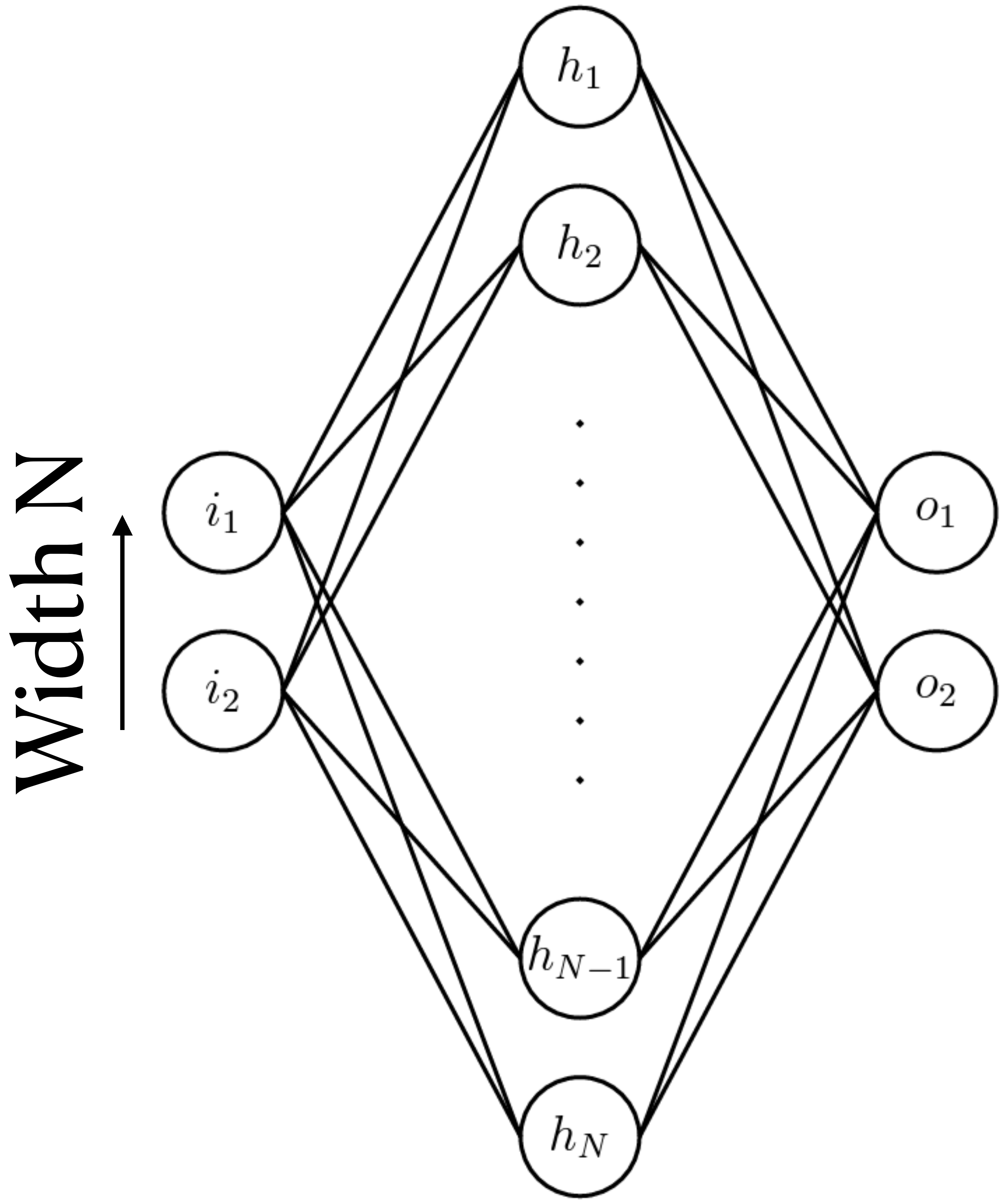
Step 2: Given some $S[\phi]$, find principles to design NN architectures to get this action from output ensembles at initialization.

Principles: Given an architecture, predict $S[\phi]$

Let's build up this principle!

First, free vs. interacting field theories in Neural Networks....

- At $\lim N \to \infty$, i.i.d. parameters $P(h; \vec{\alpha} = 0)$, each output is a sum over infinite independent variables.
- By Central Limit Theorem (CLT), each such output is a draw from some Gaussian distribution (NNGP).
- In field theory language, the outputs are drawn from a free field theory.



$$\phi(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(x)$$

Finite N, and dissimilar, correlated parameters break CLT assumptions.

Initialize NN to have large CLT breaking. Outputs are now draws from a non-perturbative field theory.

Newsflash: NN field theories are Euclidean or statistical. However, if all correlation functions of a NNFT satisfy Osterwalder-Schrader axioms, it's a *Quantum* Field Theory.

[Halverson 2021]

Next, there's a duality!

Observables, e.g. NN Correlation Functions

Evaluate in terms of NN architecture, parameters etc. Call it 'parameter space'.

$$G^{(n)}(x_1, \ldots, x_n) := \mathbb{E}[\phi(x_1) \ldots \phi(x_n)] = \int d\theta \, P(\theta) \, \phi(x_1) \ldots \phi(x_n)$$

Evaluate in terms of derivatives of $Z[\phi]$, if we know $S[\phi]$. Call it 'field space'.

$$G^{(n)}(x_1, \ldots, x_n) = \int D\phi \, e^{-S[\phi]} \, \phi(x_1) \ldots \phi(x_n)$$

- Initialize some NN with CLT-breaking.
- Action is unknown.
- Evaluate correlators in parameter space.

A few steps later...

We borrow a machinery from Statistical Physics, known as 'Edgeworth expansion'.

Using this, field density $P[\phi] = e^{-s[\phi]}$ has an alternative expression in parameter space.

$$P[\phi] = \frac{1}{Z} \exp\Big(\sum_{r=3}^{\infty} \frac{(-1)^r}{r!} \int \prod_{i=1}^{r} d^d x_i G_c^{(r)}(x_1, \cdots, x_r) \frac{\delta}{\delta\phi(x_1)} \cdots \frac{\delta}{\delta\phi(x_r)} \Big)$$
$$\times \exp\Big(-\frac{1}{2} \int d^d x_1 d^d x_2 \, \phi(x_1) G_c^{(2)}(x_1, x_2)^{-1} \phi(x_2) \Big),$$

$G_c^{(r)}$ is connected r-point function, computed in parameter space.

$$\int d^d y G_c^{(2)}(x, y)^{-1} G_c^{(2)}(y, z) = \delta^d(x - z)$$

Notice: $P[\phi] = e^{-S[\phi]} = \int dJ e^{W[J] - J\phi}$

Here $W[J] = \log Z[J]$ is generating functional for connected correlators or cumulants, called CGF.

If $P(h; \overrightarrow{\alpha} = 0)$ then $G_c^{(r)} \propto \frac{1}{N^{r/2-1}}$.

Otherwise, mixed N-scaling.

Let's explore another Duality!

Field picture:

$$e^{-S[\phi]} = P[\phi] = \frac{1}{Z}\exp\left(\sum_{r=3}^{\infty}\frac{(-1)^r}{r!}\int\prod_{i=1}^{r} d^d x_i G_c^{(r)}(x_1,\cdots,x_r)\frac{\delta}{\delta\phi(x_1)}\cdots\frac{\delta}{\delta\phi(x_r)}\right)e^{-S_{GP}[\phi]}$$

Source picture:

$$e^{W[J]} = Z[J] = c'\exp\left(\sum_{r=3}^{\infty}(-1)^r\int\prod_{i=1}^{r} d^d x_i\, g_r(x_1,\cdots,x_r)\frac{\delta}{\delta J(x_1)}\cdots\frac{\delta}{\delta J(x_r)}\right)e^{-S_0[J]}$$

Introduce Feynman rules to evaluate couplings $g_r(x_1,\cdots,x_r)$ in terms of $G_c^{(r)}$.

	Field Picture	Source Picture
Field	$\phi(x)$	$J(x)$
CGF	$W[J] = \log(Z[J])$	$S[\phi] = -\log(P[\phi])$
Cumulant	$G_c^{(r)}(x_1,\ldots,x_r)$	$g_r(x_1,\ldots,x_r)$

A Brief glimpse into

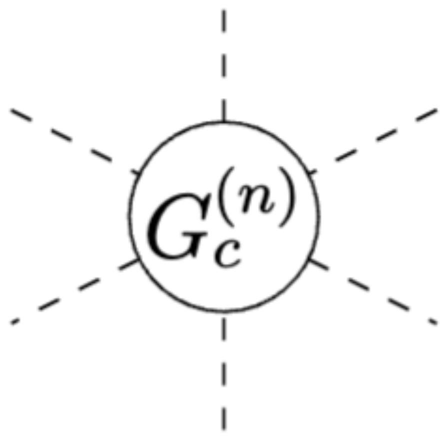
Feynman Rules for $g_r(x_1, \ldots, x_r)$.

1. Internal points associated to vertices are unlabelled, for diagrammatic simplicity. Propagators therefore connect to internal points in all possible ways.

2. For each propagator between z_i and y_j, where $z_i = x_i$ (internal) or $z_i = y_i$ (external),

$$z_i \text{ - - - - } y_j = G_c^{(2)}(z_i, y_j)^{-1}.$$

3. For each vertex,



$$= (-1)^n \int d^d y_1 \cdots d^d y_n \, G_c^{(n)}(y_1, \cdots, y_n).$$

4. Divide by symmetry factor and $(-1)^r$ factor of coupling $g_r(x_1, \cdots, x_r)$.

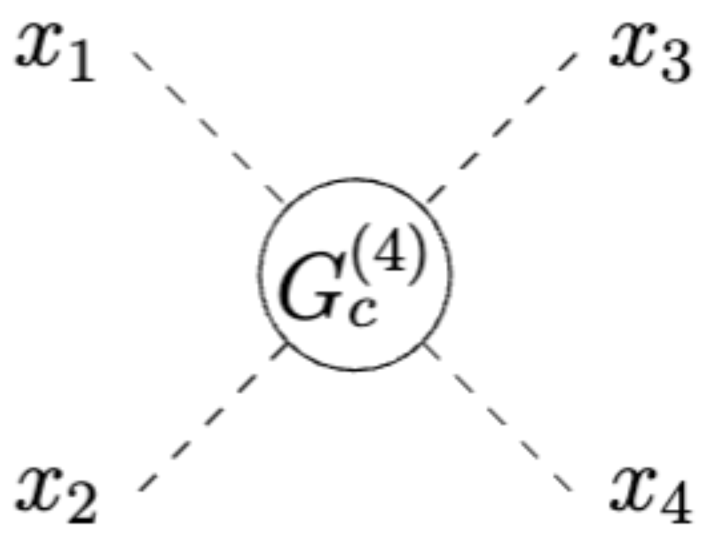
Example: Finite width, i.i.d. parameters

Quartic coupling
of NNFT

$$g_4(x_1, \ldots, x_4) = \frac{1}{4!}\Bigg[\int dy_1 dy_2 dy_3 dy_4\, G_c^{(4)}(y_1, y_2, y_3, y_4)\, G_c^{(2)}(y_1, x_1)^{-1} G_c^{(2)}(y_2, x_2)^{-1}$$

$$G_c^{(2)}(y_3, x_3)^{-1} G_c^{(2)}(y_4, x_4)^{-1} \;+ \text{Comb.}\Bigg] + O\left(\frac{1}{N^2}\right),$$

$$=$$



$$+\, O\left(\frac{1}{N^2}\right).$$

Effective action at leading order

$$S = S_{\text{GP}} - \int d^d x_1 \ldots d^d x_4\;\; g_4(x_1, \ldots, x_4)\, \phi(x_1) \ldots \phi(x_4) + O\left(\frac{1}{N^2}\right)$$

Principles: Given $S[\phi]$, find NN architecture for this field theory at initialization.

Can we deform $S[\phi]$ parametrically?

Parameter space — Dual spaces — Field space

$$Z_G[J] = \mathbb{E}_G[e^{\int d^dx\, J(x)\phi(x)}]$$

$$Z_G[J] = \int D\phi\, e^{-S_G[\phi]+\int d^dx J(x)\phi(x)}$$

How can we induce such an action deformation parametrically?

We want this outcome

$$Z[J] = \mathbb{E}_G[e^{-\lambda \int d^dx_1 \ldots d^dx_r\, \mathcal{O}_\phi(x_1,\ldots,x_r)} e^{\int d^dx\, J(x)\phi(x)}]$$

Dual descriptions

$$Z[J] = \int D\phi\, e^{-S[\phi]+\int d^dx J(x)\phi(x)}$$

$\mathcal{O}_\phi$ is nonlocal operator, may depend on ϕ and its derivatives

$$S_G[\phi] \to S[\phi] = S_G[\phi] + \lambda \int d^dx_1 \ldots d^dx_r\, \mathcal{O}_\phi(x_1,\ldots,x_r)$$

Let's try...

Partition function of NNGP, given in parameter space.

$$Z_G[J] = \int d\theta P_G(\theta)\, e^{\int d^dx J(x)\phi_\theta(x)}$$

What can we change in the architecture to get the following partition function?

$$Z[J] = \int d\theta P_G(\theta)\; e^{-\lambda \int d^dx_1 \ldots d^dx_r\, \mathcal{O}_{\phi_\theta}(x_1,\ldots,x_r)} e^{\int d^dx J(x)\phi_\theta(x)}$$

Try this: Infinite width $\lim N \to \infty$, parameters identically drawn from

$$P(\theta) := P_G(\theta)\; e^{-\lambda \int d^dx_1 \ldots d^dx_r\, \mathcal{O}_{\phi_\theta}(x_1,\ldots,x_r)}$$

Operator $\mathcal{O}_{\phi_\theta}$ violates CLT by breaking statistical independence!

Example: Scalar $\lambda\phi^4$ theory

Target action:

$$S[\phi] = \int d^d x \left[\phi(x)(\nabla^2 + m^2)\phi(x) + \frac{\lambda}{4!}\phi(x)^4\right]$$

$$\nabla^2 := \frac{\partial^2}{\partial x^2}$$

Engineer architecture for NNGP

$$\phi_{a,b,c}(x) = \sqrt{\frac{2\,\mathrm{vol}(B^d_\Lambda)}{\sigma_a^2(2\pi)^d}} \sum_{i,j} \frac{a_i \cos(b_{ij}x_j + c_i)}{\sqrt{\mathbf{b}_i^2 + m^2}}$$

Width $\lim N \to \infty$.

Parameter distributions

$$P_G(a) = \prod_i e^{-\frac{N}{2\sigma_a^2} a_i a_i}$$

$$P_G(b) = \prod_i P_G(\mathbf{b}_i) \text{ with } P_G(\mathbf{b}_i) = \mathrm{Unif}(B^d_\Lambda)$$

$$P_G(c) = \prod_i P_G(c_i) \text{ with } P_G(c_i) = \mathrm{Unif}([-\pi, \pi])$$

Example continued

2-pt function at NNGP is that of a free scalar field theory in d Euclidean dimensions and mass m.

$$G^{(2)}(p) = \frac{1}{p^2 + m^2}$$

To realize $\lambda\phi^4$ theory as an action deformation of this NNGP, we need to insert the operator,

$$e^{-\frac{\lambda}{4!}\int d^dx\, \phi_{a,b,c}(x)^4}$$

Express $\phi_{a,b,c}$ in parameter space

Let's absorb this operator insertion into new CLT-breaking parameter distribution at $\lim N \to \infty$.

$$P(a,b,c) = P_G(a)P_G(b)P_G(c)\; e^{-\frac{\lambda}{4!}\int d^dx\, \phi_{a,b,c}(x)^4}$$

New partition function is

$$Z[J] = \int da\, db\, dc\;\; P(a,b,c)\; e^{\int d^dx J(x)\, \phi_{a,b,c}(x)}$$

All new correlation functions are invariant under Euclidean invariance, permutation invariance, satisfy cluster decomposition, and reflection positivity (OS axioms). This NNFT is a NN *quantum* field theory!

Conclusion and outlooks

- Given NN architecture, we compute correlation functions in parameter space, and use the new Feynman rules to compute interaction terms in NN field action $S[\phi]$.

- Given an interacting field theory action $S[\phi]$, we can design infinite width NNs to produce it by operator insertion through correlated parameter distributions.

- As an example, we found an infinite width NN for $\lambda\phi^4$ scalar field theory.

- We still need principles to design finite width NNs that result in (local) interacting field theory actions at initialization.

Thank you!

Questions?

Feel free to get in touch:
email: amaiti@seas.harvard.edu
Twitter: @AninditaMaiti7
web: https://aninditamaiti.github.io/

Extra Slides

CGF and Edgeworth Method

Consider a sum over N random variables:

$$\phi = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$$

Moment generating function of ϕ:

$$Z_\phi[J] := \mathbb{E}[e^{J\phi}] = \mathbb{E}[e^{J \sum_i X_i/\sqrt{N}}]$$

Cumulant generating functional of ϕ is the log go the moment generating function:

$$W_\phi[J] := \log\, \mathbb{E}[e^{J\phi}] = \log\, \mathbb{E}[e^{J \sum_i X_i/\sqrt{N}}] \;= \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} J^r$$

Cumulant or connected statistical moment of ϕ:

$$\kappa_r^\phi := \left(\frac{\partial}{\partial J}\right)^r W_\phi[J]$$

Edgeworth method for probability density of ϕ in terms of its cumulants:

$$P[\phi] = \exp\left[\sum_{r=3}^{\infty} \frac{\kappa_r}{r!}(-\partial_\phi)^r\right] \int dJ e^{\kappa_2 \frac{J^2}{2} + \kappa_1 J - J\phi}$$

$$= \exp\left[\sum_{r=3}^{\infty} \frac{\kappa_r}{r!}(-\partial_\phi)^r\right] e^{-\frac{(\phi-\kappa_1)^2}{2\kappa_2}},$$

Compute NNFT quartic for finite N, non-i.i.d. parameters

For statistical independence breaking hyper parameter α, connected correlation functions at leading order

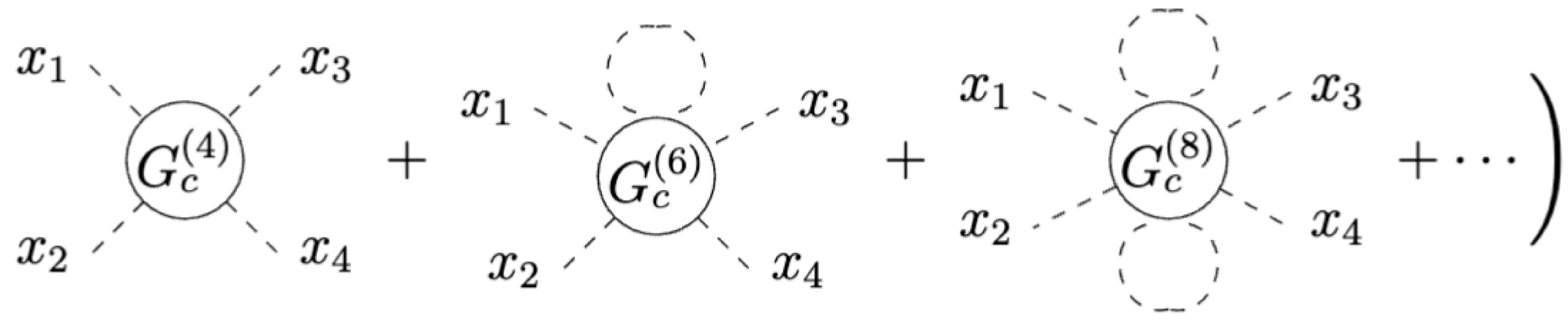
$$G_c^{(r)}(x_1, \cdots, x_r) \propto \alpha \quad \forall r > 2$$

Quartic coupling at leading order

$$g_4(x_1, x_2, x_3, x_4) = \sum_{n=2}^{\infty} \frac{(-1)^{2n-4}}{(2n)!} \Big[\int dy_1 \cdots dy_{2n}\, G_c^{(2n)}(y_1, \cdots, y_{2n})\, G_c^{(2)}(y_1, x_1)^{-1}$$
$$G_c^{(2)}(y_2, x_2)^{-1}\, G_c^{(2)}(y_3, x_3)^{-1}\, G_c^{(2)}(y_4, x_4)^{-1} \prod_{m=5}^{2n-1} G_c^{(2)}(y_m, y_{m+1})^{-1} + \text{Comb.} \Big] + O(\alpha^2)$$

It's an infinite series sum! Cannot be truncated unless we impose more conditions on NN architecture.

$$=(-1)^{2n-4} \Bigg($$



$$\Bigg) + O(\alpha^2)$$

Compute $S[\phi]$ for an actual architecture

Network output $\phi(x) = W_i^1 \cos(W_{ij}^0 x_j + b_i^0)$

Parameter distributions

$$W^0 \sim \mathcal{N}(0, \sigma_{W_0}^2/d)$$

$$W^1 \sim \mathcal{N}(0, \sigma_{W_1}^2/N)$$

$$b^0 \sim \mathrm{Unif}[-\pi, \pi]$$

NN field theory action at leading order

$$S_{\mathrm{Cos}}[\phi] = \frac{2\sigma_{W_0}}{\sigma_{W_1}^2\sqrt{d}} \int d^dx\, \phi(x)\, e^{-\frac{\sigma_{W_0}^2 \nabla_x^2}{2d}} \phi(x) - \int d^dx_1 \cdots d^dx_4 \left[\frac{4\sqrt{6}\pi^{3/2}\sigma_{W_0}^4}{Nd^2\sigma_{W_1}^4} \sum_{\mathcal{P}(abcd)} e^{-\frac{\sigma_{W_0}^2 \nabla_{r_{abcd}}^2}{6d}} - \frac{8\pi\sigma_{W_0}^4}{Nd^2\sigma_{W_1}^4} \sum_{\mathcal{P}(ab,cd)} e^{-\frac{\sigma_{W_0}^2(\nabla_{r_{ab}}^2 + \nabla_{r_{cd}}^2)}{2d}} \right] \phi(x_1)\cdots\phi(x_4) + O(1/N^2). \quad (3.51)$$