Building Quantum Field Theories out of Neurons

Jim Halverson

Based on: [J.H.],
[J.H., Maiti, Stoner], [Maiti, Stoner, J.H.]

Related WIP: [Gukov, J.H.], [J.H., Maiti, Schwartz, Stoner]

Northeastern University

The Gang

Anindita Maiti
targeting physics postdocs, start Fall 2023

Keegan Stoner
targeting ML labs, start Fall 2022

Connecting @ Physics / ML Interface

NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

one of five inaugural NSF AI research institutes, this one at the interface with physics!

MIT, Northeastern, Harvard, Tufts.

ML for physics / math discoveries?
Can physics / math help ML?

Sign up for our mailing list: www.iaifi.org.
Spring colloquia to be announced soon!

Summer school in 2022!

Physics ∩ ML

Physics Meets ML

virtual seminar series, “continuation” of 2019 meeting at Microsoft Research.

Bi-weekly seminars from physicists and CS, academia and industry.

Organizers: Bahri (Google), Krippendorf (LMU Munich), J.H., Paganini (DeepMind), Ruehle (CERN), Shiu (Madison), Yang (MSR)

Sign up at www.physicsmeetsml.org.

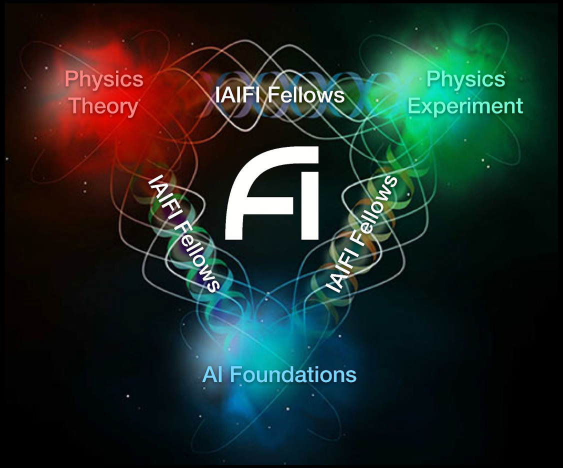
Feel free to reach out!

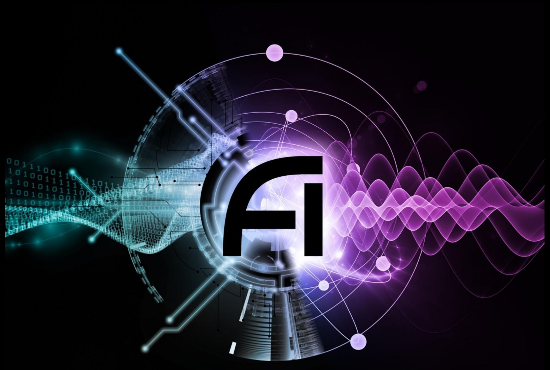
e-mail: jhh@neu.edu
Twitter: @jhhalverson
web: www.jhhalverson.com

ML for Math:
e.g. “Learning to Unknot”: 2010.16263
ML for Strings:
e.g. “Statistical Predictions in String Theory and Deep Generative Models”: 2001.00555
ML for QM + CM:
2112.00723 “Infinite Neural Network Quantum States”
NN / Field Theory:
2008.08601 “Neural Networks and Quantum Field Theory”
2106.00694 “Symmetry-via-Duality”

NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/)

Advance physics knowledge—from the smallest building blocks of nature to the largest structures in the universe—and galvanize AI research innovation





https://iaifi.org

iaifi-management@mit.edu

IAIFI Postdoctoral Fellowship

@iaifi-news

Finite-N Literature: Field Theory Ideas for ML

- [Dyer, Gur-Ari] 1909.11304 **Asymptotics of Wide Networks from Feynman Diagrams**
- [Yaida] 1910.00019 **Non-Gaussian processes and neural networks at finite widths**
- [J.H., Maiti, Stoner] 2008.08601 **Neural Networks and Quantum Field Theory**
- [Bachtis, Aarts, Lucini] 2102.09449 **Quantum field-theoretic machine learning**
- [Zavatone-Veth, Pehlevan] 2104.11734 **Exact marginal prior distributions of finite Bayesian neural networks**
- [Maiti, Stoner, J.H.] 2106.00694 **Symmetry-via-Duality: Invariant Neural Network Densities from Parameter-Space Correlators**
- [Roberts, Yaida, Hanin] 2106.10165 **The Principles of Deep Learning Theory**
- [Erdmenger, Grosvenor, Jefferson] 2107.06898
  Towards quantifying information flows: relative entropy in deep neural networks and the renormalization group
- [Erbin, Lahoche, Samary] 2108.01403 **Nonperturbative renormalization for the neural network-QFT correspondence**
- [Grosvenor, Jefferson] 2109.13247 **The edge of chaos: quantum field theory and deep neural networks**

<u>**if I missed you!**</u>
my sincere apologies,
please write me.

Finite-N Literature: Field Theory Ideas for ML

- [Dyer, Gur-Ari] 1909.11304 **Asymptotics of Wide Networks from Feynman Diagrams**
- [Yaida] 1910.00019 **Non-Gaussian processes and neural networks at finite widths**
- [J.H., Maiti, Stoner] 2008.08601 **Neural Networks and Quantum Field Theory**
- [Bachtis, Aarts, Lucini] 2102.09449 **Quantum field-theoretic machine learning**
- [Zavatone-Veth, Pehlevan] 2104.11734 **Exact marginal prior distributions of finite Bayesian neural networks**
- [**Maiti**, Stoner, J.H.] 2106.00694 **Symmetry-via-Duality: Invariant Neural Network Densities from Parameter-Space Correlators**
- [**Roberts**, Yaida, Hanin] 2106.10165 **The Principles of Deep Learning Theory**
- [Erdmenger, Grosvenor, Jefferson] 2107.06898
  Towards quantifying information flows: relative entropy in deep neural networks and the renormalization group
- [**Erbin**, Lahoche, Samary] 2108.01403 **Nonperturbative renormalization for the neural network-QFT correspondence**
- [Grosvenor, Jefferson] 2109.13247 **The edge of chaos: quantum field theory and deep neural networks**

this conference:
numerous talks on
this subject!

See talks: Maiti, Roberts, Erbin.

Goal Here: NN Approach to QFT

Q: what takes the place of an action?
Q: design principles? symmetry?
Q: interactions, from where?

Q: when can we rotate to Lorentzian signature, get QFT defined by NNs?

Outline

- Building Quantum Fields out of Neurons
- Turning on Interactions: Break the Central Limit Theorem
- Engineering Symmetries + Euclidean Invariance
- Neural Network QFTs:
- Examples: a Gaussian NN **Quantum** Field Theory and a large-N duality

Building Quantum Fields Out of Neurons

Fields from Neurons

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

$$\theta = \{a, \theta_h\}$$

$$a \sim P(a) \qquad \theta_h \sim P(\theta_h)$$

For simplicity + scaling:

$$\mathbb{E}[a^4] = \gamma_a^4/N^2 \qquad \mathbb{E}[a^2] = \sigma_a^2/N \qquad \mathbb{E}[a^{2n+1}] = 0$$

- real scalar = sum of N generally ***non-Gaussian*** neurons h_i.
- neurons identically distributed, not necessarily independently distributed.
- a priori, don't know action of fields or neurons.

Key Point:
randomness of fields from construction,
not the action.

theory def'd by architecture,
not (necessarily) a Lagrangian.

Correlation Functions: In Parameter Space

method of computing: in GPs,
"Computing with Infinite Networks" [Williams]
NeurIPS 1996

Field and Neuron Correlators:

$$G^{(n)}(x_1, \ldots, x_n) = \frac{1}{Z_\theta} \int d\theta \, \phi(x_1) \ldots \phi(x_n) P(\theta)$$

$$H^{(m)}_{i_1, \ldots, i_n}(x_1, \ldots, x_n) = \frac{1}{Z_{\theta_h}} \int d\theta_h \, h_{i_1}(x_1) \ldots h_{i_n}(x_n) P(\theta_h)$$

- **exact expressions** in parameter space.
- doesn't require knowledge of

$$S[\phi] \qquad\qquad S[h]$$

- sometimes can **evaluate exactly**! see extras.
- can **study / utilize**, e.g. for symmetries, even if you can't do the integral. see Maiti's talk.

Two-Point Function:

$$G^{(2)}(x_1, x_2) = \frac{\sigma_a^2}{N} \sum_{i=1}^{N} H^{(2)}_{ii}(x_1, x_2) = \sigma_a^2 H^{(2)}_{ii}(x_1, x_2)$$

- det'd diagonal neuron 2-pt function.
- more general: field correlators function of neuron correlators
- notice: N-scaling of variance chosen wisely.
- identical neurons allows no sum on i.

Turning on Interactions

Need non-Gaussianities?

Break the central limit theorem.

Reminder: Free Theories from Central Limit Theorem

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

Drawn from Gaussian distribution on functions, by Central Limit Theorem, when:

1) infinite sum of
2) independent random variables.

Of course: does *not* require Gaussian neurons h_i.

Neural Network Gaussian Process (NNGP)

Key Point: of slide below is that NNGPs are everywhere.

NNGP Correspondence: GP limits are ubiquitous!

Single-layer infinite width feedforward networks are GPs. [Neal], [Williams] 1990's

Deep infinite width feedforward networks are GPs. [Lee et al., 2017], [Matthews et al., 2018]

Infinite channel CNNs are GPs. [Novak et al., 2018] [Garriga-Alonso et al. 2018]

Tensor programs show any *standard* architecture admits GP limit. [Yang, 2019]

infinite channel limit [5, 6]. In [7, 8, 9], Yang developed a language for understanding which architectures admit GP limits, which was utilized to demonstrate that any standard architecture admits a GP limit, i.e. any architecture that is a composition of multilayer perceptrons, recurrent neural networks, skip connections [10, 11], convolutions [12, 13, 14, 15, 16] or graph convolutions [17, 18, 19, 20, 21, 22], pooling [15, 16], batch [23] or layer [24] normalization, and / or attention [25, 26]. Furthermore, though these results apply to randomly initialized neural networks, appropriately trained networks are also drawn from GPs [27, 28]. NNGPs have been used to model finite neural networks in [29, 30, 31], with some key differences from our work. For these reasons, we believe that an EFT approach to neural networks is possible under a wide variety of circumstances.

tons of examples cited in our paper admit GP limits

GP property persists under appropriate training. [Jacot et al., 2018] [Lee et al., 2019]

Interactions from Breaking the CLT

- connected 4-pt function takes form:

$$G_c^{(4)}(x_1,\dots,x_4) = I_N^{(4)} + I_{IB}^{(4)}$$

- First term vanishes at infinite-N:

ML Lit: see [Yaida] [J.H., Maiti, Stoner] [Roberts, Yaida, Hanin] [Erbin, Lahoche, Samary] and others for related finite-N results.

$$I_N^{(4)} = \frac{\gamma_a^4}{N} H_{iiii} - \frac{\sigma_a^4}{N}\left(H_{ii}^{12}H_{ii}^{34} + H_{ii}^{13}H_{ii}^{24} + H_{ii}^{14}H_{ii}^{23}\right)$$

$$\lim_{N\to\infty} I_N^{(4)} = 0$$

$$H_{i_1,\dots,i_n} := H^{(n)}_{i_1,\dots,i_n}(x_1,\dots,x_n)$$

- Second term vanishes in independence limit:

$$I_{IB}^{(4)} = \sigma_a^4(1-\delta_{ij})\left(1-\frac{1}{N}\right)\Big(H_{iijj} - H_{ii}^{12}H_{jj}^{34} + H_{ijij} - H_{ii}^{13}H_{jj}^{24} + H_{ijji} - H_{ii}^{14}H_{jj}^{23}\Big)$$

where $H_{iijj} = H_{ii}^{12}H_{jj}^{34}$

- Sometimes have parametric control over breaking of independence, use pert. thy.

$$\lim_{\epsilon\to 0} P(\theta_h) = \prod_{i=1}^{N} P(\theta_{h_i}) \qquad |\epsilon| \ll 1$$

Can still have symmetries. (See Maiti talk).

Engineering Symmetries

Generally, but also Euclidean invariance.

Symmetries: The Basic Idea

$$G^{(n)}(x_1, \ldots, x_n) = \frac{1}{Z_\theta} \int d\theta \, \phi(x_1) \ldots \phi(x_n) P(\theta)$$

- recall: computing corrs. in parameter space.
- demonstrate symmetries via param. space study of correlators.
- existence or non-existence of symms. arises from parameter measures and densities.

Simple Example

$\mathbb{Z}_2$ symmetry: $\phi \to -\phi$

from $\mathbb{E}[a^{2n+1}] = 0$
which ensures odd-point functions vanish.

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

$$G^{(n)}(x_1, \ldots, x_n) = \mathbb{E}[\phi(x_1) \ldots \phi(x_n)]$$

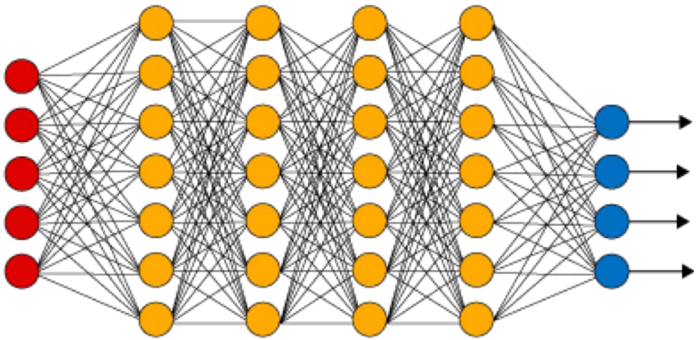
Inheritance of Spacetime Symmetries

- fields from neurons h_i.
- neurons themselves a complicated function g acting on an input layer l_j.
- **input layer invariance:**
  act on x, absorb into θ_l, transform E_l, demonstrate inv. of correlators.
- **field / neuron ensembles invariant if:**
  1) input layer ensemble is invariant.
  2) parameter constraint satisfied.

  (idea: still leaves E_l invariant).

$$\theta_h = \theta_g \cup \theta_\ell$$
$$\theta_g \cap \theta_\ell = \emptyset$$

$$h_i(x) = \sum_{j=1}^{k} g_{ij}(\ell_j(x))$$



Key Point:
ensembles of fields and deep neurons can inherit symmetries from input layer.

Euclidean Invariance

$$\ell_i(x) = F(\mathbf{b}_i) \cos\left(\sum_j b_{ij} x_j + c_i\right)$$

$$\ell : \mathbb{R}^d \to \mathbb{R}^N$$

$$b_{ij} \sim P(b_{ij}) \quad c \sim U[-\pi, \pi]$$

$$\mathbf{b}_i \quad i^{\text{th}} \text{ row of } b_{ij}$$

Translation Invariance:

$$L^{(n)}_{i_1 \dots i_n}(x_1, \dots, x_n) = \mathbb{E}_{b,c}[\ell_{i_1}(x_1) \dots \ell_{i_n}(x_n)]$$
$$= \mathbb{E}_b\Big[F(\mathbf{b}_{i_1}) \dots F(\mathbf{b}_{i_n}) \times$$
$$\mathbb{E}_c[\cos(b_{i_1 j_1} x^1_{j_1} + c_{i_1}) \dots \cos(b_{i_n j_n} x^n_{j_n} + c_{i_n})]\Big]$$

$$\mathbb{E}_c[\cos(b_{i_1 j_1}(x_{j_1} + d_{j_1}) + c_{i_1}) \dots \cos(b_{i_n j_n}(x_{j_n} + d_{j_n}) + c_{i_n})]$$
$$= \mathbb{E}_c[\cos(b_{i_1 j_1} x_{j_1} + \tilde{c}_{i_1}) \dots \cos(b_{i_n j_n} x_{j_n} + \tilde{c}_{i_n})]$$
$$= \mathbb{E}_{\tilde{c}}[\cos(b_{i_1 j_1} x_{j_1} + \tilde{c}_{i_1}) \dots \cos(b_{i_n j_n} x_{j_n} + \tilde{c}_{i_n})], \qquad (20)$$

- **want:** Euclidean invariance, so we can get Lorentz invariance after continuation.
- **left:** a family of Euclidean-invariant input layer architectures.
- **left:** demonstration of translation invariance.
- SO(d) invariance of l-ensemble follows if $F(b_i)$ and $P(b_i)$ are invariant .

Key Point:

families of examples for both deep and shallow networks.

Neural Network Quantum Field Theories

Neural Network Quantum Field Theories

So far: ensembles of Euclidean fields from NNs.

Definition: a NN architecture induces a NN-QFT if it is the Wick-rotation of a unitary Lorentzian QFT.

Q: when is a Euclidean theory the Wick rotation of a Lorentzian QFT?

Axions for Euclidean Green's Functions

Konrad Osterwalder* and Robert Schrader**

Lyman Laboratory of Physics, Harvard University, Cambridge, Mass. USA

Received December 18, 1972

Abstract. We establish necessary and sufficient conditions for Euclidean Green's functions to define a unique Wightman field theory.

Osterwalder-Schrader Axioms:
when do Euclidean correlators define a Wightman QFT?

Need to satisfy:

E0: Temperedness
E1: Euclidean covariance (just engineered)
E2: Reflection Positivity (rel. to unitarity)
E3: Permutation Symmetry (trivial for NNs)
E4: Cluster decomposition

Relax E1 for NN-QFTs that are not Lorentz invariant?

Unitarity and Reflection Positivity

early lit: [Osterwalder, Schrader]
text: [Glimm, Jaffe]

$$\mathcal{O}_E(t_E, \mathbf{x})^\dagger = \mathcal{O}_E(-t_E, \mathbf{x})$$

$$|\psi\rangle = \mathcal{O}(-t_{E1}) \cdots \mathcal{O}(-t_{En})|0\rangle$$

$$\langle\psi| = \langle 0|\mathcal{O}(t_{En}) \cdots \mathcal{O}(t_{E1})$$

$$\langle\psi|\psi\rangle \geq 0$$

yields a constraint called ***Reflection Positivity*** on Euclidean correlators.

RP gives a constraint on ***all*** Euclidean correlators.

Important for future work: general RP NNs.

Here: content ourselves to check 2-pt condition, which ensures RP of any Gaussian process, therefore NNGPs.

$$\int d^d x d^d y\, f^*(x) f(y)\, G^{(2)}(x^\theta, y) \geq 0$$

Unitarity and Reflection Positivity: Power Spectrum Rep

- interested in Euclidean-invariant theories.
- 2-pt function T-invariant.
  equiv: determined by power spectrum.
- Q: what is RP constraint on power spectrum?

Key points:
1) any T-invt NN-QFT must satisfy.
2) for T-invt NNGP, this is sufficient for RP.

Two-point RP Rephrased:

$$\int \frac{d^{d-1}\mathbf{p}}{(2\pi)^{d-1}} \left[\int d^d x \, e^{-i\mathbf{p}\cdot\mathbf{x}} f^*(x) \right] \left[\int d^d y \, e^{i\mathbf{p}\cdot\mathbf{y}} f(y) \right]$$

$$\times \rho(\mathbf{p}, \tau_x + \tau_y) \geq 0$$

$$\rho(\mathbf{p}, \tau_x + \tau_y) = \int \frac{dp_0}{2\pi} e^{ip_0(\tau_x + \tau_y)} G^{(2)}(p)$$

Sufficient condition for RP: B positive,

$$\rho = A^*(\mathbf{p}, \tau_x) A(\mathbf{p}, \tau_y) B(\mathbf{p})$$

Examples

A family of examples.

Specific examples: Gaussian NN-QFT and Large-N Duality

Single Layer Families of Examples

Thus far, we have: $$\phi = \sum_{i=1}^{N} a_i h_i(x)$$

$$\theta = \{a, \theta_h\}$$

$$a \sim P(a) \qquad \theta_h \sim P(\theta_h)$$

$$h_i(x) = \sum_{j=1}^{k} g_{ij}(\ell_j(x))$$

Euclidean-invariant input layer ensemble ensures Euclidean invariant neuron ensemble, under parameter constraint.

$$\theta_h = \theta_g \cup \theta_\ell$$
$$\theta_g \cap \theta_\ell = \emptyset$$

One Euclidean invariant input layer is

$$\ell_i(x) = F(\mathbf{b}_i)\cos\Big(\sum_j b_{ij}x_j + c_i\Big)$$

$$\ell : \mathbb{R}^d \to \mathbb{R}^N$$
$$b_{ij} \sim P(b_{ij}) \quad c \sim U[-\pi, \pi]$$
$$\mathbf{b}_i \quad i^{\text{th}} \text{ row of } b_{ij}$$

with $P(b_i)$ and $F(b_i)$ chosen to be invariant.

Single-layer family of examples: just take $g_{ij} = \delta_{ij}$

i.e., scalar field is $$\phi(x) = \sum_{i=1}^{N} a_i \ell_i(x)$$

Fully specified example: must choose $P(b_i)$ and $F(b_i)$

Neuron Correlators

Two-point Function

$$L^{(2)}_{ij}(x,y) = \frac{1}{2}\delta_{ij}\,\mathbb{E}_b[F(\mathbf{b}_i)^2 \cos(b_{ik}(x_k - y_k))]$$

Power Spectrum

$$\begin{aligned} L^{(2)}_{ij}(p) &= \frac{1}{4}\delta_{ij}\,\mathbb{E}_b\left[F(\mathbf{b}_i)^2\left(\delta^{(d)}(b_{ik}-p_k) + \delta^{(d)}(b_{ik}+p_k)\right)\right] \\ &= \frac{1}{4Z_{\mathbf{b}_i}}\delta_{ij}\left[P(\mathbf{b}_i)F(\mathbf{b}_i)^2\big|_{b_{ik}=p_k} + P(\mathbf{b}_i)F(\mathbf{b}_i)^2\big|_{b_{ik}=-p_k}\right]\end{aligned}$$

Four-point Correlators

$$L_{iikk} = \frac{1}{4}\mathbb{E}_b\left[F(\mathbf{b}_i^2)F(\mathbf{b}_j^2)\cos(b_{ij}(x_j^1 - x_j^2))\cos(b_{ij}(x_j^3 - x_j^4))\right]$$

$$L_{iiii} = \frac{1}{8}\mathbb{E}_b\left[F(\mathbf{b}_i^4)\left(\cos(b_{ij}x_j^{++--}) + 2\text{ perms.}\right)\right]$$

$$\begin{aligned} L_{ijkl} &:= L^{(4)}_{ijkl}(x_1, x_2, x_3, x_4) \\ x^{++--} &:= x_1 + x_2 - x_3 - x_4 \end{aligned}$$

- **tunable power spectrum!** via choice of $F(b_i)$ and $P(b_i)$
- manifest translation invariance.
- SO(d) invariant for symmetric $F(b_i)$ and $P(b_i)$.

Specific Example #1: A Gaussian NN-QFT

- take $g_{ij} = \delta_{ij}$ family, specify:

$$F(\mathbf{b}_i) = \frac{1}{\sqrt{\mathbf{b}_i^2 + m^2}} \qquad\qquad P(\mathbf{b}_i) = \mathcal{U}(S^d_\Lambda)$$

- power spectrum is:

$$G^{(2)}(p) = \frac{\sigma_a^2\,(2\pi)^d}{2\,\mathrm{vol}(S^d_\Lambda)}\,\frac{1}{p^2+m^2}$$

- $N \rightarrow \infty$ for Gaussianity.

Key Point:

free scalar as a NN-QFT!

Specific Example #2: Another Gaussian NN-QFT

- take $g_{ij} = \delta_{ij}$ family, d=1, specify:

$$F(b_i) = 1 \qquad\qquad P(b_i) = \frac{1}{b_i^2 + m^2}$$

- power spectrum is:

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

- N $\longrightarrow$ ∞ for Gaussianity.
- RP from structure of

$$\rho(p, \tau_x + \tau_y) = \frac{\pi m \sigma_a^2}{2} \frac{e^{-\mu(\tau_x + \tau_y)}}{\mu}$$

$$\mu = \sqrt{p^2 + m^2}$$

Key Point:

used tunability of power spectrum to get another Gaussian NN-QFT.

Specific Example #3: Large-N Duality

Cos-net:

$$\phi(x) = \sum_{i=1}^{N} a_i l_i(x), \qquad \ell_i(x) = \sum_j \cos(b_{ij} x_j + c_i)$$

$$a \sim \mathcal{N}(0, \frac{\sigma_a^2}{N}), \;\; b \sim \mathcal{N}(0, \frac{\sigma_b^2}{d}), \;\; c \sim \mathcal{U}[-\pi, \pi]$$

Gauss-net:

$$\phi(x) = \frac{\sum_{i=1}^{N} a_i g_i(x)}{\sqrt{\exp[2(\sigma_c^2 + \sigma_b^2 x^2/d)]}} \qquad g_i(x) = \sum_j \exp(b_{ij} x_j + c)$$

$$a \sim \mathcal{N}(0, \frac{\sigma_a^2}{2N}) \;\; b \sim \mathcal{N}(0, \frac{\sigma_b^2}{d}) \;\; c \sim \mathcal{N}(0, \sigma_c^2),$$

both architectures have

$$G^{(2)}(p) = \frac{\sigma_a^2}{2Z_b} e^{-\frac{1}{2}\frac{d}{\sigma_b^2}p^2}$$

there **dual theories** in NNGP limit, N → ∞.

however, at finite-N, theories and symmetries are different!

extra slides: exact connected 4-pt functions.

Miscellany:

It’s interesting to think about classical configurations and spontaneous symmetry breaking in this context.

See paper, I’m surely out of time.

Summary and Outlook

In Summary:

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

parameters a_i. drawn iid as $a \sim P(a)$

$$\mathbb{E}[a] = 0 \qquad \mathbb{E}[a^2] = \sigma_a^2/N$$

$$\mathbb{E}[a^4] = \gamma_a^4/N^2$$

neuron h_i has its own parameters.

Key points:

- fields w/ **different origin of randomness**, from neuron construction, not distribution in path integral.
- **non-Lagrangian.** often non-perturbative.
  [J.H., Maiti, Stoner] for Lagrangian modeling
- can satisfy Osterwalder-Schrader axioms, give **Lorentz-invariant, unitary theory**.
- **Interactions:** break CLT!
  i.e. from finite-N, or independence breaking.
- **Important for future:** theory of RP NNs.

Some Predictions

If built from *many nearly-independent* neurons,
QFTs are nearly-Gaussian.
Of course, often see this in Nature.

Mixed field-neuron correlators to excite individual neurons?
Possibility: neurons not just kinematically inaccessible,
also statistically inaccessible.

Potential Use Case

Lattice field theory.
usual *ML* for lattice tackles sampling problem.

This is a fundamentally different construction
where sampling is easy, but action is unknown.
But we do have design principles!

Thanks!

Questions?

Or get in touch after:

e-mail: jhh@neu.edu
Twitter: @jhhalverson
web: www.jhhalverson.com

Gauss-net and Cos-net 4-pt Functions

A Key Point:
Gauss-net not T-invariant at Finite-N
Cos-net T-invariant a Finite-N.

Gauss-net:

$$G^{(4)}|_{c,\text{Gauss}} = \frac{1}{4N}\sigma_a^4\Big[3e^{4\sigma_c^2}e^{-\frac{\sigma_b^2}{2d}[x_1^2+x_2^2+x_3^2+x_4^2-2x_1x_2-2x_1x_3-2x_1x_4-2x_2x_3-2x_2x_4-2x_3x_4]}$$
$$- e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_4)^2+(x_2-x_3)^2\right)} - e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_3)^2+(x_2-x_4)^2\right)} - e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_2)^2+(x_3-x_4)^2\right)}\Big]$$

Cos-net:

$$G^{(4)}|_{c,\text{Cos}} = \frac{1}{8N}\sigma_a^4\Big[3\left(e^{-\frac{1}{2d}\sigma_b^2(x_1+x_2-x_3-x_4)^2} + e^{-\frac{1}{2d}\sigma_b^2(x_1-x_2+x_3-x_4)^2} + e^{-\frac{1}{2d}\sigma_b^2(x_1-x_2-x_3+x_4)^2}\right)$$
$$- 2e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_4)^2+(x_2-x_3)^2\right)} - 2e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_3)^2+(x_2-x_4)^2\right)} - 2e^{-\frac{1}{2d}\sigma_b^2\left((x_1-x_2)^2+(x_3-x_4)^2\right)}\Big]$$

Concrete Single-Layer Examples + 2-pt functions K:

Erf-net: $\sigma(z) = \mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z dt\, e^{-t^2}$ $\quad K_{\mathrm{Erf}}(x,x') = \sigma_b^2 + \sigma_W^2 \frac{2}{\pi} \arcsin\left[\frac{2(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} xx')}{\sqrt{\left(1 + 2(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x^2)\right)\left(1 + 2(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x'^2)\right)}}\right]$

Gauss-net: $\sigma(x) = \frac{\exp(W\,x + b)}{\sqrt{\exp\left[2(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x^2)\right]}}$ $\quad K_{\mathrm{Gauss}}(x,x') = \sigma_b^2 + \sigma_W^2\ \exp\left[-\frac{\sigma_W^2 |x - x'|^2}{2d_{\mathrm{in}}}\right]$

ReLU-net: $\sigma(z) = \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$

$$\begin{aligned} K_{\mathrm{ReLU}}(x,x') &= \sigma_b^2 + \sigma_W^2 \frac{1}{2\pi} \sqrt{(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x \cdot x)(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x' \cdot x')} (\sin\theta + (\pi - \theta)\cos\theta), \\ \theta &= \arccos\left[\frac{\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x \cdot x'}{\sqrt{(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x \cdot x)(\sigma_b^2 + \frac{\sigma_W^2}{d_{\mathrm{in}}} x' \cdot x')}}\right], \end{aligned}$$