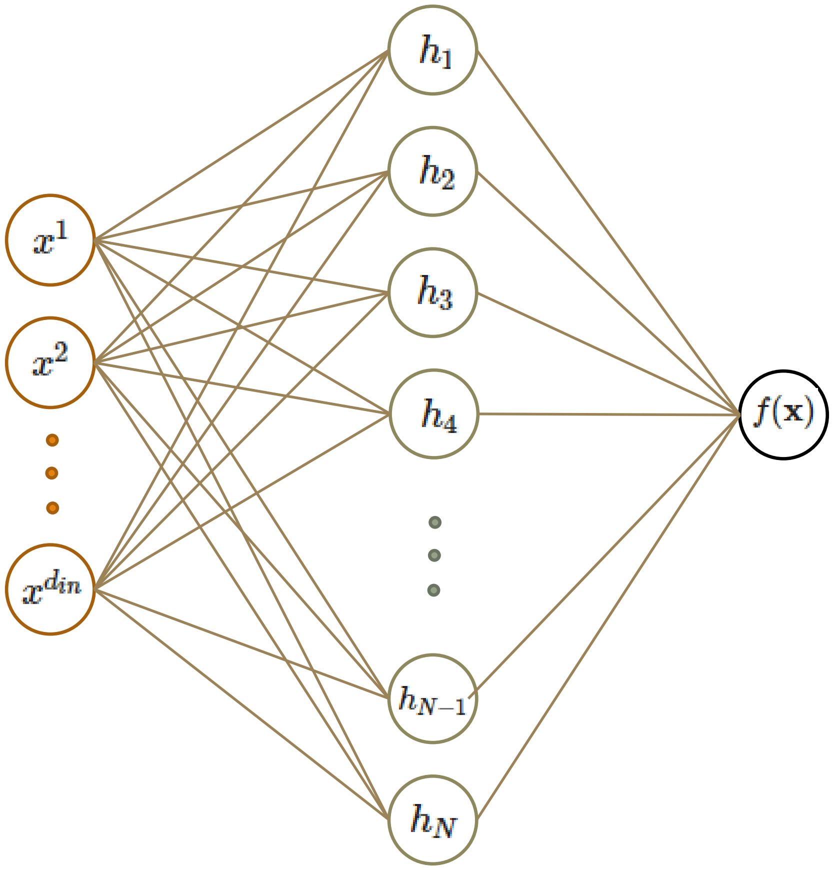
Neural networks and quantum field theory

JAMES HALVERSON, ANINDITA MAITI, AND KEEGAN STONER

HTTPS://DOI.ORG/10.1088/2632-2153/ABECA3

Outline



fully-connected neural network with one hidden layer

Neural network		Gaussian Process		Quantum field theory
infinite width	$\Longleftrightarrow$	GP	$\Longleftrightarrow$	free field theory
finite width	$\Longleftrightarrow$	NGP	$\Longleftrightarrow$	effective field theory
		Non-Gaussian Process		

Gaussian Process

Function over x : $f(x)$

$f(x)$ has a distribution: $p(f)$

x is like time/space coordinates. Infinite choices of x, but for any finite subset of x, the finite function variables $\{f_i\}_{i=1}^{m}$ has joint (zero mean) Gaussian distribution.

$$p(\mathbf{f}|\mathbf{X}) \sim \mathcal{N}$$

For neural networks, x is our input data, f is the network.

$f_{\theta,N} : R^{d_{in}} \to R^{d_{out}}$

parameter space distribution $\longrightarrow$ function space distribution

Correspondence between GP and neural networks

i.i.d.

Consider one hidden later of N units, and output dimension is 1:

$$f(\mathbf{x}) = b_1 + \sum_{j=1}^{N} W_1^j \sigma(\mathbf{x}; \mathbf{W}_0^j, b_0^j)$$

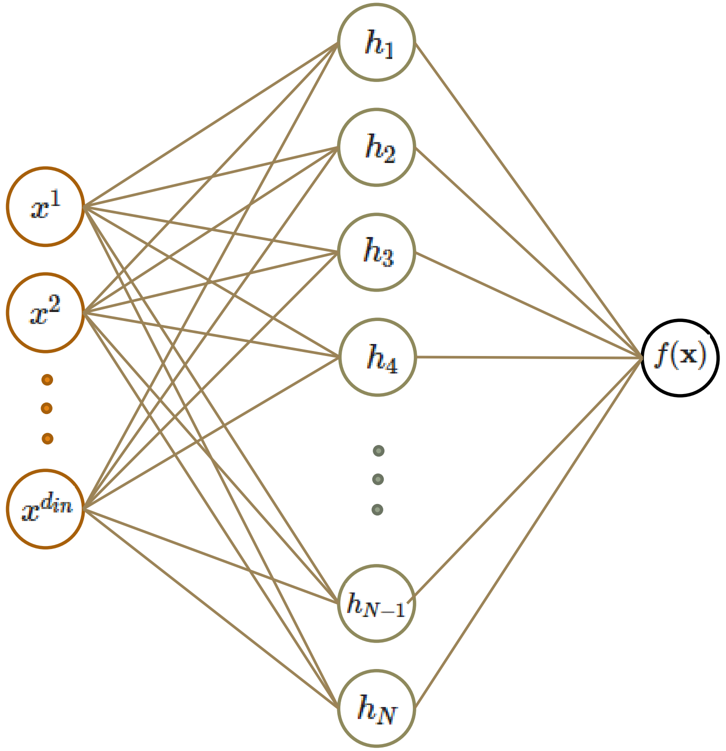
Sum over i.i.d. terms
By CLT, for infinite N
f(x) is Gaussian

$\mathbf{W}_0^j$ is i.i.d., b_0^j, b_1 and W_1^j is zero mean and independent.

In this paper, $b_0^j, b_1 \sim \mathcal{N}(0, \sigma_b^2)$, $\mathbf{W}_0^j \sim \mathcal{N}(0, \sigma_W^2/d_{in})$, $\mathbf{W}_1^j \sim \mathcal{N}(0, \sigma_W^2/N)$

$f_{\theta,N} : R^{din} \rightarrow R^{dout}$

parameter space distribution $\rightarrow$ function space distribution



Correspondence between neural networks and QFT

$$\boxed{\{f(x_1), \ldots, f(x_k)\} \sim \mathcal{N}(\mu, \Xi^{-1})}$$

assumption: $\mu = 0$

$(\Xi^{-1})_{ij} = K_{ij}$

- Correlation function (n-pt functions)

$$\boxed{G^{(n)}(x_1, \ldots, x_n) = \frac{\int df f_1 \ldots f_n e^{-S}}{Z}}$$

- partition function $Z = \int df e^{-S}$
- discrete action $S = \frac{1}{2} f_i \Xi_{ij} f_j$ (Einstein summation)
- continuous action $\boxed{S = \frac{1}{2} \int d^{d_{in}} x d^{d_{in}} x' f(x) \Xi(x, x') f(x')}$

GP/asymptotic NN	Free QFT
Input x	External space or momentum space point
Kernel $K(x_1, x_2)$	Feynman propagator
Asymptotic NN $f(x)$	Free Field
Log-likelihood	Free action S_{GP}

- quantum field $\phi(x)$
- path integral $Z = \int D\phi e^{-S[\phi]}$
- action $S[\phi] = \int d^d x \phi(x)(\Box + m^2)\phi(x)$

Free scalar field theory

Experimental Verification of NNGP

How to measure

- normalized deviation m_n

$$m_n(x_1, \ldots, x_n) = \Delta G^{(n)}(x_1, \ldots, x_n) / G_{GP}^{(n)}(x_1, \ldots, x_n)$$

$$\Delta G^{(n)}(x_1, \ldots, x_n) = G^{(n)}(x_1, \ldots, x_n) - G_{GP}^{(n)}(x_1, \ldots, x_n)$$

Experimental Measurement

$G^{(n)}(x_1, \ldots, x_n) = \mathcal{E}[f_\alpha(x_1) \ldots f_\alpha(x_n)]$ is measured in the experiment $100(\text{experiments}) * 10^5(\text{nets})$ times

weights and biases is drawn from Gaussian dist. with mean equals zero and std equals 1

$G_{GP}^{(n)}(x_1, \ldots, x_n)$ is computed using Wick contraction

$$G_{GP}^{(n)}(x_1, \ldots, x_n) = \sum_{p \in \text{Wick}(x_1, \ldots, x_n)} K(a_1, b_1) \ldots K(a_{n/2}, b_{n/2})$$

Theoretical Computation

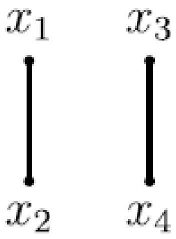
Feynman Diagram

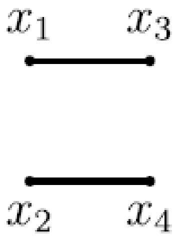
$$G^{(2)}_{\mathrm{GP}}(x_1, x_2) = K(x_1, x_2)$$

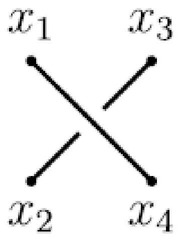
$=$ 

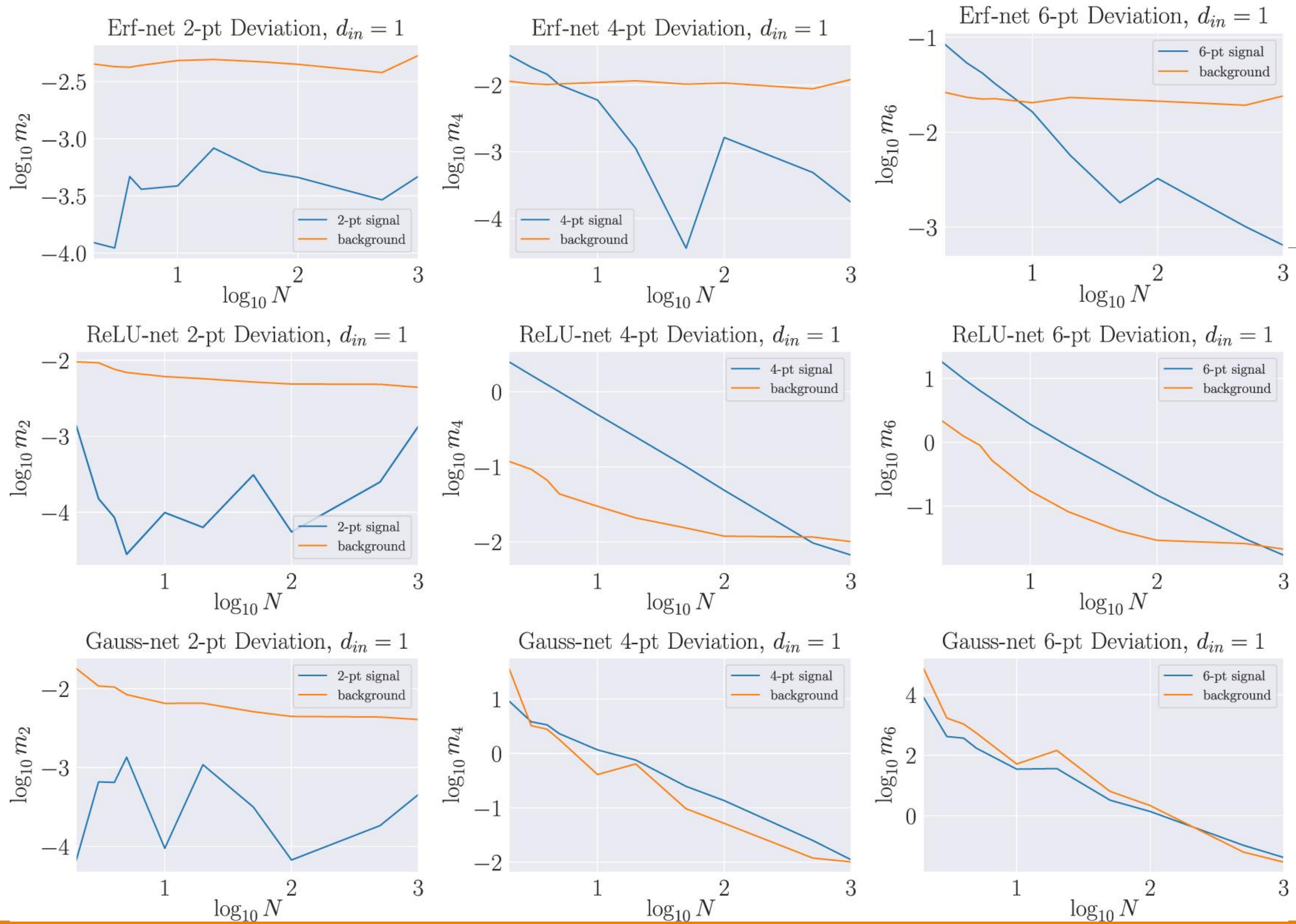
.

$$G^{(4)}_{\mathrm{GP}}(x_1, x_2, x_3, x_4) = K(x_1, x_2)K(x_3, x_4) + K(x_1, x_3)K(x_2, x_4) + K(x_1, x_4)K(x_2, x_3)$$

$=$ 

$+$ 

$+$ 



background is average std across 100 experiments of m_n

NGP contributions

$$\boxed{\Delta G^{(n)} \propto N^{-1}}$$

$$\boxed{G^{(2k)}(x_1,\ldots,x_{2k})|_{\text{connected}} \propto \frac{1}{N^{k-1}}}$$

$$\Delta G^{(2)} \propto 0$$

$$\Delta G^{(4)} \propto N^{-1}$$

$$\Delta G^{(6)} \propto N^{-1} + N^{-2} \propto N^{-1}$$

$$G^{(4)}(x_1,\ldots,x_4)|_{\text{connected}} = \Delta G^{(4)} \propto N^{-1}$$

$$G^{(6)}(x_1,\ldots,x_6)|_{\text{connected}} = \Delta G^{(6)} - \sum G^{(4)}(x_1,\ldots,x_4)|_{\text{connected}}\, G^{(2)}(x_5,x_6) \propto N^{-2}$$

From GP to NGP

NGP NN	EFT
Input x	External space or momentum space point
Kernel $K(x_1, x_2)$	Free or exact propagator
NN output $f(x)$	Interacting Field
Non-Gaussianities	Interactions
Non-Gaussian coefficients	Coupling strengths
Log-likelihood	EFT action S

Perturbation Theory

$$G^{(n)}(x_1, \ldots, x_n) = \frac{\int df f_1 \ldots f_n e^{-S}}{Z_0}$$

$$Z_0 = \int df e^{-S}$$

$$S = S_{GP} + \Delta S$$

for $\mathcal{O}_k := g_k f(x)^k$ in ΔS as $\int d^{d_{in}} x \mathcal{O}_k$

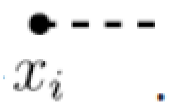
$$G^{(n)}(x_1, \ldots, x_n) = \frac{\int df f(x_1) \ldots f(x_n) \left[1 - \int d^{d_{in}} x g_k f(x)^k + O(g_k^2)\right] e^{-S_{GP}} / Z_{GP,0}}{\int df \left[1 - \int d^{d_{in}} x g_k f(x)^k + O(g_k^2)\right] e^{-S_{GP}} / Z_{GP,0}}$$

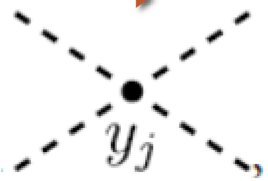
$$\Delta S = \int d^{d_{in}} x \left[\lambda f(x)^4 + \kappa f(x)^6\right]$$

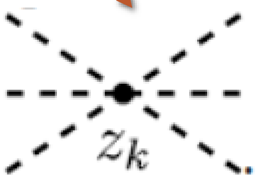
Feynman Rule

$$G^{(n)}(x_1,\ldots,x_n) = \frac{\int df\, f(x_1)\ldots f(x_n)\left[1-\int d^{d_{in}}x\left[\lambda f(x)^4+\kappa f(x)^6\right]\right]e^{-S_{GP}}/Z_{GP,0}}{\int df\left[1-\int d^{d_{in}}x\left[\lambda f(x)^4+\kappa f(x)^6\right]\right]e^{-S_{GP}}/Z_{GP,0}}$$

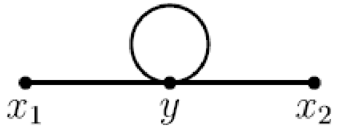
$$\Delta S = \int d^{d_{in}}x\left[\lambda f(x)^4+\kappa f(x)^6\right]$$

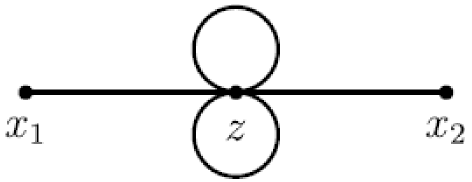






Throw away





Because no correction to 2-pt function

Coupling constants

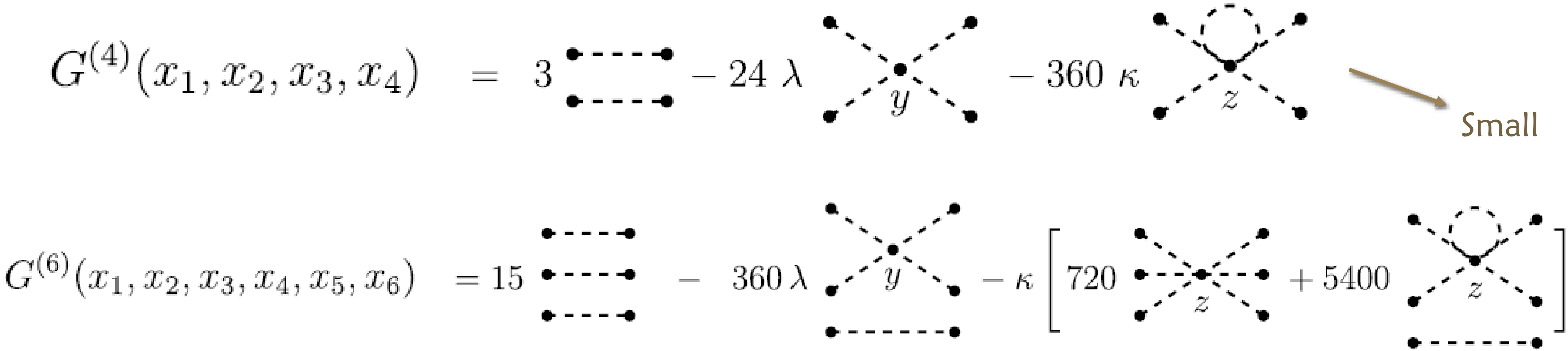
$$\Delta S = \int d^{d_{in}}x \left[\lambda f(x)^4 + \kappa f(x)^6\right]$$

$\lambda, \ \kappa$ is constants by symmetry at GP and Technical Naturalness.

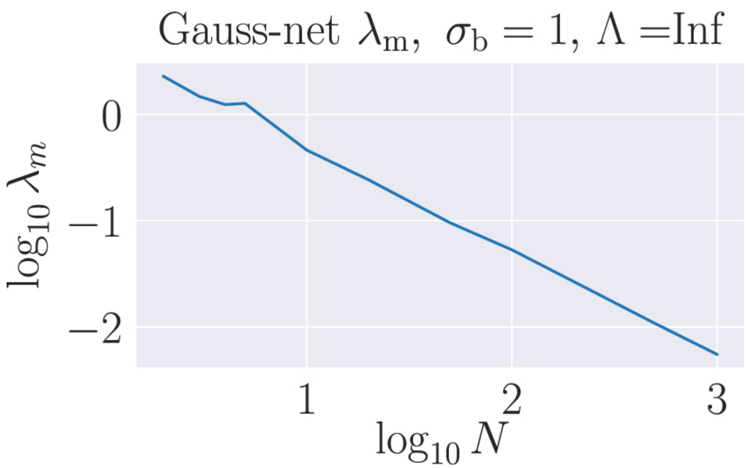
- Technical naturalness: a coupling g appearing in ΔS may be small relative to Λ if a symmetry is restored when g is set to zero.

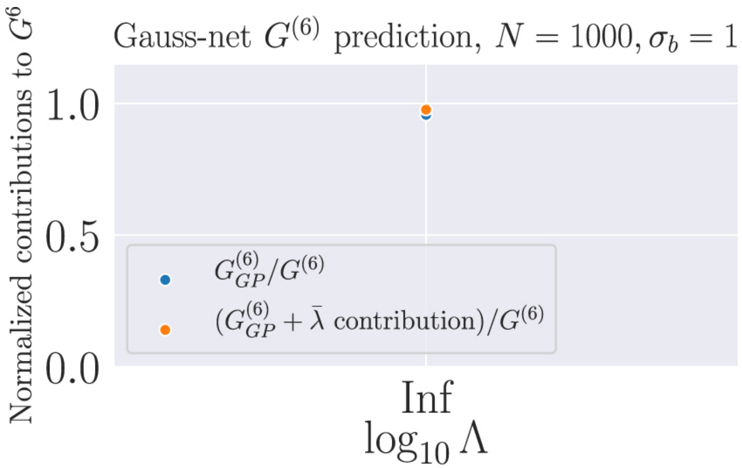
Conjecture: couplings in NGPs associated to neural network architectures are constants (or nearly constants) if the kernel $K(x, y)$ associated with their GP limit is translationally invariant.

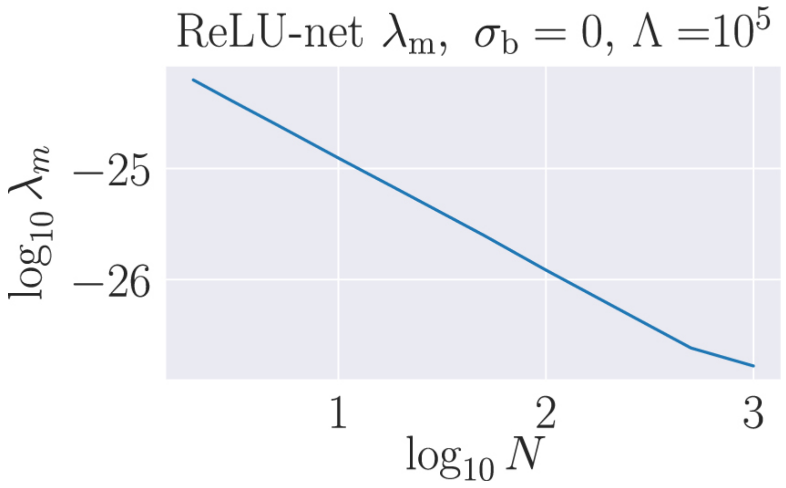
Experimental measurements of coupling constants

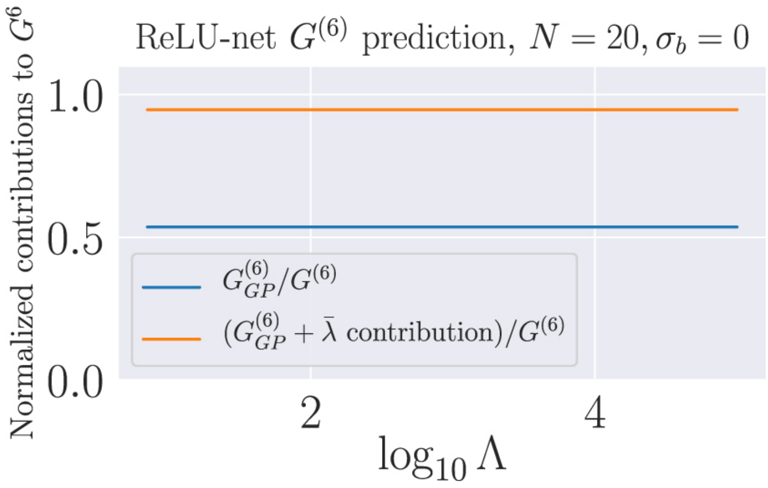


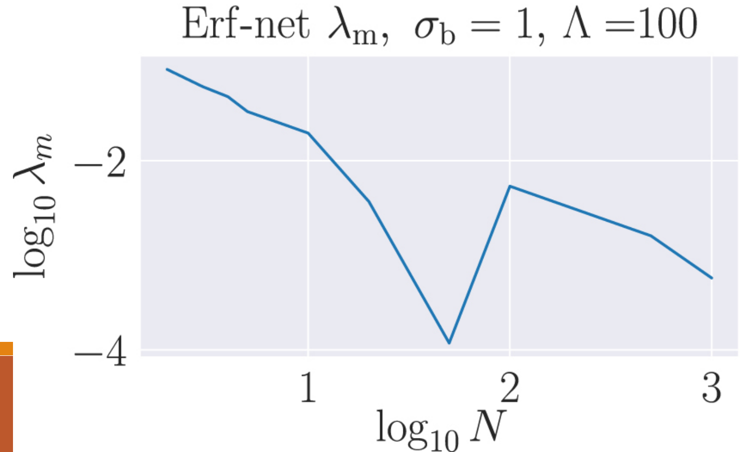
Small

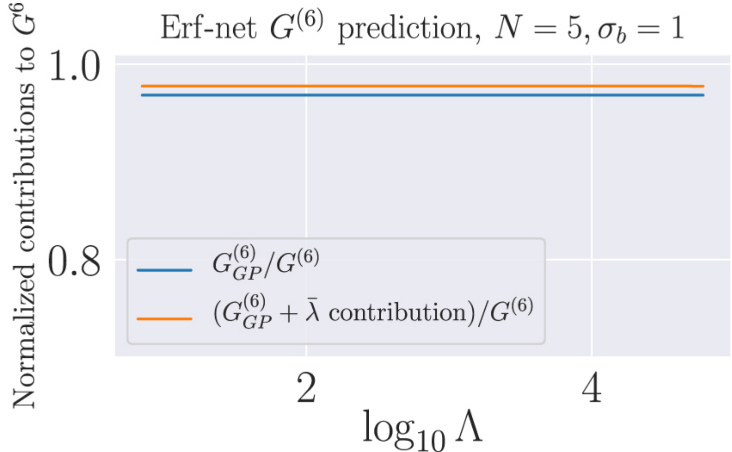

Gauss-net λ_{m}, $\sigma_{\mathrm{b}} = 1$, $\Lambda =$Inf
$\log_{10} \lambda_m$
0
−1
−2
1
2
3
$\log_{10} N$


Gauss-net $G^{(6)}$ prediction, $N = 1000, \sigma_b = 1$
Normalized contributions to G^6
1.0
0.5
0.0
$G^{(6)}_{GP}/G^{(6)}$
$(G^{(6)}_{GP} + \bar{\lambda}$ contribution$)/G^{(6)}$
Inf
$\log_{10} \Lambda$


ReLU-net λ_{m}, $\sigma_{\mathrm{b}} = 0$, $\Lambda = 10^5$
$\log_{10} \lambda_m$
−25
−26
1
2
3
$\log_{10} N$


ReLU-net $G^{(6)}$ prediction, $N = 20, \sigma_b = 0$
Normalized contributions to G^6
1.0
0.5
0.0
$G^{(6)}_{GP}/G^{(6)}$
$(G^{(6)}_{GP} + \bar{\lambda}$ contribution$)/G^{(6)}$
2
4
$\log_{10} \Lambda$


Erf-net λ_{m}, $\sigma_{\mathrm{b}} = 1$, $\Lambda =$100
$\log_{10} \lambda_m$
−2
−4
1
2
3
$\log_{10} N$


Erf-net $G^{(6)}$ prediction, $N = 5, \sigma_b = 1$
Normalized contributions to G^6
1.0
0.8
$G^{(6)}_{GP}/G^{(6)}$
$(G^{(6)}_{GP} + \bar{\lambda}$ contribution$)/G^{(6)}$
2
4
$\log_{10} \Lambda$

Experimental prediction of correlation function

Table 5. Optimized values of λ_0, λ_2, λ_{NL} and resulting MAPE and MSE loss upon application to test-sets.

	$(\lambda_0, \lambda_2, \lambda_{NL})$	Test (MAPE, MSE)
Gauss M_0	$(0.0, 0.0, 0.0)$	$(100, 0.019)$
Gauss M_1	$(0.0046, 0.0, 0.0)$	$(0.0145, 6.8 \times 10^{-10})$
Gauss M_2	$(0.0043, 0.0011, 0.0)$	$(0.0144, 6.7 \times 10^{-10})$
Gauss M_3	$(0.00062, 0.00016, 0.0015)$	$(0.0156, 7.5 \times 10^{-10})$
ReLU M_0	$(0.0, 0.0, 0.0)$	$(100, 0.003)$
ReLU M_1	$(6.2 \times 10^{-11}, 0.0, 0.0)$	$(0.0035, 7.6 \times 10^{-12})$
ReLU M_2	$(1.2 \times 10^{-18}, 8.7 \times 10^{-15}, 0.0)$	$(0.0013, 1.5 \times 10^{-12})$
ReLU M_3	$(1.2 \times 10^{-18}, 8.7 \times 10^{-15}, 6.8 \times 10^{-17})$	$(0.0012, 1.2 \times 10^{-12})$
Erf M_0	$(0.0, 0.0, 0.0)$	$(100, 0.006)$
Erf M_1	$(0.039, 0.0, 0.0)$	$(0.030, 8.3 \times 10^{-10})$
Erf M_2	$(0.040, -0.00043, 0.0)$	$(0.0042, 1.9 \times 10^{-11})$
Erf M_3	$(0.0019, -0.0054, 0.0063)$	$(0.037, 1.1 \times 10^{-9})$

Renormalization, RG flow

- goal of RG: solve divergences arising from integrals over the space of inputs

using cutoff: $\Delta S_\Lambda = \int_{-\Lambda}^{\Lambda} d^{d_{in}}x \sum_{l \leqslant k} g_{\mathcal{O}_l}(\Lambda)\mathcal{O}_l$

RGEs: $\frac{dG^{(n)}(x_1,\ldots,x_n)}{d\log\Lambda} = 0$ $\quad\quad \beta(g_{\mathcal{O}_l}) := \frac{d(g_{\mathcal{O}_l}(\Lambda))}{d\log\Lambda}$

Perturbation Theory

$$G^{(n)}(x_1,\ldots,x_n) = \frac{\int dff_1 \ldots f_n e^{-S}}{Z_0}$$

$$Z_0 = \int dfe^{-S}$$

$$S = S_{GP} + \Delta S$$

for $\mathcal{O}_k := g_k f(x)^k$ in ΔS as $\int d^{d_{in}}x\mathcal{O}_k$

$$G^{(n)}(x_1,\ldots,x_n) = \frac{\int dff(x_1)\ldots f(x_n)\left[1-\int d^{d_{in}}xg_kf(x)^k+O(g_k^2)\right]e^{-S_{GP}}/Z_{GP,0}}{\int df\left[1-\int d^{d_{in}}xg_kf(x)^k+O(g_k^2)\right]e^{-S_{GP}}/Z_{GP,0}}$$

$$\Delta S = \int d^{d_{in}}x\left[\lambda f(x)^4 + \kappa f(x)^6\right]$$