arXiv:2305.02334v1 [hep-th] 3 May 2023

神经网络有效场论的结构

Ian Banta, [1] Tianji Cai, [1] Nathaniel Craig, [1,2] 和 Zhengkang Zhang[1,2]

[1] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*

[2] *Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA*

我们提出了一种图示方法，用于研究与初始化时的深度神经网络相对应的有效场论（EFT），大幅简化了对神经元统计有限宽度修正的计算。有效场论计算的结构清晰地表明，一个条件就能决定神经元预激活的所有连通关联函数是否具有临界性。理解这类有效场论或许有助于推动深度学习和场论模拟的发展。

引言 — 近年来，机器学习（ML）经历了一场革命，应用范围从图像识别和自然语言处理，到自动驾驶汽车和围棋对弈。这些发展的核心是深度神经网络的设计，这是一类由多个人工神经元层组成的机器学习架构。这类网络显然相当复杂，可训练参数数量之多令人望而却步，因此实际应用往往依赖代价高昂的反复试错。尽管如此，研究人员仍在广泛探索，力图揭开这个黑箱。

对这类复杂系统进行理论理解之所以成为可能，是因为人们观察到，广泛的神经网络架构实际上都存在一个简单的极限：当网络宽度（每层的神经元数量）趋于无穷大时，它们会退化为高斯过程 [1–6], ，并在基于梯度的训练过程中演化为由神经切核控制的线性模型 [7–9]。然而，无限宽度的网络在实践中既不存在，也无法为深度学习提供准确的模型。因此，理解有限宽度效应至关重要，而近期已有多种方法对其进行了研究 [10–23]。

机器学习理论的这一研究方向与理论物理学之间存在着引人入胜的协同关系 [24]。尤其是，人们已经认识到，神经网络与（统计或量子）场论之间存在天然的对应关系 [25–34]。无限宽网络——也就是高斯过程——对应于自由理论，而宽网络中的有限宽度修正则可以像弱相互作用理论中那样通过微扰计算。这使得我们能够对神经网络进行可系统改进的刻画，超越少数几种可精确求解的特殊情形 [35–37]。与此同时，从有效理论的角度来看 [21], ，深度神经网络中的信息传播可以理解为重整化群（RG）流。考察 RG 不动点附近的标度行为，可以揭示将网络调节至临界性的方法 [38–40], ；这对于缓解实际应用中臭名昭著的梯度爆炸和梯度消失问题至关重要。反过来，这种协同关系也为利用神经网络研究场论带来了新的机遇 [33]。

受近期进展的启发，在这封短文中，我们进一步探究了典型深度神经网络所对应的有效场论（EFT）的结构。为此，我们发展了一种新颖的图示形式主义。[1] 我们的方法主要建立在参考文献 [21, 22], 的框架之上，这些框架使有限宽度修正的系统计算成为可能。正如我们在正文中简洁地重现已知结果，并在补充材料中给出更多包含新结果的示例所展示的那样，图示形式主义极大地简化了这些计算。有趣的是，重整化群分析中的图结构表明，神经网络有效场论是一种相当特殊的类型，其中单一条件便能支配所有神经元关联函数的临界调节。对这些有效场论的研究，或可为神经网络的性质以及新颖的场论现象带来新的认识。

深度神经网络的有效场论 — 深度神经网络的典型架构——多层感知机，可以定义为由一组神经元构成，其数值 $\phi_i^{(\ell)}$ （称为预激活值）通过以下操作确定，给定输入 $\vec{x} \in \mathbb{R}^{n_0}$:

$$\phi_i^{(1)}(\vec{x}) = \sum_{j=1}^{n_0} W_{ij}^{(1)} x_j + b_i^{(1)} ,$$
$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) + b_i^{(\ell)} \quad (\ell \geq 2) . \quad (1)$$

这里，括号中的上标标记层，下标i, j 标记层内的神经元（第 ℓ层共有 n_ℓ 个），而 $\sigma(\phi)$是激活函数（常见选择包括双曲正切(ϕ)或 ReLU$(\phi) \equiv \max(0, \phi)$）。权重 $W_{ij}^{(\ell)}$ 和偏置 $b_i^{(\ell)}$ $(\ell = 1, \ldots, L)$是网络参数，在训练过程中会对其进行调整以最小化损失函数，从而使训练后的网络能够近似所需函数。

深度神经网络有效场论的基本思想是考虑一个网络系综，其中在初始化时，$W_{ij}^{(\ell)}$ 和 $b_i^{(\ell)}$ 独立地抽取自均值为零的

[1] 另请参阅参考文献 [13, 17, 18, 26, 27, 31, 32, 41]，了解受费曼图启发的机器学习方法。

高斯分布，其方差分别为 $C_W^{(\ell)}/n_{\ell-1}$ 和 $C_b^{(\ell)}$。这个系综的统计特性既描述了以这种方式初始化的神经网络的典型行为，也描述了特定网络可能偏离典型行为的程度。用场论语言来说，这些特性由所有神经元预激活场 $\phi_i^{(\ell)}(\vec{x})$ 的欧几里得作用量 $\mathcal{S}[\phi] = -\log P(\phi)$ 描述，其中 $P(\phi)$ 是联合概率分布。正如我们在补充材料中回顾的那样，在初始化时，每一层的条件概率分布都是高斯分布：

$$P\big(\phi^{(\ell)}\big|\phi^{(\ell-1)}\big) = \big[\det\big(2\pi\mathcal{G}^{(\ell)}\big)\big]^{-\frac{n_\ell}{2}} e^{-\mathcal{S}_0^{(\ell)}}, \tag{2}$$

$$\mathcal{S}_0^{(\ell)} = \int d\vec{x}_1 d\vec{x}_2 \, \frac{1}{2}\sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\phi_i^{(\ell)}(\vec{x}_2), \tag{3}$$

where $\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2) = \frac{1}{n_{\ell-1}}\sum_{j=1}^{n_{\ell-1}} \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2)$, with

$$\begin{aligned}\mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2) &= C_b^{(\ell)} + C_W^{(\ell)}\,\sigma\big(\phi_j^{(\ell-1)}(\vec{x}_1)\big)\,\sigma\big(\phi_j^{(\ell-1)}(\vec{x}_2)\big)\\ &\equiv C_b^{(\ell)} + C_W^{(\ell)}\,\sigma_{j,\vec{x}_1}^{(\ell-1)}\,\sigma_{j,\vec{x}_2}^{(\ell-1)}\end{aligned} \tag{4}$$

对于 $\ell \geq 2$和 $\mathcal{G}_j^{(1)}(\vec{x}_1,\vec{x}_2) = C_b^{(1)} + C_W^{(1)}\,x_{1j}\,x_{2j}$。我们在输入空间中取了连续极限，以便更好地与场论分析相对应。$\big(\mathcal{G}^{(\ell)}\big)^{-1}$ 在 $\mathcal{G}^{(\ell)}$ 不可逆时，被理解为伪逆。我们看到，对于 $\ell \geq 2$，$\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)$ 是（$\ell-1$）第 th 层神经元的算子，因此，式（3）实际上是一个具有层间耦合的相互作用理论。这也意味着，式（2）中的行列式并非恒定的前因子。为计及其影响，我们引入辅助反交换场 ψ，ψ 它们类似于 Faddeev-Popov 过程中的鬼场和反鬼场。将所有层都包括在内，我们得到

$$e^{-\mathcal{S}[\phi]} = \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\; e^{-\sum_{\ell=1}^{L}\left(\mathcal{S}_0^{(\ell)}[\phi]+\mathcal{S}_\psi^{(\ell)}[\phi,\psi,\bar{\psi}]\right)}, \tag{5}$$

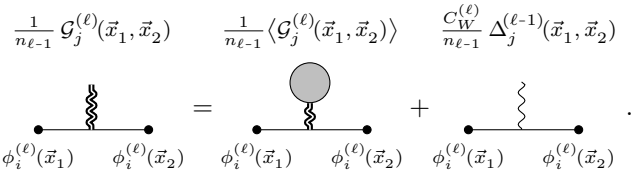
其中 $\mathcal{S}_0^{(\ell)}$ 由上面的公式（3）给出，并且

$$\mathcal{S}_\psi^{(\ell)} = -\int d\vec{x}_1 d\vec{x}_2 \sum_{i'=1}^{n_\ell/2} \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\psi_{i'}^{(\ell)}(\vec{x}_2)\,. \tag{6}$$

第 ℓ层神经元分别通过 $\mathcal{S}_0^{(\ell)}$ 和 $\mathcal{S}_0^{(\ell+1)}$ 与第（$\ell-1$）层和第（$\ell+1$）层神经元相互作用，而与之相关联的鬼场与第（$\ell-1$）层神经元具有符号相反的耦合，但不与第（$\ell+1$）层神经元发生耦合。这意味着，就它们与 $\phi^{(\ell-1)}$ 的耦合而言，$\phi^{(\ell)}$ 和 $\psi^{(\ell)}$ 环路会相互抵消。这是必然的，因为网络具有方向性——给定层的神经元不可能受到更深层中发生的事件的影响。

费曼图中的神经元统计 — 我们感兴趣的是计算神经元统计量，即*i.e.* 连通上述有效场论中神经元预激活场 $\phi_i^{(\ell)}(\vec{x})$ 的关联函数。更准确地说，我们希望追踪神经元关联函数随网络层 ℓ 的演化；网络层编码了信息在深度神经网络中的处理过程，其形式与场论中的重整化群流类似。为此，我们构建了一种高效的图示框架，以递归方式确定 ℓ 层的神经元关联函数，并将其表示为第 $(\ell-1)$ 层神经元关联函数的形式。

从作用量方程 (3) 出发，我们可以推导出以下费曼规则（详情请参见补充材料）：



(7)

如上所述，团块表示对附着于其上的算符（或算符乘积）取期望值。式（7）同时包含了我们通常所说的传播子和顶点：右侧第一项对 j 求和后，就是 $\phi_i^{(\ell)}$ 的完整传播子（或二点关联函数），

$$\big\langle\phi_{i_1}^{(\ell)}(\vec{x}_1)\,\phi_{i_2}^{(\ell)}(\vec{x}_2)\big\rangle = \delta_{i_1 i_2}\big\langle\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\big\rangle\,, \tag{8}$$

而第二项是 $\phi_i^{(\ell)}$ 双线性项与由 $\phi_j^{(\ell-1)}$ 构成的算符之间的相互作用顶点：

$$\Delta_j^{(\ell-1)}(\vec{x}_1,\vec{x}_2) \equiv \sigma_{j,\vec{x}_1}^{(\ell-1)}\,\sigma_{j,\vec{x}_2}^{(\ell-1)} - \big\langle\sigma_{j,\vec{x}_1}^{(\ell-1)}\,\sigma_{j,\vec{x}_2}^{(\ell-1)}\big\rangle \tag{9}$$

对于 $\ell \geq 2$，且 $\Delta_j^{(0)}(\vec{x}_1,\vec{x}_2) = 0$。根据式（7），显然，每个 $\phi^2\Delta$ 顶点都带有一个因子 $\frac{1}{n}$（其中n统称为 $n_1, \cdots, n_{L-1}$）。在无限宽度极限下，$n\to\infty$，有效场论是一种自由理论；而对于较大但有限的 n，我们得到一个弱相互作用理论，其中高点连通关联函数可以通过微扰方式计算，并表示为 $\frac{1}{n}$ 展开。

为了解释其运作方式，我们首先仔细考察二点关联函数公式（8）（由于我们已将 $\int\mathcal{D}\phi\, e^{-\mathcal{S}} = 1$ 归一化，因此它会自动相连，这意味着真空泡的总和为零）。我们可以将其展开为 $\frac{1}{n}$：

$$\big\langle\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\big\rangle = \sum_{p=0}^{\infty}\frac{1}{n_{\ell-1}^p}\,\mathcal{K}_p^{(\ell)}(\vec{x}_1,\vec{x}_2)\,. \tag{10}$$

领先阶（LO）项 $\mathcal{K}_0^{(\ell)}$ 称为核；它是自由理论极限下 $\phi_i^{(\ell)}$ 的传播子 $n\to\infty$。在该极限下计算 $\big\langle\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\big\rangle$，相当于对前述各项使用自由理论传播子 $\mathcal{K}_0^{(\ell-1)}$ ——

层神经元 $\phi_j^{(\ell-1)}$ 在式（7）的团块中：

$$\mathcal{K}_0^{(\ell)}(\vec{x}_1,\vec{x}_2) = \sum_j \frac{1}{n_{\ell-1}} \langle \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle_{\mathcal{K}_0^{(\ell-1)}} = C_b^{(\ell)} + C_W^{(\ell)} \langle \sigma_{\vec{x}_1}\sigma_{\vec{x}_2}\rangle_{\mathcal{K}_0^{(\ell-1)}} . \quad (11)$$

此处下标 $\mathcal{K}_0^{(\ell-1)}$ 表示期望值是用自由理论传播子 $\mathcal{K}_0^{(\ell-1)}$（参见补充材料中的式（A.15））。我们省略了 σ 上的神经元和层指标，因为 $\mathcal{K}_0^{(\ell-1)}$ 下标已指明该层，且该层中所有神经元的期望值都相同。还可以针对激活函数的特定选择进一步计算 $\langle \sigma_{\vec{x}_1}\sigma_{\vec{x}_2}\rangle_{\mathcal{K}_0^{(\ell-1)}}$，$\sigma$但在当前分析中，我们不限定激活函数。

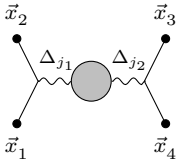
式（11）使我们能够递归地由 $\mathcal{K}_0^{(\ell)}$ 确定 $\mathcal{K}_0^{(\ell-1)}$，这在对无限宽网络的研究中早已广为人知。它也可以看作是 $\mathcal{K}_0$的重整化群流，其紫外边界条件为 $\mathcal{K}_0^{(1)}(\vec{x}_1,\vec{x}_2) = C_b^{(1)} + \frac{C_W^{(1)}}{n_0}\vec{x}_1\cdot\vec{x}_2$。将图示计算推广到 $\mathcal{K}_{p\geq 1}$很直接。我们在补充材料中给出了 $\mathcal{K}_1$ 重整化群流的一个简单推导。

接下来，考虑连通四点关联函数：

$$\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\,\phi_{i_2}^{(\ell)}(\vec{x}_2)\,\phi_{i_3}^{(\ell)}(\vec{x}_3)\,\phi_{i_4}^{(\ell)}(\vec{x}_4)\rangle_{\mathrm{C}} = \delta_{i_1 i_2}\delta_{i_3 i_4}\frac{1}{n_{\ell-1}} V_4^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4) + \text{perms.}, \quad (12)$$

其中

$$\frac{1}{n_{\ell-1}} V_4^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4) = \sum_{j_1,j_2}$$

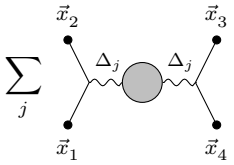


. (13)

从这里开始，我们仅用输入参数 $\vec{x}_1$、$\vec{x}_2$等标记外腿；显然，它们表示指标两两相同的第 ℓ 层神经元。为简洁起见，我们也省略 $\ell-1$ 上的“$(\Delta_j^{(\ell-1)})$”上标，并省去前因子 $\frac{C_W^{(\ell)}}{n_{\ell-1}}$。注意，式（13）中的团块根据定义是自动连通的，因为 $\langle \Delta_j^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle = 0$。

为了计算该图，我们需要考虑两种情况，$j_1 = j_2$ 和 $j_1 \neq j_2$ 对于 $j_1 = j_2 \equiv j$，团块在领先阶取其自由理论值：

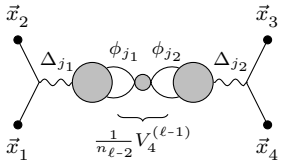
$$\sum_j$$



$$= \frac{(C_W^{(\ell)})^2}{n_{\ell-1}} \langle \Delta(\vec{x}_1,\vec{x}_2)\Delta(\vec{x}_3,\vec{x}_4)\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n^2}\Big). \quad (14)$$

如公式（11）中所示，我们省略了 Δ上的层和神经元指标。对于 $j_1 \neq j_2$，自由理论传播子无法连接 Δ_{j_1} 和Δ_{j_2}，因此主导贡献来自插入（$\ell-1$）层的一个连通四点关联函数：

$$\sum_{j_1,j_2}$$



$$= \frac{(C_W^{(\ell)})^2}{4\,n_{\ell-2}} \prod_{\alpha=1}^4 \int d\vec{y}_\alpha d\vec{z}_\alpha \,(\mathcal{K}_0^{(\ell-1)})^{-1}(\vec{y}_\alpha,\vec{z}_\alpha)\, V_4^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4) \Big\langle \Delta(\vec{x}_1,\vec{x}_2)\,\phi(\vec{z}_1)\,\phi(\vec{z}_2)\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} \Big\langle \Delta(\vec{x}_3,\vec{x}_4)\,\phi(\vec{z}_3)\,\phi(\vec{z}_4)\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n^2}\Big)$$

$$= \frac{(C_W^{(\ell)})^2}{4\,n_{\ell-2}} \prod_{\alpha=1}^4 \int d\vec{y}_\alpha\, V_4^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4) \Big\langle \frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} \Big\langle \frac{\delta^2\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_3)\delta\phi(\vec{y}_4)}\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n^2}\Big). \quad (15)$$

在该图中，内部实线表示 $\phi_{j_1}^{(\ell-1)}$、$\phi_{j_2}^{(\ell-1)}$ 传播子。交换两条 $\phi_{j_1}^{(\ell-1)}$ 线或两条 $\phi_{j_2}^{(\ell-1)}$ 线会得到相同的图，因此对称因子为 $\frac{1}{2^2} = \frac{1}{4}$。中心较小的团块（连同与其相连的传播子）表示（$\ell-1$）层的一个连通四点关联函数，$\frac{1}{n_{\ell-2}} V_{(\ell-1)\,4}$。较大的团块产生了式（15）第二行中的关联函数；由于 $\langle \Delta_j^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle = \langle \phi_j^{(\ell-1)}(\vec{x})\rangle = 0$，它们自动成为连通的。根据标准定义，关联函数 $\langle \phi(\vec{x})\ldots\rangle$ 包含传播子 $\mathcal{K}_0^{(\ell-1)}(\vec{x},\vec{y})\ldots$（其中 $\vec{y}$需要对其积分），因此当我们用关联函数来构建

图时，每条连接两个关联函数（团块）的内部传播子都会被计数两次。因此，为避免重复计数，我们在图中的每条内部线上插入一个逆传播子。这解释了式（$(\mathcal{K}_0^{(\ell-1)})^{-1}$15）第一行中的因子，它们实际上会截去连通四点关联函数（或者等价地，图中较大的团块）上的外部传播子。式（15）中的最终表达式通过 Wick 收缩得到，由此产生的因子 $\mathcal{K}_0^{(\ell-1)}$ 会抵消 $(\mathcal{K}_0^{(\ell-1)})^{-1}$。

将方程（14）与（15）相加，即可得到以 $V_4^{(\ell)}$ 表示为 $V_4^{(\ell-1)}$ 和 $\mathcal{K}_0^{(\ell-1)}$ 的最终结果，即 的重整化群流V_4，与参考文献 [14, 21] 一致。这两个方程均为$\mathcal{O}(\frac{1}{n})$，因此由方程（ $V_{(\ell)\,4}$ 定义的12）为 $\mathcal{O}$(1)。

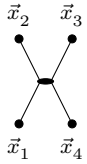
图示计算可直接推广至高点连通关联函数，并为系统分析深度神经网络中的有限宽度效应提供了简洁的框架。在补充材料中，我们还给出了新的六点和八点连通关联函数结果，作为进一步的示例。

重整化群流也可以在有效场论作用量层面进行表述。其思路是考虑一系列有效场论， $\mathcal{S}^{(\ell)}$ eff($\ell = 1, \ldots\ldots, L$)，这些有效场论是通过积分掉除第 ℓ 层以外所有层中的神经元和鬼场得到的。它们具有以下形式：

$$
\begin{aligned}
\mathcal{S}_{\text{eff}}^{(\ell)} = & \int d\vec{x}_1 d\vec{x}_2 \left(\mathcal{K}_0^{(\ell)} + \mu^{(\ell)}\right)^{-1}(\vec{x}_1, \vec{x}_2) \\
& \left[\frac{1}{2}\,\phi_i^{(\ell)}(\vec{x}_1)\,\phi_i^{(\ell)}(\vec{x}_2) - \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\,\psi_{i'}^{(\ell)}(\vec{x}_2)\right] \\
& - \int d\vec{x}_1 d\vec{x}_2 d\vec{x}_3 d\vec{x}_4\, \lambda^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{x}_3, \vec{x}_4) \\
& \left[\frac{1}{8}\,\phi_i^{(\ell)}(\vec{x}_1)\,\phi_i^{(\ell)}(\vec{x}_2)\,\phi_j^{(\ell)}(\vec{x}_3)\,\phi_j^{(\ell)}(\vec{x}_4)\right. \\
& -\frac{1}{2}\,\phi_i^{(\ell)}(\vec{x}_1)\,\phi_i^{(\ell)}(\vec{x}_2)\,\bar{\psi}_{j'}^{(\ell)}(\vec{x}_3)\,\psi_{j'}^{(\ell)}(\vec{x}_4) \\
& \left.+\frac{1}{2}\,\bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\,\psi_{i'}^{(\ell)}(\vec{x}_2)\,\bar{\psi}_{j'}^{(\ell)}(\vec{x}_3)\,\psi_{j'}^{(\ell)}(\vec{x}_4)\right] \\
& + \cdots
\end{aligned}
\tag{16}
$$

其中默认对重复指标求和。有效场论耦合 $\mu^{(\ell)}$, $\lambda^{(\ell)} \sim \mathcal{O}(\frac{1}{n})$ 可以通过连通关联函数确定，因此，它们的重整化群流可直接由上文讨论的后者的重整化群流推导得到。例如，通过匹配连通四点关联函数，可以建立 $\lambda^{(\ell)}$ 与 $V_4^{(\ell)}$ 和 $\mathcal{K}_0^{(\ell)}$ 之间的关系：

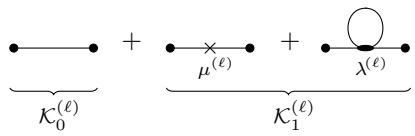
$$
\frac{1}{n_{\ell-1}}\,V_4^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{x}_3, \vec{x}_4) =
$$



$$
+ \mathcal{O}\left(\frac{1}{n^2}\right)
$$

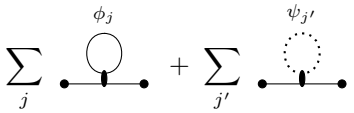
$$
= \prod_{\alpha=1}^{4} \int d\vec{y}_\alpha \mathcal{K}_0^{(\ell)}(\vec{x}_\alpha, \vec{y}_\alpha) \lambda^{(\ell)}(\vec{y}_1, \vec{y}_2; \vec{y}_3, \vec{y}_4) + \mathcal{O}\left(\frac{1}{n^2}\right), \tag{17}
$$

其中，我们用拉长的顶点表示 $\lambda^{(\ell)}$ 的四个自变量的配对。对于二点关联函数，计算涉及以下图：



(18)

最后一个图中四次顶点处腿的另一种配对会产生一个 $\mathcal{O}(\frac{n}{n})$ 贡献，但由于耦合符号相反，该贡献会被含有鬼场圈图的图抵消：



(19)

类似的抵消也解释了为何在式（ $\mathcal{O}(\frac{n}{n^2})$17）中计算连通四点关联函数时要排除圈图。[2]

重整化群流与临界性的结构 ——神经元统计的重整化群分析与深度神经网络的临界调节高度相关。在实际的深度学习应用中，调节的必要性早已得到认可，尤其是在缓解臭名昭著的梯度爆炸和梯度消失问题方面；在有限机器精度下，这些问题会使深度网络难以训练。在有效场论框架中，这与以下事实有关：超参数 $C_b^{(\ell)}$，$C_W^{(\ell)}$ 的一般选择会导致神经元关联函数在重整化群流下呈指数标度。要抑制这种指数行为，就需要通过审慎设置这些超参数 [38–40]，将网络调节到临界性。在线性核层面，参考文献 [21] 的临界性分析揭示了两个突出的普适类，各种激活函数的网络都属于其中之一：尺度不变类（包括例如 ReLU）和 $\mathcal{K}^\star = 0$ 类（包括例如 双曲正切）。在每种情况下， $\mathcal{K}_0^{(\ell)}$ 都会随着 ℓ 的增加而流向一个非平凡不动点；关键在于，不动点附近的标度呈幂律而非指数形式，这使信息能够通过各层传播，从而使网络可以从数据中学习非平凡特征。

虽然以往的临界性分析主要集中于二点关联函数，但研究高点关联函数同样十分重要，因为它们编码了系综中的涨落。换句话说，它并不

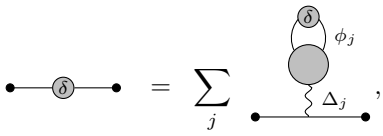
[2] 通过从参考文献中积分掉鬼场，我们可以重现其有效理论。 [21] 对式（16）中的鬼场进行积分。基于这一无鬼场有效理论的计算会产生 $\mathcal{O}(\frac{n_{\ell-1}}{n_\ell})$ 项，因此必须要么在 $n_\ell \gg n_{\ell-1} \gg 1$ 的范围内进行计算，要么将作用量对除 $\mathcal{O}$个神经元以外的所有神经元进行边缘化，这些神经元位于第（$\ell-1$）层，从而获得微扰$\frac{1}{n}$ 展开。保留鬼场可以避免这些微妙问题，使 $\mu^{(\ell)}$ 真正成为 $\mathcal{O}(\frac{1}{n})$，并使 $\lambda^{(\ell)}$ 与（截肢后的）连通四点关联子真正相差 $\mathcal{O}(\frac{1}{n^2})$。

只需要求网络在平均意义上表现良好是不够的，每个网络的标度行为还必须接近平均水平。乍一看，如果要求所有高点关联函数在任意输入点处都呈现幂律标度，那么临界性似乎施加的约束数量超过了可调超参数的数量。然而，正如我们将要展示的，重整化群流的结构体现在图解形式中，其特性使得将$\mathcal{K}_0$调节至临界性实际上能够确保不动点附近所有高点连通关联函数均呈现幂律标度。

我们先从二点关联函数开始。渐近标度行为（指数还是幂律）可以通过以下问题推断：在第$(\ell-1)$层进行无穷小变化时，

$$\langle \mathcal{G}^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle \to \langle \mathcal{G}^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle + \delta\langle \mathcal{G}^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle\,, \tag{20}$$

$\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\ \vec{x}_2)\rangle$会如何变化？用图表示，这可以按如下方式计算：



(21)

其中，标记为“δ”的团块表示（二点）关联函数的变化。结果为

$$\delta\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle = \int d\vec{y}_1 d\vec{y}_2\, \chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\, \delta\langle \mathcal{G}^{(\ell-1)}(\vec{y}_1,\vec{y}_2)\rangle\,, \tag{22}$$

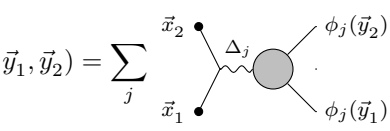
或者，等价地：

$$\frac{\delta\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle}{\delta\langle \mathcal{G}^{(\ell-1)}(\vec{y}_1,\vec{y}_2)\rangle} = \chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\,, \tag{23}$$

其中

$$\begin{aligned}
&\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\\
&= \frac{C_W^{(\ell)}}{2}\int d\vec{z}_1 d\vec{z}_2\, \big(\mathcal{K}_0^{(\ell-1)}\big)^{-1}(\vec{y}_1,\vec{z}_1)\, \big(\mathcal{K}_0^{(\ell-1)}\big)^{-1}(\vec{y}_2,\vec{z}_2)\\
&\qquad \Big\langle \Delta(\vec{x}_1,\vec{x}_2)\,\phi(\vec{z}_1)\,\phi(\vec{z}_2)\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n}\Big)\\
&= \frac{C_W^{(\ell)}}{2}\Big\langle \frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n}\Big)\,.
\end{aligned} \tag{24}$$

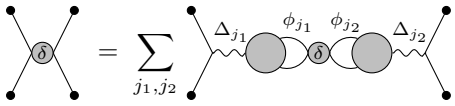
我们可以清楚地看到它与上文公式（15）在结构上的相似性。归根结底，两个公式中都出现了相同的子图，因此我们可以写成：



$$\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2) = \sum_j \quad , \tag{25}$$

其中的$\phi_j^{(\ell-1)}$腿已被截肢，并且假定完整的图中这些腿之间存在交换对称性（因此对称因子为1_2，见式（24）。

对于高点连通关联函数，同样的模式依然成立。对于连通四点关联函数，$V_{(\ell-1)\,4}(\vec{x}_1,\vec{x}_2;\ \vec{x}_3,\vec{x}_4)$的无穷小变化会导致$V_{(\ell)\,4}(\vec{x}_1,\vec{x}_2;\ \vec{x}_3,\vec{x}_4)$发生变化：



$$= \sum_{j_1,j_2} \quad , \tag{26}$$

并且我们发现：

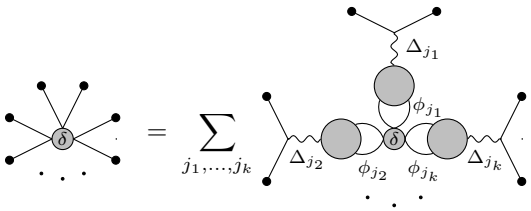
$$\begin{aligned}
&\frac{n_{\ell-2}}{n_{\ell-1}}\frac{\delta V_4^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4)}{\delta V_4^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)}\\
&= \frac{1}{2}\Big[\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\,\chi^{(\ell)}(\vec{x}_3,\vec{x}_4;\vec{y}_3,\vec{y}_4)\\
&\qquad + \chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_3,\vec{y}_4)\,\chi^{(\ell)}(\vec{x}_3,\vec{x}_4;\vec{y}_1,\vec{y}_2)\Big]\,,
\end{aligned} \tag{27}$$

其中对称化形式出现是因为$V_{(\ell)\,4}\ (\vec{x}_1,\vec{x}_2;\ \vec{x}_3,\vec{x}_4) = V_{(\ell)\,4}\ (\vec{x}_3,\vec{x}_4;\ \vec{x}_1,\vec{x}_2)$.

一般而言，连通$2k$点关联函数定义为

$$\begin{aligned}
&\big\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\ \dots\ \phi_{i_{2k}}^{(\ell)}(\vec{x}_{2k})\big\rangle_{\mathrm{C}}\\
&= \delta_{i_1 i_2}\cdots\delta_{i_{2k-1}i_{2k}}\,\frac{1}{n_{\ell-1}^{k-1}}\,V_{2k}^{(\ell)}(\vec{x}_1,\vec{x}_2;\dots;\vec{x}_{2k-1},\vec{x}_{2k})\\
&\quad + \text{perms.},
\end{aligned} \tag{28}$$

其中$V_{2k}^{(\ell)}$可以证明是$\mathcal{O}(1)$[22]。其变化由以下公式得出：



$$= \sum_{j_1,\dots,j_k} \quad , \tag{29}$$

并且我们有

$$\begin{aligned}
&\left(\frac{n_{\ell-2}}{n_{\ell-1}}\right)^{k-1}\frac{\delta V_{2k}^{(\ell)}(\vec{x}_1,\vec{x}_2;\dots;\vec{x}_{2k-1},\vec{x}_{2k})}{\delta V_{2k}^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\dots;\vec{y}_{2k-1},\vec{y}_{2k})}\\
&= \mathrm{sym.}\left[\prod_{k'=1}^{k}\chi^{(\ell)}(\vec{x}_{2k'-1},\vec{x}_{2k'};\vec{y}_{2k'-1},\vec{y}_{2k'})\right],
\end{aligned} \tag{30}$$

其中，“对称化”表示以与式（27）相同的方式对该表达式进行对称化。

上述方程中的量$\chi^{(\ell)}(\vec{x}_1,\ \vec{x}_2;\ \vec{y}_1,\ \vec{y}_2)$是对平行和垂直磁化率的一种推广，$\chi_\parallel$以及$\chi_\perp$，这两个量是在文献[21]中分析两个邻近输入这一特殊情况时引入的。在邻近输入极限下，将网络调节至临界状态意味着调整超参数$C_W^{(\ell)}$，$C_b^{(\ell)}$使得核递归方程（11）具有一个不动点$\mathcal{K}^\star$

其中 $\chi_{\parallel} = \chi_{\perp} = 1$。在补充材料中，我们证明，至少对于尺度不变普适类和 $\mathcal{K}^\star = 0$ 普适类，这种调节实际上意味着满足一个更强的条件（在 $\frac{1}{n}$ 的领先阶）：

$$\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\Big|_{\mathcal{K}_0^{(\ell-1)}=\mathcal{K}^\star} = \frac{1}{2}\Big[\delta(\vec{x}_1-\vec{y}_1)\,\delta(\vec{x}_2-\vec{y}_2)+\delta(\vec{x}_1-\vec{y}_2)\,\delta(\vec{x}_2-\vec{y}_1)\Big]. \quad (31)$$

式（31）确保围绕不动点的微扰在各层之间保持不变，这不仅适用于二点关联函数，$\delta\langle\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle = \delta\langle\mathcal{G}^{(\ell-1)}(\vec{x}_1,\vec{x}_2)\rangle$，也适用于整个高点连通关联函数塔，$\delta V^{(\ell)}_{2k}(\vec{x}_1,\ldots,\vec{x}_{2k}) = \frac{1}{n_{\ell-2}^{k-1}}\,\delta V^{(\ell-1)}_{2k}(\vec{x}_1,\ldots,\vec{x}_{2k})$。这进而意味着，一旦满足式（31）这一单一条件，所有这些关联函数的重整化群流都会以幂律而非指数的方式趋向不动点。上述讨论清楚地表明，临界性下高点连通关联函数的幂律标度（此前参考文献 [21, 22] 已在退化输入极限下观察到，直至八点阶）根植于有效场论相互作用的结构中，这一点体现在式（21）、（26）和（29）中的图所共有的结构上。

总结与展望 — 在这封信中，我们引入了一种图示形式主义，大幅简化了与典型深度神经网络相对应的有效场论中有限宽度效应的微扰计算。简洁地重现已知结果并推导出新结果，凸显了图示方法的高效；而引入鬼场则极大地简化了 $\frac{1}{n}$ 有效场论作用量中的计数。我们的分析还清晰揭示了这类有效场论的结构，这些结构是深度神经网络中临界调节取得成功的基础。事实上，在对神经元预激活值的所有高点连通关联函数进行重整化群分析时，都会出现一种普适的图示结构。这意味着，初始化时所有神经元统计量的临界性（即 幂律而非指数标度）都由单一条件决定，见式（31）。

从深度学习的角度来看，显而易见的下一步是扩展图示形式主义，将基于梯度的训练纳入其中，并简化涉及神经切核 [7, 8]及其微分 [11–13, 21]的微扰计算。从基础物理学的角度来看，我们希望能从神经网络与场论之间的紧密联系中获得更多认识。理解其他神经网络架构（例如 循环神经网络 [32, 42] 和 Transformer [43]）所对应的有效场论结构，将有助于我们进一步洞察这种联系，并可能启发用于模拟场论的新型机器学习架构设计。

致谢 ——我们特别感谢 Sho Yaida 在整个过程中给予的有益讨论。

在这项工作过程中。我们感谢 Hannah Day、Marat Freyt-sis、Boris Hanin、Yonatan Kahn 和 Anindita Maiti 对初稿提出的有益讨论和意见，并感谢 Guy Gur-Ari 进行相关讨论。本工作中的费曼图使用 tikz-feynman [44] 绘制。本工作部分由美国能源部资助，资助编号为 DE-SC0011702。本工作部分在阿斯彭物理中心开展，该中心由美国国家科学基金会资助（资助编号 NSF PHY-2210452）；另有部分在卡弗里理论物理研究所开展，该研究所由美国国家科学基金会资助（资助编号 NSF PHY-1748958）。

[1]R. M. Neal, 《无限网络的先验》, 载于 *Bayesian Learning for Neural Networks* (Springer New York, 纽约州纽约市, 1996 年), 第 29–53 页。[2] C. Williams, 《用无限网络进行计算》, 载于 *Advances in Neural Information Processing Systems* 9, , 第 [3] 卷, M. Mozer、M. Jordan 和 T. Petsche 编（MIT Press, 1996 年）。00165 [J. Lee、Y. Bahri、R. Novak、S. S. Schoenholz、J. Pennington 和 J. Sohl-Dickstein, 《深度神经网络作为高斯过程》, arXiv:1711.stat.ML[4] 。1804 11271 A. G. d. [ ] G. Matthews、M. Rowland、J. Hron、R. E. Turner 和 Z. Ghahramani, 《宽深度神经网络中的高斯过程行为》, arXiv:[5] 。stat.ML 。12478 [G. Yang, 《张量程序 I：任意架构的宽前馈或循环神经网络都是高斯过程》, arXiv:1910.]cs.NE[6] 。01562 [B. Hanin, 《无限宽度极限下的随机神经网络作为高斯过程》, arXiv:2107.] math.PR[7] 。07572 [A. Jacot、F. Gabriel 和 C. Hongler, 《神经切核：神经网络中的收敛与泛化》, 《神经信息处理系统进展》第 31 卷（2018 年）, arXiv:1806.]cs.LG[8] 。06720 [J. Lee、L. Xiao、S. Schoenholz、Y. Bahri、R. Novak、J. Sohl-Dickstein 和 J. Pennington, 《任意深度的宽神经网络在梯度下降下演化为线性模型》, 《神经信息处理系统进展》第 32 卷（2019 年）, arXiv:1902.] stat.ML[9] 。14548 [G. Yang, 《张量程序 II：适用于任意架构的神经切核》, arXiv:2006.]stat.ML[10] 。10030 [J. M. Antognini, 《神经网络高斯过程的有限尺寸修正》, arXiv:1908.]cs.LG[11] 。1909 05989 B. Hanin [ ] 和 M. Nica, 《神经切核的有限深度与宽度修正》, arXiv:[12] cs.LG 。*International conference on machine learning* 2020 4542 J. Huang 和 H.-T. Yau, 《深度神经网络的动力学与神经切层级》, 载于 08156 [（PMLR, ）, 第 – 4551 页, arXiv:1909.]cs.LG[13] 。11304 [E. Dyer 和 G. Gur-Ari, 《宽网络的渐近性：来自费曼图的洞见》, arXiv:1909.]cs.LG[14] 。00019 [S. Yaida, 《有限宽度下的非高斯过程与神经网络》, arXiv:1910.]stat.ML[15] 。G. Naveh、O. B. David、H. Sompolinsky 和 Z. Ringel, 《预测使用噪声梯度训练的有限深度神经网络的输出》, 《Physical Review E》104,

064301 (2021), arXiv:2004.01190 [stat.ML]. [16] I. Ser oussi、G. Naveh 和 Z. Ringel，尺度分离与某些卷积神经网络中特征学习的热力学描述，arXiv:2112.15383 [stat.ML]. [17] K. Aitken 和 G. Gur-Ari，具有多项式激活函数的宽网络的渐近性，arXiv: 2006 06687 . [ ] cs.LG . [18] A. Andreassen 和 E. Dyer，宽卷积神经网络的渐近性，arXiv:2008.08675 [cs.LG]. [19] J. Zavatone-Veth、A. Canatar、B. Ruben 和 C. Pehle-van，有限贝叶斯神经网络中表征学习的渐近性，《神经信息处理系统进展》34，24765（2021），arXiv:2106.00651 [cs.LG]. [20] G. Naveh 和 Z. Ringel，一种自洽理论：高斯过程捕捉有限卷积神经网络中的特征学习效应，《神经信息处理系统进展》34，21352（2021），arXiv:2106.04110 [cs.LG]. [21] D. A. Roberts、S. Yaida 和 B. Hanin， *The Principles of Deep Learning Theory* （剑桥大学出版社，2022） arXiv:2106.10165 [cs.LG]. [22] B. Hanin，有限宽度下随机全连接神经网络中的关联函数，arXiv: 2204 01058 . [ ] math.PR . [23] S. Yaida，元原则超参数缩放策略族，arXiv:2210.04909 [cs.LG]. [24] D. A. Roberts，为什么人工智能困难而物理学简单？，arXiv:2104.00008 [hep-th]. [25] S. S. Schoenholz、J. Pennington 和 J. Sohl-Dickstein，随机神经网络与统计场论之间的对应关系，arXiv:1710.06570 [stat.ML]. [26] O. Cohen、O. Malka 和 Z. Ringel，过参数化深度神经网络的学习曲线：场论视角，Physical Review Research 3，023034（2021），arXiv:1906.05301 [cs.LG]. [27] J. Halverson、A. Maiti 和 K. Stoner，神经网络与量子场论，Mach. Learn. Sci. Tech. 2，035002 2021 2008 ( )，arXiv: .08601 [cs.LG]. [28] D. Bachtis、G. Aarts 和 B. Lucini，量子场论机器学习，Phys. Rev. D 103，074510（2021），arXiv:2102.09449 [hep-lat]. [29] A. Maiti、K. Stoner 和 J. Halverson，通过对偶实现对称性：由参数空间关联函数得到的不变神经网络密度，arXiv:2106.00694 [cs.LG]. [30] J. Erdmenger、K. T. Grosvenor 和 R. Jefferson，迈向信息流的量化：深度神经网络与重整化群中的相对熵，Sci-Post Phys. 12，041（2022），arXiv:2107.06898 [hep-th]. [31] H. Erbin、V. Lahoche 和 D. O. Samary，非-

神经网络-量子场论对应关系的微扰重整化，Mach. Learn. Sci. Tech. 3, 015027 (2022)，arXiv:2108.01403 [hep-th]。[32] K. T. Grosvenor 和 R. Jefferson，混沌边缘：量子场论与深度神经网络，SciPost Physics 12, 081 (2022)，arXiv:2109.13247 [hep-th]。[33] J. Halverson，由神经元构建量子场论，arXiv:2112.04527 [hep-th]。[34] H. Erbin、V. Lahoche 和 D. O. Samary，神经网络-量子场论对应关系中的重整化，arXiv:2212.11811 [hep-th]。[35] J. Zavatone-Veth 和 C. Pehlevan，有限贝叶斯神经网络的精确边缘先验分布，神经信息处理系统进展 34, 3364 2021 2104 ( )，arXiv: . 11734 [cs.LG]。[36] L. Noci、G. Bachmann、K. Roth、S. Nowozin 和 T. Hofmann，深度 ReLU 网络先验预测分布的精确刻画，神经信息处理系统进展 34, 20851 (2021)，arXiv:2106.06615 [cs.LG]。[37] B. Hanin 和 A. Zlokapa，利用深度线性网络进行贝叶斯插值，arXiv:2212.14457 [stat.ML]。[38] M. Raghu、B. Poole、J. Kleinberg、S. Ganguli 和 J. Sohl-Dickstein，论深度神经网络的表达能力，载于 *international conference on machine learning* 2017 2847 2854, 1606 (PMLR,) pp. – arXiv: . 05336 [ ] stat.ML 。

[39] B. Poole、S. Lahiri、M. Raghu、J. Sohl-Dickstein 和 S. Ganguli，通过瞬态混沌实现深度神经网络中的指数级表达能力，神经信息处理系统进展 **29** 2016 1606 05340 ( ), [ ] arXiv: . stat.ML 。[40] S. S. Schoenholz、J. Gilmer、S. Ganguli 和 J. Sohl-Dickstein，深度信息传播，arXiv:1611.01232 [stat.ML]。[41] A. Maloney、D. A. Roberts 和 J. Sully，可解的神经网络缩放定律模型，arXiv:2210.16859 [cs.LG]。[42] K. Segadlo、B. Epping、A. van Meegen、D. Dahmen、M. Kr¨amer 和 M. Helias，深度神经网络与循环神经网络的统一场论，arXiv:2112.05589 [cond-mat.dis-nn]。[43] E. Dinan、S. Yaida 和 S. Zhang，Transformer 初始化时的有效理论，(2023 2304 02034)，arXiv: . [ ] cs.LG 。[44] J. Ellis，TikZ-Feynman：使用 TikZ 绘制费曼图，计算物理通讯 210, 103 (2017)，arXiv:1601.05437 [hep-ph]。

补充材料

有效场论作用量

本节详细介绍推导多层感知机有效场论作用量的中间步骤，该作用量由公式（5）给出。推导的初始步骤紧密遵循参考文献第4章 [21]。我们从公式（1）中的运算开始，这些运算递归地根据第 ℓ 层预激活值确定（$\ell-1$）层预激活值：

$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) + b_i^{(\ell)}\,, \tag{A.1}$$

其中约定，当 $\ell = 1$时，应将 $\sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big)$ 替换为 x_j。考虑一个网络系综，其中权重 $W_{(\ell)ij}$和偏置 $b_i^{(\ell)}$在初始化时分别独立地从某些概率分布 $P_{(\ell)W}$和 $P_b^{(\ell)}$中抽取，此时我们可以写出以下条件概率分布（暂时将 $\vec{x}$视为离散变量）：

$$P\big(\phi^{(\ell)}\big|\phi^{(\ell-1)}\big) = \prod_{i,j}\int dW_{ij}\, P_W^{(\ell)}(W_{ij}) \prod_i \int db_i\, P_b^{(\ell)}(b_i) \prod_{i,\vec{x}} \delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell-1}} W_{ij}\,\sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) - b_i\Big)\,. \tag{A.2}$$

当 $P_{(\ell)W}$和 $P_b^{(\ell)}$服从高斯分布时，可以对权重和偏置进行解析积分。具体而言，假设它们分别是方差为 $C_W^{(\ell)}/n_{\ell-1}$和 $C_b^{(\ell)}$ 的零均值高斯分布：

$$P_W^{(\ell)}(W) = \frac{1}{\sqrt{2\pi C_W^{(\ell)}/n_{\ell-1}}} \exp\left(-\frac{W^2}{2C_W^{(\ell)}/n_{\ell-1}}\right), \qquad P_b^{(\ell)}(b) = \frac{1}{\sqrt{2\pi C_b^{(\ell)}}} \exp\left(-\frac{b^2}{2C_b^{(\ell)}}\right), \tag{A.3}$$

并将式（A.2）中的狄拉克δ函数重写为

$$\delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell-1}} W_{ij}\sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) - b_i\Big) = \int \frac{d\Lambda_i(\vec{x})}{2\pi}\, \exp\Big[i\,\Lambda_i(\vec{x})\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell-1}} W_{ij}\sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) - b_i\Big)\Big] \tag{A.4}$$

我们可以配方并积分消去 W_{ij}， b_i。最后，对辅助变量 $\Lambda_i(\vec{x})$ 进行积分（此时被积函数再次呈现为高斯分布），并取连续极限，即可得到正文中的公式（2），为方便起见，在此重新列出：

$$P\big(\phi^{(\ell)}\big|\phi^{(\ell-1)}\big) = \Big[\det\Big(2\pi\mathcal{G}^{(\ell)}\Big)\Big]^{-\frac{n_\ell}{2}} \exp\left[-\int d\vec{x}_1 d\vec{x}_2\, \frac{1}{2}\sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\phi_i^{(\ell)}(\vec{x}_2)\right], \tag{A.5}$$

其中

$$\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv \frac{1}{n_{\ell-1}}\sum_{j=1}^{n_{\ell-1}} \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2)\,, \qquad \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv C_b^{(\ell)} + C_W^{(\ell)}\,\sigma_{j,\vec{x}_1}^{(\ell-1)}\,\sigma_{j,\vec{x}_2}^{(\ell-1)} = \mathcal{G}_j^{(\ell)}(\vec{x}_2,\vec{x}_1)\,, \tag{A.6}$$

其中， $\sigma_{j\ ,\ \vec{x}_1}^{(\ell-1)} \equiv \sigma\big(\phi_j^{(\ell-1)}(\vec{x}_1)\big)$ 应替换为 x_{1j} （ $\vec{x}_1$的第j个分量），对于 $\sigma_{j\ \vec{x}_2}^{(\ell-1)}$，当 $\ell = 1$时，同样如此。（伪）逆 $\big(\mathcal{G}^{(\ell)}\big)^{-1}$ 满足：

$$\int d\vec{x}_1 d\vec{x}_2\, \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_1')\,\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\mathcal{G}^{(\ell)}(\vec{x}_2,\vec{x}_2') = \mathcal{G}(\vec{x}_1',\vec{x}_2')\,. \tag{A.7}$$

受上述步骤与 Faddeev-Popov 过程之间表面类比的启发（其中式（A.1）起到“选择规范”的作用），接下来我们将泛函行列式重写为对一组反对易场的路径积分，即鬼场和反鬼场（这一过程在其他语境中也称为超对称化）：

$$\Big[\det\Big(2\pi\mathcal{G}^{(\ell)}\Big)\Big]^{-\frac{n_\ell}{2}} = \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\, \exp\left[\sum_{i'=1}^{n_\ell/2} \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\psi_{i'}^{(\ell)}(\vec{x}_2)\right]. \tag{A.8}$$

考虑所有层 $\ell = 1, \ldots, L$，只需将条件概率分布相乘：

$$e^{-\mathcal{S}} = P\big(\phi^{(1)}, \ldots, \phi^{(L)}\big) = P\big(\phi^{(1)}\big)\, P\big(\phi^{(2)}|\phi^{(1)}\big)\, \ldots\, P\big(\phi^{(L)}|\phi^{(L-1)}\big)\,, \tag{A.9}$$

并由此得到正文中的公式（5）。请注意，鬼场和反鬼场的总数与L神经元的数量相同， $\sum n_{\ell}$。因此，尽管随机初始化的权重和偏置在ℓ=1每一层引入了独立的随机性，却不存在具有独立涨落的物理“场”自由度。（Faddeev-Popov 过程中的类似表述是，考虑一族规范并不会向规范理论中引入新的物理自由度。）

费曼规则和图示计算的一般步骤

<escape></escape>为了证明式（<escape></escape>7<escape></escape>）中的费曼规则是正确的，我们首先注意到，如果 <escape></escape> $\phi_j^{(\ell-1)}$ <escape></escape>（<escape></escape>x<escape></escape>）是经典背景场，那么对于 <escape></escape> $\phi_i^{(\ell)}$ <escape></escape>（<escape></escape>$\vec{x}$ <escape></escape>）我们就只有一个自由理论，其传播子为 <escape></escape> $\mathcal{G}^{(\ell)}$<escape></escape>（<escape></escape>$\vec{x}_1$<escape></*escape*>, <*escape*></escape> $\vec{x}_2$<escape></escape>）。在这种情况下，二点关联函数由下式给出：<escape></escape>

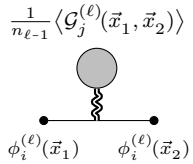
$$\big\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\, \phi_{i_2}^{(\ell)}(\vec{x}_2) \big\rangle = \delta_{i_1 i_2}\, \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)\,, \tag{A.10}$$

并且，通过简单的 Wick 收缩，四点关联函数可表示为

$$\big\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\, \phi_{i_2}^{(\ell)}(\vec{x}_2)\, \phi_{i_3}^{(\ell)}(\vec{x}_3)\, \phi_{i_4}^{(\ell)}(\vec{x}_4) \big\rangle = \delta_{i_1 i_2}\delta_{i_3 i_4}\, \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)\, \mathcal{G}^{(\ell)}(\vec{x}_3, \vec{x}_4) + \text{perms.} \tag{A.11}$$

实际上， $\phi_j^{(\ell-1)}$ （$\vec{x}$）并非经典背景场，而是存在统计涨落。这意味着我们应对方程（A.10）和（A.11）右端的表达式取系综平均，其图示表示为：

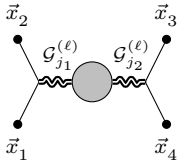
$$\big\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\, \phi_{i_2}^{(\ell)}(\vec{x}_2) \big\rangle = \delta_{i_1 i_2} \big\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \big\rangle = \delta_{i_1 i_2} \sum_j$$



, (A.12)

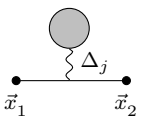
$$\big\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\, \phi_{i_2}^{(\ell)}(\vec{x}_2)\, \phi_{i_3}^{(\ell)}(\vec{x}_3)\, \phi_{i_4}^{(\ell)}(\vec{x}_4) \big\rangle = \delta_{i_1 i_2}\delta_{i_3 i_4} \big\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)\, \mathcal{G}^{(\ell)}(\vec{x}_3, \vec{x}_4) \big\rangle + \text{perms.}$$

$$= \delta_{i_1 i_2}\delta_{i_3 i_4} \sum_{j_1, j_2}$$



+ perms. (A.13)

同样的讨论也适用于高点关联函数。最终，只需规定方程（7）左端的规则，并要求每条双波浪线都必须连接到一个团块（即它不能是外线）。此外，由于我们主要关注连通关联函数，将$\mathcal{G}_j^{(\ell)}$ （$\vec{x}_1$，$\vec{x}_2$）分解为期望值部分和涨落部分（如方程（7）右端所示）会很方便；前者自动与图的其余部分断开，因此我们的计算中只需显式考虑后者。这种分解带来的另一项简化源于涨落部分没有蝌蚪图：



$$= \frac{C_W^{(\ell)}}{n_{\ell-1}} \big\langle \Delta_j^{(\ell-1)}(\vec{x}_1, \vec{x}_2) \big\rangle = 0\,. \tag{A.14}$$

这意味着下文的式（13）和式（A.23）中的团块分别涉及两个和三个 $\Delta_j^{(\ell-1)}$，且均为自动连通，因为任何将团块断开的方式都会产生一个蝌蚪图。

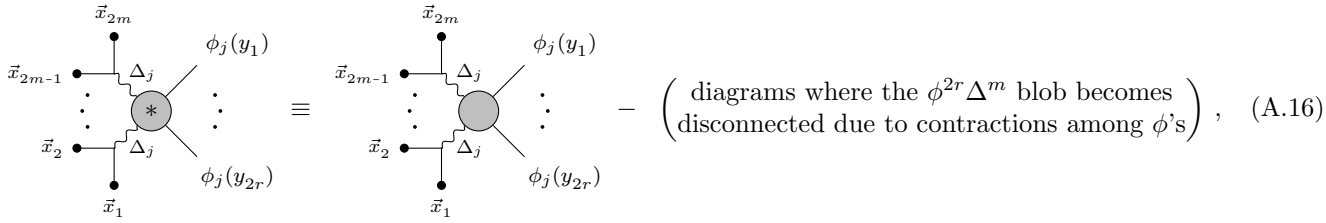
需要指出的是，我们在神经网络有效场论计算中使用费曼图的方式，或许与人们在其他情境下习惯的方式略有不同。通常，人们会推导出传播子和

相互作用顶点的费曼规则，并用它们构建图，再据此用理论参数计算关联函数。然而，在当前情形中，我们的目标是推导重整化群流，即关联函数之间的关系。这里的一般策略是，首先将 ℓ 层的 ϕ 关联函数写成 $(\ell-1)$ 层 Δ 关联函数的形式，也就是（若干个） $\Delta_j^{(\ell-1)}$ 的期望值，如费曼规则方程 (7) 所概述，并由上文方程 (A.12) 和 (A.13) 举例说明（将 $\mathcal{G}_j^{(\ell)}$ 替换为 $\Delta_j^{(\ell-1)}$，以分离出连通贡献）；然后再将这些 $(\ell-1)$ 层 Δ 关联函数写成 $(\ell-1)$ 层 ϕ 关联函数的形式。在第二步中，如果一个 Δ 关联函数涉及相同的神经元指标（例如方程 (14)），那么在领先阶，它就直接取其自由理论值，该值可用自由传播子（即二点 ϕ 关联函数）表示；如果涉及不同的神经元指标，我们就需要插入混合的 Δ-ϕ 关联函数（例如方程 (15) 中较大的团块），以连接 Δ 与四点或更高点的 ϕ 关联函数。无论哪种情况，我们都可以将结果表示为 $(\ell-1)$ 层单神经元算子的自由理论期望值：

$$\Big\langle \mathcal{O}\big(\phi_i^{(\ell-1)}(\vec{x}_1),\phi_i^{(\ell-1)}(\vec{x}_2),\dots\big)\Big\rangle_{\mathcal{K}_0^{(\ell-1)}} \equiv \frac{\int \mathcal{D}\phi\,\mathcal{O}\big(\phi(\vec{x}_1),\phi(\vec{x}_2),\dots\big)\,e^{-\int d\vec{y}_1 d\vec{y}_2\,\frac{1}{2}\phi(\vec{y}_1)\left(\mathcal{K}_0^{(\ell-1)}\right)^{-1}(\vec{y}_1,\vec{y}_2)\,\phi(\vec{y}_2)}}{\int \mathcal{D}\phi\, e^{-\int d\vec{y}_1 d\vec{y}_2\,\frac{1}{2}\phi(\vec{y}_1)\left(\mathcal{K}_0^{(\ell-1)}\right)^{-1}(\vec{y}_1,\vec{y}_2)\,\phi(\vec{y}_2)}}\,, \tag{A.15}$$

其中 $\mathcal{O}$ 表示 Δ和 ϕ的乘积（但领先阶二点关联函数 $\mathcal{K}_0$ 除外，其中 $\mathcal{O}=\sigma_{\vec{x}_1}\sigma_{\vec{x}_2}$；见式（11））。然后，我们可以通过 Wick 收缩，将这些期望值改写为 Δ的泛函导数的期望值（例如见式（15）。

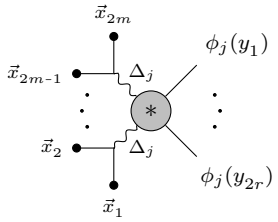
为了系统地实施 这一通用流程，方便起见，我们引入以下 $*$-团块记法 ：



(A.16)

其中，所有 m Δ和 $2r$ ϕ都带有相同的神经元指标 j。式（A.16）中被减去的图，是指其中 $\phi^{2r}\Delta^m$ 团块由于任意数量的 ϕ腿之间发生 Wick 收缩而变得不连通的那些图。在正文中，计算连通四点关联函数时，我们只遇到 $m=2$、$r=0$ 和 $m=r=1$ 这几种情况。在这些情况下，$*$-团块、完整团块和连通团块之间没有区别，因为根据式（A.14）的无蝌蚪条件，以任何方式断开这些团块都会得到零。不过，对于一般的 m、r，我们必须注意，通过 Wick 收缩 ϕ 并非断开 $\phi^{2r}\Delta^m$团块的唯一方式；而且，在一个图中断开 $\phi^{2r}\Delta^m$ 团块也不一定会使整个图变得不连通。尽管如此，我们将在下一节的例子中看到，使用 $*$-团块可以方便地组织连通 ϕ 关联函数的重整化群流推导，并妥善处理与重复计数有关的细节。$*$-团块也有简单的表达式：要求每个 ϕ 都必须与某个 Δ 发生 Wick 收缩后，我们得到领先阶 $\frac{1}{n}$ 下的以下通用公式：

–



$$= \left(\frac{C_W^{(\ell)}}{n_{\ell-1}}\right)^m \int \prod_{\alpha=1}^{2r} d\vec{z}_\alpha\, \mathcal{K}_0^{(\ell-1)}(\vec{y}_\alpha,\vec{z}_\alpha) \left\langle \frac{\delta^{2r}\big(\Delta(\vec{x}_1,\vec{x}_2)\cdots\Delta(\vec{x}_{2m-1},\vec{x}_{2m})\big)}{\delta\phi(\vec{z}_1)\cdots\delta\phi(\vec{z}_{2r})}\right\rangle_{\mathcal{K}_0^{(\ell-1)}}. \tag{A.17}$$

如正文中公式（15）下方所述，在计算完整图时，我们总是会将公式（A.17）中子图的表达式与和 ϕ 腿相关联的逆传播子进行卷积（这些腿在完整图中会成为内部线），或者换句话说，我们会截去公式（ ϕ ）中的腿。这样，右侧公式（A.17）中就只剩下期望值。在下文中，我们将使用 $*$-团块来构建图，尽管当 $r=0$ 或 $m=r=1$ 时，它们与完整团块重合（此时“$*$”标签是多余的）。

最后，我们对比一下本文结果与文献 [21], 的结果。该文献也给出了截至次领先阶的二点关联函数，以及领先阶的连通四点关联函数。文献 [21]中的结果同样以单神经元算子的自由理论期望值表示，但其中的算子是 σ和 ϕ的乘积，而我们的结果则是

以 Δ乘积的导数表示的。为了比较这些结果，可以利用定义式（9）将我们的结果改写为 σ的导数，并利用 Wick 收缩将文献 [21]中的结果化为相同的表达式。例如，按照文献 [21],的记号，期望值 $\langle \sigma_{\alpha_1}\sigma_{\alpha_2}(z_{\beta_1}z_{\beta_2} - g_{\beta_1\beta_2})\rangle_g$对应于本文记号中的 $\langle \sigma(\vec{x}_1)\,\sigma(\vec{x}_2)\big(\phi(\vec{y}_1)\,\phi(\vec{y}_2) - \mathcal{K}_0^{(\ell-1)}(\vec{y}_1,y_2)\big)\rangle_{\mathcal{K}_0^{(\ell-1)}}$，可以通过将两个 ϕ分别与 $\sigma(\vec{x}_1)\,\sigma(\vec{x}_2)$进行 Wick 收缩来进一步求值；另一种可能的 Wick 收缩——将两个 ϕ彼此收缩——会产生一项，该项与 $\mathcal{K}_0^{(\ell-1)}$ 项相抵消。因此，

$$\big\langle \sigma(\vec{x}_1)\,\sigma(\vec{x}_2)\big(\phi(\vec{y}_1)\,\phi(\vec{y}_2) - \mathcal{K}_0^{(\ell-1)}(\vec{y}_1,y_2)\big)\big\rangle_{\mathcal{K}_0^{(\ell-1)}} = \int d\vec{z}_1 d\vec{z}_2\, \mathcal{K}_0^{(\ell-1)}(\vec{y}_1,\vec{z}_1)\,\mathcal{K}_0^{(\ell-1)}(\vec{y}_2,\vec{z}_2)\left\langle \frac{\delta^2\big(\sigma(\vec{x}_1)\sigma(\vec{x}_2)\big)}{\delta\phi(\vec{z}_1)\,\delta\phi(\vec{z}_2)}\right\rangle_{\mathcal{K}_0^{(\ell-1)}}, \tag{A.18}$$

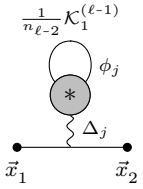
<escape>其在忽略前置系数后，与方程（A.17）右端的表达式相同，其中 $m = r = 1$ 由方程（9）代入得到（方程（9）中的最后一项本身就是一个期望值，其泛函导数为零）。</escape>

图解计算的进一步演示

现在，我们再给出三项计算，以进一步展示我们的图示方法：次领先阶（NLO）二点关联函数、领先阶（LO）连通六点关联函数和领先阶（LO）连通八点关联函数，它们分别为 $\mathcal{O}\big(\frac{1}{n}\big)$、$\mathcal{O}\big(\frac{1}{n^2}\big)$ 和 $\mathcal{O}\big(\frac{1}{n^3}\big)$。Ref. [21], 此前已计算过次领先阶二点关联函数，我们复现了其结果。据我们所知，连通六点和八点关联函数的结果尚未发表；作为独立验证，我们也采用 Ref. [22] 的代数方法计算了它们，发现结果与我们的图示计算完全一致。[3]

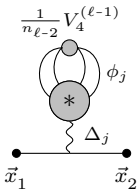
Two-point correlator at NLO

我们首先考虑次领先阶的二点关联函数计算，即 $p = 1$ 方程（10）中级数的项，$\frac{1}{n_{\ell-1}}\mathcal{K}_1^{(\ell)}(\vec{x}_1,\ \vec{x}_2)$。二点关联函数重整化群流的 $\mathcal{O}\big(\frac{1}{n}\big)$ 修正有两个来源。首先，我们可以让其中一个（$\ell-1$）层传播子的次领先阶部分取值：

$$\sum_j$$ 

$$= \frac{C_W^{(\ell)}}{2\,n_{\ell-2}}\int d\vec{y}_1\, d\vec{y}_2\, \mathcal{K}_1^{(\ell-1)}(\vec{y}_1,\vec{y}_2)\left\langle \frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\right\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n^2}\Big), \tag{A.19}$$

其中我们使用了带有 $m = r = 1$ 的公式（A.17），对称因子 $_1\ _2$ 来自交换连接到 $*$-团块的两个 ϕ_j 腿。对 j 的求和产生一个 $n_{\ell-1}$ 因子，该因子抵消了 $\phi^2\Delta$ 顶点中的 $\frac{1}{n_{\ell-1}}$ 因子。另一个贡献来自插入（$\ell-1$）层的连通四点关联函数：

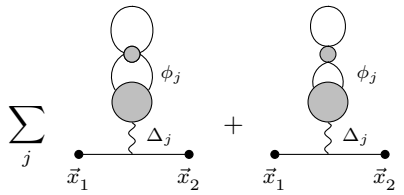
$$\sum_j$$ 

$$= \frac{C_W^{(\ell)}}{8\,n_{\ell-2}}\int\prod_{\alpha=1}^{4} d\vec{y}_\alpha\, V_4^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)\left\langle \frac{\delta^4\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)\,\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)}\right\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\Big(\frac{1}{n^2}\Big), \tag{A.20}$$

这里我们使用了方程（A.17），其中有 $m = 1,\ r = 2$；对称因子 $\frac{1}{2^3} = \frac{1}{8}$来自交换连接到$(\ell-1)$的四条腿上的 $\vec{y}_1\leftrightarrow\vec{y}_2$、$\vec{y}_3\leftrightarrow\vec{y}_4$ $\vec{y}_1\ \vec{y}_2\ \leftrightarrow\ \vec{y}_3\ \vec{y}_4$ ϕ_j V_4 $\vec{y}_1\ \vec{y}_2\ \vec{y}_3\ \vec{y}_4$ 和（，）（，）。值得注意的是，鉴于

[3] 文献 [22] 给出了退化输入极限下的连通六点和八点关联函数。修正后者 arXiv 第二版中的一些错误后，我们的结果在该极限下与文献 [22] 一致。感谢 Boris Hanin 就此问题与我们通信。

方程（A.16）中对 $*$-团块的定义，方程（A.20）中的图不包含四条ϕ_j 腿中任意两条进行维克收缩所产生的贡献：



(A.21)

收缩 ϕ_j 这样的腿会使 $\phi_j^4\Delta_j$ 子图断开连接，而整个图仍然保持连通。另一方面，这些贡献实际上已经包含在方程（A.19）中，因为这些图的上半部分（涉及（$\ell-1$）第 th 层连通四点关联函数）只是对 $\phi_j^{(\ell-1)}$ 传播子的次领先阶修正。一般而言，我们在方程（A.16）中对 $*$-团块的定义能够有效避免对此类图的重复计数。

将两者相加。　综合上述讨论的贡献，我们得到次领先阶两点相关函数的重整化群流　r:

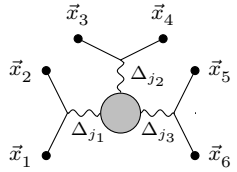
$$\frac{1}{n_{\ell\text{-}1}}\,\mathcal{K}_1^{(\ell)}(\vec{x}_1,\vec{x}_2) = \text{Eq. (A.19)} + \text{Eq. (A.20)}\,. \tag{A.22}$$

这可用于递归地根据 $\mathcal{K}_1^{(\ell\text{-}1)}$、$V_4^{(\ell\text{-}1)}$ 和 $\mathcal{K}_0^{(\ell\text{-}1)}$ 确定 $\mathcal{K}_1^{(\ell)}$。

Connected six-point correlator

接下来，我们演示领先阶连通六点关联函数的重整化群流的推导。与连通四点关联函数的式（13）类似，我们有

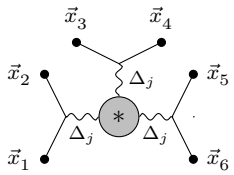
$$\frac{1}{n_{\ell\text{-}1}^2}\,V_6^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4;\vec{x}_5,\vec{x}_6) = \sum_{j_1,j_2,j_3}$$



. (A.23)

我们需要考虑三种情况。首先，如果 j_1、j_2、j_3 都相等，即 $j_1=j_2=j_3\equiv j$，我们可以直接使用自由理论传播子 $\mathcal{K}_0^{(\ell\text{-}1)}$ 连接 Δ_j 中包含的 ϕ_j 场，从而得到领先阶结果：

$$\sum_j$$

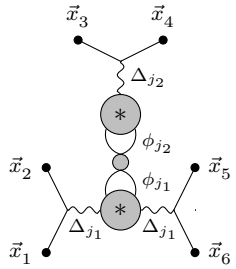


$$= \frac{(C_W^{(\ell)})^3}{n_{\ell\text{-}1}^2}\left\langle \Delta(\vec{x}_1,\vec{x}_2)\,\Delta(\vec{x}_3,\vec{x}_4)\,\Delta(\vec{x}_5,\vec{x}_6)\right\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}} + \mathcal{O}\Big(\frac{1}{n^3}\Big)\,, \tag{A.24}$$

其中，对 j 求和会产生一个因子 $n_{\ell\text{-}1}$，它与来自 $\phi^2\Delta$ 个顶点的三个 $\frac{1}{n_{\ell\text{-}1}}$ 因子中的一个相消，使结果成为 $\mathcal{O}\big(\frac{1}{n^2}\big)$。

其次，如果 j_1、j_2、j_3 取两个　对于不同的取值，我们需要在第 $\ell-1$ 层使用连通四点关联函数。

连接具有不同索引的神经元（同时仍使用自由传播子连接具有相同索引的神经元）：

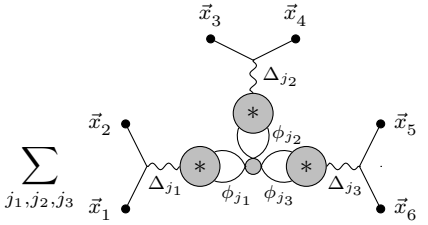
$$\sum_{j_1,j_2}$$  $$+\ \text{perms.} = \frac{\big(C_W^{(\ell)}\big)^3}{4\,n_{\ell\text{-}1}n_{\ell\text{-}2}}\int\prod_{\alpha=1}^{4}d\vec{y}_\alpha\ V_4^{(\ell\text{-}1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)$$

$$\Bigg[\Big\langle\frac{\delta^2\big(\Delta(\vec{x}_1,\vec{x}_2)\,\Delta(\vec{x}_5,\vec{x}_6)\big)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}+\Big\langle\frac{\delta^2\big(\Delta(\vec{x}_1,\vec{x}_2)\,\Delta(\vec{x}_3,\vec{x}_4)\big)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_5,\vec{x}_6)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}$$

$$+\Big\langle\frac{\delta^2\big(\Delta(\vec{x}_3,\vec{x}_4)\,\Delta(\vec{x}_5,\vec{x}_6)\big)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Bigg]+\mathcal{O}\Big(\frac{1}{n^3}\Big)\,, \tag{A.25}$$

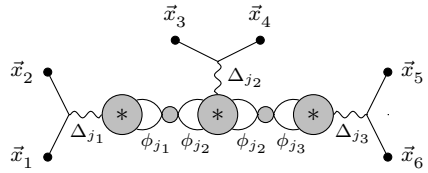
由于我们对 $*$-团块的定义公式（A.16），该图是自动连通的。为得到公式（A.25）中的表达式，我们对两个 $*$-团块分别在$(m,r)=(2,1)$和$(1,1)$的情况下使用了公式（A.17）；对称因子 $\frac{1}{2^2}=\frac{1}{4}$来自交换连接到 $V^{(\ell\text{-}1)}_4(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)$ 的四条ϕ_j腿中的 $\vec{y}_1\leftrightarrow\vec{y}_2$和 $\vec{y}_3\leftrightarrow\vec{y}_4$。与公式（A.24）中的第一项相比，此处 j 求和产生了一个额外因子 $n_{\ell-1}$，而插入的连通四点关联函数带有因子$\frac{1}{n_{\ell-2}}$，因此最终结果仍为$\mathcal{O}\big(\frac{1}{n^2}\big)$。请注意，无论求和中是否包含 $j_1=j_2$ 项，公式（A.25）都成立（正文中的公式（15）也同样如此）。

最后，如果 j_1,j_2,j_3 彼此各不相同，我们必须使用连通六点关联函数或两个连通四点关联函数来连接第（$\ell-1$）层神经元。前一种情况下，我们有

$$\sum_{j_1,j_2,j_3}$$  $$=\frac{\big(C_W^{(\ell)}\big)^3}{8\,n_{\ell\text{-}2}^2}\prod_{\alpha=1}^{6}\int d\vec{y}_\alpha\ V_6^{(\ell\text{-}1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4;\vec{y}_5,\vec{y}_6)$$

$$\Big\langle\frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_5,\vec{x}_6)}{\delta\phi(\vec{y}_5)\,\delta\phi(\vec{y}_6)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}+\mathcal{O}\Big(\frac{1}{n^3}\Big)\,, \tag{A.26}$$

而在后一种情况下，我们有

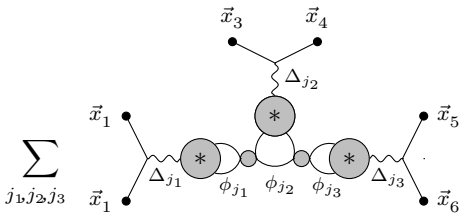
$$\sum_{j_1,j_2,j_3}$$  $$+\ \text{perms.} = \frac{\big(C_W^{(\ell)}\big)^3}{16\,n_{\ell\text{-}2}^2}\int\prod_{\alpha=1}^{8}d\vec{y}_\alpha\ V_4^{(\ell\text{-}1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)\,V_4^{(\ell\text{-}1)}(\vec{y}_5,\vec{y}_6;\vec{y}_7,\vec{y}_8)$$

$$\Bigg[\Big\langle\frac{\delta^4\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)\,\delta\phi(\vec{y}_7)\,\delta\phi(\vec{y}_8)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_5,\vec{x}_6)}{\delta\phi(\vec{y}_5)\,\delta\phi(\vec{y}_6)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}$$

$$+\Big\langle\frac{\delta^4\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)\,\delta\phi(\vec{y}_7)\,\delta\phi(\vec{y}_8)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_5,\vec{x}_6)}{\delta\phi(\vec{y}_5)\,\delta\phi(\vec{y}_6)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}$$

$$+\Big\langle\frac{\delta^4\Delta(\vec{x}_5,\vec{x}_6)}{\delta\phi(\vec{y}_3)\,\delta\phi(\vec{y}_4)\,\delta\phi(\vec{y}_7)\,\delta\phi(\vec{y}_8)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\,\delta\phi(\vec{y}_2)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Big\langle\frac{\delta^2\Delta(\vec{x}_3,\vec{x}_4)}{\delta\phi(\vec{y}_5)\,\delta\phi(\vec{y}_6)}\Big\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}\Bigg]+\mathcal{O}\Big(\frac{1}{n^3}\Big)\,. \tag{A.27}$$

对称因子和 $\frac{1}{n}$ 计数此时应该已经显而易见了。注意，使用 $*$-团块后，式（A.27）既排除了非连通图（显然这是我们应该排除的），也排除了通过收缩四条 ϕ_{j_2} 腿中的两条而得到的下述连通图（从而使 $\phi_{j_2}^4\Delta_{j_2}$ 团块断开）：



$+$ 置换。 (A.28)

另一方面，这一贡献已经在方程 (A.26) 中计入。因此，正如在上一小节中计算 $\mathcal{K}_1^{(\ell)}$ 时一样（参见方程 (A.21) 附近的讨论），使用 $*$-团块可以巧妙地避免重复计数。

总而言之，领先阶下连通六点关联函数的重整化群流仅由上文讨论的四组费曼图得出：

$$\frac{1}{n_{\ell-1}^2}V_6^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4;\vec{x}_5,\vec{x}_6) = \text{Eq. (A.24)} + \text{Eq. (A.25)} + \text{Eq. (A.26)} + \text{Eq. (A.27)}\,. \tag{A.29}$$

这使我们能够递归地根据 $V_6^{(\ell-1)}$、$V_4^{(\ell-1)}$ 和 $\mathcal{K}_0^{(\ell-1)}$ 确定 $V_{(\ell)\,6}$。

Connected eight-point correlator

最后，我们在表 I 中总结了连通八点关联函数的计算。与上面的 V_6 情况一样，我们按照求和所涉及的不同 j 指标的数量来组织计算。为简单起见，我们省略对 j_1、j_2、.....的求和，以及图中的所有标记；此时应该已经很清楚，在领先阶（LO）下，连接到同一个 $*$-团块的所有 Δ 和 ϕ 都共享相同的神经元指标，而较小的 $V_{(\ell-1)\,2k}$ 团块则可以连接不同的神经元指标。一个值得注意的新特点是，从八点阶开始，我们需要减去那些在公式 (A.16) 中 $*$-团块的定义里未被排除的非连通图。与此同时，使用 $*$-团块仍然能够以公式 (A.21) 和 (A.28) 附近讨论的相同方式避免重复计数。每一行中的连通图都清楚地展示了 $\frac{1}{n}$ 计数：每个 $\phi^2\Delta$ 顶点（共四个）都带有一个 $\frac{1}{n_{\ell-1}}$ 因子；每个 $*$-团块表示一次 j 求和，因此带有一个 $n_{\ell-1}$ 因子；而表示第 $(\ell-1)$ 层神经元的 $2k$ 点连通关联函数的较小团块则带有一个 $\frac{1}{n_{\ell-2}^{k-1}}$ 因子。所有图的最终结果均为 $\mathcal{O}\left(\frac{1}{n^3}\right)$。

对于一般（非退化）输入 $\vec{x}_1$、.....、$\vec{x}_8$，每个图的表达式都很容易直接读出，但结果相当冗长，也没有特别直观的启发。因此，为简洁起见，我们在表 I 的最后一列给出退化输入极限下的最终结果，并省略输入参数。在每个表达式中，我们将对称因子写在前面；其余数值因子来自组合计数。连通八点关联函数的重整化群流由表 I中的所有图求和给出，可用于递归确定 $V_8^{(\ell)}$，其依据是 $V_{(\ell-1)\,8}$、$V_{(\ell-1)\,6}$、$V_{(\ell-1)\,4}$ 和 $\mathcal{K}_0^{(\ell-1)}$。

关于临界性调节的进一步讨论

本节简要回顾文献 [21]中关于临界性调节的结果，并说明其与我们的方程（31）之间的联系。在邻近输入极限下，文献 [21] 确定了以下临界性条件：

$$\chi_{\parallel}^{(\ell)}(\mathcal{K}^\star) \equiv \frac{C_W^{(\ell)}}{2\,\mathcal{K}^{\star 2}}\left\langle\sigma^2(\phi^2-\mathcal{K}^\star)\right\rangle_{\mathcal{K}^\star} = \frac{C_W^{(\ell)}}{2}\left\langle(\sigma^2)''\right\rangle_{\mathcal{K}^\star} = 1\,, \tag{A.30}$$

$$\chi_{\perp}^{(\ell)}(\mathcal{K}^\star) \equiv C_W^{(\ell)}\left\langle\sigma'^2\right\rangle_{\mathcal{K}^\star} = 1, \tag{A.31}$$

其中，假设所有涉及的函数都具有共同的输入参数 $\vec{x}$，并且 $\mathcal{K}^\star(\vec{x})$ 是当 $\vec{x}_1 = \vec{x}_2 \equiv \vec{x}$时，核递归方程（11）的不动点。第一个条件方程（A.30）确保了 的范数具有幂律标度。

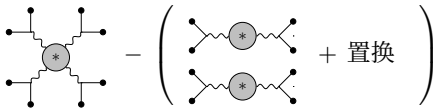
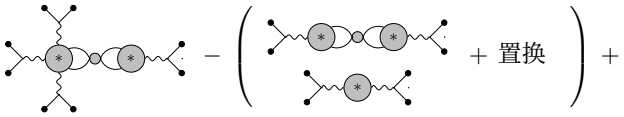
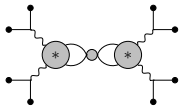
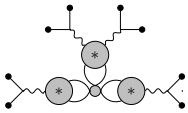
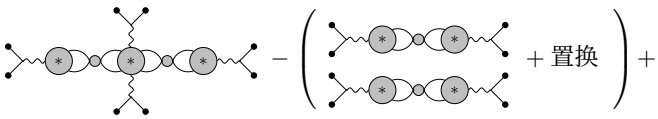
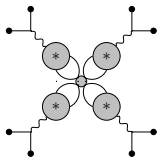
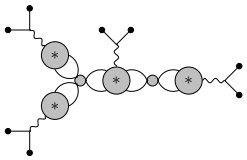
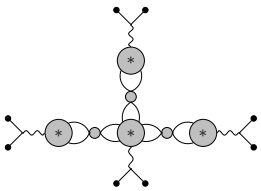
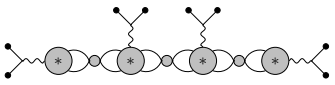
# of j 求和	图	简并极限结果 $/\left(C_W^{(\ell)}\right)^4$
1		$\frac{1}{n_{\ell-1}^3}\left(\langle\Delta^4\rangle - 3\langle\Delta^2\rangle^2\right)$
2		$\frac{1}{4}\cdot\frac{V_{4}^{(\ell-1)}}{n_{\ell-1}^2 n_{\ell-2}}\cdot 4\left(\langle\partial^2\Delta^3\rangle\langle\partial^2\Delta\rangle - 3\langle\partial^2\Delta\rangle^2\langle\Delta^2\rangle\right)$
		$\frac{1}{4}\cdot\frac{V_{4}^{(\ell-1)}}{n_{\ell-1}^2 n_{\ell-2}}\cdot 3\langle\partial^2\Delta^2\rangle^2$
3		$\frac{1}{8}\cdot\frac{V_{6}^{(\ell-1)}}{n_{\ell-1} n_{\ell-2}^2}\cdot 6\langle\partial^2\Delta^2\rangle\langle\partial^2\Delta\rangle^2$
		$\frac{1}{16}\cdot\frac{\left(V_{4}^{(\ell-1)}\right)^2}{n_{\ell-1} n_{\ell-2}^2}\cdot 12\langle\partial^4\Delta\rangle\langle\partial^2\Delta^2\rangle\langle\partial^2\Delta\rangle$
4		$\frac{1}{16}\cdot\frac{V_{8}^{(\ell-1)}}{n_{\ell-2}^3}\langle\partial^2\Delta\rangle^4$
		$\frac{1}{32}\cdot\frac{V_{6}^{(\ell-1)}V_{4}^{(\ell-1)}}{n_{\ell-2}^3}\cdot 12\langle\partial^4\Delta\rangle\langle\partial^2\Delta\rangle^3$
		$\frac{1}{64}\cdot\frac{\left(V_{4}^{(\ell-1)}\right)^3}{n_{\ell-2}^3}\cdot 4\langle\partial^6\Delta\rangle\langle\partial^2\Delta\rangle^3$
		$\frac{1}{64}\cdot\frac{\left(V_{4}^{(\ell-1)}\right)^3}{n_{\ell-2}^3}\cdot 12\langle\partial^4\Delta\rangle^2\langle\partial^2\Delta\rangle^2$

表 I. 连通八点关联函数的图示计算。

给定输入的预激活值 $\vec{x}$，而第二个条件，即方程（A.31）（早期研究中的参考文献 [38–40]也指出了这一点），确保范数相同且彼此无限接近的输入所对应的预激活值之间的距离遵循幂律标度。这些条件意味着需要对初始化超参数进行如下调节：

- 对于尺度不变激活函数，$\sigma(\phi) = \begin{cases} a_-\phi & (\phi < 0) \\ a_+\phi & (\phi \geq 0) \end{cases}$（包括 ReLU，对应于 $a_- = 0, a_+ = 1$），对任意输入 $\vec{x}$ 存在有限不动点 $\mathcal{K}^*$ 要求 $C(\ell)_b = 0$。同时，$\chi(\ell)_\parallel(\mathcal{K}^\star) = \chi(\ell)_\perp(\mathcal{K}^\star) = C(\ell)_W \frac{1}{2}(a_+^2 + a_-^2)$，与 $\mathcal{K}^\star$无关。因此，设置 $C(\ell)_W = \frac{2}{a_+^2 + a_-^2}$ 可确保对于任意输入 $\vec{x}$，方程（A.30）和（A.31）均得到满足。

- 对于满足 $\sigma(0) = 0, \sigma'(0) \neq 0$ 的光滑激活函数，我们可以对 $\sigma(\phi) = \sigma_1\phi + \frac{1}{2}\sigma_2\phi^2 + \ldots$ 进行泰勒展开，并证明当我们设定 $C(\ell)_b = 0$ 时，对于任意 $\vec{x}$，$\mathcal{K}^\star = 0$ 都是不动点。同时，$\chi(\ell)_\parallel(\mathcal{K}^\star) = \chi(\ell)_\perp(\mathcal{K}^\star) = C(\ell)_W \sigma_1^2$，这意味着我们应设定 $C(\ell)_W = \frac{1}{\sigma_1^2}$，以满足方程（A.30）和（A.31）。

值得注意的是，尽管临界性条件方程（A.30）和（A.31）看起来似乎约束过多，因为它们需要对每个输入 $\vec{x}$ 都成立，而只有两个超参数 $C_W^{(\ell)}$，$C_b^{(\ell)}$ 我们可以在每一层进行调节；但实际上，对于上述两类激活函数（称为尺度不变类和 $\mathcal{K}^\star = 0$ 普适类），至少可以满足这些条件。

我们现在证明，上述超参数调节实际上确保了更强的条件（式（31）在 $\frac{1}{n}$的领先阶也成立。对 $\chi^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{y}_1, \vec{y}_2)$ 的定义（式（24））中的泛函导数进行显式计算，可得

$$\begin{aligned}
\chi^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{y}_1, \vec{y}_2) &= \frac{C_W^{(\ell)}}{2} \left\langle \frac{\delta^2 \left[\sigma\left(\phi(\vec{x}_1)\right) \sigma\left(\phi(\vec{x}_2)\right)\right]}{\delta\phi(\vec{y}_1)\, \delta\phi(\vec{y}_2)} \right\rangle_{\mathcal{K}_0^{(\ell-1)}} + \mathcal{O}\left(\frac{1}{n}\right) \\
&= \frac{C_W^{(\ell)}}{2} \Big\{ \Big[\delta(\vec{x}_1 - \vec{y}_1)\, \delta(\vec{x}_2 - \vec{y}_2) + \delta(\vec{x}_1 - \vec{y}_2)\, \delta(\vec{x}_2 - \vec{y}_1)\Big] \left\langle \sigma'\left(\phi(\vec{x}_1)\right) \sigma'\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}_0^{(\ell-1)}} \\
&\quad + \delta(\vec{x}_1 - \vec{y}_1)\, \delta(\vec{x}_1 - \vec{y}_2) \left\langle \sigma''\left(\phi(\vec{x}_1)\right) \sigma\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}_0^{(\ell-1)}} \\
&\quad + \delta(\vec{x}_2 - \vec{y}_1)\, \delta(\vec{x}_2 - \vec{y}_2) \left\langle \sigma\left(\phi(\vec{x}_1)\right) \sigma''\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}_0^{(\ell-1)}} \Big\} + \mathcal{O}\left(\frac{1}{n}\right). \qquad \text{(A.32)}
\end{aligned}$$

因此，在 $\frac{1}{n}$ 的领先阶，式（31）可等价地写为

$$C_W^{(\ell)} \left\langle \sigma'\left(\phi(\vec{x}_1)\right) \sigma'\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}^\star} = 1\,, \qquad \left\langle \sigma''\left(\phi(\vec{x}_1)\right) \sigma\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}^\star} = 0 \qquad (\forall \vec{x}_1, \vec{x}_2)\,, \qquad \text{(A.33)}$$

or

$$C_W^{(\ell)} \left\langle \sigma'\left(\phi(\vec{x}_1)\right) \sigma'\left(\phi(\vec{x}_2)\right) \right\rangle_{\mathcal{K}^\star} = \frac{C_W^{(\ell)}}{2} \left\langle \left[\sigma\left(\phi(\vec{x}_1)\right) \sigma\left(\phi(\vec{x}_2)\right)\right]'' \right\rangle_{\mathcal{K}^\star} = 1 \qquad (\forall \vec{x}_1, \vec{x}_2)\,. \qquad \text{(A.34)}$$

后一种形式清楚地表明，式（A.30）和（A.31）是一般而言更强的条件（式（31）在退化输入极限下的形式。重要的是，对于尺度不变激活函数和 $\mathcal{K}^\star = 0$ 激活函数，式（A.34）中的期望值与 $\vec{x}_1$、$\vec{x}_2$无关，因此，从近邻输入分析推导出的超参数调节意味着式（31）也成立。

最后，我们注意到，在 $\frac{1}{n}$ 中将超参数调节至满足领先阶 LO 下的临界性条件式（31），足以确保连通关联函数呈现幂律标度，即 即 $\delta\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)\rangle \sim \ell^\gamma$ 以及 $V^{(\ell)}_{2k}(\vec{x}_1, \ldots \vec{x}_{2k}) \sim \ell^{\gamma_k}$（而非 $e^{\pm\ell}$），这些结果来自求解 RG 方程，其中 γ，γ_k 是临界指数。尽管 $\mathcal{O}\left(\frac{1}{n}\right)$ 修正乍看之下会导致$\delta\langle \mathcal{G}^{(\ell)}\rangle$，$V^{(\ell)}_{2k} \sim e^{\pm\frac{\ell}{n}}$），但当网络深度为 $L \ll n$ 时，这种标度实际上并非指数形式。另一方面，$\mathcal{O}\left(\frac{1}{n}\right)$ 修正会改变临界指数，而将 $C_W^{(\ell)}$，$C_b^{(\ell)}$ 调节 $\mathcal{O}\left(\frac{1}{n}\right)$ 个量级，则可能带来更有利的幂律标度（例如 涨落增长更慢） [21]。

临界性处幂律标度的数值验证

在最后这一节中，我们进行数值实验，以验证当超参数 $C_W^{(\ell)}, C_b^{(\ell)}$ 取前一节讨论的临界值时，连通关联函数的幂律标度。对于每项实验，我们

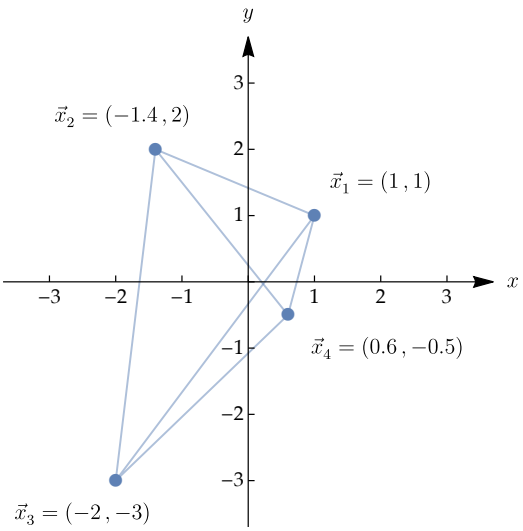


图 1. 数值实验中用于说明的输入点。

生成一个由 $N_{\text{net}} = 1{,}000$ 个神经网络构成的系综，这些网络的隐藏层宽度为 $n = 300$，深度为 $L = 30$，并按照式（A.3）进行初始化。我们考虑一个 $n_0 = 2$ 维输入空间，并选取四个点 $\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4$，如图 1 所示。激活函数选用 ReLU 或双曲正切。对于每个网络，我们根据式（1）计算每一层的神经元预激活值。随后通过系综平均得到连通二点和四点关联函数：

$$\left\langle \mathcal{G}^{(\ell)}(\vec{x}_\alpha, \vec{x}_\beta) \right\rangle = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \phi_{I,i}^{(\ell)}(\vec{x}_\alpha)\, \phi_{I,i}^{(\ell)}(\vec{x}_\beta)\,, \tag{A.35}$$

$$\begin{aligned} V_4^{(\ell)}(\vec{x}_\alpha, \vec{x}_\beta; \vec{x}_\gamma, \vec{x}_\delta) = {} & \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{1}{n_\ell} \sum_{i_1, i_2 = 1}^{n_\ell} \phi_{I,i_1}^{(\ell)}(\vec{x}_\alpha)\, \phi_{I,i_1}^{(\ell)}(\vec{x}_\beta)\, \phi_{I,i_2}^{(\ell)}(\vec{x}_\gamma)\, \phi_{I,i_2}^{(\ell)}(\vec{x}_\delta) - n_\ell \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\alpha, \vec{x}_\beta) \right\rangle \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\gamma, \vec{x}_\delta) \right\rangle \\ & - \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\alpha, \vec{x}_\gamma) \right\rangle \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\beta, \vec{x}_\delta) \right\rangle - \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\alpha, \vec{x}_\delta) \right\rangle \left\langle \mathcal{G}^{(\ell)}(\vec{x}_\beta, \vec{x}_\gamma) \right\rangle + \mathcal{O}\Big(\frac{1}{n}\Big)\,, \end{aligned} \tag{A.36}$$

其中，I 标记系综中的网络。我们分别对 ReLU 网络和 tanh 网络进行了 10 次实验，并将这 10 次实验的标准差作为所计算各关联函数的数值不确定性，在下方图中以竖线表示。

在图 2 和 3中，我们分别展示了 ReLU 网络和 tanh 网络二点关联函数的数值结果。我们选取 $\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\ \vec{x}_1) \rangle$（上排）和 $\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\ \vec{x}_2) \rangle$（下排）作为简并和非简并输入的示例；我们已确认，所有其他二点关联函数也表现出类似行为，并且同样符合下文讨论的理论预期。在每幅图中，中间面板对应将超参数设为其临界值 $C_W^{(\ell)} = C_W^{\text{c}}$、$C_b^{(\ell)} = 0$所得的结果；左侧和右侧面板则分别对应将$C_W^{(\ell)} = \frac{1}{4} C_W^{\text{c}}$和 $4C_W^{\text{c}}$设为相应数值（对应于 $W_{(\ell)\,ij}$的标准差分别为临界值的一半和两倍），同时保持 $C_b^{(\ell)} = 0$不变所得的结果。这里的重要结论是，所有中间面板中的标度均呈幂律，而所有左侧和右侧面板中的标度均呈指数形式。下面我们将更详细地解释这些图：

- 对于 ReLU 激活函数（图 2），我们预计任意输入的二点关联函数都会在 $C_{(\ell)\,W} < C_{\text{c}W}$ ($C_{(\ell)\,W} > C_{\text{c}W}$ 时，指数地流向 $\mathcal{K}^\star = 0$ ($\mathcal{K}^\star = \infty$) 处的平凡不动点；这与图 2 的左（右）面板一致。另一方面，在临界性 $C_{(\ell)\,W} = C_{\text{c}_W} = 2$ 处，存在一个非平凡不动点 [2, 21, 27]

$$\mathcal{K}^\star(\vec{x}_\alpha, \vec{x}_\beta) = \frac{C_W^{\text{c}}}{n_0} \sqrt{|\vec{x}_\alpha||\vec{x}_\beta|} = \sqrt{|\vec{x}_\alpha||\vec{x}_\beta|}\,. \tag{A.37}$$

  对于简并输入，二点关联函数在紫外边界 $\ell = 1$ 处已经位于该不动点，并且预计在 $\frac{1}{n}$ 中的领先阶下，对所有 ℓ 都保持不变。顶排中间的面板与这一预期一致。对于非退化输入，底排中间的面板证实，流向该不动点的重整化群流如预期呈幂律。

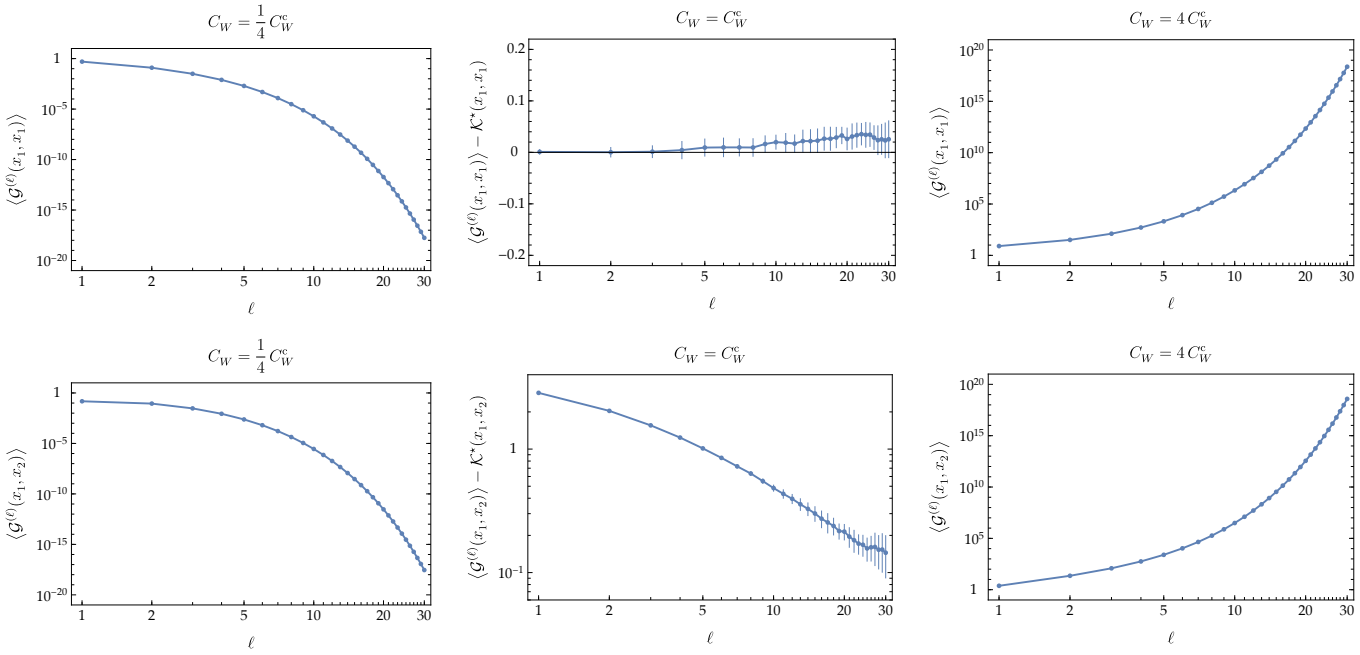


图 2。 **ReLU网络的二点关联函数**。 趋向不动点的渐近标度在临界性处呈幂律（中间面板），而偏离临界性时呈指数形式（左侧和右侧面板）。中间面板中的不动点 $\mathcal{K}^\star$ 由公式（A.37）给出。详情请参阅正文。

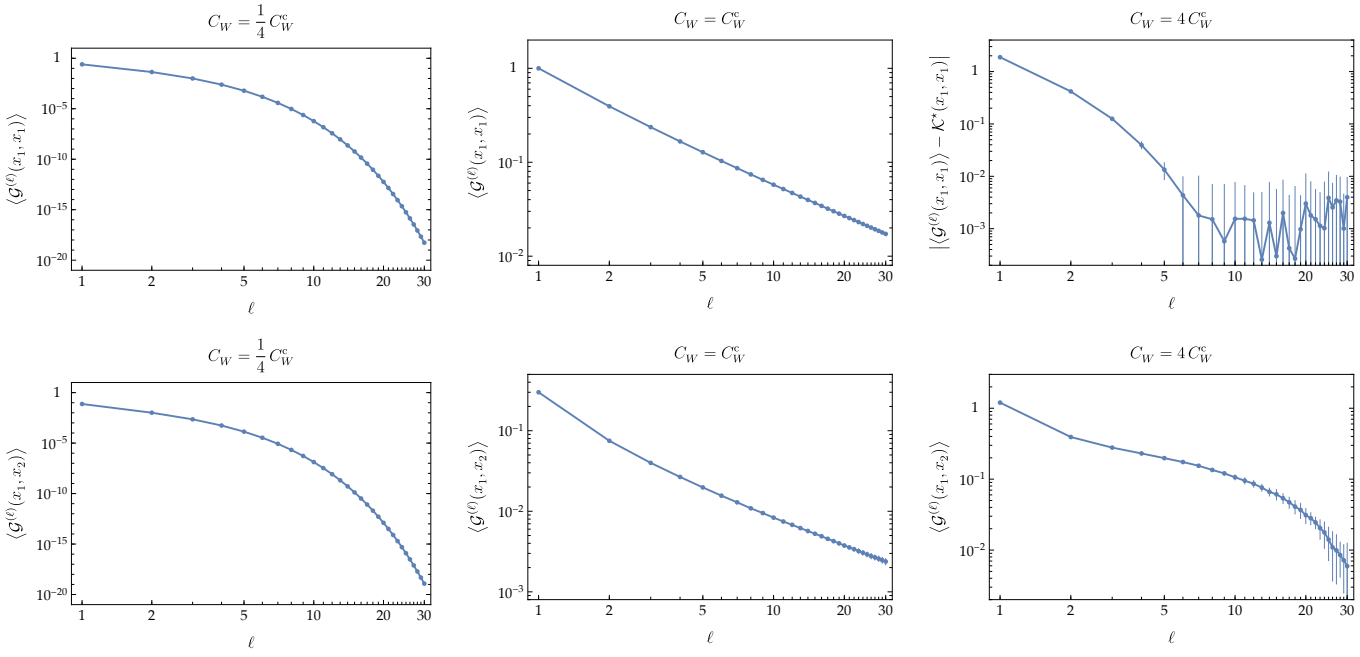


图 3。 **tanh网络的二点关联函数**。 向不动点的渐近标度在临界性处呈幂律（中间面板），在偏离临界性时呈指数形式（左侧和右侧面板）。右上面板中的不动点 $\mathcal{K}^\star$ 由方程（A.38）的非零解给出。详情见正文。

- 对于双曲正切激活函数（图 3），首先注意到，在临界性处 $C_{(\ell)\ W} = C_{\mathrm{c}W} = 1$ （中间图），重整化群流如预期般以幂律趋向不动点 $\mathcal{K}^\star = 0$ 。其次，对于 $C_{(\ell)\ W} < C_{\mathrm{c}W}$ （左侧图），二点关联函数流向同一个不动点 $\mathcal{K}^\star = 0$，但标度呈指数形式。最后一种情况 $C_{(\ell)\ W} > C_{\mathrm{c}W}$ （右侧图）更为微妙：对于非退化输入，我们再次观察到指数式流向 $\mathcal{K}^\star = 0$；对于简并输入，不动点 $\mathcal{K}^\star = 0$ 具有排斥性，而二点函数实际上流向另一个满足下式的不动点：

$$\mathcal{K}^\star = C_W \left\langle \tanh^2\phi \right\rangle_{\mathcal{K}^\star} = \frac{C_W}{\sqrt{2\pi\mathcal{K}^\star}} \int_{-\infty}^{\infty} d\phi\ \tanh^2\phi\, e^{-\frac{\phi^2}{2\mathcal{K}^\star}}\,, \tag{A.38}$$

其中 C_W 是 $C_W^{(\ell)}$ 的共同取值。在后一种情况下，标度同样呈指数形式，这与远离临界性时的预期一致。

接下来，在图 4 和 5中，我们分别展示了 ReLU 网络和 tanh 网络在临界性处一组具有代表性的连通四点关联函数的数值结果。这些结果包括四个输入全部相同，以及取两个、三个和四个不同值的情形。我们绘制 $|V_4^{(\ell)}|$ ，因为 $V_4^{(\ell)}$ 可能为正，也可能为负，并且在某些情况下实际上会沿着重整化群流改变符号。由于连通四点关联函数在第一层为零，图中的起点为 $\ell = 2$ ；数值结果与 $V_4^{(1)} = 0$一致。在图 4 和 5所示的所有情形中，我们都观察到了清晰的渐近幂律标度。这些结果证实，将 $C_W^{(\ell)}$ 、 $C_b^{(\ell)}$ 调至临界性，可以确保一般（非退化）输入下所有连通四点关联函数都呈现幂律标度。我们还确认，其他连通四点关联函数在临界性处也呈现幂律标度。另一方面，远离临界性时，我们观察到连通四点关联函数呈现指数标度行为，这与图 2 和 3左右面板中的二点关联函数类似。

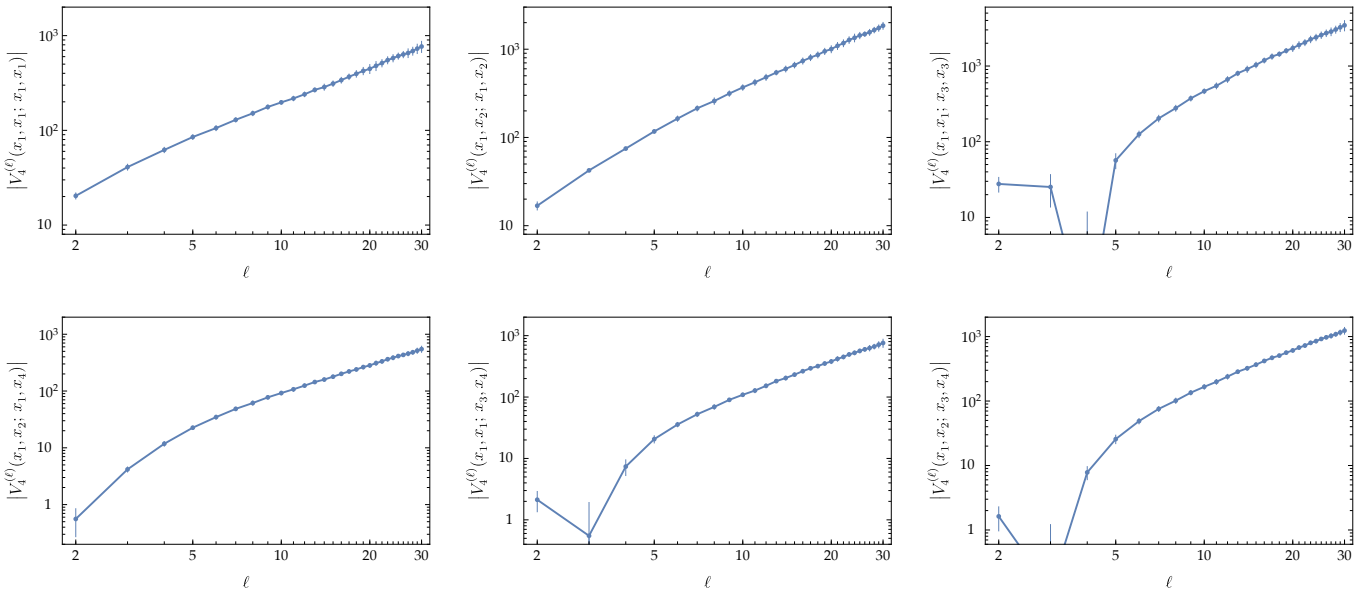


图4。**临界性下 ReLU网络的连通四点关联函数**。对于所有输入选择，渐近标度均呈幂律。详情见正文。

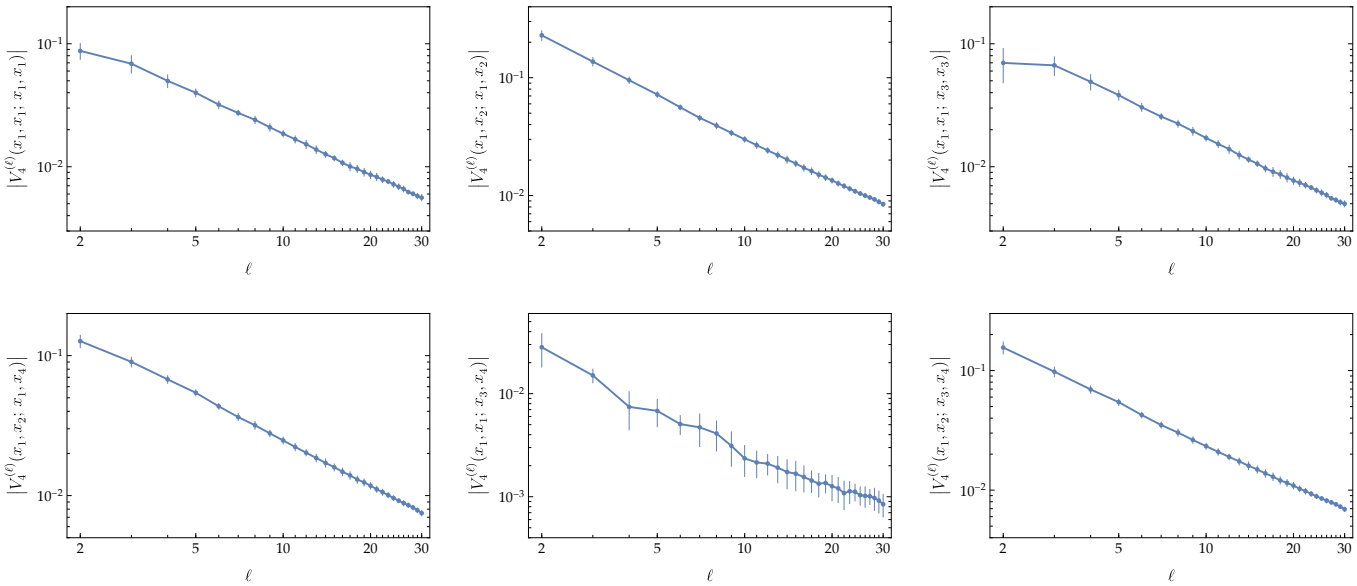


图5。**临界性下 tanh网络的连通四点关联函数**。对于所有输入选择，渐近标度均呈幂律。详情见正文。