CRITICALITY AND SATURATION IN ORTHOGONAL NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

It has been known for a long time that initializing weight matrices to be orthogonal instead of having i.i.d. Gaussian components can improve training performance. This phenomenon can be analyzed using finite-width corrections, where the infinite-width statistics are supplemented by a power series in $1/\mathrm{width}$. In particular, recent empirical results by Day et al. show that the tensors appearing in this treatment stabilize for large depth, as opposed to the tensors of i.i.d.-initialized networks. In this article, we derive explicit layer-wise recursion relations for the tensors appearing in the finite-width expansion of the network statistics in the case of orthogonal initializations. We also provide an extension of recently-introduced Feynman diagrams for the corresponding recursions in the i.i.d.-case which are valid to all orders in $1/\mathrm{width}$. Finally, we show explicitly that the recursions we derive reproduce the stability of the finite-width tensors which was observed for activation functions with vanishing fixed point. This work therefore provides a theoretical explanation for the stability of nonlinear networks of finite width initialized with orthogonal weights, closing a long-standing gap in the literature. We validate our theoretical results experimentally by showing that numerical solutions of our recursion relations and their analytical large-depth expansions agree excellently with Monte-Carlo estimates from network ensembles.

1 INTRODUCTION

Neural network training usually starts from randomly initialized parameters and a good choice of initialization distribution is critical to ensure stable training dynamics to avoid exploding or vanishing gradients. Typically, the initial weight components are sampled i.i.d. from Gaussian distributions whose variance scales inversely with the width of the network (Glorot & Bengio, 2010; He et al., 2015). However, it was found that initializing the weight matrices to be orthogonal (i.e. to satisfy $WW^\top = \mathbb{I}$, yielding eigenvalues all exactly one) improves performance (Saxe et al., 2014; Mishkin & Matas, 2016; Day et al., 2025).

This phenomenon has been studied theoretically for many years. However, these studies are largely restricted to linear networks (Saxe et al., 2014; Hu et al., 2019) or mean-field theory approaches in the large-width limit (Pennington et al., 2017; 2018; Xiao et al., 2018). Also the neural tangent kernel (NTK) at infinite width (Jacot et al., 2018) has been used to analyze orthogonal initializations. However, at infinite width, the NTKs of networks initialized with orthogonal and Gaussian weights agree (Huang et al., 2021), rendering the infinite-width NTK an unsuitable tool to understand the observed differences between the two initializations.

In this article, we instead use $1/\mathrm{width}$ corrections to the neural network statistics at infinite-width in order to investigate the effect of orthogonal initializations for nonlinear networks at finite width. In this framework (Roberts & Yaida, 2022), the network statistics are studied in terms of the cumulants of the preactivations and derivatives. The cumulants are then decomposed into tensors which are expanded in a power series in $1/\mathrm{width}$, with the zero-order contribution corresponding to the infinite-width limit. For orthogonal initializations, the first corrections to the statistics of preactivations were worked out analytically and corrections to the remaining tensors appearing in the first correction to the infinite-width training dynamics were sampled numerically (Day et al., 2025). These results show empirically that the network statistics stabilize at large depth for orthogonal initializations, in contrast to the case of Gaussian initializations. In summary, no theoretical analysis was so far able

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

to show that orthogonal initializations of nonlinear, finite-width neural networks lead to improved performance over Gaussian initializations. In this article, we fill this gap.

In particular, the finite-width corrections to the tensors governing the network statistics obey layer-wise recursion relations. These have been derived only for the cumulant of four preactivations (Day et al., 2025) (the V_4 tensor). Here, we derive the recursion relations for the additional ten fundamental tensors (D, F, A, B, P, Q, R, S, T and U). Specifically, D, F, A and B govern joint preactivation-NTK statistics and P, Q, R, S, T and U capture the statistics of higher-derivative analogues of the NTK, the so-called dNTK and ddNTK tensors. For complete definitions of these tensors, see Appendix A. Together, these ten tensors determine the training dynamics completely to order $1/\text{width}$ (Roberts & Yaida, 2022). The orthogonal statistics are captured by Weingarten functions (Weingarten, 1978) which add additional terms to the recursion relations compared to the Gaussian case. We then show by explicit iteration of our recursions and by expanding their solutions around the large-depth limit that our theoretical results reproduce the empirically observed stability of these tensors. Since these recursion relations are laborious to derive algebraically, we also extend a recently-introduced framework to facilitate these computations using Feynman diagrams (Guillen et al., 2025) to the case of orthogonal weights. This extension, which introduces a novel propagator and charge, simplifies the computation of recursion relations dramatically. We prove the completeness of our Feynman diagrams to all orders in 1width and demonstrate their power by computing the correction to order $1/\text{width}^2$ of the recursion relation of the cumulant of six preactivations (the V_6 tensor).

Limitations Although our analysis is limited to MLPs, there are no conceptual obstacles for extensions to other architectures. Our experiments and the large-depth expansion are restricted to the $\tanh$ activation function, although our theoretical results are general and apply to any smooth activation function. Finally, our explicit calculations are mostly restricted to the order $1/\text{width}$, although our Feynman diagrams are valid to all orders.

Our main contributions are:

- We derive for the case of orthogonal initializations the recursion relations for the tensors D, F, A, B, P, Q, R, S, T and U at order $1/n$, where n is the width of the network. We furthermore derive the recursion relation for the tensor V_6 at order $1/n^2$.
- We provide Feynman rules that simplify the computation of these recursion relations dramatically and prove that they reproduce the correct algebraic expressions to all orders.
- We solve our recursion relations by iteration in the single-input case and show that the solution matches empirical results for tanh-MLPs. We analytically compute a large-depth expansion of the solution and show that it supports the saturation observed empirically. We furthermore extend criticality results for the neural network Gaussian process kernel (NNGP) and the NTK from Gaussian to orthogonal initializations.

2 Related work

Orthogonal initializations of neural network weights have been considered for a long time (Ngiam et al., 2010; Saxe et al., 2014). An ablation showing performance boosts by orthogonal initializations can be found in Mishkin & Matas (2016). See Day et al. (2025) for a more recent direct Gaussian–orthogonal comparison.

Theoretical studies of orthogonal initializations mainly consider linear neural networks (Saxe et al., 2014) or use a mean-field theory approach which requires the infinite-width limit. In the latter case, a considerable amount of work has been done on dynamical isometry, which requires orthogonal initializations and supposes that the singular values of the Jacobian of the network are all close to one. With the help of free probability theory, it could be shown that ReLU nonlinearities destroy the dynamical isometry present in deep linear networks with orthogonal initialization, but the sigmoid nonlinearity retains it (Pennington et al., 2017). These results have been extended to further nonlinearities (Pennington et al., 2018) and to convolutional networks (Xiao et al., 2018). All these studies focus on initialization but convergence rates for orthogonally initialized networks were also derived in the linear setting (Hu et al., 2019), confirming the faster convergence in this case.

The NTK for orthogonal initializations at infinite width was found to agree with that of Gaussian initializations (Huang et al., 2021), motivating the study of finite-width corrections. The approach used in this article was introduced in Yaida (2020), with further details available in a textbook (Roberts & Yaida, 2022), yielding results about symmetries (Maiti et al.), initialization stability (Banta et al., 2024) and scaling laws (Maloney et al., 2022; Zhang, 2025) among others. Our Feynman-diagrammatic treatment of the finite-width corrections is based on the rules introduced in Guillen et al. (2025) which include all tensors appearing in the first correction to the training dynamics. Earlier papers using Feynman diagrams to capture preactivation statistics include Dyer & Gur-Ari (2020); Banta et al. (2024); Halverson et al. (2021); Grosvenor & Jefferson (2022); Demirtas et al. (2024); Maloney et al. (2022); Zhang (2025).

This article is most closely related to Day et al. (2025), where finite-width corrections were first considered in the orthogonal case. We substantially extend the results in this reference, as outlined above.

3 Orthogonal neural networks

We consider an L-layer multilayer perceptron (MLP) $\mathcal{N} : \mathbb{R}^{n_{\text{in}}} \to \mathbb{R}^{n_{\text{out}}}$, defined recursively by the feed-forward equations $z_i^{(\ell)}(x) = \sum_{j=1}^{n} W_{ij}^{(\ell)}\, \sigma\Big(z_j^{(\ell-1)}(x)\Big), i = 1, \ldots, n$, where $x \in \mathbb{R}^{n_{\text{in}}}$ denotes the input, indices $i, j, \ldots$ label neurons, σ is a pointwise activation function, and the weight matrices $W^{(\ell)} \in \mathbb{R}^{n\times n}$ are sampled independently from the orthogonal group with respect to the Haar measure. For simplicity, we assume that all hidden layers have the same width $n_\ell = n$, and that biases are omitted.

The statistics of matrices sampled from the orthogonal group are characterized by so-called Weingarten functions $\mathcal{W}$ (Collins & Śniady, 2006; Collins & Matsumoto, 2009), see Appendix B for a summary. Specifically, the joint moments of the entries of a matrix W drawn from the orthogonal Haar measure $\text{Haar}(\text{O}(n))$ and scaled by $\sqrt{C_W}$ are given by

$$\mathbb{E}\Big[W_{i_1 j_1}^{(\ell)} \cdots W_{i_{2k} j_{2k}}^{(\ell)}\Big] = (C_W)^k \sum_{\pi,\sigma\in\mathcal{P}_2(2k)} \delta_i^\pi\, \delta_j^\sigma\, \mathcal{W}\big(\pi^{-1}\sigma\big)\,, \tag{1}$$

where C_W is a layer-independent hyperparameter, $\mathcal{P}_2(2k)$ denotes the set of pair partitions of $\{1, \ldots, 2k\}$, and $i = (i_1, \ldots, i_{2k})$ is a tuple of free indices. For each pairing $\pi \in \mathcal{P}_2(2k)$, we define $\delta_i^\pi \coloneqq \prod_{\{a,b\}\in\pi} \delta_{i_a i_b}$, where $\delta_{i_a i_b}$ is the standard Kronecker delta. Thus, $\delta_i^\pi = 1$ if $i_a = i_b$ for every pair $\{a, b\} \in \pi$, and $\delta_i^\pi = 0$ otherwise. For illustration, the second- and fourth-order moments are given by

$$\mathbb{E}[W_{i_1 j_1}^{(\ell)} W_{i_2 j_2}^{(\ell)}] = C_W (12)_i (12)_j \mathcal{W}[1] \tag{2}$$

$$\mathbb{E}[W_{i_1 j_1}^{(\ell)} W_{i_2 j_2}^{(\ell)} W_{i_3 j_3}^{(\ell)} W_{i_4 j_4}^{(\ell)}] = (C_W)^2 \begin{pmatrix} (12)(34)_i \\ (13)(24)_i \\ (14)(23)_i \end{pmatrix}^T \begin{pmatrix} \mathcal{W}[1,1] & \mathcal{W}[2] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[1,1] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[2] & \mathcal{W}[1,1] \end{pmatrix} \begin{pmatrix} (12)(34)_j \\ (13)(24)_j \\ (14)(23)_j \end{pmatrix} \tag{3}$$

where we employ the compact notation $(12)(34)\cdots(2k-1\,2k)_i \equiv \delta_{i_1 i_2}\, \delta_{i_3 i_4} \cdots \delta_{i_{2k-1} i_{2k}}$, and the corresponding orthogonal Weingarten functions are given by

$$\mathcal{W}[1] = \frac{1}{n}, \qquad \mathcal{W}[1,1] = \frac{n+1}{(n-1)n(n+2)}, \qquad \mathcal{W}[2] = -\frac{1}{(n-1)n(n+2)}. \tag{4}$$

Consequently, in contrast to the standard Gaussian setting where Wicks theorem implies vanishing higher-order cumulants, the cumulants of the orthogonal weights W, i.e. the connected parts of their joint moments, are nonzero and play a central role in determining the finite-width statistics of the neural network ensemble $\mathcal{N}$.

As discussed in Roberts & Yaida (2022), effective field theory (EFT) techniques enable the systematic computation of preactivation statistics to all orders in $1/n$. Within this framework, the neural network Gaussian process (NNGP) kernel $\widehat{K}_{i_1 i_2}^{(\ell+1)}(x_1, x_2) = z_{i_1}^{(\ell+1)}(x_1) z_{i_2}^{(\ell+1)}(x_2)$, defines the expectation $\mathbb{E}[\widehat{K}_{ii}^{(\ell+1)}(x_1, x_2)]$. At leading order in $1/n$, this kernel satisfies the recursion

$$K^{(\ell+1)}(x_1, x_2) = C_W \langle \sigma_1^{(\ell)} \sigma_2^{(\ell)} \rangle_{K^{(\ell)}} \tag{5}$$

where we use the shorthands $K^{(\ell)} = \lim_{n\to\infty} \mathbb{E}[\widehat{K}^{(\ell)}_{ii}]$, $\sigma^{(\ell)}_a \equiv \sigma\big(z^{(\ell)}(x_a)\big)$, and $\langle\cdot\rangle_K$ denotes Gaussian expectation with covariance K.

At next-to-leading order, the fourth-order cumulant of the preactivations can be decomposed as

$$\mathbb{E}^c[z^{(\ell)}_{i_1}(x_1)z^{(\ell)}_{i_2}(x_2)z^{(\ell)}_{i_3}(x_3)z^{(\ell)}_{i_4}(x_4)]=\frac{1}{n}\Big[(12)(34)_i V^{(\ell)}_{1234}+(13)(24)_i V^{(\ell)}_{1324}+(14)(23)_i V^{(\ell)}_{1423}\Big] \quad (6)$$

where $\mathbb{E}^c$ denotes the joint cumulant and we introduced $V^{(\ell+1)}_{1234} = V^{(\ell+1)}(x_1,x_2,x_3,x_4)$ for the coefficients. The function $V^{(\ell+1)}_{1234}$ obeys the recursion (Day et al., 2025)

$$V^{(\ell+1)}_{1234}=(C_W)^2\Big[\langle\sigma^{(\ell)}_1\sigma^{(\ell)}_2\sigma^{(\ell)}_3\sigma^{(\ell)}_4\rangle_{K^{(\ell)}}-\langle\sigma^{(\ell)}_1\sigma^{(\ell)}_2\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_3\sigma^{(\ell)}_4\rangle_{K^{(\ell)}}-\langle\sigma^{(\ell)}_1\sigma^{(\ell)}_3\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_2\sigma^{(\ell)}_4\rangle_{K^{(\ell)}}$$
$$-\langle\sigma^{(\ell)}_1\sigma^{(\ell)}_4\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_2\sigma^{(\ell)}_3\rangle_{K^{(\ell)}}\Big]+\frac{(C_W)^2}{4}\sum_{\substack{\beta_i\in\{1,2,3,4\}\\ i=1,\dots,4}} V^{(\ell)}_{\beta_1\beta_2\beta_3\beta_4}\left\langle\frac{d^2(\sigma^{(\ell)}_1\sigma^{(\ell)}_2)}{dz^{(\ell)}_{\beta_1}dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle\frac{d^2(\sigma^{(\ell)}_3\sigma^{(\ell)}_4)}{dz^{(\ell)}_{\beta_3}dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}} \quad (7)$$

As noted in Day et al. (2025), the recursion equation 7 differs from its Gaussian counterpart by the last two terms in the square brackets. As we show below, this small change has significant consequences for information processing in orthogonal networks.

3.1 The NTK at finite-width

The layer-ℓ empirical (NTK) associated with the network $\mathcal{N}$ is defined as

$$\widehat{\Theta}^{(\ell)}_{ij}(x_1,x_2) = \sum_{\mu,\ell'} \lambda^{(\ell')}_\theta \frac{\partial z^{(\ell)}_i(x_1)}{\partial\theta^{(\ell')}_\mu}\frac{\partial z^{(\ell)}_j(x_2)}{\partial\theta^{(\ell')}_\mu}, \quad (8)$$

where $\lambda^{(\ell')}_\theta$ is a layer-dependent training hyperparameter, $\theta^{(\ell')}_\mu$ denotes the weight parameters at layer ℓ', and the sum runs over all parameters in layers $\ell' \le \ell$. The NTK of the full network is given by $\widehat{\Theta} = \widehat{\Theta}^{(L)}$. As is well known, this kernel is a nonlinear, initialization-dependent object that evolves during training. In the small learning-rate regime, the NTK provides a leading-order description of the training dynamics. Consequently, understanding its behavior is important for analyzing neural network learning dynamics (Jacot et al., 2018).

To compute the statistics of the NTK, we employ the chain rule to derive the following forward recursion:

$$\widehat{\Theta}^{(\ell+1)}_{i_1i_2}(x_1,x_2)=(12)_i\left(\lambda^{(\ell+1)}_b+\frac{\lambda^{(\ell+1)}_W}{n}\sum_{j=1}^n\sigma^{(\ell)}_{j,1}\sigma^{(\ell)}_{j,2}\right)+\sum_{j_1,j_2=1}^n W^{(\ell+1)}_{i_1j_1}W^{(\ell+1)}_{i_2j_2}\sigma'^{(\ell)}_{j_1,1}\sigma'^{(\ell)}_{j_2,2}\widehat{\Theta}^{(\ell)}_{j_1j_2}(x_1,x_2), \quad (9)$$

where the standard rescaling $\lambda^{(\ell+1)}_W \mapsto \lambda^{(\ell+1)}_W/n$ has been applied to ensure that all network parameters contribute at the same order. As discussed in Roberts & Yaida (2022), an appropriate choice of the depth-dependence of the training hyperparameters depends on the activation function under consideration.

The fluctuations of the NTK are defined as $\widehat{\Delta\Theta}^{(\ell)} \equiv \widehat{\Theta}^{(\ell)} - \mathbb{E}[\widehat{\Theta}^{(\ell)}]$ and we use the shorthand $\Theta^{(\ell)} = \lim_{n\to\infty}\mathbb{E}[\widehat{\Theta}^{(\ell)}]$. Applying the EFT techniques developed in Roberts & Yaida (2022) to the recursion equation 9 yields the following leading-order equation for the NTK mean

$$\Theta^{(\ell+1)}_{12} = \lambda^{(\ell+1)}_b + \lambda^{(\ell+1)}_W\big\langle\sigma^{(\ell)}_1\sigma^{(\ell)}_2\big\rangle_{K^{(\ell)}} + C_W\big\langle\sigma'^{(\ell)}_1\sigma'^{(\ell)}_2\big\rangle_{K^{(\ell)}}\Theta^{(\ell)}_{12} \quad (10)$$

As in the Gaussian case, the leading-order statistics of the NTK in $1/n$ are determined by the cross-cumulant between the preactivations and the NTK fluctuation, together with the NTK variance (Roberts & Yaida, 2022). The former admits the following compact expression

$$\mathbb{E}^c_\theta[z^{(\ell)}_{i_1}(x_1)z^{(\ell)}_{i_2}(x_2)\widehat{\Delta\Theta}^{(\ell)}_{i_3i_4}(x_3,x_4)]=\frac{1}{n}\Big(D^{(\ell)}_{12\mathbf{34}}(12)(34)_i+F^{(\ell)}_{13\mathbf{24}}(13)(24)_i+F^{(\ell)}_{14\mathbf{23}}(14)(23)_i\Big) \quad (11)$$

where the tensors D and F are defined in Appendix A together with all other tensors we use. As shown in Appendix C, a careful algebraic analysis yields, for instance, the following leading-order $1/n$ recursion relations for the tensor F

$$F^{(\ell+1)}_{1324} = (C_W)^2 \left[\langle \sigma_1^{(\ell)} \sigma_2^{(\ell)} \sigma_3'^{(\ell)} \sigma_4'^{(\ell)} \rangle_{K^{(\ell)}} - \langle \sigma_1^{(\ell)} \sigma_2^{(\ell)} \rangle_{K^{(\ell)}} \langle \sigma_3'^{(\ell)} \sigma_4'^{(\ell)} \rangle_{K^{(\ell)}} \right] \Theta^{(\ell)}_{34} + (C_W)^2 \sum_{\alpha,\beta,\gamma,\delta=1}^{4} \langle \sigma_1^{(\ell)} \sigma_3'^{(\ell)} z_\alpha^{(\ell)} \rangle_{K^{(\ell)}} \langle \sigma_2^{(\ell)} \sigma_4'^{(\ell)} z_\beta^{(\ell)} \rangle_{K^{(\ell)}} K^{\alpha\gamma}_{(\ell)} K^{\beta\delta}_{(\ell)} F^{(\ell)}_{\gamma 3 \delta 4} + \mathcal{O}\left(\frac{1}{n}\right) . \quad (12)$$

As we will show, this recursion can be derived efficiently using a diagrammatic framework based on a consistent set of Feynman rules and graph-theoretic principles, yielding a substantial simplification over conventional algebraic approaches. An analogous construction applies to the tensor D, as well as to the NTK variance tensors A and B (Roberts & Yaida, 2022), see Appendix C for details.

4 Feynman Diagrams

A novel diagrammatic framework, inspired by the Feynman-diagram description of interactions among fundamental particles in high-energy physics, was recently introduced for the study of deep neural networks with Gaussian initializations (Guillen et al., 2025; Banta et al., 2024). In this section, we extend this framework to incorporate the orthogonality properties of the network parameters and present a self-contained set of Feynman rules for the orthogonal case.

4.1 Feynman Rules

Without loss of generality, consider the channel $(12)_i(34)_i \dots (2k\,2k-1)_i$, corresponding to the pairing $(12)(34)\dots(2k\,2k-1)$. The Feynman rules reproducing tensors with this channel structure can be organized into two groups. The first group implements the orthogonality constraints as follows:

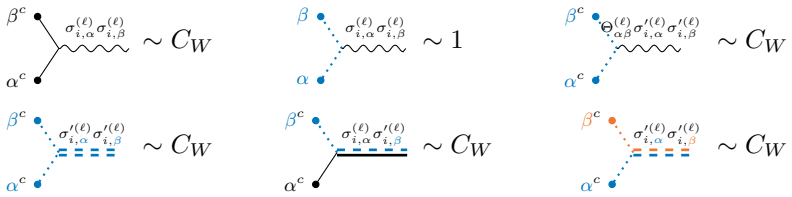
1. Preactivations and NTKs are denoted by external lines, as illustrated below.



(13)

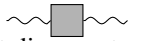
   In equation 13, a colored line represents a single NTK label. Distinct colors are used for external dotted lines associated with different NTKs. The colors of the dotted lines correspond to the θ-indices appearing in these definitions. Since the θ-indices are contracted in pairs, distinct colors encode the corresponding contraction pattern.

2. Define the cubic vertices as



(14)

   Here the superscript c, referred to as the orthogonality charge, keeps track of the label's orthogonal character, and lines that do not end in a dot represent internal lines.

3. Draw a square propagator connecting internal lines in all possible ways, consistent with the chosen pairing. The square  represents the full expectation value $\mathbb{E}[\cdot]$. This procedure generates distinct diagram types, both connected and disconnected. The connected diagrams are further classified as s-class diagrams, defined by the number s of square propagators apppearing in the diagram. The square propagator obeys an additional selection rule: all k-class diagrams formed from pairs of external lines corresponding to different object types vanish.

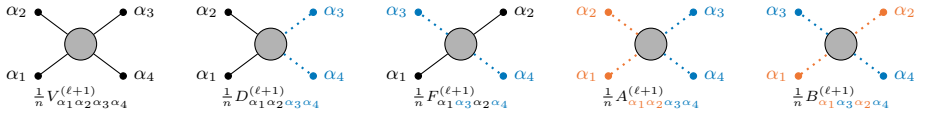
4. For each s-class diagram, generate all inequivalent permutations of its $2m$ external labels carrying orthogonality charge. Multiply each resulting diagram by $\frac{1}{n}$ for every uncharged pairing, and by the appropriate m-class Weingarten function $\mathcal{W}$, determined by the relative ordering τ of the diagram's labels with respect to the original pairing $\pi = (12)(34)\dots(2k\,2k-1)$:

$\mathcal{W}[\tau, \pi] = \mathcal{W}[\ell(\tau \circ \pi)]$, where $\circ$ denotes ordinary permutation multiplication and $\ell(e)$ denotes the cycle length of e. See Appendix B for a detailed discussion of Weingarten functions and their properties.

5. Multiply each s-class contribution by the Möbius coefficient $(-1)^{s-1}(s-1)!$, and sum over all classes.

The second group implements the effective field theory techniques developed in Roberts & Yaida (2022), applied to the square propagator in the diagrammatic construction of the previous step, through the following set of Feynman rules analogous to those introduced in Guillen et al. (2025); Banta et al. (2024):

6. We define the bare propagator as $\langle \quad \rangle_{K^{(\ell)}} \equiv$  where $\langle \quad \rangle_{K^{(\ell)}}$ denotes a zero-mean Gaussian expectation with covariance specified by $K^{(\ell)}$. The expectation value is taken over the decorations of the internal lines attached to the propagator, which obeys the same selection rules described in Guillen et al. (2025). These rules are summarized as follows:
   (a) Propagators may only connect to internal lines emanating from cubic vertices or from the internal quartic vertices introduced below. In particular, propagators cannot be directly connected to other propagators.
   (b) Dotted lines attached to a propagator do not enter the Gaussian expectation value, as they carry no decorations.
   (c) Each preactivation line decorated with z_i acts as a derivative with respect to z_i acting on the argument of the Gaussian expectation value.
   (d) The neural indices of all internal lines connected to a propagator must be identical.
   (e) If both dotted and dashed lines of the same color are attached to the propagator, they must appear in pairs carrying the same sample index. The two lines in each pair attach to different vertices. Moreover, if both vertices are drawn in the orientation specified in the Feynman rules, the top-to-bottom ordering of the sample indices (and therefore the colors) of the lines connected to the two vertices must coincide.
   (f) A pair of dashed lines of the same color connected to the propagator contributes a factor $\Theta_{\alpha\beta}$ when the two lines attach to different vertices, where α and β denote the sample indices of the pair.
7. Quartic vertices are defined analogously, following Guillen et al. (2025). Explicitly,



(15)

8. Higher-order NTK and preactivation tensors are introduced via a natural generalization of the vertices in equation 15.
9. The square propagator decomposes into all connected and disconnected diagrams built from the bare propagator, quartic vertices, and higher-order vertices, with internal lines remaining undotted. This decomposition respects the selection rules (a)-(f).

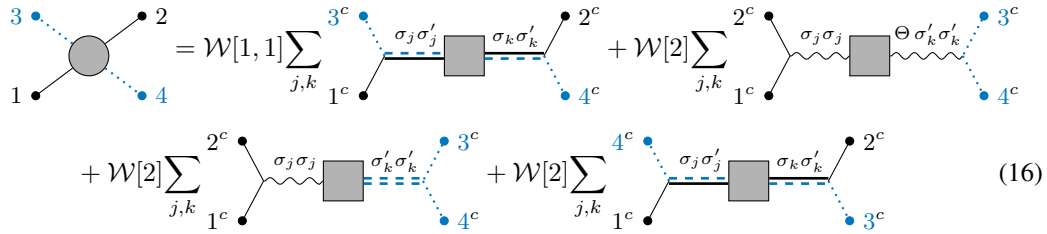
In Appendix G.3, we prove that the Feynman rules (1)-(9) are complete to all orders in the 1/n expansion. We further derive a simplified formulation valid at leading order in $1/n$ in Appendix F. As a consistency check, we recover the results of Banta et al. (2024) and Guillen et al. (2025) for preactivations and the NTK with Gaussian weights by restricting to the leading diagonal Weingarten contributions.

These Feynman rules satisfy the following theorem:

Theorem 4.1. *The Feynman rules stated in items 1-9, uniquely determine the recursion relations governing the layerwise evolution of the orthogonal NTK tensors D, F, A, B at order $1/n$.*

Proof. We illustrate the proof using the tensor F, deriving the recursion relation equation 12 directly from the Feynman rules above. The analysis for the tensors D, A and B is analogous and deferred to Appendix G.1, completing the proof.

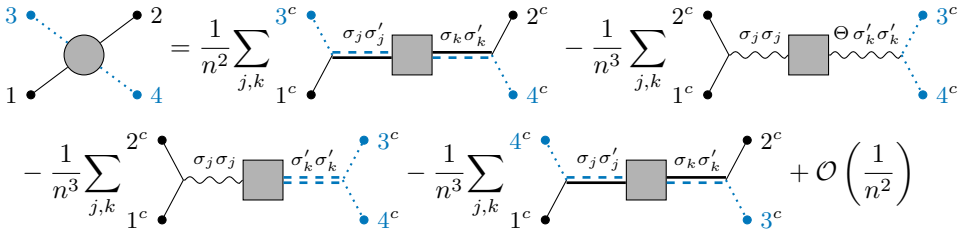
The Feynman rules (1)-(5) produce the following diagrams for the tensor F:



(16)

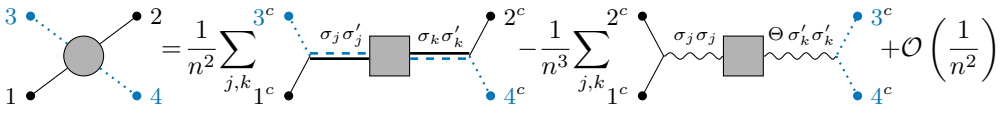
The absence of 2-class diagrams in equation 16 follows from the fact that the square propagator vanishes when acting on a pair of lines of different type. The 1-class diagrams are generated from the inequivalent permutations of the reference pairing $(13)(24)$, namely $\{(13)(24), (12)(34), (14)(23)\}$. The corresponding $k = 2$ Weingarten functions multiplying the subdiagrams are determined by the relative ordering of the reference pairing τ and the pairing π appearing in the subdiagram: $\mathcal{W}[\pi, \tau] = \mathcal{W}[1, 1]$ for $\pi = \tau$, $\mathcal{W}[\pi, \tau] = \mathcal{W}[2]$ for $\pi \neq \tau$. The overall factor of 1 follows from the Möbius relation for a single-block partition.

At first order in $1/n$, the Weingarten functions can be approximated by their leading terms in the expansions equation 54 and equation 55. In this manner, equation 16 simplifies to



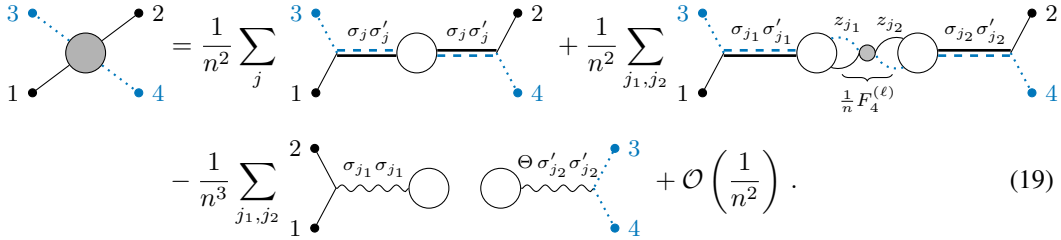
(17)

We now apply the second set of Feynman rules (6)–(9). As an immediate consequence, the last terms in equation 17 vanish at order $1/n$. A nonzero contribution would require the corresponding diagrams to produce a factor of n^2, which occurs when $j \neq k$ and a square propagator decomposes into two bare propagators forming disconnected subdiagrams. However, such configurations are forbidden by the selection rules (a)–(f), as each subdiagram violates color conservation. Consequently, we are left with



(18)

We now expand the diagrams in equation 18 in terms of the bare propagator and quartic vertices by analyzing the diagonal and off-diagonal neural components of each subdiagram, together with the selection rules (a)-(f). This results in



(19)

When $i = j$, the first diagram in equation 18 contributes as the first diagram in equation 19, while when $i \neq j$, the first and second diagrams in equation 18 reduce respectively to the second and third diagrams in equation 19. The $1/n$ scaling of the diagrams follows from straightforward power counting of the respective summations: When $i = j$, the summation contributes factor of n, whereas when $i \neq j$ it yields a factor of n^2. Combined with the explicit coefficients on the right-hand side of equation 19, these factors produce an overall contribution of order $1/n$. It is straightforward to verify that the first and second lines on the RHS of equation 19 reproduce the corresponding lines on the RHS of equation 12. □

In the same manner, the full Feynman rules can be used to reproduce the recursion relations governing higher-derivative versions of the NTK, as detailed below.

Theorem 4.2. *The set of Feynman rules presented above, in conjunction with those defined in Appendix D, reproduces the recursion relations governing the orthogonal dNTK and ddNTK tensors: P, Q and R, S, T, U, respectively, at order $1/n$.*

Proof. See Appendix G.2. □

The preceding results establish that Feynman diagrams are complete at order $1/n$. This construction extends straightforwardly to all orders in $1/n$ by allowing the vertices in 15 to admit an arbitrary number of external legs, without requiring any further modifications. The following theorem formalizes this extension.

Theorem 4.3. *The Feynman rules defined above, augmented by the higher-order generalizations of the tensors V, D, F, A, B, P, Q, R, S, T, U, provide a complete characterization of the statistics of the orthogonal NTK and its descendants at arbitrary order in $1/n$.*

Proof. See Appendix G.3. □

This result demonstrates that our diagrammatic formalism enables systematic computation of preactivation and NTK statistics to arbitrary order in $1/n$ from a concise set of rules. As a concrete example, we derive in Appendix I the recursion relation for the V_6 tensor at order $1/n^2$, in contrast to the cumbersome algebraic manipulations demanded by direct methods.

5 Applications

In this section, we deduce a set of results for orthogonal neural networks from the Feynman-diagrammatic framework developed above.

5.1 Criticality and stability at finite width

Deep neural network outputs are highly sensitive to initialization: naive architectural choices can cause preactivations to explode or vanish exponentially. At infinite width, this stability is characterized by the susceptibility χ, which controls the growth of both the NNGP and the NTK. This analysis extends to finite-width Gaussian-initialized networks, where criticality at infinite width enforces criticality at finite width for both preactivation and NTK statistics (Guillen et al., 2025; Banta et al., 2024).

Next, we combine the ideas developed in Guillen et al. (2025) with the Feynman rules of Section 4 to show that

Theorem 5.1. *Criticality of the infinite-width NNGP and NTK implies criticality of orthogonal preactivation and NTK cumulants.*

Proof. See Appendix G.4. □

In Figure 1, we empirically verify this by sampling the NTK components of a $\tanh$ orthogonal MLP ($n = 50$, $L = 30$) with inputs drawn i.i.d. from the interval $(0, 1)$ across varying C_W (see Appendix K). This illustrates the effectiveness of the diagrammatic framework for deriving results to all orders in $1/n$.

5.2 Single-input solutions

Once derived within the Feynman diagrammatic framework, the recursion relations can be solved for given initial conditions. We focus on the single-input setting with $\tanh$ activation and solve the resulting recursions for the NTK tensors (see Appendix H for an explicit derivation). We consider networks of width $n = 50$ and depth $L = 10$, initialized at criticality $C_W = 1$. Gaussian expectations are evaluated numerically in *Mathematica*. We analyze normalized cumulants, defined as ratios to their

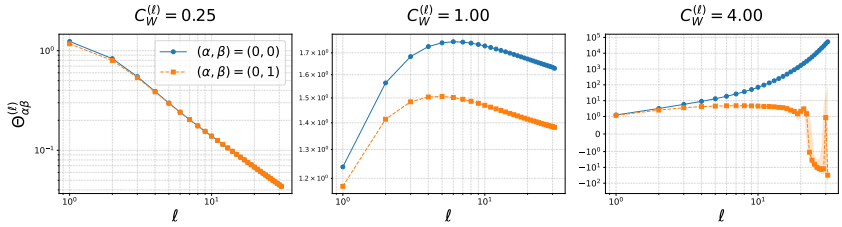


Figure 1: *Gradient stability.* Monte Carlo estimates of the NTK $\Theta_{\alpha\beta}^{(\ell)}$ for a $\tanh$ orthogonal network ($n = 50$, $L = 30$) across varying C_W, with the critical case $C_W = 1$ shown in the center. Means are computed over 600 initializations; shaded regions denote standard errors (typically not visible). Results are consistent with Theorem 5.1.

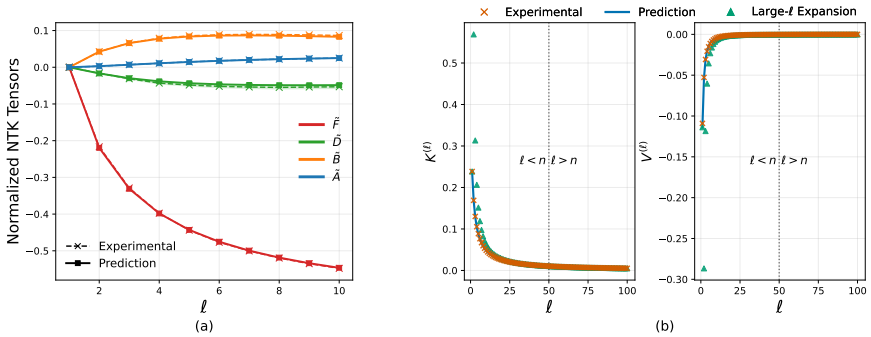


Figure 2: *Orthogonal saturation.* (a) Normalized NTK tensors from Monte Carlo simulations of $\tanh$ orthogonal networks ($n = 50$, $L = 10$) at criticality, compared with theoretical predictions. Means are computed over 600 initializations; shaded regions denote standard deviations (typically not visible), showing quantitative agreement. (b) Large-ℓ expansions of the NNGP and quartic vertex V, compared with exact solutions and simulations. While discrepancies appear at small ℓ, asymptotic predictions are in precise agreement at large ℓ. The theory captures the regime $\ell > n$, consistent with Day et al. (2025), where orthogonal networks exhibit saturation at large depth, unlike the Gaussian case.

infinite-width counterparts; for example, $\tilde{D}^{(\ell)} = D^{(\ell)}/(K^{(\ell)}\Theta^{(\ell)})$ and $\tilde{F}^{(\ell)} = F^{(\ell)}/(K^{(\ell)}\Theta^{(\ell)})$. The solutions are shown in Figure 2 (a), where we observe excellent agreement with finite-width simulations, supporting the validity of our framework. See Appendix H.1 for a complete analysis, including NTK derivatives. Our results are consistent with the empirical findings of Day et al. (2025).

5.3 Large-ℓ Expansions

The recursion relations derived above admit analytic solutions in the large-depth regime. In this limit, one can expand any function $\mathcal{O}(\ell)$ in a large-depth asymptotic (polyhomogeneous) expansion as (Olver, 1997; Melrose, 1993)

$$\mathcal{O}(\ell) = \ell^{-p_{\mathcal{O}}} \sum_{p,q=1}^{\infty} c_{p,q}^{\mathcal{O}} \frac{\log^p \ell}{\ell^q} \tag{20}$$

into the recursion equations and determine the coefficients $c_{p,q}^{\mathcal{O}}$ systematically once the initial conditions are fixed. We derive large-ℓ expansions for the NNGP and the quartic vertex V in the single-input setting, and compare them with exact solutions and Monte Carlo simulations in Figure 2 (b). The analytical expansion agrees well with the empirical observations for large ℓ. Complete derivations of the expansions equation 20 for the NTK tensors and their derivatives are provided in Appendix J.

Reproducibility Statement

The theoretical analysis is self-contained: all tensors entering the finite-width expansion are explicitly defined, and the complete Feynman rules used to derive the recursion relations are presented in the main text and appendices. Detailed derivations and proofs are provided in the appendices. The network architecture, initialization scheme, hyperparameters, and sampling procedure used in the numerical experiments are also specified, enabling reproduction of the reported results.

References

Ian Banta, Tianji Cai, Nathaniel Craig, and Zhengkang Zhang. Structures of neural network effective theories. *Physical Review D*, 109(10):105007, May 2024. doi: 10.1103/PhysRevD.109.105007.

Benoît Collins and Sho Matsumoto. On some properties of orthogonal weingarten functions. *Journal of Mathematical Physics*, 50(11):113516, 2009. doi: 10.1063/1.3251304.

Benoît Collins and Piotr Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic groups. *Communications in Mathematical Physics*, 264(3):773–795, 2006. doi: 10.1007/s00220-006-1554-3.

Hannah Day, Yonatan Kahn, and Daniel A Roberts. Feature learning and generalization in deep networks with orthogonal weights. *Machine Learning: Science and Technology*, 6(3):035027, August 2025. ISSN 2632-2153. doi: 10.1088/2632-2153/adf278.

Mehmet Demirtas, James Halverson, Anindita Maiti, Matthew D Schwartz, and Keegan Stoner. Neural network field theories: Non-Gaussianity, actions, and locality. *Machine Learning: Science and Technology*, 5(1):015002, January 2024. ISSN 2632-2153. doi: 10.1088/2632-2153/ad17d3.

Ethan Dyer and Guy Gur-Ari. Asymptotics of wide networks from feynman diagrams. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1gFvANKDS.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, March 2010.

Kevin Grosvenor and Ro Jefferson. The edge of chaos: Quantum field theory and deep neural networks. *SciPost Physics*, 12(3):081, March 2022. ISSN 2542-4653. doi: 10.21468/SciPostPhys.12.3.081.

Max Guillen, Philipp Misof, and Jan E. Gerken. Finite-Width Neural Tangent Kernels from Feynman Diagrams, August 2025.

James Halverson, Anindita Maiti, and Keegan Stoner. Neural Networks and Quantum Field Theory. *Machine Learning: Science and Technology*, 2(3):035002, September 2021. ISSN 2632-2153. doi: 10.1088/2632-2153/abeca3.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

Wei Hu, Lechao Xiao, and Jeffrey Pennington. Provable Benefit of Orthogonal Initialization in Optimizing Deep Linear Networks. In *International Conference on Learning Representations*, September 2019.

Wei Huang, Weitao Du, and Richard Yi Da Xu. On the Neural Tangent Kernel of Deep Networks with Orthogonal Initialization. In *Twenty-Ninth International Joint Conference on Artificial Intelligence*, volume 3, pp. 2577–2583, August 2021. doi: 10.24963/ijcai.2021/355.

Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural Tangent Kernel: Convergence and Generalization in Neural Networks. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Anindita Maiti, Keegan Stoner, and James Halverson. Symmetry-via-duality: Invariant neural network densities from parameter-space correlators. In *Machine Learning in Pure Mathematics and Theoretical Physics*, chapter Chapter 8, pp. 293–330. doi: 10.1142/9781800613706_0008.

Alexander Maloney, Daniel A. Roberts, and James Sully. A Solvable Model of Neural Scaling Laws, October 2022.

Richard B. Melrose. *The Atiyah–Patodi–Singer Index Theorem*. A K Peters, Wellesley, MA, 1993.

Dmytro Mishkin and Jiri Matas. All you need is a good init. In *International Conference on Learning Representations 2016*. arXiv, February 2016.

Jiquan Ngiam, Zhenghao Chen, Daniel Chia, Pang Koh, Quoc Le, and Andrew Ng. Tiled convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

Frank W. J. Olver. *Asymptotics and Special Functions*. A K Peters, Wellesley, MA, 1997. Reprint of the 1974 original.

Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. Resurrecting the sigmoid in deep learning through dynamical isometry: Theory and practice. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. The emergence of spectral universality in deep networks. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pp. 1924–1932. PMLR, March 2018.

Roger Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413, 1955. doi: 10.1017/S0305004100030401.

Daniel A. Roberts and Sho Yaida. *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks*. Cambridge University Press, Cambridge, 2022. ISBN 978-1-316-51933-2. doi: 10.1017/9781009023405.

Andrew M. Saxe, James L. McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks, February 2014.

Don Weingarten. Asymptotic behavior of group integrals in the limit of infinite rank. *Journal of Mathematical Physics*, 19(5):999–1001, May 1978. ISSN 0022-2488. doi: 10.1063/1.523807.

Lechao Xiao, Yasaman Bahri, Jascha Sohl-Dickstein, Samuel Schoenholz, and Jeffrey Pennington. Dynamical Isometry and a Mean Field Theory of CNNs: How to Train 10,000-Layer Vanilla Convolutional Neural Networks. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 5393–5402. PMLR, July 2018.

Sho Yaida. Non-Gaussian processes and neural networks at finite widths. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, pp. 165–192. PMLR, August 2020.

Zhengkang Zhang. Neural scaling laws from large-N field theory: Solvable model beyond the ridgeless limit. *Machine Learning: Science and Technology*, 6(2):025010, April 2025. ISSN 2632-2153. doi: 10.1088/2632-2153/adc872.

A Definitions of all tensors

In this section, we collect the definitions of the tensors used throughout this work, both at infinite width and at the leading finite-width order.

The ensemble means of the empirical NNGP and NTK receive corrections at order $1/n$. We define these corrections by

$$\mathbb{E}[\widehat{K}^{(\ell)}_{\alpha_1\alpha_2}] = K^{(\ell)}_{\alpha_1\alpha_2} + \frac{1}{n_{\ell-1}} K^{\{1\}(\ell)}_{\alpha_1\alpha_2} + O\left(\frac{1}{n^2}\right) \tag{21}$$

$$\mathbb{E}[\widehat{\Theta}^{(\ell)}_{\alpha_1\alpha_2}] = \Theta^{(\ell)}_{\alpha_1\alpha_2} + \frac{1}{n_{\ell-1}}\Theta^{\{1\}(\ell)}_{\alpha_1\alpha_2} + O\left(\frac{1}{n^2}\right), \tag{22}$$

Here, the unhatted quantities denote the corresponding infinite-width limits.

The cumulant of four preactivations defines the tensor V through the decomposition

$$\mathbb{E}^c_\theta\left[z^{(\ell)}_{i_1;\alpha_1} z^{(\ell)}_{i_2;\alpha_2} z^{(\ell)}_{i_3;\alpha_3} z^{(\ell)}_{i_4;\alpha_4}\right] = \frac{1}{n_{\ell-1}}\Big[(12)(34)_i V^{(\ell)}_{\alpha_1\alpha_2\alpha_3\alpha_4} + (13)(24)_i V^{(\ell)}_{\alpha_1\alpha_3\alpha_2\alpha_4} + (14)(23)_i V^{(\ell)}_{\alpha_1\alpha_4\alpha_2\alpha_3}\Big]. \tag{23}$$

We use the shorthand $(ab)(cd)_i \equiv \delta_{i_a i_b}\delta_{i_c i_d}$ for Kronecker-delta contractions of the neuron indices, as introduced in Section 3.

The cumulant of two NTK fluctuations defines the tensors A and B:

$$\mathbb{E}^c_\theta\left[\widehat{\Delta\Theta}^{(\ell)}_{i_1 i_2;\alpha_1\alpha_2}\widehat{\Delta\Theta}^{(\ell)}_{i_3 i_4;\alpha_3\alpha_4}\right] = \frac{1}{n_{\ell-1}}\Big[(12)(34)_i A^{(\ell)}_{\alpha_1\alpha_2\alpha_3\alpha_4} + (13)(24)_i B^{(\ell)}_{\alpha_1\alpha_3\alpha_2\alpha_4} + (14)(23)_i B^{(\ell)}_{\alpha_1\alpha_4\alpha_2\alpha_3}\Big]. \tag{24}$$

A consistent analysis of the training dynamics through order $1/n$ also requires the following higher-derivative analogues of the NTK (Roberts & Yaida, 2022):

$$\widehat{\mathrm{d}\Theta}^{(\ell)}_{i_0 i_1 i_2;\delta_0\delta_1\delta_2} = \sum_{\ell_1,\ell_2=1}^{\ell}\sum_{\mu_1,\mu_2}\frac{d^2 z^{(\ell)}_{i_0;\delta_0}}{d\theta^{(\ell_1)}_{\mu_1} d\theta^{(\ell_2)}_{\mu_2}}\frac{dz^{(\ell)}_{i_1;\delta_1}}{d\theta^{(\ell_1)}_{\mu_1}}\frac{dz^{(\ell)}_{i_2;\delta_2}}{d\theta^{(\ell_2)}_{\mu_2}} \tag{25}$$

$$\widehat{\mathrm{dd_I}\Theta}^{(\ell)}_{i_0 i_1 i_2 i_3;\delta_0\delta_1\delta_2\delta_3} = \sum_{\ell_1,\ell_2,\ell_3=1}^{\ell}\sum_{\mu_1,\mu_2,\mu_3}\frac{d^3 z^{(\ell)}_{i_0;\delta_0}}{d\theta^{(\ell_1)}_{\mu_1} d\theta^{(\ell_2)}_{\mu_2} d\theta^{(\ell_3)}_{\mu_3}}\frac{dz^{(\ell)}_{i_1;\delta_1}}{d\theta^{(\ell_1)}_{\mu_1}}\frac{dz^{(\ell)}_{i_2;\delta_2}}{d\theta^{(\ell_2)}_{\mu_2}}\frac{dz^{(\ell)}_{i_3;\delta_3}}{d\theta^{(\ell_3)}_{\mu_3}} \tag{26}$$

$$\widehat{\mathrm{dd_{II}}\Theta}^{(\ell)}_{i_1 i_2 i_3 i_4;\delta_1\delta_2\delta_3\delta_4} = \sum_{\ell_1,\ell_2,\ell_3=1}^{\ell}\sum_{\mu_1,\mu_2,\mu_3}\frac{d^2 z^{(\ell)}_{i_1;\delta_1}}{d\theta^{(\ell_1)}_{\mu_1} d\theta^{(\ell_3)}_{\mu_3}}\frac{d^2 z^{(\ell)}_{i_2;\delta_2}}{d\theta^{(\ell_2)}_{\mu_2} d\theta^{(\ell_3)}_{\mu_3}}\frac{dz^{(\ell)}_{i_3;\delta_3}}{d\theta^{(\ell_1)}_{\mu_1}}\frac{dz^{(\ell)}_{i_4;\delta_4}}{d\theta^{(\ell_2)}_{\mu_2}}. \tag{27}$$

The definitions of their cumulants define the tensors P, Q, R, S, T, and U:

$$\mathbb{E}^c_\theta\left[\widehat{\mathrm{d}\Theta}^{(\ell)}_{i_0 i_1 i_2;\delta_0\delta_1\delta_2} z^{(\ell)}_{i_3;\delta_3}\right] = \frac{1}{n_{\ell-1}}\left[(03)(12)_i P^{(\ell)}_{\delta_0\delta_1\delta_2\delta_3} + (01)(23)_i Q^{(\ell)}_{\delta_0\delta_5\delta_6\delta_7} + (02)(13)_i Q^{(\ell)}_{\delta_0\delta_2\delta_1\delta_3}\right] \tag{28}$$

$$\mathbb{E}^c_\theta\left[\widehat{\mathrm{dd_I}\Theta}^{(\ell)}_{i_0 i_1 i_2 i_3;\delta_0\delta_1\delta_2\delta_3}\right] = \frac{1}{n_{\ell-1}}\left[(01)(23)_i R^{(\ell)}_{\delta_0\delta_1\delta_2\delta_3} + (02)(13)_i R^{(\ell)}_{\delta_0\delta_2\delta_3\delta_1} + (03)(12)_i R^{(\ell)}_{\delta_0\delta_3\delta_1\delta_2}\right] \tag{29}$$

$$\mathbb{E}^c_\theta\left[\widehat{\mathrm{dd_{II}}\Theta}^{(\ell)}_{i_1 i_2 i_3 i_4;\delta_1\delta_2\delta_3\delta_4}\right] = \frac{1}{n_{\ell-1}}\left[(12)(34)_i S^{(\ell)}_{\delta_1\delta_2\delta_3\delta_4} + (13)(24)_i T^{(\ell)}_{\delta_1\delta_3\delta_4\delta_2} + (14)(23)_i U^{(\ell)}_{\delta_1\delta_4\delta_2\delta_3}\right]. \tag{30}$$

B Orthogonal Weingarten calculus

We briefly review the Weingarten calculus used in the main text. This calculus provides a systematic method for computing joint moments of the entries of a Haar-distributed orthogonal matrix. See Collins & Śniady (2006); Collins & Matsumoto (2009) for detailed treatments.

Pair partitions and Kronecker constraints. Let $\mathcal{P}_2(2k)$ denote the set of pair partitions, or pairings, of $\{1,\dots,2k\}$. Each $\pi \in \mathcal{P}_2(2k)$ consists of k disjoint unordered pairs. For example,

$$\mathcal{P}_2(4) = \big\{(12)(34),\ (13)(24),\ (14)(23)\big\}. \tag{31}$$

For a tuple of free indices $i = (i_1, \ldots, i_{2k})$, define

$$\delta_i^\pi := \prod_{\{a,b\}\in\pi} \delta_{i_a i_b}, \tag{32}$$

where $\delta_{i_a i_b}$ is the standard Kronecker delta. Thus, $\delta_i^\pi = 1$ if $i_a = i_b$ for every pair $\{a, b\} \in \pi$, and $\delta_i^\pi = 0$ otherwise. For example,

$$\delta_i^{(12)(34)} = \delta_{i_1 i_2}\delta_{i_3 i_4}. \tag{33}$$

The quantity δ_j^τ is defined analogously for $j = (j_1, \ldots, j_{2k})$ and $\tau \in \mathcal{P}_2(2k)$.

Haar integration formula. Let $Q \sim \mathrm{Haar}(\mathrm{O}(n))$. All odd-order moments of its entries vanish. Its even-order moments are given by

$$\mathbb{E}[Q_{i_1 j_1} \cdots Q_{i_{2k} j_{2k}}] = \sum_{\pi,\tau\in\mathcal{P}_2(2k)} \delta_i^\pi\, \delta_j^\tau\, \mathrm{Wg}^{\mathrm{O}(n)}(\pi, \tau), \tag{34}$$

where $\mathrm{Wg}^{\mathrm{O}(n)}(\pi, \tau)$ is the orthogonal Weingarten function.

If the weights are scaled as

$$W = \sqrt{C_W}\, Q, \tag{35}$$

then each moment of order $2k$ acquires a factor $(C_W)^k$:

$$\mathbb{E}[W_{i_1 j_1} \cdots W_{i_{2k} j_{2k}}] = (C_W)^k \sum_{\pi,\tau\in\mathcal{P}_2(2k)} \delta_i^\pi\, \delta_j^\tau\, \mathrm{Wg}^{\mathrm{O}(n)}(\pi, \tau). \tag{36}$$

Gram matrix and Weingarten coefficients. Recall that (ab) denotes a pair containing the labels a and b. Given two pairings $\pi, \tau \in \mathcal{P}_2(2k)$, we group together all labels that can be related by repeatedly following pairs from either π or τ. The resulting groups form a partition of $\{1, \ldots, 2k\}$, denoted by $\pi \vee \tau$, and $\#(\pi \vee \tau)$ denotes the number of such groups.

For example, consider

$$\pi = (12)(34), \qquad \tau = (13)(24).$$

Starting from 1, the pairing π relates 1 to 2; the pairing τ then relates 2 to 4; and π relates 4 to 3. Hence, all four labels can be reached from one another:

$$1 \overset{\pi}{\leftrightarrow} 2 \overset{\tau}{\leftrightarrow} 4 \overset{\pi}{\leftrightarrow} 3 \overset{\tau}{\leftrightarrow} 1.$$

They therefore form a single group,

$$\pi \vee \tau = \big\{\{1, 2, 3, 4\}\big\}, \qquad \#(\pi \vee \tau) = 1.$$

By contrast, if

$$\pi = \tau = (12)(34),$$

both pairings relate 1 only to 2 and 3 only to 4. There is no sequence of pairs connecting a label in $\{1, 2\}$ to a label in $\{3, 4\}$. Consequently, there are two separate groups:

$$\pi \vee \tau = \big\{\{1, 2\}, \{3, 4\}\big\}, \qquad \#(\pi \vee \tau) = 2.$$

The Gram matrix indexed by pairings is defined by

$$G_{\pi,\tau} := n^{\#(\pi\vee\tau)}. \tag{37}$$

When G is invertible, the orthogonal Weingarten matrix is

$$\mathrm{Wg}^{\mathrm{O}(n)} = G^{-1}. \tag{38}$$

More generally, including dimensions for which G is singular, the Weingarten matrix is defined by

$$\mathrm{Wg}^{\mathrm{O}(n)} = G^{+}, \tag{39}$$

where G^+ denotes the Moore–Penrose pseudoinverse (Penrose, 1955; Collins & Matsumoto, 2009).

Coset type and the notation $\mathcal{W}[\lambda]$. Let $B_1, \ldots, B_s$ be the groups forming $\pi \vee \tau$. Each group B_r contains an even number of labels, which we write as

$$|B_r| = 2\lambda_r.$$

Because the groups contain all $2k$ labels, the positive integers $\lambda_1, \ldots, \lambda_s$ satisfy

$$\lambda_1 + \cdots + \lambda_s = k.$$

The partition

$$\lambda = (\lambda_1, \ldots, \lambda_s)$$

is called the *coset type* of the pair (π, τ).

For example, consider again the pairings of four labels. If

$$\pi = \tau = (12)(34),$$

then

$$\pi \vee \tau = \big\{\{1,2\},\{3,4\}\big\}.$$

There are two groups, each of size $2 = 2 \times 1$. Hence,

$$\lambda = (1,1).$$

By contrast, if

$$\pi = (12)(34), \qquad \tau = (13)(24),$$

then

$$\pi \vee \tau = \big\{\{1,2,3,4\}\big\}.$$

There is a single group of size $4 = 2 \times 2$, and therefore

$$\lambda = (2).$$

The orthogonal Weingarten coefficient $\mathrm{Wg}^{\mathrm{O}(n)}(\pi, \tau)$ depends on π and τ only through their coset type. We consequently write

$$\mathcal{W}[\lambda] \equiv \mathrm{Wg}^{\mathrm{O}(n)}(\pi, \tau) \qquad \text{when } (\pi, \tau) \text{ has coset type } \lambda. \tag{40}$$

Thus, the two examples above correspond respectively to $\mathcal{W}[1,1]$ and $\mathcal{W}[2]$. The expression inside the square brackets labels a partition of k, rather than a particular pairing or permutation.

The three coset types appearing at sixth order can be understood in the same way. Fix

$$\pi = (12)(34)(56).$$

Then:

$$\tau = (12)(34)(56) \quad \Longrightarrow \quad \pi \vee \tau = \big\{\{1,2\},\{3,4\},\{5,6\}\big\}, \qquad \lambda = (1,1,1), \tag{41}$$

$$\tau = (13)(24)(56) \quad \Longrightarrow \quad \pi \vee \tau = \big\{\{1,2,3,4\},\{5,6\}\big\}, \qquad \lambda = (2,1), \tag{42}$$

$$\tau = (16)(23)(45) \quad \Longrightarrow \quad \pi \vee \tau = \big\{\{1,2,3,4,5,6\}\big\}, \qquad \lambda = (3). \tag{43}$$

These three cases give the coefficients $\mathcal{W}[1,1,1]$, $\mathcal{W}[2,1]$, and $\mathcal{W}[3]$, respectively.

Fourth-order Gram matrix. For $k = 2$, we use the ordered basis

$$\big((12)(34), (13)(24), (14)(23)\big)$$

of $\mathcal{P}_2(4)$. If the same pairing is chosen for π and τ, then $\pi \vee \tau$ has two blocks, each containing two labels. Hence, the corresponding Gram-matrix entry is $n^{\#(\pi \vee \tau)} = n^2$. If π and τ are distinct, then $\pi \vee \tau$ has a single block containing all four labels, and the corresponding entry is n. Therefore,

$$G = \begin{pmatrix} n^2 & n & n \\ n & n^2 & n \\ n & n & n^2 \end{pmatrix}. \tag{44}$$

The diagonal entries of G^{-1} correspond to the coset type $\lambda = (1,1)$, while its off-diagonal entries correspond to the coset type $\lambda = (2)$. Thus,

$$G^{-1} = \begin{pmatrix} \mathcal{W}[1,1] & \mathcal{W}[2] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[1,1] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[2] & \mathcal{W}[1,1] \end{pmatrix}. \tag{45}$$

Here, $\mathcal{W}[1,1]$ and $\mathcal{W}[2]$ denote the Weingarten coefficients associated with the coset types $\lambda = (1,1)$ and $\lambda = (2)$, respectively. Inverting the Gram matrix gives

$$\mathcal{W}[1,1] = \frac{n+1}{(n-1)n(n+2)}, \qquad \mathcal{W}[2] = -\frac{1}{(n-1)n(n+2)}. \tag{46}$$

Coefficients through sixth order. For completeness, the orthogonal Weingarten coefficients required in this work are

$$\mathcal{W}[1] = \frac{1}{n}, \tag{47}$$

$$\mathcal{W}[1,1] = \frac{n+1}{(n-1)n(n+2)}, \qquad \mathcal{W}[2] = -\frac{1}{(n-1)n(n+2)}, \tag{48}$$

$$\mathcal{W}[1,1,1] = \frac{n^2+3n-2}{(n-2)(n-1)n(n+2)(n+4)}, \qquad \mathcal{W}[2,1] = -\frac{1}{(n-2)(n-1)n(n+4)},$$
$$\mathcal{W}[3] = \frac{2}{(n-2)(n-1)n(n+2)(n+4)}. \tag{49}$$

As illustrated above, the possible sixth-order coset types are

$$\lambda = (1,1,1), \qquad \lambda = (2,1), \qquad \lambda = (3).$$

These correspond, respectively, to three blocks of size 2, one block of size 4 together with one block of size 2, and one block of size 6. The associated Weingarten coefficients are denoted by

$$\mathcal{W}[1,1,1], \qquad \mathcal{W}[2,1], \qquad \mathcal{W}[3].$$

Thus, each term in a sixth-order moment is assigned one of the three coefficients in equation 49, according to the coset type of its two pairings.

Relation to the special orthogonal group. The orthogonal group can be decomposed as

$$\mathrm{O}(n) = \{Q \in \mathrm{O}(n) : \det Q = +1\} \cup \{Q \in \mathrm{O}(n) : \det Q = -1\},$$

where the first subset is the special orthogonal group $\mathrm{SO}(n)$. For tensor moments of order $m < n$, Haar integration over $\mathrm{SO}(n)$ agrees with Haar integration over $\mathrm{O}(n)$. In this range, the same pairing formula and the same orthogonal Weingarten coefficients therefore apply to both groups.

At orders $m \geq n$, additional $\mathrm{SO}(n)$-invariant contractions involving the Levi–Civita symbol may contribute, so the moments over $\mathrm{SO}(n)$ and $\mathrm{O}(n)$ need not coincide. This distinction does not arise in the present work, where the weight matrices are sampled directly from $\mathrm{O}(n)$.

C Orthogonal NTK tensors

The statistics of the joint distribution of preactivations and the NTK in orthogonal neural networks are characterized by cumulants involving both quantities. At first order in $1/n$, these statistics are fully determined by the two cumulants shown in equation 11 and equation 60. In this appendix, we derive explicit recursion relations for the tensors D, F, A, and B at all orders in $1/n$, and then present their first-order $1/n$ approximations.

C.1 NTK-preactivation cross-correlator

The cumulant defined in equation 11 admits the following convenient representation

$$\begin{aligned}&\mathbb{E}_\theta^c[z_{i_1}^{(\ell+1)}(x_1) z_{i_2}^{(\ell+1)}(x_2) \widehat{\Delta\Theta}_{i_3 i_4}^{(\ell+1)}(x_3,x_4)] \\ &\quad = \mathbb{E}_\theta^c[z_{i_1}^{(\ell+1)}(x_1) z_{i_2}^{(\ell+1)}(x_2) \widehat{\Theta}_{i_3 i_4}^{(\ell+1)}(x_3,x_4)] - \mathbb{E}_\theta^c[z_{i_1}^{(\ell+1)}(x_1) z_{i_2}^{(\ell+1)}(x_2)] \mathbb{E}_\theta^c[\widehat{\Theta}_{i_3 i_4}^{(\ell+1)}(x_3,x_4)]\end{aligned} \tag{50}$$

Substituting the preactivation recursion equation 3 and the NTK recursion equation 9, one finds that

$$\begin{aligned}&\mathbb{E}_\theta^c[z_{i_1}^{(\ell+1)}(x_1) z_{i_2}^{(\ell+1)}(x_2) \widehat{\Delta\Theta}_{i_3 i_4}^{(\ell+1)}(x_3,x_4)] \\ &\quad = (12)_i (34)_i \left\{ \frac{\lambda_W^{(\ell+1)} C_W}{n^2} \sum_{j,k=1}^{n} \left(\mathbb{E}[\sigma_{j,1}^{(\ell)} \sigma_{j,2}^{(\ell)} \sigma_{k,3}^{(\ell)} \sigma_{k,4}^{(\ell)}] - \mathbb{E}[\sigma_{j,1}^{(\ell)} \sigma_{j,2}^{(\ell)}] \mathbb{E}[\sigma_{k,3}^{(\ell)} \sigma_{k,4}^{(\ell)}] \right)\right.\end{aligned}$$

$$+ (C_W)^2 \sum_{j,k=1}^{n} \left(\mathbb{E}\left[\left(\mathcal{W}[1,1]\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)} - \frac{1}{n^2}\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}] \right) \sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)} \right] + 2\mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}^{(\ell)}] \right) \Bigg\}$$

$$+ (13)(24)_i \Bigg\{ (C_W)^2 \sum_{j,k=1}^{n} \Big(\mathcal{W}[1,1]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}]$$

$$+ \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}] + \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\widehat{\Theta}_{jk,34}] \Big) \Bigg\}$$

$$+ (14)(23)_i \Bigg\{ (C_W)^2 \sum_{j,k=1}^{n} \Big(\mathcal{W}[1,1]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\widehat{\Theta}_{jk,34}]$$

$$+ \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}] + \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}] \Big) \Bigg\} \tag{51}$$

Here $\mathcal{W}[1,1]$ and $\mathcal{W}[2]$ denote the $k=2$ Weingarten functions appearing in equation 1, and we have used the index symmetry of the NTK sample labels. A direct comparison of equation 51 and equation 11 then yields

$$\frac{1}{n}D_{1234}^{(\ell+1)} = \frac{\lambda_W^{(\ell+1)}C_W}{n^2} \sum_{j,k=1}^{n} \left(\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{(\ell)}\sigma_{k,4}^{(\ell)}] - \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}]\mathbb{E}[\sigma_{k,3}^{(\ell)}\sigma_{k,4}^{(\ell)}] \right)$$

$$+ (C_W)^2 \sum_{j,k=1}^{n} \left(\mathbb{E}\left[\left(\mathcal{W}[1,1]\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)} - \frac{1}{n^2}\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}] \right) \sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)} \right] \right.$$

$$\left. + 2\mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}^{(\ell)}] \right) \tag{52}$$

$$\frac{1}{n}F_{1324}^{(\ell+1)} = (C_W)^2 \sum_{j,k=1}^{n} \Big(\mathcal{W}[1,1]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}^{(\ell)}]$$

$$+ \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)}] + \mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\widehat{\Theta}_{jk,34}^{(\ell)}] \Big) \tag{53}$$

To obtain the first-order $1/n$ corrections to these tensors, we expand the $k=2$ Weingarten functions in the large-n regime. Explicitly,

$$\mathcal{W}[1,1] = \frac{n+1}{n(n-1)(n+2)} = \frac{1}{n^2} + \frac{2}{n^4} - \frac{2}{n^5} + \mathcal{O}\left(\frac{1}{n^6}\right) \tag{54}$$

$$\mathcal{W}[2] = -\frac{1}{n(n-1)(n+2)} = -\frac{1}{n^3} + \frac{1}{n^4} - \frac{3}{n^5} + \mathcal{O}\left(\frac{1}{n^6}\right) \tag{55}$$

Upon substituting equation 54 and equation 55 into equation 52 and equation 53, we obtain

$$\frac{1}{n}D_{1234}^{(\ell+1)} = \frac{\lambda_W^{(\ell+1)}C_W}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\widehat{\Delta K}_{j,12}^{(\ell)}\widehat{\Delta K}_{k,34}^{(\ell)}] + \frac{(C_W)^2}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\widehat{\Delta K}_{j,12}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}]\Theta_{34}^{(\ell)}$$

$$+ \frac{(C_W)^2}{n^2} \sum_{j,k=1}^{n} \mathbb{E}[\widehat{\Delta K}_{j,12}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Delta\Theta}_{kk,34}^{(\ell)}] - 2\frac{(C_W)^2}{n^3} \sum_{j=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}]\Theta_{34}^{(\ell)}$$

$$- 2\frac{(C_W)^2}{n^3} \sum_{j,k=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Delta\Theta}_{jk,34}^{(\ell)}] + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{56}$$

$$\frac{1}{n}F_{1324}^{(\ell+1)} = \frac{(C_W)^2}{n^2}\left[\sum_{j=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}]\Theta_{34}^{(\ell)} + \sum_{j,k=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Delta\Theta}_{jk,34}^{(\ell)}] \right]$$

$$- \frac{(C_W)^2}{n^3}\left[\sum_{j,k=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}]\Theta_{34}^{(\ell)} + \sum_{j,k=1}^{n} \mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Delta\Theta}_{jk,34}^{(\ell)}] \right]$$

$$- \frac{(C_W)^2}{n^3}\left[\sum_{j=1}^{n}\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}\sigma_{j,2}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}]\Theta_{34}^{(\ell)} + \sum_{j,k=1}^{n}\mathbb{E}[\sigma_{j,1}^{(\ell)}\sigma_{j,4}^{\prime(\ell)}\sigma_{k,2}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\widehat{\Delta\Theta}_{jk,34}^{(\ell)}]\right] + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{57}$$

Here we introduce the tensor $\widehat{\Delta K}_{i,\alpha\beta}^{(\ell)} = \sigma_{i,\alpha}^{(\ell)}\sigma_{i,\beta}^{(\ell)} - \mathbb{E}[\sigma_{i,\alpha}^{(\ell)}\sigma_{i,\beta}^{(\ell)}]$, and decompose the NTK tensor into its mean Θ and fluctuations $\widehat{\Delta\Theta}$. One can then directly apply the effective field theory techniques developed in Roberts & Yaida (2022) to expand non-Gaussian expectation values with an interaction controlled by the four-point vertex V_4. After lengthy and technically involved algebraic computations, one arrives at

$$\begin{aligned}
D_{1234}^{(\ell+1)} &= C_W\langle\widehat{\Delta G}_{12}^{(\ell)}\widehat{\Delta\Omega}_{34}^{(\ell+1)}\rangle_{K^{(\ell)}} \\
&\quad + \frac{C_W}{4}\sum_{\beta_1,\beta_2,\beta_3,\beta_4\in\{1,2,3,4\}} V_{(\beta_1\beta_2)(\beta_3\beta_4)}^{(\ell)}\left\langle\frac{d^2(\widehat{\Delta G}_{12}^{(\ell)})}{dz_{\beta_1}^{(\ell)}dz_{\beta_2}^{(\ell)}}\right\rangle_{K^{(\ell)}}\left\langle\frac{d^2(\widehat{\Delta\Omega}_{34}^{(\ell+1)})}{dz_{\beta_1}^{(\ell)}dz_{\beta_2}^{(\ell)}}\right\rangle_{K^{(\ell)}} \\
&\quad + (C_W)^2\sum_{\beta_1,\beta_2\in\{1,2,3,4\}}\left\langle\frac{d^2(\widehat{\Delta G}_{12}^{(\ell)})}{dz_{\beta_1}^{(\ell)}dz_{\beta_2}^{(\ell)}}\right\rangle_{K^{(\ell)}}\langle\sigma_3^{\prime(\ell)}\sigma_4^{\prime(\ell)}\rangle_{K^{(\ell)}}D_{\beta_1\beta_2 34}^{(\ell)} + \mathcal{O}\left(\frac{1}{n}\right),
\end{aligned} \tag{58}$$

$$\begin{aligned}
F_{1324}^{(\ell+1)} &= (C_W)^2\left[\langle\sigma_1^{(\ell)}\sigma_2^{(\ell)}\sigma_3^{\prime(\ell)}\sigma_4^{\prime(\ell)}\rangle_{K^{(\ell)}} - \langle\sigma_1^{(\ell)}\sigma_2^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_3^{\prime(\ell)}\sigma_4^{\prime(\ell)}\rangle_{K^{(\ell)}}\right]\Theta_{34}^{(\ell)} \\
&\quad + (C_W)^2\sum_{\alpha,\beta,\gamma,\delta=1}^{4}\left\langle\frac{d(\sigma_1^{(\ell)}\sigma_3^{\prime(\ell)})}{dz_\gamma^{(\ell)}}\right\rangle_{K^{(\ell)}}\left\langle\frac{d(\sigma_2^{(\ell)}\sigma_4^{\prime(\ell)})}{dz_\delta^{(\ell)}}\right\rangle_{K^{(\ell)}}F_{\gamma 3\delta 4}^{(\ell)} + \mathcal{O}\left(\frac{1}{n}\right),
\end{aligned} \tag{59}$$

where $\widehat{\Delta G}_{\alpha\beta}^{(\ell)} = \sigma_{i,\alpha}^{(\ell)}\sigma_{i,\beta}^{(\ell)} - \langle\sigma_{i,\alpha}^{(\ell)}\sigma_{i,\beta}^{(\ell)}\rangle_{K^{(\ell)}}$, and $\widehat{\Omega}_{i,\alpha\beta}^{(\ell+1)} = \lambda_W^{(\ell+1)}\sigma_{i,\alpha}^{(\ell)}\sigma_{i,\beta}^{(\ell)} + C_W\Theta_{\alpha\beta}^{(\ell)}\sigma_{i,\alpha}^{\prime(\ell)}\sigma_{i,\beta}^{\prime(\ell)}$ with $\widehat{\Delta\Omega}_{i,\alpha\beta}^{(\ell+1)} = \widehat{\Omega}_{i,\alpha\beta}^{(\ell+1)} - \langle\widehat{\Omega}_{i,\alpha\beta}^{(\ell+1)}\rangle_{K^{(\ell)}}$.

We observe that the recursion relation for the D tensor coincides with its Gaussian counterpart, while the recursion relation for the F tensor acquires an additional term proportional to the NTK mean. The expression in equation 12 is obtained from equation 59 by performing integrations-by-part inside the Gaussian expectation values of derivative terms.

C.2 NTK VARIANCE

The NTK variance admits the compact representation

$$\begin{aligned}
&\mathbb{E}_\theta^c[\widehat{\Delta\Theta}_{i_1i_2}^{(\ell+1)}(x_1,x_2)\widehat{\Delta\Theta}_{i_3i_4}^{(\ell+1)}(x_3,x_4)] \\
&\quad = \frac{1}{n}\Big(A_{1234}^{(\ell+1)}(12)(34)_i + B_{1324}^{(\ell+1)}(13)(24)_i + B_{1423}^{(\ell+1)}(14)(23)_i\Big),
\end{aligned} \tag{60}$$

where the tensors A and B are fully determined by $\hat{K}$ and $\hat{\Theta}$. To see this, we rewrite the cumulant equation 60 in the equivalent form

$$\begin{aligned}
&\mathbb{E}_\theta^c[\widehat{\Delta\Theta}_{i_1i_2}^{(\ell+1)}(x_1,x_2)\widehat{\Delta\Theta}_{i_3i_4}^{(\ell+1)}(x_3,x_4)] \\
&\quad = \mathbb{E}_\theta^c[\widehat{\Theta}_{i_1i_2}^{(\ell+1)}(x_1,x_2)\widehat{\Theta}_{i_3i_4}^{(\ell+1)}(x_3,x_4)] - \mathbb{E}_\theta^c[\widehat{\Theta}_{i_1i_2}^{(\ell+1)}(x_1,x_2)]\mathbb{E}_\theta^c[\widehat{\Theta}_{i_3i_4}^{(\ell+1)}(x_3,x_4)]
\end{aligned} \tag{61}$$

Substituting the recursion of the NTK tensor from equation 9, we obtain that

$$\begin{aligned}
&\mathbb{E}_\theta^c[\widehat{\Delta\Theta}_{i_1i_2}^{(\ell+1)}(x_1,x_2)\widehat{\Delta\Theta}_{i_3i_4}^{(\ell+1)}(x_3,x_4)] \\
&\quad = (12)(34)_i\Bigg\{\frac{(\lambda_W^{(\ell+1)})^2}{n^2}\sum_{j,k}\mathbb{E}[\widehat{\Delta K}_{j,12}^{(\ell)}\widehat{\Delta K}_{j,34}^{(\ell)}] + \frac{\lambda_W^{(\ell+1)}C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta K}_{j,12}^{(\ell)}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)}] \\
&+ \frac{\lambda_W^{(\ell+1)}C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta K}_{j,34}^{(\ell)}\sigma_{k,1}^{\prime(\ell)}\sigma_{k,2}^{\prime(\ell)}\widehat{\Theta}_{kk,12}^{(\ell)}] + (C_W)^2\Bigg[\sum_{j,k=1}^{n}\mathcal{W}[1,1]\mathbb{E}[\sigma_{j,1}^{\prime(\ell)}\sigma_{j,2}^{\prime(\ell)}\widehat{\Theta}_{jj,12}\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)}] \\
&- \frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\sigma_{j,1}^{\prime(\ell)}\sigma_{j,2}^{\prime(\ell)}\widehat{\Theta}_{jj,12}^{(\ell)}]\mathbb{E}[\sigma_{k,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{kk,34}^{(\ell)}] + \sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma_{j,1}^{\prime(\ell)}\sigma_{k,2}^{\prime(\ell)}\widehat{\Theta}_{jk,12}^{(\ell)}\sigma_{j,3}^{\prime(\ell)}\sigma_{k,4}^{\prime(\ell)}\widehat{\Theta}_{jk,34}^{(\ell)}]
\end{aligned}$$

$$+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Theta}^{(\ell)}_{kj,34}]\Big]\Big\}$$

$$+(13)(24)_i(C_W)^2\Big\{\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]$$

$$+\sum_{j,k=1}^{n}\mathcal{W}[1,1]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{jk,34}]+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Theta}^{(\ell)}_{kj,34}]\Big\}$$

$$+(14)(23)_i(C_W)^2\Big\{\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]$$

$$+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{jk,34}]+\sum_{j,k=1}^{n}\mathcal{W}[1,1]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Theta}^{(\ell)}_{kj,34}]\Big\} \tag{62}$$

A direct comparison with equation 60 then yields

$$\begin{aligned}\frac{1}{n}A^{(\ell+1)}_{1234}=&\frac{(\lambda_W^{(\ell+1)})^2}{n^2}\sum_{j,k}\mathbb{E}[\widehat{\Delta K}^{(\ell)}_{j,12}\widehat{\Delta K}^{(\ell)}_{j,34}]+\frac{\lambda_W^{(\ell+1)}C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta K}^{(\ell)}_{j,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]\\&+\frac{\lambda_W^{(\ell+1)}C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta K}^{(\ell)}_{j,34}\sigma'^{(\ell)}_{k,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{kk,12}]+(C_W)^2\Big[\sum_{j,k=1}^{n}\mathcal{W}[1,1]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]\\&-\frac{1}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Theta}^{(\ell)}_{jj,12}]\mathbb{E}[\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{jk,34}]\\&+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Theta}^{(\ell)}_{kj,34}]\Big]\end{aligned} \tag{63}$$

$$\begin{aligned}\frac{1}{n}B^{(\ell+1)}_{1324}=&(C_W)^2\Big\{\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{kk,34}]\\&+\sum_{j,k=1}^{n}\mathcal{W}[1,1]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Theta}^{(\ell)}_{jk,34}]+\sum_{j,k=1}^{n}\mathcal{W}[2]\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Theta}^{(\ell)}_{kj,34}]\Big\}\end{aligned} \tag{64}$$

The use of the $1/n$-expansions equation 54, equation 55 allows the following simplifications

$$\begin{aligned}\frac{1}{n}A^{(\ell+1)}_{1234}=&\frac{1}{n^2}\sum_{j,k}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,12}\widehat{\Delta\Omega}^{(\ell+1)}_{j,34}]+\frac{C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]\\&+\frac{C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,34}\sigma'^{(\ell)}_{k,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{kk,12}]+\frac{(C_W)^2}{n^2}\Big[\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]\\&-\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}]\mathbb{E}[\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]-\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34}\\&-\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}\widehat{\Delta\Theta}^{(\ell)}_{jj,34}]\Theta^{(\ell)}_{12}-\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{34}\\&-\frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{jk,34}]-\frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Delta\Theta}^{(\ell)}_{kj,34}]\Big]\\&+\mathcal{O}\left(\frac{1}{n^2}\right)\end{aligned} \tag{65}$$

$$\frac{1}{n}B^{(\ell+1)}_{1324} = \frac{(C_W)^2}{n^2}\Bigg[-\frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34} - \frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]\Theta^{(\ell)}_{12}$$
$$-\frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}]\Theta^{(\ell)}_{34} - \frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]$$
$$+\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34} + \sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{34}$$
$$+\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}\widehat{\Delta\Theta}^{(\ell)}_{jj,34}]\Theta^{(\ell)}_{12} + \sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{jk,34}]$$
$$-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34} - \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{34}$$
$$-\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}\widehat{\Delta\Theta}^{(\ell)}_{jj,34}]\Theta^{(\ell)}_{12} - \frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{j,4}\widehat{\Delta\Theta}^{(\ell)}_{kj,34}]\Bigg]$$
$$+\mathcal{O}\left(\frac{1}{n^2}\right) \tag{66}$$

Here, we again decompose the NTK tensor into its mean Θ and fluctuations $\widehat{\Delta\Theta}$. These expressions can be further simplified by noting that not all contributions arising from the off-diagonal entries of the Weingarten matrix appear at order $1/n$. In particular, the terms proportional to $1/n$ inside the square brackets in equation 65 scale at most linearly with n, as follows from equation 11 and equation 60, yielding an overall contribution of order $\mathcal{O}\left(\frac{1}{n^2}\right)$. This observation leads to the formula

$$\frac{1}{n}A^{(\ell+1)}_{1234} = \frac{1}{n^2}\sum_{j,k}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,12}\widehat{\Delta\Omega}^{(\ell+1)}_{j,34}] + \frac{C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]$$
$$+\frac{C_W}{n^2}\sum_{j,k=1}^{n}\mathbb{E}[\widehat{\Delta\Omega}^{(\ell+1)}_{j,34}\sigma'^{(\ell)}_{k,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{kk,12}] + \frac{(C_W)^2}{n^2}\Bigg[\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]$$
$$-\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\widehat{\Delta\Theta}^{(\ell)}_{jj,12}]\mathbb{E}[\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{kk,34}]\Bigg] + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{67}$$

Similarly, the only term that can yield a non-zero contribution at order $\frac{1}{n}$ in equation 66 is the first term, $\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34}$, through its off-diagonal components. Likewise, the terms of order n^0 that vanish at order $\frac{1}{n}$ are precisely those containing a single NTK fluctuation inside the expectation value, in agreement with equation 11. We therefore obtain

$$\frac{1}{n}B^{(\ell+1)}_{1324} = \frac{(C_W)^2}{n^2}\Bigg[-\frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{k,3}\sigma'^{(\ell)}_{k,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34} + \sum_{j=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{j,2}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{j,4}]\Theta^{(\ell)}_{12}\Theta^{(\ell)}_{34}$$
$$+\sum_{j,k=1}^{n}\mathbb{E}[\sigma'^{(\ell)}_{j,1}\sigma'^{(\ell)}_{k,2}\widehat{\Delta\Theta}^{(\ell)}_{jk,12}\sigma'^{(\ell)}_{j,3}\sigma'^{(\ell)}_{k,4}\widehat{\Delta\Theta}^{(\ell)}_{jk,34}]\Bigg] + \mathcal{O}\left(\frac{1}{n^2}\right) \tag{68}$$

One can apply the effective field theory techniques of Roberts & Yaida (2022) to compute the first-order $1/n$ corrections to the expectation values appearing in equation 67 and equation 68. Using equation 11 and equation 60, together with extensive path-integral manipulations and algebraic identities, one finds that

$$A^{(\ell+1)}_{1234} = \langle\widehat{\Delta\Omega}^{(\ell+1)}_{12}\widehat{\Delta\Omega}^{(\ell+1)}_{34}\rangle_{K^{(\ell)}}$$
$$+\frac{1}{4}\sum_{\beta_1,\beta_2,\beta_3,\beta_4\in\{1,2,3,4\}}V^{(\ell)}_{(\beta_1\beta_2)(\beta_3\beta_4)}\left\langle\frac{d^2(\widehat{\Delta\Omega}^{(\ell+1)}_{12})}{dz^{(\ell)}_{\beta_1}dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle\frac{d^2(\widehat{\Delta\Omega}^{(\ell+1)}_{34})}{dz^{(\ell)}_{\beta_1}dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}$$

$$+\frac{C_W}{2}\sum_{\beta_1,\beta_2\in\{1,2,3,4\}}\left\langle\frac{d^2(\widehat{\Delta\Omega}_{12}^{(\ell+1)})}{dz_{\beta_1}^{(\ell)}dz_{\beta_2}^{(\ell)}}\right\rangle_{K^{(\ell)}}\langle\sigma_3'^{(\ell)}\sigma_4'^{(\ell)}\rangle_{K^{(\ell)}}D_{\beta_1\beta_2 34}^{(\ell)}$$

$$+\frac{C_W}{2}\sum_{\beta_3,\beta_4\in\{1,2,3,4\}}\left\langle\frac{d^2(\widehat{\Delta\Omega}_{34}^{(\ell+1)})}{dz_{\beta_3}^{(\ell)}dz_{\beta_4}^{(\ell)}}\right\rangle_{K^{(\ell)}}\langle\sigma_1'^{(\ell)}\sigma_2'^{(\ell)}\rangle_{K^{(\ell)}}D_{12\beta_3\beta_4}^{(\ell)}$$

$$+(C_W)^2\langle\sigma_1'^{(\ell)}\sigma_2'^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_3'^{(\ell)}\sigma_4'^{(\ell)}\rangle_{K^{(\ell)}}A_{1234}^{(\ell)}+\mathcal{O}\left(\frac{1}{n}\right), \tag{69}$$

$$B_{1324}^{(\ell+1)} = (C_W)^2\left[\langle\sigma_1'^{(\ell)}\sigma_2'^{(\ell)}\sigma_3'^{(\ell)}\sigma_4'^{(\ell)}\rangle_{K^{(\ell)}}-\langle\sigma_1'^{(\ell)}\sigma_2'^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_3'^{(\ell)}\sigma_4'^{(\ell)}\rangle_{K^{(\ell)}}\right]\Theta_{12}^{(\ell)}\Theta_{34}^{(\ell)}$$

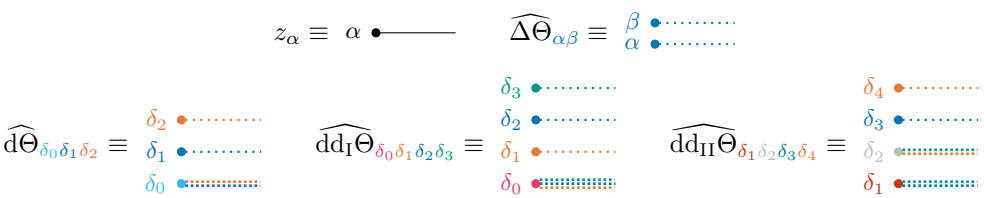
$$+(C_W)^2\langle\sigma_1'^{(\ell)}\sigma_3'^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_2'^{(\ell)}\sigma_4'^{(\ell)}\rangle_{K^{(\ell)}}B_{1324}^{(\ell)}+\mathcal{O}\left(\frac{1}{n}\right). \tag{70}$$

Here we observe that the recursion relation for the tensor A coincides with its Gaussian counterpart, while that for the tensor B acquires an additional term proportional to the NTK mean.

D Generalized Feynman rules

Without loss of generality, consider the channel $(12)_i(34)_i\dots(2k\,2k-1)_i$, corresponding to the pairing $(12)(34)\dots(2k\,2k-1)$. The Feynman rules reproducing tensors with this channel structure can be organized into two groups. The first group implements the orthogonality constraints as follows:

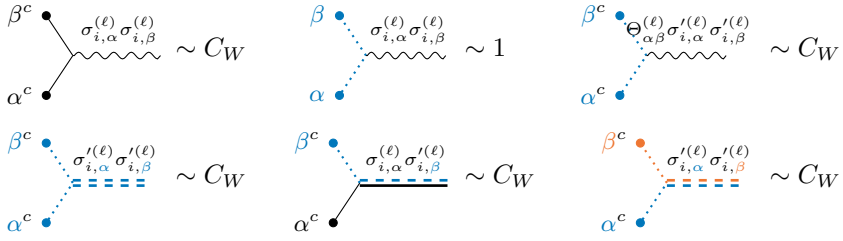
1. Preactivations, NTKs, dNTKs and ddNTKs are represented by external lines, as illustrated below.

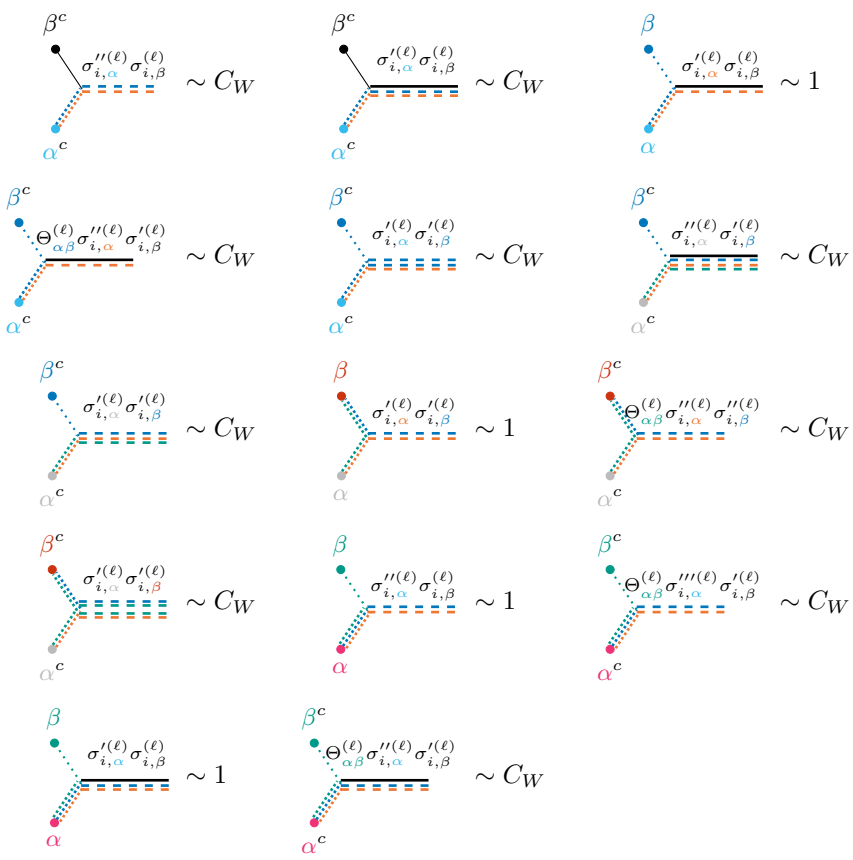


(71)

In the first line of equation 71, a colored line represents a single NTK label. Distinct colors are used for external dotted lines associated with different NTKs. In the second line, we introduce the diagrammatic representations of the dNTK: $\widehat{\mathrm{d}\Theta}_{\delta_0\delta_1\delta_2} = \sum_{\mu\nu}\frac{d^2z_{\delta_0}}{d\theta_\mu d\theta_\nu}\frac{dz_{\delta_1}}{d\theta_\mu}\frac{dz_{\delta_2}}{d\theta_\nu}$, the $\mathrm{dd_I}$NTK: $\widehat{\mathrm{dd_I}\Theta}_{\delta_0\delta_1\delta_2\delta_3} = \sum_{\mu\nu\rho}\frac{d^3z_{\delta_0}}{d\theta_\mu d\theta_\nu d\theta_\rho}\frac{dz_{\delta_1}}{d\theta_\mu}\frac{dz_{\delta_2}}{d\theta_\nu}\frac{dz_{\delta_3}}{d\theta_\rho}$, and the $\mathrm{dd_{II}}$NTK: $\widehat{\mathrm{dd_{II}}\Theta}_{\delta_1\delta_2\delta_3\delta_4} = \sum_{\mu\nu\rho}\frac{d^2z_{\delta_1}}{d\theta_\mu d\theta_\nu}\frac{d^2z_{\delta_2}}{d\theta_\rho d\theta_\nu}\frac{dz_{\delta_3}}{d\theta_\mu}\frac{dz_{\delta_4}}{d\theta_\rho}$. The colors of the dotted lines correspond to the θ-indices appearing in these definitions. Since the θ-indices are contracted in pairs, distinct colors encode the corresponding contraction pattern. In the diagrammatic notation, triple lines represent third derivatives, double lines represent second derivatives, and single lines represent first derivatives.

2. Define the cubic vertices as

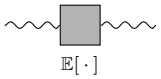




(72)

Here the superscript c, referred to as the orthogonality charge, keeps track of the label's orthogonal character, and lines that do not end in a dot represent internal lines.

3. Draw a square propagator connecting internal lines in all possible ways, consistent with the chosen pairing. The square represents the full expectation value.



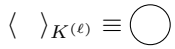
(73)

This procedure generates distinct diagram types, both connected and disconnected. The connected diagrams are further classified as s-class diagrams, defined by the number s of square propagators appearing in the diagram. The square propagator obeys an additional selection rule: all k-class diagrams formed from pairs of external lines corresponding to different object types vanish.

4. For each s-class diagram, generate all inequivalent permutations of its $2m$ external labels carrying orthogonality charge. Multiply each resulting diagram by $1/n$ for every uncharged pairing, and by the appropriate m-class Weingarten function $\mathcal{W}$, determined by the relative ordering τ of the diagram's labels with respect to the original pairing $\pi = (12)(34)\dots(2k\,2k-1)$: $\mathcal{W}[\tau,\pi] = \mathcal{W}[\ell(\tau\circ\pi)]$, where $\circ$ denotes ordinary permutation multiplication and $\ell(e)$ denotes the cycle length of e. See Appendix B for a detailed discussion of Weingarten functions and their properties.
5. Multiply each s-class contribution by the Möbius coefficient $(-1)^{s-1}(s-1)!$, and sum over all classes.

The second group implements the effective field theory techniques developed in Roberts & Yaida (2022), applied to the square propagator in the diagrammatic construction of the previous step, through the following set of Feynman rules analogous to those introduced in Guillen et al. (2025):

6. We define the bare propagator as



(74)

where $\langle \quad \rangle_{K^{(\ell)}}$ denotes a zero-mean Gaussian expectation with covariance specified by $K^{(\ell)}$. The expectation value is taken over the decorations of the internal lines attached to the propagator, which obeys the same selection rules described in Guillen et al. (2025). These rules are summarized as follows:

(a) Propagators may only connect to internal lines emanating from cubic vertices or from the internal quartic vertices introduced below. In particular, propagators cannot be directly connected to other propagators.

(b) Dotted lines attached to a propagator do not enter the Gaussian expectation value, as they carry no decorations.

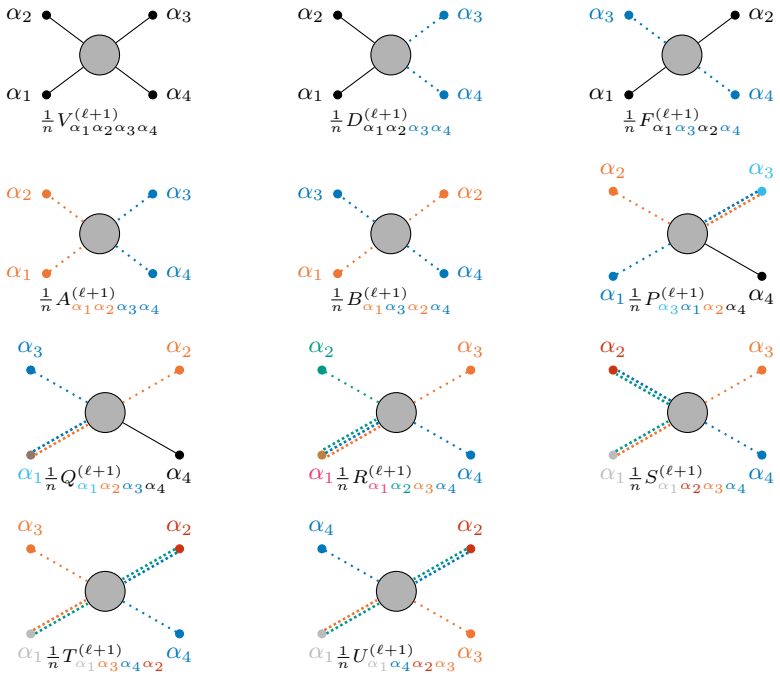
(c) Each preactivation line decorated with z_i acts as a derivative with respect to z_i acting on the argument of the Gaussian expectation value.

(d) The neural indices of all internal lines connected to a propagator must be identical.

(e) If both dotted and dashed lines of the same color are attached to the propagator, they must appear in pairs carrying the same sample index. The two lines in each pair attach to different vertices. Moreover, if both vertices are drawn in the orientation specified in the Feynman rules, the top-to-bottom ordering of the sample indices (and therefore the colors) of the lines connected to the two vertices must coincide.

(f) A pair of dashed lines of the same color connected to the propagator contributes a factor $\Theta_{\alpha\beta}$ when the two lines attach to different vertices, where α and β denote the sample indices of the pair.

7. Quartic vertices are defined analogously, following Guillen et al. (2025). Explicitly,



(75)

8. Higher-order NTK and preactivation tensors are introduced via a natural generalization of the vertices in equation 75.

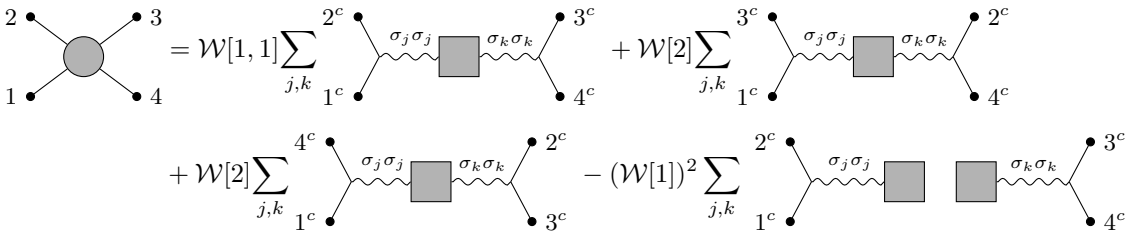
9. The square propagator decomposes into all connected and disconnected diagrams built from the bare propagator, quartic vertices, and higher-order vertices, with internal lines remaining undotted. This decomposition respects the selection rules (a)-(f).

E FEYNMAN RULES IN ACTION

In this appendix, we explicitly apply the Feynman rules (1)-(9) of Section 4 to the tensors V, F, and B.

E.1 THE QUARTIC VERTEX: V_4

The Feynman rules (1)-(5) generate the following diagrams for the quartic vertex V_4:

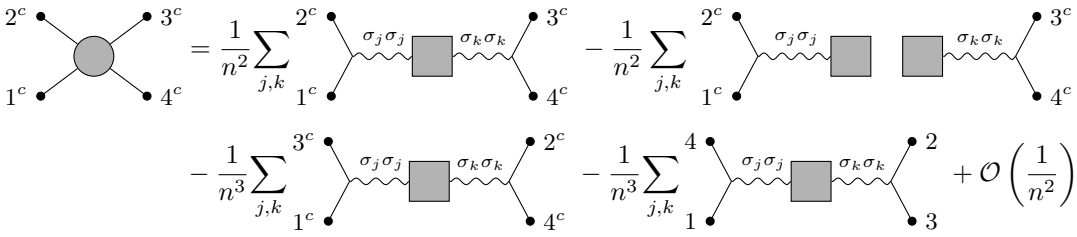


(76)

The first three terms in equation 76 are 1-class diagrams, as they are rendered fully connected by the square propagator. They arise from the three inequivalent pairings obtained by permuting the reference pairing $(12)(34)$, namely $\{(12)(34), (13)(24), (14)(23)\}$. The $k = 2$ Weingarten functions are selected according to the relative ordering between the pairing π appearing in the diagram and the reference pairing τ: $\mathcal{W}[\pi, \tau] = \mathcal{W}[1, 1]$ for $\pi = \tau$, $\mathcal{W}[\pi, \tau] = \mathcal{W}[2]$ for $\pi \neq \tau$. The coefficient multiplying each contribution is fixed by the Möbius formula evaluated on the single-block partition $s = 1$.

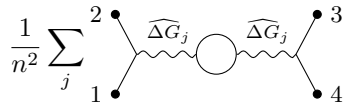
By contrast, the last diagram in equation 76 belongs to the 2-class, since the two cubic vertices are disconnected. Each subdiagram admits a unique pairing, yielding the $k = 1$ Weingarten function $W[1]$. The negative coefficient associated with this contribution follows from the Möbius formula applied to the two-block partition $s = 2$.

We now apply the second set of Feynman rules, (6)–(9), to the square propagators in equation 76. Since we are interested in the first-order $\frac{1}{n}$ correction to the quartic vertex, we approximate the Weingarten functions by their leading terms in the $\frac{1}{n}$ expansions equation 54 and equation 55. Consequently, equation 76 simplifies to



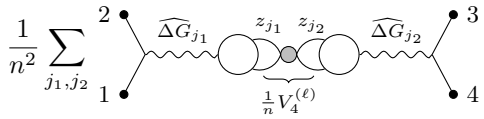
(77)

To evaluate equation 77, we distinguish two cases. When $i = j$, the only nontrivial contribution arises from the first line of equation 77, which reads



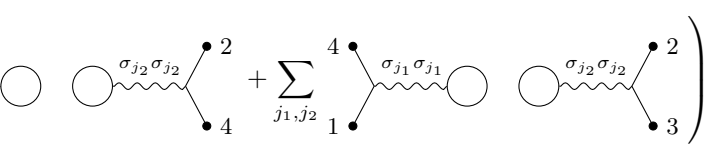
(78)

When $i \neq j$, the first line of equation 77 yields the following diagram:



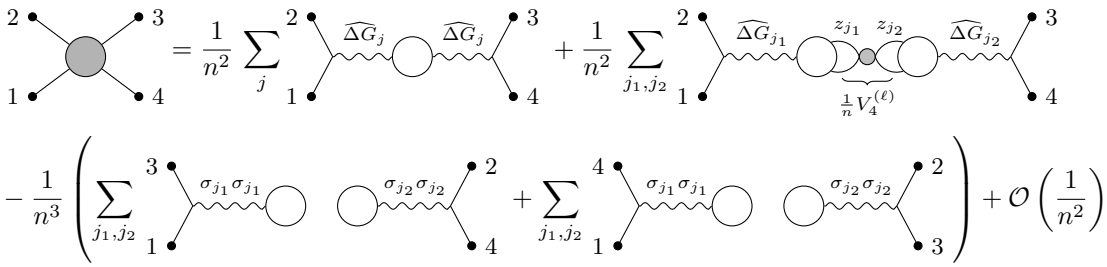
(79)

whereas the second line produces the following diagram:



(80)

Substituting equation 78, equation 79, and equation 80 into equation 77, we obtain



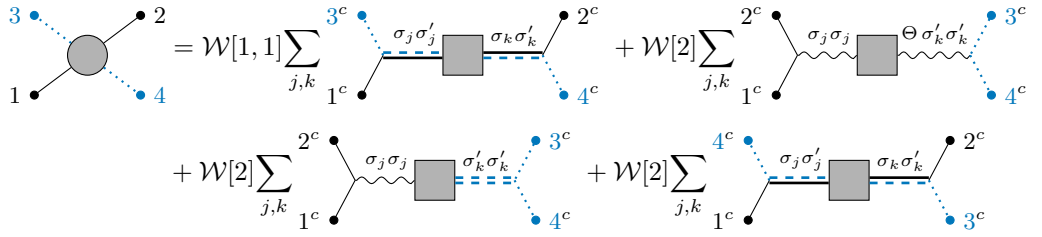
(81)

The $1/n$ scaling of the diagrams on the right-hand side can be verified by straightforward power counting: the summation in the first diagram yields a factor of n, while the summations in the remaining diagrams yield factors of n^2, which combine with the explicit prefactors to produce an overall $1/n$ contribution. Consequently, the expression equation 81 reproduces the recursion relation for the quartic vertex shown in equation 7, previously derived in Banta et al. (2024) through direct algebraic manipulations.

E.2 Joint NTK-preactivation cumulant: F

We already applied the Feynman rules of Section 4 to F in the main text; for completeness, we repeat the derivation here.

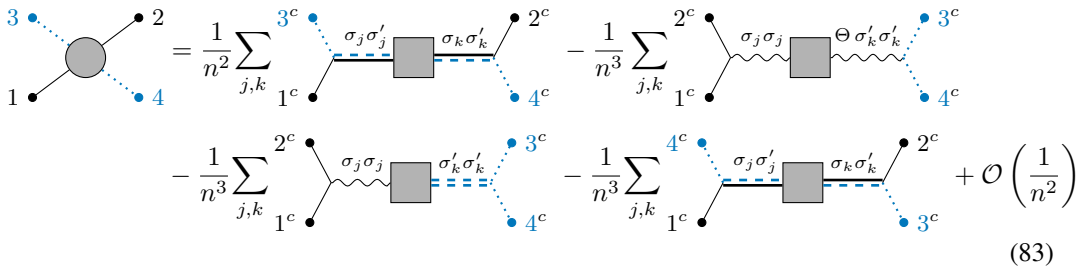
The Feynman rules (1)-(5) produce the following diagrams for the tensor F:



(82)

The absence of 2-class diagrams in equation 82 follows from the fact that the square propagator vanishes when acting on a pair of lines of different type. The 1-class diagrams are generated from the inequivalent permutations of the reference pairing $(13)(24)$, namely $\{(13)(24), (12)(34), (14)(23)\}$. The corresponding $k = 2$ Weingarten functions multiplying the subdiagrams are determined by the relative ordering of the reference pairing τ and the pairing π appearing in the subdiagram: $\mathcal{W}[\pi, \tau] = \mathcal{W}[1, 1]$ for $\pi = \tau$, $\mathcal{W}[\pi, \tau] = \mathcal{W}[2]$ for $\pi \neq \tau$. The overall factor of 1 follows from the Möbius relation for a single-block partition.

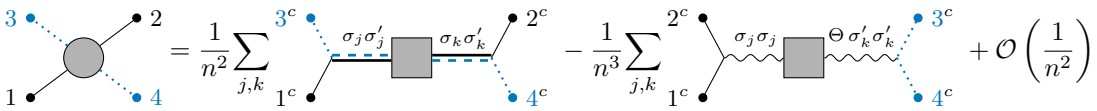
At first order in $1/n$, the Weingarten functions can be approximated by their leading terms in the expansions equation 54 and equation 55. In this manner, equation 82 simplifies to



(83)

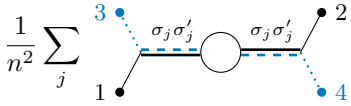
We now apply the second set of Feynman rules (6)–(9). As an immediate consequence, the last terms in equation 83 vanish at order $1/n$. A nonzero contribution would require the corresponding diagrams to produce a factor of n^2, which occurs when $j \neq k$ and a square propagator decomposes into two bare propagators forming disconnected subdiagrams. However, such configurations are forbidden by

the selection rules (a)–(f), as each subdiagram violates color conservation. Consequently, we are left with



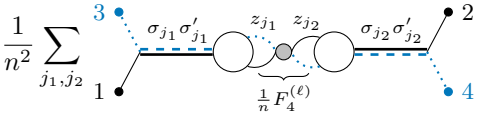
(84)

We now expand the diagrams in equation 84 in terms of the bare propagator and quartic vertices by analyzing the diagonal and off-diagonal neural components of each subdiagram, together with the selection rules (a)-(f). When $i = j$, the first diagram in equation 84 contributes as



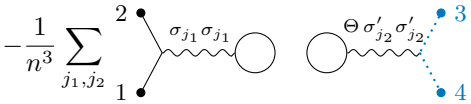
(85)

When $i \neq j$, the first diagram reduces to



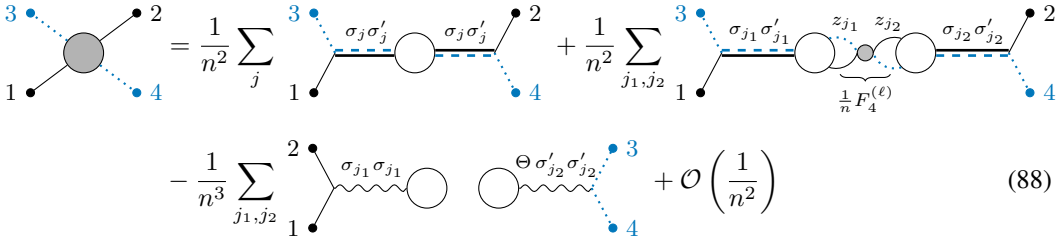
(86)

whereas the second diagram becomes



(87)

Substituting equation 85, equation 86, and equation 87 into equation 84, we obtain

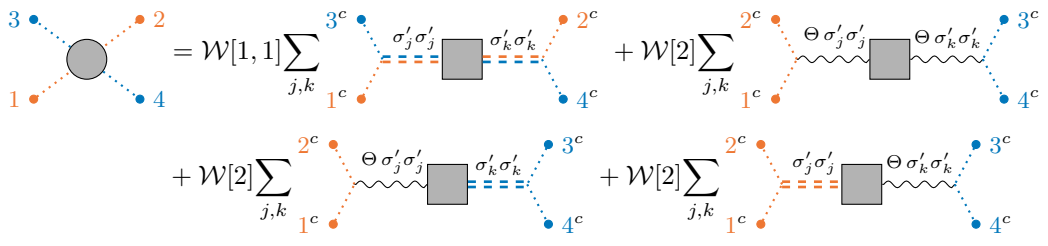


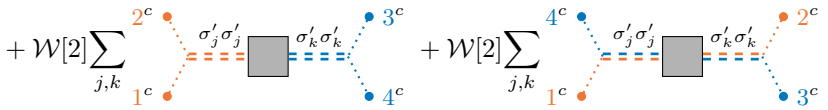
(88)

The $1/n$ scaling of the diagrams follows from straightforward power counting of the respective summations: When $i = j$, the summation contributes factor of n, whereas when $i \neq j$ it yields a factor of n^2. Combined with the explicit coefficients on the right-hand side of equation 88, these factors produce an overall contribution of order $1/n^2$. It is straightforward to verify that the first and second lines on the RHS of equation 88 reproduce the corresponding lines on the RHS of equation 12.

E.3 NTK VARIANCE: B

The Feynman rules (1)-(5) yield the following diagrams for the tensor B:

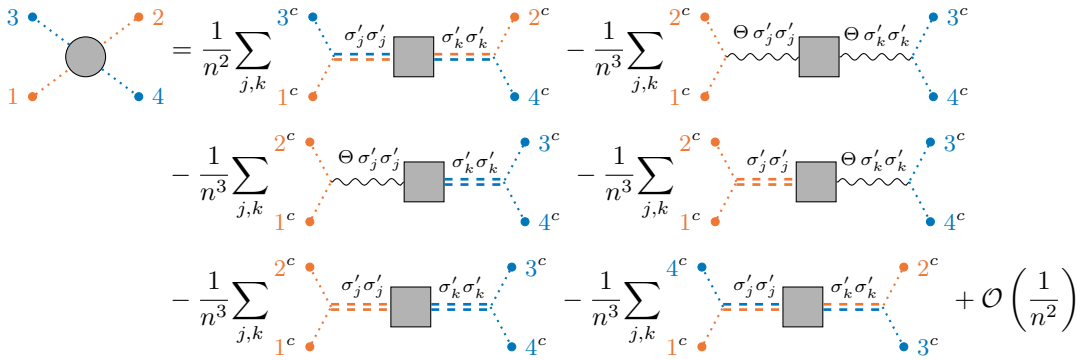




(89)

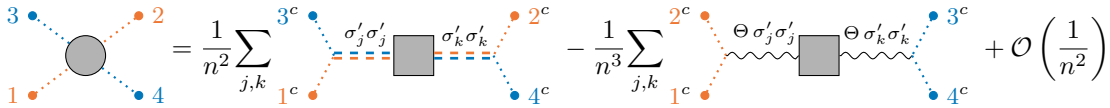
In equation 89, the application of Feyman rule (3) eliminates the presence of 2-class diagrams. Applying rule (4) to the inequivalent permutations of the reference pairing $(13)(24)$, namely $\{(13)(24)\}$, $\{(12)(34)\}$, $\{(14)(23)\}$ generates the 1-class diagrams shown in equation 89. The corresponding Weingarten coefficients multiplying each subdiagram are determined by the relative ordering between the reference pairing τ and the pairing π defining the subdiagram: $\mathcal{W}[\pi, \tau] = \mathcal{W}[1, 1]$ for $\pi = \tau$, $\mathcal{W}[\pi, \tau] = \mathcal{W}[2]$ for $\pi \neq \tau$. The overall factor of 1 follows from the Möbius formula for a single-block partition.

Restricting the analysis to first order in $1/n$, the expansions equation 54 and equation 55 reduce equation 89 to



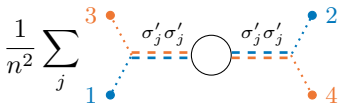
(90)

We now apply the second set of Feynman rules (6)-(9). In particular, the selection rules (a)-(f) eliminate the last four terms in equation 90. To cancel the prefactor $1/n^3$, the corresponding diagrams would have to contribute a factor of n^2, which could only arise from two disconnected bare propagators. However, such subdiagrams violate the color-conservation property of the propagator. Consequently, we are left with



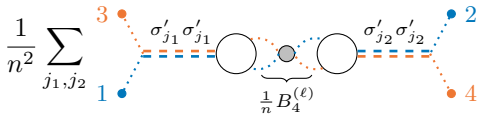
(91)

Applying Feynman rules (6)–(9) to the diagonal and off-diagonal neural components on the right-hand side of equation 91 generates the following diagrams: When $i = j$, the first term yields



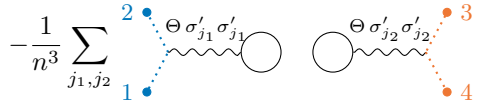
(92)

When $i \neq j$, the first term produces



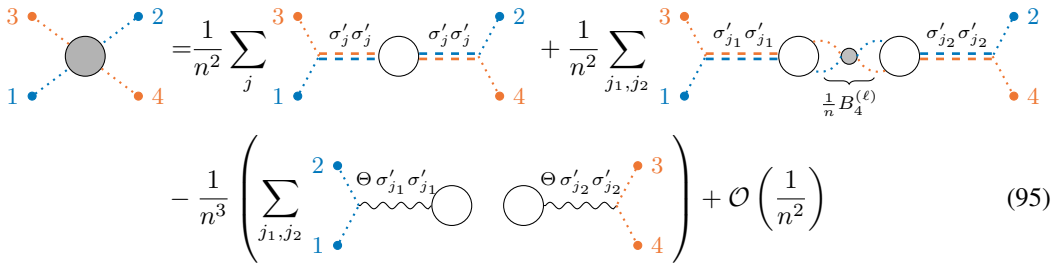
(93)

whereas the second term becomes



(94)

Substituting equation 92, equation 93, and equation 94 into equation 91, we obtain



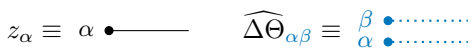
(95)

The linear $1/n$ scaling of the right-hand side of equation 95 follows from simple power counting of the associated summations. Diagrams involving a single neural index contribute a factor of n, whereas those involving two distinct neural indices contribute a factor of n^2. Combined with the explicit prefactors in equation 95, each term on the right-hand side yields an overall contribution of order $1/n$. It is straightforward to verify that the first and second lines on the RHS of equation 95 reproduce the corresponding lines on the RHS of equation 70.

F FEYNMAN RULES AT LEADING ORDER IN $1/n$

We present an alternative formulation of the Feynman rules introduced in Section 4, valid at leading order in $1/n$, which closely parallels the diagrammatic framework of Guillen et al. (2025) in the Gaussian setting. For simplicity, we restrict attention to the statistics of preactivations and the NTK. We adopt the same conventions and notation as in Guillen et al. (2025):

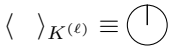
1. Preactivations and NTKs are represented by external lines, as illustrated below.



(96)

   where a colored line corresponds to a single NTK label.

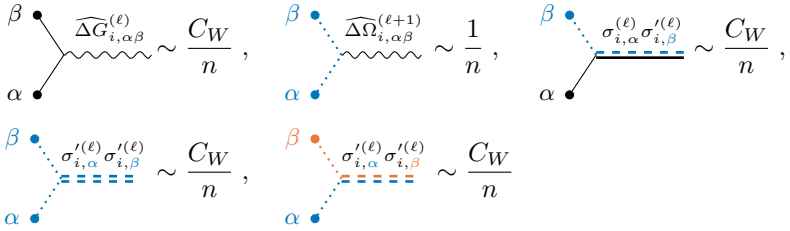
2. The propagator is represented by



(97)

   where $\langle\ \ \rangle_{K^{(\ell)}}$ denotes a zero-mean Gaussian expectation with covariance specified by $K^{(\ell)}$. The expectation value is taken over the decorations of the internal lines attached to the propagator, which satisfies the set of selection rules (a)-(f) listed in Rule (6) of Section 4.

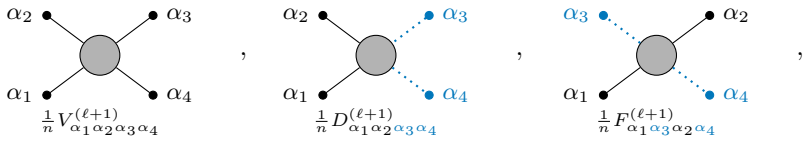
3. Cubic vertices are defined as in Guillen et al. (2025). Explicitly,

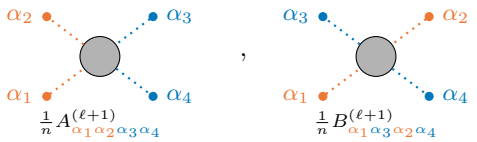


(98)

   where $\widehat{\Omega}^{(\ell+1)}_{i,\alpha\beta} = \sigma^{(\ell)}_{i,\alpha}\sigma^{(\ell)}_{i,\beta} + C_W\Theta^{(\ell)}_{\alpha\beta}\sigma'^{(\ell)}_{i,\alpha}\sigma'^{(\ell)}_{i,\beta}$ and $\widehat{\Delta\Omega}^{(\ell+1)}_{i,\alpha\beta} = \widehat{\Omega}^{(\ell+1)}_{i,\alpha\beta} - \langle\widehat{\Omega}^{(\ell+1)}_{i,\alpha\beta}\rangle_{K^{(\ell)}}$. Lines without a dot at one end are internal lines.

4. Quartic vertices are defined analogously, following Guillen et al. (2025). Explicitly,





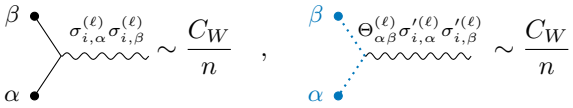
(99)

5. The orthogonal diagram describing the $2m$-point cumulant for the reference pairing $\pi = (12)(34)\cdots(2m-1\ 2m)$ is obtained from the Gaussian one, using the Feynman rules equation 98 and equation 99, by summing over all reconnections of the external labels:

$$V^{\text{orth}}_{2m,\pi} = V^{\text{gauss}}_{2m,\pi} + \sum_{\lambda\vdash m,\ \lambda\neq(1,\ldots,1)} \frac{\beta_\lambda}{n^{\,m-\ell(\lambda)}} \sum_{\tau\in\mathcal{C}_\lambda(\pi)} \prod_{j=1}^{\ell(\lambda)} V^{\text{conn}}_{2\lambda_j} \tag{100}$$

Here:

- $\lambda = (\lambda_1, \ldots, \lambda_{\ell(\lambda)}) \vdash m$ is a partition of m, and $\ell(\lambda)$ is its number of parts;
- $\mathcal{C}_\lambda(\pi)$ denotes the set of pairings τ such that, when compared with π, the external labels split into blocks of sizes $2\lambda_1, \ldots, 2\lambda_{\ell(\lambda)}$;
- each block of size $2\lambda_j$ contributes a connected correlator $V^{\text{conn}}_{2\lambda_j}$, with V^{conn}_2 given by



(101)

- β_λ is the leading coefficient of the orthogonal Weingarten function for pairings in the class λ, namely

$$\mathcal{W}[\tau,\pi] = \frac{\beta_\lambda}{n^{\,m-\ell(\lambda)}} + O\Big(n^{-(m-\ell(\lambda)+1)}\Big), \qquad \tau\in\mathcal{C}_\lambda(\pi). \tag{102}$$

Its general expression is given by

$$\beta_\lambda = \prod_{i=1}^{\ell} (-1)^{\lambda_i - 1}\, C_{\lambda_i - 1} \tag{103}$$

where C_k is the k-th Catalan number, $C_k = \frac{1}{k+1}\binom{2k}{k}$. For low orders one finds

$$\begin{aligned} &\beta_{(2)} = -1, \\ &\beta_{(2,1)} = -1, \qquad \beta_{(3)} = 2, \\ &\beta_{(2,1,1)} = -1, \qquad \beta_{(2,2)} = 1, \qquad \beta_{(3,1)} = 2, \qquad \beta_{(4)} = -5. \end{aligned} \tag{104}$$

We next apply these rules to derive the recursion relations for the tensors F and B, recovering the expressions previously obtained from the general Feynman rules of Section 4, namely equation 19 and equation 95.

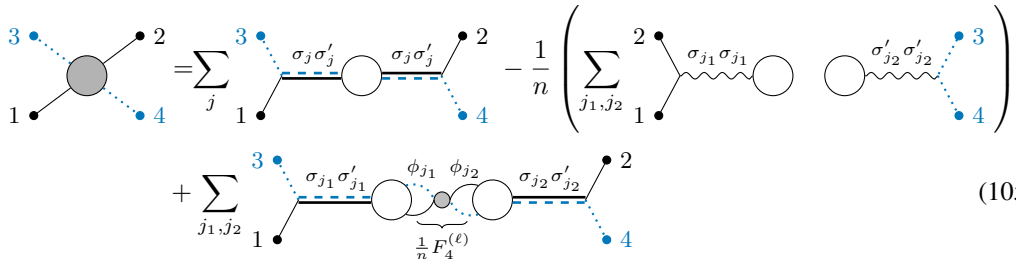
The layer-$(\ell+1)$ tensor on the LHS of equation 12 is represented by the third quartic vertex in equation 99. We enumerate all Feynman diagrams compatible with these external lines at order $1/n$. Since the fifth Feynman rule mixes pairing orderings, the analysis splits into two cases.

For the ordering 1324, the Gaussian contraction scheme applies: the only admissible cubic vertex connecting a dotted (NTK) line with a solid (preactivation) line is the third vertex in equation 98. This introduces two internal solid-dashed lines weighted by $\sigma^{(\ell)}_i \sigma'^{(\ell)}_i$, which may carry either identical or distinct channel indices. In the former case, the lines are connected by a single propagator; in the latter, by two propagators and an internal F-tensor at layer ℓ.

For all other orderings, the fifth Feynman rule applies and requires subtracting the contributions associated with the orderings 1234 and 1423. The pairings $(12)(34)$ and $(14)(23)$, when composed with the reference pairing $(13)(24)$, yield the partition $\lambda = (2)$, with length $\ell(\lambda) = 1$. Consequently, these contributions scale as $1/n$, with coefficient $\beta_{(2)} = -1$. Each pairing contributes a factor V_2, leading to a product of two such vertices. By the selection rules (a)-(f) of Section 4, the $(14)(23)$ term

vanishes at order $1/n$ due to violation of color-line preservation. The only non-vanishing contribution therefore arises from $(12)(34)$, yielding a product of two cubic vertices: one involving NTK lines and one involving two preactivation lines, as specified in equation 101.

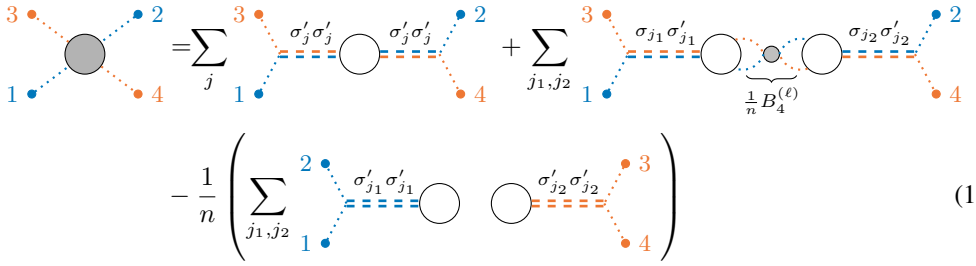
The diagrams corresponding to equation 12 are thus given by



(105)

The sums over channel indices are included in equation 105. These are the only diagrams contributing at order $1/n$; all others are either higher order or excluded by the selection rules.

The recursion relation for the tensor B at order $1/n$ follows analogously. In the Gaussian sector, ordering 1324, the only compatible cubic vertex is the fifth vertex in equation 98, yielding two diagrams: one with identical channel indices connected by a single propagator, and one with distinct indices connected by two propagators and an internal B-tensor at layer ℓ. For the remaining orderings, the 1423 contribution vanishes by color-line preservation. The 1234 ordering, constructed from the cubic vertex in equation 101, instead yields a nonvanishing contribution. The associated prefactor is determined by the cycle structure of the pairing $(12)(34)$ relative to $(13)(24)$, namely $\lambda(2)$ with $\beta_{(2)} = -1$. We therefore obtain



(106)

The sums over channel indices are included in equation 106. These exhaust all contributions at order $1/n$; all remaining diagrams are either higher order or ruled out by the selection rules.

We have verified that the leading-order Feynman rules (1)-(5) remain valid at order $1/n^2$, as shown in Appendix I.

A useful application of the leading-order $1/n$ Feynman rules (1)-(5) is an alternative proof of Theorem 5.1. At this order, orthogonality enters only through the introduction of lower-rank tensors relative to the tensor under consideration, as prescribed by rule (5). Consequently, criticality can be determined by analyzing the first term in equation 100, corresponding to the Gaussian-like sector. In this way, network stability follows as in the Gaussian case. In particular, if the infinite-width NNGP and NTK are at criticality, then the orthogonal network is also at criticality.

G Proof of Main Theorems

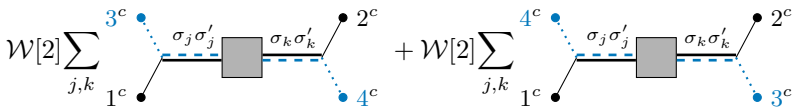
G.1 Proof of Theorem 4.1

Theorem 4.1. *The Feynman rules stated in items 1-9, uniquely determine the recursion relations governing the layerwise evolution of the orthogonal NTK tensors D, F, A, B at order $1/n$.*

This statement was already proved for the tensors F and B in Appendix E. To show that the Feynman rules of Appendix D also determine the recursion relations for the tensors D and A at order $1/n$, it suffices to argue that any corrections arising from orthogonality are at least of order $1/n^2$, as can be inferred from equation 58 and equation 69. We now verify this case by case.

G.1.1 JOINT NTK-PREACTIVATION CUMULANT

Proof. Orthogonality enters only through Feynman rule (4), which introduces the Weingarten function $\mathcal{W}[2]$ for pairings different from the reference pairing. For the tensor D, the reference pairing is $\tau = (12)(34)$. The diagrams associated with pairings inequivalent to τ are therefore given by

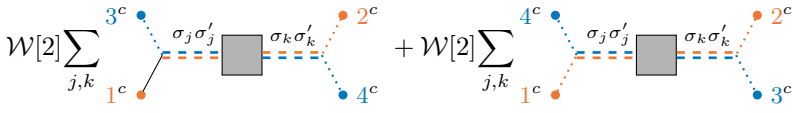


(107)

Since the leading term of $\mathcal{W}[2]$ scales as $1/n^3$, these contributions can survive at order $1/n$ only if the corresponding diagrams produce factors of n^2. This would occur if the square propagator decomposed into two bare propagators with distinct neural indices j and k. However, such configurations violate color conservation in the propagator, as specified by the selection rules (a)–(f). Consequently, these diagrams vanish at order $1/n$, implying that the recursion relation for the orthogonal tensor D coincides with its Gaussian counterpart at order $1/n$. □

G.1.2 NTK VARIANCE

Proof. Similarly to equation G.1.1, orthogonality induces the following diagrams for the tensor A with reference pairing (12)(34):



(108)

Once again, the leading scaling $\mathcal{W}[2] \sim 1/n^3$ requires the diagrams to contribute a factor n^2. This would occur if the square propagator decomposed into two bare propagators with distinct neural indices j and k. However, such configurations violate color conservation in the propagators and therefore vanish at order $1/n$. Consequently, the recursion relation for the orthogonal tensor A coincides with its Gaussian counterpart at order $1/n$. □

G.2 PROOF OF THEOREM 4.2

Theorem 4.2. *The set of Feynman rules presented above, in conjunction with those defined in Appendix D, reproduces the recursion relations governing the orthogonal dNTK and ddNTK tensors: P, Q and R, S, T, U, respectively, at order $1/n$.*

Proof. We prove this statement by showing that the recursion relations for the dNTK and ddNTK tensors coincide with those derived in the Gaussian case. As discussed in the main text, this result is further supported by the numerical simulations presented in Section 5.2 for the single-input setting.

To see this, note that orthogonality enters through Feynman rule (4), which introduces the Weingarten function $W[2]$ multiplying the 1-class diagrams obtained from inequivalent permutations with respect to the reference pairing. Since the leading scaling of $W[2]$ is $1/n^3$, such contributions could survive at order $1/n$ only if the square propagator decomposed into two bare propagators, so that the corresponding Gaussian expectation values with distinct neural indices produce a factor of n^2. However, for the preactivation-dNTK cumulant and the mean of the dNTK this configuration is forbidden, as it violates color conservation in the bare propagators. Indeed, since the dNTK and ddNTKs are represented by colored lines as in equation 71, and the cubic vertices involving these lines always include an internal line carrying at least one color, see equation equation 72, no bare propagator can attach to such vertices. Consequently, the diagrams proportional to $\mathcal{W}[2]$ in the recursion relations of the dNTK and ddNTK tensors decouple at order $1/n$. Therefore, the recursion relations for the orthogonal dNTK and ddNTK tensors coincide with those obtained in the Gaussian setup. □

G.3 PROOF OF THEOREM 4.3

Theorem 4.3. *The Feynman rules defined above, augmented by the higher-order generalizations of the tensors V, D, F, A, B, P, Q, R, S, T, U, provide a complete characterization of the statistics of the orthogonal NTK and its descendants at arbitrary order in $1/n$.*

Proof. We begin by observing that the restriction to cubic vertices with two external legs and one internal line in our Feynman rules follows from two structural properties of orthogonal weights: the vanishing of all odd moments, and the fact that weight contractions always occur in pairs of neural indices.

The square propagator introduced in Section 4 generates all possible pairings of internal lines. This is equivalent to summing over all connected and disconnected diagrams with an even number of external legs. In parallel, the permutation of external labels carrying orthogonal charge, weighted by the corresponding Weingarten functions, implements the orthogonal contractions of the weight parameters appearing in the definition of the observables. The signs associated with each diagram match those prescribed by the Möbius relation (Rule (5)), thereby reproducing the standard combinatorial structure of cumulants.

Concretely, applying this procedure to n preactivations with a fixed pairing structure generates all even-point diagrams, each weighted by the appropriate Weingarten function. Since these weights encode orthogonal contractions, the square propagators can be identified with expectation values over the weights multiplying the corresponding activations. This allows one to rewrite the full expression as products of parameter expectation values of preactivations $z_{\alpha,i} = \sum_j W_{ij}\sigma_{\alpha,j}$, each weighted by Möbius factors determined by the associated set partition. This coincides with the standard definition of the cumulant of n preactivations.

As an illustration, consider six preactivations with reference pairing $(12)(34)(56)$. Following the discussion above, the square propagator generates all possible even-point connected and disconnected diagrams, weighted by the corresponding Weingarten functions $\mathcal{W}[1,1,1]$, $\mathcal{W}[2,1]$, $\mathcal{W}[3]$, $\mathcal{W}[2]$, $\mathcal{W}[1,1]$, $\mathcal{W}[1]$ (see Appendix I.2 for details). These weights and their associated index contractions coincide with those obtained from the orthogonality of the parameters, allowing one to rewrite the resulting contributions as products of expectation values such as

$$\begin{gathered}\mathbb{E}[z_1z_2z_3z_4z_5z_6] \quad , \quad \mathbb{E}[z_1z_2z_3z_4]\mathbb{E}[z_5z_6] \quad , \quad \mathbb{E}[z_1z_2z_5z_6]\mathbb{E}[z_3z_4] \quad , \\ \mathbb{E}[z_3z_4z_5z_6]\mathbb{E}[z_1z_2] \quad , \quad \mathbb{E}[z_1z_2]\mathbb{E}[z_3z_4]\mathbb{E}[z_5z_6] \end{gathered} \tag{109}$$

Upon attaching the corresponding Möbius prefactors and summing over all partitions, one recovers the contribution of the channel $(12)(34)(56)$ to the standard sextic cumulant.

The use of the multileg definitions equation 71, together with the Feynman rules with coloured lines in equation 72, allows one to straightforwardly extend this argument to composite observables such as the NTK, dNTK, and ddNTK, which are multilinear in the weights.

The second set of Feynman rules implements an effective field theory expansion for non-Gaussian probability distributions. In this framework, the square propagator (full expectation value) decomposes into a Gaussian (free) propagator and interaction vertices arising from the $1/n$-expansion of the distribution. As shown in prior work (Guillen et al., 2025), this expansion is governed by a unique cubic vertex structure obtained by Gaussianizing the argument of the square propagator, together with higher-order interaction vertices. In particular, no new cubic vertices appear at higher orders, highlighting the uniqueness of our Feynman rules.

Combining these observations, Rules (1)-(5) reproduce the cumulant structure with orthogonal contractions, while Rules (6)-(9) generate the $1/n$-expansion of the expectation values entering the cumulant in terms of Gaussian propagators and higher-order interaction vertices. This completes the proof. □

G.4 Proof of Theorem 5.1

Theorem 5.1. *Criticality of the infinite-width NNGP and NTK implies criticality of orthogonal preactivation and NTK cumulants.*

Proof. The proof is twofold. First, as established in Guillen et al. (2025), the stability analysis admits a bootstrap structure: criticality of lower-rank tensors suffices to control higher-rank tensors, whose

behavior can then be analyzed systematically using the Feynman rules of Section 4 or Appendix D. Second, at any fixed order in $1/n$, orthogonality modifies the recursion of a given tensor only through the introduction of lower-rank tensors, as prescribed by Rule 4 in Section 4.1. This rule introduces Weingarten functions multypling diagrams generated by inequivalent permutations relative to the reference pairing. By definition, these contributions are subleading compared to the identity permutation. Since these lower-rank tensors are already assumed to be critical under the bootstrap hypothesis, the resulting stability analysis for orthogonal tensors coincides with that of the Gaussian case. □

H Analytical recursion relations: Single-input case

In this appendix, we list the recursion relations obtained from the Feynman diagram approach in the single-input setting for the NTK, dNTK, and ddNTK tensors. The expressions for the tensors D, F, A and B follow directly from equation 58, equation 19, equation 69, and equation 70, respectively. The relations for the tensors P, Q, R, S, T, U can be derived analogously through the systematic application of the Feynman rules in Appendix D. Since these relations coincide with their Gaussian counterparts at order $1/n$, one may also use the results of Roberts & Yaida (2022). Explicitly, we obtain

$$D^{(\ell+1)} = \chi_{\perp}^{(\ell)}\chi_{\parallel}^{(\ell)} D^{(\ell)} + \left(\frac{\lambda_W^{(\ell+1)}}{C_W}\right)\Big[C_W^2\,\langle\sigma(z)\sigma(z)\sigma(z)\sigma(z)\rangle_{K^{(\ell)}} - \big(C_W g^{(\ell)}\big)^2 + \big(\chi_{\parallel}^{(\ell)}\big)^2 V^{(\ell)}\Big] + \Theta^{(\ell)}\Big[C_W^2\,\langle\sigma(z)\sigma(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} - C_W g^{(\ell)}\,\chi_{\perp}^{(\ell)} + 2h^{(\ell)}\,\chi_{\parallel}^{(\ell)}\,V^{(\ell)}\Big], \tag{110}$$

$$F^{(\ell+1)} = \big(\chi_{\parallel}^{(\ell)}\big)^2 F^{(\ell)} + C_W^2\left(\langle\sigma(z)\sigma(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} - \langle\sigma(z)\sigma(z)\rangle_{K^{(\ell)}}\langle\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}}\right)\Theta^{(\ell)}, \tag{111}$$

$$\begin{aligned} A^{(\ell+1)} = {} & \big(\chi_{\perp}^{(\ell)}\big)^2 A^{(\ell)} + \left(\frac{\lambda_W^{(\ell+1)}}{C_W}\right)^2\Big[C_W^2\,\langle\sigma(z)\sigma(z)\sigma(z)\sigma(z)\rangle_{K^{(\ell)}} - \big(C_W g^{(\ell)}\big)^2 + \big(\chi_{\parallel}^{(\ell)}\big)^2 V^{(\ell)}\Big] \\ & + 2\left(\frac{\lambda_W^{(\ell+1)}}{C_W}\right)\Theta^{(\ell)}\Big[C_W^2\,\langle\sigma(z)\sigma(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} - C_W g^{(\ell)}\,\chi_{\perp}^{(\ell)} + 2h^{(\ell)}\,\chi_{\parallel}^{(\ell)}\,V^{(\ell)}\Big] \\ & + 2\left(\frac{\lambda_W^{(\ell+1)}}{C_W}\right)\chi_{\perp}^{(\ell)}\chi_{\parallel}^{(\ell)} D^{(\ell)} + 4h^{(\ell)}\chi_{\perp}^{(\ell)}\Theta^{(\ell)} D^{(\ell)} \\ & + \big(\Theta^{(\ell)}\big)^2\Big[C_W^2\,\langle\sigma'(z)\sigma'(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} - \big(\chi_{\perp}^{(\ell)}\big)^2 + \big(2h^{(\ell)}\big)^2 V^{(\ell)}\Big], \end{aligned} \tag{112}$$

$$B^{(\ell+1)} = \big(\chi_{\perp}^{(\ell)}\big)^2 B^{(\ell)} + C_W^2\left(\langle\sigma'(z)\sigma'(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} - \langle\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}}\langle\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}}\right)\big(\Theta^{(\ell)}\big)^2, \tag{113}$$

$$P^{(\ell+1)} = C_W^2\left\langle\sigma''(z)\sigma'(z)\sigma'(z)\sigma(z)\right\rangle_{K^{(\ell)}}(\Theta^{(\ell)})^2 + C_W\,\chi_{\perp}^{(\ell)}\left\langle\sigma''(z)\sigma(z)\right\rangle_{K^{(\ell)}} B^{(\ell)} + \Big[C_W\,\chi_{\perp}^{(\ell)}\left\langle\sigma''(z)\sigma(z)\right\rangle_{K^{(\ell)}} + \big(\chi_{\perp}^{(\ell)}\big)^2\Big]P^{(\ell)}, \tag{114}$$

$$Q^{(\ell+1)} = C_W^2\left\langle\sigma''(z)\sigma'(z)\sigma'(z)\sigma(z)\right\rangle_{K^{(\ell)}}(\Theta^{(\ell)})^2 + \frac{\lambda_W^{(\ell+1)}}{C_W}\,F^{(\ell+1)} + 2h^{(\ell)}\,\chi_{\parallel}^{(\ell)}\,\Theta^{(\ell)}\,F^{(\ell)} + \Big[C_W\,\chi_{\perp}^{(\ell)}\left\langle\sigma''(z)\sigma(z)\right\rangle_{K^{(\ell)}} + \big(\chi_{\perp}^{(\ell)}\big)^2\Big]Q^{(\ell)} \tag{115}$$

$$
\begin{aligned}
R^{(\ell+1)} &= \left(\chi_{\perp}^{(\ell)}\right)^2 R^{(\ell)} + \lambda_W^{(\ell+1)} C_W \langle \sigma''(z)\sigma'(z)\sigma'(z)\sigma(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^2 \\
&\quad + C_W^2 \langle \sigma'''(z)\sigma'(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^3 \\
&\quad + \chi_{\perp}^{(\ell)} \Big(\lambda_W^{(\ell+1)} \langle \sigma''(z)\sigma(z)\rangle_{K^{(\ell)}} + C_W \Theta^{(\ell)} \langle \sigma'''(z)\sigma'(z)\rangle_{K^{(\ell)}} \Big) \Big(B^{(\ell)} + P^{(\ell)} \Big) \\
&\quad + \chi_{\perp}^{(\ell)} \Big(\lambda_W^{(\ell+1)} \langle \sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} + C_W \Theta^{(\ell)} \langle \sigma''(z)\sigma''(z)\rangle_{K^{(\ell)}} \Big) P^{(\ell)},
\end{aligned} \tag{116}
$$

$$
\begin{aligned}
S^{(\ell+1)} &= \left(\chi_{\perp}^{(\ell)}\right)^2 S^{(\ell)} + \lambda_W^{(\ell+1)} C_W \langle \sigma'(z)\sigma'(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^2 \\
&\quad + C_W^2 \langle \sigma''(z)\sigma''(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^3 \\
&\quad + \chi_{\perp}^{(\ell)} \Big[\lambda_W^{(\ell+1)} \langle \sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} + C_W \Theta^{(\ell)} \langle \sigma''(z)\sigma''(z)\rangle_{K^{(\ell)}} \Big] B^{(\ell)},
\end{aligned} \tag{117}
$$

$$
\begin{aligned}
T^{(\ell+1)} &= \left(\chi_{\perp}^{(\ell)}\right)^2 T^{(\ell)} + 2 C_W \lambda_W^{(\ell+1)} \langle \sigma''(z)\sigma'(z)\sigma'(z)\sigma(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^2 \\
&\quad + C_W^2 \langle \sigma''(z)\sigma''(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^3 \\
&\quad + (\lambda_W^{(\ell+1)})^2 \Theta^{(\ell)} \langle \sigma'(z)\sigma'(z)\sigma(z)\sigma(z)\rangle_{K^{(\ell)}} \\
&\quad + \Big(\lambda_W^{(\ell+1)} \langle z\sigma'(z)\sigma(z)\rangle_{K^{(\ell)}} + C_W \Theta^{(\ell)} \langle z\sigma''(z)\sigma'(z)\rangle_{K^{(\ell)}} \Big)^2 \frac{F^{(\ell)}}{(K^{(\ell)})^2} \\
&\quad + 2\chi_{\perp}^{(\ell)} \Big[\lambda_W^{(\ell+1)} \Big(\langle \sigma''(z)\sigma(z)\rangle_{K^{(\ell)}} + \langle \sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} \Big) \\
&\quad + C_W \Theta^{(\ell)} \Big(\langle \sigma'''(z)\sigma'(z)\rangle_{K^{(\ell)}} + \langle \sigma''(z)\sigma''(z)\rangle_{K^{(\ell)}} \Big) \Big] Q^{(\ell)},
\end{aligned} \tag{118}
$$

$$
U^{(\ell+1)} = \left(\chi_{\perp}^{(\ell)}\right)^2 U^{(\ell)} + C_W^2 \langle \sigma''(z)\sigma''(z)\sigma'(z)\sigma'(z)\rangle_{K^{(\ell)}} (\Theta^{(\ell)})^3, \tag{119}
$$

where the susceptibilities $\chi_{||}^{(\ell)}$, $\chi_{\perp}^{(\ell)}$, and the auxiliary functions $g^{(\ell)}(K)$, $h^{(\ell)}(K)$ are defined as follows:

$$
\begin{aligned}
\chi_{||}^{(\ell)} &= \frac{C_W}{K^{(\ell)}} \langle z^{(\ell)} \sigma^{(\ell)}(z) \sigma'^{(\ell)}(z)\rangle_{K^{(\ell)}}, \quad \chi_{\perp}^{(\ell)} = C_W \langle \sigma'^{(\ell)}(z)\sigma'^{(\ell)}(z)\rangle_{K^{(\ell)}} \\
h^{(\ell)}(K) &= \frac{C_W}{4(K^{(\ell)})^2} \langle ((z^{(\ell)})^2 - K^{(\ell)})(\sigma'^{(\ell)}(z)\sigma'^{(\ell)}(z))\rangle_{K^{(\ell)}}, \quad g^{(\ell)}(K) = \langle \sigma^{(\ell)}\sigma^{(\ell)}\rangle_{K^{(\ell)}}
\end{aligned} \tag{120}
$$

H.1 Solution to the single-input recursions

In this subsection, we analyze the solutions of the recursions equation 110–equation 119, shown in Figure 9. As in Section 5.2, we consider a square neural network with $\tanh$ activation, width $n = 50$, and depth $L = 10$, and perform both numerical and symbolic computations in *Mathematica*. The input vector $x_0 \in \mathbb{R}^{50}$ is taken to be

$$
x_0 = \begin{pmatrix} 0.934738 \\ 0.26696 \\ 0.784097 \\ 0.656448 \\ 0.305308 \\ 0.401958 \\ 0.894594 \\ 0.0559893 \\ 0.000643274 \\ 0.0274513 \end{pmatrix} \oplus \begin{pmatrix} 0.377754 \\ 0.127474 \\ 0.879907 \\ 0.710555 \\ 0.509949 \\ 0.312682 \\ 0.0854376 \\ 0.869372 \\ 0.114232 \\ 0.0851646 \end{pmatrix} \oplus \begin{pmatrix} 0.254697 \\ 0.560475 \\ 0.508664 \\ 0.0271565 \\ 0.426426 \\ 0.457646 \\ 0.913778 \\ 0.40436 \\ 0.407187 \\ 0.0644401 \end{pmatrix} \oplus \begin{pmatrix} 0.256718 \\ 0.869761 \\ 0.0406222 \\ 0.431362 \\ 0.906228 \\ 0.55979 \\ 0.275852 \\ 0.553722 \\ 0.235762 \\ 0.751627 \end{pmatrix} \oplus \begin{pmatrix} 0.178558 \\ 0.411167 \\ 0.100846 \\ 0.220264 \\ 0.215917 \\ 0.490943 \\ 0.596323 \\ 0.0799147 \\ 0.205998 \\ 0.0372218 \end{pmatrix}. \tag{121}
$$

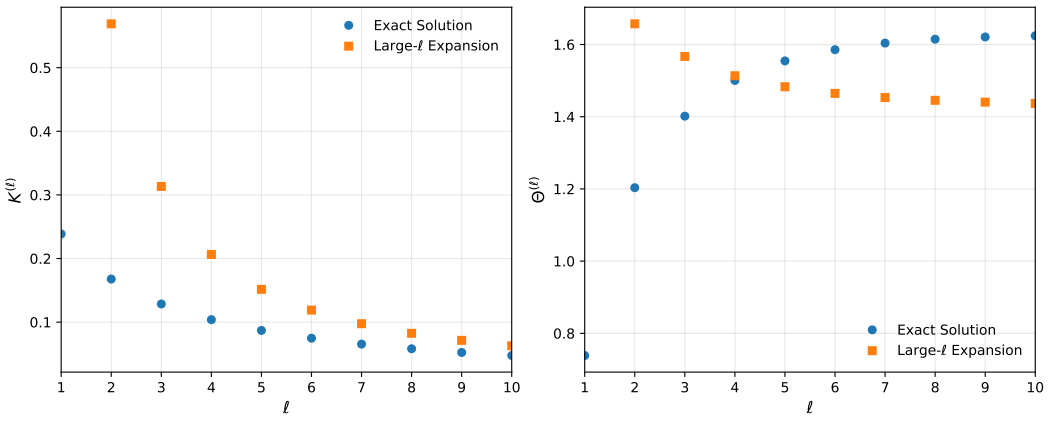


Figure 3: *The NNGP and NTK.* Solution of the single-input recursion relations equation 5 and equation 10. We consider a $\tanh$ network of width $n = 50$ and depth $L = 10$ with inputs drawn from $(0, 1)$. Blue points denote the exact solutions of equation 165 and equation 10, while orange boxes represent the large-ℓ expansions equation 181 and equation 183, evaluated at integer ℓ. The tensor magnitudes remain smaller than their Gaussian counterparts and exhibit early-layer saturation.

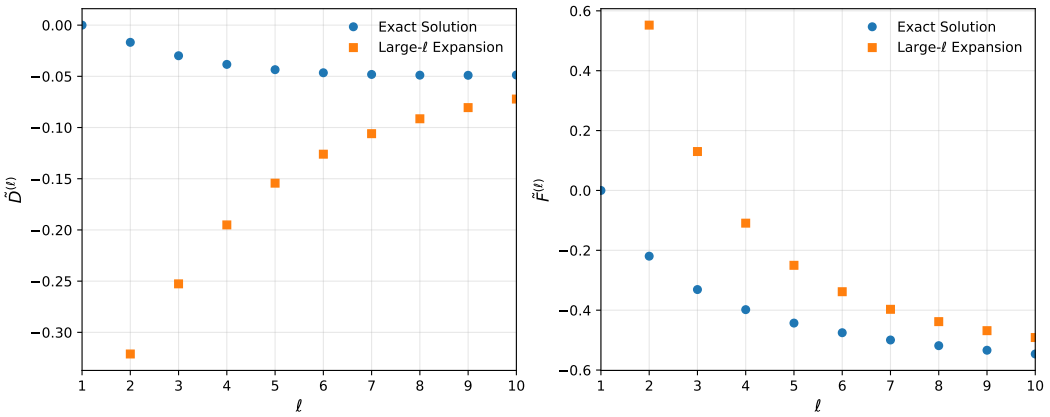


Figure 4: *The NTK mixed tensors.* Solution of the single-input recursion relations equation 110 and equation 111. We consider a $\tanh$ network of width $n = 50$ and depth $L = 10$ with inputs drawn from $(0, 1)$. Blue points denote the exact solutions of equation 110 and equation 111, while orange boxes represent the large-ℓ expansions equation 184 and equation 185, evaluated at integer ℓ. The tensor magnitudes remain smaller than their Gaussian counterparts and exhibit early-layer saturation.

The normalized tensors are defined by

$$\tilde{B}^{(\ell)} = \frac{B^{(\ell)}}{\left(\Theta^{(\ell)}\right)^2}, \qquad \tilde{A}^{(\ell)} = \frac{A^{(\ell)}}{\left(\Theta^{(\ell)}\right)^2}, \qquad \tilde{P}^{(\ell)} = \frac{P^{(\ell)}}{\left(\Theta^{(\ell)}\right)^2}, \qquad \tilde{Q}^{(\ell)} = \frac{Q^{(\ell)}}{\left(\Theta^{(\ell)}\right)^2},$$
$$\tilde{R}^{(\ell)} = \frac{R^{(\ell)}K^{(\ell)}}{\left(\Theta^{(\ell)}\right)^3}, \qquad \tilde{S}^{(\ell)} = \frac{S^{(\ell)}K^{(\ell)}}{\left(\Theta^{(\ell)}\right)^3}, \qquad \tilde{T}^{(\ell)} = \frac{T^{(\ell)}K^{(\ell)}}{\left(\Theta^{(\ell)}\right)^3}, \qquad \tilde{U}^{(\ell)} = \frac{U^{(\ell)}K^{(\ell)}}{\left(\Theta^{(\ell)}\right)^3}. \tag{122}$$

We report results for the normalized tensors in the following order: the NTK-mixed tensors $\tilde{D}$ and $\tilde{F}$ (Figure 4); the NTK variance tensors $\tilde{A}$ and $\tilde{B}$ (Figure 5); the dNTK tensors $\tilde{P}$ and $\tilde{Q}$ (Figure 6); the $\mathrm{dd_I NTK}$ tensor $\tilde{R}$ (Figure 7); and the $\mathrm{dd_{II} NTK}$ tensors $\tilde{S}$, $\tilde{T}$, and $\tilde{U}$ (Figure 8). The corresponding large-ℓ expansions, derived in Appendix J, are shown for comparison.

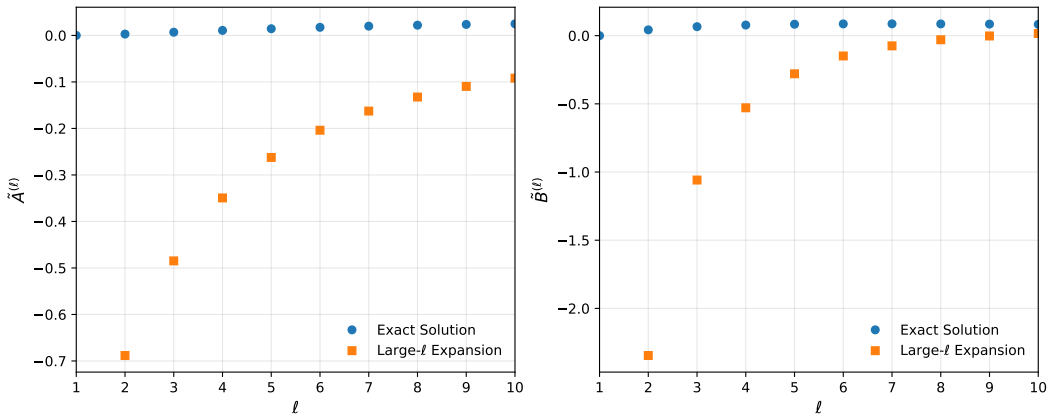


Figure 5: *The NTK variance tensors.* Solution of the single-input recursion relations equation 112 and equation 113. We consider a $\tanh$ network of width $n = 50$ and depth $L = 10$ with inputs drawn from $(0, 1)$. Blue points denote the exact solutions of equation 112 and equation 113, while orange boxes represent the large-ℓ expansion equation 186 and equation 187, evaluated at integer ℓ. The magnitude of the tensors remains smaller than its Gaussian counterpart and exhibits early-layer saturation.

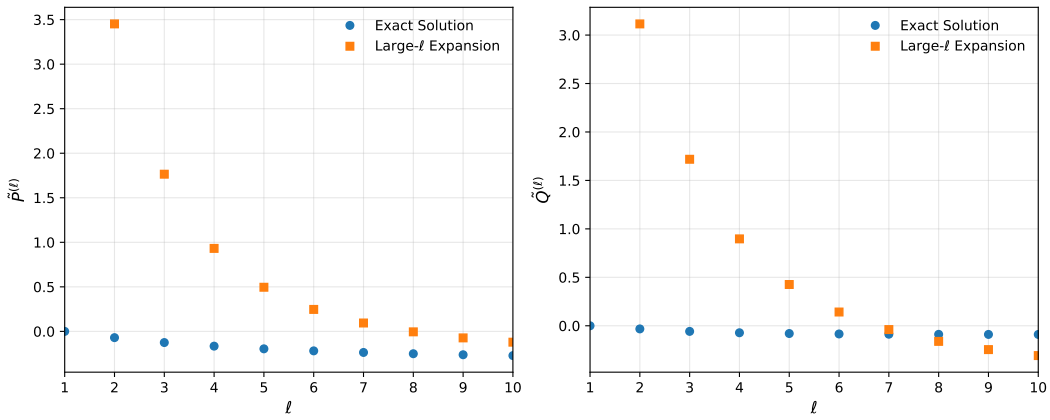


Figure 6: *The dNTK tensors.* Solutions of the single-input recursion relations equation 114 and 115. We consider a $\tanh$ network of width $n = 50$ and depth $L = 10$ with inputs drawn from $(0, 1)$. Blue points denote the exact solution of equation 114 and 115, while orange boxes represent the large-ℓ expansions equation 188 and equation 189, evaluated at integer ℓ. The magnitude of the tensors remains smaller than its Gaussian counterpart and exhibits early-layer saturation.

For ease of comparison, we provide a combined plot of all NTK, dNTK, and ddNTK tensors in Figure 9. The results show that all tensors attain smaller magnitudes than their Gaussian counterparts and exhibit early-layer saturation, consistent with the empirical findings of Day et al. (2025), which are quantitatively reproduced here.

I THE V_6 TENSOR AT ORDER $1/n^2$

In this appendix, we compute the recursion relation governing the layer evolution of the sextic vertex V_6 using both algebraic and diagrammatic methods.

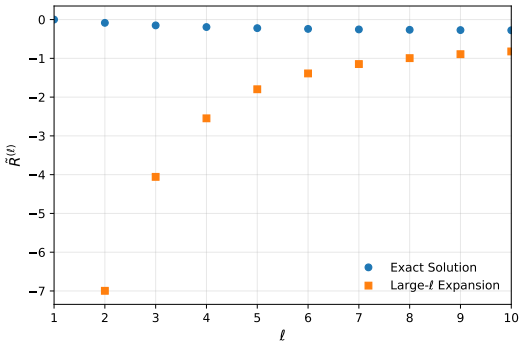


Figure 7: *The d_INTK tensor.* Solution of the single-input recursion relation equation 116. We consider a $\tanh$ network of width $n = 50$ and depth $L = 10$ with inputs drawn from $(0, 1)$. Blue points denote the exact solution of equation 116, while orange boxes represent the large-ℓ expansion equation 190, evaluated at integer ℓ. The magnitude of the tensor remains smaller than its Gaussian counterpart and exhibits early-layer saturation.

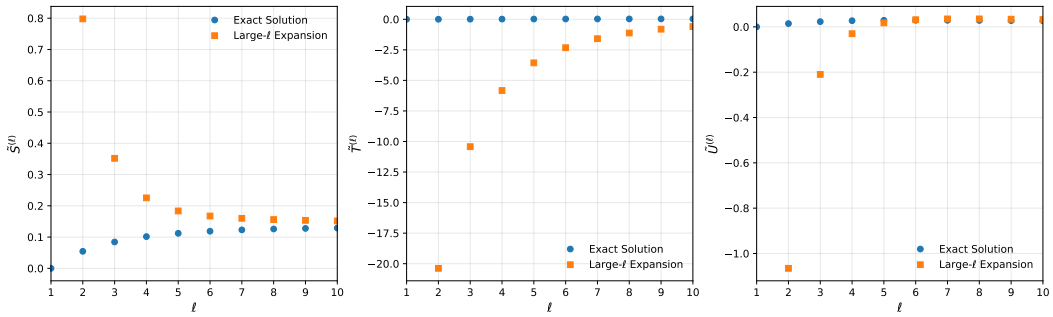


Figure 8: *The dd_{II}NTK tensors.* Solutions to the single-input recursion relations equation 117, equation 118, and equation 119. We consider a $\tanh$ network with width $n = 50$ and depth $L = 10$, with inputs drawn from $(0, 1)$. Blue points show the exact solutions of equation 117, equation 118, and equation 119, while orange boxes represent the corresponding large-ℓ expansions equation 191, equation 192, and equation 193, evaluated at integer ℓ. The tensor magnitude remains below its Gaussian counterpart and exhibits early-layer saturation.

I.1 Algebraic derivation

The sextic vertex is defined from the six-point cumulant of the preactivations as follows

$$\begin{aligned}
\mathbb{E}^c_\theta[z^{(\ell+1)}_{i_1,\alpha_1} \dots z^{(\ell+1)}_{i_6,\alpha_6}] &= \mathbb{E}_\theta[z^{(\ell+1)}_{i_1,\alpha_1} \dots z^{(\ell+1)}_{i_6,\alpha_6}] \\
&\quad - \Big[\mathbb{E}_\theta[z^{(\ell+1)}_{i_1,\alpha_1} \dots z^{(\ell+1)}_{i_4,\alpha_4}]\mathbb{E}_\theta[z^{(\ell+1)}_{i_5,\alpha_5} z^{(\ell+1)}_{i_6,\alpha_6}] + \text{14 other (4,2) subdivisions}\Big] \\
&\quad + 2\Big[\mathbb{E}_\theta[z^{(\ell+1)}_{i_1,\alpha_1} z^{(\ell+1)}_{i_2,\alpha_2}]\mathbb{E}_\theta[z^{(\ell+1)}_{i_3,\alpha_3} z^{(\ell+1)}_{i_4,\alpha_4}]\mathbb{E}_\theta[z^{(\ell+1)}_{i_5,\alpha_5} z^{(\ell+1)}_{i_6,\alpha_6}] + \text{14 other (2,2,2) subdivisions}\Big] \\
&\quad = \frac{1}{n_\ell^2}\Big[\delta_{i_1 i_2}\delta_{i_3 i_4}\delta_{i_5 i_6} V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} + \text{14 other (2,2,2) subdivisions}\Big] \qquad (123)
\end{aligned}$$

To obtain an explicit expression for $V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$, we evaluate the expectation values of the orthogonal weights defining the preactivations $z_{i,\alpha}$. In the case of four weights, we introduce the basis $e_1 = (12)(34)$, $e_2 = (13)(24)$, $e_3 = (14)(23)$. The general definition of the Weingarten function, $\mathcal{W}[\tau, \pi] = \mathcal{W}[\ell(\tau \circ \pi)]$, where $\circ$ denotes ordinary permutation multiplication and $\ell(e)$ denotes the cycle length of e, implies that $\mathcal{W}[e_i, e_j] = \mathcal{W}[1, 1]$ if $e_i = e_j$, and $\mathcal{W}[e_i, e_j] = \mathcal{W}[2]$ if

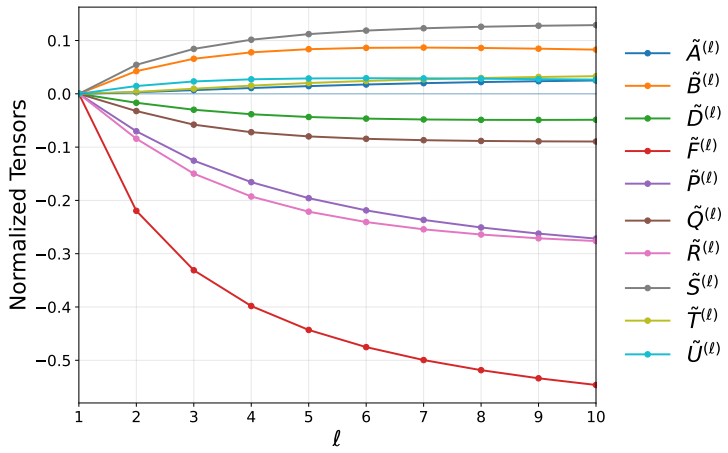


Figure 9: *Orthogonal saturation.* Exact solutions of the single-input recursion relations at order $1/n$ (see Theorems 4.1 and 4.2). We consider a $\tanh$ network of width $n = 50$ with inputs drawn from the real interval $(0, 1)$. All tensors attain smaller values than their Gaussian counterparts and exhibit early-layer saturation. These predictions are consistent with the empirical results of Day et al. (2025).

$e_i \neq e_j$. Therefore, in the basis (e_1, e_2, e_3), the $k = 2$ Weingarten matrix takes the form

$$M = \begin{pmatrix} \mathcal{W}[1,1] & \mathcal{W}[2] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[1,1] & \mathcal{W}[2] \\ \mathcal{W}[2] & \mathcal{W}[2] & \mathcal{W}[1,1] \end{pmatrix} \tag{124}$$

For six weights, the corresponding basis consists of 15 elements, namely

$$\begin{aligned} &\tilde{e}_1 = (12)(34)(56), \quad \tilde{e}_2 = (12)(35)(46), \quad \tilde{e}_3 = (12)(36)(45), \\ &\tilde{e}_4 = (13)(24)(56), \quad \tilde{e}_5 = (13)(25)(46), \quad \tilde{e}_6 = (13)(26)(45), \\ &\tilde{e}_7 = (14)(23)(56), \quad \tilde{e}_8 = (14)(25)(36), \quad \tilde{e}_9 = (14)(26)(35), \\ &\tilde{e}_{10} = (15)(23)(46), \quad \tilde{e}_{11} = (15)(24)(36), \quad \tilde{e}_{12} = (15)(26)(34), \\ &\tilde{e}_{13} = (16)(23)(45), \quad \tilde{e}_{14} = (16)(24)(35), \quad \tilde{e}_{15} = (16)(25)(34). \end{aligned} \tag{125}$$

In this case, the cycle lengths can take only three possible values, namely $(1, 1, 1)$, $(2, 1)$, and (3), depending on the structure resulting from the product of three permutations. Consequently, in the basis equation 125, the $k = 3$ Weingarten matrix is represented by a 15×15 matrix whose first row reads

$$\begin{aligned} &\big(\mathcal{W}[1,1,1],\ \mathcal{W}[2,1],\ \mathcal{W}[2,1],\ \mathcal{W}[2,1],\ \mathcal{W}[3],\ \mathcal{W}[3],\ \mathcal{W}[2,1], \\ &\mathcal{W}[3],\ \mathcal{W}[3],\ \mathcal{W}[3],\ \mathcal{W}[3],\ \mathcal{W}[2,1],\ \mathcal{W}[3],\ \mathcal{W}[3],\ \mathcal{W}[2,1]\big) \end{aligned} \tag{126}$$

Using equation 124 and equation 126, we then obtain

$$V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} = V^{\mathcal{G}(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} + \Delta V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} \tag{127}$$

where $V^{\mathcal{G}(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ is defined as

$$\begin{aligned} \frac{1}{n_\ell^2} V^{\mathcal{G}(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} &= (C_W^{(\ell+1)})^3 \sum_{i,j,k=1}^{n} \Big[\mathcal{W}[1,1,1]\mathbb{E}[\sigma^{(\ell)}_{i,1}\sigma^{(\ell)}_{i,2}\sigma^{(\ell)}_{j,3}\sigma^{(\ell)}_{j,4}\sigma^{(\ell)}_{k,5}\sigma^{(\ell)}_{k,6}] \\ &- \mathcal{W}[1,1]\cdot\mathcal{W}[1]\Big(\mathbb{E}[\sigma^{(\ell)}_{i,1}\sigma^{(\ell)}_{i,2}\sigma^{(\ell)}_{j,3}\sigma^{(\ell)}_{j,4}]\mathbb{E}[\sigma^{(\ell)}_{k,5}\sigma^{(\ell)}_{k,6}] + \mathbb{E}[\sigma^{(\ell)}_{j,3}\sigma^{(\ell)}_{j,4}\sigma^{(\ell)}_{k,5}\sigma^{(\ell)}_{k,6}]\mathbb{E}[\sigma^{(\ell)}_{i,1}\sigma^{(\ell)}_{i,2}] \\ &+ \mathbb{E}[\sigma^{(\ell)}_{i,1}\sigma^{(\ell)}_{i,2}\sigma^{(\ell)}_{k,5}\sigma^{(\ell)}_{k,6}]\mathbb{E}[\sigma^{(\ell)}_{j,3}\sigma^{(\ell)}_{j,4}]\Big) + 2(\mathcal{W}[1])^3\mathbb{E}[\sigma^{(\ell)}_{i,1}\sigma^{(\ell)}_{i,2}]\mathbb{E}[\sigma^{(\ell)}_{j,3}\sigma^{(\ell)}_{j,4}]\mathbb{E}[\sigma^{(\ell)}_{k,5}\sigma^{(\ell)}_{k,6}]\Big] \end{aligned} \tag{128}$$

and $\Delta V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ is given by

$$\begin{aligned}
\frac{1}{n_\ell^2}\Delta V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} = (C_W^{(\ell+1)})^3 \sum_{i,j,k=1}^{n} \Big[& \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,i}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,k}] \\
& + \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,i}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,j}] + \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,j}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,k}] \\
& + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,k}] + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,j}] \\
& + \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,i}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,k}] + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,i}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,k}] \\
& + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,i}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,j}] + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,i}\sigma^{(\ell)}_{\alpha_6,k}] \\
& + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,j}\sigma^{(\ell)}_{\alpha_5,i}\sigma^{(\ell)}_{\alpha_6,k}] + \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,i}\sigma^{(\ell)}_{\alpha_6,j}] \\
& + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,i}] + \mathcal{W}[3]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,j}\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,i}] \\
& + \mathcal{W}[2,1]\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,k}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,i}] \\
& - \mathcal{W}[2]\cdot\mathcal{W}[1]\Big(\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,j}] + \mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_3,j}\sigma^{(\ell)}_{\alpha_4,i}]\Big)\mathbb{E}[\sigma^{(\ell)}_{\alpha_5,k}\sigma^{(\ell)}_{\alpha_6,k}] \\
& - \mathcal{W}[2]\cdot\mathcal{W}[1]\Big(\mathbb{E}[\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,j}\sigma^{(\ell)}_{\alpha_5,i}\sigma^{(\ell)}_{\alpha_6,j}] + \mathbb{E}[\sigma^{(\ell)}_{\alpha_3,i}\sigma^{(\ell)}_{\alpha_4,j}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,i}]\Big)\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,k}\sigma^{(\ell)}_{\alpha_2,k}] \\
& - \mathcal{W}[2]\mathcal{W}[1]\Big(\mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_5,i}\sigma^{(\ell)}_{\alpha_6,j}] + \mathbb{E}[\sigma^{(\ell)}_{\alpha_1,i}\sigma^{(\ell)}_{\alpha_2,j}\sigma^{(\ell)}_{\alpha_5,j}\sigma^{(\ell)}_{\alpha_6,i}]\Big)\mathbb{E}[\sigma^{(\ell)}_{\alpha_3,k}\sigma^{(\ell)}_{\alpha_4,k}]\Big] \qquad (129)
\end{aligned}$$

The term $V^{\mathcal{G}(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ captures the contribution of the diagonal components of the Weingarten matrices and reduces to the Gaussian case in the infinite-width limit, whereas $\Delta V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ encodes the off-diagonal contributions.

We now compute $V^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ at order $1/n^2$. To this end, we expand the $k=3$ Weingarten functions as follows

$$\begin{aligned}
\mathcal{W}[1,1,1] &= \frac{1}{n^3} + \frac{6}{n^5} + O\left(\frac{1}{n^6}\right), \\
\mathcal{W}[2,1] &= -\frac{1}{n^4} + \frac{1}{n^5} + O\left(\frac{1}{n^6}\right), \qquad (130) \\
\mathcal{W}[3] &= \frac{2}{n^5} + O\left(\frac{1}{n^6}\right).
\end{aligned}$$

Combining equation 130, equation 54 and equation 55, we obtain

$$\begin{aligned}
\frac{1}{n_\ell^2}V^{\mathcal{G}(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} &= \frac{1}{n_\ell^2}V^{G(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} + \frac{6}{n_\ell^2}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}} \\
&\quad - \frac{3\cdot 2}{n_\ell^2}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}} \\
&= \frac{1}{n_\ell^2}\mathcal{V}^{G(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} \qquad (131)
\end{aligned}$$

The factor 6 in the first equality of equation 131 originates from $\mathcal{W}[1,1,1]$ in equation 128, while the factor $2\cdot 3$ arises from the three contributions of the form $\mathcal{W}[1,1]\cdot\mathcal{W}[1]$. The tensor $V^{G(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ denotes the Gaussian six-point cumulant at order $\frac{1}{n^2}$. The corresponding recursion relation was derived explicitly in Banta et al. (2024) and reads

$$\begin{aligned}
\frac{1}{n_\ell^2}V^{G(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6} &= \frac{\left(C_W^{(\ell+1)}\right)^3}{n_\ell^2}\left\langle \widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}\,\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}\,\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}\right\rangle_{K^{(\ell)}} \\
&\quad + \frac{\left(C_W^{(\ell+1)}\right)^3}{4\,n_\ell n_{\ell-1}}\sum_{\beta_i\in\{\alpha_1,\ldots,\alpha_6\}} V^{(\ell)}_{\beta_1\beta_2\beta_3\beta_4}\left[\left\langle\frac{d^2\left(\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}\,\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}\right)}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle\frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}}\right.
\end{aligned}$$

$$
\begin{aligned}
&+\left\langle \frac{d^2\big(\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}\,\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}\big)}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}}+\left\langle \frac{d^2\big(\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}\,\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}\big)}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}}\Bigg]\\
&+\frac{\big(C_W^{(\ell+1)}\big)^3}{8\,n_{\ell-1}^2}\sum_{\beta_i\in\{\alpha_1,\ldots,\alpha_6\}}V^{(\ell)}_{\beta_1\beta_2\beta_3\beta_4\beta_5\beta_6}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}}{dz^{(\ell)}_{\beta_5}\,dz^{(\ell)}_{\beta_6}}\right\rangle_{K^{(\ell)}}\\
&+\frac{\big(C_W^{(\ell+1)}\big)^3}{16\,n_{\ell-1}^2}\sum_{\beta_i\in\{\alpha_1,\ldots\alpha_6\}}V^{(\ell)}_{\beta_1\beta_2\beta_3\beta_4}\,V^{(\ell)}_{\beta_5\beta_6\beta_7\beta_8}\Bigg[\left\langle \frac{d^4\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}\,dz^{(\ell)}_{\beta_7}\,dz^{(\ell)}_{\beta_8}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}}{dz^{(\ell)}_{\beta_5}\,dz^{(\ell)}_{\beta_6}}\right\rangle_{K^{(\ell)}}\\
&+\left\langle \frac{d^4\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}\,dz^{(\ell)}_{\beta_7}\,dz^{(\ell)}_{\beta_8}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}}{dz^{(\ell)}_{\beta_5}\,dz^{(\ell)}_{\beta_6}}\right\rangle_{K^{(\ell)}}\\
&+\left\langle \frac{d^4\widehat{\Delta G}^{(\ell)}_{\alpha_5\alpha_6}}{dz^{(\ell)}_{\beta_3}\,dz^{(\ell)}_{\beta_4}\,dz^{(\ell)}_{\beta_7}\,dz^{(\ell)}_{\beta_8}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_1\alpha_2}}{dz^{(\ell)}_{\beta_1}\,dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}\left\langle \frac{d^2\widehat{\Delta G}^{(\ell)}_{\alpha_3\alpha_4}}{dz^{(\ell)}_{\beta_5}\,dz^{(\ell)}_{\beta_6}}\right\rangle_{K^{(\ell)}}\Bigg]. \qquad (132)
\end{aligned}
$$

where $\widehat{\Delta G}^{(\ell)}_{\alpha\beta}=\sigma^{(\ell)}_\alpha\sigma^{(\ell)}_\beta-\langle\sigma^{(\ell)}_\alpha\sigma^{(\ell)}_\beta\rangle_{K^{(\ell)}}$.

Likewise, the off-diagonal contribution $\Delta\mathcal{V}^{(\ell+1)}_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}$ can be systematically computed as follows:

1. In the $\frac{1}{n}$-expansion of $\mathcal{W}[2,1]$, only the terms of order $\frac{1}{n^5}$ and $\frac{1}{n^4}$ contribute nontrivially. The former becomes relevant when all neural indices are distinct, yielding

$$
\begin{aligned}
&\frac{1}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big]
\end{aligned} \qquad (133)
$$

   The latter splits into two cases:

   - The first case arises when all neural indices are distinct:

$$
\begin{aligned}
&\frac{3}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big]
\end{aligned} \qquad (134)
$$

   - The second case occurs when exactly two neural indices coincide. In this case, we find two types of contributions: one that does not involve the four-point cumulant,

$$
\begin{aligned}
&-\frac{1}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\\
&\quad+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}
\end{aligned}
$$

$$+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\Big] \tag{135}$$

and one that explicitly involves V_4:

$$-\frac{1}{n_\ell n_{\ell-1}}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\Big] \tag{136}$$

where we have introduced the notation $\langle\sigma_\alpha\sigma_\beta|\sigma_\delta\sigma_\epsilon\rangle_V$ as shorthand for

$$\langle\sigma^{(\ell)}_\alpha\sigma^{(\ell)}_\beta|\sigma^{(\ell)}_\delta\sigma^{(\ell)}_\epsilon\rangle_{V^{(\ell)}}=\frac{1}{4}\sum_{\beta_i\in\{\alpha,\beta,\delta,\epsilon\}}\left\langle\frac{d^2(\sigma^{(\ell)}_\alpha\sigma^{(\ell)}_\beta)}{dz^{(\ell)}_{\beta_1}dz^{(\ell)}_{\beta_2}}\right\rangle_{K^{(\ell)}}V^{(\ell)}_{\beta_1\beta_2\beta_3\beta_4}\left\langle\frac{d^2(\sigma^{(\ell)}_\delta\sigma^{(\ell)}_\epsilon)}{dz^{(\ell)}_{\beta_3}dz^{(\ell)}_{\beta_4}}\right\rangle_{K^{(\ell)}} \tag{137}$$

2. In the $\frac{1}{n}$-expansion of $\mathcal{W}[3]$, only the term of order $\frac{1}{n^5}$ contributes. This occurs when all neural indices are distinct:

$$\frac{2}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\Big] \tag{138}$$

3. In the $\frac{1}{n}$-expansion of $\mathcal{W}[2]$, only the terms of order $\frac{1}{n^4}$ and $\frac{1}{n^3}$ contribute nontrivially. The $\frac{1}{n^4}$ term arises when all neural indices are distinct

$$-\frac{1}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big] \tag{139}$$

The $\frac{1}{n^3}$ contribution, in turn, splits into two cases:

- one in which all neural indices are distinct:

$$-\frac{1}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}$$

$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big] \tag{140}$$

- and one in which exactly two neural indices coincide. In the latter case, we again obtain two types of contributions: one that does not involve the four-point cumulant,

$$\frac{1}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big] \tag{141}$$

  and one that explicitly involves V_4

$$\frac{1}{n_\ell n_{\ell-1}}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\rangle_{K^{(\ell)}}$$
$$+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\Big] \tag{142}$$

Combining all contributions, we obtain

$$\begin{aligned}
\frac{1}{n_\ell^2}\Delta V^{(\ell+1)}_{\alpha_1\ldots\alpha_6} = &-\frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\Big]\\
&-\frac{(C_W^{(\ell+1)})^3}{n_\ell n_{\ell-1}}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_2}|\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}|\sigma^{(\ell)}_{\alpha_5}\sigma^{(\ell)}_{\alpha_6}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}|\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_4}\rangle_{V^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\Big]\\
&+\frac{2(C_W^{(\ell+1)})^3}{n_\ell^2}\Big[\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\\
&+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_3}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_4}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}+\langle\sigma^{(\ell)}_{\alpha_1}\sigma^{(\ell)}_{\alpha_6}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_2}\sigma^{(\ell)}_{\alpha_4}\rangle_{K^{(\ell)}}\langle\sigma^{(\ell)}_{\alpha_3}\sigma^{(\ell)}_{\alpha_5}\rangle_{K^{(\ell)}}\Big]
\end{aligned}$$

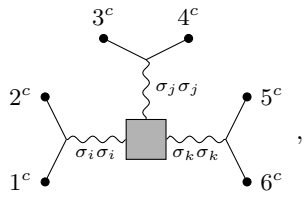
$$
\begin{aligned}
&+ \frac{2(C_W^{(\ell+1)})^3}{n_\ell^2}\Big[\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_2}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_2}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_5}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_5}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\Big]
\end{aligned}
\tag{143}
$$

or, more compactly

$$
\begin{aligned}
\frac{1}{n_\ell^2}\Delta V_{\alpha_1\ldots\alpha_6}^{(\ell+1)} &= -\frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\Big[\langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}\widehat{\Delta G}_{\alpha_4\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}\widehat{\Delta G}_{\alpha_4\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_3}^{(\ell)}\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_4}^{(\ell)}\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_4}^{(\ell)}\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_3}^{(\ell)}\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_5}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_6}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_6}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_5}^{(\ell)}\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\Big] \\
&- \frac{(C_W^{(\ell+1)})^3}{n_\ell n_{\ell-1}}\Big[\langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_5}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}|\widehat{\Delta G}_{\alpha_4\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_1\alpha_2}^{(\ell)}|\widehat{\Delta G}_{\alpha_4\alpha_5}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_3}^{(\ell)}|\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_4}^{(\ell)}|\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_4}^{(\ell)}|\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_3}^{(\ell)}|\widehat{\Delta G}_{\alpha_5\alpha_6}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_5}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_6}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\widehat{\Delta G}_{\alpha_1\alpha_6}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\widehat{\Delta G}_{\alpha_2\alpha_5}^{(\ell)}|\widehat{\Delta G}_{\alpha_3\alpha_4}^{(\ell)}\rangle_{V^{(\ell)}}\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\Big] \\
&+ \frac{2(C_W^{(\ell+1)})^3}{n_\ell^2}\Big[\langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}} \\
&+ \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_3}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_4}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}} + \langle\sigma_{\alpha_1}^{(\ell)}\sigma_{\alpha_6}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_2}^{(\ell)}\sigma_{\alpha_4}^{(\ell)}\rangle_{K^{(\ell)}}\langle\sigma_{\alpha_3}^{(\ell)}\sigma_{\alpha_5}^{(\ell)}\rangle_{K^{(\ell)}}\Big]
\end{aligned}
\tag{144}
$$

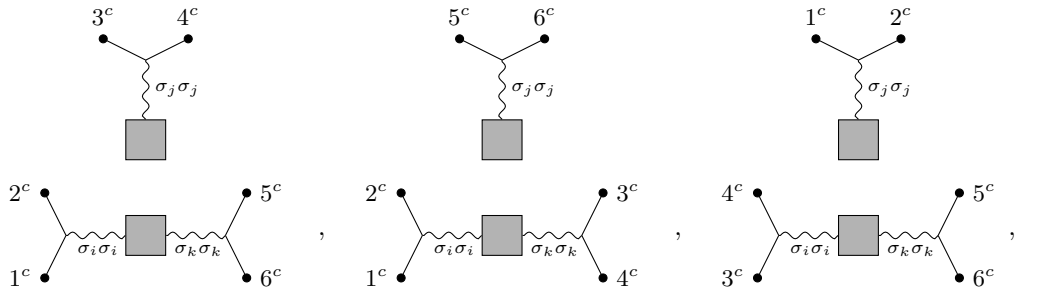
I.2 Diagrammatic derivation

We now derive the recursion relation for the sextic vertex equation 144 using the Feynman rules of Appendix D. Taking the reference pairing to be $(12)(34)(56)$, the application of rule (3) generates the 1-class diagram



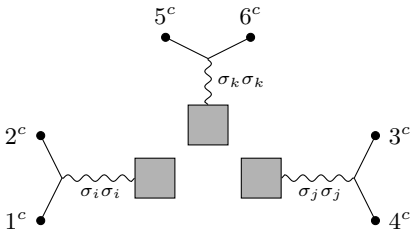
, (145)

as well as the corresponding 2-class diagrams



(146)

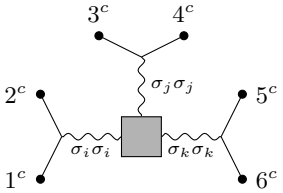
and the 3-class diagram



. (147)

The application of rule (4) introduces additional diagrams, each accompanied by a Weingarten function determined by the cycle structure resulting from the composition of the corresponding pairing with the reference pairing. As discussed in the previous subsection, for 1-class diagrams there are three possible cycle types, namely $(1,1,1)$, $(2,1)$, and (3). The fifteen 3-class diagrams can thus be organized as follows: one diagram of type $(1,1,1)$

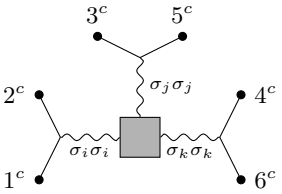
$\mathcal{W}[1,1,1]$



, (148)

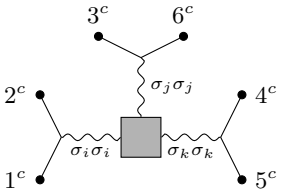
six diagrams of type $(2,1)$

$\mathcal{W}[2,1]$



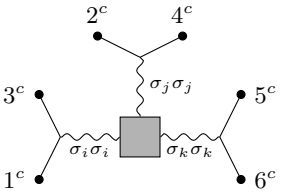
,

$\mathcal{W}[2,1]$



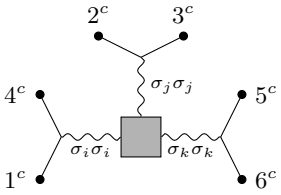
,

$\mathcal{W}[2,1]$



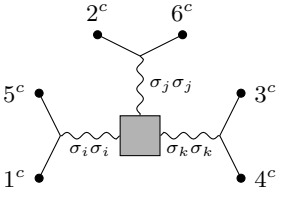
,

$\mathcal{W}[2,1]$



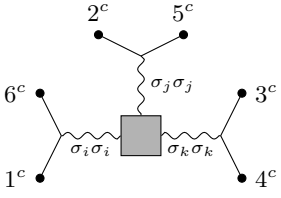
,

$\mathcal{W}[2,1]$



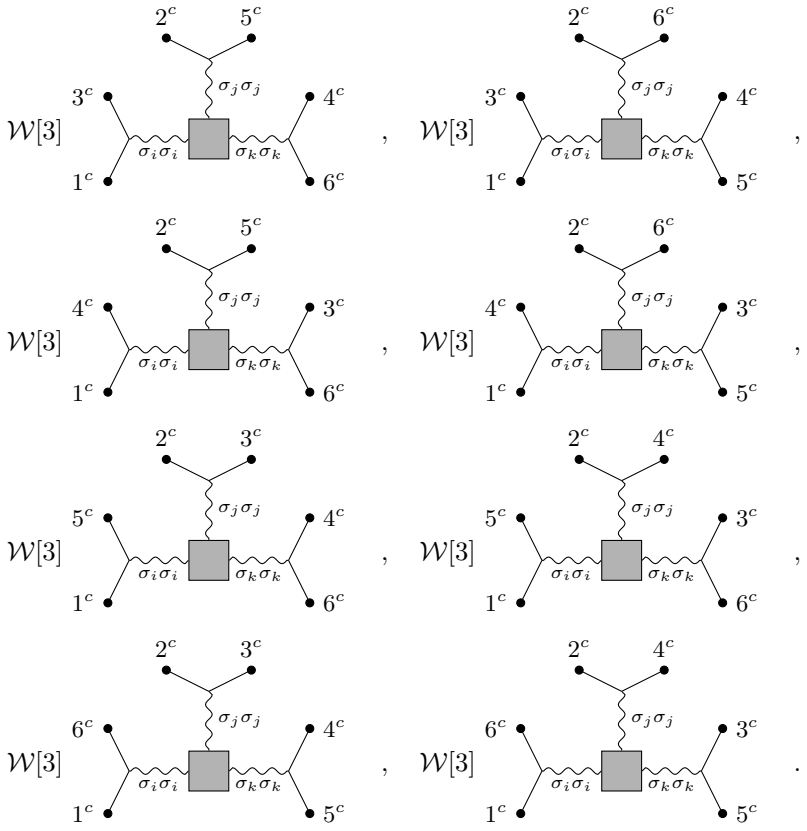
,

$\mathcal{W}[2,1]$

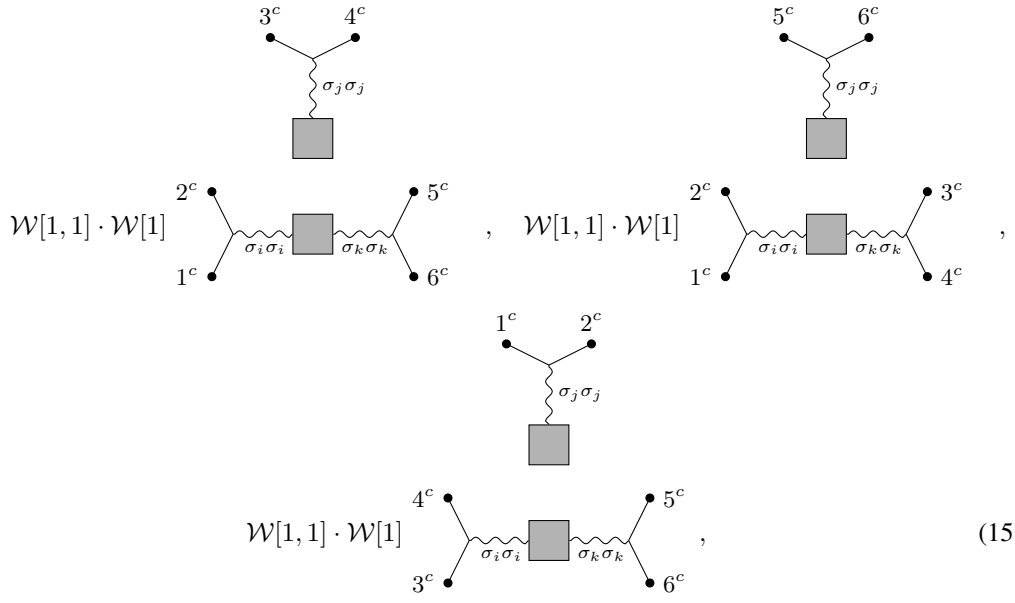


, (149)

and eight diagrams of type (3)

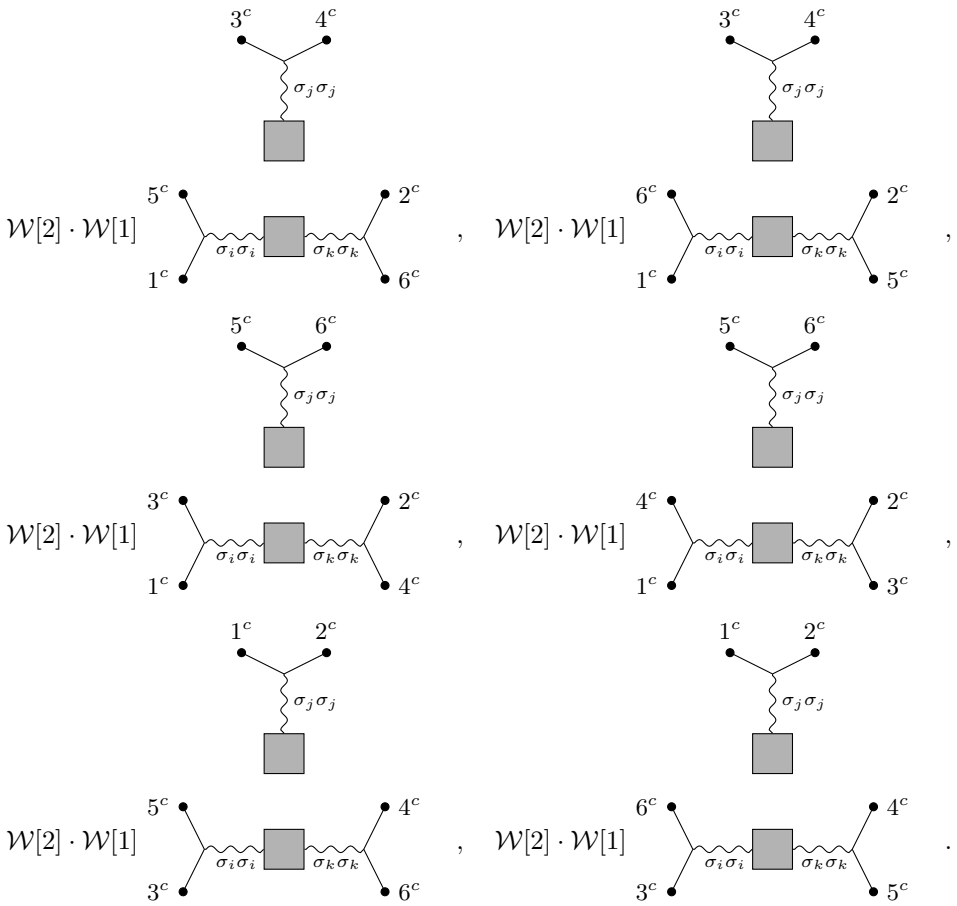


Similarly, the nine 2-class diagrams generated by rule (4) can be classified according to the associated $k = 2$ Weingarten factor. Explicitly, three diagrams are weighted by $\mathcal{W}[1,1]$



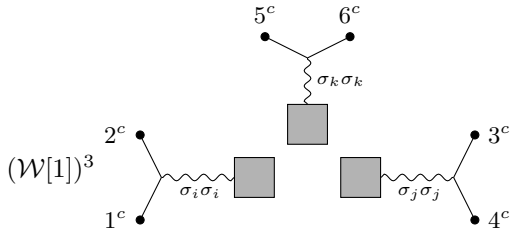
(151)

and six diagrams are weighted by $\mathcal{W}[2]$



(152)

The single 3-class diagram contributes as



(153)

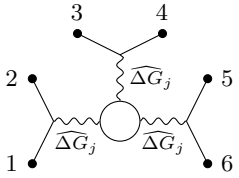
Möbius factors are incorporated through rule (5): 1-class diagrams are multiplied by 1, 2-class diagrams by -1, and 3-class diagrams by 2.

After summing over all neural indices and all diagrams described above, one recovers equation 127, with $V^{\mathcal{G}}$ given by equation 128 and ΔV defined in equation 129.

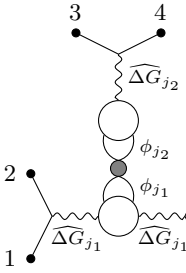
We now apply the Feynman rules (6)–(9) to the square propagator. In particular, we focus on the layer evolution of the sextic vertex at order $1/n^2$. To this end, the Weingarten functions are expanded as in equation 54, equation 55, and equation 130. The bare propagators, together with the corresponding quartic and sextic vertices, are then introduced in accordance with the selection rules (a)–(f). After a careful, albeit straightforward, implementation of these rules, we arrive at the following diagrammatic

decomposition of the tensor $V_6^{(\ell+1)}$:

$$\frac{1}{n_\ell^2} V^{(\ell+1)}_{123456} = \frac{1}{n_\ell^3} \sum_j$$

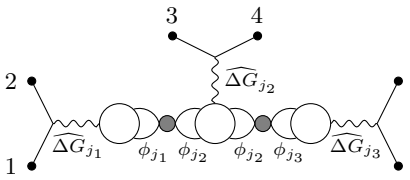


$$+ \frac{1}{n_\ell^3} \left(\sum_{j_1, j_2}$$



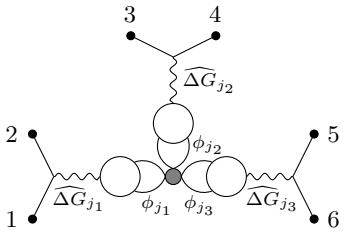
$$+ P(12,34,56) \Bigg)$$

$$+ \frac{1}{n_\ell^3} \left(\sum_{j_1, j_2, j_3}$$

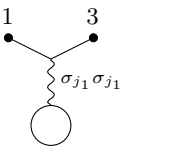


$$+ P(12,34,56) \Bigg)$$

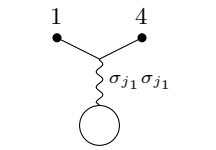
$$+ \frac{1}{n_\ell^3} \sum_{j_1, j_2, j_3}$$



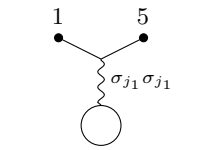
$$- \frac{1}{n_\ell^4} \Bigg(\sum_{j_1, j_2}$$

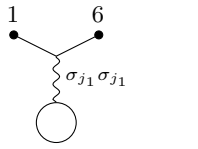


$$+ \sum_{j_1, j_2}$$

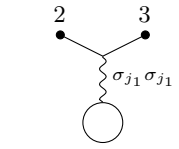


$$+ \sum_{j_1, j_2}$$

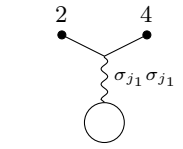




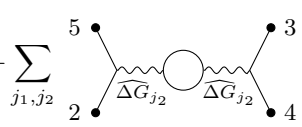
$$+ \sum_{j_1, j_2}$$



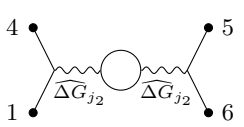
$$+ \sum_{j_1, j_2}$$



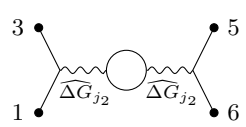
$$+ \sum_{j_1, j_2}$$

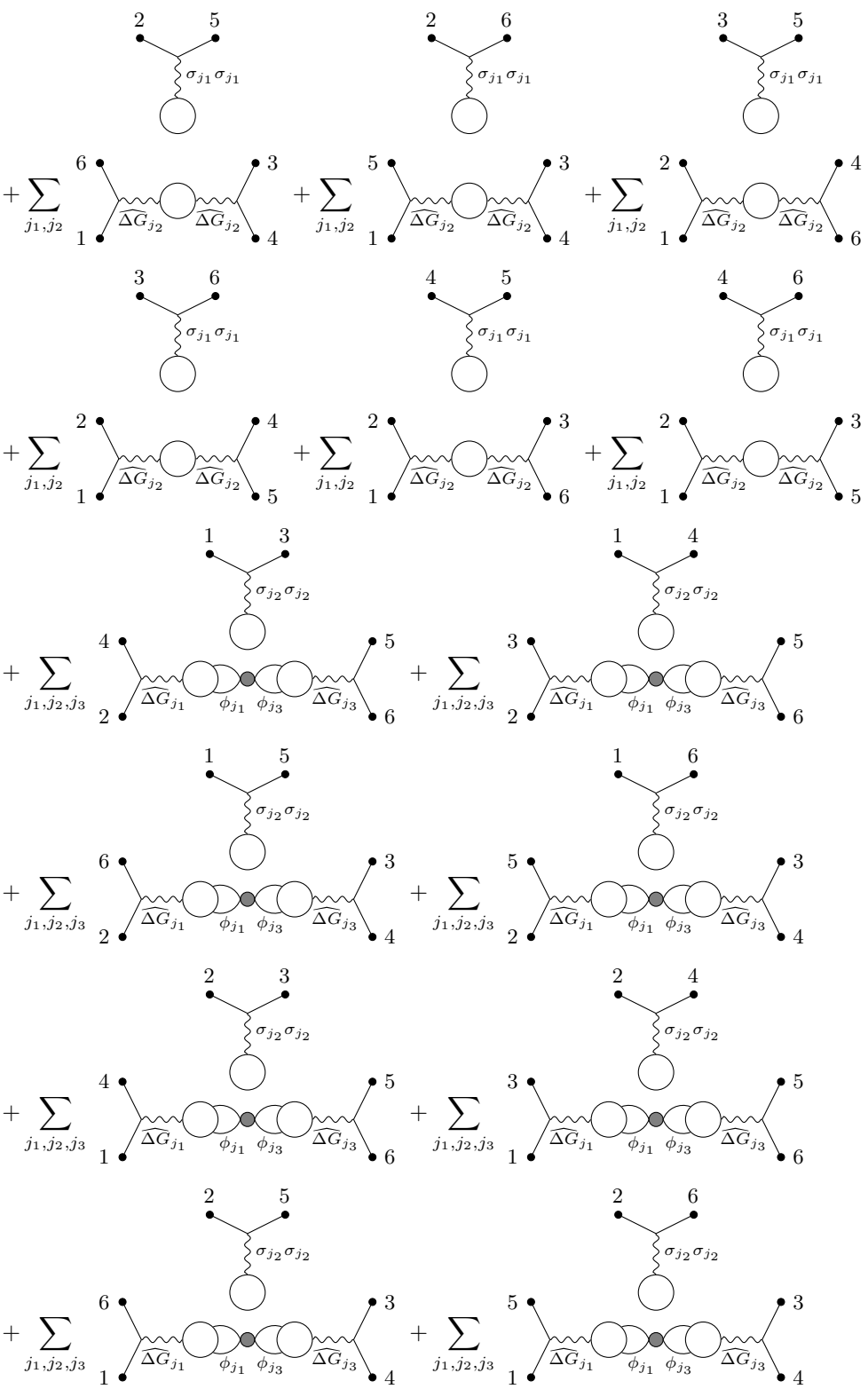


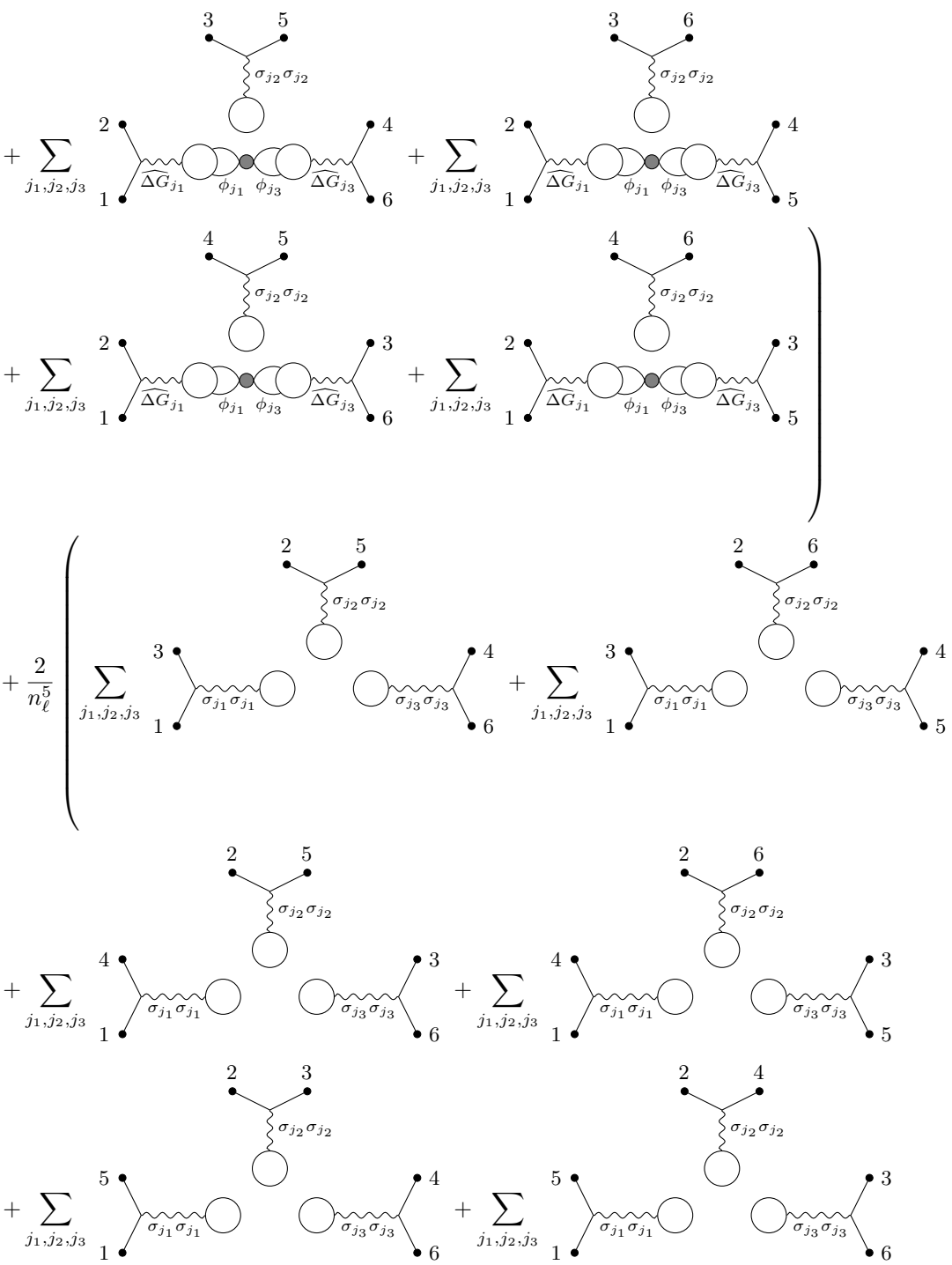
$$+ \sum_{j_1, j_2}$$

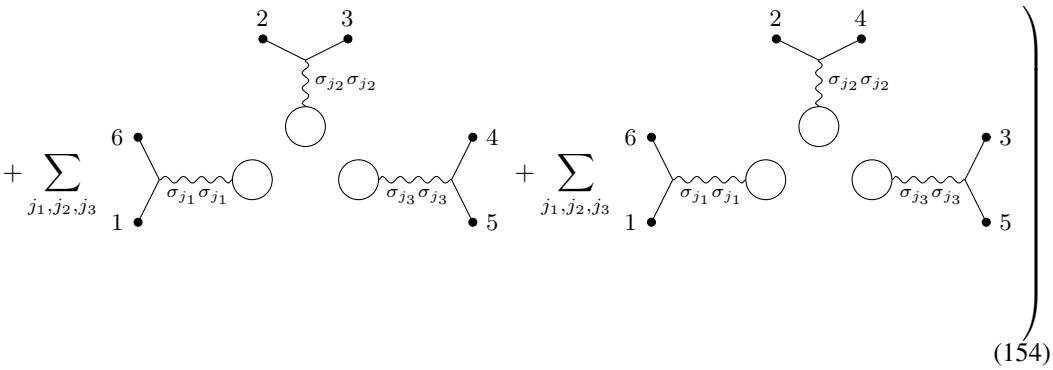


$$+ \sum_{j_1, j_2}$$




2 5
$\sigma_{j_1}\sigma_{j_1}$
2 6
$\sigma_{j_1}\sigma_{j_1}$
3 5
$\sigma_{j_1}\sigma_{j_1}$
$+\sum_{j_1,j_2}$
6 3
1 4
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
$+\sum_{j_1,j_2}$
5 3
1 4
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
$+\sum_{j_1,j_2}$
2 4
1 6
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
3 6
$\sigma_{j_1}\sigma_{j_1}$
4 5
$\sigma_{j_1}\sigma_{j_1}$
4 6
$\sigma_{j_1}\sigma_{j_1}$
$+\sum_{j_1,j_2}$
2 4
1 5
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
$+\sum_{j_1,j_2}$
2 3
1 6
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
$+\sum_{j_1,j_2}$
2 3
1 5
$\widehat{\Delta G}_{j_2}$ $\widehat{\Delta G}_{j_2}$
1 3
$\sigma_{j_2}\sigma_{j_2}$
1 4
$\sigma_{j_2}\sigma_{j_2}$
$+\sum_{j_1,j_2,j_3}$
4 5
2 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
3 5
2 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
1 5
$\sigma_{j_2}\sigma_{j_2}$
1 6
$\sigma_{j_2}\sigma_{j_2}$
$+\sum_{j_1,j_2,j_3}$
6 3
2 4
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
5 3
2 4
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
2 3
$\sigma_{j_2}\sigma_{j_2}$
2 4
$\sigma_{j_2}\sigma_{j_2}$
$+\sum_{j_1,j_2,j_3}$
4 5
1 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
3 5
1 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
2 5
$\sigma_{j_2}\sigma_{j_2}$
2 6
$\sigma_{j_2}\sigma_{j_2}$
$+\sum_{j_1,j_2,j_3}$
6 3
1 4
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
5 3
1 4
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$


3 5
$\sigma_{j_2}\sigma_{j_2}$
2 4
1 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
3 6
$\sigma_{j_2}\sigma_{j_2}$
2 4
1 5
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
4 5
$\sigma_{j_2}\sigma_{j_2}$
2 3
1 6
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\sum_{j_1,j_2,j_3}$
4 6
$\sigma_{j_2}\sigma_{j_2}$
2 3
1 5
$\widehat{\Delta G}_{j_1}$ ϕ_{j_1} ϕ_{j_3} $\widehat{\Delta G}_{j_3}$
$+\frac{2}{n_\ell^5}$
$\sum_{j_1,j_2,j_3}$
2 5
$\sigma_{j_2}\sigma_{j_2}$
3 4
1 6
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$
$+\sum_{j_1,j_2,j_3}$
2 6
$\sigma_{j_2}\sigma_{j_2}$
3 4
1 5
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$
$+\sum_{j_1,j_2,j_3}$
2 5
$\sigma_{j_2}\sigma_{j_2}$
4 3
1 6
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$
$+\sum_{j_1,j_2,j_3}$
2 6
$\sigma_{j_2}\sigma_{j_2}$
4 3
1 5
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$
$+\sum_{j_1,j_2,j_3}$
2 3
$\sigma_{j_2}\sigma_{j_2}$
5 4
1 6
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$
$+\sum_{j_1,j_2,j_3}$
2 4
$\sigma_{j_2}\sigma_{j_2}$
5 3
1 6
$\sigma_{j_1}\sigma_{j_1}$ $\sigma_{j_3}\sigma_{j_3}$



(154)

Remarkably, these diagrams exactly reproduce the algebraic expression derived in the previous subsection at order $1/n^2$, see equation 131.

Moreover, equation 154 provides a direct validation of the leading-order (1/n) Feynman rules presented in Appendix F. Specifically, the first three lines of equation 154 follow from the Gaussian-like Feynman rules (1)-(4). The remaining contributions, namely the two groups of terms in parentheses multiplied by $-1/n^4$ and $2/n^5$, arise from rule (5): the former corresponds to $\lambda = (2,1)$ with $\beta_{(2,1)} = -1$, while the latter corresponds to $\lambda = (3)$ with $\beta_{(3)} = 2$. In both cases, the cubic vertices defined in equation 101 appear as disconnected components, as prescribed by rule (5).

I.3 Single-input case

In the single-input setting, the recursion relation for $V_6^{(\ell)}$ takes a simple and compact form. Explicitly, $V_6^{\mathcal{G}}$ is readily obtained as

$$\begin{aligned}\frac{1}{n_\ell^2}V_6^{\mathcal{G}(\ell+1)} &= \frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\left(\langle\sigma^6\rangle_K - 3\langle\sigma^4\rangle_K\langle\sigma^2\rangle_K + 2(\langle\sigma^2\rangle_K)^3\right)\\ &+ 6\frac{(C_W^{(\ell+1)})^2}{n_\ell n_{\ell-1}}V_4^{(\ell)}\chi_{||}^{(\ell)}\left[3\langle\sigma^2(\sigma')^2\rangle_K + \langle\sigma^3\sigma''\rangle_K - \langle(\sigma')^2\rangle_K\langle\sigma^2\rangle_K - \langle\sigma^2\rangle_K\langle\sigma\sigma''\rangle_K\right]\\ &+ \frac{3C_W^{(\ell+1)}}{2n_{\ell-1}^2}(\chi_{||}^{(\ell)})^2(V_4^{(\ell)})^2\left[3\langle(\sigma'')^2\rangle_K + 4\langle\sigma'\sigma'''\rangle_K + \langle\sigma\sigma^{''''}\rangle_K\right] + \frac{1}{n_{\ell-1}^2}(\chi_{||}^{(\ell)})^3V_6^{(\ell)}\end{aligned} \tag{155}$$

whereas ΔV_6 is given by

$$\begin{aligned}\frac{1}{n_\ell^2}\Delta V_6^{(\ell+1)} &= \frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\left[-12\langle\sigma^4\rangle_K\langle\sigma^2\rangle_K + 28(\langle\sigma^2\rangle_K)^3 - 12\left(\frac{n_\ell}{n_{\ell-1}}\right)\langle\sigma^2|\sigma^2\rangle_{V^{(\ell)}}\langle\sigma^2\rangle_K\right]\\ &= \frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\left[-12\langle\sigma^4\rangle_K\langle\sigma^2\rangle_K + 28(\langle\sigma^2\rangle_K)^3 - \frac{12}{(C_W^{(\ell+1)})^2}\left(\frac{n_\ell}{n_{\ell-1}}\right)V^{(\ell)}(\chi_{||}^{(\ell)})^2\langle\sigma^2\rangle_K\right]\end{aligned} \tag{156}$$

Therefore, the orthogonal six-point cumulant reads

$$\begin{aligned}\frac{1}{n_\ell^2}V_6^{(\ell+1)} &= \frac{(C_W^{(\ell+1)})^3}{n_\ell^2}\left(\langle\sigma^6\rangle_K - 15\langle\sigma^4\rangle_K\langle\sigma^2\rangle_K + 30(\langle\sigma^2\rangle_K)^3\right)\\ &+ 6\frac{(C_W^{(\ell+1)})^2}{n_\ell n_{\ell-1}}V_4^{(\ell)}\chi_{||}^{(\ell)}\left[3\langle\sigma^2(\sigma')^2\rangle_K + \langle\sigma^3\sigma''\rangle_K - 3\langle(\sigma')^2\rangle_K\langle\sigma^2\rangle_K - 3\langle\sigma^2\rangle_K\langle\sigma\sigma''\rangle_K\right]\\ &+ \frac{3C_W^{(\ell+1)}}{2n_{\ell-1}^2}(\chi_{||}^{(\ell)})^2(V_4^{(\ell)})^2\left[3\langle(\sigma'')^2\rangle_K + 4\langle\sigma'\sigma'''\rangle_K + \langle\sigma\sigma^{''''}\rangle_K\right] + \frac{1}{n_{\ell-1}^2}(\chi_{||}^{(\ell)})^3V_6^{(\ell)}\end{aligned} \tag{157}$$

We use *Mathematica* to solve equation 157 for a rectangular $\tanh$ MLP of width $n = 50$, depth $L = 30$ and input vector as in equation 121, both exactly and asymptotically; see Appendix J for details. The results for the normalized tensor $\tilde{V}_6^{(\ell)} = V_6^{(\ell)}/(K^{(\ell)})^3$ are shown in Figure 10.

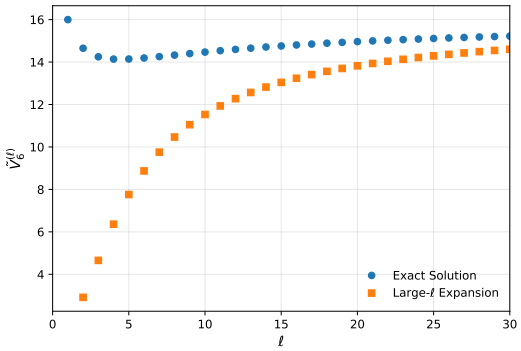


Figure 10: *The sextic vertex.* Solution to the single-input recursion relation equation 157 at order $1/n^2$. We consider a $\tanh$ network with width $n = 50$ and depth $L = 30$, with inputs drawn from $(0, 1)$. Blue points show the exact solution of equation 157, while orange boxes represent the large-ℓ expansion equation 194, evaluated at integer ℓ. The tensor magnitude remains below its Gaussian counterpart and exhibits early-layer saturation, consistent with the behavior of lower-rank orthogonal tensors.

J LARGE-ℓ EXPANSION

In this appendix, we solve the single-input recursion relations equation 110-equation 119 and equation 157 for $\tanh$ networks in the large-ℓ regime. We adopt a standard renormalization-inspired ansatz and expand a generic tensor $\mathcal{O}^{(\ell)}$ as

$$\mathcal{O}^{(\ell)} = \ell^{-p_{\mathcal{O}}} \sum_{i,j=0}^{k} c_{i,j}^{\mathcal{O}} \frac{\log^j(\ell)}{\ell^i} \tag{158}$$

where $p_{\mathcal{O}}$ is the critical exponent and k controls the truncation order. Throughout, we set $k = 5$.

The coefficients $c_{i,j}^{\mathcal{O}}$ in equation 158 depend on both the input data and the network parameters through the initial conditions

$$K^{(1)} = \frac{C_W}{n} \sum_{i=1}^{n} x_i \cdot x_i, \quad \Theta^{(1)} = \lambda_b^{(1)} + \frac{\lambda_W^{(1)}}{n} \sum_{i=1}^{n} x_i \cdot x_i, \quad V^{(1)} = -2\big(K^{(1)}\big)^2,$$
$$D^{(1)} = 0, \quad F^{(1)} = 0, \quad A^{(1)} = 0, \quad B^{(1)} = 0, \quad P^{(1)} = 0,$$
$$Q^{(1)} = 0, \quad R^{(1)} = 0, \quad S^{(1)} = 0, \quad T^{(1)} = 0, \quad U^{(1)} = 0. \tag{159}$$

In addition, they depend parametrically on the choice of activation function. We compute these coefficients using *Mathematica*, combining symbolic manipulation with numerical evaluation. We report below the critical exponents $p_{\mathcal{O}}$ and the corresponding coefficients for all tensors considered in this work.

We consider a square neural network with activation function $\tanh$, and fix $C_W = 1$, thereby focusing on the critical regime (Roberts & Yaida, 2022). As discussed in Roberts & Yaida (2022), for the $\tanh$ nonlinearity the training hyperparameters can be chosen as

$$\lambda_b^{(\ell)} = \frac{1}{\ell} \quad , \quad \lambda_W^{(\ell)} = 1 \tag{160}$$

Since the recursion relations equation 110-equation 119 connect quantities at consecutive layers, i.e. at depths ℓ and $\ell + 1$, the following large-ℓ asymptotic expansions are useful:

$$\frac{1}{(1+\ell)^a} = \frac{1}{\ell^a} \sum_{i=0}^{\infty} \binom{-a}{i} \left(\frac{1}{\ell}\right)^i \tag{161}$$

$$\log(1+\ell) = \log(\ell) - \sum_{i=0}^{\infty} \frac{1}{i} \left(-\frac{1}{\ell}\right)^i \tag{162}$$

The Gaussian expectations appearing in equation 110-equation 119 are evaluated by expanding the activation function as

$$\sigma(z) = \sum_{m=0}^{\infty} \sigma_m \frac{z^m}{m!} \tag{163}$$

where the coefficients $\sigma_m = \sigma^{(m)}(0)$ are determined by derivatives of σ at the origin. For $\sigma = \tanh$, these coefficients are known in closed form.

For the trucantion order $k = 5$ in equation 158, retaining a finite number of terms in the expansion equation 163 is sufficient to obtain a consistent large-ℓ expansion. In practice, we include terms up to $m = 20$, which ensures stability of all coefficients at the desired order.

Concretely, the non-vanishing coefficients are

$$\begin{aligned} \sigma_1 = 1 \quad , \quad \sigma_3 = -2 \quad , \quad \sigma_5 = 16 \quad , \quad \sigma_7 = -272 \quad , \quad \sigma_9 = 7936 \quad , \quad \sigma_{11} = -353792 \\ \sigma_{13} = 22368256 \quad , \quad \sigma_{15} = -1903757312 \\ \sigma_{17} = 209865342976 \quad , \quad \sigma_{19} = -29088885112832 \end{aligned} \tag{164}$$

with $\sigma_m = 0$ for m even.

J.1 The NNGP

To illustrate the procedure, we derive the expansion of the NNGP kernel $K^{(\ell)}$. Substituting the ansatz equation 158 into the recursion equation 5 yields

$$\begin{aligned} K^{(\ell)} = &\frac{1}{2\ell} + \frac{1}{\ell^2}\left(\frac{5\log\ell}{24} - \frac{5\log\ell_0}{24}\right) \\ &+ \frac{1}{\ell^3}\left(\frac{25(\log\ell)^2}{288} + \frac{5\log\ell\,(-5-10\log\ell_0)}{288} + \frac{53+25\log\ell_0+25(\log\ell_0)^2}{288}\right) \\ &+ \frac{1}{\ell^4}\left(\frac{125(\log\ell)^3}{3456} + \frac{25(\log\ell)^2\,(-25-30\log\ell_0)}{6912} + \frac{5\log\ell\left(23+\frac{125}{8}\log\ell_0+\frac{75}{8}(\log\ell_0)^2\right)}{432}\right. \\ &\qquad \left. + \frac{-8597-9200\log\ell_0-3125(\log\ell_0)^2-1250(\log\ell_0)^3}{34560}\right) \\ &+ \frac{1}{\ell^5}\left(\frac{625(\log\ell)^4}{41472} + \frac{125(\log\ell)^3\,(-65-60\log\ell_0)}{124416} + \frac{25(\log\ell)^2\left(287+\frac{650}{3}\log\ell_0+100(\log\ell_0)^2\right)}{27648}\right. \\ &\qquad + \frac{\log\ell\left(-10897-\frac{21525}{2}\log\ell_0-\frac{8125}{2}(\log\ell_0)^2-1250(\log\ell_0)^3\right)}{20736} \\ &\qquad \left. + \frac{-2479663+653820\log\ell_0+322875(\log\ell_0)^2+81250(\log\ell_0)^3+18750(\log\ell_0)^4}{1244160}\right). \end{aligned} \tag{165}$$

where we defined $c_{1,0}^K = -\frac{5}{24}\log(\ell_0)$. Consistency of the expansion fixes the critical exponent to $p_K = 1$, ensuring that the recursion equation 5 admits a solution within the ansatz. The first two lines of equation 165 reproduce the known large-ℓ expansion of the midpoint kernel up to order ℓ^{-3} (Roberts & Yaida, 2022). The scale ℓ_0 is fixed by the initial condition equation 159 after replacing $\ell = 1$ in equation 165.

J.2 The quartic vertex

Having obtained the large-ℓ expansion of $K^{(\ell)}$, we now compute the corresponding expansion of the quartic vertex $V_4^{(\ell)}$ from equation 7. We again adopt the ansatz equation 158 together with the initial conditions equation 159. Using the large-ℓ identities equation 161, equation 162, the $\tanh$ coefficients equation 164, and fixing the critical exponent to $p_V = 2$, we obtain

$$V_4^{(\ell)} = -\frac{1}{2\ell^2} + \frac{1}{\ell^3}\left(-\frac{5}{12}\log\ell + \frac{1}{12}\left(8+5\log\ell_0\right)\right)$$

$$+\frac{1}{\ell^4}\left(-\frac{25}{96}(\log\ell)^2+\frac{1}{144}\log\ell\,(-167+75\log\ell_0)+c^V_{2,0}\right). \tag{166}$$

Although the ansatz formally includes coefficients $c^V_{i,j}$ for $i,j=0,\dots,5$, the recursion equation 7 only constrains a subset of these coefficients within our truncation scheme. In particular, the Gaussian expectation values in the first two terms on the left-hand side of equation 7, which depend on $K^{(\ell)}$, contribute at most at order ℓ^{-5}. Since $p_V=2$, this implies that only coefficients with $i=0,1,2$ can be determined. The coefficient $c^V_{2,0}$ is fixed by the initial conditions equation 159 after replacing $\ell=1$ in equation 166.

J.3 THE NTK

We next determine the large-ℓ behavior of the frozen NTK $\Theta^{(\ell)}$ from equation 9. Substituting the expansion of $K^{(\ell)}$ from equation 165 and proceeding as in the previous cases, we find that the critical exponent is $p_\Theta=0$, yielding

$$\begin{aligned}
\Theta^{(\ell)} = \frac{3}{2}&+\frac{1}{\ell}\Big[-\frac{5}{24}(\log\ell)^2+\frac{1}{24}\log\ell(27+10\log\ell_0)+c^\Theta_{1,0}\Big]\\
&+\frac{1}{\ell^2}\Big[-\frac{25}{288}(\log\ell)^3+\frac{5}{288}(\log\ell)^2(31+15\log\ell_0)+\frac{447+205\log\ell_0}{288}\\
&\quad+\log\ell\Big(\frac{-313-175\log\ell_0-50(\log\ell_0)^2}{288}\Big)+\log\ell\Big(\frac{5}{12}c^\Theta_{1,0}\Big)\\
&\quad-\frac{1}{12}(4+5\log\ell_0)c^\Theta_{1,0}\Big]\\
&+\frac{1}{\ell^3}\Big[-\frac{125}{3456}(\log\ell)^4+\frac{125}{864}(\log\ell)^3(2+\log\ell_0)\\
&\quad+\frac{-183829-116850\log\ell_0-26500(\log\ell_0)^2}{69120}\\
&\quad+(\log\ell)^2\Big(-\frac{25(143+93\log\ell_0+25(\log\ell_0)^2)}{3456}\Big)+(\log\ell)^2\Big(\frac{25}{144}c^\Theta_{1,0}\Big)\\
&\quad+\frac{99+65\log\ell_0+25(\log\ell_0)^2}{144}\,c^\Theta_{1,0}\\
&\quad+\log\ell\Big(\frac{17031+10790\log\ell_0+2650(\log\ell_0)^2+500(\log\ell_0)^3}{6912}\Big)\\
&\quad+\log\ell\Big(-\frac{5}{144}(13+10\log\ell_0)\,c^\Theta_{1,0}\Big)\Big]\\
&+\frac{1}{\ell^4}\Big[-\frac{625}{41472}(\log\ell)^5+\frac{125(\log\ell)^4(103+50\log\ell_0)}{82944}\\
&\quad+\frac{287109682+282873525\log\ell_0+85689000(\log\ell_0)^2}{67184640}\\
&\quad+\frac{12240000(\log\ell_0)^3}{67184640}\\
&\quad+(\log\ell)^3\Big(-\frac{25(23863+16110\log\ell_0+4050(\log\ell_0)^2)}{746496}\Big)\\
&\quad+(\log\ell)^3\Big(\frac{125}{1728}c^\Theta_{1,0}\Big)\\
&\quad-\frac{21949+18100\log\ell_0+6125(\log\ell_0)^2}{17280}\,c^\Theta_{1,0}\\
&\quad-\frac{1250(\log\ell_0)^3}{17280}\,c^\Theta_{1,0}\\
&\quad+(\log\ell)^2\Big(\frac{2029301+1492050\log\ell_0+457875(\log\ell_0)^2}{746496}\Big)\\
&\quad+(\log\ell)^2\Big(\frac{78750(\log\ell_0)^3}{746496}-\frac{25(49+30\log\ell_0)}{3456}c^\Theta_{1,0}\Big)
\end{aligned}$$

$$+\log\ell\Big(\frac{-126292817-95368260\log\ell_0-30944250(\log\ell_0)^2}{22394880}\Big)$$
$$+\log\ell\Big(\frac{-5130000(\log\ell_0)^3-675000(\log\ell_0)^4}{22394880}\Big)$$
$$+\log\ell\Big(\frac{5(362+245\log\ell_0+75(\log\ell_0)^2)}{1728}c^{\Theta}_{1,0}\Big)\Big]. \tag{167}$$

As in the previous cases, the truncation order k together with the scaling $p_\Theta = 0$ restricts the set of coefficients $c^{\Theta}_{i,j}$ that can be determined within this approximation. The coefficient $c^{\Theta}_{1,0}$ is determined by the initial conditions equation 159 after replacing $\ell = 1$ in equation 167.

J.4 The NTK-preactivation mixed cumulant

We next determine the preactivation-NTK mixed cumulants $D^{(\ell)}$ and $F^{(\ell)}$. Applying the same procedure as above yields the critical exponents $p_D = 2$ and $p_F = 1$, together with the large-ℓ expansions

$$D^{(\ell)} = -\frac{4}{3\ell^2}+\frac{1}{\ell^3}\Bigg[\frac{5}{54}(\log\ell)^3+\frac{1}{36}(\log\ell)^2(51-10\log\ell_0)+c^{D}_{1,0}$$
$$+\frac{1}{144}\log\ell\left(-274-290\log\ell_0-75(\log\ell_0)^2-192c^{\Theta}_{1,0}-288c^{V}_{2,0}\right)\Bigg], \tag{168}$$

$$F^{(\ell)} = -\frac{1}{2\ell}+\frac{1}{\ell^2}\left[\frac{5}{48}(\log\ell)^2+\frac{1}{24}\log\ell\left(-21-5\log\ell_0\right)+\frac{1}{16}\left(27+5\log\ell_0-8c^{\Theta}_{1,0}\right)\right]$$
$$+\frac{1}{\ell^3}\Bigg[\frac{25}{288}(\log\ell)^3-\frac{25}{288}(\log\ell)^2(10+3\log\ell_0)$$
$$+\frac{1}{288}\log\ell\left(1061+365\log\ell_0+50(\log\ell_0)^2-120c^{\Theta}_{1,0}\right)$$
$$+\frac{1}{288}\left(-3321-575\log\ell_0-25(\log\ell_0)^2+432c^{\Theta}_{1,0}+120\log\ell_0\,c^{\Theta}_{1,0}\right)\Bigg]$$
$$+\frac{1}{\ell^4}\Bigg[\frac{625}{13824}(\log\ell)^4-\frac{5}{3456}(\log\ell)^3(284+125\log\ell_0)+c^{F}_{3,0}$$
$$+\frac{(\log\ell)^2}{13824}\left(23833+12990\log\ell_0+3000(\log\ell_0)^2-3600c^{\Theta}_{1,0}\right)$$
$$+\frac{\log\ell}{69120}\left(1107057-450550\log\ell_0-84000(\log\ell_0)^2-5000(\log\ell_0)^3\right.$$
$$\left.-188640c^{\Theta}_{1,0}+36000\log\ell_0\,c^{\Theta}_{1,0}\right)\Bigg]. \tag{169}$$

The coefficients $c^{D}_{1,0}$ and $c^{F}_{3,0}$ are fixed by the initial conditions equation 159 together with the previously determined parameters $\log\ell_0$, $c^{V}_{2,0}$, and $c^{\Theta}_{1,0}$.

J.5 The NTK-variance

We now turn to the NTK variance tensors $A^{(\ell)}$ and $B^{(\ell)}$. Using the expansion of $D^{(\ell)}$ in equation 168, we find the critical exponents $p_A = 1$ and $p_B = 1$ together with

$$A^{(\ell)} = \frac{16}{3\ell}+\frac{1}{\ell^2}\Bigg[-\frac{5}{54}(\log\ell)^4+\frac{1}{27}(\log\ell)^3(-61+10\log\ell_0)+c^{A}_{1,0}$$
$$+\frac{4}{9}\log\ell\left(-15+10\log\ell_0-9dd_{1,0}+12c^{\Theta}_{1,0}\right)$$

$$+\frac{1}{72}(\log\ell)^2\left(330+370\log\ell_0+75(\log\ell_0)^2+192c^{\Theta}_{1,0}+288c^{V}_{2,0}\right)\Bigg], \tag{170}$$

$$\begin{aligned}
B^{(\ell)} = \frac{9}{2\ell} &+ \frac{1}{\ell^2}\Bigg[-\frac{5}{12}(\log\ell)^3+\frac{1}{8}(\log\ell)^2\left(27+10\log\ell_0\right)+c^{B}_{1,0}\\
&\quad+\frac{3}{4}\log\ell\left(-21+8c^{\Theta}_{1,0}\right)\Bigg]\\
&+\frac{1}{\ell^3}\Bigg[-\frac{125}{288}(\log\ell)^4+\frac{5}{144}(\log\ell)^3\left(121+50\log\ell_0\right)\\
&\quad-\frac{1}{6}(4+5\log\ell_0)\,c^{B}_{1,0}\\
&\quad+\frac{1}{72}(\log\ell)^2\left(-1683-435\log\ell_0-100(\log\ell_0)^2+420c^{\Theta}_{1,0}\right)\\
&\quad+\frac{1}{288}\left(-18516-690\log\ell_0-125(\log\ell_0)^2+5664c^{\Theta}_{1,0}+240\log\ell_0\,c^{\Theta}_{1,0}-576(c^{\Theta}_{1,0})^2\right)\\
&\quad+\log\ell\left(\frac{5}{6}c^{B}_{1,0}+\frac{5043+3330\log\ell_0+50(\log\ell_0)^2-1344c^{\Theta}_{1,0}-960\log\ell_0\,c^{\Theta}_{1,0}}{144}\right)\Bigg]\\
&+\frac{1}{\ell^4}\Bigg[-\frac{125}{384}(\log\ell)^5+\frac{125(\log\ell)^4(211+90\log\ell_0)}{6912}\\
&\quad+\frac{214+170\log\ell_0+75(\log\ell_0)^2}{144}\,c^{B}_{1,0}\\
&\quad-\frac{5}{3456}(\log\ell)^3\left(17317+7310\log\ell_0+1650(\log\ell_0)^2-2880c^{\Theta}_{1,0}\right)\\
&\quad+\frac{15676647+5933100\log\ell_0+221500(\log\ell_0)^2+22500(\log\ell_0)^3}{69120}\\
&\quad-\frac{4405920+1816800\log\ell_0+60000(\log\ell_0)^2}{69120}\,c^{\Theta}_{1,0}\\
&\quad+\frac{339840+172800\log\ell_0}{69120}\,(c^{\Theta}_{1,0})^2\\
&\quad+(\log\ell)^2\left(\frac{25}{48}c^{B}_{1,0}+\frac{174843+113770\log\ell_0+16900(\log\ell_0)^2+2500(\log\ell_0)^3}{2304}\right.\\
&\qquad\left.-\frac{36000+21600\log\ell_0}{2304}\,c^{\Theta}_{1,0}\right)\\
&\quad+\log\ell\Bigg(-\frac{5}{72}(17+15\log\ell_0)\,c^{B}_{1,0}\\
&\qquad+\frac{-625380-280395\log\ell_0-72950(\log\ell_0)^2-1250(\log\ell_0)^3}{3456}\\
&\qquad+\frac{159888+64080\log\ell_0+18000(\log\ell_0)^2}{3456}\,c^{\Theta}_{1,0}-\frac{8640}{3456}\,(c^{\Theta}_{1,0})^2\Bigg)\Bigg].
\end{aligned} \tag{171}$$

The parameters $c^{A}_{1,0}$ and $c^{B}_{1,0}$ are again fixed by the initial conditions equation 159.

J.6 The dNTK-preactivation mixed cumulant

The asymptotic behavior of the preactivation-dNTK cumulant tensors $P^{(\ell)}$ and $Q^{(\ell)}$ follows from the same procedure. This yields the critical exponents $p_P = 0$ and $p_Q = 0$, together with the large-ℓ

expansions

$$
\begin{aligned}
P^{(\ell)} = & -\frac{3}{4}+\frac{1}{\ell}\Big[\frac{5}{16}(\log\ell)^2+\frac{1}{8}\log\ell(-16-5\log\ell_0)+\frac{1}{16}(31+5\log\ell_0-24c^{\Theta}_{1,0})\Big] \\
& +\frac{1}{\ell^2}\Big[-\frac{25}{576}(\log\ell)^4+\frac{5}{576}(\log\ell)^3(137+20\log\ell_0) \\
& \quad +\frac{(\log\ell)^2}{576}\big(-5748-1785\log\ell_0-100(\log\ell_0)^2+240c^{\Theta}_{1,0}\big) \\
& \quad +\frac{\log\ell}{576}\big(26199+5205\log\ell_0+350(\log\ell_0)^2\big) \\
& \quad +\frac{\log\ell}{576}\big(-5592c^{\Theta}_{1,0}-480\log\ell_0\,c^{\Theta}_{1,0}\big) \\
& \quad +\frac{1}{576}\big(-40257-6300\log\ell_0-125(\log\ell_0)^2-576c^{B}_{1,0}\big) \\
& \quad +\frac{1}{576}\big(9624c^{\Theta}_{1,0}+840\log\ell_0\,c^{\Theta}_{1,0}-576(c^{\Theta}_{1,0})^2\big)\Big] \\
& +\frac{1}{\ell^3}\Big[-\frac{55}{3456}(\log\ell)^5+\frac{5}{13824}(\log\ell)^4(2009+220\log\ell_0) \\
& \quad +\frac{(\log\ell)^3}{20736}\big(-128583-62430\log\ell_0-3700(\log\ell_0)^2-1920c^{\Theta}_{1,0}\big) \\
& \quad +\frac{(\log\ell)^2}{13824}\big(396027+182700\log\ell_0+50800(\log\ell_0)^2\big) \\
& \quad +\frac{(\log\ell)^2}{13824}\big(3000(\log\ell_0)^3-66192c^{\Theta}_{1,0}+3840\log\ell_0\,c^{\Theta}_{1,0}\big) \\
& \quad +\frac{\log\ell}{23040}\big(1355239-1599000\log\ell_0-260500(\log\ell_0)^2\big) \\
& \quad +\frac{\log\ell}{23040}\big(-10000(\log\ell_0)^3-19200c^{B}_{1,0}-516720c^{\Theta}_{1,0}\big) \\
& \quad +\frac{\log\ell}{23040}\big(269600\log\ell_0\,c^{\Theta}_{1,0}+24000(\log\ell_0)^2c^{\Theta}_{1,0}+72960(c^{\Theta}_{1,0})^2\big) \\
& \quad +c^{P}_{3,0}\Big]. \qquad (172)
\end{aligned}
$$

$$
\begin{aligned}
Q^{(\ell)} = & -\frac{17}{12}+\frac{1}{\ell}\Big[\frac{5}{8}(\log\ell)^2+\frac{1}{24}\log\ell(-101-30\log\ell_0)+\frac{1}{12}(93+10\log\ell_0-36c^{\Theta}_{1,0})\Big] \\
& +\frac{1}{\ell^2}\Big[-\frac{25}{288}(\log\ell)^4+\frac{5}{576}(\log\ell)^3(183+40\log\ell_0) \\
& \quad +\frac{(\log\ell)^2}{576}\big(-6013-2205\log\ell_0-200(\log\ell_0)^2+480c^{\Theta}_{1,0}\big) \\
& \quad +\frac{\log\ell}{576}\big(20897+7085\log\ell_0+750(\log\ell_0)^2\big) \\
& \quad +\frac{\log\ell}{576}\big(-4392c^{\Theta}_{1,0}-960\log\ell_0\,c^{\Theta}_{1,0}\big) \\
& \quad +\frac{1}{576}\big(-53509-8045\log\ell_0-150(\log\ell_0)^2\big) \\
& \quad +\frac{1}{576}\big(11424c^{\Theta}_{1,0}+1800\log\ell_0\,c^{\Theta}_{1,0}-1152(c^{\Theta}_{1,0})^2\big)\Big] \\
& +\frac{1}{\ell^3}\Big[-\frac{305}{3456}(\log\ell)^5+\frac{5}{13824}(\log\ell)^4(4273+1220\log\ell_0) \\
& \quad +\frac{(\log\ell)^3}{20736}\big(-209617-115500\log\ell_0-15950(\log\ell_0)^2+11280c^{\Theta}_{1,0}\big) \\
& \quad +\frac{(\log\ell)^2}{17280}\big(1546071+268150\log\ell_0+81125(\log\ell_0)^2\big) \\
& \quad +\frac{(\log\ell)^2}{17280}\big(9375(\log\ell_0)^3-227460c^{\Theta}_{1,0}-28200\log\ell_0\,c^{\Theta}_{1,0}\big)
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{\log \ell}{6912}\big(-697585 - 864585\log \ell_0 - 137400(\log \ell_0)^2\big) \\
&+ \frac{\log \ell}{6912}\big(-6250(\log \ell_0)^3 + 27648 c^F_{3,0} - 62712 c^\Theta_{1,0}\big) \\
&+ \frac{\log \ell}{6912}\big(129840\log \ell_0\, c^\Theta_{1,0} + 22800(\log \ell_0)^2 c^\Theta_{1,0} + 10368(c^\Theta_{1,0})^2\big) \\
&+ c^Q_{3,0}\Big]. \qquad (173)
\end{aligned}
$$

The coefficients $c^P_{3,0}$ and $c^Q_{3,0}$ follow from the initial conditions equation 159.

J.7 THE DD$_\text{I}$NTK MEAN

Applying this framework to the dd$_\text{I}$NTK mean tensor $R^{(\ell)}$ fixes the critical exponent to $p_R = -1$, yielding

$$
\begin{aligned}
R^{(\ell)} = \; & -0.43821 - 1.75\,\ell + 0.9375(\log \ell)^2 + \log \ell\,(-6.67172 - 2.5\log \ell_0) \\
&+ 0.520833\log \ell_0 - 6c^\Theta_{1,0} \\
&+ \frac{1}{\ell}\Big[-30.8931 - 0.173611(\log \ell)^4 - 39.3424\log \ell_0 - 2.08333(\log \ell_0)^2 \\
&\quad + (\log \ell)^3(4.17447 + 1.30208\log \ell_0) - 3c^B_{1,0} + 49.7536 c^\Theta_{1,0} + 7.5\log \ell_0\, c^\Theta_{1,0} - 9(c^\Theta_{1,0})^2 \\
&\quad + (\log \ell)^2\big(-27.3051 - 16.1153\log \ell_0 - 1.5625(\log \ell_0)^2 + 3.125 c^\Theta_{1,0}\big) \\
&\quad + \log \ell\big(75.8169 + 33.0744\log \ell_0 + 3.125(\log \ell_0)^2\big) \\
&\quad + \log \ell\big(-43.9268 c^\Theta_{1,0} - 7.5\log \ell_0\, c^\Theta_{1,0}\big)\Big] \\
&+ \frac{1}{\ell^2}\Big[(\log \ell)^5\big(0.256462 + 0.336123\log \ell_0 + 0.0434028(\log \ell_0)^2 - 0.0347222 c^\Theta_{1,0}\big) \\
&\quad + (\log \ell)^4\big(-0.898367 - 1.47402\log \ell_0 - 0.580512(\log \ell_0)^2 - 0.036169(\log \ell_0)^3 \\
&\qquad + 1.24491 c^\Theta_{1,0} + 0.260417\log \ell_0\, c^\Theta_{1,0}\big) \\
&\quad + (\log \ell)^3\big(-7.80677 - 6.49926\log \ell_0 - 1.31955(\log \ell_0)^2 + 0.138889 c^B_{1,0} \\
&\qquad - 7.24624 c^\Theta_{1,0} - 4.46181\log \ell_0\, c^\Theta_{1,0} - 0.347222(\log \ell_0)^2 c^\Theta_{1,0} + 0.416667(c^\Theta_{1,0})^2\big) \\
&\quad + (\log \ell)^2\big(-111.626 + 21.3942\log \ell_0 + 13.3665(\log \ell_0)^2 + 1.30208(\log \ell_0)^3 \\
&\qquad - 1.125 c^B_{1,0} - 0.416667\log \ell_0\, c^B_{1,0} - 3.46714 c^\Theta_{1,0} - 1.02641\log \ell_0\, c^\Theta_{1,0} \\
&\qquad - 9.375(c^\Theta_{1,0})^2 - 1.25\log \ell_0\,(c^\Theta_{1,0})^2\big) \\
&\quad + \log \ell\big(131.914 - 39.0856\log \ell_0 - 32.2968(\log \ell_0)^2 - 2.17014(\log \ell_0)^3\big) \\
&\quad + \log \ell\big(+c^B_{1,0} - 6.96453 c^\Theta_{1,0} + 32.6793\log \ell_0\, c^\Theta_{1,0} + 6.25(\log \ell_0)^2 c^\Theta_{1,0}\big) \\
&\quad + \log \ell\big(-2c^B_{1,0} c^\Theta_{1,0} + 11(c^\Theta_{1,0})^2 - 2(c^\Theta_{1,0})^3 - 2c^P_{3,0}\big) \\
&\quad + c^R_{3,0}\Big]. \qquad (174)
\end{aligned}
$$

The parameter $c^R_{3,0}$ is determined analogously from the initial conditions equation 159.

J.8 THE DD$_\text{II}$NTK MEAN

Applying the same analysis to the dd$_\text{II}$NTK mean tensors $S^{(\ell)}$, $T^{(\ell)}$ and $U^{(\ell)}$, yields the critical exponents $p_S = -1$, $p_T = -1$, and $p_U = 0$, together with

$$
\begin{aligned}
S^{(\ell)} = \; & \frac{3}{4}\,\ell - \frac{5}{16}(\log \ell)^2 + \frac{1}{16}\log \ell\,(27 + 10\log \ell_0) + \frac{3}{32}(45 + 16 c^\Theta_{1,0}) \\
&+ \frac{1}{\ell}\Big[\frac{25}{576}(\log \ell)^4 - \frac{5}{288}(\log \ell)^3(61 + 10\log \ell_0) \\
&\quad + \frac{1}{576}(\log \ell)^2\big(2763 + 1560\log \ell_0 + 100(\log \ell_0)^2 - 240 c^\Theta_{1,0}\big)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \frac{1}{576}\log\ell\big(-7137 + 360\log\ell_0 - 200(\log\ell_0)^2 + 5232c^{\Theta}_{1,0} + 480\log\ell_0\, c^{\Theta}_{1,0}\big)\\
&\quad + \frac{1}{576}\big(6 + 765\log\ell_0 + 50(\log\ell_0)^2 + 576c^{B}_{1,0} + 2400c^{\Theta}_{1,0}\\
&\qquad - 480\log\ell_0\, c^{\Theta}_{1,0} + 576(c^{\Theta}_{1,0})^2\big)\Big]\\
&+ \frac{1}{\ell^2}\Big[\frac{335}{3456}(\log\ell)^5 - \frac{5}{3456}(\log\ell)^4(1328 + 335\log\ell_0)\\
&\quad + \frac{5}{3456}(\log\ell)^3\big(8685 + 4286\log\ell_0 + 480(\log\ell_0)^2 - 912c^{\Theta}_{1,0}\big)\\
&\quad + \frac{1}{6912}(\log\ell)^2\big(-412047 - 75465\log\ell_0 - 14625(\log\ell_0)^2 - 1000(\log\ell_0)^3\\
&\qquad + 164520c^{\Theta}_{1,0} + 27360\log\ell_0\, c^{\Theta}_{1,0}\big)\\
&\quad + \frac{1}{13824}\log\ell\big(1165869 + 298080\log\ell_0 + 1200(\log\ell_0)^2 + 2500(\log\ell_0)^3\\
&\qquad + 52992c^{B}_{1,0} - 458208c^{\Theta}_{1,0} - 96480\log\ell_0\, c^{\Theta}_{1,0} - 9600(\log\ell_0)^2 c^{\Theta}_{1,0}\\
&\qquad + 108288(c^{\Theta}_{1,0})^2\big) + c^{S}_{3,0}\Big],
\end{aligned}
\tag{175}
$$

$$
\begin{aligned}
T^{(\ell)} = {}& -18.981 + 1.88889\,\ell - 1.84028(\log\ell)^2 + 0.196759\log\ell_0\\
&+ \log\ell\big(5.80387 + 1.18056\log\ell_0\big) + 2.83333\, c^{\Theta}_{1,0}\\
&+ \frac{1}{\ell}\Big[440.657 + 0.868056(\log\ell)^4 + (\log\ell)^3\big(-8.44101 - 1.04167\log\ell_0\big)\\
&\quad + 8.02013\log\ell_0 - 0.491898(\log\ell_0)^2 + (\log\ell)^2\big(46.0699 + 3.38594\log\ell_0 - 2.5\, c^{\Theta}_{1,0}\big)\\
&\quad - 32.6732\, c^{\Theta}_{1,0} + \log\ell\big(-158.347 - 13.6717\log\ell_0 + 8.12626\, c^{\Theta}_{1,0}\big)\Big]\\
&+ \frac{1}{\ell^2}\Big[(\log\ell)^5\big(-0.1989 - 0.0241419\log\ell_0 + 0.0694444\, c^{\Theta}_{1,0}\big)\\
&\quad + (\log\ell)^4\big(-2.50973 - 1.10439\log\ell_0 - 0.108507(\log\ell_0)^2 - 0.376245\, c^{\Theta}_{1,0}\big)\\
&\quad + (\log\ell)^3\big(-75.8647 + 8.00378\log\ell_0 + 0.962168(\log\ell_0)^2 + 1.63315\, c^{\Theta}_{1,0}\\
&\qquad - 0.347222\log\ell_0\, c^{\Theta}_{1,0}\big)\\
&\quad + (\log\ell)^2\big(-256.807 - 13.8107\log\ell_0 - 2.88587(\log\ell_0)^2 - 10.1288\, c^{\Theta}_{1,0}\\
&\qquad + 3.15131\log\ell_0\, c^{\Theta}_{1,0}\big)\\
&\quad + \log\ell\big(-297.85 + 308.089\log\ell_0 + 8.69265(\log\ell_0)^2 - 0.204958(\log\ell_0)^3\\
&\qquad + 229.472\, c^{\Theta}_{1,0} - 10.9002\log\ell_0\, c^{\Theta}_{1,0} + (c^{\Theta}_{1,0})^2 - 4\, c^{Q}_{3,0}\big)\\
&\quad + c^{T}_{3,0}\Big],
\end{aligned}
\tag{176}
$$

$$
\begin{aligned}
U^{(\ell)} = {}& \frac{27}{8} + \frac{1}{\ell}\Big[-\frac{45}{16}(\log\ell)^2 + \frac{9}{16}\log\ell\,(37 + 10\log\ell_0)\\
&\quad - \frac{9}{8}\big(47 + 5\log\ell_0 - 12c^{\Theta}_{1,0}\big)\Big]\\
&+ \frac{1}{\ell^2}\Big[\frac{5}{64}(\log\ell)^5 - \frac{5}{128}(\log\ell)^4(27 + 10\log\ell_0)\\
&\quad + \frac{1}{192}(\log\ell)^3\big(1119 + 540\log\ell_0 + 100(\log\ell_0)^2 - 240c^{\Theta}_{1,0}\big)\\
&\quad + \frac{3}{128}(\log\ell)^2\big(-837 - 260\log\ell_0 + 432c^{\Theta}_{1,0} + 160\log\ell_0\, c^{\Theta}_{1,0}\big)\\
&\quad - \frac{3}{64}\log\ell\big(-2733 + 405\log\ell_0 + 100(\log\ell_0)^2 + 1104c^{\Theta}_{1,0} - 192(c^{\Theta}_{1,0})^2\big)\\
&\quad + c^{U}_{2,0}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{\ell^3}\Big[-\frac{5}{2304}(\log\ell)^5(819+230\log\ell_0)\\
&\quad+\frac{5}{1152}(\log\ell)^4\big(1593+1135\log\ell_0+225(\log\ell_0)^2-300c^{\Theta}_{1,0}\big)\\
&\quad+\frac{1}{3456}(\log\ell)^3\big(-187047-81075\log\ell_0-14850(\log\ell_0)^2-2000(\log\ell_0)^3\\
&\qquad+43560c^{\Theta}_{1,0}+18000\log\ell_0\,c^{\Theta}_{1,0}\big)\\
&\quad+\frac{1}{2304}(\log\ell)^2\big(460782+110370\log\ell_0+21525(\log\ell_0)^2+500(\log\ell_0)^3\\
&\qquad-173088c^{\Theta}_{1,0}-41760\log\ell_0\,c^{\Theta}_{1,0}-9600(\log\ell_0)^2c^{\Theta}_{1,0}+20160(c^{\Theta}_{1,0})^2\big)\\
&\quad+\frac{1}{46080}\big(30188151-884850\log\ell_0+309000(\log\ell_0)^2+65000(\log\ell_0)^3\\
&\qquad-12139200c^{\Theta}_{1,0}-302400\log\ell_0\,c^{\Theta}_{1,0}-60000(\log\ell_0)^2c^{\Theta}_{1,0}+1497600(c^{\Theta}_{1,0})^2\\
&\qquad+57600\log\ell_0\,(c^{\Theta}_{1,0})^2-92160(c^{\Theta}_{1,0})^3\big)\\
&\quad+\log\ell\Big(\frac{-1946727-1035240\log\ell_0+48450(\log\ell_0)^2+15500(\log\ell_0)^3}{4608}\\
&\qquad+\frac{526176+348480\log\ell_0+4800(\log\ell_0)^2}{4608}\,c^{\Theta}_{1,0}-\frac{64512+46080\log\ell_0}{4608}\,(c^{\Theta}_{1,0})^2\\
&\qquad+\frac{5}{6}c^{U}_{2,0}\Big)-\frac{1}{6}(4+5\log\ell_0)\,c^{U}_{2,0}\Big].
\end{aligned}
\tag{177}
$$

The coefficients $c^{S}_{3,0}$, $c^{T}_{3,0}$, and $c^{U}_{2,0}$ are fixed by the initial conditions equation 159.

J.9 The sextic vertex

Finally, for the sextic vertex $V_6^{(\ell)}$, we obtain the critical exponent $p_{V_6}=3$, and the large-ℓ expansion

$$
V_6^{(\ell)}=\frac{2}{\ell^3}+\frac{1}{\ell^4}\left(-0.573233+2.5\log\ell-1.25\log\ell_0\right). \tag{178}
$$

Within our approximation scheme, $V_6^{(\ell)}$ is entirely determined by the lower-rank tensors $V_4^{(\ell)}$ and $K^{(\ell)}$.

J.10 Application

We apply the above results to a square neural network of width $n=50$, using the same input vector as in Section H. This choice fixes the initial conditions

$$
K^{(1)}=0.238565\quad,\quad \Theta^{(1)}=1.23857\quad,\quad V_4^{(1)}=-0.113827 \tag{179}
$$

These determine the scale ℓ_0 via $\log(\ell_0)=-2.74141$, as well as the coefficients $c^{\mathcal{O}}$ in equation 165-equation 177:

$$
\begin{aligned}
&c^{V}_{2,0}=0.861763, && c^{\Theta}_{1,0}=0.829493, && c^{D}_{1,0}=\frac{4}{3}, && c^{F}_{3,0}=6.49751,\\
&c^{A}_{1,0}=-\frac{16}{3}, && c^{B}_{1,0}=7.19961, && c^{P}_{3,0}=39.7952, && c^{Q}_{3,0}=47.0357,\\
&c^{R}_{3,0}=-49.1347, && c^{S}_{3,0}=-16.4737, && c^{T}_{3,0}=-372.59, && c^{U}_{2,0}=-196.103
\end{aligned}
\tag{180}
$$

Substituting these values into equation 165–equation 177 yields the corresponding large-ℓ expansions. We now present these expressions explicitly. The NNGP takes the form

$$
\begin{aligned}
K^{(\ell)}=&\frac{1}{2\ell}+\frac{1}{\ell^2}\left(0.571128+\frac{5}{24}\log\ell\right)\\
&+\frac{1}{\ell^3}\left(0.598432+0.389134\log\ell+\frac{25}{288}(\log\ell)^2\right)\\
&+\frac{1}{\ell^4}\left(0.54664+0.585901\log\ell\right.
\end{aligned}
$$

Table 1: Coefficients of the large-ℓ expansion for the tensor $K^{(\ell)}$.

p_K	$c^K_{i,j}$						
1	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{1}{2}$	0	0	0	0	0
	1	0.571128	$\frac{5}{24}$	0	0	0	0
	2	0.598432	0.389134	$\frac{25}{288}$	0	0	0
	3	0.54664	0.585901	0.20704	$\frac{125}{3456}$	0	0
	4	-1.97764	0.666942	0.401984	0.0999518	$\frac{625}{41472}$	0

Table 2: Coefficients of the large-ℓ expansion for the tensor $V^{(\ell)}$.

p_V	$c^V_{i,j}$						
2	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{1}{2}$	0	0	0	0	0
	1	-0.475589	$-\frac{5}{12}$	0	0	0	0
	2	0.861763	-2.58754	$-\frac{25}{96}$	0	0	0

$$
\begin{aligned}
&\quad +0.20704(\log\ell)^2 + \frac{125}{3456}(\log\ell)^3\Big) \\
&+ \frac{1}{\ell^5}\,(-1.97764 + 0.666942\log\ell \\
&\quad +0.401984(\log\ell)^2 + 0.0999518(\log\ell)^3 + \frac{625}{41472}(\log\ell)^4\Big)\,. \qquad (181)
\end{aligned}
$$

For completeness, the parameters p_K and $c^K_{i,j}$ are summarized in Table 1. The quartic vertex reads

$$
\begin{aligned}
V_4^{(\ell)} = &-\frac{1}{2\ell^2} + \frac{1}{\ell^3}\left(-0.475589 - \frac{5}{12}\log\ell\right) \\
&+ \frac{1}{\ell^4}\left(0.861763 - 2.58754\log\ell - \frac{25}{96}(\log\ell)^2\right), \qquad (182)
\end{aligned}
$$

with coefficients given in Table 2. The NTK takes the form

$$
\begin{aligned}
\Theta^{(\ell)} = &\frac{3}{2} + \frac{1}{\ell}\left(0.829493 - 0.0172557\log\ell - \frac{5}{24}(\log\ell)^2\right) \\
&+ \frac{1}{\ell^2}\left(0.271725 - 0.380142\log\ell - 0.175715(\log\ell)^2 - \frac{25}{288}(\log\ell)^3\right) \\
&+ \frac{1}{\ell^3}\,(-0.280312 - 0.00940075\log\ell \\
&\quad -0.405269(\log\ell)^2 - 0.107265(\log\ell)^3 - \frac{125}{3456}(\log\ell)^4\Big) \\
&+ \frac{1}{\ell^4}\,(-1.08234 - 0.722724\log\ell \\
&\quad -0.125245(\log\ell)^2 - 0.279448(\log\ell)^3 - 0.0513459(\log\ell)^4 - \frac{625}{41472}(\log\ell)^5\Big)\,, \qquad (183)
\end{aligned}
$$

with coefficients summarized in Table 3.

The NTK-preactivation mixed cumulants are

$$
D^{(\ell)} = -\frac{4}{3\ell^2} + \frac{1}{\ell^3}\left(\frac{4}{3} - 3.12564\log\ell + 2.17817(\log\ell)^2 + \frac{5}{54}(\log\ell)^3\right), \qquad (184)
$$

Table 3: Coefficients of the large-ℓ expansion for the tensor $\Theta^{(\ell)}$.

p_Θ	$c^\Theta_{i,j}$						
0	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{3}{2}$	0	0	0	0	0
	1	0.829493	-0.0172557	$-\frac{5}{24}$	0	0	0
	2	0.271725	-0.380142	-0.175715	$-\frac{25}{288}$	0	0
	3	-0.280312	-0.00940075	-0.405269	-0.107265	$-\frac{125}{3456}$	0
	4	-1.08234	-0.722724	-0.125245	-0.279448	-0.0513459	$-\frac{625}{41472}$

Table 4: Coefficients of the large-ℓ expansion for the tensor $D^{(\ell)}$.

p_D	$c^D_{i,j}$						
2	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{4}{3}$	0	0	0	0	0
	1	$\frac{4}{3}$	-3.12564	2.17817	$\frac{5}{54}$	0	0

$$
\begin{aligned}
F^{(\ell)} = &-\frac{1}{2\ell} + \frac{1}{\ell^2}\left(0.416062 - 0.303872\log\ell + \frac{5}{48}(\log\ell)^2\right) \\
&+ \frac{1}{\ell^3}\left(-6.41357 + 1.16879\log\ell\right. \\
&\qquad \left. -0.154146(\log\ell)^2 + \frac{25}{288}(\log\ell)^3\right) \\
&+ \frac{1}{\ell^4}\left(6.49751 + 22.7949\log\ell\right. \\
&\qquad \left. +0.562927(\log\ell)^2 + 0.0848911(\log\ell)^3 + \frac{625}{13824}(\log\ell)^4\right), \qquad (185)
\end{aligned}
$$

with coefficients listed in Tables 4 and 5.

The NTK variance tensors are

$$
A^{(\ell)} = \frac{16}{3\ell} + \frac{1}{\ell^2}\left(-\frac{16}{3} - 19.7601\log\ell + 3.98303(\log\ell)^2 - 3.2746(\log\ell)^3 - \frac{5}{54}(\log\ell)^4\right), \qquad (186)
$$

$$
\begin{aligned}
B^{(\ell)} = &\frac{9}{2\ell} + \frac{1}{\ell^2}\left(7.19961 - 10.773\log\ell - 0.0517672(\log\ell)^2 - \frac{5}{12}(\log\ell)^3\right) \\
&+ \frac{1}{\ell^3}\left(-36.2955 - 12.3472\log\ell\right. \\
&\qquad \left. -12.4116(\log\ell)^2 - 0.55801(\log\ell)^3 - \frac{125}{288}(\log\ell)^4\right) \\
&+ \frac{1}{\ell^4}\left(24.5959 - 70.6998\log\ell\right. \\
&\qquad \left. -14.6047(\log\ell)^2 - 10.5449(\log\ell)^3 - 0.646109(\log\ell)^4 - \frac{125}{384}(\log\ell)^5\right), \qquad (187)
\end{aligned}
$$

with coefficients reported in Tables 6 and 7.

The dNTK tensors read

$$
\begin{aligned}
P^{(\ell)} = &-\frac{3}{4} + \frac{1}{\ell}\left(-0.163431 - 0.286616\log\ell + \frac{5}{16}(\log\ell)^2\right) \\
&+ \frac{1}{\ell^2}\left(-38.8818 + 19.1203\log\ell\right.
\end{aligned}
$$

Table 5: Coefficients of the large-ℓ expansion for the tensor $F^{(\ell)}$.

p_F	$c^F_{i,j}$						
1	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{1}{2}$	0	0	0	0	0
	1	0.416062	−0.303872	$\frac{5}{48}$	0	0	0
	2	−6.41357	1.16879	−0.154146	$\frac{25}{288}$	0	0
	3	6.49751	22.7949	0.562927	0.0848911	$\frac{625}{13824}$	0

Table 6: Coefficients of the large-ℓ expansion for the tensor $A^{(\ell)}$.

p_A	$c^A_{i,j}$						
1	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{16}{3}$	0	0	0	0	0
	1	$-\frac{16}{3}$	−19.7601	3.98303	−3.2746	$-\frac{5}{54}$	0

$$
\begin{aligned}
&\quad -2.44277(\log\ell)^2 + 0.713296(\log\ell)^3 - \frac{25}{576}(\log\ell)^4\Big) \\
&+ \frac{1}{\ell^3}\,(39.7952 + 130.51\log\ell \\
&\quad +10.9595(\log\ell)^2 + 0.63484(\log\ell)^3 + 0.508496(\log\ell)^4 - \frac{55}{3456}(\log\ell)^5\Big)\,,
\end{aligned} \tag{188}
$$

$$
\begin{aligned}
Q^{(\ell)} = &-\frac{17}{12} + \frac{1}{\ell}\left(2.97701 - 0.781566\log\ell + \frac{5}{8}(\log\ell)^2\right) \\
&+ \frac{1}{\ell^2}\,(-48.596 + 9.80987\log\ell \\
&\quad -1.86301(\log\ell)^2 + 0.636662(\log\ell)^3 - \frac{25}{288}(\log\ell)^4\Big) \\
&+ \frac{1}{\ell^3}\,(47.9186 + 108.564\log\ell \\
&\quad +63.8277(\log\ell)^2 - 0.168637(\log\ell)^3 + 0.33582(\log\ell)^4 - \frac{305}{3456}(\log\ell)^5\Big)\,,
\end{aligned} \tag{189}
$$

with coefficients given in Tables 8 and 9.

The $\mathrm{dd_I NTK}$ tensor is

$$
\begin{aligned}
R^{(\ell)} = &-6.84299 - 1.75\,\ell + 0.181818\log\ell + 0.9375(\log\ell)^2 \\
&+ \frac{1}{\ell}\,(57.7277 - 10.7504\log\ell \\
&\quad +7.72305(\log\ell)^2 + 0.604921(\log\ell)^3 - 0.173611(\log\ell)^4\big) \\
&+ \frac{1}{\ell^2}\,(-49.1347 - 77.9822\log\ell \\
&\quad -101.16(\log\ell)^2 + 3.35098(\log\ell)^3 - 0.034582(\log\ell)^4 - 0.367605(\log\ell)^5\big)\,,
\end{aligned} \tag{190}
$$

with coefficients listed in Table 10. The $\mathrm{dd_{II} NTK}$ tensors are

$$
\begin{aligned}
S^{(\ell)} = &\,5.46299 + \frac{3}{4}\,\ell - 0.0258836\log\ell - \frac{5}{16}(\log\ell)^2 \\
&+ \frac{1}{\ell}\,(10.2607 - 11.0739\log\ell
\end{aligned}
$$

Table 7: Coefficients of the large-ℓ expansion for the tensor $B^{(\ell)}$.

p_B	$c^B_{i,j}$						
1	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{9}{2}$	0	0	0	0	0
	1	7.19961	-10.773	-0.0517672	$-\frac{5}{12}$	0	0
	2	-36.2955	-12.3472	-12.4116	-0.55801	$-\frac{125}{288}$	0
	3	24.5959	-70.6998	-14.6047	-10.5449	-0.646109	$-\frac{125}{384}$

Table 8: Coefficients of the large-ℓ expansion for the tensor $P^{(\ell)}$.

p_P	$c^P_{i,j}$						
0	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{3}{4}$	0	0	0	0	0
	1	-0.163431	-0.286616	$\frac{5}{16}$	0	0	0
	2	-38.8818	19.1203	-2.44277	0.713296	$-\frac{25}{576}$	0
	3	39.7952	130.51	10.9595	0.63484	0.508496	$-\frac{55}{3456}$

$$
\begin{aligned}
&\quad -1.66866(\log\ell)^2 - 0.583088(\log\ell)^3 + \frac{25}{576}(\log\ell)^4\Big) \\
&+ \frac{1}{\ell^2}\,(-16.4737 + 39.1868\log\ell \\
&\quad -31.861(\log\ell)^2 - 0.309359(\log\ell)^3 - 0.592631(\log\ell)^4 + \frac{335}{3456}(\log\ell)^5\Big)\,,
\end{aligned}
\tag{191}
$$

$$
\begin{aligned}
T^{(\ell)} &= -17.1701 + 1.88889\,\ell + 2.56748\log\ell - 1.84028(\log\ell)^2 \\
&+ \frac{1}{\ell}\,(387.871 - 114.126\log\ell \\
&\quad +34.7139(\log\ell)^2 - 5.58537(\log\ell)^3 + 0.868056(\log\ell)^4) \\
&+ \frac{1}{\ell^2}\,(-372.59 - 1045.22\log\ell \\
&\quad -256.202(\log\ell)^2 - 88.4311(\log\ell)^3 - 0.609713(\log\ell)^4 - 0.0751136(\log\ell)^5)\,,
\end{aligned}
\tag{192}
$$

$$
\begin{aligned}
U^{(\ell)} &= \frac{27}{8} + \frac{1}{\ell}\left(-26.2564 + 5.39205\log\ell - \frac{45}{16}(\log\ell)^2\right) \\
&+ \frac{1}{\ell^2}\,(-196.103 + 108.191\log\ell \\
&\quad -3.04052(\log\ell)^2 + 0.995277(\log\ell)^3 + 0.0161773(\log\ell)^4 + \frac{5}{64}(\log\ell)^5\Big) \\
&+ \frac{1}{\ell^3}\,(218.985 - 21.8079\log\ell \\
&\quad +93.3553(\log\ell)^2 - 11.5693(\log\ell)^3 - 0.331592(\log\ell)^4 - 0.409017(\log\ell)^5)\,,
\end{aligned}
\tag{193}
$$

with coefficients given in Tables 11-13.

Finally, the sextic vertex reads

$$
V_6^{(\ell)} = \frac{2}{\ell^3} + \frac{1}{\ell^4}\,(2.85353 + 2.5\log\ell)\,. \tag{194}
$$

with coefficients reported in Table 14.

Table 9: Coefficients of the large-ℓ expansion for the tensor $Q^{(\ell)}$.

p_Q	$c^Q_{i,j}$						
0	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{17}{12}$	0	0	0	0	0
	1	2.97701	−0.781566	$\frac{5}{8}$	0	0	0
	2	−48.596	9.80987	−1.86301	0.636662	$-\frac{25}{288}$	0
	3	47.9186	108.564	63.8277	−0.168637	0.33582	$-\frac{305}{3456}$

Table 10: Coefficients of the large-ℓ expansion for the tensor $R^{(\ell)}$.

p_R	$c^R_{i,j}$						
−1	$i\backslash j$	0	1	2	3	4	5
	0	$-\frac{7}{4}$	0	0	0	0	0
	1	−6.84299	0.181818	0.9375	0	0	0
	2	57.7277	−10.7504	7.72305	0.604921	−0.173611	0
	3	−49.1347	−77.9822	−101.16	3.35098	−0.034582	−0.367605

K Experimental setup

K.1 Stability analysis

In this section, we provide details of the experiments in Section 5 and present additional results validating the stability at criticality of the NNGP K, the four-point vertex V_4, and the NTK tensors D, F, A, and B as the network depth ℓ increases.

We estimate the preactivation and NTK tensors via Monte Carlo sampling. For this purpose, individual networks are initialized in JAX using `neural-tangents stax` layers. Computing NTK-related quantities requires the Jacobian $\frac{\partial z_i^{(\ell)}(x_\alpha)}{\partial \theta_\mu}$, which we obtain via JAX's automatic differentiation. Since forming the full Jacobian is computationally and memory intensive for wide networks, we instead compute layer-wise statistics sequentially and propagate them to deeper layers using the chain rule. Sampling across initializations is parallelized with `vmap`. Orthogonal weight initialization is implemented using JAX's `orthogonal` initializer. As tanh is not directly available in this setup, we approximate it using `ElementwiseNumerical` with degree 80.

All experiments are conducted on tanh MLPs without biases. Inputs are drawn with i.i.d. standard normal components, resulting in

$$x_0 = \begin{pmatrix} 0.934738 \\ 0.26696 \\ 0.784097 \\ 0.656448 \\ 0.305308 \\ 0.401958 \\ 0.894594 \\ 0.0559893 \\ 0.000643274 \\ 0.0274513 \end{pmatrix} \oplus \begin{pmatrix} 0.377754 \\ 0.127474 \\ 0.879907 \\ 0.710555 \\ 0.509949 \\ 0.312682 \\ 0.0854376 \\ 0.869372 \\ 0.114232 \\ 0.0851646 \end{pmatrix} \oplus \begin{pmatrix} 0.254697 \\ 0.560475 \\ 0.508664 \\ 0.0271565 \\ 0.426426 \\ 0.457646 \\ 0.913778 \\ 0.40436 \\ 0.407187 \\ 0.0644401 \end{pmatrix} \oplus \begin{pmatrix} 0.256718 \\ 0.869761 \\ 0.0406222 \\ 0.431362 \\ 0.906228 \\ 0.55979 \\ 0.275852 \\ 0.553722 \\ 0.235762 \\ 0.751627 \end{pmatrix} \oplus \begin{pmatrix} 0.178558 \\ 0.411167 \\ 0.100846 \\ 0.220264 \\ 0.215917 \\ 0.490943 \\ 0.596323 \\ 0.0799147 \\ 0.205998 \\ 0.0372218 \end{pmatrix}, \tag{195}$$

Table 11: Coefficients of the large-ℓ expansion for the tensor $S^{(\ell)}$.

p_S	$c^S_{i,j}$						
-1	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{3}{4}$	0	0	0	0	0
	1	5.46299	-0.0258836	$-\frac{5}{16}$	0	0	0
	2	10.2607	-11.0739	-1.66866	-0.583088	$\frac{25}{576}$	0
	3	-16.4737	39.1868	-31.861	-0.309359	-0.592631	$\frac{335}{3456}$

Table 12: Coefficients of the large-ℓ expansion for the tensor $T^{(\ell)}$.

p_T	$c^T_{i,j}$						
-1	$i\backslash j$	0	1	2	3	4	5
	0	1.88889	0	0	0	0	0
	1	-17.1701	2.56748	-1.84028	0	0	0
	2	387.871	-114.126	34.7139	-5.58537	0.868056	0
	3	-372.59	-1045.22	-256.202	-88.4311	-0.609713	-0.0751136

$$x_1 = \begin{pmatrix}0.986304\\0.396331\\0.829442\\0.163461\\0.0583318\\0.0385021\\0.885504\\0.160447\\0.583711\\0.389812\end{pmatrix} \oplus \begin{pmatrix}0.193679\\0.632048\\0.2954\\0.12661\\0.440952\\0.949114\\0.824362\\0.373311\\0.386399\\0.101208\end{pmatrix} \oplus \begin{pmatrix}0.815355\\0.475999\\0.493653\\0.267819\\0.0133166\\0.814708\\0.315126\\0.47199\\0.992467\\0.570161\end{pmatrix} \oplus \begin{pmatrix}0.23285\\0.183045\\0.92565\\0.199642\\0.38384\\0.184987\\0.518954\\0.078869\\0.456603\\0.333712\end{pmatrix} \oplus \begin{pmatrix}0.752504\\0.0212271\\0.805301\\0.696779\\0.0208987\\0.71065\\0.335234\\0.908037\\0.94978\\0.318431\end{pmatrix}, \tag{196}$$

$$x_2 = \begin{pmatrix}0.268954\\0.486888\\0.22212\\0.915653\\0.795563\\0.797374\\0.826671\\0.11106\\0.853911\\0.298177\end{pmatrix} \oplus \begin{pmatrix}0.0628893\\0.679924\\0.792366\\0.986086\\0.936489\\0.273049\\0.360604\\0.921974\\0.820319\\0.53683\end{pmatrix} \oplus \begin{pmatrix}0.631468\\0.779267\\0.0763499\\0.669761\\0.155254\\0.343043\\0.907627\\0.0647726\\0.0988953\\0.761094\end{pmatrix} \oplus \begin{pmatrix}0.901974\\0.196523\\0.939642\\0.794076\\0.225599\\0.66191\\0.912069\\0.0218007\\0.303759\\0.0650933\end{pmatrix} \oplus \begin{pmatrix}0.345452\\0.888144\\0.360429\\0.207685\\0.600649\\0.377979\\0.827548\\0.871934\\0.364661\\0.384977\end{pmatrix}, \tag{197}$$

$$x_3 = \begin{pmatrix}0.0682295\\0.140096\\0.340508\\0.359852\\0.613812\\0.19288\\0.134531\\0.49299\\0.591431\\0.893789\end{pmatrix} \oplus \begin{pmatrix}0.773795\\0.604593\\0.842464\\0.882483\\0.93913\\0.872499\\0.317102\\0.289739\\0.0178811\\0.160385\end{pmatrix} \oplus \begin{pmatrix}0.814678\\0.0561101\\0.193425\\0.413272\\0.622813\\0.668977\\0.771141\\0.884678\\0.516841\\0.54484\end{pmatrix} \oplus \begin{pmatrix}0.54987\\0.101687\\0.224264\\0.370917\\0.731496\\0.234507\\0.906523\\0.431445\\0.115241\\0.275449\end{pmatrix} \oplus \begin{pmatrix}0.810621\\0.0049892\\0.133457\\0.778419\\0.169455\\0.0719057\\0.1728\\0.906299\\0.322143\\0.847235\end{pmatrix}. \tag{198}$$

Table 13: Coefficients of the large-ℓ expansion for the tensor $U^{(\ell)}$.

p_U	$c^U_{i,j}$						
0	$i\backslash j$	0	1	2	3	4	5
	0	$\frac{27}{8}$	0	0	0	0	0
	1	-26.2564	5.39205	$-\frac{45}{16}$	0	0	0
	2	-196.103	108.191	-3.04052	0.995277	0.0161773	$\frac{5}{64}$
	3	218.985	-21.8079	93.3553	-11.5693	-0.331592	-0.409017

Table 14: Coefficients of the large-ℓ expansion for the tensor $V_6^{(\ell)}$.

p_{V_6}	$c^{V_6}_{i,j}$						
3	$i\backslash j$	0	1	2	3	4	5
	0	2	0	0	0	0	0
	1	2.85353	2.5	0	0	0	0

For the two-dimensional tensors K and Θ, computations are restricted to the input pair (x_0, x_1).

Monte Carlo estimation of the kernels. We consider an ensemble of N_{net} neural networks with orthogonal weight initialization. The empirical NNGP and NTK are estimated as

$$\overline{K}^{(\ell)}_{\alpha\beta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} z^{(\ell)}_{I;i,\alpha} z^{(\ell)}_{I;i,\beta} \,, \tag{199}$$

$$\overline{\Theta}^{(\ell)}_{\alpha\beta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \left(\sum_{\mu} \frac{\partial z^{(\ell)}_{I;i,\alpha}}{\partial \theta_\mu} \frac{\partial z^{(\ell)}_{I;i,\beta}}{\partial \theta_\mu} \right) , \tag{200}$$

where i denotes a fixed channel and I indexes network initializations. Error bars indicate standard errors, computed as the sample standard deviation divided by $\sqrt{N_{\text{net}}}$.

Variance stability. Figure 11 demonstrates variance stability under the same setup used for the gradient analysis in Figure 1. At the critical value $C_W = 1$, both the orthogonal NNGP and NTK remain stable. While variance stability has been studied numerically for Gaussian initializations in Banta et al. (2024), our results show that the corresponding criticality conditions transfer directly to the orthogonal setting.

Furthermore, at criticality, the empirical NNGP remains stable beyond the perturbative regime $\ell < n$. In Figure 12 (a), we compare its diagonal component with both the single-input exact solution (Section H.1) and the large-ℓ expansion (Section J). The Monte Carlo estimate agrees quantitatively

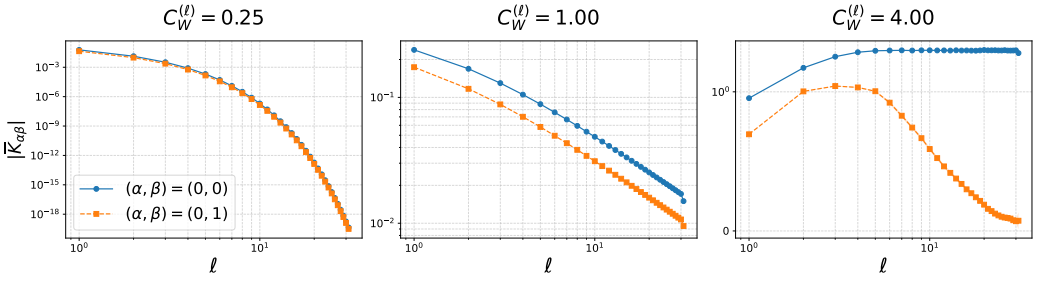


Figure 11: *Variance stability.* Components of the Monte Carlo estimate of the NNGP $\overline{K}_{\alpha\beta}$ for a $\tanh$ MLP, shown as a function of layer depth ℓ for identical and distinct inputs, across three values of $C_W^{(\ell)}$. Hidden layers have width 50. Means are computed over 600 initializations for both non-critical (left, right) and critical (middle) cases. Error bars are shown in all panels (see text).

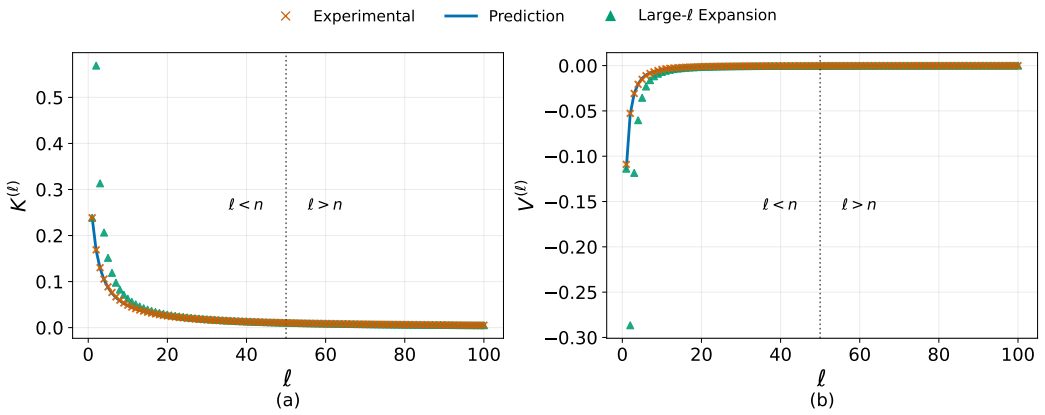


Figure 12: *Stability beyond the perturbative regime.* Comparison of the diagonal components of the Monte Carlo estimate, single-input exact solution, and large-ℓ expansion for the NNGP K and quartic vertex V in a tanh MLP with orthogonal initialization. Hidden layers have width 50; means are computed over 600 initializations. (a) The NNGP estimates are in quantitative agreement with the exact solution at both small and large depths. The large-ℓ expansion is inaccurate at small ℓ, as expected, but becomes accurate after a few layers. Stability persists up to $\ell = 100$, well beyond the perturbative regime $\ell < n$. (b) The quartic vertex V exhibits analogous scaling behavior.

with the exact solution across all depths, and matches the large-ℓ expansion at large ℓ. This confirms the stability of the orthogonal NNGP in the regime $\ell > n$, consistent with the observations of Day et al. (2025).

Monte–Carlo estimation of the four-point cumulant V_4. We estimate the quartic vertex V_4 to leading order as

$$\overline{V}^{(\ell)}_{\alpha\beta\gamma\delta} = \left(\frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{n_{\ell-1}}{n_\ell(n_\ell-1)} \sum_{\substack{i,j=1 \\ i\neq j}}^{n_\ell} z^{(\ell)}_{I;i,\alpha} z^{(\ell)}_{I;i,\beta} z^{(\ell)}_{I;j,\gamma} z^{(\ell)}_{I;j,\delta} \right) - n_{\ell-1} \overline{K}^{(\ell)}_{\alpha\beta} \overline{K}^{(\ell)}_{\gamma\delta} + \mathcal{O}\left(\frac{1}{n}\right). \tag{201}$$

where $\overline{K}$ is defined in equation 199. Here we exploit channel symmetry and estimate

$$\overline{K}^{(\ell)}_{\alpha\beta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} z^{(\ell)}_{I;i,\alpha} z^{(\ell)}_{I;i,\beta}. \tag{202}$$

To estimate uncertainties, we repeat the Monte Carlo procedure N_{stats} times and report the mean and standard deviation. Results up to $\ell = 30$ are shown in Figures 13 (critical) and 14 (non-critical). In the critical case, fitted asymptotic power laws confirm the stability of V_4.

Moreover, at criticality, the quartic vertex V remains well-behaved beyond the perturbative regime $\ell < n$. In Figure 12 (b), we compare its diagonal component with the single-input exact solution (Section H.1) and the large-ℓ expansion (Section J). The Monte Carlo estimates closely track the exact solution across all depths and align with the asymptotic expansion at large ℓ. This provides further evidence for the stability of orthogonal networks in the regime $\ell > n$, consistent with Day et al. (2025).

Monte–Carlo estimation of the tensors D, F, A and B. Similarly to V, these tensors are estimated to leading order as

$$\overline{D}^{(\ell)}_{\alpha\beta\gamma\delta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{n_{\ell-1}}{n_\ell^2} \sum_{i,j=1}^{n_\ell} z^{(\ell)}_{I;i,\alpha} z^{(\ell)}_{I;i,\beta} \widehat{\Delta\Theta}^{(\ell)}_{I;jj,\gamma\delta} + \mathcal{O}\left(\frac{1}{n}\right) \tag{203}$$

$$\overline{F}^{(\ell)}_{\alpha\gamma\beta\delta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{n_{\ell-1}}{n_\ell^2} \sum_{i,j=1}^{n_\ell} z^{(\ell)}_{I;i,\alpha} z^{(\ell)}_{I;j,\beta} \widehat{\Delta\Theta}^{(\ell)}_{I;ij,\gamma\delta} + \mathcal{O}\left(\frac{1}{n}\right) \tag{204}$$

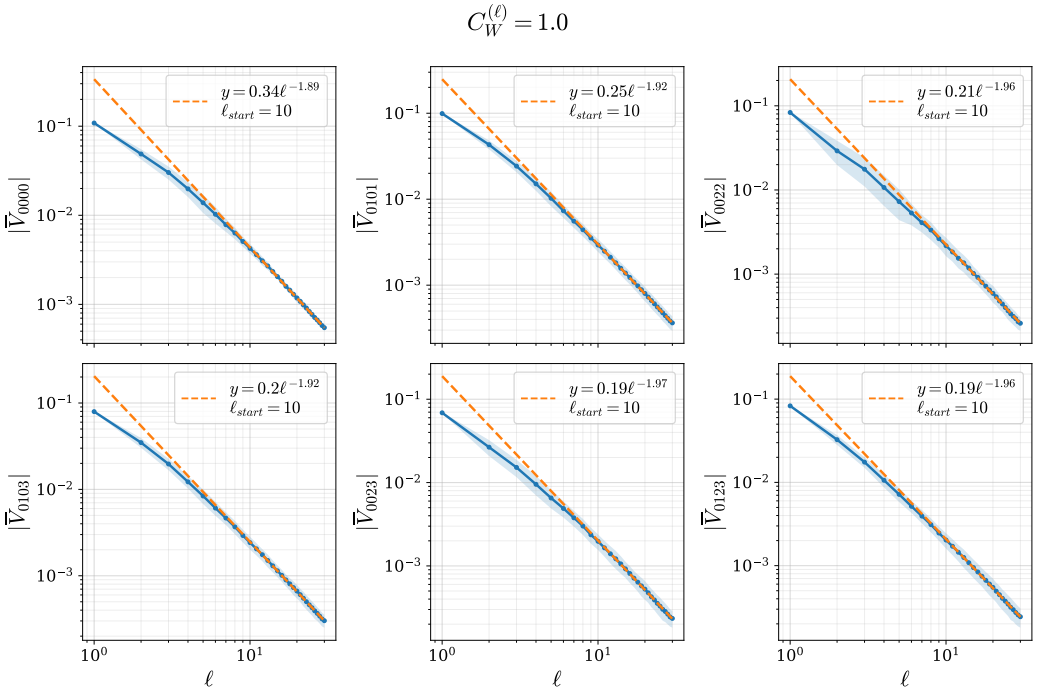


Figure 13: *Stability of the four-point cumulant V_4 at criticality.* Selected components of the Monte Carlo estimate $\overline{V}_4$ for a $\tanh$ MLP are shown as a function of layer depth ℓ at the critical value $C_W^{(\ell)} = 1$. Hidden layers have width 50. An asymptotic power-law fit is shown in orange, with the fit starting at ℓ_{start}. Estimates are obtained from $N_{\text{net}} = 600$ initializations, with means and error bars computed over $N_{\text{stats}} = 10$ repetitions (see text).

$$\overline{A}^{(\ell)}_{\alpha\beta\gamma\delta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{n_{\ell-1}}{n_\ell^2} \sum_{i,j=1}^{n_\ell} \widehat{\Delta\Theta}^{(\ell)}_{I;ii,\alpha\beta} \widehat{\Delta\Theta}^{(\ell)}_{I;jj,\gamma\delta} + \mathcal{O}\left(\frac{1}{n}\right) \tag{205}$$

$$\overline{B}^{(\ell)}_{\alpha\gamma\beta\delta} = \frac{1}{N_{\text{net}}} \sum_{I=1}^{N_{\text{net}}} \frac{n_{\ell-1}}{n_\ell^2} \sum_{i,j=1}^{n_\ell} \widehat{\Delta\Theta}^{(\ell)}_{I;ij,\alpha\beta} \widehat{\Delta\Theta}^{(\ell)}_{I;ij,\gamma\delta} + \mathcal{O}\left(\frac{1}{n}\right) , \tag{206}$$

where the NTK fluctuation $\widehat{\Delta\Theta}^{(\ell)}_{ij,\alpha\beta}$ was introduced in Section 3.1. The computation is repeated N_{stats} times to estimate the mean and standard deviation. For each tensor, both critical and non-critical cases are evaluated using the same input configurations as for V, up to $\ell = 30$. Results for D, F, A, and B are shown in Figures 15–22. In the critical case, regressions of the asymptotic power laws (shown in orange) confirm the expected power-law behavior. This provides consistent evidence that the NTK tensors D, F, A, and B are stabilized under the same criticality conditions as in the infinite-width limit.

Similarly, at criticality, the NTK tensors remain well-behaved beyond the perturbative regime $\ell < n$. In Figure 23, we compare the diagonal components of the tensors D and F with their single-input exact solutions (Section H.1) and large-ℓ expansions (Section J). The Monte Carlo estimates closely track the exact solutions across all depths and align with the asymptotic expansions at large ℓ. This provides further evidence for the stability of orthogonal networks in the regime $\ell > n$, consistent with Day et al. (2025).

K.2 Compute resources

All experiments were conducted on a laptop equipped with an AMD Radeon Pro 5500M (4 GB VRAM). Computations were performed on a single GPU. The stability analysis for the NNGP and V_4 was completed within several hours for all three values of C_W. Gradient-dependent tensors are

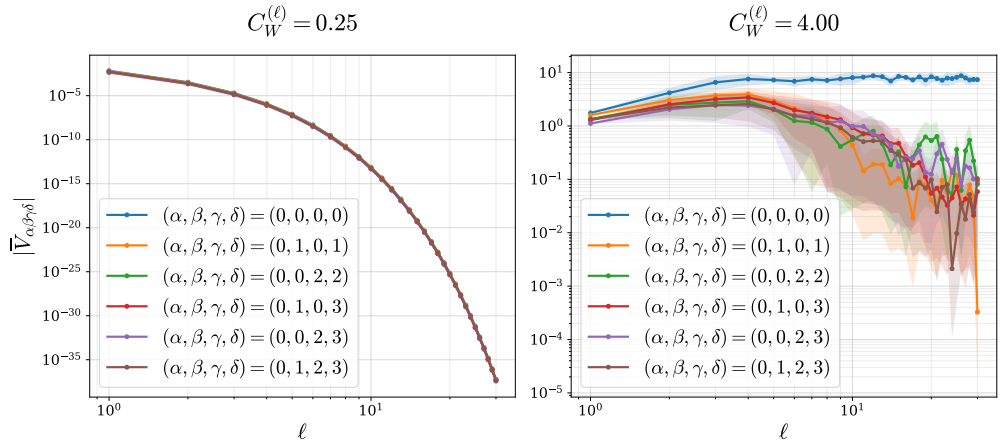


Figure 14: *Instability of the four-point cumulant V_4 away from criticality.*Selected components of the Monte Carlo estimate $\overline{V}_4$ are shown for $C_W^{(\ell)} < 1$ (left) and $C_W^{(\ell)} > 1$ (right). Estimates are computed from $N_{\text{net}} = 600$ initializations, with means and error bars obtained from $N_{\text{stats}} = 10$ repetitions (see text).

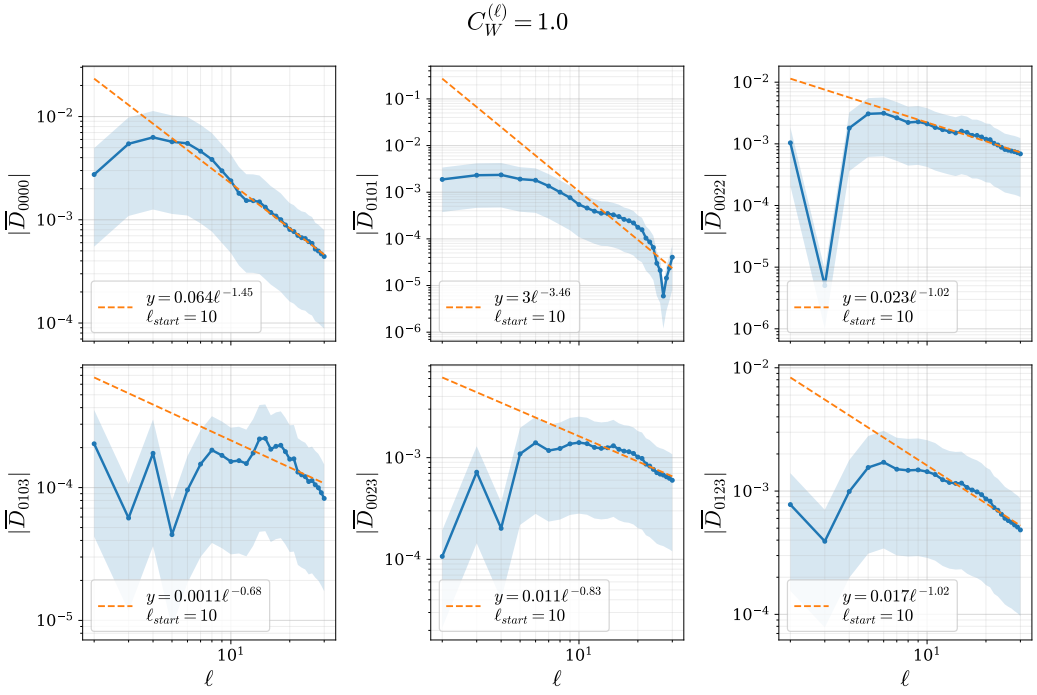


Figure 15: *Stability of the D tensor at criticality.* Selected components of the Monte Carlo estimate $\overline{D}$ for a $\tanh$ MLP are shown as a function of layer depth ℓ at the critical value $C_W^{(\ell)} = 1$. Hidden layers have width 50. An asymptotic power-law fit is shown in orange, with the fit starting at ℓ_{start}. Estimates are obtained from $N_{\text{net}} = 600$ initializations, with means and error bars computed over $N_{\text{stats}} = 10$ repetitions (see text).

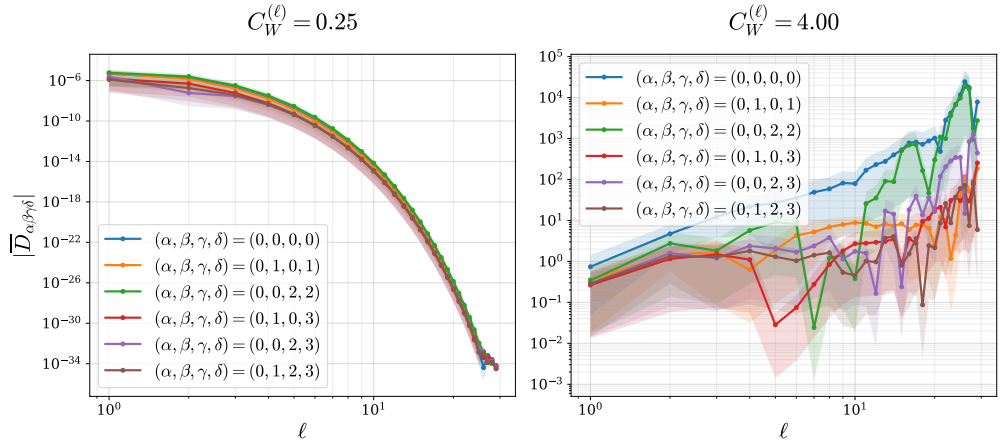


Figure 16: *Instability of the D tensor away from criticality.* Selected components of the Monte Carlo estimate $\overline{D}$ are shown for $C_W^{(\ell)} < 1$ (left) and $C_W^{(\ell)} > 1$ (right). Estimates are computed from $N_{\text{net}} = 600$ initializations, with means and error bars obtained from $N_{\text{stats}} = 10$ repetitions (see text).

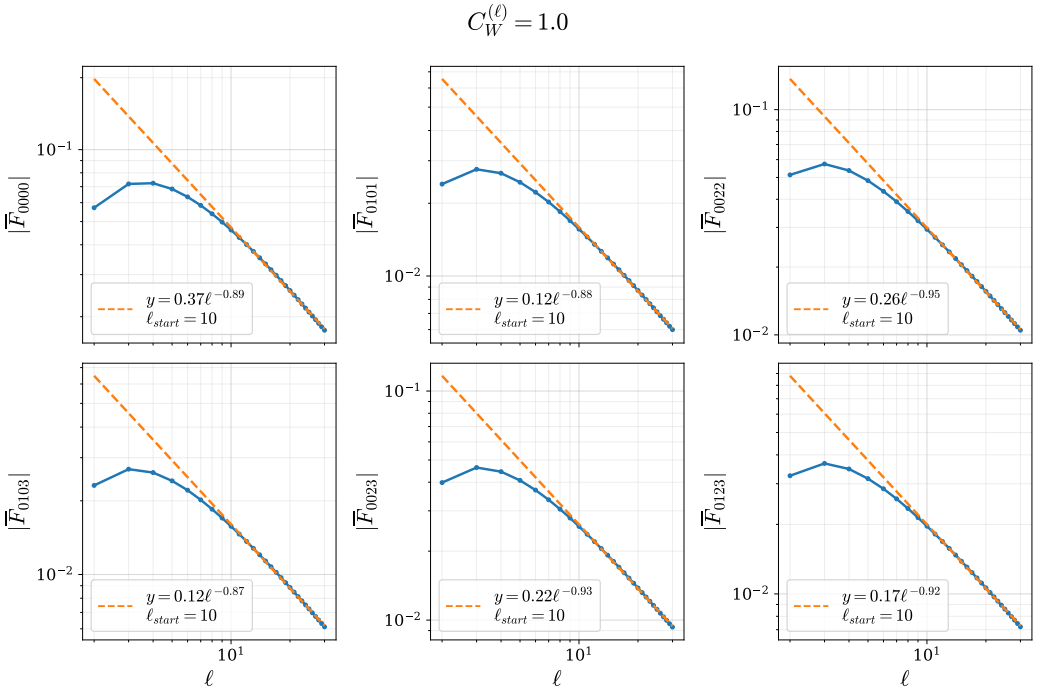


Figure 17: *Stability of the F tensor at criticality.* Selected components of the Monte Carlo estimate $\overline{F}$ for a $\tanh$ MLP are shown as a function of layer depth ℓ at the critical value $C_W^{(\ell)} = 1$. Hidden layers have width 50. An asymptotic power-law fit is shown in orange, with the fit starting at ℓ_{start}. Estimates are obtained from $N_{\text{net}} = 600$ initializations, with means and error bars computed over $N_{\text{stats}} = 10$ repetitions (see text).

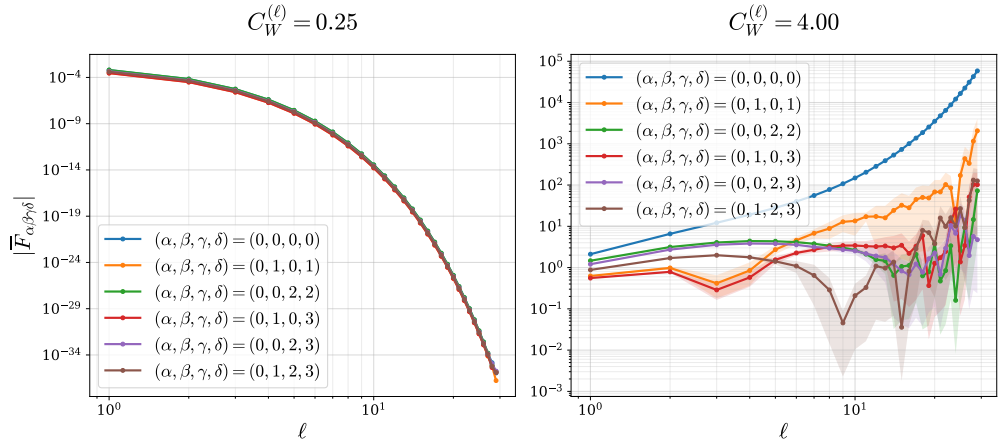


Figure 18: *Instability of the F tensor away from criticality.* Selected components of the Monte Carlo estimate $\overline{F}$ are shown for $C_W^{(\ell)} < 1$ (left) and $C_W^{(\ell)} > 1$ (right). Estimates are computed from $N_{\text{net}} = 600$ initializations, with means and error bars obtained from $N_{\text{stats}} = 10$ repetitions (see text).

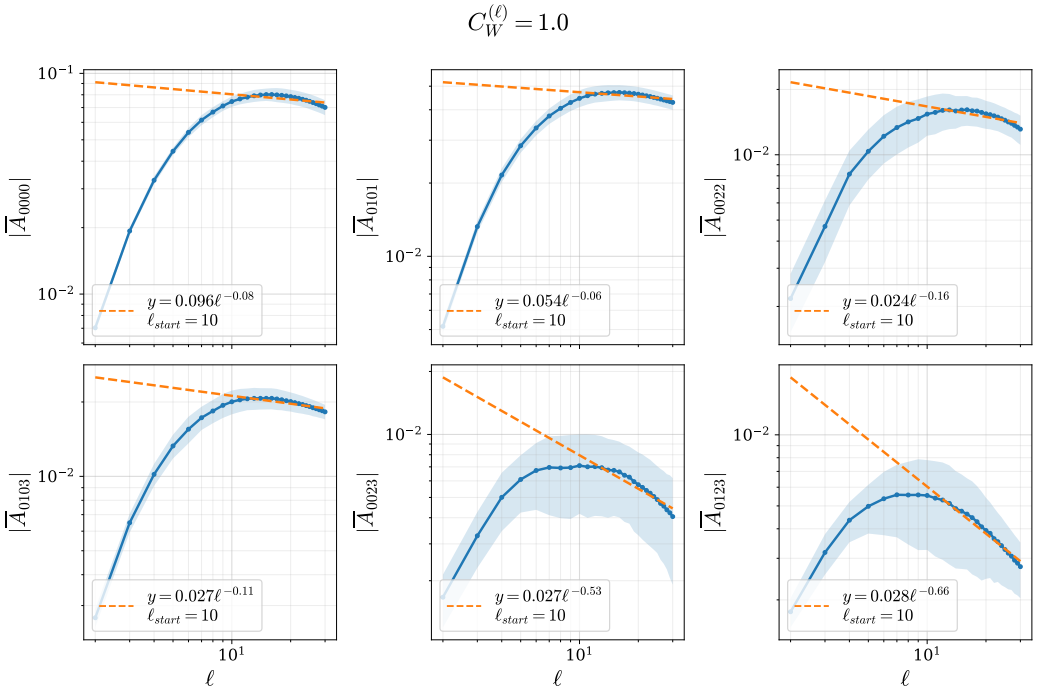


Figure 19: *Stability of the A tensor at criticality.* Selected components of the Monte Carlo estimate $\overline{A}$ for a $\tanh$ MLP are shown as a function of layer depth ℓ at the critical value $C_W^{(\ell)} = 1$. Hidden layers have width 50. An asymptotic power-law fit is shown in orange, with the fit starting at ℓ_{start}. Estimates are obtained from $N_{\text{net}} = 600$ initializations, with means and error bars computed over $N_{\text{stats}} = 10$ repetitions (see text).

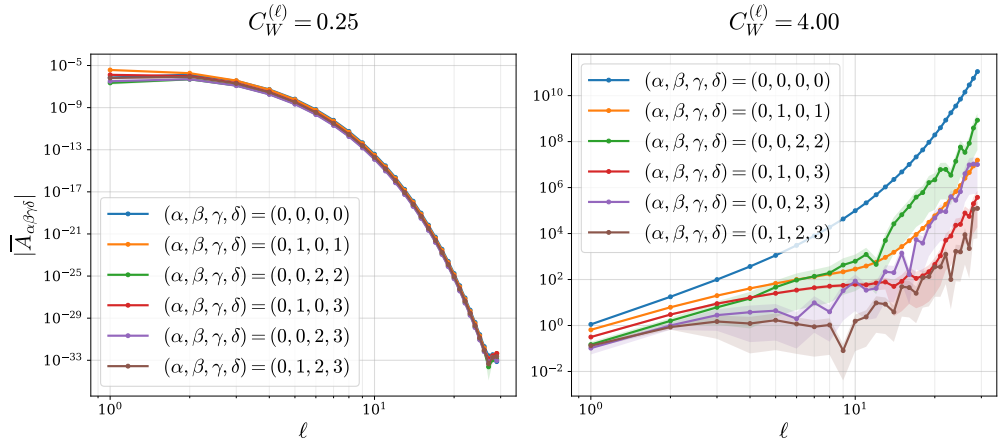


Figure 20: *Instability of the A tensor away from criticality.* Selected components of the Monte Carlo estimate $\overline{A}$ are shown for $C_W^{(\ell)} < 1$ (left) and $C_W^{(\ell)} > 1$ (right). Estimates are computed from $N_{\text{net}} = 600$ initializations, with means and error bars obtained from $N_{\text{stats}} = 10$ repetitions (see text).

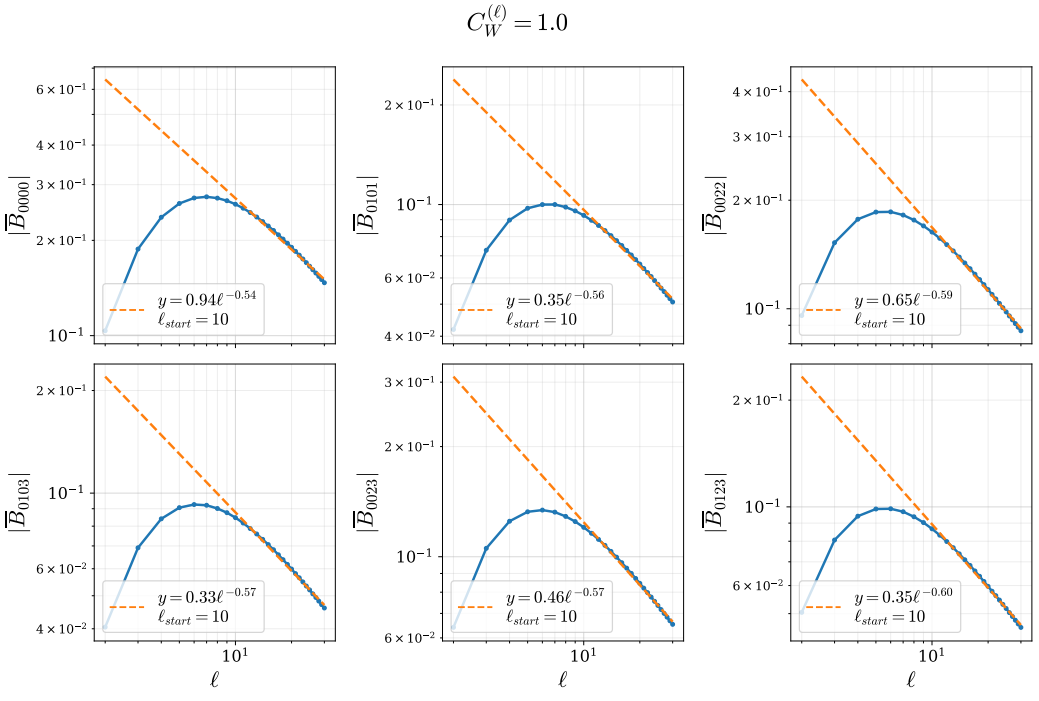


Figure 21: *Stability of the B tensor at criticality.* Selected components of the Monte Carlo estimate $\overline{B}$ for a $\tanh$ MLP are shown as a function of layer depth ℓ at the critical value $C_W^{(\ell)} = 1$. Hidden layers have width 50. An asymptotic power-law fit is shown in orange, with the fit starting at ℓ_{start}. Estimates are obtained from $N_{\text{net}} = 600$ initializations, with means and error bars computed over $N_{\text{stats}} = 10$ repetitions (see text).

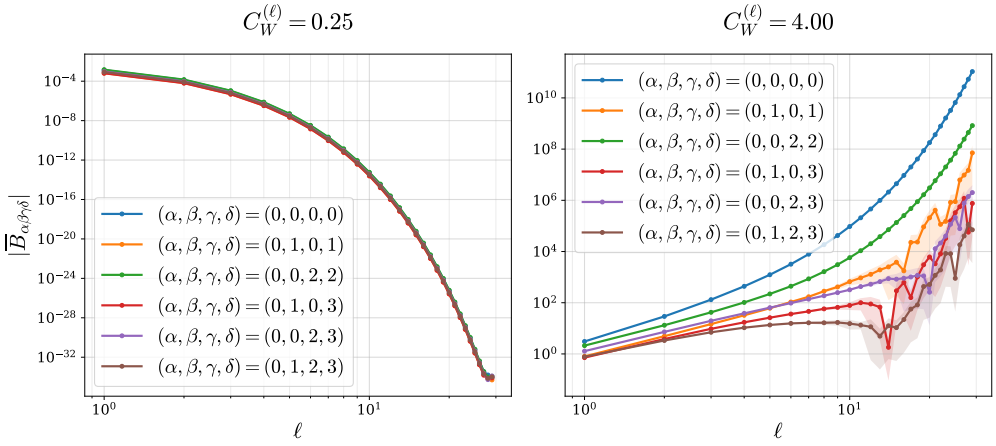


Figure 22: *Instability of the B tensor away from criticality.* Selected components of the Monte Carlo estimate $\overline{B}$ are shown for $C_W^{(\ell)} < 1$ (left) and $C_W^{(\ell)} > 1$ (right). Estimates are computed from $N_{\text{net}} = 600$ initializations, with means and error bars obtained from $N_{\text{stats}} = 10$ repetitions (see text).

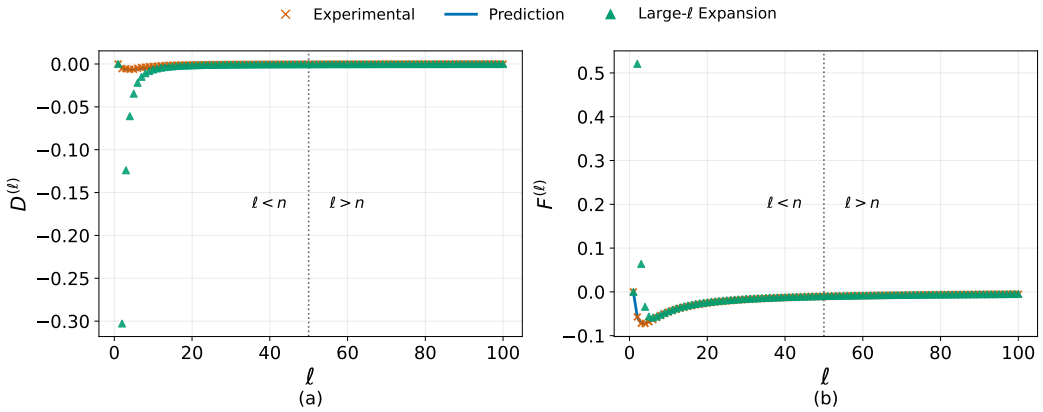


Figure 23: *Stability beyond the perturbative regime.* Comparison of the diagonal components of the Monte Carlo estimate, single-input exact solution, and large-ℓ expansion for the NTK tensors D and F in a tanh MLP with orthogonal initialization. Hidden layers have width 50; means are computed over 600 initializations. (a) The tensor D estimates are in quantitative agreement with the exact solution at both small and large depths. The large-ℓ expansion is inaccurate at small ℓ, as expected, but becomes accurate after a few layers. Stability persists up to $\ell = 100$, well beyond the perturbative regime $\ell < n$. (b) The tensor F exhibits analogous scaling behavior.

more computationally demanding; accordingly, the layer width was reduced to 50, and each tensor (for three values of C_W) required up to $\mathcal{O}(10)$ hours to compute.