Mutual Information of Neural Network Initializations: a Random Matrix Theory Study

A Preprint

‡ **Giuseppe Ughi** , † **Jared Tanner**
‡ QuantumBlack, London, UK
† Mathematical Institute, University of Oxford, UK
giuseppe.ughi@quantumblack.com
tanner@maths.ox.ac.uk

April 30, 2023

Abstract

The properties of randomly initialised feed-forward neural networks are known to be influenced by the variance of the weight matrices and biases, as well as the choice of the nonlinear activation function. This phenomenon was first studied from a geometric perspective in [19] and from an information-theoretical perspective in [21]. Specifically, [21] introduced a lower bound of the mutual information between an input and its hidden layers' outputs when the activation functions are odd. Here, the same lower bound is analyzed using more advanced techniques from random matrix theory to model the eigen-distribution of the random matrices determining the bound when no bias is considered.

Keywords Neural Nework Initialization, Mutual Information, Random Matrix Theory

1 Introduction

Starting with the work in [6], the research on initialising a feed-forward neural network (henceforth abridged to neural network) focused on analysing the properties of random neural networks. Random neural networks are random nonlinear functions from which neural networks are sampled and subsequently trained to map a training set of inputs to known outputs. Specifically, the following recursive relation is considered

$$\mathbf{h}^{(\ell)} = \mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell-1)}) + \mathbf{b}^{(\ell)} \tag{1}$$

where $\mathbf{h}^{(0)} = \mathbf{x}$ is the input, and $\mathbf{W}_{ij}^{(\ell)} \sim \mathcal{N}(0, \sigma_w^2/n_{\ell-1})$ and $\mathbf{b}_i^{(\ell)} \sim \mathcal{N}(0, \sigma_b^2)$ for $1 < i < n_\ell$ and $1 < j < n_{\ell-1}$, with n_ℓ being the width of layer ℓ, and $\phi(\cdot)$ the activation function.

The way the choice of $(\sigma_w, \sigma_b, \phi(\cdot))$ affects the properties of a neural network sampled from (1) has been studied from a diverse set of perspectives [19, 14, 12, 1, 23]. Especially notabily is the work in [19] which pioneered that, for a given nonlinear activation function $\phi(\cdot)$, it is possible to select the parameters (σ_w, σ_b) such that the sampled neural networks preserve geometric information about the inputs and this choice of parameters typically leads to a superior initial training. This analysis relied on geometric considerations of how the distribution of intermediate hidden layers converges to their limiting distribution.

The work in [21] conducted an alternative investigation on the flow of information through the layers of random neural networks with odd activation functions from an information-theoretical perspective; specifically, they studied the decay of the mutual information between an input and its hidden layers' output, building on the results of [1, 20]. Following the methodology in [7, 20, 17, 7, 5], the contributions in [21] considered a noise term $\mathbf{n}^{(\ell)} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ added before the activation function

$$\mathbf{h}^{(\ell)} = \mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell-1)}) + \mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)}, \tag{2}$$

and introduced the following lower bound on the mutual information $I(\mathbf{x};\mathbf{h}^{(\ell)})$ between the Gaussian input $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2\mathbf{I})$, as also modelled in [4], and the hidden layer $\mathbf{h}^{(\ell)}$

$$I(\mathbf{x};\mathbf{h}^{(\ell)}) \geq I^{\mathcal{G}}(\mathbf{x};\mathbf{h}^{(\ell)}) = \mathbb{E}_{\mathcal{W}^{:\ell}}\left[\frac{1}{2}\log\left(\frac{|\mathbf{\Lambda}_{h^{(\ell)}}|}{\left|\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}\right|}\right)\right], \tag{3}$$

where $|\cdot|$ denotes the matrix determinant, $\mathcal{W}^{:\ell} = \{\mathbf{W}^{(i)}, \mathbf{b}^{(i)}\}_{i=1}^{\ell}$, and

$$\mathbb{V}ar_{\mathbf{x},\{\mathbf{n}^{(l)}\}_{l=1}^{\ell}|\mathcal{W}^{:\ell}}\begin{bmatrix}\mathbf{x}\\ \mathbf{h}^{(\ell)}\end{bmatrix} = \begin{bmatrix}\sigma_x^2\mathbf{I} & \mathbf{\Sigma}_{xh^{(\ell)}}\\ \mathbf{\Sigma}_{xh^{(\ell)}}^\top & \mathbf{\Lambda}_{h^{(\ell)}}\end{bmatrix}. \tag{4}$$

The lower bound in (3) was then approximated via the mean-field theory as a function of $(\sigma_w, \sigma_b, \phi(\cdot))$, suggesting that in some cases the initialisations are optimal from both a training and a mutual information perspective. However, although the mean-field approximation was qualitatively accurate, there was a consistent error in the approximation due to the strong assumptions that it entailed.

Here, an alternative approximation of the lower bound $I^{\mathcal{G}}(\mathbf{x};\mathbf{h}^{(\ell)})$ in (3) is proposed by taking into consideration the eigen-spectrum of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$. In general, considering a function $F(\mathbf{A})$ that can be written as a power series on an n-dimensional random matrix $\mathbf{A}$ with eigen-distribution $\rho_{\mathbf{A}}$, the limiting normalised trace for $n \to \infty$ of the function $F(\mathbf{A})$ can be evaluated as an expectation on $\rho_{\mathbf{A}}$;

$$\lim_{n\to\infty}\frac{1}{n}trF(\mathbf{A}) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}F(\lambda_j) = \mathbb{E}_{\rho_{\mathbf{A}}}[F(\lambda)]. \tag{5}$$

This property is applied to the log-determinant case. As a matter of fact, $log|\mathbf{A}| = tr(log(\mathbf{A}))$ and if the eigen-distribution $\rho_{\mathbf{A}}(\lambda)$ of matrix $\mathbf{A}$ is known, then

$$\lim_{n\to\infty}\log(|\mathbf{A}|)/n = \lim_{n\to\infty}\log\left(\prod_i\lambda_i\right)/n = \lim_{n\to\infty}\sum_i\log(\lambda_i)/n = \int\log(x)\,d\rho_{\mathbf{A}}(x). \tag{6}$$

Therefore, by learning the eigen-spectrum of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ one is able to compute their log-determinants whose difference defines the mutual information lower bound (3).

1.1 Outline and main contributions

The manuscript focuses on calculating the lower bound of the mutual information between an input and its representation as a hidden layer vector at layer ℓ; that is the lower bound in (3). Section 2 considers approximating the eigen-spectra of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ with the Marchenko-Pastur distribution and show that the resulting lower bounds of (3) are equal to those preveously derived by the authors in [21] where instead a mean-field approximation was used and the spectra distribution was treated as a point distribution; see Proposition 1 and (15). Section 3 improves upon the calculation of the spectra of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ by making use of the Stieltjes Transform. In particular, subsection 3.1 computes the spectra of $\mathbf{\Lambda}_{h^{(\ell)}}$ as given by Theorem 3.1; see Figure 1 for plots of the spectra with the demonstrated improvements. Section 3.2 computes the spectra of $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ as given by Theorem 3.2 and Figure 2 shows calculations of the spectra which illustrate the improved fit to empirical observations. Section 3.4 then makes use of these estimates of the spectra to compute the associated lower bound of the mutual information in (3); see Figure 3. Figures 4 shows separate contributions of the mutual information bound from each of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ as functions of the weight variance σ_w and for different layer depths. Figure 5 shows that the mutual information bound converges towards a single point in σ_w as depth increases. These results give an alternative perspective on the geometric focused edge-of-chaos theory in [11].

2 Marchenko-Pastur Approximation

The Marchenko-Pastur law determines via an analytic expression the eigen-spectrum of a specific type of random sample-covariance matrices when the dimensions of the sampled space and the number of samples go to infinity. The tractability of this expression makes the approximation of the eigen-spectra of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top\mathbf{\Sigma}_{xh^{(\ell)}}$ with the Marchenko-Pastur law a valid baseline for understanding the impact of using the eigen-spectra when computing the mutual information lower bound. In Figures 1 and 2 it is shown how the eigen-spectra of the empirical covariance matrices compares to the Marchenko-Pastur distribution alongside the approximations introduced in Section 3.

Specifically, consider an $n \times m$ matrix $\mathbf{X}$ with entries $\mathbf{X}_{ij}$ that are i.i.d. real random variables such that $\mathbb{E}[\mathbf{X}_{ij}] = 0$ and $\mathbb{E}[\mathbf{X}_{ij}^2] = \sigma^2$, and denote by $\mathbf{M}$ the $n \times n$ matrix

$$\mathbf{M} = \frac{1}{m}\mathbf{X}\mathbf{X}^\top \in \mathbb{R}^{n\times n} \tag{7}$$

with $0 \leq \lambda_1 \leq \ldots \leq \lambda_n$ being the eigenvalues of $\mathbf{M}$. Defining the random spectral distribution by

$$\mu_{\mathbf{M}}(x) := \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{x>\lambda_j}, \tag{8}$$

with $\mathbb{1}$ being the indicator function, the Marchenko-Pastur law is defined in Theorem 2.1.

Theorem 2.1 (Marchenko-Pastur Law [9]). *For $\boldsymbol{M}$ and $\mu_{\boldsymbol{M}}$ as defined above, when $n \to \infty$ and $m \to \infty$ such that $n/m \to \gamma \in (0, 1]$, $\mu_{\boldsymbol{M}} \to \mu$ in expectation and almost surely, where μ is the deterministic measure satisfying*

$$\frac{d\mu}{dx} = \begin{cases} \frac{1}{2\pi\sigma^2\gamma x}\sqrt{(\sigma^2 a_+ - x)(x - \sigma^2 a_-)} & \text{if } a_- \leq x \leq a_+ \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

with $a_\pm = (1 \pm \sqrt{\gamma})^2$.

The log-determinant of a matrix whose eigenvalues are distributed according to the Marchenko-Pastur distribution [2] is defined as

$$\mathbb{E}_\mu\left[\frac{log(|X|)}{n}\right] = (1 - \frac{1}{\gamma})\log(1-\gamma) - 1 + \log(\sigma^2). \tag{10}$$

Consequently, the log-determinant of a matrix, whose eigenvalues are distributed according to the Marchenko-Pastur distribution, is determined uniquely by the variance of the eigenvalues, $\mathbb{V}ar_\mu[\lambda] = \sigma^2$, and the shape of the matrices γ. Further, for the Marchenko-Pastur distribution the variance of the eigenvalues is the same as their expectation, i.e. $\mathbb{E}_\mu[\lambda] = \sigma^2$. Relative to the mutual information lower bound estimation, the shape γ is equal to 1[1], and it was shown in [21] that the expected eigenvalues of the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ are the variables $q^{(\ell)}$ and $q_c^{(\ell)}$, defined by the following relations

$$\begin{cases} q^{(\ell)} = \sigma_w^2 \int \phi(\sqrt{q^{(\ell-1)}}z)\mathcal{D}z + \sigma_b^2 + \sigma_n^2 \\ q^{(1)} = \sigma_w^2\sigma_x^2 + \sigma_b^2 + \sigma_n^2 \end{cases} \tag{11}$$

and

$$\begin{cases} q_c^{(\ell)} = q^{(\ell)} - \sigma_w^2 q^{(\ell)} \left(\int \rho^{(\ell-1)}\phi'\left(\sqrt{q^{(\ell)}}z\right)\mathcal{D}z\right)^2 \\ \rho^{(\ell)} = \frac{1}{n\sqrt{q^{(\ell)}}}\sqrt{q^{(\ell)} - q_c^{(\ell)}} \end{cases} \quad \text{with } \rho^{(1)} = 1, \tag{12}$$

see [15] for other interpretations of $q^{(\ell)}$. Therefore it is possible to compute explicitly the log-determinant of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ under the Marchenko-Pastur distribution assumption.

Proposition 1. *Assuming that the eigen-distributions of the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ follow the Marchenko-Pastur distribution with means $q^{(\ell)}$ and $q_c^{(\ell)}$ respectively, the mutual information is lower bounded by*

$$I(\boldsymbol{x}; \boldsymbol{h}^{(\ell)}) \geq \frac{n}{2}\log\left(\frac{q^{(\ell)}}{q_c^{(\ell)}}\right) \tag{13}$$

Proof. By considering equation (10) with $\gamma = 1$, $\mathbb{V}ar_{\mathbf{\Lambda}_{h^{(\ell)}}}[\lambda] = \mathbb{E}_{\mathbf{\Lambda}_{h^{(\ell)}}}[\lambda] = q^{(\ell)}$, and $\mathbb{V}ar_{\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}}[\lambda] = \mathbb{E}_{\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}}[\lambda] = q_c^{(\ell)}$, it follows that (3) can be expressed as follows

$$I(\mathbf{x}; \mathbf{y}) \geq \frac{1}{2}\mathbb{E}_{\mathcal{W}^{:\ell}}\left[\log\left(|\mathbf{\Lambda}_{h^{(\ell)}}|\right)\right] - \mathbb{E}_{\mathcal{W}^{:\ell}}\left[\frac{1}{2}\log\left(\left|\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}\right|\right)\right] \tag{14}$$

[1]Following the practise in information theory $0\log(0) = 0$. This singularity is due to having no noise included in the distribution.

$$= \frac{n}{2}\log\left(q^{(\ell)}\right) - \frac{n}{2} - \frac{n}{2}\log\left(q_c^{(\ell)}\right) + \frac{n}{2} = \frac{n}{2}\log\left(\frac{q^{(\ell)}}{q_c^{(\ell)}}\right) \tag{15}$$

□

This shows that the lower bound defined under the mean-field approximation in [21] is equivalent to the one arising when assuming that the eigen-distributions of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ are defined by the Marchenko-Pastur distribution.

3 Spectra calculations using the Stieltjes Transform

The Stieltjes transform allows to study the eigen-distribution resulting from the interaction of random matrices; therefore it is possible to rely on this transform to model accurately the eigen-spectrum of the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$, as it was done in [13, 3], and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$, as it is introduced in this work.

More in detail, when considering a random matrix $\mathbf{X}$ the Stieltjes transform is defined by the limiting eigen-distribution $\rho_{\mathbf{X}}$ as follows

$$G_{\mathbf{X}} = \int_{\mathbb{R}} \frac{\rho_{\mathbf{X}}(t)}{z-t} dt = \sum_{k=0}^{\infty} \frac{\mathfrak{m}_k}{z^{k+1}} \qquad z \in \mathbb{C} \setminus \mathbb{R} \tag{16}$$

where $\mathfrak{m}_k = \int \lambda^k \rho_{\mathbf{X}}(\lambda) d\lambda$. If the Stieltjes transform is known, it is also possible to retrieve the underlying distribution $\rho_{\mathbf{X}}$ by considering the Sokhotski-Plemelj formula

$$\rho_{\mathbf{X}}(t) = \frac{1}{\pi} \lim_{\nu \to 0^+} Im\left(G_{\mathbf{X}}(t - i\nu)\right). \tag{17}$$

The Stieltjes transform is especially useful when considering matrices that are free, a generalisation of independence in the matrix space; for more details consider [10].

Following [16], to identify how free matrices interact the moment generating function $M_{\mathbf{X}}$ is introduced,

$$M_{\mathbf{X}}(z) = zG_{\mathbf{X}}(z) - 1 \tag{18}$$

and then the S-transform and the R-transform are defined as

$$S_{\mathbf{X}}(z) = \frac{1+z}{zM_{\mathbf{X}}^{-1}(z)}, \qquad R_{\mathbf{X}}(z) = G_{\mathbf{X}}^{-1}(z) - \frac{1}{z}. \tag{19}$$

Specifically, if two matrices $\mathbf{M}$ and $\mathbf{C}$ are free relative to each other, it is possible to compute the Stieltjes transform of $\mathbf{CMC}^\top$, using the multiplicativity of the S-transform

$$S_{\mathbf{CMC}^\top}(z) = S_{\mathbf{CC}^\top}(z) S_{\mathbf{M}}(z). \tag{20}$$

and [16] retrieves the following implicit expression

$$M_{\mathbf{CMC}^\top}(\hat{z}) = M_{\mathbf{M}}(S_{\mathbf{CC}^\top}(M_{\mathbf{CMC}^\top}(\hat{z}))\hat{z}). \tag{21}$$

If $M_{\mathbf{M}}$ and $S_{\mathbf{CC}^\top}$ are known, the algorithm to compute the density of the eigenspectrum of $\mathbf{CMC}^\top$ is described in Algorithm 1; the property of the Stieltjes transform of behaving as $1/z$ for $|z| \to \infty$ is used to initialise the limit.

Furthermore, if $\mathbf{M}$ and $\mathbf{C}$ are free relative to each other, then it is possible to obtain the Stieltjes transform of $\mathbf{M} + \mathbf{C}$ via the additivity of the R-transform which is also based on the Stieltjes transform

$$R_{\mathbf{M}+\mathbf{C}}(z) = R_{\mathbf{M}}(z) + R_{\mathbf{C}}(z). \tag{22}$$

As also shown in [16], the following implicit equation holds

$$G_{\mathbf{M}+\mathbf{C}}(\bar{z}) = G_{\mathbf{M}}\left(\bar{z} - R_{\mathbf{C}}(G_{\mathbf{M}+\mathbf{C}}(\bar{z}))\right) \tag{23}$$

and consequently the computation of $\rho_{\mathbf{M}+\mathbf{C}}(x)$ is described in Algorithm 2.

These properties of the Stieltjes transform of free matrices are going to be central to the modelling of the spectral distribution of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$.

Algorithm 1: $\rho_{\mathbf{WAW}}(\lambda)$ [11]

Choose 2N steps and $b > 1$;
Initialize $z_0 = \lambda - ib^N$ and $G = 1/z$;
for $k \in \{1, ..., 2N\}$ **do**
  $z_k \leftarrow \lambda - ib^{N-k}$;
  $M_k \leftarrow$ Root of (21) nearest to $z_k G - 1$;
  $G \leftarrow (M_k + 1)/z_k$
end
Return $\frac{1}{\pi} Im\,(G)$

Algorithm 2: $\rho_{\mathbf{A+B}}(\lambda)$

Choose 2N steps and $b > 1$;
Initialize $z_0 = \lambda - ib^N$ and $G = 1/z$;
for $k \in \{1, ..., 2N\}$ **do**
  $z_k \leftarrow \lambda - ib^{N-k}$;
  $G \leftarrow$ Root of (23) nearest to G;
end
Return $\frac{1}{\pi} Im\,(G)$;

3.1 Eigenspectrum of $\mathbf{\Lambda}_{h^{(\ell)}}$ with $\sigma_b = 0$ and $\sigma_n = 0$

Given the weights $\mathbf{W}^{(\ell)}$ and $\mathbf{b}^{(\ell)}$, the matrix $\mathbf{\Lambda}_{h^{(\ell)}}$ is defined as

$$\Lambda_{\mathbf{h}^{(\ell+1)}} = \mathbb{E}_x[(\mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell)}) + \mathbf{n}^{(\ell)} + \mathbf{b}^{(\ell)})(\mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell)}) + \mathbf{n}^{(\ell)} + \mathbf{b}^{(\ell)})^\top] \tag{24}$$

$$= \frac{\sigma_w^2}{n}\boldsymbol{W}^{(\ell)}\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]\boldsymbol{W}^{(\ell)^\top} + (\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})(\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})^\top, \tag{25}$$

where the last equality is due to considering only odd activation functions ϕ and therefore the vector $\mathbb{E}[\phi(\mathbf{h}^{\ell})]$ being null.

The matrix $\mathbf{W}^{(\ell)}\mathbf{W}^{(\ell)^\top}$, whose S-transform is $S_{\mathbf{WW}^\top}(z) = \frac{1}{\sigma_w^2(1+z)}$ [16, Chapter 15.2], is free relative to $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]$, and it is possible to study the Stieltjes transform of their product with Algorithm 1 if the expression of the moment generating function $M_{\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]}$ is known. Moreover, since $\mathbf{W}^{(\ell)}\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]\mathbf{W}^{(\ell)^\top}$ is free relative to $(\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})(\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})^\top$, it is then possible to study the full spectrum of the matrix $\Lambda_{\mathbf{h}^{(\ell+1)}}$ with Algorithm 2.

Therefore, the crucial point is computing the Stieltjes transform of $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]$. For the second layer, it is possible to rely on the work in [13]; this work identified with Theorem 3.1 the Stieltjes transform of $\frac{1}{m}\phi(\mathbf{XW})\phi(\mathbf{WX})^{\top 2}$, where the input matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$ is considered with the columns being m i.i.d. vectors $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \sigma_x^2\mathbf{I})$, and the expectation $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]$ corresponds to the limit for $m \to \infty$. The following work in [3], extended this work to compute the spectral distribution of the covariance matrices after the second layer. Specifically, the Stieltjes transform of $\frac{1}{m}\mathbf{Y}^{(\ell+1)}\mathbf{Y}^{(\ell+1)\top}$ with $\mathbf{Y}^{(\ell+1)} = \phi\left(\mathbf{WY}^{(\ell)}\right)$, can be approximated by implementing the same expression for the second layer as in [13] considering $\mathbf{X} = \mathbf{Y}^{(\ell)}$ and $\sigma_x^2 = q^{(\ell)}/\sigma_w^2$.

Theorem 3.1 ([13, Theorem 1])**.** *Consider an odd activation function ϕ satisfying*

$$\left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^k(\sigma_w\sigma_x z)\right| < \infty \quad \textit{and} \quad \left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^{(k)}(\sigma_w\sigma_x z)\right| < \infty \tag{26}$$

for $k > 1$ with $\phi^{(k)}$ being the k-th derivative of ϕ, an input $\boldsymbol{X} \in \mathbb{R}^{n_0 \times m}$ and $\boldsymbol{W} \in \mathbb{R}^{n_1 \times n_o}$ with their respective columns being sampled as follows $\boldsymbol{w}_i^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, \frac{\sigma_w^2}{n_{\ell-1}}\boldsymbol{I})$ and $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_x^2\boldsymbol{I})$, and with $\boldsymbol{Y} = \phi(\boldsymbol{WX})$ and

$$\psi = \frac{n_0}{m}, \quad \varphi = \frac{n_0}{n_1}. \tag{27}$$

Then the eigen-distribution of the empirical covariance matrix $\boldsymbol{M} = \frac{1}{m}\boldsymbol{YY}^\top$ is asymptotically defined by

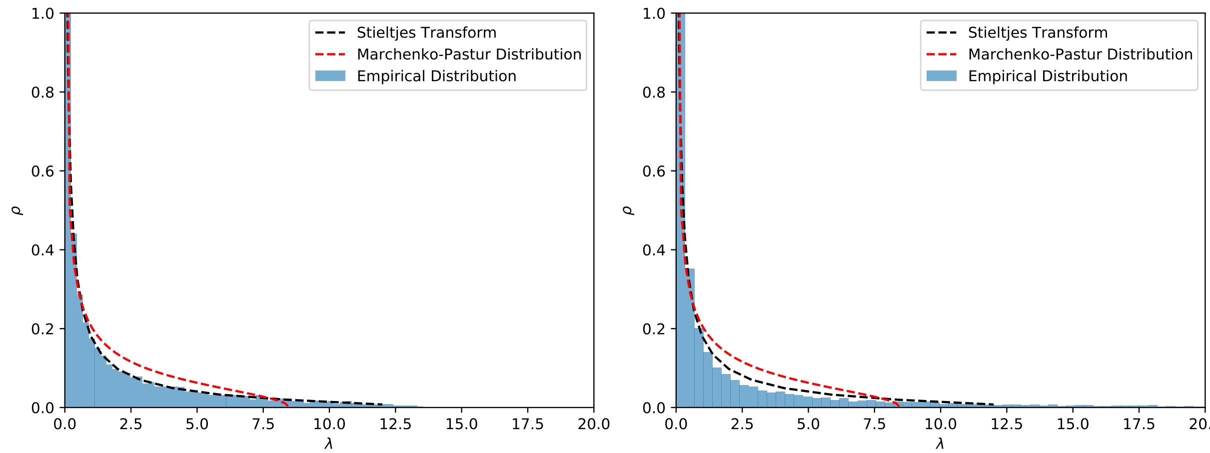
$$G_{\boldsymbol{M}}(z) \simeq \frac{\varphi}{z}H(\varphi z) + \frac{1-\varphi}{z} \tag{28}$$

where the generating function $H(z)$ satisfies the following recursive relation

$$H(z) = 1 + \frac{(\theta_1 - \theta_2)H_\varphi(z)H_\psi(z)}{z} + \frac{H_\varphi(z)H_\psi(z)\theta_2}{z - H_\varphi(z)H_\psi(z)\theta_2} \tag{29}$$

where $H_\psi(z) = 1 + (H(z) - 1)\psi$, $H_\varphi(z) = 1 + (H(z) - 1)\varphi$,

$$\theta_1 = \int \frac{1}{\sqrt{2\pi}}\phi(\sigma_w\sigma_x z)^2 e^{-\frac{z^2}{2}}dz, \quad \textit{and} \quad \theta_2 = \left(\int \frac{\sigma_w\sigma_x}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\phi'(\sigma_w\sigma_x z)dz\right)^2. \tag{30}$$



(a) 2nd Layer (b) 4th Layer

Figure 1: Approximation of the eigen-spectrum for the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ with $n = 100$ and $(\sigma_w, \sigma_b, \sigma_n) = (2, 0, 0.01)$ at the second and fourth layers. The histogram (Empirical Results) corresponds to the eigenvalues obtained on 200 different simulations and it is compared to the Marchenko-Pastur distribution with the same mean, to the distributions obtained with the Stieltjes transform for $\mathbf{\Lambda}_{h^{(\ell)}}$ in Theorem 3.1.

In Figure 1, an example is shown of how by relying on the Stieltjes transform in Theorem 3.1 it is possible to improve the description of the eigen-distribution of the matrix $\mathbf{\Lambda}_{h^{(\ell)}}$. Specifically, it is shown that for the second layer the approximation is very accurate, while some error creeps in for deeper layers.

3.2 Eigenspectrum of $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ with $\sigma_b = 0$ and $\sigma_n = 0$

Given the weights $\mathbf{W}^{(\ell)}$ and $\mathbf{b}^{(\ell)}$, the matrix $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ is defined as follows

$$\begin{aligned}
\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}} &= \mathbb{E}_x[(\mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell)}) + \mathbf{n}^{(\ell)} + \mathbf{b}^{(\ell)})(\mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell)}) + \mathbf{n}^{(\ell)} + \mathbf{b}^{(\ell)})^\top] \\
&\quad - \frac{1}{\sigma_x^2}\mathbb{E}_x[\mathbf{W}^{(\ell)}\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top {\mathbf{W}^{(\ell)}}^\top] \qquad (31) \\
&= \mathbf{W}^{(\ell)}\left(\mathbb{E}_\mathbf{x}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \frac{1}{\sigma_x^2}\mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]\right){\mathbf{W}^{(\ell)}}^\top \\
&\quad + (\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})(\mathbf{b}^{(\ell)} + \mathbf{n}^{(\ell)})^\top, \qquad (32)
\end{aligned}$$

where the last equality is due to considering only odd activation functions ϕ and therefore the vector $\mathbb{E}[\phi(\mathbf{h}^{(\ell)})]$ being null.

The two matrices $\mathbb{E}_\mathbf{x}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top]$ and $\mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$ are not free relative to each other and the additivity property of the R-transform cannot be implemented; therefore the $\mathbb{E}_\mathbf{x}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$ matrix has to be considered as one. The changes to the eigen-spectrum due to the remaining operations in (32) follow the same logic as for the ones in (25). Thus the only unknown is the Stieltjes transform of $\mathbb{E}_\mathbf{x}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_\mathbf{x}[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$ which is the $m \to \infty$ case for the expression introduced in Theorem 3.2.

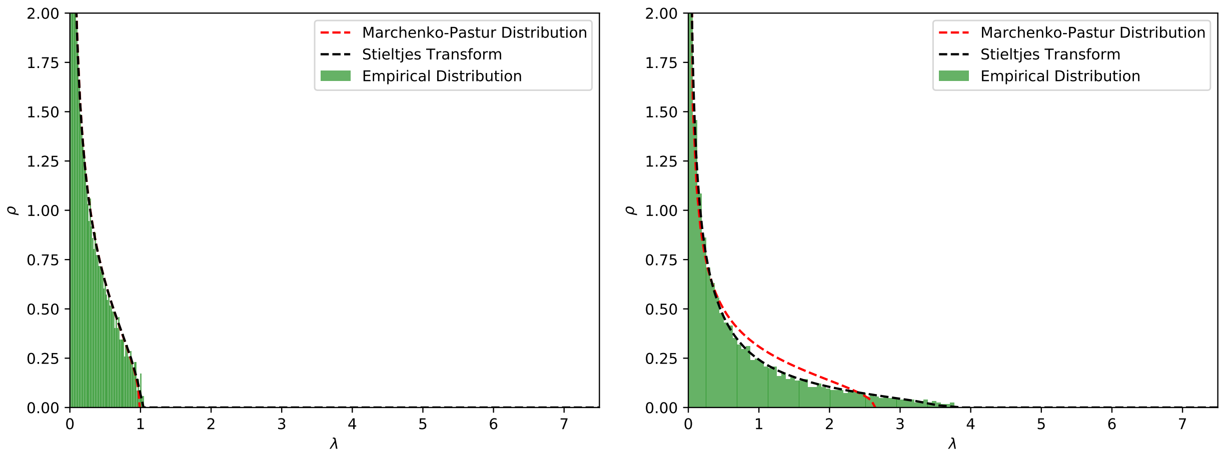
Theorem 3.2. *Consider the odd activation function ϕ for which the following holds*

$$\left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^k(\sigma_w \sigma_x z)\right| < \infty \quad \textit{and} \quad \left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^{(k)}(\sigma_w \sigma_x z)\right| < \infty \qquad (33)$$

for $k > 1$ with $\phi^{(k)}$ being the k-th derivative of ϕ, for each layer the matrices $\boldsymbol{W}^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$, an input matrix $\boldsymbol{X} \in \mathbb{R}^{n_0 \times m}$ with their respective columns being sampled as follows $\boldsymbol{w}_i^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, \frac{\sigma_w^2}{n_{\ell-1}}\boldsymbol{I})$ and $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_x^2\boldsymbol{I})$. Then define $\boldsymbol{H}_{i,p}^{(\ell)} = \sum_{k_\ell=1}^n \boldsymbol{W}_{i,k_\ell}^{(\ell)}\boldsymbol{Y}_{k_\ell,p}^{(\ell)}$ with $\boldsymbol{Y}_{i,p}^{(\ell)} = \phi(\boldsymbol{H}_{i,p}^{(\ell-1)})$ and with $\boldsymbol{Y}^{(1)} = \boldsymbol{X}$, $\psi = \frac{n_0}{m}$, and $\varphi = \frac{n_0}{n_1}$. Under the assumption that each column in $\boldsymbol{Y}^{(\ell)}$ is distributed according to $\boldsymbol{Y}_{:p}^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{I})$, the matrix

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_\boldsymbol{X}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_\boldsymbol{X}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_\boldsymbol{X}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \qquad (34)$$

[2] In this subsection the superscript $^{(\ell)}$ is omitted for clarity.



(a) 2nd Layer   (b) 4th Layer

Figure 2: Approximation of the eigen-spectrum for the matrices $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ with $n = 100$ and $(\sigma_w, \sigma_b, \sigma_n) = (2, 0, 0.01)$ at the second and fourth layers. The histogram (Empirical Results) corresponds to the eigenvalues obtained on 100 different simulations and it is compared to the Marchenko-Pastur distribution with the same mean, to the distributions obtained with the Stieltjes transform for $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ in Theorem 3.2.

has an eigen-distribution whose Stieltjes transform is asymptotically defined by

$$G_{\boldsymbol{M}}(z) \simeq \frac{1-\varphi}{z} + \frac{\varphi}{z} H(z) \tag{35}$$

where

$$\begin{aligned} H(z) =& \frac{H_{\psi b}(z) H_\varphi(z) \left(\theta_1^{(\ell)} - \sigma_x^2 \sigma_w^{2\ell} \left(\prod_{l=1}^{\ell} \theta_3^{(l)} \right)^2 \right)}{\varphi z} - \frac{H_{\psi c}(z) H_\varphi(z) \left(\theta_2^{(\ell)} - \sigma_x^2 \sigma_w^{2\ell} \left(\prod_{l=1}^{\ell} \theta_3^{(l)} \right)^2 \right)}{\varphi z} \\ &+ \frac{H_{\psi c}(z) H_\varphi(z) \left(\theta_2^{(\ell)} - \sigma_x^2 \sigma_w^{2\ell} \left(\prod_{l=1}^{\ell} \theta_3^{(l)} \right)^2 \right)}{\varphi z - H_{\psi c}(z) H_\varphi(z) \left(\theta_2^{(\ell)} - \sigma_x^2 \sigma_w^{2\ell} \left(\prod_{l=1}^{\ell} \theta_3^{(l)} \right)^2 \right)} + 1 \end{aligned} \tag{36}$$

with

$$\theta_1^{(\ell)} = \int \frac{1}{\sqrt{2\pi}} \phi(\sigma_w \sqrt{q^{(\ell)}} z)^2 e^{-\frac{z^2}{2}} dz, \qquad \theta_2^{(\ell)} = \left(\int \frac{\sqrt{q^{(\ell)}}}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi'(\sqrt{q^{(\ell)}} z) dz \right)^2, \tag{37}$$

$$\theta_3^{(\ell)} = \int \frac{\phi'\left(\sqrt{q^{(\ell)}} z\right)}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, \tag{38}$$

$H_{\psi\alpha}(z) = 1 + \kappa_\alpha \psi(H(z) - 1)$, $H_\varphi = 1 + \varphi(H(z) - 1)$, $\alpha \in \{b, c\}$, $\kappa_c = 1 + \frac{2\sigma_x^2\sigma_w^2 \theta_3^{(\ell)^2} \theta_2^{(\ell)} - \sigma_x^4 \sigma_w^4 \theta_3^{(\ell)^4}}{(\theta_2^{(\ell)} - \sigma_x^2 \sigma_w^2 \theta_3^{(\ell)^2})^2}$, *and* $\kappa_b = 1 + \frac{\theta_1^{(\ell)} \sigma_x^2 \sigma_w^2 \theta_3^{(\ell)^2} + \sigma_x^2 \sigma_w^2 \theta_3^{(\ell)^2} \theta_2^{(\ell)} - \sigma_x^4 \sigma_w^4 \theta_3^{(\ell)^4}}{(\theta_1^{(\ell)} - \sigma_x^2 \sigma_w^2 \theta_3^{(\ell)^2})(\theta_2^{(\ell)} - \sigma_x^2 \sigma_w^2 \theta_3^{(\ell)^2})}$.

The proof is included in Appendix A which relies on Appendices B and C.

In Figure 2, the empirical eigen-distribution of the matrices $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ at the second and fourth layers is compared with the approximation that is retrieved by considering the Stieltjes transform in Theorem 3.2. As for $\mathbf{\Lambda}_{h^{(\ell)}}$, the second layer's eigen-distribution is visually approximated very accurately, while there remains a visually observable difference for deeper layers, although of a smaller entity than for $\mathbf{\Lambda}_{h^{(\ell)}}$.

3.3 Noise Application

The expectation on which $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ are computed also considers the random vector $\mathbf{n}$. Therefore, as for the input $\mathbf{x}$ it is necessary to consider an input matrix $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_m]$ with $m \to \infty$ and $m/n \to \infty$, and therefore the following cases are considered

$$\Lambda_{\mathbf{h}^{(\ell+1)}} = \frac{\sigma_w^2}{n} \boldsymbol{W}^{(\ell)} \mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)}) \phi(\mathbf{h}^{(\ell)})^\top] \boldsymbol{W}^{(\ell)^\top} + \mathbf{N}\mathbf{N}^\top \tag{39}$$

and

$$\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}} = \mathbf{W}^{(\ell)} \left(\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \frac{1}{\sigma_x^2}\mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top] \right) \mathbf{W}^{(\ell)\top} + \mathbf{N}\mathbf{N}^\top. \tag{40}$$

Crucially, for computational purposes, the noise at the previous layers is assumed to not affect the distribution significantly to be modelled since only small noises are considered.

The Stieltjes transform of the perturbation is

$$G_{\mathbf{N}\mathbf{N}^\top}(z) = \frac{1}{z - \sigma_n^2} \tag{41}$$

and thus by considering the implicit equation (23), the Stieltjes transform of the perturbed matrix is

$$G_{\mathbf{A}+\mathbf{N}\mathbf{N}^\top}(z) = G_{\mathbf{A}}(z - \sigma_n^2). \tag{42}$$

This corresponds to a shift to the right of the distribution of the matrices that are used for the computation of the mutual information. Thanks to this shift the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ are guaranteed of not being singular, and thus allow the computation of the mutual information.

3.4 Estimation of the mutual information

Relying on the determination of the Stieltjes transform of the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ in Sections 3.1 and 3.2 with $\sigma_b = 0$, and on the distribution shift caused by the noise application determined in Section 3.3, it is now possible to numerically compute the lower bound of the mutual information thanks to the property reported in (6) when no bias is considered.

In Figure 3 the empirical mutual information lower bound for different layers of a neural network is compared to the approximation attained with different approaches when σ_w varies, $\sigma_b = 0$, $\sigma_n = 0.1$, and the input is considered with $\sigma_x = 1$. The empirical lower bound[3] is generated by considering 200 initialisations of neural networks with 100-dimensional matrices and with $100,000$ samples. It is possible to notice that the analysis conducted by considering the Stieltjes transform is exact on the second layer, but as soon as the layers increase the bound is over-estimated; however, this approach yields to more accurate approximations of the lower bound introduced in [21]. In Figure 4 the behaviour of the different approximations of the terms $\log(|\mathbf{\Lambda}_{h^{(\ell)}}|)/n$ and $\log(|\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}|)/n$ for the second and fifth layers are compared in the same settings, and it is possible to infer how the error observed in Figure 3 is primarily due to the inaccurate estimation of the log-determinant of $\mathbf{\Lambda}_{h^{(\ell)}}$; thus giving a direction for future improvements.

Similarly to [21], the approximation with the Stieltjes transform of the mutual information lower bound allows to rely on an analytical expression to study the behaviour of the bound at deep layers; thus avoiding running expensive simulations. Therefore, it is possible to obtain results such as those reported in Figure 5, where the bound approximation for very deep layers supports that selecting $\sigma_w = 1$ when $\sigma_b = 0$ and $\phi(\cdot) = tanh(\cdot)$ is optimal from a mutual information perspective. This conclusion follows from how the bound decreases the slowest for this kind of initialisation parameters, consistently with [21]. Furthermore, as it was observed in [21], the same choice of parameters $(\sigma_w, \sigma_b, \phi(\cdot)) = (1, 0, tanh(\cdot))$ is on the edge of chaos [15], thus suggesting that initialisations that are optimal from a training perspective also yield maximal information propagation.

4 Final Considerations

This work explored how the approximation of the lower bound of the mutual information for a neural network at initialisation without bias can be approached from a random matrix perspective; which was shown to consist in modelling the eigen-distribution of the matrices $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$.

In Section 2 a simple approximation of the eigen-spectra with the Marchenko-Pastur distribution was introduced and it was shown that the results with this approximation correspond to those relative to the mean-field approximation in [21]. In Section 3 the eigen-distribution of $\mathbf{\Lambda}_{h^{(\ell)}}$ and $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ was approximated relying on the Stieltjes

[3]The log determinant of the sampled matrices $\log(|\mathbf{\Lambda}_{h^{(\ell)}}|)/n$ and $\log(|\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}|)/n$ was computed considering the following identity $\log|\mathbf{A}| = \log|\mathbf{L}||\mathbf{L}^\top| = 2\log|\mathbf{L}| = 2\log\left(\prod_{j=1}^n \mathbf{L}_{ii}\right) = 2\sum_{i=1}^n \log(\mathbf{L}_{ii})$.

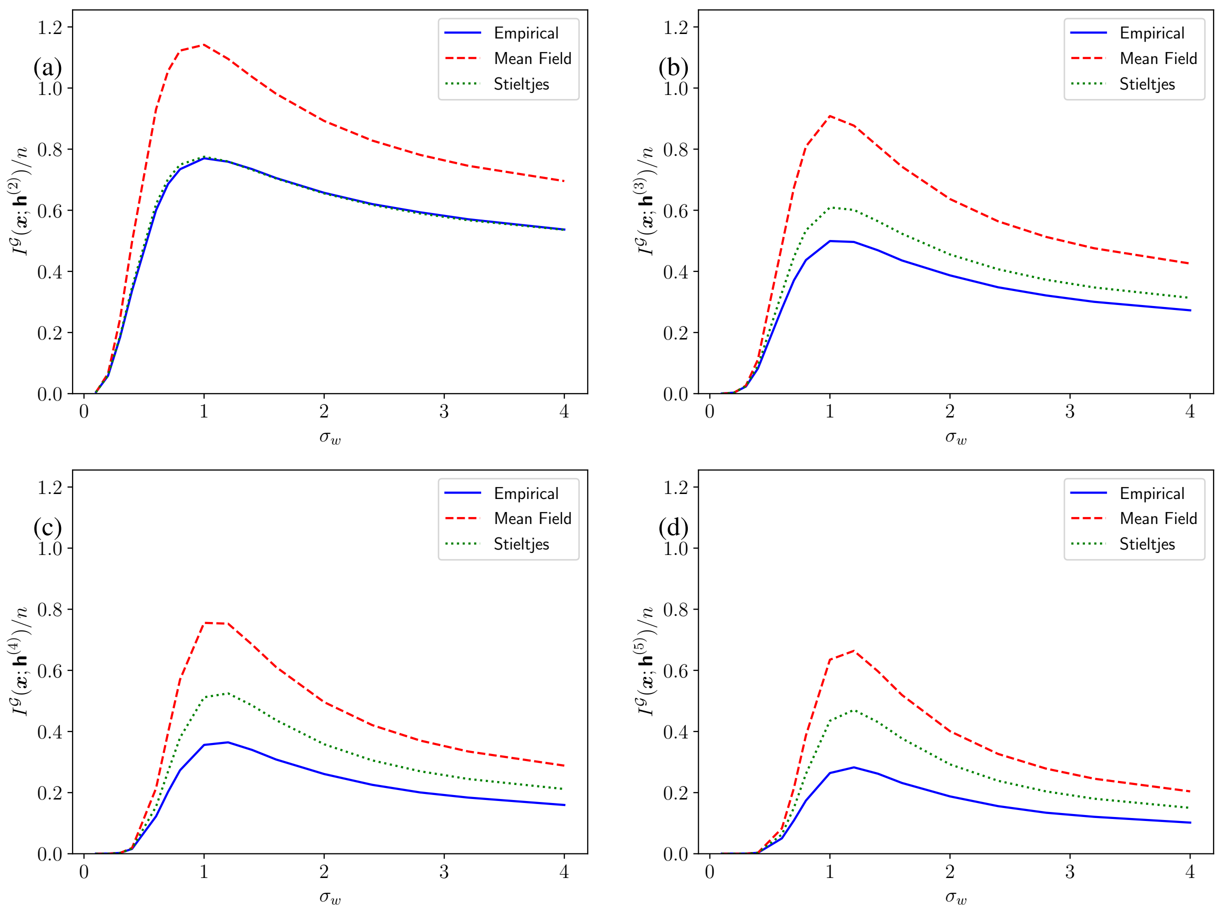


Figure 3: Comparison of the empirical mutual information lower bound between the input and the second (a), third (b), fourth (c), and fifth (d) hidden layer with its approximation with the mean-field approximation and the Stieltjes transform.

transform and its properties for free matrices. Specifically, in Section 3.1 the modelling of the eigen-distribution of $\mathbf{\Lambda}_{h^{(\ell)}}$ in [13, 3] was considered, and in Section 3.2 an approximation for $\mathbf{\Lambda}_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$ was introduced. This approach resulted in an improved estimation of the mutual information in Section 3.4, despite some limitations were identified in the computation of the log-determinant of $\mathbf{\Lambda}_{h^{(\ell)}}$. The Stieltjes transform based method also identified the same behaviour as in [21] for which there is a choice of parameters that optimises the decay of the mutual information lower bound and at the same time is optimal from a training perspective according to the edge of chaos theory [12].

In conclusion, this work showed that it is possible to rely on techniques of random matrix theory to improve on the estimation of the mutual information lower bound introduced in [21]. By expanding this work in the direction that it was proposed within the text, it might be possible to model exactly how the mutual information lower bound decreases through the layers, and this might be beneficial to the understanding of how to train neural networks optimally.

Acknowledgements

This work has been supported by the collaboration between the "Centre for Intelligent Multidimensional Data Analysis" at City University of Hong Kong and the University of Oxford. The authors are extremely grateful to Lucas Benigni, Ewout van den Berg, and Harald Oberhauser for the fruitful discussions.

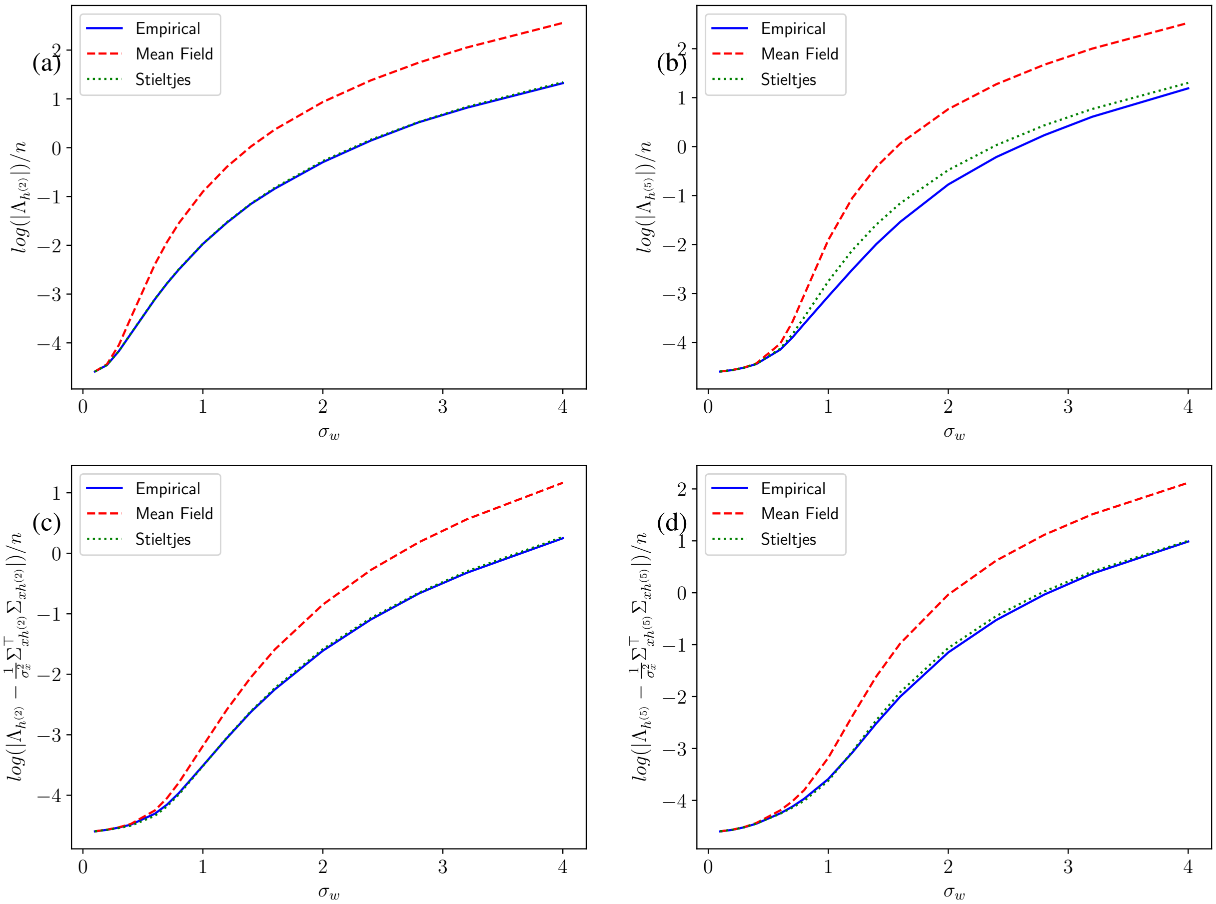


Figure 4: Comparison of the empirical estimation of $\log(|\Lambda_{h^{(\ell)}}|)/n$ at the second (a) and fifth (b) layers, and of $\log(|\Lambda_{h^{(\ell)}} - \frac{1}{\sigma_x^2}\Sigma_{xh^{(\ell)}}^\top \Sigma_{xh^{(\ell)}}|)/n$ at the second (c) and fifth (d) layers with its approximation with the mean-field approximation, and the Stieltjes transform.

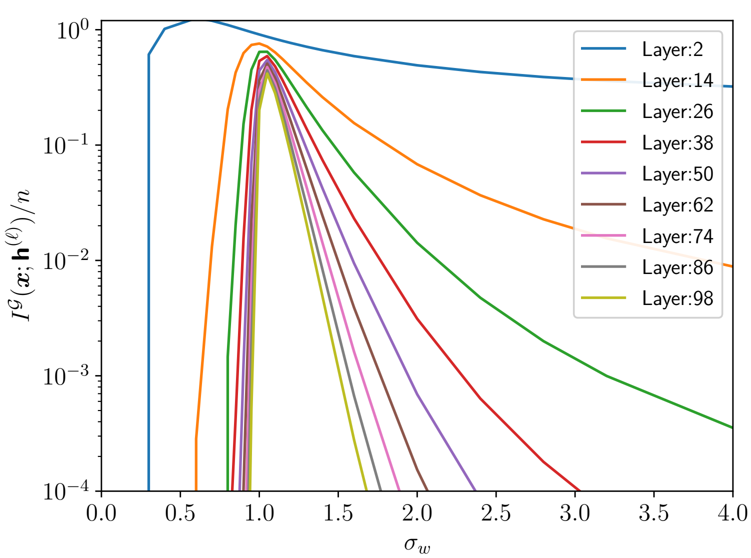


Figure 5: Convergence for increasing layers of the mutual information lower bound approximated with the Stieltjes transform approach.

References

[1] ABROL, V., AND TANNER, J. Information-bottleneck under mean field initialization. *International Conference on Machine Learning (ICML) Workshop on Uncertainty and Robustness in Deep Learning* (2020).

[2] ANECDOTE (HTTPS://MATH.STACKEXCHANGE.COM/USERS/121127/ANECDOTE). Integration over the marchenko-pastur distribution. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/2254998 (version: 2017-04-27).

[3] BENIGNI, L., AND PÉCHÉ, S. Eigenvalue distribution of some nonlinear models of random matrices. *Electronic Journal of Probability 26* (2021), 1–37.

[4] CHOROMANSKA, A., HENAFF, M., MATHIEU, M., AROUS, G. B., AND LECUN, Y. The loss surfaces of multilayer networks. In *Artificial intelligence and statistics* (2015), PMLR, pp. 192–204.

[5] FOGGO, B., AND YU, N. On the maximum mutual information capacity of neural architectures. *arXiv preprint arXiv:2006.06037* (2020).

[6] GLOROT, X., AND BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (2010), JMLR Workshop and Conference Proceedings, pp. 249–256.

[7] GOLDFELD, Z., VAN DEN BERG, E., GREENEWALD, K., MELNYK, I., NGUYEN, N., KINGSBURY, B., AND POLYANSKIY, Y. Estimating information flow in deep neural networks. *International Conference on Machine Learning (ICML)* (2018).

[8] ISSERLIS, L. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika 12*, 1/2 (1918), 134–139.

[9] MARČENKO, V. A., AND PASTUR, L. A. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik 1*, 4 (1967), 457.

[10] MINGO, J. A., AND SPEICHER, R. *Free probability and random matrices*, vol. 35. Springer, 2017.

[11] PENNINGTON, J., AND BAHRI, Y. Geometry of neural network loss surfaces via random matrix theory. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (2017), JMLR. org, pp. 2798–2806.

[12] PENNINGTON, J., SCHOENHOLZ, S. S., AND GANGULI, S. Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017), pp. 4788–4798.

[13] PENNINGTON, J., AND WORAH, P. Nonlinear random matrix theory for deep learning. *Neural Information Processing Systems (NeurIPS)* (2017).

[14] POOLE, B., LAHIRI, S., RAGHU, M., SOHL-DICKSTEIN, J., AND GANGULI, S. Exponential expressivity in deep neural networks through transient chaos. In *Advances in neural information processing systems* (2016), pp. 3360–3368.

[15] POOLE, B., LAHIRI, S., RAGHU, M., SOHL-DICKSTEIN, J., AND GANGULI, S. Exponential expressivity in deep neural networks through transient chaos. *Neural Information Processing Systems (NeurIPS)* (2016).

[16] POTTERS, M., AND BOUCHAUD, J.-P. *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

[17] SAXE, A. M., BANSAL, Y., DAPELLO, J., ADVANI, M., KOLCHINSKY, A., TRACEY, B. D., AND COX, D. D. On the information bottleneck theory of deep learning. *Journal of Statistical Mechanics: Theory and Experiment 2019*, 12 (2019), 124020.

[18] SCHMÜDGEN, K. *The moment problem*, vol. 9. Springer, 2017.

[19] SCHOENHOLZ, S. S., GILMER, J., GANGULI, S., AND SOHL-DICKSTEIN, J. Deep information propagation. *International Conference on Learning Representations (ICLR)* (2017).

[20] SHWARTZ-ZIV, R., AND TISHBY, N. Opening the black box of deep neural networks via information.

[21] TANNER, J., AND UGHI, G. Mutual information of neural network initialisations: Mean field approximations. In *2021 IEEE International Symposium on Information Theory (ISIT)* (2021), pp. 813–818.

[22] WINKELBAUER, A. Moments and absolute moments of the normal distribution. *CoRR abs/1209.4340* (2012).

[23] ZHANG, G., BOTEV, A., AND MARTENS, J. Deep learning without shortcuts: Shaping the kernel with tailored rectifiers. *arXiv preprint arXiv:2203.08120* (2022).

A Proof of Theorem 3.2

The proof of Theorem 3.2 relies on the use of the method of moments detailed in Section A.1. This method allows to identify an implicit relation that the Stieltjes transform has to satisfy and that allows to compute the transform efficiently. Therefore, this method is first used to prove the recursive relation at the second layer in Section A.2, and followingly it is expanded to the subsequent layers in Section A.3. Appendix B and C are in support of these proofs.

A.1 The Method of Moments

The method of moments was used in [13] and [3] to find an implicit recursive relation for the Stieltjes transform of matrix $\mathbf{\Lambda}_{h^{(2)}}$, $G_{\mathbf{\Lambda}_{h^{(2)}}}(z)$[4] in the limit of $n_\ell \to \infty$. Given a general random matrix $\mathbf{M}$, this method aims at evaluating the identity

$$G_{\mathbf{M}}(z) = \sum_{k=0}^{\infty} \frac{\mathfrak{m}_k}{z^{k+1}} \tag{43}$$

by determining exactly the moments, $\mathfrak{m}_k$, of the eigen-distribution $\rho_{\mathbf{M}}$ of $\mathbf{M}$. When considering the matrices relevant to the computation of the mutual information lower bound, the final expression of $\mathfrak{m}_k$ allows to identify the leading contribution and therefore to show that there is a unique measure yielding the identified transform; but the expression is too computationally complex to be computed in practice. However, by computing the moments $\mathfrak{m}_k$ it is possible to identify an implicit relation that the Stieltjes transform has to satisfy and that allows to compute the transform efficiently. Therefore, computing the moments will indirectly allow to identify the Stieltjes transform.

In this work similarly to [13] and [3], the method of moments is applied to matrices $\mathbf{M} \in \mathbb{R}^{n_1 \times n_1}$ that are the outer product of two matrices $\mathbf{M} = \frac{1}{m}\mathbf{Y}\mathbf{Y}^\top$ where the elements of $\mathbf{Y} \in \mathbb{R}^{n_1 \times m}$ that share an index are not independent, i.e. $\mathbf{Y}_{i,j} \perp\!\!\!\perp \mathbf{Y}_{p,q}$ but $\mathbf{Y}_{i,j} \not\perp\!\!\!\perp \mathbf{Y}_{i,q}$ and $\mathbf{Y}_{i,j} \not\perp\!\!\!\perp \mathbf{Y}_{p,j}$ for $i \neq p$ and $j \neq q$. Specifically, in this context the method of moments aims to identify $G(z)$ by considering the limiting moments $\mathfrak{m}_k = \lim_{n\to\infty} \mathfrak{m}_k^{(n)}$ [5] where

$$\mathfrak{m}_k^{(n)} = \frac{1}{n_1}\mathbb{E}[tr(\mathbf{M}^k)] = \frac{1}{n_1}\mathbb{E}\left[\sum_{i_1,\dots,i_k\in[n]} \mathbf{M}_{i_1 i_2}\mathbf{M}_{i_2 i_3}\dots\mathbf{M}_{i_k i_1}\right] \tag{44}$$

$$= \frac{1}{n_1 m^k}\mathbb{E}\left[\sum_{\substack{i_1,\dots,i_k\in[n]\\ \mu_1,\dots,\mu_k\in[m]}} \mathbf{Y}_{i_1\mu_1}\mathbf{Y}_{i_2\mu_1}\mathbf{Y}_{i_2\mu_2}\mathbf{Y}_{i_3\mu_2}\dots\mathbf{Y}_{i_k\mu_k}\mathbf{Y}_{i_1\mu_k}\right] \tag{45}$$

$$= \frac{1}{n_1 m^k}\sum_{\substack{i_1,\dots,i_k\in[n]\\ \mu_1,\dots,\mu_k\in[m]}} \mathbb{E}\left[\mathbf{Y}_{i_1\mu_1}\mathbf{Y}_{i_2\mu_1}\mathbf{Y}_{i_2\mu_2}\mathbf{Y}_{i_3\mu_2}\dots\mathbf{Y}_{i_k\mu_k}\mathbf{Y}_{i_1\mu_k}\right] \tag{46}$$

and these finite dimensional moments $\mathfrak{m}_k^{(n)}$ are computed by associating each sequence of indices $\{(i_1,\mu_1),\dots,(i_1,\mu_k)\}$ to a pattern which is going to be represented with a graph. These graphs are such that despite having different indices, all the addends in (46) sharing the same pattern have the same expected value because of the independence of the weights. Therefore, the addends can be grouped according to their pattern, and, once the contribution of each pattern[6] is established, it is necessary to only quantify the pattern's frequency in the outer sum of (46).

The patterns that arise in (46) are identifiable as graphs by associating the indices, i_ξ and μ_ξ, to the vertex indices of a graph whose edges are defined by $Y_{i_\xi \mu_\xi}$. The work in [13] and [3] proves that (46) is dominated by the terms in the sum whose index pattern is associated to a connected outer-planar graphs in which all blocks are simple even cycles, and these graphs are defined as *admissible graphs*; for details consider Definition A.1. For example, all the admissible patterns for $k = 3$ are shown in Figure 6; all the admissible graphs would consist in permutations of the indices.

Definition A.1 ([13]). *For any positive integer k, a $2k$-cycle, is an admissible graph. Start by labelling the vertices in the $2k$-cycle as $1,\dots,2k$ in a clockwise fashion. Consider any pair of vertices v_1 and v_2 of the same parity, one may*

[4] In the upcoming work the subscript on the Stieltjes transform identifying the corresponding matrix on which the transform is computed is going to be dropped unless the context requires it.

[5] Note that with the expression $n \to \infty$ we assume that all of the layer dimension of the nets are proportional to each other and therefore $n_\ell \to \infty$ for any ℓ.

[6] From now on, the contribution of a pattern or a graph will correspond to the expected value of any term in (46) whose indices define the same pattern or graph.

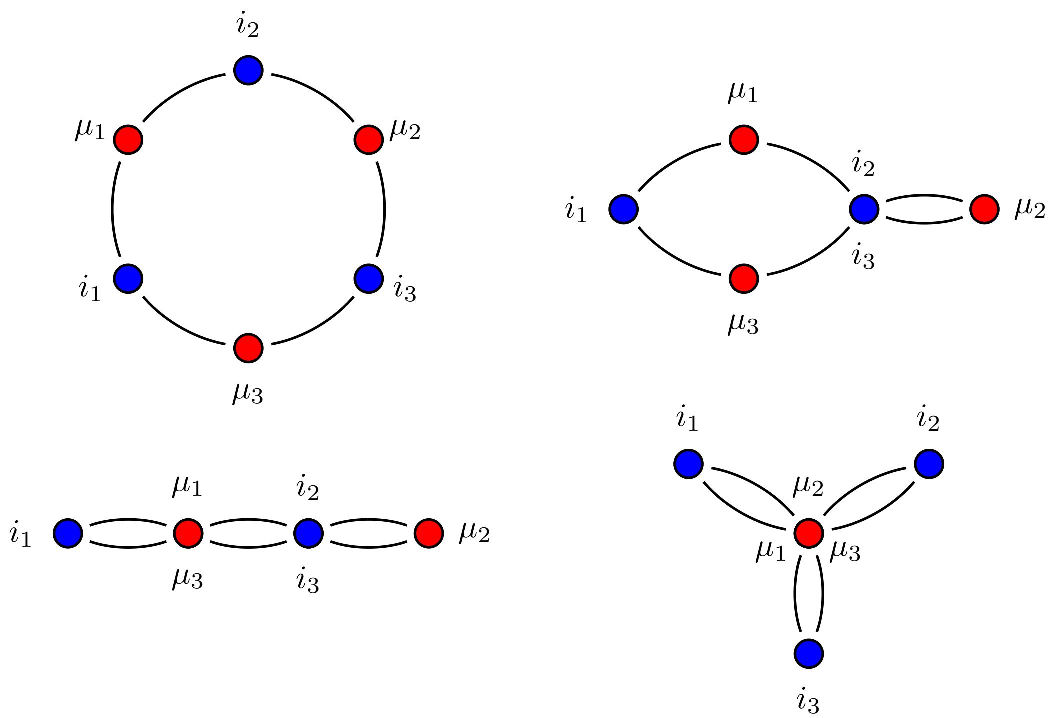


Figure 6: Admissible patterns according to Definition A.1 when considering k=3. Any admissible graph with $k = 3$ consist of a permutation of the indices in the illustrated patterns. This Figures reproduces parts of [13, Figure S1].

obtain an admissible graph from another admissible graph by identifying v_1 and v_2 if there exist two vertex-disjoint paths between v_1 and v_2. The merged vertex is assigned the same parity as that on v_1 and v_2.

The above definition of admissible graphs, allows to compute the moments $\mathfrak{m}_k^{(n)}$ for finite n-dimensional matrices $\mathbf{M}$ by first considering simple cycles and then aggregating their contributions. At leading order, the contribution of a graph is multiplicative in the contributions of its constituent cycles.

When considering the limiting moment $\mathfrak{m}_k = \lim_{n\to\infty} \mathfrak{m}_k^{(n)}$, the explicit formulation of the moments allow to explicitly determine the Stieltjes transform of the matrices relevant to the mutual information lower bound estimation, however, this formulation is too complex to be used in practice. Nevertheless, thanks to the explicit formulation of the Stiletjes transform, it is possible to retrieve an implicit formulation of the Stieltjes transform that can be computed in practice. To find this implicit formulation, two properties of the limiting moments $\mathfrak{m}_k$ are considered: the fact that there exists a generating function based on $\mathfrak{m}_k$ that is related to the limiting Stieltjes transform G, and the fact that the multiplicativity of the constituent cycles used to compute $\mathfrak{m}_k^{(n)}$ is preserved.

A.2 Stieltjes Transform of the Matrix $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(1)})\phi(\mathbf{h}^{(1)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(1)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(1)})^\top]$ with $\sigma_b = 0$ and $\sigma_n = 0$

To rely on the moment of methods as in Section 3.1, it is necessary to study the expectation on the input as an empirical expectation with $m \to \infty$ samples. By considering an input matrix $\mathbf{X}$, the Stieltjes transform of $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(1)})\phi(\mathbf{h}^{(1)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(1)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(1)})^\top]$ is determined by studying the eigen-distribution of

$$\mathbf{M} = \frac{1}{m}\mathbb{E}_{\mathbf{X}}\left[\phi(\mathbf{W}^{(1)}\mathbf{X})\phi(\mathbf{W}^{(1)}\mathbf{X})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}\left[\phi(\mathbf{W}^{(1)}\mathbf{X})\mathbf{X}^\top\right]\mathbb{E}\left[\phi(\mathbf{W}^{(1)}\mathbf{X})\mathbf{X}^\top\right]^\top \tag{47}$$

when $m \to \infty$. The definition of $\mathbf{M}$ already considers the expectation over $\mathbf{X}$ and this is necessary because of the nature of the matrix product $\mathbf{\Sigma}_{xh^{(\ell)}}^\top \mathbf{\Sigma}_{xh^{(\ell)}}$.

Considering the admissible graphs of Definition A.1, shown to be the leading order contribution also in this setting in Appendix B.4, we determine an implicit relation for the Stieltjes transform of $\mathbf{M}$, defined in Theorem A.1. When $m/n \to \infty$ Theorem A.1 holds for the matrix $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$.

Theorem A.1. *Consider the odd activation function* ϕ

$$\left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^k(\sigma_w\sigma_x z)\right| < \infty \quad \text{and} \quad \left|\int \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \phi^{(k)}(\sigma_w\sigma_x z)\right| < \infty \tag{48}$$

for $k > 1$ *with* $\phi^{(k)}$ *being the k-th derivative of* ϕ *and the matrices* $\boldsymbol{W} \in \mathbb{R}^{n_1 \times n_0}$ *and* $\boldsymbol{X} \in \mathbb{R}^{n_0 \times m}$ *with their respective columns being sampled as follows* $\boldsymbol{w}_i \sim \mathcal{N}(\boldsymbol{0}, \frac{\sigma_w^2}{n}\boldsymbol{I})$ *and* $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \sigma_x^2\boldsymbol{I})$. *Then, defining* $\psi = \frac{n_0}{m}$ *and* $\varphi = \frac{n_0}{n_1}$, *the matrix* $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\boldsymbol{Y}^\top\right]$ *with* $\boldsymbol{Y} = \phi(\boldsymbol{W}\boldsymbol{X})$ *has an eigen-distribution whose Stieltjes transform is asymptotically defined by*

$$G_{\boldsymbol{M}}(z) \simeq \frac{1-\varphi}{z} + \frac{\varphi}{z}H(z) \tag{49}$$

where

$$\begin{aligned} H(z) = 1 &+ \frac{H_{\psi b}(z)H_\varphi(z)(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} - \frac{H_{\psi c}(z)H_\varphi(z)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} \\ &+ \frac{H_{\psi c}(z)H_\varphi(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z - H_{\psi c}(z)H_\varphi(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)} \end{aligned} \tag{50}$$

with

$$\theta_1 = \int \frac{1}{\sqrt{2\pi}}\phi(\sigma_w\sigma_x z)^2 e^{-\frac{z^2}{2}}dz, \tag{51}$$

$$\theta_2 = \left(\int \frac{\sigma_w\sigma_x}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\phi'(\sigma_w\sigma_x z)dz\right)^2, \tag{52}$$

$$\theta_3 = \int_{z_1} \frac{\phi'\left(\sigma_x\sigma_w z\right)}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}dz \tag{53}$$

$H_{\psi\alpha}(z) = 1 + \psi\kappa_\alpha(H(z)-1)$, $H_\varphi = 1 + \varphi(H(z)-1)$, $\alpha \in \{b, c\}$, *and* $\kappa_b = 1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}$, $\kappa_c = 1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2}$.

In the following subsections, the proof of Theorem A.1 determines the moments of $\mathbf{M}$ defined according to the following expectation

$$\frac{1}{n}\mathbb{E}_{\mathbf{W}}[tr(\mathbf{M}^k)] = \frac{1}{n}\mathbb{E}_{\mathbf{W}}\left[\sum_{i_1,\ldots,i_k\in[n]} \mathbf{M}_{i_1i_2}\mathbf{M}_{i_2i_3}\ldots\mathbf{M}_{i_ki_1}\right] \tag{54}$$

$$= \frac{1}{nm^k}\mathbb{E}_{\mathbf{W}}\left[\sum_{\substack{i_1,\ldots,i_k\in[n]\\ \mu_1,\ldots,\mu_k\in[m]}} \mathbf{M}^{\mu_1}_{i_1i_2}\mathbf{M}^{\mu_2}_{i_2i_3}\ldots\mathbf{M}^{\mu_k}_{i_ki_1}\right] \tag{55}$$

where

$$\begin{aligned} \mathbf{M}^{\mu_1}_{i_1i_2} &= \mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l\mu_1})\phi(\sum_l \mathbf{W}_{i_2,l}\mathbf{X}_{l\mu_1})\right] \\ &- \frac{1}{m\sigma_x^2}\sum_{p,q=1}^{n,m}\mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l\mu_1})\mathbf{X}_{p\mu_1}\right]\mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}_{i_2,l}\mathbf{X}_{lq})\mathbf{X}_{pq}\right]. \end{aligned} \tag{56}$$

In Subsection A.2.1 the contribution of single cycle patterns are determined and in Subsection A.2.2 they are aggregated to identify the contribution to the moments of an admissible graph. This allows to define a generating function and to determine the contribution of two graphs that are connected via a vertex. These observations are then used in Subsection A.2.3 to prove Theorem A.1.

An interesting consequence on the Theorem A.1 being applied at the first layer is reported in Corollary A.1.1 whose proof is included in Appendix B.5.

Corollary A.1.1. *For* $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ *and* $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ *as defined in Theorem A.1, the eigen-distribution of the matrix* $\boldsymbol{W}\boldsymbol{M}\boldsymbol{W}^\top$ *asymptotically follows the Marchenko-Pastur distribution with mean* $\sigma_w^2(\theta_1 - \theta_2)$ *and shape* $\gamma = \frac{n_1}{n_0}$ *when* $\psi = 0$ *and* $\varphi = 1$.

A.2.1 $2k$-cycle Contributions

Similarly the heterogeneous input proof, the contribution of cycles of length $2k$ is first computed to then study the contribution of any admissible graph. We thus start by considering cycles of length $2k$, i.e. in Equation (55) $i_1 \neq i_2 \neq ... \neq i_k$ and $\mu_1 \neq \mu_2 \neq ... \neq \mu_k$.

$$E_{2k}^{(n)} = \mathbb{E}_{\mathbf{W}}\left[\mathbf{M}_{i_1 i_2}^{\mu_1}\mathbf{M}_{i_2 i_3}^{\mu_2}...\mathbf{M}_{i_k i_1}^{\mu_k}\right] \tag{57}$$

$$\begin{aligned}
&= \int_{\mathbf{W}}\Bigg\{\Bigg(\int \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l\mu_1})\phi(\sum_l \mathbf{W}_{i_2,l}\mathbf{X}_{l\mu_1})\mathcal{D}\mathbf{X} \\
&\quad -\frac{1}{m\sigma_x^2}\sum_{p,q=1}^{n,m}\int \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l\mu_1})\mathbf{X}_{p\mu_1}\mathcal{D}\mathbf{X}\int \phi(\sum_l \mathbf{W}_{i_2,l}\mathbf{X}_{lq})\mathbf{X}_{pq}\mathcal{D}\mathbf{X}\Bigg) \\
&\quad ... \\
&\quad \Bigg(\int \phi(\sum_l \mathbf{W}_{i_k,l}\mathbf{X}_{l,\mu_k})\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_k})\mathcal{D}\mathbf{X} \\
&\quad -\frac{1}{m\sigma_x^2}\sum_{p,q=1}^{n,m}\int \phi(\sum_l \mathbf{W}_{i_k,l}\mathbf{X}_{l\mu_k})\mathbf{X}_{p\mu_k}\mathcal{D}\mathbf{X}\int \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{lq})\mathbf{X}_{pq}\mathcal{D}\mathbf{X}\Bigg)\Bigg\}\mathcal{D}\mathbf{W}
\end{aligned} \tag{58}$$

It is possible to reformulate the right addend thanks to the following identity

$$\int_{\mathbf{X}} \phi(\sum_{l\neq\beta}\mathbf{W}_{\alpha l}\mathbf{X}_{lp} + \mathbf{W}_{\alpha\beta}\mathbf{X}_{\beta,p})\mathbf{X}_{\beta,p}\mathcal{D}\mathbf{X} = \int_{\mathbf{X}} \sigma_x^2\phi'(\sum_l \mathbf{W}_{\alpha l}\mathbf{X}_{lp})\mathbf{W}_{\alpha\beta}\mathcal{D}\mathbf{X} \tag{59}$$

$$= \sigma_x^2\mathbf{W}_{\alpha\beta}\int \phi'\left(\sigma_x\sigma_w z\right)\mathcal{D}z = \sigma_x^2\mathbf{W}_{\alpha\beta}\theta_3 \tag{60}$$

where $\mathcal{D}z$ is the standard normal measure.

This leads to the following expression

$$E_{2k}^{(n)} = \int_{\mathbf{W}}\Bigg\{\Bigg(\int \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l\mu_1})\phi(\sum_l \mathbf{W}_{i_2,l}\mathbf{X}_{l\mu_1})\mathcal{D}\mathbf{X} - \frac{\sigma_x^2}{m}\sum_{p,q=1}^{n,m}\mathbf{W}_{i_1 p}\mathbf{W}_{i_2 p}\theta_3^2\Bigg) \tag{61}$$

$$...$$

$$\Bigg(\int \phi(\sum_l \mathbf{W}_{i_k,l}\mathbf{X}_{l,\mu_k})\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_k})\mathcal{D}\mathbf{X} - \frac{\sigma_x^2}{m}\sum_{p,q=1}^{n,m}\mathbf{W}_{i_k p}\mathbf{W}_{i_1 p}\theta_3^2\Bigg)\Bigg\}\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{62}$$

$$= \int_{W,X}\left\{\prod_{\xi=1}^{k}\left(\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi}) - \sigma_x^2\sum_{p=1}^{n}\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}p}\theta_3^2\right)\right\}\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}. \tag{63}$$

In the case of $k = 1$, the integral can be split in two due to its linearity, and this leads to

Lemma A.2. *For* $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ *and* $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ *as defined in Theorem A.1, when* $k = 1$ *then contribution of an admissible cycle is*

$$E_2^{(n)} = \theta_1\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) - \sigma_w^2\sigma_x^2\theta_3^2 \tag{64}$$

with

$$\theta_1 = \int \frac{1}{\sqrt{2\pi}}\phi(\sigma_w\sigma_x z)^2 e^{-\frac{z^2}{2}}dz, \tag{65}$$

$$\theta_3 = \int \frac{1}{\sqrt{2\pi}}\phi'\left(\sigma_x\sigma_w z\right)e^{-\frac{z^2}{2}}dz. \tag{66}$$

Proof. See proof in Section B.1. □

However, when considering $k > 1$, we have to consider also mixed products. More complex considerations allow to determine the contribution of a general $2k$-cycle.

Lemma A.3. *For $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\,\boldsymbol{X}\right)$ as defined in Theorem A.1, when $k > 1$ then contribution of an admissible cycle is*

$$E_{2k}^{(n)} = n_0^{1-k}\left(\theta_2 - \sigma_w^2\sigma_x^2\theta_3^2\right)^k\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{67}$$

with

$$\theta_2 = \left(\int \frac{\sigma_w\sigma_x}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\phi'(\sigma_w\sigma_x z)z\right)^2, \tag{68}$$

$$\theta_3 = \int \frac{1}{\sqrt{2\pi}}\phi'\left(\sigma_x\sigma_w z\right)e^{-\frac{z^2}{2}}dz \tag{69}$$

Proof. See proof in Section B.2. □

A.2.2 Agglomeration of the Contributions

To compute the moments of the distribution, we follow the exact same process as in the heterogeneous case for $\boldsymbol{\Lambda}_{h^l}$ where we identify all of the possible patterns, identify their cardinality (each pattern can be achieved by selecting different vertices), and their Expected contribution thanks to the previous subsection.

To count the different patterns, after naming I_i and I_μ the identifications in Definition A.1 of respectively odd and even parity, the following variable is considered similarly to [13, 3]:

Definition A.2. *$\mathcal{C}(k, I_i, I_\mu, b, b^\mu, c^\mu)$ is the number of admissible pattern with $2k$ edges, I_i i-identifications, $I_\mu = b^\mu + c^\mu$ μ-identifications where c^μ are the identifications that do not define any two dimensional cycle, and with exactly b cycles of size 2, similarly to [13, 3].*

Therefore, it is possible to compute the k-th moment of the eigen-distribution

Proposition 2. *For $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, the k-th moment of the eigen-distribution, for $k > 1$, is*

$$\mathfrak{m}_k^{(n)} = \mathfrak{m}_k\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{70}$$

with

$$\mathfrak{m}_k = \lim_{n\to\infty}\mathfrak{m}_k^{(n)} \tag{71}$$

$$= \varphi^{1-k}\sum_{I_i, I_\mu = 1}^{k}\;\sum_{b=0}^{I_i + I_\mu + 1}\;\sum_{b^\mu = 0}^{b}\frac{\mathcal{C}(k, I_i, I_\mu, b, b^\mu, c^\mu)}{2\pi\sqrt{(k - I_i)(k - I_\mu)}}(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^b\cdot$$

$$\cdot\,(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^{k-b}\kappa_b^{b^\mu}\kappa_c^{I_\mu - b^\mu}\varphi^{I_1}\psi^{I_\mu}\left(\frac{e}{k - I_i}\right)^{k - I_1}\left(\frac{e}{k - I_\mu}\right)^{k - I_\mu} \tag{72}$$

$$= \varphi^{1-k}H_k \tag{73}$$

where we define $\kappa_b = 1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}$, $\kappa_c = 1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2}$, and based on these finally $H_k = \sum_{I_i, I_\mu = 1}^{k}\sum_{b=0}^{I_i + I_\mu + 1}\sum_{b^\mu = 0}^{b}\mathcal{C}(k, I_i, I_\mu, b, b^\mu, c^\mu)(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^b(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^{k-b}\kappa_b^{b^\mu}\kappa_c^{c^\mu}\varphi^{I_1}\phi^{I_\mu}$.

Proof. See proof in Section B.3. □

This allows to compute the Stieltjes transform by introducing the generating function H.

Corollary A.3.1. *For $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, the Stieltjes transform of the eigen-distribution of $\boldsymbol{M}$ asymptotically satisfies the following equation*

$$G_{\boldsymbol{M}}(z) \simeq \frac{1-\varphi}{z} + \frac{\varphi}{z}H(z) \tag{74}$$

where $H(z)$ is the generating function defined as $H(z) = 1 + \sum_{k=1}^{\infty}\frac{H_k}{z^k\varphi^k} = 1 + \frac{1}{\varphi}\sum_{k=1}^{\infty}\frac{\mathfrak{m}_k}{z^k}$.

Proof. To start with we notice that using the definition of the Stieltjes transform (43)

$$G_{\mathbf{M}}(z) = \sum_{k=0}^{\infty}\frac{\mathfrak{m}_k^{(n)}}{z^{k+1}} = \sum_{k=0}^{\infty}\frac{\mathfrak{m}_k}{z^{k+1}}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{75}$$

$$= G(z)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{76}$$

By relying on equations (43) and (71)-(73)

$$G(z) = \frac{1}{z} + \sum_{k=1}^{\infty}\frac{\mathfrak{m}_k}{z^{k+1}} = \frac{1}{z} + \frac{\varphi}{z}\sum_{k=1}^{\infty}\frac{H_k}{z^k\varphi^k} \tag{77}$$

$$= \frac{1-\varphi}{z} + \frac{\varphi}{z}\left(1+\sum_{k=1}^{\infty}\frac{H_k}{z^k\varphi^k}\right) = \frac{1-\varphi}{z} + \frac{\varphi}{z}H(z). \tag{78}$$

□

Moreover, from Proposition 2 it follows that the total contribution to the moment of a graph which is defined by the junction of two graphs, is equal to the contribution of each defining block with a penalty term that depends on the identification joining the two graphs.

Corollary A.3.2. *Consider $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, and consider a graph G_3 with $2(p+q)$ edges that is defined, via etiher an i or μ identification, by two graphs G_1 and G_2 with $2p$ and $2q$ edges. Then the asymptotic contribution, $\mathfrak{m}_{p+q}^{G_3}$, of the graph G_3 to the moment $\mathfrak{m}_{p+q}$ is equal to the product of the contributions of the block cycle graphs that define G_3 with a correction term. This is, they are joined with an i-identification*

$$\mathfrak{m}_{p+q}^{G_3} = \mathfrak{m}_p^{G_1}\frac{n_1}{n_1}\mathfrak{m}_p^{G_2} \tag{79}$$

and if they are joined with a μ-identification

$$\mathfrak{m}_{p+q}^{G_3} = \mathfrak{m}_p^{G_1}\frac{n_1\kappa_\alpha}{m}\mathfrak{m}_p^{G_2} \tag{80}$$

where if either one of the block cycles connected to the identification are of dimension 2, then $\alpha = b$ or otherwise $\alpha = c$, and $\kappa_b = 1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)}$, $\kappa_c = 1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^2}$.

A.2.3 Proof of the Implicit Relation in Theorem A.1

Proof. As shown in Corollary A.3.2, a graph G_3, that is defined by a graph G_1 and G_2, contributes to the total moment according to the following identity.

$$\mathfrak{m}_{q+p}^{G_3} = \mathfrak{m}_q^{G_1}\frac{n_1}{n_1}\mathfrak{m}_p^{G_2} \tag{81}$$

when considering an i-identification or

$$\mathfrak{m}_{q+p}^{G_3} = \mathfrak{m}_q^{G_1}\frac{n_1\kappa_\alpha}{m}\mathfrak{m}_p^{G_2} \tag{82}$$

when considering a μ-identification where $\alpha = b$ if either of the adjacent blocks to the identifications is a 2 dimensional cycle, or $\alpha = c$ otherwise.

If a vertex v is fixed, it is possible to use the moment generating function to define the contributions of all the possible patterns that include the chosen vertex. For each of these patterns, the vertex is going to be in a cycle of dimension 2ℓ, with a contribution $\mathfrak{m}_\ell^{2\ell-cycle}$, and to each of the remaining vertices either nothing or a further graph is connected with a correction depending on the identification. By considering the limiting variables $E_{2\ell} = \lim_{n\to\infty} E_{2\ell}^{(n)}$, the total contribution, $C_{2\ell}$, to the different moments for graphs in which the vertex v is included in a 2ℓ cycle is defined by

$$C_{2\ell} = \frac{\mathfrak{m}_\ell^{2\ell-cycle}}{z^\ell}\left(1+\sum_{G_\alpha}\frac{\mathfrak{m}_{|G_\alpha|}^{G_\alpha}}{z^{|G_\alpha|}}\right)^\ell\left(1+\sum_{G_\alpha}\frac{\kappa_\alpha n_1}{m}\frac{\mathfrak{m}_{|G_\alpha|}^{G_\alpha}}{z^{|G_\alpha|}}\right)^\ell \tag{83}$$

$$= \frac{1}{n_1 m^\ell z^\ell}E_{2\ell}n_1^{\ell-1}m^\ell\left(1+\sum_{G_\alpha}\frac{\mathfrak{m}_{|G_\alpha|}^{G_\alpha}}{z^{|G_\alpha|}}\right)^\ell\left(1+\sum_{G_\alpha}\frac{\kappa_\alpha n_1}{m}\frac{\mathfrak{m}_{|G_\alpha|}^{G_\alpha}}{z^{|G_\alpha|}}\right)^\ell \tag{84}$$

$$= \frac{1}{n_1 m^\ell z^\ell}E_{2\ell}n_1^{\ell-1}m^\ell\left(1+\frac{n_1}{n_1}\varphi(H(z)-1)\right)^\ell\left(1+\frac{\kappa_\alpha n_1}{m}\varphi(H(z)-1)\right)^\ell \tag{85}$$

$$= \frac{1}{n_1 m^\ell z^\ell}E_{2\ell}n_1^{\ell-1}m^\ell\left(1+\varphi(H(z)-1)\right)^\ell\left(1+\kappa_\alpha\psi(H(z)-1)\right)^\ell \tag{86}$$

$$= \frac{1}{n_1 m^\ell z^\ell}E_{2\ell}n_1^{\ell-1}m^\ell H_\varphi^\ell H_{\psi\alpha}^\ell \tag{87}$$

where G_α is any possible graph, $|G_\alpha|$ is its dimension, and the following definition of the generating function was used

$$H(z)-1 = \frac{1}{\varphi}\sum_{G_\alpha}\frac{\mathfrak{m}_{|G_\alpha|}^{G_\alpha}}{z^{|G_\alpha|}}, \tag{88}$$

the following variables were introduced

$$H_\varphi(z) = 1+\varphi(H(z)-1) \text{ and } H_{\psi\alpha}(z) = 1+\kappa_\alpha\psi(H(z)-1), \tag{89}$$

and κ_α was considered to be dependent uniquely on the cycle that contains vertex v is part of. If $\ell > 1$, then $\alpha = b$ only if the block of G_α to which the cycle is attached is of dimension 2. However, the cardinality of the connection to these kinds of graphs is of a lower order than to any other kind, thus, for any cycle with $\ell > 1$ it is going to be assumed that $\alpha = c$.

Therefore, by summing the contribution $C_{2\ell}$ over all the possible fixed vertices v and over all the possible cycle dimensions ℓ, the moment generating function is retrieved.

$$\varphi(H(z)-1) = \sum_k\frac{\mathfrak{m}_k}{z^k} = \sum_{i=1}^{n_1}\sum_{\ell=1}^{\infty}C_{2\ell} \tag{90}$$

$$= \sum_{\ell=1}^{\infty}\frac{E_{2\ell}n_1^{\ell-1}}{z^\ell}H_{\psi\alpha}(z)^\ell H_\varphi(z)^\ell \tag{91}$$

$$= \frac{\theta_1-\sigma_x^2\sigma_w^2\theta_3^2}{z}H_{\psi b}(z)H_\varphi(z) + \sum_{\ell=2}^{\infty}n_0^{1-\ell}\frac{(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^\ell}{z^\ell}n_1^{\ell-1}H_{\psi c}(z)^\ell H_\varphi(z)^\ell \tag{92}$$

$$= \frac{\theta_1-\sigma_x^2\sigma_w^2\theta_3^2}{z}H_{\psi b}(z)H_\varphi(z) + \sum_{\ell=2}^{\infty}\frac{\varphi^{1-\ell}(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^\ell}{z^\ell}H_{\psi c}(z)^\ell H_\varphi(z)^\ell. \tag{93}$$

Therefore the recursive formula is

$$H(z) = 1+\frac{H_{\psi b}(z)H_\varphi(z)(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} + \sum_{q_0=2}^{\infty}\left(\frac{H_{\psi c}(z)H_\varphi(\theta_2-\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z}\right)^{q_0} \tag{94}$$

$$= 1+\frac{H_{\psi b}(z)H_\varphi(z)(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} - \frac{H_{\psi c}(z)H_\varphi(z)(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} + \sum_{q_0=1}^{\infty}\left(\frac{H_{\psi c}(z)H_\varphi(\theta_2-\theta_3)}{\varphi z}\right)^{q_0} \tag{95}$$

$$= 1+\frac{H_{\psi b}(z)H_\varphi(z)(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} - \frac{H_{\psi c}(z)H_\varphi(z)(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} + \frac{1}{1-\left(\frac{H_{\psi c}(z)H_\varphi(\theta_2-\theta_3)}{\varphi z}\right)} - 1 \tag{96}$$

$$= 1 + \frac{H_{\psi b}(z)H_{\varphi}(z)(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} - \frac{H_{\psi c}(z)H_{\varphi}(z)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} + \frac{\varphi z}{\varphi z - H_{\psi c}(z)H_{\varphi}(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)} - 1 \tag{97}$$

$$= 1 + \frac{H_{\psi b}(z)H_{\varphi}(z)(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} - \frac{H_{\psi c}(z)H_{\varphi}(z)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z} + \frac{H_{\psi c}(z)H_{\varphi}(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}{\varphi z - H_{\psi c}(z)H_{\varphi}(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)} \tag{98}$$

□

A.3 Stieltjes Transform of the Matrix $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$ with $\sigma_b = 0$ and $\sigma_n = 0$

In this section, the matrix $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(\ell)})\phi(\mathbf{h}^{(\ell)})^\top] - \mathbb{E}_x[\phi(\mathbf{h}^{(\ell)})\mathbf{x}^\top]\mathbb{E}_x[\mathbf{x}\phi(\mathbf{h}^{(\ell)})^\top]$ is studied at deeper layers following a similar argument to the one for $\mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(1)})\phi(\mathbf{h}^{(1)})^\top] - \mathbb{E}_{\mathbf{x}}[\phi(\mathbf{h}^{(1)})\mathbf{x}^\top]\mathbb{E}_{\mathbf{x}}[\mathbf{x}\phi(\mathbf{h}^{(1)})^\top]$. Therefore we study the following matrix

$$\mathbf{M} = \frac{1}{m}\mathbb{E}_{\mathbf{X}}\left[\phi(\mathbf{W}^{(\ell)}\mathbf{Y}^{(\ell)})\phi(\mathbf{W}^{(\ell)}\mathbf{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\mathbf{x}}\left[\phi(\mathbf{W}^{(\ell)}\mathbf{Y}^{(\ell)})\mathbf{X}^\top\right]\mathbb{E}_{\mathbf{x}}\left[\phi(\mathbf{W}^{(\ell)}\mathbf{Y}^{(\ell)})\mathbf{X}^\top\right]. \tag{99}$$

The consequent Stieltjes transform introduced in Theorem 3.2 however has to rely on the assumption that all the elements in the hidden layer are independent since it is not amenable to keeping track of the covariance through the layers, as shown in Lemma A.5.

The proof of Theorem 3.2 relies on the proof of Theorem A.1 to implement the moments method. This subsection will focus on implementing the method of moments by considering the following expectation

$$\frac{1}{n}\mathbb{E}_{\mathbf{W}}[tr(\mathbf{M}^k)] = \frac{1}{n}\mathbb{E}_{\mathbf{W}}\left[\sum_{i_1,\dots,i_k\in[n]}\mathbf{M}_{i_1i_2}\mathbf{M}_{i_2i_3}\dots\mathbf{M}_{i_ki_1}\right] \tag{100}$$

$$= \frac{1}{nm^k}\mathbb{E}_{\mathbf{W}}\left[\sum_{\substack{i_1,\dots,i_k\in[n]\\ \mu_1,\dots,\mu_k\in[m]}}\mathbf{M}^{\mu_1}_{i_1i_2}\mathbf{M}^{\mu_2}_{i_2i_3}\dots\mathbf{M}^{\mu_k}_{i_ki_1}\right] \tag{101}$$

where

$$\mathbf{M}^{\mu_1}_{i_1i_2} = \mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l}\mathbf{Y}^{(\ell)}_{l\mu_1})\phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l}\mathbf{Y}^{(\ell)}_{l\mu_1})\right] - \frac{1}{m\sigma_x^2}\sum_{p,q=1}^{n,m}\mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l}\mathbf{Y}^{(\ell)}_{l\mu_1})\mathbf{X}_{p\mu_1}\right]\mathbb{E}_{\mathbf{X}}\left[\phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l}\mathbf{Y}^{(\ell)}_{lq})\mathbf{X}_{pq}\right]. \tag{102}$$

In Section A.3.1 the contribution of simple cycles is computed and it is shown how it is necessary to include the assumption of the elements in the hidden layer being independent, i.e. $\mathbf{Y}^{(\ell)}_{:p} \sim \mathcal{N}(\mathbf{0}, q^{(\ell)}\mathbf{I})$, to retrieve a tractable analytical formulation. Once this assumption is considered, then in Section A.3.2 two updated Lemmas from Section A.2.2 are presented and their expression allows to state that the proof of Theorem 3.2 is the same as for Theorem A.1. The results in Section A.3.2 are not proven explicitly since their proofs trivially consist in renaming some variable in the proofs for Theorem A.1 .

A.3.1 $2k$-cycle Contributions

The contribution of cycles of length $2k$ is first computed to then study the contribution of any admissible graph. We then consider cycles of length $2k$, i.e. in Equation (101) $i_1 \neq i_2 \neq \dots \neq i_k$ and $\mu_1 \neq \mu_2 \neq \dots \neq \mu_k$

$$E^{(\ell,n)}_{2k} = \mathbb{E}_{\mathbf{W}}\left[\mathbf{M}^{\mu_1}_{i_1i_2}\mathbf{M}^{\mu_2}_{i_2i_3}\dots\mathbf{M}^{\mu_k}_{i_ki_1}\right] \tag{103}$$

$$
\begin{aligned}
&= \int_{\mathbf{W}} \Bigg\{ \Bigg(\int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \mathcal{D}\mathbf{X} \\
&\quad - \frac{1}{m\sigma_x^2} \sum_{p,q=1}^{n,m} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \mathbf{X}_{p\mu_1} \mathcal{D}\mathbf{X} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l} \mathbf{Y}^{(\ell)}_{lq}) \mathbf{X}_{pq} \mathcal{D}\mathbf{X} \Bigg) \\
&\quad \dots \\
&\quad \Bigg(\int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_k,l} \mathbf{Y}^{(\ell)}_{l\mu_k}) \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_k}) \mathcal{D}\mathbf{X} \\
&\quad - \frac{1}{m\sigma_x^2} \sum_{p,q=1}^{n,m} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_k,l} \mathbf{Y}^{(\ell)}_{l\mu_k}) \mathbf{X}_{p\mu_k} \mathcal{D}\mathbf{X} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{lq}) \mathbf{X}_{pq} \mathcal{D}\mathbf{X} \Bigg) \Bigg\} \mathcal{D}\mathbf{W} \qquad (104)
\end{aligned}
$$

This can be rewritten thanks to the following identity.

$$
\int_{\mathbf{X}} \phi(\mathbf{H}^{(\ell)}_{\alpha p}) \mathbf{X}_{\beta p} \mathcal{D}\mathbf{X}_{\beta p} = \int_{\mathbf{X}} \phi(\mathbf{H}^{(\ell)}_{\alpha p}) \mathbf{X}_{\beta p} e^{-\frac{\mathbf{X}^2_{\beta p}}{2\sigma_x^2}} d\mathbf{X} \qquad (105)
$$

$$
= -\sigma_x^2 \left[\phi(\mathbf{H}^{(\ell)}_{\alpha p}) \mathbf{X}_{\beta p} e^{-\frac{\mathbf{X}^2_{\beta p}}{2\sigma_x^2}} \right]_{-\infty}^{+\infty} + \sigma_x^2 \int_{\mathbf{X}} \frac{\partial \left(\phi(\mathbf{H}^{(\ell)}_{\alpha p}) \right)}{\partial \mathbf{X}_{\beta p}} e^{-\frac{\mathbf{X}^2_{\beta p}}{2\sigma_x^2}} d\mathbf{X}_{\beta p} \qquad (106)
$$

$$
= \sigma_x^2 \int_{\mathbf{X}} \phi'(\mathbf{H}^{(\ell)}_{\alpha p}) \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \frac{\partial \left(\phi(\mathbf{H}^{(\ell-1)}_{k_\ell p}) \right)}{\partial \mathbf{X}_{\beta p}} \mathcal{D}\mathbf{X} \qquad (107)
$$

$$
= \sigma_x^2 \int_{\mathcal{X}} \phi'(\mathbf{H}^{(\ell)}_{k_\ell p}) \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \left(\prod_{j=1}^{\ell-2} \left(\phi'(\mathbf{H}^{(\ell-j)}_{k_{\ell-j+1} p}) \sum_{k_{\ell-j}} \mathbf{W}^{(\ell-j)}_{k_{\ell-j+1} k_{\ell-j}}) \right) \right) \frac{\partial \left(\phi(\mathbf{H}^{(1)}_{\alpha p}) \right)}{\partial \mathbf{X}_{\beta p}} \mathcal{D}\mathbf{X} \qquad (108)
$$

$$
= \sigma_x^2 \int_{\mathbf{X}} \phi'(\mathbf{H}^{(\ell)}_{k_\ell p}) \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \left(\prod_{j=1}^{\ell-2} \left(\phi'(\mathbf{H}^{(\ell-j)}_{k_{\ell-j+1} p}) \sum_{k_{\ell-j}} \mathbf{W}^{(\ell-j)}_{k_{\ell-j+1} k_{\ell-j}}) \right) \right) \phi'(\mathbf{H}^{(1)}_{k_1 p}) \mathbf{W}^{(1)}_{k_2 \beta} \mathcal{D}\mathbf{X} \qquad (109)
$$

$$
= \sigma_x^2 \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \left(\prod_{j=1}^{\ell-2} \left(\sum_{k_{\ell-j}} \mathbf{W}^{(\ell-j)}_{k_{\ell-j+1} k_{\ell-j}}) \right) \right) \mathbf{W}^{(1)}_{k_2 \beta} \int_{\mathbf{X}} \left(\prod_{j=1}^{\ell} \phi'(\mathbf{H}^{(\ell-j)}_{k_{\ell-j+1} p}) \right) \mathcal{D}\mathbf{X} \qquad (110)
$$

$$
= \sigma_x^2 \mathfrak{W}^{(\ell)}_{\alpha\beta} \int_{\mathbf{X}} \left(\prod_{j=1}^{\ell} \phi'(\mathbf{H}^{(\ell-j)}_{k_{\ell-j+1} p}) \right) \mathcal{D}\mathbf{X} \qquad (111)
$$

where the new variable $\mathfrak{W}^{(\ell)}_{\alpha\beta} = \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \left(\prod_{j=1}^{\ell-2} \left(\sum_{k_{\ell-j}} \mathbf{W}^{(\ell-j)}_{k_{\ell-j+1} k_{\ell-j}}) \right) \right) \mathbf{W}^{(1)}_{k_2 \beta} = \sum_{k_\ell} \mathbf{W}^{(\ell)}_{\alpha k_\ell} \mathfrak{W}^{(\ell-1)}_{k_\ell \beta}$ is introduced and the symmetry of the problem for which $\mathbf{H}^{(\ell+1)}_{\alpha\beta} \sim \mathbf{H}^{(\ell)}_{\gamma\delta}$ for any α, β, γ, and δ, is used to isolate the terms in $\int_{\mathbf{X}} \left(\prod_{j=1}^{\ell} \phi'(\mathbf{H}^{(\ell-j)}_{k_{\ell-j+1} p}) \right) \mathcal{D}\mathbf{X}$.

Therefore, by defining

$$
\theta_3^{(\ell)} = \int_{\mathbf{X}} \phi'(\mathbf{H}^{(\ell)}_{\alpha\beta}) \mathcal{D}\mathbf{X} = \int \phi'(\sqrt{q^{(\ell)}} z \mathcal{D}z)
$$

we can consider

$$
\begin{aligned}
\mathbf{M}^{\mu_1}_{i_1 i_2} &= \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \mathcal{D}\mathbf{X} \\
&\quad - \frac{1}{m\sigma_x^2} \sum_{p,q=1}^{n,m} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \mathbf{X}_{p\mu_1} \mathcal{D}\mathbf{X} \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l} \mathbf{Y}^{(\ell)}_{lq}) \mathbf{X}_{pq} \mathcal{D}\mathbf{X} \qquad (112) \\
&= \int \phi(\sum_l \mathbf{W}^{(\ell)}_{i_1,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \phi(\sum_l \mathbf{W}^{(\ell)}_{i_2,l} \mathbf{Y}^{(\ell)}_{l\mu_1}) \mathcal{D}\mathbf{X}
\end{aligned}
$$

$$- \frac{\sigma_x^2}{m} \sum_{p,q=1}^{n,m} \mathfrak{W}_{i_1 p}^{(\ell)} \int_{\mathbf{X}} \left(\prod_{j=1}^{\ell} \phi'(\mathbf{H}_{k_\ell \mu_1}^{(\ell-1)}) \right) \mathcal{D}\mathbf{X} \mathfrak{W}_{i_2 p}^{(\ell)} \int_{\mathbf{X}} \left(\prod_{j=1}^{\ell} \phi'(\mathbf{H}_{k_\ell q}^{(\ell-1)}) \right) \mathcal{D}\mathbf{X} \tag{113}$$

$$= \int \phi(\sum_l \mathbf{W}_{i_1,l}^{(\ell)} \mathbf{Y}_{l\mu_1}^{(\ell)}) \phi(\sum_l \mathbf{W}_{i_2,l}^{(\ell)} \mathbf{Y}_{l\mu_1}^{(\ell)}) \mathcal{D}\mathbf{X} - \sigma_x^2 \sum_p^n \mathfrak{W}_{i_1 p}^{(\ell)} \mathfrak{W}_{i_2 p}^{(\ell)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \tag{114}$$

and we therefore aim at computing

$$E_{2k}^{(\ell,n)} = \int_{\mathbf{WX}} \left\{ \prod_{\xi=1}^{k} \left(\phi(\sum_l \mathbf{W}_{i_\xi,l}^{(\ell)} \mathbf{Y}_{l\mu_\xi}^{(\ell)}) \phi(\sum_l \mathbf{W}_{i_{\xi+1},l}^{(\ell)} \mathbf{Y}_{l\mu_\xi}^{(\ell)}) - \sigma_x^2 \sum_p^n \mathfrak{W}_{i_\xi p}^{(\ell)} \mathfrak{W}_{i_{\xi+1} p}^{(\ell)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \right) \right\} \mathcal{D}\mathbf{X}\mathcal{D}\mathbf{W} \tag{115}$$

$$= \sum_{n_w=0}^{k} \binom{k}{n_w} \sum_{\substack{\mathbf{n}_\phi \ s.t. \\ \|\mathbf{n}_\phi\|_1 = n - n_w}} \sum_{p=1}^{n} E_w^{(k,n_w,\mathbf{n}_\phi,p)} \tag{116}$$

where n_w is the number of terms of the type $\mathfrak{W}_{i_\xi p}^{(\ell)} \mathfrak{W}_{i_{\xi+1} p}^{(\ell)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2$ in one addend of the expansion of the product in $E_{2k}^{(\ell,n)}$, and $\mathbf{n}_\phi$ corresponds to a vector whose entries define how many terms of the type $\phi(\sum_l \mathbf{W}_{i_\xi,l}^{(\ell)} \mathbf{Y}_{l\mu_\xi}^{(\ell)}) \phi(\sum_l \mathbf{W}_{i_{\xi+1},l}^{(\ell)} \mathbf{Y}_{l\mu_\xi}^{(\ell)})$ are sequentially included between two terms of the type $\mathfrak{W}_{i_\xi p}^{(\ell)} \mathfrak{W}_{i_{\xi+1} p}^{(\ell)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2$.

To achieve accurate modelling of the contribution of $E_{2k}^{(\ell,n)}$ for deeper layers, it is necessary to to keep into consideration the covariance matrix of each hidden layer $\mathbf{Y}^{(\ell)}$. Therefore, at each layer the matrix $\tilde{\mathbf{W}}^{(\ell)} \mathbf{\Sigma}^{(\ell)^{1/2}} \tilde{\mathbf{Y}}$ is going to be considered instead of $\mathbf{W}^{(\ell)} \mathbf{Y}^{(\ell)}$. For $k = 1$, the contribution of a two dimensional cycle is determined as follows

Lemma A.4. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m} \mathbb{E}_{\boldsymbol{X}} \left[\phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)}) \phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)})^\top \right] - \frac{1}{m^2 \sigma_x^2} \mathbb{E}_{\boldsymbol{x}} \left[\phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)}) \boldsymbol{X}^\top \right] \mathbb{E}_{\boldsymbol{x}} \left[\phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)}) \boldsymbol{X}^\top \right] \tag{117}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ as defined in Theorem 3.2 without the hypothesis of independence for the elements of $\boldsymbol{Y}^{(\ell)}$, i.e. $\boldsymbol{Y}_{:p}^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)} \mathbf{\Sigma}^{(\ell)})$, when $k = 1$ then the contribution of an admissible cycle is

$$E_2^{(\ell,n)} = \int \tilde{\theta}_1^{(\ell,n)} \mathcal{D}\mathbf{\Sigma}^{(\ell)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right) \right) - \sigma_x^2 \sigma_w^{2\ell} \prod_{l=1}^{\ell} (\theta_3^{(l)})^2 \tag{118}$$

where

$$\tilde{\theta}_1^{(\ell,n)} = \int \frac{1}{\sqrt{2\pi}} \phi(\sqrt{q^{(\ell)}} \sqrt{\frac{tr(\mathbf{\Sigma}^{(\ell)})}{n}} z)^2 e^{-\frac{z^2}{2}} dz \tag{119}$$

$$\theta_3^{(\ell)} = \int \phi'(\sqrt{q^{(\ell)}} z) e^{-\frac{z^2}{2}} dz. \tag{120}$$

Proof. See proof in Section C.1. □

For $k > 1$, this modelling leads to the following result for an addend ω in the expansion of (116) with $\mathbf{n}_\phi$ and n_w

Lemma A.5. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m} \mathbb{E}_{\boldsymbol{X}} \left[\phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)}) \phi(\boldsymbol{W}^{(\ell)} \boldsymbol{Y}^{(\ell)})^\top \right]$$

$$- \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{121}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2 without the hypothesis of independence for the elements of* $\boldsymbol{Y}^{(\ell)}$*, i.e.* $\boldsymbol{Y}_{:p}^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{\Sigma}^{(\ell)})$*, when* $k > 1$ *then the contribution of a single term in the expanded sum of* (116) *is*

$$E_\omega^{(n,k,n_w,\boldsymbol{n}_\phi,p)} = \int\left(-\frac{\sigma_w^2\sigma_x^2\left(\prod_{l=1}^{\ell}\theta_3^{(l)}\right)^2}{n}\right)^{n_w}\left(\frac{\tilde{\theta}_2^{(\ell,n)}}{n}\right)^{k-n_w}\left(\prod_{\xi=1}\mu^{(n)}_{\boldsymbol{n}_\phi^{(\xi)}}\right) \tag{122}$$

where $\mu_k^{(n)} = \frac{tr(\boldsymbol{\Sigma}^{(\ell)^k})}{n}$ *and*

$$\tilde{\theta}_2^{(\ell,n)} = \left(\int\frac{\sqrt{q^{(\ell)}}}{\sqrt{2\pi}}e^{-\frac{\tilde{z}^2}{2}}\phi'(\sqrt{q^{(\ell)}}\sqrt{\frac{tr(\boldsymbol{\Sigma}^{(\ell)})}{n}}\tilde{z})d\tilde{z}\right)^2, \tag{123}$$

$$\theta_3^{(\ell)} = \int\phi'(\sqrt{q^{(\ell)}}z)e^{-\frac{z^2}{2}}dz. \tag{124}$$

Proof. See proof in Section C.2. □

The dependence of $E_\omega^{(n,k,n_w,n_\phi,p)}$ on the moments of $\boldsymbol{\Sigma}^{(\ell)}$ is such that it is not possible to aggregate explicitly all of the contributions in (116). For this reason, it is necessary to introduce a new assumption that $\boldsymbol{\Sigma} = \mathbf{I}$. With this new assumption, it is then possible to compute explicitly the contribution of cycles with $k > 1$.

Lemma A.6. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{125}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2, i.e.* $\boldsymbol{Y}_{:p}^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{I})$*, when* $k > 1$ *then contribution of an admissible cycle is*

$$E_{2k} = n_0^{1-k}\left(\theta_2^{(\ell)} - \sigma_w^{2\ell}\sigma_x^2\left(\prod_{l=1}^{\ell}\theta_3^{(\ell)}\right)^2\right)^k\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{126}$$

where

$$\theta_2^{(\ell)} = \left(\int\frac{\sqrt{q^{(\ell)}}}{\sqrt{2\pi}}e^{-\frac{\tilde{z}^2}{2}}\phi'(\sqrt{q^{(\ell)}}\tilde{z})d\tilde{z}\right)^2, \tag{127}$$

$$\theta_3^{(\ell)} = \int\phi'(\sqrt{q^{(\ell)}}z)e^{-\frac{z^2}{2}}dz. \tag{128}$$

Proof. See proof in Section C.3. □

When considering $\boldsymbol{\Sigma}^{(\ell)} = \mathbf{I}$ then Lemma A.4 is updated as follows

Lemma A.7. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{129}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2 without the hypothesis of independence for the elements of* $\boldsymbol{Y}^{(\ell)}$*, i.e.* $\boldsymbol{Y}_{:p}^{(\ell)} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{I})$*, when* $k=1$ *then the contribution of an admissible cycle is*

$$E_2^{(\ell,n)} = \theta_1^{(\ell)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) - \sigma_x^2\sigma_w^{2\ell}\prod_{l=1}^{\ell}(\theta_3^{(l)})^2 \tag{130}$$

where

$$\theta_1^{(\ell)} = \int \frac{1}{\sqrt{2\pi}}\phi(\sqrt{q^{(\ell)}}z)^2 e^{-\frac{z^2}{2}}dz \tag{131}$$

$$\theta_3^{(\ell)} = \int \phi'(\sqrt{q^{(\ell)}}z)e^{-\frac{z^2}{2}}dz. \tag{132}$$

A.3.2 Aggregation of the Contributions and Implicit Relation in Theorem 3.2

Once the hypothesis of independence of the elements in the hidden layers is considered, the propositions for the aggregation of the contributions are achieved in the same exact way as for the first layer of $\boldsymbol{\Lambda}_{h^{(1)}} - \frac{1}{\sigma_x^2}\boldsymbol{\Sigma}_{xh^{(1)}}^\top\boldsymbol{\Sigma}_{xh^{(1)}}$, leading to the following two Lemmas.

Lemma A.8. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{133}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2, the Stieltjes transform of the eigen-distribution of* $\boldsymbol{M}$ *asymptotically satisfies the following equation*

$$G_{\boldsymbol{M}}(z) \simeq \frac{1-\varphi}{z} + \frac{\varphi}{z}H(z) \tag{134}$$

where $H(z)$ *is the generating function defined as* $H(z) = 1 + \sum_{k=1}^{\infty}\frac{H_k}{z^k\varphi^k} = 1 + \frac{1}{\varphi}\sum_{k=1}^{\infty}\frac{\mathfrak{m}_k}{z^k}$.

Lemma A.9. *Consider the matrix*

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{135}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2, and consider a graph* G_3 *with* $2(p+q)$ *edges that is defined, via either an* i *or* μ *identification, by two graphs* G_1 *and* G_2 *with* $2p$ *and* $2q$ *edges. Then the asymptotic contribution,* $\mathfrak{m}_{p+q}^{G_3}$*, of the graph* G_3 *to the moment* $\mathfrak{m}_{p+q}$ *is equal to the product of the contributions of the block cycle graphs that define* G_3 *with a correction term. This is, they are joined with an* i*-identification*

$$\mathfrak{m}_{p+q}^{G_3} = \mathfrak{m}_p^{G_1}\frac{n_1}{n_1}\mathfrak{m}_p^{G_2} \tag{136}$$

and if they are joined with a μ*-identification*

$$\mathfrak{m}_{p+q}^{G_3} = \mathfrak{m}_p^{G_1}\frac{n_1\kappa_\alpha}{m}\mathfrak{m}_p^{G_2} \tag{137}$$

where if either one of the block cycles connected to the identification are of dimension 2*, then* $\alpha = b$ *or otherwise* $\alpha = c$ *and* $\kappa_b = 1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)}$, $\kappa_c = 1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^2}$.

By using Lemmas A.8 and A.9 it is then possible to proof Theorem 3.2 by using the same exact process as for Theorem A.1.

B Proofs Relevant to Theorem A.1

B.1 Lemma A.2

The expectation of any $2k$ cycle can be expressed in a different format thanks to Lemma B.1. From such a formulation it is then possible to identify the expected contribution for $k = 1$. First thought the following result is proven

Lemma B.1. *For $Y = \phi(\boldsymbol{WX})$ as defined in Theorem A.1, when $k = 1$ then the following identity holds*

$$\int_{\boldsymbol{W},\boldsymbol{X}} \phi(\sum_l \boldsymbol{W}_{i_1,l}\boldsymbol{X}_{l,\mu_1})\phi(\sum_l \boldsymbol{W}_{i_1,l}\boldsymbol{X}_{l,\mu_1})\mathcal{D}\boldsymbol{W}\mathcal{D}\boldsymbol{X}$$
$$= \int \frac{1}{\sqrt{2\pi}}\phi(\sigma_w\sigma_w z)^2 e^{-\frac{z^2}{2}}dz\,(1+\mathcal{O}(1/n)) = \theta_1\,(1+\mathcal{O}(1/n)) \tag{138}$$

with

$$\theta_1 = \int \frac{1}{\sqrt{2\pi}}\phi(\sigma_w\sigma_w z)^2 e^{-\frac{z^2}{2}}dz. \tag{139}$$

Proof. Let's start by considering the general case with $i_1 \neq i_2 \neq ... \neq i_k$ and $\mu_1 \neq \mu_2 \neq ... \neq \mu_k$ when computing

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_1}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right]. \tag{140}$$

After expanding the expectation, we consider auxiliary integrals over z, by adding delta functions enforcing $\mathbf{Z} = \mathbf{W}\,\mathbf{X}$ with

$$\mathbf{Z}_{i\mu} = \begin{cases} z_{i\mu} & \text{if } (i,\mu)\in\mathcal{Z} \\ 0 & \text{otherwise.}\end{cases} \tag{141}$$

where $\mathcal{Z}$ denotes the set of unique pairs (i,μ) in equation (140):

$$\int \phi(\sum_l \mathbf{W}_{i_1 l}\mathbf{X}_{l\mu 1})\phi(\sum_l \mathbf{W}_{i_2 l}\mathbf{X}_{l\mu 1})...\phi(\sum_l \mathbf{W}_{i_1 l}\mathbf{X}_{l\mu_k})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{142}$$

$$= \int \prod_{(\alpha,\beta)\in\mathcal{Z}} \delta(z_{\alpha\beta} - \sum_k \mathbf{W}_{\alpha k}\mathbf{X}_{k\beta})\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{143}$$

where

$$\mathcal{D}z = \prod_{(\alpha,\beta)\in\mathcal{Z}} dz_{\alpha\beta}. \tag{144}$$

Now we consider the Fourier expression of the Dirac δ

$$\delta(x) = \frac{1}{2\pi}\int e^{i\lambda x}d\lambda \tag{145}$$

and therefore introduced the matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times m}$ whose entries are

$$\mathbf{\Lambda}_{i\mu} = \begin{cases} \lambda_{i\mu} & \text{if } (i,\mu)\in\mathcal{Z} \\ 0 & \text{otherwise.}\end{cases} \tag{146}$$

with

$$\mathcal{D}\lambda = \prod_{(\alpha,\beta)\in\mathcal{Z}} \frac{d\lambda_{\alpha\beta}}{2\pi}. \tag{147}$$

to obtain

$$\int \phi(\sum_l \mathbf{W}_{i_1 l}\mathbf{X}_{l\mu 1})\phi(\sum_l \mathbf{W}_{i_2 l}\mathbf{X}_{l\mu 1})...\phi(\sum_l \mathbf{W}_{i_1 l}\mathbf{X}_{l\mu_k})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{148}$$

$$= \int \prod_{z_{\alpha\beta}\in\mathcal{Z}} \delta(z_{\alpha\beta} - \sum_k \mathbf{W}_{\alpha k}\mathbf{X}_{k\beta})\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{149}$$

$$= \int \prod_{z_{\alpha\beta}\in\mathcal{Z}} exp\left(-i\lambda_{\alpha\beta}\left(\sum_k \mathbf{W}_{\alpha k}\mathbf{X}_{k\beta} - z_{\alpha\beta}\right)\right)\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{150}$$

$$= \int exp\left(-i \sum_{z_{\alpha\beta}\in\mathcal{Z}} \lambda_{\alpha\beta} \underbrace{\left(\sum_k \mathbf{W}_{\alpha k}\mathbf{X}_{k\beta} - z_{\alpha\beta}\right)}_{(\mathbf{WX}-\mathbf{Z})_{\alpha\beta}}\right) \phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{151}$$

$$= \int exp\left(-i\sum_{\alpha,\beta=1}^{n,m} \mathbf{\Lambda}_{\alpha\beta}(\mathbf{WX}-\mathbf{Z})_{\alpha\beta}\right)\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{152}$$

$$= \int e^{-i\, tr\left(\Lambda^\top(\mathbf{WX}-\mathbf{Z})\right)}\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{153}$$

where $tr()$ corresponds to the trace function.

Now we first integrate over $\mathbf{X}$ the factors of (153) that depend on it

$$\int e^{-itr(\mathbf{\Lambda}^\top\mathbf{WX})}\mathcal{D}\mathbf{X} = \prod_{b,c=1}^{m,n}\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\mathbf{X}_{cb}^2 - i\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{X}_{cb}\right]$$

$$= \prod_{b,c=1}^{m,n}\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\left(\mathbf{X}_{cb} + i\sigma_x^2(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})\right)^2 - \frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2\right] \tag{154}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2\right] \underbrace{\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\left(\mathbf{X}_{cb} + i\sigma_x^2(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})\right)^2\right]}_{=1} \tag{155}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2\right] = exp\left[-\frac{\sigma_x^2}{2}\sum_{b,c=1}^{m,n}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2\right] \tag{156}$$

$$= exp\left[-\frac{\sigma_x^2}{2}\|\mathbf{\Lambda}^\top\mathbf{W}\|_F^2\right] = e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{157}$$

where in equation (155) we used the property the complex integral of $z = x + iy$ over the closed cycle $(-\infty, \infty, i\mu + \infty, i\mu - \infty)$ of the analytical function $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-iy)^2/(2\sigma^2)}$ is null and therefore

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-i\mu)^2/(2\sigma^2)}dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x)^2/(2\sigma^2)}dx = 1.$$

Now we integrate over $\mathbf{W}$

$$\int e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})}\mathcal{D}\mathbf{W}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}e^{-\frac{n\mathbf{W}_{ij}^2}{2\sigma_w^2}}\right)e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{158}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{n\sum_{i,j=1}^{n}\mathbf{W}_{ji}\mathbf{W}_{ij}}{2\sigma_w^2}}e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{159}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{ntr\mathbf{W}^\top\mathbf{W}}{2\sigma_w^2}}e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{160}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{ntr\mathbf{W}^\top\mathbf{W}}{2\sigma_w^2}-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{161}$$

$$= \int \left(\prod_{j=1}^{n} \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} \left(\frac{n}{2\sigma_w^2} \mathbf{w}_j^\top \mathbf{w}_j + \frac{\sigma_x^2}{2} \mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j \right)} \tag{162}$$

$$= \int \left(\prod_{j=1}^{n} \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} e^{-\left(\frac{n}{2\sigma_w^2} \mathbf{w}_j^\top \mathbf{w}_j + \frac{\sigma_x^2}{2} \mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j \right)} \right) \tag{163}$$

$$= \prod_{j=1}^{n} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j} \tag{164}$$

$$= \prod_{j=1}^{n} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j} \tag{165}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \tag{166}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j}}{det\left((\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1} \right)^{1/2}} \tag{167}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi)^{n/2}} \frac{e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j}}{det\left(\sigma_w^2/n(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1} \right)^{1/2}} \tag{168}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \underbrace{\int \frac{d^n \mathbf{w}_j}{(2\pi)^{n/2}} \frac{e^{-\frac{1}{2} \mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \mathbf{w}_j}}{det\left((\frac{n}{\sigma_w^2} \mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1} \right)^{1/2}}}_{=1} \tag{169}$$

$$= \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}} \tag{170}$$

where we consider $\mathbf{w}_j$ as the jth column of $\mathbf{W}$ and we used the property that for a general non-singular matrix $det(\mathbf{A}^{-1}) = det^{-1}(\mathbf{A})$.

This implies that by considering $F(z) = \prod_{(\alpha,\beta)\in\mathcal{Z}} \phi(z_{\alpha\beta})$

$$\int \phi(\sum_l \mathbf{W}_{i_1 l} \mathbf{X}_{l\mu 1}) \phi(\sum_l \mathbf{W}_{i_2 l} \mathbf{X}_{l\mu 1}) ... \phi(\sum_l \mathbf{W}_{i_1 l} \mathbf{X}_{l\mu_k}) \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{171}$$

$$= \int \left(\frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}} \right) e^{-i\, tr(\mathbf{\Lambda}^\top \mathbf{Z})} F(z) \mathcal{D}z\mathcal{D}\lambda \tag{172}$$

$$= \int exp\left(-\frac{n}{2} \log det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top) - itr\mathbf{\Lambda}^\top \mathbf{Z} \right) F(z) \mathcal{D}z\mathcal{D}\lambda \tag{173}$$

Now we will consider the integration over the $\lambda_{\alpha\beta}$ variables. Since the eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}^\top$ are non-negative, as a matter of fact for any pair $(\lambda, \mathbf{v})$ the following holds $\lambda = \frac{\mathbf{v}^T \mathbf{\Lambda}^T \mathbf{\Lambda} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\|\mathbf{\Lambda}\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \geq 0$, the maximizer of the argument in the exponential is $\mathbf{\Lambda} = 0$, by the saddle point approximation we can consider only an expansion around $\mathbf{\Lambda} = 0$. We can then use the same analysis done in [13] and decompose the log determinant via $logdet|\mathbf{I} + \mathbf{X}| = \sum_{\xi=1} \frac{(-1)^{\xi+1}}{\xi} tr(\mathbf{X}^\xi)$. Then it follows that

$$\int \phi(\sum_l \mathbf{W}_{i_1 l} \mathbf{X}_{l\mu 1}) \phi(\sum_l \mathbf{W}_{i_2 l} \mathbf{X}_{l\mu 1}) ... \phi(\sum_l \mathbf{W}_{i_1 l} \mathbf{X}_{l\mu_k}) \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{174}$$

$$= \int exp\left(-\frac{n}{2}\log det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top) - itr\mathbf{\Lambda}^\top\mathbf{Z}\right)F(z)\mathcal{D}z\mathcal{D}\lambda \tag{175}$$

$$= \int e^{-\frac{\sigma_x^2\sigma_w^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - \frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}z\mathcal{D}\lambda \tag{176}$$

$$= \int e^{-\frac{\sigma_x^2\sigma_w^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)}e^{-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}z\mathcal{D}\lambda \tag{177}$$

We now consider the following change of variable

$$\bar{\lambda}_{ij} = \frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{ij} \tag{178}$$

$$\mathcal{D}\bar{\mathbf{\Lambda}} = \prod_{(\alpha,\beta)\in\mathcal{Z}}\frac{d\bar{\lambda}_{\alpha\beta}}{2\pi\sigma_x\sigma_w/\sqrt{n}} \tag{179}$$

therefore

$$\int \phi(\sum_l \mathbf{W}_{i_1l}\mathbf{X}_{l\mu 1})\phi(\sum_l \mathbf{W}_{i_2l}\mathbf{X}_{l\mu 1})...\phi(\sum_l \mathbf{W}_{i_1l}\mathbf{X}_{l\mu_k})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{180}$$

$$= \int e^{-\frac{\sigma_x^2\sigma_w^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)}e^{-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top Z}F(z)\mathcal{D}z\mathcal{D}\lambda \tag{181}$$

$$= \int e^{-\frac{n}{2}tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)}e^{-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right) - itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}^\top\mathbf{Z}}F(z)\mathcal{D}z\mathcal{D}\bar{\mathbf{\Lambda}}. \tag{182}$$

We will study the contribution of the exponential $e^{-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)}$ by considering its Taylor expansion $e^x = \sum_{\nu=0}^\infty x^\nu/\nu!$

$$\int \phi(\sum_l \mathbf{W}_{i_1l}\mathbf{X}_{l\mu 1})\phi(\sum_l \mathbf{W}_{i_2l}\mathbf{X}_{l\mu 1})...\phi(\sum_l \mathbf{W}_{i_1l}\mathbf{X}_{l\mu_k})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{183}$$

$$= \int\left(\sum_{\nu=0}^\infty\frac{\left(-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right)e^{-\frac{n}{2}tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top) - itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}^\top\mathbf{Z}}F(z)\mathcal{D}z\mathcal{D}\bar{\mathbf{\Lambda}} \tag{184}$$

$$= \int\left(\sum_{\nu=0}^\infty\frac{\left(-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right)e^{-\sum_{\lambda_{\alpha\beta}\in\mathcal{Z}}\left(\frac{n}{2}\bar{\lambda}_{\alpha\beta}^2 - i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}_{\alpha\beta}z_{\alpha\beta}\right)}F(z)\mathcal{D}z\mathcal{D}\bar{\mathbf{\Lambda}} \tag{185}$$

$$= \int\left(\sum_{\nu=0}^\infty\frac{\left(-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right)\left(\prod_{\lambda_{\alpha\beta}\in\mathcal{Z}}e^{-\frac{n}{2}\bar{\lambda}_{\alpha\beta}^2 - i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}_{\alpha\beta}z_{\alpha\beta}}\right)F(z)\mathcal{D}z\mathcal{D}\bar{\mathbf{\Lambda}} \tag{186}$$

When $k = 1$, following [11] the zeroth order expansion of the Taylor series

$$\left(\sum_{\nu=0}^\infty\frac{\left(-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right)$$

is the leading contribution to $\int_{\mathbf{W},\mathbf{X}}\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1})\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}$.

The zeroth order contribution is determined as follows.

$$\int\left(\frac{\sqrt{n}}{2\pi\sigma_x\sigma_w}e^{-\frac{n}{2}tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)}e^{-itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}\mathbf{Z}}F(z)\right)dz_{i_1\mu_1}d\bar{\lambda}_{i_1\mu_1}$$

$$= \int \left(\int \frac{\sqrt{n}}{2\pi\sigma_x\sigma_w} e^{-\frac{n}{2}\lambda^2 - i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}z} \phi(z)^2 dz d\bar{\lambda} \right) \tag{187}$$

$$= \int \left(\int \frac{\sqrt{n}}{2\pi\sigma_x\sigma_w} e^{-\frac{z^2}{2\sigma_w^2\sigma_x^2}} \frac{\sqrt{2\pi}}{\sqrt{n}} \phi(z)^2 \left(\int \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}\left(\lambda + \frac{i\sqrt{n}}{n\sigma_w\sigma_x}z\right)^2} d\bar{\lambda} \right) dz \right) \tag{188}$$

$$= \int \frac{\sqrt{n}}{2\pi\sigma_x\sigma_w} \phi(z)^2 e^{-\frac{z^2}{2\sigma_w^2\sigma_x^2}} \frac{\sqrt{2\pi}}{\sqrt{n}} dz = \int \frac{1}{\sqrt{2\pi}\sigma_x\sigma_w} \phi(z)^2 e^{-\frac{z^2}{2\sigma_w^2\sigma_x^2}} dz \tag{189}$$

$$= \int \frac{1}{\sqrt{2\pi}\sigma_x\sigma_w} \phi(\sigma_w\sigma_x\tilde{z})^2 e^{-\frac{\tilde{z}^2}{2}} \sigma_w\sigma_x d\tilde{z} = \int \frac{1}{\sqrt{2\pi}} \phi(\sigma_w\sigma_x\tilde{z})^2 e^{-\frac{\tilde{z}^2}{2}} d\tilde{z} = \theta_1 \tag{190}$$

For higher order terms of the exponential Taylor series we first notice that

$$\left(-\frac{n}{2} \sum_{\xi \geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi \right) \right)^\nu = \left(-\frac{n}{2} \sum_{\xi \geq 2} \frac{(-1)^{\xi+1}}{\xi} \bar{\lambda}^{2\xi} \right)^\nu .$$

For each of the expansion terms, the same steps in (187)-(189) lead to the following $2\xi\nu$-th non-central moment of a Gaussian to be considered

$$\int \frac{\sqrt{n}}{\sqrt{2\pi}} \bar{\lambda}^{2\xi\nu} e^{-\frac{n}{2}(\bar{\lambda} + \frac{i\sqrt{n}z}{n\sigma_x\sigma_w})} d\bar{\lambda} = \left(\frac{1}{n}\right)^{\xi\nu/2} 2^{\xi\nu} \frac{\Gamma(\frac{2\xi\nu+1}{2})}{\sqrt{\pi}} \Phi(-\frac{2\xi\nu}{2}; \frac{1}{2}; -\frac{z^2}{\sigma_w^2\sigma_x^2}) \tag{191}$$

where the solution of the non-central moment is given in [22] with

$$\Phi(-\frac{2\xi\nu}{2}; \frac{1}{2}; -\frac{z^2}{\sigma_w^2\sigma_x^2}) = \sum_{i=1}^{\infty} \frac{1}{i!} \frac{(-\frac{2\xi\nu}{2})(-\frac{2\xi\nu}{2}+1)...(-\frac{2\xi\nu\nu}{2}+i-1)}{(-\frac{1}{2})(-\frac{1}{2}+1)...(-\frac{1}{2}+i-1)} \left(\frac{-z^2}{\sigma_w^2\sigma_x^2}\right)^i . \tag{192}$$

This leads to the computation of the following integral

$$\int \sum_i \alpha_i \left(\frac{-z^2}{\sigma_w^2\sigma_x^2}\right)^i \phi^2(\sigma_x^2\sigma_w^2 z)\mathcal{D}z \tag{193}$$

which is finite since $\Phi(-\frac{2\xi\nu}{2}; \frac{1}{2}; -\frac{z^2}{\sigma_w^2\sigma_x^2})$ is an entire function of $\xi\nu$ and z, and when solving it by parts we compute the following integrals that are finite by hypothesis

$$\int \phi^{(k)}(\sigma_x^2\sigma_w^2 z)\mathcal{D}z \tag{194}$$

for any derivative k. However, because of the non-central moment, we also gain a factor $\left(\frac{1}{n}\right)^{\xi\nu/2}$ and therefore these terms have a contribution that is $\mathcal{O}(\frac{1}{n})$ relative to the zeroth-order one.

Therefore it follows that

$$\int_{\mathbf{W},\mathbf{X}} \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1})\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1})\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}$$
$$= \int \frac{1}{\sqrt{2\pi}} \phi(\sigma_w\sigma_w z)^2 e^{-\frac{z^2}{2}} dz\, (1 + \mathcal{O}(1/n)) = \theta_1\, (1 + \mathcal{O}(1/n)) \tag{195}$$

□

B.1.1 Proof of Lemma A.2

Proof. In the case where $k = 1$,

$$E_2^{(n)} = \int_{\mathbf{W},\mathbf{X}} \phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1})\phi(\sum_l \mathbf{W}_{i_1,l}\mathbf{X}_{l,\mu_1}) - \sigma_x^2 \sum_{p=1}^{n} \mathbf{W}_{i_1 p}\mathbf{W}_{i_1 p}\theta_3^2 \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{196}$$
$$= \theta_1(1 + \mathcal{O}(1/n)) - \sigma_w^2\sigma_x^2\theta_3^2 \tag{197}$$

where the first term follows from LemmaB.1 and $\theta_3 = \int \phi'(\sigma_w\sigma_w z)e^{-\frac{z^2}{2}} dz$ □

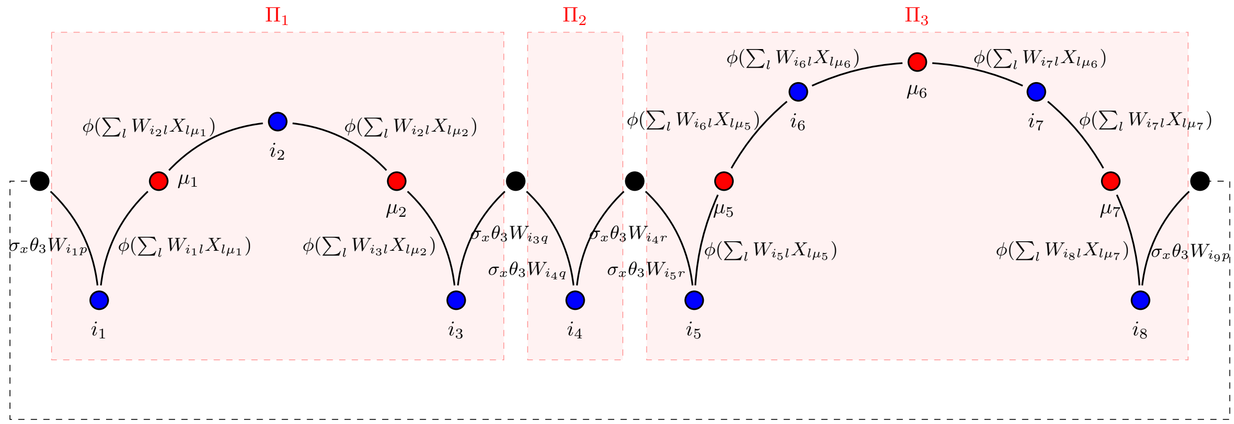


Figure 7: Example of how in the Schur complement the contribution to the momentum is defined on as many independent blocks of products, Π_j, as the number of factors in the form $\sigma_x^2 \mathbf{W}_{i_\xi p} \mathbf{W}_{i_{\xi+1} p} \theta_3^2$.

B.2 Lemma A.3

When considering $k > 1$ we need to consider all the mixed products in (63). As a matter of fact each term in the expansion of the product in E_{2k} consists of successions of products of the kind $\phi(\sum_l \mathbf{W}_{i_\xi,l} \mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\xi})$ alternated with products of the kind $\sigma_x^2 \theta_3^2 \sum_{p=1}^n \mathbf{W}_{i_\xi p} \mathbf{W}_{i_{\xi+1} p}$. Each element of the kind $\sigma_x^2 \theta_3^2 \mathbf{W}_{i_\xi p} \mathbf{W}_{i_{\xi+1} p}$ divides the sequence of products into independent blocks Π_j. As we see in Figure 7, with three factors of the kind $\sigma_x^2 \theta_3^2 \mathbf{W}_{i_\xi p} \mathbf{W}_{i_{\xi+1} p}$ we generate three independent blocks Π_j that are independent among themselves.

Lemma B.2. *For $\boldsymbol{M} = \frac{1}{m} \mathbb{E}_{\boldsymbol{X}} \left[\boldsymbol{Y}\boldsymbol{Y}^\top \right] - \frac{1}{\sigma_x^2 m^2} \mathbb{E}_{\boldsymbol{x}} \left[\boldsymbol{Y}\boldsymbol{X}^\top \right] \mathbb{E}_{\boldsymbol{X}} \left[\boldsymbol{X}\, \boldsymbol{Y}^\top \right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, when $k > 1$ each independent block Π_j^l, with l terms $\phi(\sum_l \mathbf{W}_{i_\xi,l} \mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\xi})$, generates the following contribution*

$$E^{(n)}_{\Pi_i^{(l)}} = -\frac{\sigma_w^2 \sigma_x^2 \theta_3^2 \theta_2^l}{n^{1+l}}. \tag{198}$$

with

$$\theta_2 = \left(\int \frac{\sigma_w \sigma_x}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} \phi'(\sigma_w \sigma_x \tilde{z}) d\tilde{z} \right)^2, \tag{199}$$

$$\theta_3 = \int_{z_1} \phi'\left(\sigma_x \sigma_w z_1\right) \mathcal{D} z_1 \tag{200}$$

Proof. We focus on the integration of one independent block $\Pi_j^{(l)}$, where l identifies the numbers of factors $\phi(\sum_l \mathbf{W}_{i_\xi,l} \mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\xi})$ between two of the $\sigma_x \theta_3 \mathbf{W}_{i_\xi p}$ kind, i.e.

$$E^{(n)}_{\Pi_i^{(l)}} = -\int_{\mathbf{W},\mathbf{X}} \left\{ \sigma_x \theta_3 \mathbf{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(\sum_l \mathbf{W}_{i_\xi,l} \mathbf{X}_{l,\mu_\xi}) \phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\xi}) \right) \sigma_x \theta_3 \mathbf{W}_{i_{l+1} q} \right\} \mathcal{D}\mathbf{W} \mathcal{D}\mathbf{X}. \tag{201}$$

We are going to compute this expectation following the structure of the proof in [13]. We will introduce a dummy variable z with a delta Dirac function within each ϕ element, and then introduce a Fourier representation for all the arguments in the ϕ functions. We introduce for each factor Π_j the set $\mathcal{Z}_{\Pi_j} \subset \mathcal{Z}$ which contains only the combinations (i_ξ, μ_ν) that are included in the $\phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\nu})$ arguments. Thus we consider auxiliary integrals over z, by adding delta functions enforcing $\mathbf{Z} = \mathbf{W}\mathbf{X}$ with

$$\mathbf{Z}_{i\mu} = \begin{cases} z_{i\mu} & \text{if } (i,\mu) \in \mathcal{Z}_{\Pi_j} \\ 0 & \text{otherwise.} \end{cases} \tag{202}$$

and consequently

$$E^{(n)}_{\Pi_i^{(l)}} = -\int \left\{ \sigma_x \theta_3 \mathbf{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(\sum_l \mathbf{W}_{i_\xi,l} \mathbf{X}_{l,\mu_\xi}) \phi(\sum_l \mathbf{W}_{i_{\xi+1},l} \mathbf{X}_{l,\mu_\xi}) \right) \sigma_x \theta_3 \mathbf{W}_{i_{l+1} q} \right\} \mathcal{D}\mathbf{W} \mathcal{D}\mathbf{X} \tag{203}$$

$$= -\int \prod_{(\alpha,\beta)\in\mathcal{Z}_{\Pi_j}} \delta(z_{\alpha\beta} - \sum_k \mathbf{W}_{\alpha k}\mathbf{X}_{k\beta}) \left\{ \sigma_x^2\theta_3^2 \mathbf{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi})\phi(z_{i_\xi \mu_\xi})\right) \mathbf{W}_{i_{l+1}q} \right\} \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}z \tag{204}$$

$$= -\int e^{-itr\mathbf{\Lambda}^\top(\mathbf{WX}-\mathbf{Z})} \sigma_x^2\theta_3^2 \mathbf{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi})\phi(z_{i_{\xi+1} \mu_\xi})\right) \mathbf{W}_{i_{l+1}q} \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\lambda\mathcal{D}z \tag{205}$$

where

$$\mathcal{D}z = \prod_{(\alpha,\beta)\in\mathcal{Z}_{\Pi_j}} dz_{\alpha\beta} \tag{206}$$

and in the second equality we used the property that

$$\delta(x) = \frac{1}{2\pi}\int e^{i\lambda x} d\lambda \tag{207}$$

and therefore introduced the matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times m}$ whose entries are

$$\mathbf{\Lambda}_{i\mu} = \begin{cases} \lambda_{i\mu} & \text{if } (i,\mu)\in\mathcal{Z}_{\Pi_j} \\ 0 & \text{otherwise.} \end{cases} \tag{208}$$

and therefore

$$\mathcal{D}\lambda = \prod_{(\alpha,\beta)\in\mathcal{Z}_{\Pi_j}} \frac{d\lambda_{\alpha\beta}}{2\pi}. \tag{209}$$

We integrate over $\mathbf{X}$

$$\int e^{-itr(\mathbf{\Lambda}^\top\mathbf{WX})}\mathcal{D}\mathbf{X} = \prod_{b,c=1}^{m,n} \int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\mathbf{X}_{cb}^2 - i\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{X}_{cb} \right] \tag{210}$$

$$= \prod_{b,c=1}^{m,n} \int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\left(\mathbf{X}_{cb} + i\sigma_x^2(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})\right)^2 - \frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2 \right] \tag{211}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2 \right] = exp\left[-\frac{\sigma_x^2}{2}\sum_{b,c=1}^{m,n}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac})^2 \right] \tag{212}$$

$$= exp\left[-\frac{\sigma_x^2}{2}\|\mathbf{\Lambda}^\top\mathbf{W}\|_F^2 \right] = e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})}. \tag{213}$$

We also integrate over $\mathbf{W}$, first considering the case where $p \neq q$

$$\int_{\mathbf{W}} \sigma_x^2\theta_3^2 e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{W}^\top)} \mathbf{W}_{i_1 p}\mathbf{W}_{i_{l+1}q}\mathcal{D}\mathbf{W} \tag{214}$$

$$= \sigma_x^2\theta_3^2 \int \left(\prod_{i,j=1}^{n} \frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}} \right) e^{-\frac{ntr(\mathbf{W}^\top\mathbf{W})}{2\sigma_w^2}} e^{-\frac{\sigma_x^2}{2}tr(\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \mathbf{W}_{i_1 p}\mathbf{W}_{i_{l+1}q} \tag{215}$$

$$= \sigma_x^2\theta_3^2 \int \left(\prod_{j=1}^{n} \frac{d^n\mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n}(\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top\mathbf{w}_j + \frac{\sigma_x^2}{2}\mathbf{w}_j^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{w}_j)} \mathbf{w}_p^{(i_1)}\mathbf{w}_q^{(i_{l+1})} \tag{216}$$

$$= \sigma_x^2\theta_3^2 \int \left(\prod_{\substack{j=1\\j\neq p\\j\neq q}}^{n} \frac{d^n\mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{\substack{j=1\\j\neq p\\j\neq q}}^{n}(\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top\mathbf{w}_j + \frac{\sigma_x^2}{2}\mathbf{w}_j^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{w}_j)}$$

$$\cdot \underbrace{\int \frac{d^n\mathbf{w}_p}{(2\pi\sigma_w^2/n)^{n/2}} \mathbf{w}_p^{(i_1)} e^{-\frac{n}{2\sigma_w^2}\mathbf{w}_p^\top\mathbf{w}_p - \frac{\sigma_x^2}{2}\mathbf{w}_p^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{w}_p}}_{=0}$$

$$\cdot \underbrace{\int \frac{d^n \mathbf{w}_p}{(2\pi\sigma_w^2/n)^{n/2}} \mathbf{w}_q^{(i_{l+1})} e^{-\frac{n}{2\sigma_w^2}\mathbf{w}_q^\top \mathbf{w}_q - \frac{\sigma_x^2}{2}\mathbf{w}_q^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_q}}_{=0} \tag{217}$$

$$= 0. \tag{218}$$

While now we will consider the case where $q = p$

$$\int_{\mathbf{W}} e^{-\frac{\sigma_x^2}{2} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{W}\mathbf{W}^\top)} \sigma_x^2 \theta_3^2 \mathbf{W}_{i_1 p} \mathbf{W}_{i_{l+1} q} \mathcal{D}\mathbf{W}$$

$$= \sigma_x^2 \theta_3^2 \int \left(\prod_{i,j=1}^n \frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}} \right) e^{-\frac{ntr(\mathbf{W}^\top \mathbf{W})}{2\sigma_w^2}} e^{-\frac{\sigma_x^2}{2} tr(\mathbf{W}^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{W})} \mathbf{W}_{i_1 p} \mathbf{W}_{i_{l+1} q} \tag{219}$$

$$= \sigma_x^2 \theta_3^2 \int \left(\prod_{j=1}^n \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^n (\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top \mathbf{w}_j + \frac{\sigma_x^2}{2}\mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j)} \mathbf{w}_p^{(i_1)} \mathbf{w}_q^{(i_{l+1})} \tag{220}$$

$$= \sigma_x^2 \theta_3^2 \int \left(\prod_{\substack{j=1 \\ j\neq p}}^n \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{\substack{j=1 \\ j\neq p}}^n (\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top \mathbf{w}_j + \frac{\sigma_x^2}{2}\mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j)}$$

$$\cdot \int \frac{d^n \mathbf{w}_p}{(2\pi\sigma_w^2/n)^{n/2}} \mathbf{w}_p^{(i_1)} \mathbf{w}_p^{(i_{l+1})} e^{-\frac{n}{2\sigma_w^2}\mathbf{w}_p^\top \mathbf{w}_p - \frac{\sigma_x^2}{2}\mathbf{w}_p^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_p} \tag{221}$$

$$= \sigma_x^2 \theta_3^2 \prod_{\substack{j=1 \\ j\neq p}}^n \underbrace{\left(\int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} e^{-\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top \mathbf{w}_j - \frac{\sigma_x^2}{2}\mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j)} \right)}_{=\frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}}$$

$$\cdot \int \frac{d^n \mathbf{w}_p}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \mathbf{w}_p^{(i_1)} \mathbf{w}_p^{(i_{l+1})} e^{-\frac{1}{2}\mathbf{w}_p^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_p} \tag{222}$$

$$= \frac{\sigma_x^2 \theta_3^2}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}} Cov\left[\mathbf{w}_p^{(i_1)} \mathbf{w}_p^{(i_{l+1})}\right]. \tag{223}$$

Consequently, each non-zero contribution to the momentum is going to be defined by addends in (63) whose elements in the form $\sigma_x \theta_3 \mathbf{W}_{i_\xi p}$ have the same pedex p.

We notice that the covariance matrix of the vector w_j is

$$\left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{-1} = \frac{\sigma_w^2}{n} \sum_{k=0}^\infty (-\frac{\sigma_w^2\sigma_x^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k \tag{224}$$

therefore assuming that n is large enough

$$Cov\left[\mathbf{w}_j^{(i_1)} \mathbf{w}_j^{(i_{l+1})}\right] = \left[\left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{-1}\right]_{i_1 i_{l+1}} = \left[\frac{\sigma_w^2}{n} \sum_{k=0}^\infty (-\frac{\sigma_w^2\sigma_x^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k\right]_{i_1 i_{l+1}}. \tag{225}$$

Since $\mathbf{\Lambda}$ is defined on $\mathcal{Z}_{\Pi_j}$, the non-zero terms in $\left[(\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k\right]_{i_1 i_l}$ correspond to products $\prod_{j=1}^l \left(\lambda_{i_j \mu_j} \lambda_{i_{j+1}\mu_j}\right)^{\tilde{l}_j}$. The sequence of exponents $\overrightarrow{l} = (\tilde{l}_1, \ldots, \tilde{l}_l)$ can be determined by considering diagrams as the one reported in Figure 8 with $l = 5$. For each power k, the sequence of edges that are non-zero correspond to the paths that connect i_1 to i_l in exactly k steps; the exponent $\tilde{l}_j$ correspond to the number of times that the path goes through the interval (i_j, i_{j+1}).

Therefore, $k = l$ is the only case where all of the variables $\lambda_{i_j \mu_j} \lambda_{i_{j+1}\mu_j}$ have exponent of exactly one and this allows to write the covariance in the following way.

$$Cov\left[\mathbf{w}_j^{(i_1)} \mathbf{w}_j^{(i_{l+1})}\right] = \left[\frac{\sigma_w^2}{n} \sum_{k=0}^\infty (-\frac{\sigma_w^2\sigma_x^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k\right]_{i_1 i_{l+1}} \tag{226}$$

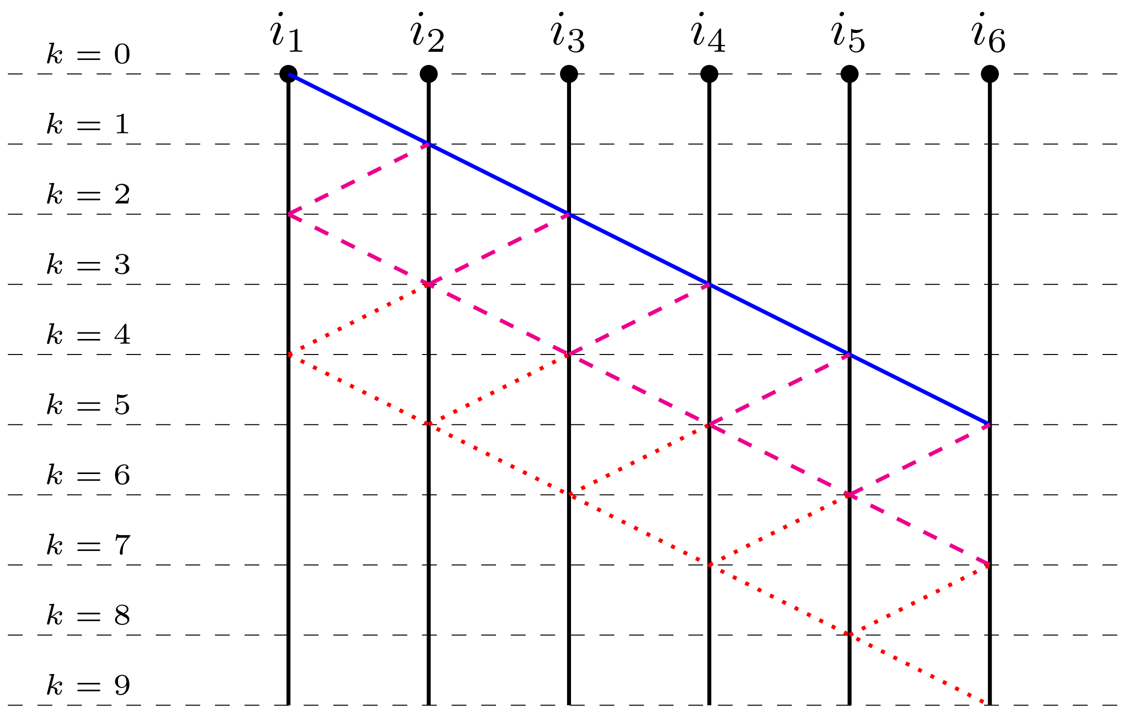


Figure 8: Diagram of sequences of paths that allow to connect i_1 to i_{l+1}, with $l = 5$, for different number of steps k. The blue line corresponds to the unique path connecting i_0 to i_l with $k = l$ steps; the dashed magenta lines corresponds to the l paths connecting i_0 to i_l with $k = l + 2$ steps; and the dotted red lines corresponds to the l^2 paths connecting i_0 to i_l with $k = l + 4$ steps.

$$= (-1)^l \frac{\sigma_w^2}{n} \prod_{j=1}^{l} \frac{\sigma_w^2 \sigma_x^2}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1} \mu_j} + \sum_{\nu=1}^{\infty} (-1)^{l+\nu} \frac{\sigma_w^2}{n} \sum_{\substack{\vec{l}\ s.t.\\ \|\vec{l}\|_1 = \nu + l\\ \tilde{l}_j \in \mathbb{N}_0}} \prod_{j=1}^{l} \left(\frac{\sigma_w^2 \sigma_x^2}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1} \mu_j} \right)^{\tilde{l}_j} \tag{227}$$

Let's now consider the integral on $\mathcal{D}\lambda\mathcal{D}z$ of the first addend in (227) and denote it with $E^{(n)\ L.O.}_{\Pi_i^{(l)}}$. By considering $F(z) = \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi}) \phi(z_{i_{\xi+1} \mu_\xi}) \right)$

$$E^{(n)\ L.O.}_{\Pi_i^{(l)}} =$$

$$- \int \sigma_x^2 \theta_3^2 (-1)^l \frac{\sigma_w^2}{n} \left(\prod_{j=1}^{l} \frac{\sigma_w^2 \sigma_x^2}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1} \mu_j} \right) \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2 \sigma_w^2}{n} \mathbf{\Lambda} \mathbf{\Lambda}^\top)^{n/2}} e^{-itr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \tag{228}$$

$$= - \int \sigma_x^2 \theta_3^2 (-1)^l \frac{\sigma_w^2}{n} \left(\prod_{j=1}^{l} \frac{\sigma_w^2 \sigma_x^2}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1} \mu_j} \right) e^{-\frac{n}{2} \log det(\mathbf{I} + \frac{\sigma_x^2 \sigma_w^2}{n} \mathbf{\Lambda} \mathbf{\Lambda}^\top)} e^{-itr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \tag{229}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \int \left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \frac{\sigma_w \sigma_x}{\sqrt{n}} \lambda_{\alpha\beta} \right)$$
$$e^{-\frac{\sigma_x^2 \sigma_w^2}{2} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - \frac{n}{2} \sum_{\xi \geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\left(\frac{\sigma_x^2 \sigma_w^2}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top \right)^{\xi} - itr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \tag{230}$$

where in the last equation we used the Taylor expansion $logdet|\mathbf{I} + \mathbf{X}| = \sum_{\xi=1} \frac{(-1)^{\xi+1}}{\xi} tr(\mathbf{X}^\xi)$. However, differently from the study of the covariance matrix $\mathbf{\Lambda}_{h^{(\ell)}}$, thanks to the $\mathbf{W}$ integral, we already have a factor containing all of the variables $\lambda_{\alpha\beta}$ exactly once. Therefore, we can consider only the first order expansion, ie. $\xi = 1$. Further we are also going to consider the following change of variable,

$$\tilde{\lambda}_{\alpha\beta} = \sigma_w \sigma_x \lambda_{\alpha\beta} \tag{231}$$

therefore

$$E_{\Pi_i^{(l)}}^{(n)\ L.O.} \approx -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \int \left(\prod_{(\alpha,\beta)\in \mathcal{Z}_{\Pi_j}} \frac{\sigma_w \sigma_x}{\sqrt{n}} \lambda_{\alpha\beta} \right) e^{-\frac{\sigma_x^2 \sigma_w^2}{2} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - itr\mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \tag{232}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \int \left(\prod_{(\alpha,\beta)\in \mathcal{Z}_{\Pi_j}} \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} \right)$$

$$e^{-\sum_{(\alpha,\beta)\in \mathcal{Z}_{\Pi_j}} \left(\frac{\tilde{\lambda}_{\alpha\beta}^2}{2} + i \frac{\tilde{\lambda}_{\alpha\beta} z_{\alpha\beta}}{\sigma_x \sigma_w} \right)} F(z) \left(\prod_{(\alpha,\beta)\in \mathcal{Z}_{\Pi_j}} \frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi \sigma_w \sigma_x} \right) \left(\prod_{(\alpha,\beta)\in \mathcal{Z}_{\Pi_j}} dz_{\alpha\beta} \right) \tag{233}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \prod_{(\alpha\beta)\in \mathcal{Z}_{\Pi_j}} \left(\int \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} e^{-\frac{\tilde{\lambda}_{\alpha\beta}^2}{2} - i \frac{\tilde{\lambda}_{\alpha\beta} z_{\alpha\beta}}{\sigma_x \sigma_w}} \phi(z_{\alpha\beta}) \frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi \sigma_w \sigma_x} dz_{\alpha\beta} \right) \tag{234}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \left(\int \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} e^{-\frac{1}{2}\left(\tilde{\lambda}_{\alpha\beta} + \frac{iz_{\alpha\beta}}{\sigma_x \sigma_w}\right)^2 - \frac{z_{\alpha\beta}^2}{2\sigma_x^2 \sigma_w^2}} \phi(z_{\alpha\beta}) \frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi \sigma_w \sigma_x} dz_{\alpha\beta} \right)^{2l} \tag{235}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \left(\int \frac{\phi(z)}{\sqrt{2\pi}\sqrt{n}\sigma_w \sigma_x} e^{-\frac{z^2}{2\sigma_x^2 \sigma_w^2}} \left(\int \frac{\tilde{\lambda}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\tilde{\lambda} + \frac{iz}{\sigma_x \sigma_w}\right)^2} d\tilde{\lambda} \right) dz \right)^{2l} \tag{236}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \left(\int \frac{\phi(z)}{\sqrt{2\pi}\sqrt{n}\sigma_w \sigma_x} e^{-\frac{z^2}{2\sigma_x^2 \sigma_w^2}} \left(-\frac{iz}{\sigma_x \sigma_w} \right) dz \right)^{2l} \tag{237}$$

$$= -(-1)^l \frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n} \left(\frac{-1}{n} \right)^l \left(\int_{\mathbf{Z}} \frac{\phi(z)}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma_x^2 \sigma_w^2}} \frac{z}{\sigma_x^2 \sigma_w^2} dz \right)^{2l}. \tag{238}$$

Now we introduce the following change of variable

$$\tilde{z} = \frac{z}{\sigma_x \sigma_w} \tag{239}$$

and then

$$E_{\Pi_i^{(l)}}^{(n)\ L.O.} = -\frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n^{1+l}} \left(\int_{\mathbf{Z}} \frac{\tilde{z}\phi(\sigma_x \sigma_w \tilde{z})}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} d\tilde{z} \right)^{2l} \tag{240}$$

$$= -\frac{\sigma_w^2 \sigma_x^2 \theta_3^2}{n^{1+l}} \left(\sigma_w \sigma_x \int_{\mathbf{Z}} \frac{\phi'(\sigma_x \sigma_w \tilde{z})}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} d\tilde{z} \right)^{2l} \tag{241}$$

$$= -\frac{\sigma_w^2 \sigma_x^2 \theta_3^2 \theta_2^l}{n^{1+l}}. \tag{242}$$

The integral on $\mathcal{D}\lambda\mathcal{D}z$ for the remaining addends in (227) lead to the computation of higher moments of the gaussian variable λ, which are finite. However, $1/n$ is elevated to the same power as λ and therefore the contribution of the remaining addends in (227) is $E_{\Pi_i^{(l)}}^{(n)\ L.O.} \mathcal{O}(1/n)$.

It then follows that

$$E_{\Pi_i^{(l)}}^{(n)} = E_{\Pi_i^{(l)}}^{(n)\ L.O.} + E_{\Pi_i^{(l)}}^{(n)\ L.O.} \left(\mathcal{O}\left(\frac{1}{n} \right) \right) \tag{243}$$

$$= -\frac{\sigma_w^2 \sigma_x^2 \theta_3^2 \theta_2^l}{n^{1+l}} \left(1 + \mathcal{O}\left(\frac{1}{n} \right) \right). \tag{244}$$

□

B.2.1 Proof of Lemma A.3

Proof. Now we can use Lemma B.2 for the computation of E_{2k}. Let's assume $k > 1$, and compute the contribution of each term ω in the expansion of (63). Specifically each term has an associated number n_w of independent blocks Π_j, and we denote with $n_\phi^{(\xi)}$ and $n_w^{(\phi)}$ the number of factors of the type $\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ and $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}p}$ respectively within each block Π_ξ, note that $k = \sum_\xi (n_w^{(\xi)} + n_\phi^{(\xi)})$, and $n_w^{(\xi)} = 1$ for each block. Further, we also fix the index p in the factors $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}p}$, and then

$$E_\omega^{(k,n_w,n_\phi,p)} = \int_{\mathbf{W},\mathbf{X}} \prod_{\xi=1}^{n_w} \left(\sigma_x\theta_3\mathbf{W}_{i_\xi p} \left(\prod_{\ell=1}^{n_\phi^{(\xi)}} \phi(\sum_l \mathbf{W}_{i_{\xi+\ell},l}\mathbf{X}_{l,\mu_{\xi+\ell}})\phi(\sum_l \mathbf{W}_{i_{\xi+\ell+1},l}\mathbf{X}_{l,\mu_{\xi+\ell}}) \right) \sigma_x\theta_3\mathbf{W}_{i_{\xi+1}p} \right) \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{245}$$

$$= \prod_{\xi=1}^{n_w} E_{\Pi_\xi\left(n_\phi^{(\xi)}\right)} = \prod_{\xi=1}^{n_w} \frac{-\sigma_w^2\sigma_x^2\theta_3^2\theta_2^{n_\phi^{(\xi)}}}{n^{1+n_\phi^{(\xi)}}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{246}$$

$$= \left(\frac{-\sigma_w^2\sigma_x^2\theta_3^2}{n}\right)^{n_w}\left(\frac{\theta_2}{n}\right)^{\sum_{\xi=1}^{n_w} n_\phi^{(\xi)}}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{247}$$

$$= \left(-\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\right)^{n_w}\left(\frac{\theta_2}{n}\right)^{k-n_w}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{248}$$

The expected contribution for the $2k$ cycle is defined by considering the contribution $\tilde{E}_{2k}^\phi = n_0^{1-k}\theta_2^k\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$ of the addend ω whose $n_w = 0$ and of the addends for which $n_w \neq 0$

$$E_{2k}^{(n)} = \tilde{E}_{2k}^\phi + \sum_{n_w=1}^{k}\binom{k}{n_w}\sum_{p=1}^{n} E_\omega^{(k,n_w,n_\phi,p)} \tag{249}$$

$$= n_0^{1-k}\theta_2^k\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) + n_0^{1-k}\sum_{n_w=1}^{k}\binom{k}{n_w}\left(-\sigma_w^2\sigma_x^2\theta_3^2\right)^{n_w}\theta_2^{k-n_w}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{250}$$

$$= n_0^{1-k}\left(\theta_2 - \sigma_w^2\sigma_x^2\theta_3^2\right)^k\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{251}$$

□

B.3 Proof of Proposition 2

For each possible pattern we have to consider how many graphs, $\#_1$ and $\#_2$ we can build by varying respectively the I_i i-identifications, and the I_μ μ-identifications. Since, we have to assign for each of the $k - I_i$ i-identification a value among the n_1 available without reinsertion

$$\#_1 = \binom{n_1}{k - I_i} = \left(\frac{n_1 e}{k - I_i}\right)^{k-I_1}(2\pi(k - I_i))^{-1/2}(1 + o(\frac{1}{n_1})) \tag{252}$$

and equivalently

$$\#_2 = \binom{m}{k - I_\mu} = \left(\frac{me}{k - I_\mu}\right)^{k-I_\mu}(2\pi(k - I_\mu))^{-1/2}(1 + o(\frac{1}{m})). \tag{253}$$

Thanks to the formulation for the contribution of $2k$-cycles in Lemmas A.2 and A.3, we are now able to determine the contribution of any admissible graph G.

Proposition 3. *For $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, an admissible graph G with $2k$ edges, b blocks of size 1, c blocks of dimensions $\overrightarrow{k} = \{k_1, \ldots, k_c\}$, and b^μ of the I_μ identifications define a one dimensional cycle, while the remaining c^μ of the I_μ identifications define larger cycles, is such that the contribution of the graph, $E_G^{(n)}$, grows as*

$$E_G^{(n)} = n_0^{c-\sum_i k_i}\left(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{\sum_i k_i}\left(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{b^i+b^\mu}$$

$$\left[1+\frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2-\sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^2}\right]^{c^\mu}\left[1+\frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2+\sigma_x^2\sigma_w^2\theta_3^2\theta_2-\sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)}\right]^{b^\mu} \cdot\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{254}$$

Proof. When considering the contribution of a generic composite graph we have to keep into consideration what kind of identification defines it. Specifically, if an identification is of the i-type, then this allows to consider the two connected sub-graphs separately. However, if it is a μ-type contribution, then there is a correction term that has to be kept into consideration.

To give an intuition on why the correction term arises, and to introduce the cases that we have to investigate, we are going to consider a graph G generated by two cycles. When they are connected by an i-identification, the factors $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ always appear as a pair in each constituent cycle of the graph. However, if they are connected with a μ-identification, this is not the case; the factors $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ can be split by the identification in the two different cycles. We will show that if none of the four edges connected to the identification vertex are of the $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ kind, then the cycles can be considered separately, while if not, they have to be considered as one. Considering separately two cycles implies that it is possible to choose the column p of the matrices $\mathbf{w}$ independently, thus leading to a n_0 fold increase in the estimation.

Now we are going to consider the expected contribution of a general graph G, which is defined by the following integral

$$E_G^{(n)}=\int\prod_{\xi=1}^{k}\left(\phi(\sum_l\mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l\mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})-\sigma_x^2\sum_{p=1}^{n}\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}p}\theta_3^2\right)\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}. \tag{255}$$

We are going to compute the contribution $E_G^{(n)}$ by considering each term, ω, in the resulting sum that results from the expansion of (255). However, to simplify the computations, we are first going to consider the contribution of the different kinds of blocks of dependent variable Π_j that determine the addend ω. As in Lemma B.2, if there are no identifications between two successive terms $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}p}$ then a block is isolated to the factors of the type $\phi(\sum_l\mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l\mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ between two of the type $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ and their contribution is the same as in Lemma B.2. We refer to the contribution of these terms as ${}_0E_{\Pi_j}^{(n)}$.

Differently from the cycle case, if there are some identifications the blocks might be defined by more than two $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ factors. We consider the following three kind of identifications:

- in the block Π_j there are only two factors of the type $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ and there are some identifications on the vertices in between the two blocks, see Figure 9a.
- there are multiple factors $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ defining the block Π_j, but none is such that μ_ξ is an identification, see Figure 9b as an example;
- there is at least one factor $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ such that μ_ξ is an identification, see Figure 9c as an example. This corresponds to the term $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ being split into two different cycles.

We are now going to identify all of their contributions.

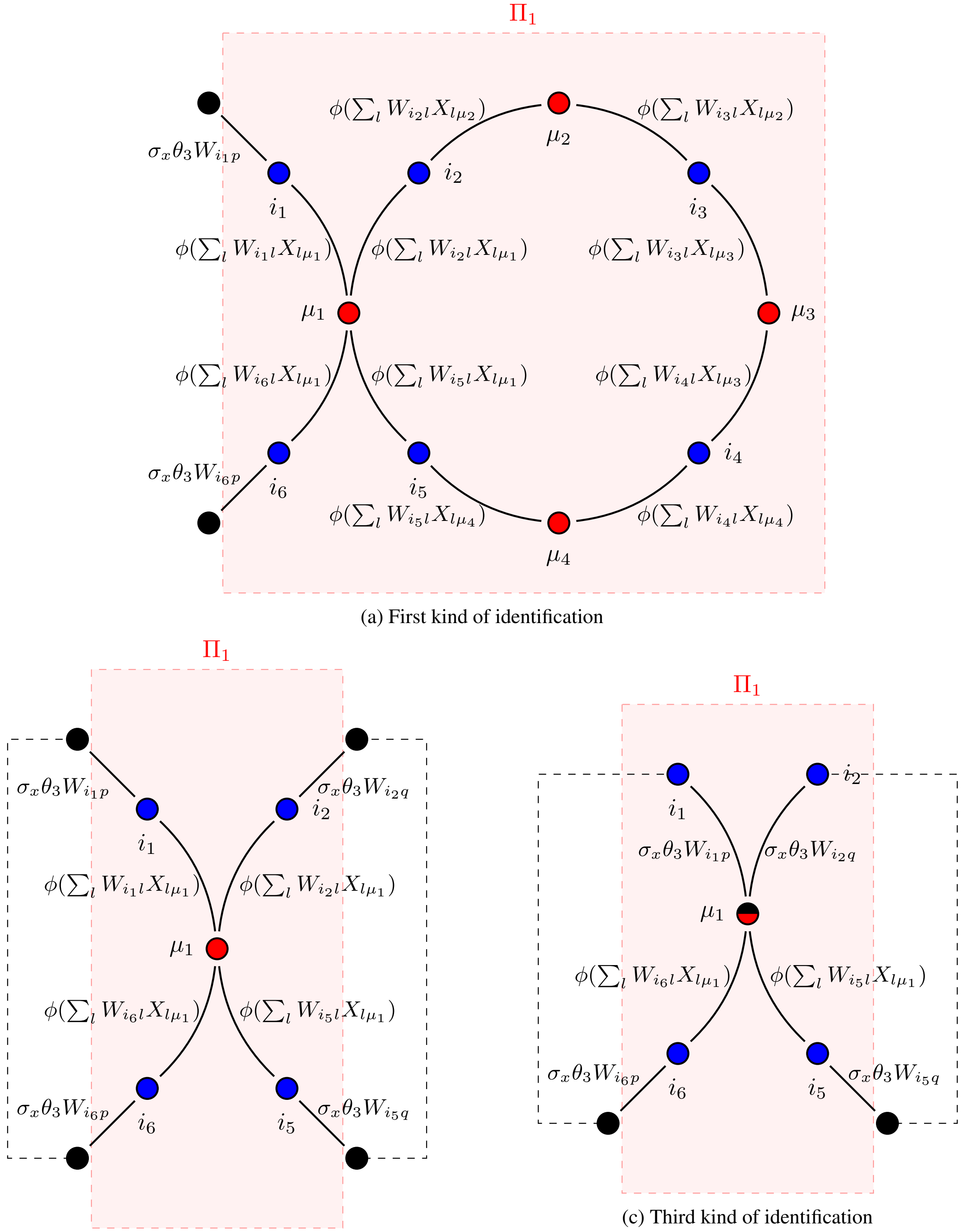
First Kind: Let's consider the integration over the block $\Pi_j^{(l,p)}$ with l factors $\phi(\sum_l\mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l\mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ and p specifying the column considered of $\mathbf{W}$[7]. Then

$${}_IE_{\Pi_i^{(l,p)}}^{(n)}=-\int\sigma_x\theta_3\mathbf{W}_{i_1p}\prod_{\xi=1}^{l}\left(\phi(\sum_l\mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l\mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})\right)\sigma_x\theta_3\mathbf{W}_{i_{l+1}p}\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}. \tag{256}$$

and we define n_{ind} as the number of identifications within the block; this corresponds to the number of complete cycles between the two factors $\sigma_x\theta_3\mathbf{W}_{i_1p}$ and $\sigma_x\theta_3\mathbf{W}_{i_{l+1}p}$. By integrating over $\mathbf{X}$ and $\mathbf{W}$ as done in Lemma B.2 to retrieve (223), we confirm that the two factors $\mathbf{W}_{i_1p}$ and $\mathbf{W}_{i_{l+1}p}$ must be relative to the same column and thus

$${}_IE_{\Pi_i^{(l,p)}}^{(n)}=-\int\sigma_x^2\theta_3^2(-1)^l\frac{\sigma_w^2}{n}Cov\left[\mathbf{w}_j^{i_0}\mathbf{w}_j^{(i_{l+1})}\right]\frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}}e^{-itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{257}$$

[7]If the two factors were relative to different columns p and q the contribution would be null



(a) First kind of identification

(b) Second kind of identification

(c) Third kind of identification

Figure 9: Illustration of some examples of how the $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1} q}$ factors might be arranged around some identifications.

Let's now denote the set of indices that are respectively on the shortest path between the two factors $\sigma_x\theta_3\mathbf{W}_{i_1p}$ and $\sigma_x\theta_3\mathbf{W}_{i_{l+1}p}$ or in the ith cycle in Π_j as $\mathfrak{D}$ and $\mathfrak{C}_i$. Then, by referring to all of the possible combinations of the different $\mathfrak{C}_i$ as $\pi(\mathfrak{C}_i)$, the covariance is

$$Cov\left[\mathbf{w}_j^{i_0}\mathbf{w}_j^{(i_{l+1})}\right] = \left[\left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{-1}\right]_{i_0 i_{l+1}} = \left[\frac{\sigma_w^2}{n}\sum_{k=0}^{\infty}(-\frac{\sigma_w^2\sigma_x^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k\right]_{i_0 i_{l+1}} \tag{258}$$

$$= \frac{\sigma_w^2}{n}\sum_{k=l}^{\infty}\left[(-\frac{\sigma_w^2\sigma_x^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^k\right]_{i_0 i_{l+1}} \tag{259}$$

$$= (-1)^{|\mathfrak{D}|}\frac{\sigma_w^2}{n}\prod_{(\alpha\beta)\in\mathfrak{D}}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta} + \sum_{\mathfrak{C}\in\pi(\mathfrak{C}_i)}(-1)^{|\mathfrak{D}|+|\mathfrak{C}|}\frac{\sigma_w^2}{n}\prod_{(\alpha\beta)\in\mathfrak{D}\cup\mathfrak{C}}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}$$
$$+\sum_{\nu=1}^{\infty}(-1)^{l+\nu}\frac{\sigma_w^2}{n}\left(\prod_{(\gamma,\delta)\in\mathfrak{D}}\left(\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\gamma\delta}\right)^{\tilde{l}_j}\right)\sum_{\substack{\overrightarrow{l}\ s.t.\\ \|\overrightarrow{l}\|_1=\nu\\ \tilde{l}_j\in\mathbb{N}_0}}\prod_{(\alpha,\beta)\in\cup_i\mathfrak{C}_i\cup\mathfrak{D}}\left(\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)^{\tilde{l}_j}. \tag{260}$$

where $|\mathfrak{C}|$ and $|\mathfrak{D}|$ correspond to the cardinality of the two set, hence twice the number of factors $\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ identified in the sets.

Because of the same argument as in Lemma B.2, the terms of the covariance where the variables $\lambda_{\alpha\beta}$ have exponential at maximum equal to one are the leading contribution of $E^{(n)}_{\Pi_j^{(l,p)}}$ up to a term $\mathcal{O}(1/n)$. Among these terms we now show that the leading contribution in the covariance is given by the term just relying on $\mathfrak{D}$; each time a cycle is considered, the contribution incurs a cost of $1/n$. By approximating the covariance with the contribution of the shortest path between the two factors of the $\sigma_x\theta_3\mathbf{W}_{i_{l+1}p}$ kind, we find the following contribution

$$-\int\sigma_x^2\theta_3^2(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2}{n}\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)\frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}}e^{-itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z$$
$$= -\int\sigma_x^2\theta_3^2(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2}{n}\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)e^{-\frac{n}{2}\log det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)}e^{-itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{261}$$
$$= -(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\int\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)$$
$$e^{-\frac{\sigma_x^2\sigma_w^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)-\frac{n}{2}\sum_{\xi\geq2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{\xi}-itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{262}$$

The leading order of the contribution ${}_IE^{(n)}_{\Pi_i^{(l,\mathbf{p})}}$ is defined with a relative error $\mathcal{O}(\frac{1}{n})$ by considering the different cycles separately. Moreover, as for Lemma B.2 there is a relative error of $\mathcal{O}(1/n)$ by considering only the zeroth order approximation of the exponent $e^{\frac{n}{2}\sum_{\xi\geq2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{\xi}}$ when considering the direct arch $\mathcal{D}$ and first order for the remaining cycles. Therefore

$$-(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\int\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)F(z)$$
$$exp\left\{-\sum_{\alpha\beta\in\mathfrak{D}\cup_{i=1}^{n_{ind}}\mathfrak{C}_i}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2 - i\lambda_{\alpha\beta}z_{\alpha\beta}\right)-\frac{n}{2}\sum_{\xi\geq2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^{\xi}\right\}\mathcal{D}\lambda\mathcal{D}z$$
$$= -(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\int\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\right)F(z)$$

$$exp\left\{-\sum_{\alpha\beta\in\mathfrak{D}\cup_{i=1}^{n_{ind}}\mathfrak{C}_i}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2-i\lambda_{\alpha\beta}z_{\alpha\beta}\right)-\frac{n}{2}\sum_{i=1}^{n_{ind}}\frac{(-1)^{|\mathfrak{C}_i|+1}}{|\mathfrak{C}_i|}\prod_{(\alpha\beta)\in\mathfrak{C}_i}\frac{\sigma_x\sigma_w}{\sqrt{n}}\lambda_{\alpha\beta}\right\}\mathcal{D}\lambda\mathcal{D}z$$

$$\cdot\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{263}$$

$$=-(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\int\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\phi(z_{\alpha\beta})e^{-\sum_{\alpha\beta\in\mathfrak{D}}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2-i\lambda_{\alpha\beta}z_{\alpha\beta}\right)}\right)$$

$$\left(\prod_{i=1}^{n_{ind}}(\prod_{\alpha\beta\in\mathfrak{C}_i}\phi(z_{\alpha\beta}))e^{-\sum_{\alpha\beta\in\mathfrak{C}_i}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2-i\lambda_{\alpha\beta}z_{\alpha\beta}\right)+\frac{n}{2}\frac{(-1)^{|\mathfrak{C}_i|}}{|\mathfrak{C}_i|}2|\mathfrak{C}_i|\prod_{(\alpha\beta)\in\mathfrak{C}_i}\frac{\sigma_x\sigma_w}{\sqrt{n}}\lambda_{\alpha\beta}}\right)\mathcal{D}\lambda\mathcal{D}z$$

$$\cdot\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{264}$$

$$=-(-1)^{|\mathfrak{D}|}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\int\underbrace{\left(\prod_{(\alpha\beta)\in\mathfrak{D}}^{|\mathfrak{D}|}\frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{\alpha\beta}\phi(z_{\alpha\beta})e^{-\sum_{\alpha\beta\in\mathfrak{D}}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2-i\lambda_{\alpha\beta}z_{\alpha\beta}\right)}\right)}_{=(-\theta_3/n)^{|\mathfrak{D}|}}$$

$$\left(\prod_{i=1}^{n_{ind}}(\prod_{\alpha\beta\in\mathfrak{C}_i}n(-1)^{|\mathfrak{C}_i|}\underbrace{\phi(z_{\alpha\beta})\frac{\sigma_x\sigma_w}{\sqrt{n}}\lambda_{\alpha\beta})exp\left\{-\sum_{\alpha\beta\in\mathfrak{C}_i}\left(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2-i\lambda_{\alpha\beta}z_{\alpha\beta}\right)\right\}}_{=(-\theta_2/n)^{|\mathfrak{C}_i|}}\right)\mathcal{D}\lambda\mathcal{D}z$$

$$\cdot\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{265}$$

$$=\frac{-\sigma_w^2\sigma_x^2\theta_3^2}{n}n^{n_{ind}}\left(\frac{\theta_2}{n}\right)^{|\mathfrak{D}|+\sum_{i=1}^{n_{ind}}|\mathfrak{C}_i|}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{266}$$

$$=\frac{-\sigma_w^2\sigma_x^2\theta_3^2}{n}\left(\frac{\theta_2}{n}\right)^{|\mathfrak{D}|}\prod_{i=1}^{n_{ind}}E_{2|\mathfrak{C}_i|}^{(n)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{267}$$

where the identities in (265) are taken from the study of the $2k$ cycles in (234)-(244), and $E_{2|\mathfrak{C}_i|}^{(n)}$ is the contribution of a cycle of dimension $2|\mathfrak{C}_i|$. Note that the last step is justified since if $|\mathfrak{C}_i|=1$ we would have retrieved θ_1 rather than $n^{1-|\mathfrak{C}_i|}\theta_2^{|\mathfrak{C}_i|}$.

If instead we considered one of the expansion terms of the covariance also including one of the sets $\mathfrak{C}_i$, then the relative variables $\lambda_{\alpha\beta}$ would already appear in line (261), and the contribution in the log-determinant would not be relevant. Consequently the n term resulting from the expansion of the exponential of the log-determinant would not appear and we would have a relative contribution of $\mathcal{O}(1/n)$.

Therefore the contribution if the first case is

$$_IE_{\Pi_j^{(l,p)}}^{(n)}=-n^{n_{ind}}\frac{\sigma_w^2\sigma_x^2\theta_3^2}{n}\left(\frac{\theta_2}{n}\right)^l\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{268}$$

and crucially it shows that the contribution of a cycle with no factors of the type $\sigma_x\theta_3\mathbf{W}_{i_1p}$ can be considered independently from the block.

Second Kind: Since the block of dependent variables might contain multiple weights in the form of $\sigma_x\theta_3\mathbf{W}_{i_1p}$ we have to first identify the arches within the block, i.e. sequences of successive i_ξ that are all within the block. With this purpose, we consider the set of indices that define the beginning and the end of each arch

$$\mathfrak{A}=\{(i_\xi,i_{\xi+l})\mid\exists\{\mu_j\}_{j=1}^l s.t.\cup_{j=1}^l\left((i_{\xi+j-1},\mu_j)\cup(i_{\xi+j},\mu_j)\right)\subset\Pi_j\}. \tag{269}$$

This allows to determine the contribution of $\Pi_j^{(l,\mathbf{p})}$ where l is the number of factors $\phi(\sum_l\mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l\mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ in Π_j and $\mathbf{p}$ is a vector containing the indices of the columns of $\mathbf{W}$

that are considered in each cycle that Π_j intersects, and therefore has dimension $|\mathfrak{A}|$. We rewrite the expected contribution in terms of the arches which define Π_j

$$ {}_{II}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}} = \int_{\mathbf{W},\mathbf{X}} \prod_{(i_\alpha,i_{\alpha+l})\in\mathfrak{A}} \Bigg(-\sigma_x^2\theta_3^2\mathbf{W}_{i_\alpha p_\alpha} \cdot \prod_{\xi=1}^{l}\left(\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})\right)\mathbf{W}_{i_{\alpha+l+1}q_\alpha}\Bigg)\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}. \tag{270}$$

where p_α and q_α are the dependent on the cycle in which $\mathbf{W}_{i_\alpha p}$ is. By integrating over $\mathbf{X}$ and $\mathbf{W}$ we implement similar steps as in Lemma B.2 were taken to to determine (223), we find that

$$ {}_{II}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}} = (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\int \mathbb{E}\left[\prod_{(i_\alpha,i_{\alpha+l})\in\mathfrak{A}} \mathbf{w}_{p_\alpha}^{(i_\alpha)}\mathbf{w}_{q_\alpha}^{(i_\alpha+l)}\right]\frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}}e^{tr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z. \tag{271}$$

The weight vectors $\mathbf{w}_{p_\alpha}$ are distributed according to a Gaussian, therefore we can rely on Isserlis' theorem [8]. If $(X_1,\dots,X_n)$ is a zero-mean multivariate normal random vector, then

$$\mathbb{E}\left[X_1X_2\cdots X_n\right] = \sum_{p\in P_n^2}\prod_{\{i,j\}\in p}\mathbb{E}\left[X_iX_j\right] = \sum_{p\in P_n^2}\prod_{\{i,j\}\in p} Cov\left(X_i,X_j\right) \tag{272}$$

where the sum is over all the pairings of $\{1,\dots,n\}$. For example, when we consider four variables then we get that

$$\begin{aligned}\mathbb{E}\left[X_1X_2X_3X_4\right] &= Cov\left(X_1,X_2\right)Cov\left(X_3,X_4\right)\\ &\quad + Cov\left(X_1,X_3\right)Cov\left(X_2,X_4\right) + Cov\left(X_1,X_4\right)Cov\left(X_2,X_3\right).\end{aligned} \tag{273}$$

We are going to consider the term in the expansion (272) where the covariances are only between elements in the same cycle, i.e. p_α is the same within the same cycle. The other terms have a lower order contribution to the moments because they imply two separate cycles being relative to the same sample p_α, and this implies that total contribution incurs in a cost of at least $1/n_0$.

Therefore, let's compute the contribution by defining the set of all the indices in a cycle as $\mathfrak{C}_i$, then

$$ {}_{II}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}} = (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\int \mathbb{E}\left[\prod_{(i_\alpha,i_{\alpha+l})\in\mathfrak{A}} \mathbf{w}_{p_\alpha}^{(i_\alpha)}\mathbf{w}_{q_\alpha}^{(i_\alpha+l)}\right]\frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}}e^{tr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{274}$$

$$= (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\int\left(\prod_{i=1}^{n_{int}} Cov\left(\mathbf{w}_{p_\alpha}^{i_\alpha}\mathbf{w}_{p_\alpha}^{i_\beta}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \frac{1}{det(\mathbf{I}+\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{n/2}}e^{tr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{275}$$

$$= (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\int\left(\prod_{i=1}^{n_{int}} Cov\left(\mathbf{w}_{p_\alpha}^{i_\alpha}\mathbf{w}_{p_\alpha}^{i_\beta}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) e^{-\frac{\sigma_x^2\sigma_w^2}{2}tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)-\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi - itr\mathbf{\Lambda}^\top\mathbf{Z}}F(z)\mathcal{D}\lambda\mathcal{D}z \tag{276}$$

The leading order of the contribution ${}_{II}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}}$ is defined with a relative error $\mathcal{O}(\frac{1}{n})$ by considering the different cycles separately. Moreover, as for Lemma B.2 there is a relative error of $\mathcal{O}(1/n)$ by considering only the zeroth order approximation of the exponent $e^{\frac{n}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi}$. Therefore

$$ {}_{II}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}} = (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\int\left(\prod_{i=1}^{n_{int}} Cov\left(\mathbf{w}_{p_\alpha}^{i_\alpha}\mathbf{w}_{p_\alpha}^{i_\beta}\right)\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) e^{-\sum_{i=1}^{n_{int}}\left(\sum_{\alpha\beta\in\mathfrak{C}_i}(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2+i\lambda_{\alpha\beta}z_{\alpha\beta})\right)}F(z)\mathcal{D}\lambda\mathcal{D}z\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{277}$$

$$= (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\prod_{i=1}^{n_{int}}\int\left(Cov\left(\mathbf{w}_{p_\alpha}^{i_\alpha}\mathbf{w}_{p_\alpha}^{i_\beta}\right)\right)$$
$$e^{-\sum_{\alpha\beta\in\mathfrak{C}_i}(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2+i\lambda_{\alpha\beta}z_{\alpha\beta})}F(z)\mathcal{D}\lambda\mathcal{D}z\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{278}$$

$$= (-\sigma_x^2\theta_3^2)^{|\mathfrak{A}|}\prod_{i=1}^{n_{int}}\int\frac{\sigma_w^2}{n}\left(\prod_{\alpha\beta\in\mathfrak{C}_i}\lambda_{\alpha\beta}\phi(z_{\alpha\beta})\right)$$
$$e^{-\sum_{\alpha\beta\in\mathfrak{C}_i}(\frac{\sigma_x^2\sigma_w^2}{2}\lambda_{\alpha\beta}^2+i\lambda_{\alpha\beta}z_{\alpha\beta})}\mathcal{D}\lambda\mathcal{D}z\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{279}$$

$$= (-\frac{\sigma_x^2\theta_3^2\sigma_w^2}{n})^{|\mathfrak{A}|}\left(\frac{\theta_2}{n}\right)^l\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{280}$$

Note that if there is a complete cycle without factors of the type $\sigma_x\theta_3\mathbf{W}_{i_{l+1}p}$ within Π_j, then we can rely on the first case to separate its contribution obtaining a further term n_0.

Third Kind: This case is uniquely due to μ-identifications. In this case, if we have an identification at μ_ξ with $(i_\xi, i_{\xi+1}) \in \mathfrak{A}$ then we have to keep the same sample p_α in both the cycles in which $\mathfrak{A}$ is defined. Therefore

$${}_{III}E^{(n)}_{\Pi_i^{(l,\mathbf{p})}} = (-\frac{\sigma_x^2\theta_3^2\sigma_w^2}{n})^{|\mathfrak{A}|}\left(\frac{\theta_2}{n}\right)^l\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)\mathbb{1}_{\mathbf{p}=p\mathbf{1}} \tag{281}$$

where

$$\mathbb{1}_{\mathbf{p}=p\mathbf{1}} = \begin{cases}1 & \text{if } \mathbf{p}=p\mathbf{1} \text{ with } p\in[n_0]\\ 0 & \text{otherwise.}\end{cases} \tag{282}$$

and $\mathbf{1}$ is a vector whose entries are all 1.

Combination of the Contributions Let's now consider how the contribution of each identification affects the contribution of one term ω of the expansion of (255). For simplicity let's start by considering the case where there are no 2-dimensional cycles in the graph G, and denote with $\mathbf{p} = (p_1, \ldots, p_c)$ the column index of $\mathbf{W}$ for each of the c cycle, then

$$E_\omega^{(n,\mathbf{p})} = \prod_{j=1}^{n_{groups}}\left({}_{type(j)}E^{(n)}_{\Pi_j^{(l_j,p_j)}}\mathbb{1}_{p=\mathbf{p}_{\alpha(j)}}\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{283}$$

$$= \left(\prod_{j=1}^{n_{groups}}\left(n_0^{n_{ind}^{(j)}}(-\frac{\sigma_x^2\theta_3^2\sigma_w^2}{n})^{|\mathfrak{A}_j|}\left(\frac{\theta_2}{n}\right)^{l_j}\mathbb{1}_{p=\mathbf{p}_{\alpha(j)}}\right)\right)\left(\prod_{(\alpha,\beta)\in\{\substack{\text{cycles connected}\\ \text{by third kind}\\ \text{identifications}}\}}\mathbb{1}_{\mathbf{p}_\alpha=\mathbf{p}_\beta}\right)$$
$$\cdot\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{284}$$

where the functions $type(\cdot) : [n_{groups}] \to \{0, I, II, III\}$ and $\alpha(\cdot) : [n_{groups}] \to [n_0]$ assign to each block Π_j the typology of identification in it and the index of the column of $\mathbf{W}$ considered, and $\mathfrak{A}_j$ identifies the set of arches within a cycle for each block Π_j. This suggests that when considering the contribution of a term ω of a graph with no 2-dimensional cycles we need to keep track of the third kind of identifications since they are the only case requiring two adjacent cycles to have the same index p_α.

When considering also cycles of dimension 2, we have to distinguish between 3 cases if they are connected with a μ-identification: if the two edges within the 2-dimensional cycle are of the $\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ type, if they are of the $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ type, or if they are alternate. In the first case, their contribution can be seen as a simple cycle and therefore we gain a factor θ_1/n rather than $\left(\frac{\theta_2}{n}\right)^l$ in equation 284. In the second case we consider the 2 cycle as a simple block Π_j of dimension one with contribution $(-\frac{\sigma_x^2\theta_3^2\sigma_w^2}{n})$, but since this is a third case identification we need to ensure that the index p is the same between the two adjacent cycles. Finally, in the third case, we are also considering a third case identification where the intra-dependent variables of the block are the $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ edges in the

2-dimensional cycle and in the adjacent one, and all of the $\phi(\sum_l \mathbf{W}_{i_\xi,l}\mathbf{X}_{l,\mu_\xi})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}\mathbf{X}_{l,\mu_\xi})$ edges in-between them. This implies that we gain a factor $(-\frac{\sigma_x^2\theta_3^2\sigma_w^2}{n})\left(\frac{\theta_2}{n}\right)^l\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$ with the constraining that the index of the weight is the same as in the adjacent cycle. Note that in this case we do not have any θ_1 contribution although there is a 2-dimensional cycle.

From the above statements, it follows that there is no need to keep track of all the independent blocks Π_j to compute $E_\omega^{(n,\mathbf{p})}$; it is sufficient to identify the terms $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$, the third kind identifications, and the structure of 2-dimensional cycles. To keep track of all of these cases we will consider the following notation for each graph. We define b^μ as the number μ-identifications where at least one of the two joined cycles is of dimension 2, and c^μ as the remaining μ-identifications, i.e. $I_\mu = c^\mu + b^\mu$. We then introduce the variables $b^\mu_w \in \{0,\dots,b^\mu\}$ and $c^\mu_w \in \{0,\dots,c^\mu_w\}$ which determine how many of the b^μ and c^μ identifications have at least two edges of the $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ kind connected to the identification and therefore are of the third kind. Finally, we introduce the variables $b_w^{1/2}\in\{0,\dots,b^\mu_w\}$ and $c_w^{1/2}\in\{0,\dots,c^\mu_w\}$ which indicate if the remaining two edges are of the $\sigma_x\theta_3\mathbf{W}_{i_\xi p}$ type.

Therefore, an addend ω of a graph with $2k$ edges defined by $b^i,\ b^i_w,\ b^\mu,\ b^\mu_w,\ b_w^{1/2},\ c^i,\ c^i_w,\ c^\mu,\ c^\mu_w,\ c_w^{1/2}$ generates the following contribution

$$
\begin{aligned}
&E_\omega^{(k,\mathbf{p},b^i,b^i_w,b^\mu,b^\mu_w,b_w^{1/2},c^i,c^i_w,c^\mu,c^\mu_w,c_w^{1/2})} = \\
&= \underbrace{\underbrace{\left(\frac{\theta_1}{n}\right)^{k-b^i-2b^\mu-2c^\mu-(n_w-b^\mu_w-b_w^{1/2}-c^\mu_w-c_w^{1/2})}}_{\text{Edges of the }\phi\phi\text{ type}}\underbrace{\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{(n_w-b^\mu_w-b_w^{1/2}-c^\mu_w-c_w^{1/2})}}_{\text{Edges of the ww type}}}_{\text{Edges not adjacent to }\mu\text{-identification and in cycle larger than 2}} \\
&\cdot\underbrace{\underbrace{\left(\frac{\theta_2}{n}\right)^{b^\mu_w-b_w^{1/2}}\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^\mu_w-b_w^{1/2}}}_{\text{Mixed Case}}\underbrace{\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{2b_w^{1/2}}}_{\text{Edges of the }ww\text{ type}}\underbrace{\left(\frac{\theta_1}{n}\right)^{b^\mu-b^\mu_w}\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^\mu-b^\mu_w}}_{\text{Edges of the }\phi\phi\text{ type}}}_{\text{Edges adjacent to }\mu\text{-identification with a cycle of dimension 2}} \\
&\cdot\underbrace{\underbrace{\left(\frac{\theta_2}{n}\right)^{c^\mu_w-c_w^{1/2}}\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{c^\mu_w-c_w^{1/2}}}_{\text{Mixed Case}}\underbrace{\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{2c_w^{1/2}}}_{\text{Edges of the }ww\text{ type}}\underbrace{\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{2(c^\mu-c^\mu_w)}}_{\text{Edges of the }\phi\phi\text{ type}}}_{\text{Edges adjacent to }\mu\text{-identification and in cycles larger than 2}} \\
&\cdot\underbrace{\left(\frac{\theta_1}{n}\right)^{b^i-b^i_w}\left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^i_w}}_{\substack{\text{Adjacent to i-identification}\\ \text{and with a cycle of dimension 2}}}\left(\prod_{\substack{(\alpha,\beta)\in\{\text{cycles connected}\\ \text{by third kind}\\ \text{identifications}\}}}\mathbb{1}_{\mathbf{p}_\alpha=\mathbf{p}_\beta}\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right).
\end{aligned}
\tag{285}
$$

For each of the addends ω, there are $n_0^{1+I_\mu+I_i}$ ways to choose the indices p for each of the $1+I_\mu+I_i$ cycles. However, for all the identifications of third kind there is a constraint on adjacent cycles having the same index p, therefore there are actually $1+I_\mu+I_i-b^\mu_w-c^\mu_w$ non zero terms ω sharing the same structure $(k,\cdot,b^i,b^i_w,b^\mu,b^\mu_w,b_w^{1/2},c^i,c^i_w,c^\mu,c^\mu_w,c_w^{1/2})$ and therefore having the same contribution.

Now let's consider the contribution of the full graph by considering all of the combinations on how the structure $(k,\mathbf{p},b^i,b^i_w,b^\mu,b^\mu_w,b_w^{1/2},c^i,c^i_w,c^\mu,c^\mu_w,c_w^{1/2})$ may change. Let's start by considering the edges located in any of the $\sum_i k_i - b^\mu - 2^{c^\mu}$ pair of edges that are not adjacent to a μ-identification and are in cycles bigger than 2, there are $\binom{k-b^i-2b^\mu-2^{c^\mu}}{l}$ ways for l terms $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ to be located. Similarly, there are $\binom{b^i}{b^i_w}$ ways to have b^i_w 2-dimensional cycles whose edges are of the $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ kind. Considering the I_μ identifications, there are $\binom{b^\mu}{b^\mu_w}$ and $\binom{c^\mu}{c^\mu_w}$ ways to choose only one pair in the identification to be of the $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ type for the case with 2-dimensional cycles or higher dimensions. Finally if a μ-identification is such that all the four edges are of the $\sigma_x^2\theta_3^2\mathbf{W}_{i_\xi p}\mathbf{W}_{i_{\xi+1}q}$ kind, there are $\binom{b^\mu}{b_w^{1/2}}$ and $\binom{c^\mu}{c_w^{1/2}}$ choices in the $b_w^{1/2}$ and $c_w^{1/2}$ cases.

Combining all of the contributions above, it is possible to compute the contribution of a graph G with $2k$ edges and with I_μ and I_i identifications, of which b^μ and b^i generate a 2-dimensional cycle.

$$E_{\mathcal{G}} = \sum_{l}^{k-b^i-2b^\mu-2c^\mu} \sum_{c_w^\mu}^{c^\mu} \sum_{c_w^{1/2}}^{c_w^\mu} \sum_{b_w^\mu}^{b^\mu} \sum_{b_w^{1/2}}^{b_w^\mu} \sum_{b_w^i}^{b^i} \binom{k-b^i-2b^\mu-2c^\mu}{l} \binom{c^\mu}{c_w^\mu} \binom{c_w^\mu}{c_w^{1/2}} \binom{b^\mu}{b_w^\mu} \binom{b_w^\mu}{b_w^{1/2}} \binom{b^i}{b_w^i}$$
$$\cdot E_\omega^{(k,\mathbf{p},b^i,b_w^i,b^\mu,b_w^\mu,b_w^{1/2},c^i,c_w^i,c^\mu,c_w^\mu,c_w^{1/2})} n_0^{1+I_\mu+I_i-b_w^\mu-c_w^\mu} \tag{286}$$
$$= \sum_{l}^{k-b^i-2b^\mu-2c^\mu} \sum_{c_w^\mu}^{c^\mu} \sum_{c_w^{1/2}}^{c_w^\mu} \sum_{b_w^\mu}^{b^\mu} \sum_{b_w^{1/2}}^{b_w^\mu} \sum_{b_w^i}^{b^i} \binom{k-b^i-2b^\mu-2c^\mu}{l} \left(\frac{\theta_2}{n}\right)^{k-b^i-2b^\mu-2c^\mu-l} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{l}$$
$$\binom{c^\mu}{c_w^\mu} \left(\frac{\theta_2}{n}\right)^{c^\mu-c_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{c_w^\mu} \binom{c_w^\mu}{c_w^{1/2}} \left(\frac{\theta_2}{n}\right)^{c^\mu-c_w^{1/2}} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{c_w^{1/2}}$$
$$\binom{b^\mu}{b_w^\mu} \left(\frac{\theta_1}{n}\right)^{b^\mu-b_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^\mu} \binom{b_w^\mu}{b_w^{1/2}} \left(\frac{\theta_2}{n}\right)^{b^\mu-b_w^{1/2}} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^{1/2}}$$
$$\binom{b^i}{b_w^i} \left(\frac{\theta_1}{n}\right)^{b^i-b_w^i} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^i} n_0^{1+I_\mu+I_i-b_w^\mu-c_w^\mu} \left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{287}$$

Let's now group all the factors that are independent from each other and consider their sum and consider $\left(\frac{\theta_2}{n}\right)^{c^\mu-c_w^{1/2}} = \left(\frac{\theta_2}{n}\right)^{c^\mu-c_w^\mu}\left(\frac{\theta_2}{n}\right)^{c_w^\mu-c_w^{1/2}}$ and $\left(\frac{\theta_2}{n}\right)^{b^\mu-b_w^{1/2}} = \left(\frac{\theta_2}{n}\right)^{b^\mu-b_w^\mu}\left(\frac{\theta_2}{n}\right)^{b_w^\mu-b_w^{1/2}}$.

$$E_{\mathcal{G}} = n_0^{1+I_\mu+I_i} \left[\sum_{l}^{k-b^i-2b^\mu-2c^\mu} \binom{k-b^i-2b^\mu-2c^\mu}{l} \left(\frac{\theta_2}{n}\right)^{k-b^i-2b^\mu-2c^\mu-l} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{l}\right]$$
$$\cdot \left[\sum_{b_w^i}^{b^i} \binom{b^i}{b_w^i} \left(\frac{\theta_1}{n}\right)^{b^i-b_w^i} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^i}\right]$$
$$\cdot \left[\sum_{c_w^\mu}^{c^\mu} \binom{c^\mu}{c_w^\mu} \left(\frac{\theta_2^2}{n^2}\right)^{c^\mu-c_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\right)^{c_w^\mu}\right.$$
$$\left. \cdot \left(\sum_{c_w^{1/2}}^{c_w^\mu} \binom{c_w^\mu}{c_w^{1/2}} \left(\frac{\theta_2}{n}\right)^{c_w^\mu-c_w^{1/2}} \left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{c_w^{1/2}}\right)\right]$$
$$\cdot \left[\sum_{b_w^\mu}^{b^\mu} \binom{b^\mu}{b_w^\mu} \left(\frac{\theta_1\theta_2}{n^2}\right)^{b^\mu-b_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\right)^{b_w^\mu}\right.$$
$$\left. \cdot \left(\sum_{b_w^{1/2}}^{b_w^\mu} \binom{b_w^\mu}{b_w^{1/2}} \left(\frac{\theta_2}{n}\right)^{b_w^\mu-b_w^{1/2}} \left(\frac{-\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^{1/2}}\right)\right] \left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{288}$$
$$= n_0^{1+I_\mu+I_i} \left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{k-b^i-2b^\mu-2c^\mu} \left(\frac{\theta_1}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^i}$$
$$\cdot \left[\sum_{c_w^\mu}^{c^\mu} \binom{c^\mu}{c_w^\mu} \left(\frac{\theta_2^2}{n^2}\right)^{c^\mu-c_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\right)^{c_w^\mu} \left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{c_w^\mu}\right]$$
$$\cdot \left[\sum_{b_w^\mu}^{b^\mu} \binom{b^\mu}{b_w^\mu} \left(\frac{\theta_1\theta_2}{n^2}\right)^{b^\mu-b_w^\mu} \left(-\frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\right)^{b_w^\mu} \left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b_w^\mu}\right] \left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{289}$$
$$= n_0^{1+I_\mu+I_i} \left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{k-b^i-2b^\mu-2c^\mu} \left(\frac{\theta_1}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^i}$$

$$\cdot \left[\frac{\theta_2^2}{n^2} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)\right]^{c^\mu} \left[\frac{\theta_1\theta_2}{n^2} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n^2}\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)\right]^{b^\mu}$$
$$\cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{290}$$
$$= n_0^{1+I_\mu+I_i}\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{k-b^i-2b^\mu-2c^\mu}\left(\frac{\theta_1}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^i}$$
$$\cdot \left[\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^2\left(1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2} + \mathcal{O}\left(\frac{1}{n}\right)\right)\right]^{c^\mu}$$
$$\cdot \left[\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)\left(\frac{\theta_1}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)\left(1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)} + \mathcal{O}\left(\frac{1}{n}\right)\right)\right]^{b^\mu}$$
$$\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{291}$$
$$= n_0^{1+I_\mu+I_i}\left(\frac{\theta_2}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{k-b^i-b^\mu}\left(\frac{\theta_1}{n} - \frac{\sigma_x^2\sigma_w^2\theta_3^2}{n}\right)^{b^i+b^\mu}$$
$$\left[1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2} + \mathcal{O}\left(\frac{1}{n}\right)\right]^{c^\mu}$$
$$\left[1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)} + \mathcal{O}\left(\frac{1}{n}\right)\right]^{b^\mu}$$
$$\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{292}$$
$$= n_0^{1+I_\mu+I_i-k}\left(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{\sum_i k_i}\left(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{b^i+b^\mu}$$
$$\left[1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2}\right]^{c^\mu}\left[1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}\right]^{b^\mu}$$
$$\cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{293}$$
$$= n_0^{c-\sum_i k_i}\left(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{\sum_i k_i}\left(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2\right)^{b^i+b^\mu}$$
$$\left[1 + \frac{2\sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^2}\right]^{c^\mu}\left[1 + \frac{\theta_1\sigma_x^2\sigma_w^2\theta_3^2 + \sigma_x^2\sigma_w^2\theta_3^2\theta_2 - \sigma_x^4\sigma_w^4\theta_3^4}{(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)(\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)}\right]^{b^\mu}$$
$$\cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{294}$$

where in the last inequality we used the two equalities $k = \sum_i k_i + b^i + b^\mu$ and $1 + I_\mu + I_i = c + b^i + b^\mu$. □

B.3.1 Proof of Proposition 2

Proof.

$$\mathfrak{m}_k^{(n)} = \frac{1}{n_1 m^k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)E_{\mathcal{G}(k,I_i,I_\mu,b,b^\mu,c^\mu)}\#_1\#_2 \tag{295}$$
$$= \frac{1}{n_1 m^k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)n_0^{1+I_i+I_\mu-b-(k-b)}.$$
$$\cdot (\theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^{k-b}(\theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^b\kappa_b^{b^\mu}\kappa_c^{I_\mu-b^\mu}n_1^{k-I_1}m^{k-I_\mu}\left(1 + \mathcal{O}\left(\frac{1}{n_0}\right)\right)$$
$$\cdot \left(\frac{e}{k-I_i}\right)^{k-I_1}(2\pi(k-I_i))^{-1/2}\left(\frac{e}{k-I_\mu}\right)^{k-I_\mu}(2\pi(k-I_\mu))^{-1/2}$$

$$\cdot\left(1+o\left(\frac{1}{m}\right)\right)\left(1+o\left(\frac{1}{n_1}\right)\right) \tag{296}$$

$$= \sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)^b\cdot$$
$$\cdot(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^{k-b}\frac{n_1^{k-1}}{n_0^{k-1}}\frac{n_0^{I_1}}{n_1^{I_1}}\frac{n_0^{I_\mu}}{m^{I_\mu}}\kappa_b^{b^\mu}\kappa_c^{I_\mu-b^\mu}$$
$$\cdot\left(\frac{e}{k-I_i}\right)^{k-I_1}\left(\frac{e}{k-I_\mu}\right)^{k-I_\mu}\frac{1}{2\pi\sqrt{(k-I_i)(k-I_\mu)}}$$
$$\cdot\left(1+\mathcal{O}\left(\frac{1}{n_0}\right)\right) \tag{297}$$

$$= \varphi^{1-k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\frac{\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)}{2\pi\sqrt{(k-I_i)(k-I_\mu)}}(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)^b\cdot$$
$$\cdot(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^{k-b}\kappa_b^{b^\mu}\kappa_c^{I_\mu-b^\mu}\varphi^{I_1}\psi^{I_\mu}$$
$$\cdot\left(\frac{e}{k-I_i}\right)^{k-I_1}\left(\frac{e}{k-I_\mu}\right)^{k-I_\mu}\left(1+\mathcal{O}\left(\frac{1}{n_0}\right)\right) \tag{298}$$

where we used the coefficients $\varphi = n_0/n_1$, $\psi = n_0/m$. Therefore

$$\mathfrak{m}_k = \lim_{n\to\infty}\mathfrak{m}_k^{(n)} \tag{299}$$
$$= \varphi^{1-k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\frac{\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)}{2\pi\sqrt{(k-I_i)(k-I_\mu)}}(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)^b\cdot$$
$$\cdot(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^{k-b}\kappa_b^{b^\mu}\kappa_c^{I_\mu-b^\mu}\varphi^{I_1}\psi^{I_\mu}\left(\frac{e}{k-I_i}\right)^{k-I_1}\left(\frac{e}{k-I_\mu}\right)^{k-I_\mu} \tag{300}$$

and $|\mathfrak{m}_k^{(n)}-\mathfrak{m}_k| = \mathcal{O}\left(\frac{1}{n}\right)$. □

B.4 Side Theorems

Convergence to the empirical spectral measure As for the case with the correlated input, the explicit Stieltjes transform in Corollary A.3.1 is too computationally intensive to be used to compute the transform. However, it allows to check that the Carleman's condition [18, Theorem 4.3] for the spectral measure defining the moments to be unique [18]. This is done, by ensuring that $\mathfrak{m}_k < C^k$.

Lemma B.3. *For $\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\,\boldsymbol{Y}^\top\right]$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\boldsymbol{X}\right)$ as defined in Theorem A.1, there exists a variable C such that*

$$\mathfrak{m}_k \le C^k. \tag{301}$$

Proof. Following a similar logic as in [3], there is a constant $\bar{C}$ such that $\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu) < \bar{C}\frac{3\tau}{4\sqrt{\pi}}\frac{\nu^{k-3/2}}{k^{5/2}}$ and consequently

$$\mathfrak{m}_k = \varphi^{1-k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\frac{\mathcal{C}(k,I_i,I_\mu,b,b^\mu,c^\mu)}{2\pi\sqrt{(k-I_i)(k-I_\mu)}}(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)^b(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^{k-b}$$
$$\kappa_b^{b^\mu}\kappa_c^{I_\mu-b^\mu}\varphi^{I_1}\psi^{I_\mu}\left(\frac{e}{k-I_i}\right)^{k-I_1}\left(\frac{e}{k-I_\mu}\right)^{k-I_\mu} \tag{302}$$
$$\le \varphi^{1-k}\sum_{I_i,I_\mu=1}^{k}\sum_{b=0}^{I_i+I_\mu+1}\sum_{b^\mu=0}^{b}\frac{3\bar{C}\tau}{2\pi 4\sqrt{\pi}}\frac{\nu^{k-3/2}}{k^{5/2}}(\theta_1-\sigma_x^2\sigma_w^2\theta_3^2)^b(\theta_2-\sigma_x^2\sigma_w^2\theta_3^2)^{k-b}$$

$$\kappa_b^{b^\mu} \kappa_c^{I_\mu - b^\mu} \varphi^{I_1} \psi^{I_\mu} (e)^{k-I_1} (e)^{k-I_\mu} \tag{303}$$

$$\leq \varphi^{1-k} \sum_{I_i, I_\mu=1}^{k} \sum_{b=0}^{I_i+I_\mu+1} \sum_{b^\mu=0}^{b} \frac{3\bar{C}\tau}{2\pi 4\sqrt{\pi}} \frac{\nu^{k-3/2}}{k^{5/2}} \max(1, \theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^k \max(1, \theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^k$$

$$\max(1,\kappa_b)^k \max(1,\kappa_c)^{I_\mu} \left(\frac{\varphi}{e}\right)^{I_1} \left(\frac{\psi}{e}\right)^{I_\mu} e^{2k} \tag{304}$$

$$\leq \varphi^{1-k} \frac{3\bar{C}\tau}{2\pi 4\sqrt{\pi}} \frac{\nu^{k-3/2}}{k^{5/2}} \max(1, \theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^k \max(1, \theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^k \max(1,\kappa_b)^k e^{2k}$$

$$\sum_{I_i, I_\mu=1}^{k} \max(1,\kappa_c)^{I_\mu} \left(\frac{\varphi}{e}\right)^{I_1} \left(\frac{\psi}{e}\right)^{I_\mu} (I_i + I_\mu + 1)(I_i + I_\mu + 2)/2 \tag{305}$$

$$\leq \varphi^{1-k} \frac{3\bar{C}\tau}{2\pi 4\sqrt{\pi}} \frac{\nu^{k-3/2}}{k^{5/2}} \max(1, \theta_1 - \sigma_x^2\sigma_w^2\theta_3^2)^k \max(1, \theta_2 - \sigma_x^2\sigma_w^2\theta_3^2)^k \max(1,\kappa_b)^k$$

$$\max(1,\kappa_c)^k \left(\frac{\varphi}{e}\right)^k \left(\frac{\psi}{e}\right)^k (2k+1)(2k+2)/2 \tag{306}$$

$$\leq C^k \tag{307}$$

and therefore the Carleman's condition is satisfied and there is a unique distribution μ defined by the considered Stieltjes transform. □

Contribution of non-admissible graphs The computation of the moments relies on the set of admissible graphs, as defined in Definition A.1. The contributions of the addends in (46) with non-admissible graphs is not significant to the moments of the eigen-distribution as in the corrleated input case. This is because the argument in [13, Supplementary Material 1.2.1] and [3, Section 3.1.4] showing that the leading order contribution for non-admissible graphs is $\mathcal{O}(n_0^{c-1-k})$ rather than $\mathcal{O}(n_0^{c-k})$ as for the admissible graphs, relies on the fact that non-admissible graphs require a further identification and this does not change in this case. This statement relies on the assumption that the activation function is such that $\left|\int \phi^k(\sigma_w\sigma_x z)\mathcal{D}z\right| < \infty$ since otherwise the contribution of a graph that consists in going through the same 2-dimensional cycle k-times could blow up. Therefore, non-admissible graphs generate a negligible contribution for $n \to \infty$.

B.5 Proof Corollary A.1.1

Proof. Becuase of $\theta_2 = \sigma_x^2\sigma_w^2\theta_3^2$ at the first layer and of considering $\psi = 0$ and $\varphi = 1$, then we find that the recursive relation for H is

$$H(z) = 1 + \frac{H_{\psi b}(z) H_\varphi(z)(\theta_1 - \sigma_x^2\sigma_{\mathbf{W}}^2\theta_3^2)}{\varphi z} \tag{308}$$

$$= 1 + \frac{H(z)(\theta_1 - \theta_2)}{z} \tag{309}$$

and therefore

$$\frac{H}{z} = \frac{1}{z - (\theta_1 - \theta_2)}. \tag{310}$$

By expliciting the related Stieltjes transform we find that

$$G(z) = \frac{H}{z} = \frac{1}{z - (\theta_1 - \theta_2)}. \tag{311}$$

which implies that the covariance matrix $\frac{1}{m}\mathbb{E}_{\mathbf{X}}\left[\mathbf{Y}\mathbf{Y}^\top\right] - \frac{1}{\sigma_x^2 m^2}\mathbb{E}_{\mathbf{x}}\left[\mathbf{Y}\mathbf{X}^\top\right]\mathbb{E}_{\mathbf{X}}\left[\mathbf{X}\,\mathbf{Y}^\top\right]$ is the identity matrix scaled by $(\theta_1 - \theta_2)$. Therefore, the distribution of the post-activation layer is going to be defined by **WW**. □

C Proofs Relevant to Theorem 3.2

In this section, it is going to be necessary computing expectations over the variable

$$\mathfrak{W}_{\alpha\beta}^{(\ell)}=\sum_{k_\ell}\mathbf{W}_{\alpha k_\ell}^{(\ell)}\left(\prod_{j=1}^{\ell-2}\left(\sum_{k_{\ell-j}}\mathbf{W}_{k_{\ell-j+1}k_{\ell-j}}^{(\ell-j)})\right)\right)\mathbf{W}_{k_2\beta}^{(1)}.$$

Specifically the following Lemma is going to be used.

Lemma C.1. *Consider a succession of independent matrices $\{\boldsymbol{W}^{(l)}\}_{l=1}^{\ell}$ such that $\boldsymbol{W}^{(l)}\in\mathbb{R}^{n_l\times n_{l-1}}$ and define the following variable*

$$\mathfrak{W}_{\alpha\beta}^{(\ell)}=\sum_{k_\ell}\boldsymbol{W}_{\alpha k_\ell}^{(\ell)}\left(\prod_{j=1}^{\ell-2}\left(\sum_{k_{\ell-j}}\boldsymbol{W}_{k_{\ell-j+1}k_{\ell-j}}^{(\ell-j)})\right)\right)\boldsymbol{W}_{k_2\beta}^{(1)}.$$

Then

$$\int\mathfrak{W}_{\alpha\beta}^{(\ell)}\mathfrak{W}_{\gamma\beta}^{(\ell)}\mathcal{D}\mathfrak{W}^{(\ell)}=\mathbb{1}_{\{\gamma\}}(\alpha)\frac{\sigma_w^{2\ell}}{n} \tag{312}$$

where $\mathbb{1}_{\{\gamma\}}$ is the indicator function, i.e. $\mathbb{1}_{\{\gamma\}}(\alpha)=1$ if $\alpha=\gamma$ and $\mathbb{1}_{\{\gamma\}}(\alpha)=0$ if $\alpha\neq\gamma$.

Proof. Consider $\alpha\neq\gamma$, then

$$\int\mathfrak{W}_{\alpha\beta}^{(\ell)}\mathfrak{W}_{\gamma\beta}^{(\ell)}\mathcal{D}\mathfrak{W}^{(\ell)}=\int\sum_{k_1,k_2=1}^{n}\mathbf{W}_{\alpha k_1}^{(\ell)}\mathfrak{W}_{k_1\beta}^{(\ell-1)}\mathbf{W}_{\gamma k_2}^{(\ell)}\mathfrak{W}_{k_2\beta}^{(\ell-1)}\mathcal{D}\mathbf{W}^{(\ell)}\mathcal{D}\mathfrak{W}^{(\ell-1)} \tag{313}$$

$$=\sum_{k_1,k_2=1}^{n}\underbrace{\int\mathbf{W}_{\alpha k_1}^{(\ell)}\mathbf{W}_{\gamma k_2}^{(\ell)}\mathcal{D}\mathbf{W}^{(\ell)}}_{=0}\int\mathfrak{W}_{k_1\beta}^{(\ell-1)}\mathfrak{W}_{k_2\beta}^{(\ell-1)}\mathcal{D}\mathfrak{W}^{(\ell-1)}=0. \tag{314}$$

While, if $\alpha=\gamma$

$$\int{\mathfrak{W}_{\alpha\beta}^{(\ell)}}^2\mathcal{D}\mathfrak{W}^{(\ell)}=\int\sum_{k_1,k_2,\ldots,k_{\ell-1}=1}^{n}\left(\mathbf{W}_{\alpha k_1}^{(\ell)}\mathbf{W}_{k_1k_2}^{(\ell-1)}\ldots\mathbf{W}_{k_{\ell-1}\beta}^{(1)}\right)^2\mathcal{D}\mathbf{W}^{(\ell)}\mathcal{D}\mathbf{W}^{(\ell-1)}\ldots\mathcal{D}\mathbf{W}^{(1)} \tag{315}$$

$$=\sum_{k_1,k_2,\ldots,k_{\ell-1}=1}^{n}\left(\int{\mathbf{W}_{\alpha k_1}^{(\ell)}}^2\mathcal{D}\mathbf{W}^{(\ell)}\int{\mathbf{W}_{k_1k_2}^{(\ell-1)}}^2\mathcal{D}\mathbf{W}^{(\ell-1)}\ldots\int{\mathbf{W}_{k_{\ell-1}\beta}^{(1)}}^2\mathcal{D}\mathbf{W}^{(1)}\right) \tag{316}$$

$$=\sum_{k_1,k_2,\ldots,k_{\ell-1}=1}^{n}\left(\frac{\sigma_w^2}{n}\right)^{\ell}=n^{\ell-1}\left(\frac{\sigma_w^2}{n}\right)^{\ell}=\frac{\sigma_w^{2\ell}}{n} \tag{317}$$

where we used in (315) the i.i.d. nature of the terms in $\mathfrak{W}_{\alpha\beta}$ for whom the following statement holds: if two variables X_1 and X_2 are independent then $\mathbb{E}[(X_1+X_2)^2]=\mathbb{E}[X_1^2]+\mathbb{E}[X_2^2]$ since $Cov[X_1X_2]=0$. □

C.1 Lemma A.4

To prove Lemma A.4 it is first necessary to study the expected contribution of 2 cycles to the moment of the covariance matrix.

C.1.1 Supporting Lemmas

Lemma C.2. *For $\boldsymbol{M}=\frac{1}{m}\boldsymbol{Y}\boldsymbol{Y}^\top$ and $\boldsymbol{Y}=\phi\left(\boldsymbol{W}\tilde{\boldsymbol{X}}\right)$ as defined in Theorem 3.2 without the hypothesis of independence for the elements of $\boldsymbol{Y}^{(\ell)}$, i.e. $\boldsymbol{Y}_{:p}^{(\ell)}\sim\mathcal{N}(\boldsymbol{0},q^{(\ell)}\boldsymbol{\Sigma}^{(\ell)})$,, then*

$$\mathbb{E}\left[\boldsymbol{Y}_{i_1\mu_2}\boldsymbol{Y}_{i_2\mu_1}...\boldsymbol{Y}_{\mu_k i_1}\right] \tag{318}$$

$$=\int\left(\sum_{\nu=0}^{\infty}\frac{\left(-\frac{1}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\boldsymbol{\Sigma}^{\xi}tr\left(\left(\bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\Lambda}}^\top\right)^{\xi}\right)\right)^{\nu}}{\nu!}\right).$$

$$\cdot\left(\prod_{\lambda_{\alpha\beta}\in\mathcal{Z}} e^{-\frac{tr\Sigma}{2}\bar{\lambda}^2_{\alpha\beta}-i\,\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}_{\alpha\beta}z_{\alpha\beta}}\right)F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\bar{\lambda} \tag{319}$$

where $\mathcal{Z}$ is the set of combinations $\{(i_1,\mu_1),(i_2,\mu_1),\ldots,(i_1,\mu_k)\}$ on which the function $F(z)=\prod_{(\alpha\beta)\in\mathcal{Z}}\phi(z_{\alpha\beta})$, the variables

$$\mathbf{Z}_{i\mu}=\begin{cases}z_{i\mu} & \text{if }(i,\mu)\in\mathcal{Z}\\ 0 & \text{otherwise.}\end{cases}\qquad \text{and}\qquad \bar{\mathbf{\Lambda}}_{i\mu}=\begin{cases}\frac{\sigma_x\sigma_w}{\sqrt{n}}\bar{\lambda}_{i\mu} & \text{if }(i,\mu)\in\mathcal{Z}\\ 0 & \text{otherwise.}\end{cases} \tag{320}$$

and the measures

$$\mathcal{D}z=\prod_{z_{\alpha\beta}\in\mathcal{Z}}dz_{\alpha\beta}\qquad \text{and}\qquad \mathcal{D}\bar{\lambda}=\prod_{\lambda_{\alpha\beta}\in\mathcal{Z}}\frac{d\lambda_{\alpha\beta}}{2\pi\sigma_x\sigma_w/\sqrt{n}} \tag{321}$$

are defined.

Proof. Now we consider auxiliary integrals over z, by adding delta functions enforcing $\mathbf{Z}=\mathbf{W}\mathbf{\Sigma}^{1/2}\,\mathbf{X}$ with

$$\mathbf{Z}_{i\mu}=\begin{cases}z_{i\mu} & \text{if }(i,\mu)\in\mathcal{Z}\\ 0 & \text{otherwise.}\end{cases} \tag{322}$$

where $\mathcal{Z}$ denotes the set of unique pairs (i,μ) in equation (140):

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{323}$$

$$=\int\prod_{(\alpha,\beta)\in\mathcal{Z}}\delta(z_{\alpha\beta}-\sum_k\mathbf{W}_{\alpha k}\mathbf{\Sigma}^{1/2}_{kk}\mathbf{X}_{k\beta})\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X} \tag{324}$$

where

$$\mathcal{D}z=\prod_{(\alpha,\beta)\in\mathcal{Z}}dz_{\alpha\beta}. \tag{325}$$

Now we consider the Fourier expression of the Dirac δ

$$\delta(x)=\frac{1}{2\pi}\int e^{i\lambda x}d\lambda \tag{326}$$

and therefore introduced the matrix $\mathbf{\Lambda}\in\mathbb{R}^{n\times m}$ whose entries are

$$\mathbf{\Lambda}_{i\mu}=\begin{cases}\lambda_{i\mu} & \text{if }(i,\mu)\in\mathcal{Z}\\ 0 & \text{otherwise.}\end{cases} \tag{327}$$

with

$$\mathcal{D}\lambda=\prod_{(\alpha,\beta)\in\mathcal{Z}}\frac{d\lambda_{\alpha\beta}}{2\pi}. \tag{328}$$

to obtain

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{329}$$

$$=\int\prod_{z_{\alpha\beta}\in\mathcal{Z}}\delta(z_{\alpha\beta}-\sum_k\mathbf{W}_{\alpha k}\mathbf{\Sigma}^{1/2}_{kk}\mathbf{X}_{k\beta})\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X} \tag{330}$$

$$=\int\prod_{z_{\alpha\beta}\in\mathcal{Z}}exp\left(-i\lambda_{\alpha\beta}\left(\sum_k\mathbf{W}_{\alpha k}\mathbf{\Sigma}^{1/2}_{kk}\mathbf{X}_{k\beta}-z_{\alpha\beta}\right)\right)\cdot$$

$$\cdot\,\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{331}$$

$$=\int exp\left(-i\sum_{z_{\alpha\beta}\in\mathcal{Z}}\lambda_{\alpha\beta}\underbrace{\left(\sum_k\mathbf{W}_{\alpha k}\mathbf{\Sigma}^{1/2}_{kk}\mathbf{X}_{k\beta}-z_{\alpha\beta}\right)}_{(\mathbf{W}\mathbf{\Sigma}^{1/2}\mathbf{X}-\mathbf{Z})_{\alpha\beta}}\right)\cdot$$

$$\cdot\,\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{332}$$

$$= \int exp \left(-i \sum_{\alpha,\beta=1}^{n,m} \mathbf{\Lambda}_{\alpha\beta}(\mathbf{W}\mathbf{\Sigma}^{1/2}\mathbf{X} - \mathbf{Z})_{\alpha\beta} \right) \cdot$$

$$\cdot \phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{333}$$

$$= \int e^{-i\, tr\left(\Lambda^\top(\mathbf{W}\mathbf{\Sigma}^{1/2}\mathbf{X}-\mathbf{Z})\right)}\phi(z_{i_1\mu_1})\phi(z_{i_2\mu_1})...\phi(z_{i_1\mu_k})\mathcal{D}z\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{\Sigma}\mathcal{D}\mathbf{X}\mathcal{D}\lambda \tag{334}$$

where $tr()$ corresponds to the trace function.

Now we first integrate over **X** the factors of (334) that depend on it

$$\int e^{-itr(\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2}\mathbf{X})}\mathcal{D}\mathbf{X} = \prod_{b,c=1}^{m,n}\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\mathbf{X}_{cb}^2 - i\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2}\mathbf{X}_{cb}\right]$$

$$= \prod_{b,c=1}^{m,n}\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\left(\mathbf{X}_{cb} + i\sigma_x^2(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})\right)^2 - \frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})^2\right] \tag{335}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})^2\right]$$

$$\underbrace{\int \frac{d\mathbf{X}_{cb}}{\sqrt{2\pi\sigma_x^2}} exp\left[-\frac{1}{2\sigma_x^2}\left(\mathbf{X}_{cb} + i\sigma_x^2(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})\right)^2\right]}_{=1} \tag{336}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{\sigma_x^2}{2}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})^2\right] = exp\left[-\frac{\sigma_x^2}{2}\sum_{b,c=1}^{m,n}(\sum_{a=1}^{n}\lambda_{ab}\mathbf{W}_{ac}\mathbf{\Sigma}_{cc}^{1/2})^2\right] \tag{337}$$

$$= exp\left[-\frac{\sigma_x^2}{2}\|\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2}\|_F^2\right] = e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2})} \tag{338}$$

where in equation (336) we used the property the complex integral of $z = x + iy$ over the closed cycle $(-\infty, \infty, i\mu + \infty, i\mu - \infty)$ of the analytical function $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-iy)^2/(2\sigma^2)}$ is null and therefore

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-i\mu)^2/(2\sigma^2)}dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x)^2/(2\sigma^2)}dx = 1.$$

Now we integrate over **W**

$$\int e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2})}\mathcal{D}\mathbf{W}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}e^{-\frac{n\mathbf{W}_{ij}^2}{2\sigma_w^2}}\right)e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2})} \tag{339}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{n\sum_{i,j=1}^{n}\mathbf{W}_{ji}\mathbf{W}_{ij}}{2\sigma_w^2}}e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2})} \tag{340}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{ntr\mathbf{W}^\top\mathbf{W}}{2\sigma_w^2}}e^{-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}\mathbf{\Sigma}^{1/2})} \tag{341}$$

$$= \int\left(\prod_{i,j=1}^{n}\frac{d\mathbf{W}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right)e^{-\frac{ntr\mathbf{W}^\top\mathbf{W}}{2\sigma_w^2}-\frac{\sigma_x^2}{2}tr(\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2\top}\mathbf{W}^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W})} \tag{342}$$

$$= \int\left(\prod_{j=1}^{n}\frac{d^n\mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}}\right)e^{-\sum_{j=1}^{n}\left(\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top\mathbf{w}_j+\frac{\sigma_x^2}{2}\mathbf{\Sigma}_{jj}\mathbf{w}_j^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{w}_j\right)} \tag{343}$$

$$= \int \left(\prod_{j=1}^{n} \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} e^{-\left(\frac{n}{2\sigma_w^2}\mathbf{w}_j^\top \mathbf{w}_j + \frac{\sigma_x^2}{2}\mathbf{\Sigma}_{jj}\mathbf{w}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{w}_j\right)} \right) \tag{344}$$

$$= \prod_{j=1}^{n} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j} \tag{345}$$

$$= \prod_{j=1}^{n} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j} \tag{346}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j}}{det^{-1}(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \tag{347}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j}}{det\left((\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1}\right)^{1/2}} \tag{348}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \int \frac{d^n \mathbf{w}_j}{(2\pi)^{n/2}} \frac{e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j}}{det\left(\sigma_w^2/n(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1}\right)^{1/2}} \tag{349}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \underbrace{\int \frac{d^n \mathbf{w}_j}{(2\pi)^{n/2}} \frac{e^{-\frac{1}{2}\mathbf{w}_j^\top \left(\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2 \mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)\mathbf{w}_j}}{det\left((\frac{n}{\sigma_w^2}\mathbf{I} + \sigma_x^2\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{-1}\right)^{1/2}}}_{=1} \tag{350}$$

$$= \prod_{j=1}^{n} \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \tag{351}$$

where we consider $\mathbf{w}_j$ as the jth column of $\mathbf{W}$ and we used the property that for a general non-singular matrix $det(\mathbf{A}^{-1}) = det^{-1}(\mathbf{A})$.

This implies that by considering $F(z) = \prod_{(\alpha,\beta)\in\mathcal{Z}} \phi(z_{\alpha\beta})$

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{352}$$

$$= \int \left(\prod_j \frac{1}{det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \right) e^{-i\, tr(\mathbf{\Lambda}^\top \mathbf{Z})} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{353}$$

$$= \int exp\left(-\sum_j \frac{1}{2}\log det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top) - itr\mathbf{\Lambda}^\top\mathbf{Z} \right) F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{354}$$

Now we will consider the integration over the $\lambda_{\alpha\beta}$ variables. Since $\mathbf{\Sigma}_{jj} > 0$ and the eigenvalues of $\mathbf{\Lambda}\mathbf{\Lambda}^\top$ are non-negative, as a matter of fact for any pair $(\lambda, \mathbf{v})$ the following holds $\lambda = \frac{\mathbf{v}^T\mathbf{\Lambda}^T\mathbf{\Lambda}\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \frac{\|\mathbf{\Lambda}\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \geq 0$, the maximizer of the argument in the exponential is $\mathbf{\Lambda} = 0$, and since the argument is going to be summed over the $n \to \infty$ eigenvalues $\mathbf{\Sigma}_{jj}$, by the saddle point approximation we can consider only an expansion around $\mathbf{\Lambda} = 0$. We can then use the same analysis done in [13] and decompose the log determinant via $logdet|\mathbf{I} + \mathbf{X}| = \sum_{\xi=1} \frac{(-1)^{\xi+1}}{\xi} tr(\mathbf{X}^\xi)$.

Then it follows that

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{355}$$

$$= \int exp\left(-\sum_{j=1}^{n} \frac{1}{2}\log det(\mathbf{I} + \frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top) - itr\mathbf{\Lambda}^\top\mathbf{Z} \right) F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{356}$$

$$= \int e^{-\sum_j \frac{\sigma_x^2\sigma_w^2}{2n}\mathbf{\Sigma}_{jj} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - \frac{1}{2}\sum_{j=1}^n \sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Sigma}_{jj}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top\mathbf{Z}} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{357}$$

$$= \int e^{-\frac{\sigma_x^2\sigma_w^2}{2n} tr\mathbf{\Sigma} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)} e^{-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr(\mathbf{\Sigma}^\xi) tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top\mathbf{Z}} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{358}$$

We now consider the following change of variable

$$\bar{\lambda}_{ij} = \frac{\sigma_w\sigma_x}{\sqrt{n}}\lambda_{ij} \tag{359}$$

$$\mathcal{D}\bar{\mathbf{\Lambda}} = \prod_{(\alpha,\beta)\in\mathcal{Z}} \frac{d\bar{\lambda}_{\alpha\beta}}{2\pi\sigma_x\sigma_w/\sqrt{n}} \tag{360}$$

therefore

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{361}$$

$$= \int e^{-\frac{\sigma_x^2\sigma_w^2}{2n} tr\mathbf{\Sigma} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top)} e^{-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left(\left(\frac{\sigma_x^2\sigma_w^2}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi\right) - itr\mathbf{\Lambda}^\top Z} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\lambda \tag{362}$$

$$= \int e^{-\frac{tr\mathbf{\Sigma}}{2} tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)} e^{-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right) - itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}^\top\mathbf{Z}} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\bar{\mathbf{\Lambda}}. \tag{363}$$

We will study the contribution of the exponential $e^{-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)}$ by considering its Taylor expansion $e^x = \sum_{\nu=0}^\infty x^\nu/\nu!$

$$\mathbb{E}\left[\mathbf{Y}_{i_1\mu_2}\mathbf{Y}_{i_2\mu_1}...\mathbf{Y}_{\mu_k i_1}\right] \tag{364}$$

$$= \int \left(\sum_{\nu=0}^\infty \frac{\left(-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right) \cdot e^{-\frac{tr\mathbf{\Sigma}}{2} tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top) - itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}^\top\mathbf{Z}} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\bar{\mathbf{\Lambda}} \tag{365}$$

$$= \int \left(\sum_{\nu=0}^\infty \frac{\left(-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right) \cdot e^{-\sum_{\lambda_{\alpha\beta}\in\mathcal{Z}}\left(\frac{tr\mathbf{\Sigma}}{2}\bar{\lambda}_{\alpha\beta}^2 - i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}_{\alpha\beta}z_{\alpha\beta}\right)} F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\bar{\mathbf{\Lambda}} \tag{366}$$

$$= \int \left(\sum_{\nu=0}^\infty \frac{\left(-\frac{1}{2}\sum_{\xi\geq 2} \frac{(-1)^{\xi+1}}{\xi} tr\mathbf{\Sigma}^\xi tr\left((\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^\top)^\xi\right)\right)^\nu}{\nu!}\right) \cdot \left(\prod_{\lambda_{\alpha\beta}\in\mathcal{Z}} e^{-\frac{tr\mathbf{\Sigma}}{2}\bar{\lambda}_{\alpha\beta}^2 - i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}_{\alpha\beta}z_{\alpha\beta}}\right) F(z)\mathcal{D}z\mathcal{D}\mathbf{\Sigma}\mathcal{D}\bar{\mathbf{\Lambda}} \tag{367}$$

□

Lemma C.3. *For $\boldsymbol{M} = \frac{1}{m}\boldsymbol{Y}\boldsymbol{Y}^\top$ and $\boldsymbol{Y} = \phi\left(\boldsymbol{W}\tilde{\boldsymbol{X}}\right)$ as defined in Theorem 3.2 without the hypothesis of independence for the elements of $\boldsymbol{Y}^{(\ell)}$, i.e. $\boldsymbol{Y}^{(\ell)}_{:p} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{\Sigma}^{(\ell)})$,, then for $k=1$*

$$\mathbb{E}\left[\boldsymbol{Y}_{i_1\mu_2}\boldsymbol{Y}_{i_2\mu_1}...\boldsymbol{Y}_{\mu_k i_1}\right] = \int \tilde{\theta}_1^{(n)}\mathcal{D}\mathbf{\Sigma}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{368}$$

with

$$\tilde{\theta}_1^{(n)} = \int \frac{1}{\sqrt{2\pi}}\phi(\sqrt{\tilde{\mu}_1^{(n)}}\sigma_w\sigma_x\tilde{z})^2 e^{-\frac{\tilde{z}^2}{2}}d\tilde{z} \tag{369}$$

where $\tilde{\mu}_k^{(n)} = tr(\mathbf{\Sigma}^k)/n$.

Proof. When $k=1$, following [11] the zeroth order expansion of the Taylor series

$$\left(\sum_{\nu=0}^{\infty}\frac{\left(-\frac{1}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\mathbf{\Sigma}^{\xi}tr\left(\left(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^{\top}\right)^{\xi}\right)\right)^{\nu}}{\nu!}\right)$$

in the integral of Lemma C.2 is the leading contribution to $E_2^{(n)}$. Defining the trace of the power matrices as $\tilde{\mu}_k^{(n)}$, i.e. $\tilde{\mu}_k^{(n)} = tr(\mathbf{\Sigma}^k)/n$, the zeroth order is determined as follows.

$$\int\left(\frac{\sqrt{n}}{2\pi\sigma_x\sigma_w}e^{-\frac{n\tilde{\mu}_1^{(n)}}{2}tr(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^{\top})}e^{-itr\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\mathbf{\Lambda}}\mathbf{Z}}F(z)\right)dz_{i_1\mu_1}d\bar{\lambda}_{i_1\mu_1}\mathcal{D}\mathbf{\Sigma}$$

$$=\int\left(\int\frac{\sqrt{n}}{2\pi\sigma_x\sigma_w}e^{-\frac{n\tilde{\mu}_1^{(n)}}{2}\lambda^2-i\frac{\sqrt{n}}{\sigma_w\sigma_x}\bar{\lambda}z}\phi(z)^2dzd\bar{\lambda}\right)\mathcal{D}\mathbf{\Sigma} \tag{370}$$

$$=\int\left(\int\frac{\sqrt{n}}{2\pi\sigma_x\sigma_w}e^{-\frac{z^2}{2\tilde{\mu}_1^{(n)}\sigma_w^2\sigma_x^2}}\frac{\sqrt{2\pi}}{\sqrt{n\tilde{\mu}_1^{(n)}}}\phi(z)^2\left(\int\frac{\sqrt{n\tilde{\mu}_1^{(n)}}}{\sqrt{2\pi}}e^{-\frac{n\tilde{\mu}_1^{(n)}}{2}\left(\lambda+\frac{i\sqrt{n}}{n\tilde{\mu}_1^{(n)}\sigma_w\sigma_x}z\right)^2}d\bar{\lambda}\right)dz\right)\mathcal{D}\mathbf{\Sigma} \tag{371}$$

$$=\int\frac{\sqrt{n}}{2\pi\sigma_x\sigma_w}\phi(z)^2e^{-\frac{z^2}{2\tilde{\mu}_1^{(n)}\sigma_w^2\sigma_x^2}}\frac{\sqrt{2\pi}}{\sqrt{n\tilde{\mu}_1^{(n)}}}dz\mathcal{D}\mathbf{\Sigma} \tag{372}$$

$$=\int\frac{1}{\sqrt{2\pi\tilde{\mu}_1^{(n)}}\sigma_x\sigma_w}\phi(z)^2e^{-\frac{z^2}{2\tilde{\mu}_1^{(n)}\sigma_w^2\sigma_x^2}}dz\mathcal{D}\mathbf{\Sigma} \tag{373}$$

$$=\int\frac{1}{\sqrt{2\pi\tilde{\mu}_1^{(n)}}\sigma_x\sigma_w}\phi(\sqrt{\tilde{\mu}_1^{(n)}}\sigma_w\sigma_x\tilde{z})^2e^{-\frac{\tilde{z}^2}{2}}\sqrt{\tilde{\mu}_1^{(n)}}\sigma_w\sigma_xd\tilde{z}\mathcal{D}\mathbf{\Sigma} \tag{374}$$

$$=\int\frac{1}{\sqrt{2\pi}}\phi(\sqrt{\tilde{\mu}_1^{(n)}}\sigma_w\sigma_x\tilde{z})^2e^{-\frac{\tilde{z}^2}{2}}d\tilde{z}\mathcal{D}\mathbf{\Sigma}=\int\tilde{\theta}_1{}^{(n)}\mathcal{D}\mathbf{\Sigma} \tag{375}$$

For higher order terms of the exponential Taylor series we first notice that

$$\left(-\frac{1}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}tr\mathbf{\Sigma}^{\xi}tr\left(\left(\bar{\mathbf{\Lambda}}\bar{\mathbf{\Lambda}}^{\top}\right)^{\xi}\right)\right)^{\nu}=\left(-\frac{1}{2}\sum_{\xi\geq 2}\frac{(-1)^{\xi+1}}{\xi}(tr\mathbf{\Sigma}^{\xi})\bar{\lambda}^{2\xi}\right)^{\nu}.$$

For each of the expansion terms, the same steps in (370)-(373) lead to the following $2\xi\nu$-th non-central moment of a Gaussian to be considered

$$\int\frac{\sqrt{n\tilde{\mu}_1^{(n)}}}{\sqrt{2\pi}}\bar{\lambda}^{2\xi\nu}e^{-\frac{n\tilde{\mu}_1^{(n)}}{2}(\bar{\lambda}+\frac{i\sqrt{n}z}{n\sigma_x\sigma_w\tilde{\mu}_1^{(n)}})}d\bar{\lambda}=\left(\frac{1}{n\tilde{\mu}_1^{(n)}}\right)^{\xi\nu/2}2^{\xi\nu}\frac{\Gamma(\frac{2\xi\nu+1}{2})}{\sqrt{\pi}}\Phi(-\frac{2\xi\nu}{2};\frac{1}{2};-\frac{z^2}{\sigma_w^2\sigma_x^2\tilde{\mu}_1^{(n)}}) \tag{376}$$

where the solution of the non-central moment is given in [22] with

$$\Phi(-\frac{2\xi\nu}{2};\frac{1}{2};-\frac{z^2}{\sigma_w^2\sigma_x^2\tilde{\mu}_1^{(n)}})=\sum_{i=1}^{\infty}\frac{1}{i!}\frac{(-\frac{2\xi\nu}{2})(-\frac{2\xi\nu}{2}+1)...(-\frac{2\xi\nu\nu}{2}+i-1)}{(-\frac{1}{2})(-\frac{1}{2}+1)...(-\frac{1}{2}+i-1)}\left(\frac{-z^2}{\sigma_w^2\sigma_x^2\tilde{\mu}_1^{(n)}}\right)^i. \tag{377}$$

This leads to the computation of the following integral

$$\int\sum_i\alpha_i\left(\frac{-z^2}{\sigma_w^2\sigma_x^2\tilde{\mu}_1^{(n)}}\right)^i\phi^2(\sigma_x^2\sigma_w^2\tilde{\mu}_1^{(n)}z)\mathcal{D}z \tag{378}$$

which is finite since $\Phi(-\frac{2\xi\nu}{2}; \frac{1}{2}; -\frac{z^2}{\sigma_w^2\sigma_x^2})$ is an entire function of $\xi\nu$ and z, and when solving it by parts we compute the following integrals that are finite by hypothesis

$$\int \phi^{(k)}(\sigma_x^2\sigma_w^2\tilde{\mu}_1^{(n)} z)\mathcal{D}z \tag{379}$$

for any derivative k. However, because of the non-central moment, we also gain a factor $\left(\frac{1}{n\tilde{\mu}_1^{(n)}}\right)^{\xi\nu/2}$ and therefore these terms have a contribution that is $\mathcal{O}(\frac{1}{n})$ relative to the zeroth-order one.

Therefore the contribution of a 2-dimensional cycle is

$$E_2^{(n)} = \int \tilde{\theta}_1^{(n)}\mathcal{D}\mathbf{\Sigma}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right). \tag{380}$$

□

C.1.2 Proof of Lemma A.4

Proof. In the case where $k = 1$ we then find

$$E_2^{(n,\ell)} = \int_{\mathbf{W},\mathbf{Y}} \phi(\sum_l \mathbf{W}_{i_\xi,l}^{(\ell)}\mathbf{Y}_{l,\mu_\xi}^{(\ell)})\phi(\sum_l \mathbf{W}_{i_\xi,l}^{(\ell)}\mathbf{Y}_{l,\mu_\xi}^{(\ell)}) - \sigma_x^2\sum_{p=1}^n \mathfrak{W}_{i_\xi p}^{(\ell)}\mathfrak{W}_{i_\xi p}^{(\ell)}\left(\prod_{j=1}^{\ell}\phi'(\mathbf{H}_{k_\ell p}^{(\ell)})\right)\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X} \tag{381}$$

$$= \int\tilde{\theta}_1^{(n,\ell)}\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) - \sigma_x^2\prod_{l=1}^{\ell}(\theta_3^{(\ell)})^2\int_{\mathbf{W}}\sum_{p=1}^n \mathfrak{W}_{i_\xi p}^{(\ell)}\mathfrak{W}_{i_\xi p}^{(\ell)}\mathcal{D}\mathbf{W} \tag{382}$$

$$= \int\tilde{\theta}_1^{(n,\ell)}\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) - \sigma_x^2\prod_{l=1}^{\ell}(\theta_3^{(\ell)})^2\sum_{p=1}^n\frac{\sigma_w^{2\ell}}{n} \tag{383}$$

$$= \int\tilde{\theta}_1^{(n,\ell)}\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) - \sigma_x^2\sigma_w^{2\ell}\prod_{l=1}^{\ell}(\theta_3^{(\ell)})^2 \tag{384}$$

where the first term is due to Lemma C.3 with

$$\tilde{\theta}_1^{(n,\ell)} = \frac{1}{\sqrt{2\pi}}\phi(\sqrt{q^{(\ell)}}\sqrt{\frac{tr(\mathbf{\Sigma}^{(\ell)})}{n}}z)^2 e^{-\frac{z^2}{2}}dz \tag{385}$$

$$\theta_3^{(\ell)} = \int\phi'(\sqrt{q^{(\ell)}}z)e^{-\frac{z^2}{2}}dz. \tag{386}$$

□

C.2 Lemma A.5

When considering $k > 1$ it is necessary to consider all the mixed products. As a matter of fact each term in the expansion of the product in $E_{2k}^{(n,\ell)}$ consists of successions of products of the kind $\phi(\sum_l \mathbf{W}_{i_\xi,l}^{(\ell)}\mathbf{Y}_{l,\mu_\xi}^{(\ell)})\phi(\sum_l \mathbf{W}_{i_{\xi+1},l}^{(\ell)}\mathbf{Y}_{l,\mu_\xi}^{(\ell)})$ alternated with products of the kind $\sigma_x^2\sum_{p=1}^n \mathfrak{W}_{i_\xi p}\mathfrak{W}_{i_{\xi+1}p}\prod_{j=1}^{\ell}\theta_3^{(j)2}$. Each element of the kind $\sigma_x^2\sum_{p=1}^n \mathfrak{W}_{i_\xi p}\mathfrak{W}_{i_{\xi+1}p}\prod_{j=1}^{\ell}\theta_3^{(j)2}$ divides the sequence of products into independent blocks $\Pi_j^{(n,l)}$ as it was for the case in Appendix B. In this section we are also considering the covariance $\mathbf{\Sigma}^{(\ell)}$ and the contribution of each addend $E_\omega^{(n,k,n_w,n_\phi,p)}$ is considered in the following form

$$E_\omega^{(n,k,n_w,n_\phi,p)} = \int\prod_j E_{\Pi_i^{(n,n_\phi^{(j)})}}\mathcal{D}\mathbf{\Sigma}^{(\ell)} \tag{387}$$

and therefore in the following expression of $E_{\Pi_i^{(n,n_\phi^{(j)})}}$, $\mathbf{\Sigma}^{(\ell)}$ should be considered as a random variable.

Lemma C.4. *For the matrix*

$$\boldsymbol{M} = \frac{1}{m}\mathbb{E}_{\boldsymbol{X}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})^\top\right] - \frac{1}{m^2\sigma_x^2}\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right]\mathbb{E}_{\boldsymbol{x}}\left[\phi(\boldsymbol{W}^{(\ell)}\boldsymbol{Y}^{(\ell)})\boldsymbol{X}^\top\right] \tag{388}$$

and $\boldsymbol{Y}^{(\ell)} = \phi\left(\boldsymbol{H}^{(\ell-1)}\right)$ *as defined in Theorem 3.2 without considering the elements in* $\boldsymbol{H}^{(\ell)}$ *independent, i.e.* $\boldsymbol{Y}^{(\ell)}_{:p} \sim \mathcal{N}(\boldsymbol{0}, q^{(\ell)}\boldsymbol{\Sigma}^{(\ell)})$*, when* $k > 1$ *each block* $\Pi_j^{(l)}$ *generates the following contribution*

$$E^{(n,\ell)}_{\Pi_i^{(l)}} = -\frac{\sigma_w^{2\ell}\sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2\tilde{\theta}_2^{(n)l}}{n^{1+l}}\left(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\Sigma}^{\ell}_{kk}{}^{l}\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{389}$$

with

$$\tilde{\theta}_2^{(n,\ell)} = \left(\int\frac{\sqrt{q^{(\ell)}}}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\phi'(\sqrt{q^{(\ell)}}\sqrt{\frac{tr(\boldsymbol{\Sigma}^{\ell})}{n}}z)dz\right)^2, \tag{390}$$

$$\theta_3 = \int_{z_1}\phi'\left(\sqrt{q^{(\ell)}}z_1\right)\mathcal{D}z_1 \tag{391}$$

and $\boldsymbol{\Sigma}^{(\ell)}$ *being the covariance matrix of the hidden layers* $\boldsymbol{Y}^{(\ell)}$.

Proof. Focusing on the integration of one independent block $\Pi_j^{(l)}$, where l identifies the numbers of factors $\phi(\sum_p \mathbf{W}^{(\ell)}_{i_\xi,p}\mathbf{Y}^{(\ell)}_{p,\mu_\xi})\phi(\sum_p \mathbf{W}^{(\ell)}_{i_{\xi+1},p}\mathbf{Y}^{(\ell)}_{p,\mu_\xi})$ between two of the $\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}^{(\ell)}_{i_\xi p}$ kind, i.e.

$$E^{(n,\ell)}_{\Pi_i^{(l)}} = -\int_{\mathbf{W},\mathbf{Y}}\left\{\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}^{(\ell)}_{i_1 p}\prod_{\xi=1}^{l}\left(\phi(\sum_p \mathbf{W}^{(\ell)}_{i_\xi,p}\mathbf{Y}^{(\ell)}_{p,\mu_\xi})\phi(\sum_p \mathbf{W}^{(\ell)}_{i_{\xi+1},p}\mathbf{Y}^{(\ell)}_{p,\mu_\xi})\right)\cdot\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}^{(\ell)}_{i_{l+1}q}\right\}\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{Y}. \tag{392}$$

By defining as $\boldsymbol{\Sigma}^{(\ell)}$ the covariance of the post-activation layer, we are now able to consider the following equality

$$\sum_p \mathbf{W}^{(\ell)}_{i_\xi,p}\mathbf{Y}^{(\ell)}_{p,\mu_\xi} = \sum_p \tilde{\mathbf{W}}^{(\ell)}_{i_\xi,p}\boldsymbol{\Sigma}^{(\ell)}_{pp}{}^{1/2}\tilde{\mathbf{Y}}_{p,\mu_\xi} \tag{393}$$

where $\tilde{\mathbf{W}}^{(\ell)} \sim \mathbf{W}^{(\ell)}$ and $\tilde{\mathbf{Y}}_{ij} \sim \mathcal{N}\left(0, \frac{q^{(\ell)}}{\sigma_w^2}\right)$ and this allows to run analysis similar to Appendix B.

The computation of this expectation follows the structure of the proof in [13]. A dummy variable z is introduced with a delta Dirac function within each ϕ element, and then a Fourier representation for all the arguments of the ϕ functions is introduced. For each factor Π_j the set $\mathcal{Z}_{\Pi_j} \subset \mathcal{Z}$ is considered containing only the combinations (i_ξ, μ_ν) that are included in the $\phi(\sum_l \tilde{\mathbf{W}}_{i_\xi,l}\boldsymbol{\Sigma}^{(\ell)}_{ll}{}^{1/2}\tilde{\mathbf{Y}}_{l,\mu_\nu})$ arguments. Thus auxiliary integrals over z are considered by adding delta functions enforcing $\mathbf{Z} = \mathbf{W}^{(\ell)}\mathbf{Y}^{(\ell)} = \tilde{\mathbf{W}}^{(\ell)}\boldsymbol{\Sigma}^{(\ell)}{}^{1/2}\tilde{\mathbf{Y}}$ with

$$\mathbf{Z}_{i\mu} = \begin{cases} z_{i\mu} & \text{if } (i,\mu) \in \mathcal{Z}_{\Pi_j} \\ 0 & \text{otherwise.}\end{cases} \tag{394}$$

and consequently

$$E^{(n,\ell)}_{\Pi_i^{(l)}} = -\int\left\{\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}_{i_1 p}\prod_{\xi=1}^{l}\left(\phi(\sum_p \tilde{\mathbf{W}}^{(\ell)}_{i_\xi,p}\boldsymbol{\Sigma}^{(\ell)}_{pp}{}^{1/2}\tilde{\mathbf{Y}}_{p,\mu_\xi})\phi(\sum_p \tilde{\mathbf{W}}^{(\ell)}_{i_{\xi+1},p}\boldsymbol{\Sigma}^{(\ell)}_{pp}{}^{1/2}\tilde{\mathbf{Y}}_{p,\mu_\xi})\right)\cdot\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}_{i_{l+1}q}\right\}\mathcal{D}\mathbf{W}\mathcal{D}\tilde{\mathbf{Y}} \tag{395}$$

$$= -\int \prod_{z_{\alpha\beta}\in\mathcal{Z}} \delta(z_{\alpha\beta} - \sum_k \tilde{\mathbf{W}}^{(\ell)}_{\alpha k} {\mathbf{\Sigma}^{(\ell)}_{kk}}^{1/2} \tilde{\mathbf{Y}}_{k\beta}) \cdot$$

$$\cdot \left\{ \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \mathfrak{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi})\phi(z_{i_\xi \mu_\xi})\right) \mathfrak{W}_{i_{l+1} q} \right\} \mathcal{D}\mathbf{W}\mathcal{D}\tilde{\mathbf{Y}}\mathcal{D}z \tag{396}$$

$$= -\int e^{-itr\mathbf{\Lambda}^\top(\tilde{\mathbf{W}}^{(\ell)}\mathbf{\Sigma}^{(\ell)\,1/2}\tilde{\mathbf{Y}} - \mathbf{Z})} \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \cdot$$

$$\cdot \mathfrak{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi})\phi(z_{i_{\xi+1} \mu_\xi})\right) \mathfrak{W}_{i_{l+1} q} \mathcal{D}\mathbf{W}\mathcal{D}\tilde{\mathbf{Y}}\mathcal{D}\lambda\mathcal{D}z \tag{397}$$

where

$$\mathcal{D}z = \prod_{(\alpha,\beta)\in\mathcal{Z}_{\Pi_i}} dz_{\alpha\beta} \tag{398}$$

and in the second equality the following property is used

$$\delta(x) = \frac{1}{2\pi}\int e^{i\lambda x} d\lambda \tag{399}$$

and therefore the matrix $\mathbf{\Lambda} \in \mathbb{R}^{n\times m}$ whose entries are

$$\mathbf{\Lambda}_{i\mu} = \begin{cases} \lambda_{i\mu} & \text{if } (i,\mu)\in\mathcal{Z}_{\Pi_i} \\ 0 & \text{otherwise.} \end{cases} \tag{400}$$

is considered and therefore

$$\mathcal{D}\lambda = \prod_{(\alpha,\beta)\in\mathcal{Z}_{\Pi_i}} \frac{d\lambda_{\alpha\beta}}{2\pi}. \tag{401}$$

To start with, the integration over $\tilde{\mathbf{Y}}$ is considered

$$\int e^{-itr(\mathbf{\Lambda}^\top\tilde{\mathbf{W}}^{(\ell)}\mathbf{\Sigma}^{(\ell)\,1/2}\tilde{\mathbf{Y}})}\mathcal{D}\tilde{\mathbf{Y}} = \prod_{b,c=1}^{m,n}\int \frac{d\tilde{\mathbf{Y}}_{cb}}{\sqrt{2\pi q^{(\ell)}}} exp\left[-\frac{\sigma_w^2}{2q^\ell}\tilde{\mathbf{Y}}_{cb}^2 - i\sum_{a=1}^{n}\lambda_{ab}\tilde{\mathbf{W}}^{(\ell)}_{ac}{\mathbf{\Sigma}^{(\ell)}_{cc}}^{1/2}\tilde{\mathbf{Y}}_{cb}\right]$$

$$= \prod_{b,c=1}^{m,n}\int \frac{d\tilde{\mathbf{Y}}_{cb}}{\sqrt{2\pi q^{(\ell)}}} exp\left[-\frac{\sigma_w^2}{2q^{(\ell)}}\left(\tilde{\mathbf{Y}}_{cb} + i\frac{q^{(\ell)}}{\sigma_w^2}(\sum_{a=1}^{n}\lambda_{ab}\tilde{\mathbf{W}}^{(\ell)}_{ac}{\mathbf{\Sigma}^{(\ell)}_{cc}}^{1/2})\right)^2 - \frac{q^{(\ell)}}{2\sigma_w^2}(\sum_{a=1}^{n}\lambda_{ab}\tilde{\mathbf{W}}^{(\ell)}_{ac})^2\right] \tag{402}$$

$$= \prod_{b,c=1}^{m,n} exp\left[-\frac{q^{(\ell)}}{2\sigma_w^2}(\sum_{a=1}^{n}\lambda_{ab}\tilde{\mathbf{W}}^{(\ell)}_{ac}{\mathbf{\Sigma}^{(\ell)}_{cc}}^{1/2})^2\right] = exp\left[-\frac{q^{(\ell)}}{2\sigma_w^2}\sum_{b,c=1}^{m,n}(\sum_{a=1}^{n}\lambda_{ab}\tilde{\mathbf{W}}^{(\ell)}_{ac}{\mathbf{\Sigma}^{(\ell)}_{cc}}^{1/2})^2\right] \tag{403}$$

$$= exp\left[-\frac{q^{(\ell)}}{2\sigma_w^2}\|\mathbf{\Lambda}^\top\tilde{\mathbf{W}}^{(\ell)}\mathbf{\Sigma}^{(\ell)\,1/2}\|_F^2\right] = e^{-\frac{q^{(\ell)}}{2\sigma_w^2}tr(\mathbf{\Sigma}^{(\ell)\top\,1/2}\tilde{\mathbf{W}}^{(\ell)\top}\mathbf{\Lambda}\mathbf{\Lambda}^\top\tilde{\mathbf{W}}^{(\ell)}\mathbf{\Sigma}^{(\ell)\,1/2})}. \tag{404}$$

and this allows to integrate over $\tilde{\mathbf{W}}^{(\ell)}$. Note that $\tilde{\mathbf{W}}^{(\ell)} = \mathbf{W}^{(\ell)}\mathbf{O}^\top$ and the elements in $\mathfrak{W}^{(\ell)}$ are dependent on $\mathbf{W}^{(\ell)}$, therefore it is also necessary to integrate over the possible orthonormal transform $\mathbf{O}$. The case where $p \neq q$ is now considered

$$\int_{\tilde{\mathbf{W}},\mathbf{O}} \sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 e^{-\frac{q^{(\ell)}}{2\sigma_w^2}tr(\mathbf{\Sigma}^{(\ell)}\tilde{\mathbf{W}}^{(\ell)\top}\mathbf{\Lambda}\mathbf{\Lambda}^\top\mathbf{W}^{(\ell)})}\mathfrak{W}^{(\ell)}_{i_1 p}\mathfrak{W}^{(\ell)}_{i_{l+1} q}\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{O}$$

$$= \sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \int\left(\prod_{i,j=1}^{n}\frac{d\tilde{\mathbf{W}}^{(\ell)}_{ij}}{\sqrt{2\pi\sigma_w^2/n}}\right) e^{-\frac{ntr(\tilde{\mathbf{W}}^{(\ell)\top}\tilde{\mathbf{W}}^{(\ell)})}{2\sigma_w^2}} e^{-\frac{q^{(\ell)}}{2\sigma_w^2}tr(\mathbf{\Sigma}^{(\ell)}\tilde{\mathbf{W}}^{(\ell)\top}\mathbf{\Lambda}\mathbf{\Lambda}^\top\tilde{\mathbf{W}}^{(\ell)})}\mathfrak{W}^{(\ell)}_{i_1 p}\mathfrak{W}^{(\ell)}_{i_{l+1} q} \tag{405}$$

$$= \sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \int\left(\prod_{j=1}^{n}\frac{d^n\tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}}\right) e^{-\sum_{j=1}^{n}(\frac{n}{2\sigma_w^2}\tilde{\mathbf{w}}_j^\top\tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2}\mathbf{\Sigma}^{(\ell)}_{jj}\tilde{\mathbf{w}}_j^\top\mathbf{\Lambda}\mathbf{\Lambda}^\top\tilde{\mathbf{w}}_j)}.$$

$$\cdot \left(\sum_{\alpha,\beta=1}^{n} \mathbf{w}_{\alpha}^{(i_1)} \mathbf{w}_{\beta}^{(i_{l+1})} \mathfrak{W}_{\alpha p}^{(\ell-1)} \mathfrak{W}_{\beta q}^{(\ell-1)} \right) \tag{406}$$

$$= \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \left(\sum_{\alpha,\beta=1}^{n} \int \left(\prod_{j=1}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} (\frac{n}{2\sigma_w^2} \tilde{w}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2} \mathbf{\Sigma}_{jj}^{(\ell)} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)} \cdot \right.$$

$$\left. \cdot \mathbf{w}_{\alpha}^{(i_1)} \mathbf{w}_{\beta}^{(i_{l+1})} \underbrace{\int \mathfrak{W}_{\alpha p}^{(\ell-1)} \mathfrak{W}_{\beta q}^{(\ell-1)}}_{=0} \right) \tag{407}$$

$$= 0. \tag{408}$$

While now the case where $q = p$ will be considered

$$\int_{\tilde{\mathbf{W}},\mathbf{O}} \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 e^{-\frac{\sigma_x^2}{2} tr(\mathbf{\Sigma}^{(\ell)} \tilde{\mathbf{W}}^{(\ell)\top} \mathbf{\Lambda}\mathbf{\Lambda}^\top \mathbf{W}^{(\ell)})} \mathfrak{W}_{i_1 p}^{(\ell)} \mathfrak{W}_{i_{l+1} p}^{(\ell)} \mathcal{D}\mathbf{W}\mathcal{D}\mathbf{O}$$

$$= \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \int \left(\prod_{i,j=1}^{n} \frac{d\tilde{\mathbf{W}}_{ij}^{(\ell)}}{\sqrt{2\pi\sigma_w^2/n}} \right) e^{-\frac{n\,tr(\tilde{\mathbf{W}}^{(\ell)\top}\tilde{\mathbf{W}}^{(\ell)})}{2\sigma_w^2}} e^{-\frac{q^{(\ell)}}{2\sigma_w^2} tr(\mathbf{\Sigma}^{(\ell)} \tilde{\mathbf{W}}^{(\ell)\top} \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{W}}^{(\ell)})}$$

$$\cdot \mathfrak{W}_{i_1 p}^{(\ell)} \mathfrak{W}_{i_{l+1} p}^{(\ell)} \mathcal{D}\mathbf{O} \tag{409}$$

$$= \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \int \left(\prod_{j=1}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} (\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2} \mathbf{\Sigma}_{jj} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)}$$

$$\cdot \left(\sum_{\alpha,\beta=1}^{n} \mathbf{w}_{\alpha}^{(i_1)} \mathfrak{W}_{\alpha p}^{(\ell-1)} \mathbf{w}_{\beta}^{(i_{l+1})} \mathfrak{W}_{\beta p}^{(\ell-1)} \right) \mathcal{D}\mathbf{O} \tag{410}$$

$$= \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \sum_{\alpha=1}^{n} \left(\frac{\sigma_w^{2(\ell-1)}}{n} \int \left(\prod_{j=1}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} (\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2} \mathbf{\Sigma}_{jj} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)} \right.$$

$$\left. \cdot \mathbf{w}_{\alpha}^{(i_1)} \mathbf{w}_{\alpha}^{(i_{l+1})} \right) \mathcal{D}\mathbf{O} \tag{411}$$

$$= \sigma_x^2 \sigma_w^{2(\ell-1)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \int \left(\prod_{j=1}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} (\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2} \mathbf{\Sigma}_{jj} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)}.$$

$$\cdot (\sum_{k=1}^{n} \mathbf{O}_{\alpha k} \tilde{\mathbf{w}}_k^{(i_1)}) (\sum_{k=1}^{n} \mathbf{O}_{\alpha k} \tilde{\mathbf{w}}_k^{(i_{l+1})}) \mathcal{D}\mathbf{O} \tag{412}$$

$$= \sum_{k=1}^{n} \sigma_x^2 \sigma_w^{2(\ell-1)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \cdot$$

$$\cdot \int \left(\prod_{j=1}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{j=1}^{n} (\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)}}{2\sigma_w^2} \mathbf{\Sigma}_{jj} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)} \tilde{\mathbf{w}}_k^{(i_1)} \tilde{\mathbf{w}}_k^{(i_{l+1})}$$

$$\cdot \underbrace{\int_{\mathbf{O}} \mathbf{O}_{\alpha k}^2 \mathcal{D}\mathbf{O}}_{=1/n} \tag{413}$$

$$= \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2(\ell-1)}}{n} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \int \left(\prod_{\substack{j=1 \\ j\neq k}}^{n} \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \right) e^{-\sum_{\substack{j=1\\j\neq k}}^{n} \left(\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j + \frac{q^{(\ell)} \mathbf{\Sigma}_{jj}}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j \right)}$$

$$\cdot \int \frac{d^n \tilde{\mathbf{w}}_k}{(2\pi\sigma_w^2/n)^{n/2}} \tilde{\mathbf{w}}_k^{(i_1)} \tilde{\mathbf{w}}_k^{(i_{l+1})} e^{-\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_k^\top \tilde{\mathbf{w}}_k - \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{2\sigma_w^2} \tilde{\mathbf{w}}_k^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_k} \tag{414}$$

$$= \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2(\ell-1)}}{n} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \cdot$$

$$\cdot \prod_{\substack{j=1\\j\neq k}}^{n} \underbrace{\left(\int \frac{d^n \tilde{\mathbf{w}}_j}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} e^{-\frac{n}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \tilde{\mathbf{w}}_j - \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}}{2\sigma_w^2} \tilde{\mathbf{w}}_j^\top \mathbf{\Lambda}\mathbf{\Lambda}^\top \tilde{\mathbf{w}}_j)} \right)}_{= \frac{1}{det(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}}{n}\mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}}$$

$$\cdot \int \frac{d^n \tilde{\mathbf{w}}_k}{(2\pi\sigma_w^2/n)^{n/2}} \frac{det^{-1}(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}}{det^{-1}(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} \tilde{\mathbf{w}}_k^{(i_0)} \tilde{\mathbf{w}}_k^{(i_{l+1})} e^{-\frac{1}{2}\tilde{\mathbf{w}}_k^\top \left(\frac{n}{\sigma_w^2} I + \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{\sigma_w^2} \mathbf{\Lambda}\mathbf{\Lambda}^\top \right) \tilde{\mathbf{w}}_k} \tag{415}$$

$$= \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2(\ell-1)} \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2}{n \prod_{j=1}^{n} det(\mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} Cov\left[\tilde{\mathbf{w}}_k^{(i_0)} \tilde{\mathbf{w}}_k^{(i_{l+1})} \right]. \tag{416}$$

The covariance matrix of the vector $\tilde{\mathbf{w}}_k$ is

$$\left(\frac{n}{\sigma_w^2} \mathbf{I} + \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{\sigma_w^2} \mathbf{\Lambda}\mathbf{\Lambda}^\top \right)^{-1} = \frac{\sigma_w^2}{n} \sum_{k=0}^{\infty} (-\frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^k \tag{417}$$

therefore

$$Cov\left[\tilde{\mathbf{w}}_k^{(i_0)} \tilde{\mathbf{w}}_k^{(i_{l+1})} \right] = \left[\left(\frac{n}{\sigma_w^2} I + \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{\sigma_w^2} \mathbf{\Lambda}\mathbf{\Lambda}^\top \right)^{-1} \right]_{i_0 i_{l+1}} = \left[\frac{\sigma_w^2}{n} \sum_{\nu=0}^{\infty} (-\frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^\nu \right]_{i_0 i_{l+1}} \tag{418}$$

$$= \frac{\sigma_w^2}{n} \sum_{\nu=l}^{\infty} \left[(-\frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^\nu \right]_{i_0 i_{l+1}} \tag{419}$$

$$= (-1)^l \frac{\sigma_w^2}{n} \prod_{j=1}^{l} \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1}\mu_j} + \sum_{\nu=1}^{\infty} (-1)^{l+\nu} \frac{\sigma_w^2}{n} \sum_{\substack{\vec{l}\ s.t.\\ \|\vec{l}\|_1 = \nu + l \\ \tilde{l}_j \in \mathbb{N}_0}} \prod_{j=1}^{l} \left(\frac{q^{(\ell)}\mathbf{\Sigma}_{kk}}{n} \lambda_{i_j\mu_j} \lambda_{i_{j+1}\mu_j} \right)^{\tilde{l}_j}. \tag{420}$$

Exactly as shown in Lemma A.3, the first addend in (420) determines the contribution $E^{(n,\ell)}_{\Pi_i^{(l)}}$ up to a relative error of $\mathcal{O}(1/n)$. This implies that by considering $F(z) = \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi}) \phi(z_{i_{\xi+1}\mu_\xi}) \right)$

$$E^{(n,\ell)}_{\Pi_i^{(l)}} = -\int e^{-itr\mathbf{\Lambda}^\top (\tilde{\mathbf{W}}^{(\ell)} \mathbf{\Sigma}^{(\ell)1/2} \tilde{\mathbf{Y}} - \mathbf{Z})} \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \cdot$$

$$\cdot \mathfrak{W}_{i_1 p} \prod_{\xi=1}^{l} \left(\phi(z_{i_\xi \mu_\xi}) \phi(z_{i_{\xi+1}\mu_\xi}) \right) \mathfrak{W}_{i_{l+1} q} \mathcal{D}\mathbf{W} \mathcal{D}X \mathcal{D}\lambda \mathcal{D}z \tag{421}$$

$$= -\sum_{k=1}^{n} \int \frac{\sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)} \right)^2 \sigma_w^{2(\ell-1)}}{n} (-1)^l \frac{\sigma_w^2}{n} \left(\prod_{j=1}^{l} \frac{q^{(\ell)}\mathbf{\Sigma}_{kk}^{(\ell)}}{n} \lambda_{i_j\mu_j} \lambda_{i_{j+1}\mu_j} \right)$$

$$\frac{1}{\prod_{j=1}^{n} det(I + \frac{q^{(\ell)}\mathbf{\Sigma}_{jj}^{(\ell)}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)^{1/2}} e^{tr\mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \left(1 + \mathcal{O}\left(\frac{1}{n}\right) \right) \tag{422}$$

$$= -\sum_{k=1}^{n} \int \frac{\sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2}{n} (-1)^l \frac{\sigma_w^{2\ell}}{n} \left(\prod_{j=1}^{l} \frac{q^{(\ell)} \mathbf{\Sigma}_{kk}^{(\ell)}}{n} \lambda_{i_j \mu_j} \lambda_{i_{j+1} \mu_j} \right)$$
$$e^{-\sum_{j=1}^{n} \frac{1}{2} \log det(I + \frac{q^{(\ell)} \mathbf{\Sigma}_{jj}^{(\ell)}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top)} e^{tr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{423}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2\ell} \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2}{n^2} \int \left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \frac{\sqrt{q^{(\ell)} \mathbf{\Sigma}_{kk}^{(\ell)}}}{\sqrt{n}} \lambda_{\alpha\beta} \right)$$
$$e^{-\frac{q^{(\ell)} tr \mathbf{\Sigma}^{(\ell)}}{2n} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - \frac{1}{2} \sum_{\xi \geq 2} \frac{(-1)^{\xi+1}}{\xi} tr \mathbf{\Sigma}^{(\ell)\xi} tr\left(\frac{q^{(\ell)}}{n} \mathbf{\Lambda}\mathbf{\Lambda}^\top\right)^\xi - i tr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z$$
$$\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{424}$$

where in the last equation the Taylor expansion $logdet|\mathbf{I} + \mathbf{X}| = \sum_{\xi=1} \frac{(-1)^{\xi+1}}{\xi} tr(\mathbf{X}^\xi)$ was used. Similar to Lemma A.3, the first order expansion of the log-determinant. $\xi = 1$, is the leading order contribution with a relative error of $\mathcal{O}(1/n)$. By introducing the following change of variable,

$$\tilde{\lambda}_{\alpha\beta} = \sqrt{q^{(\ell)}} \lambda_{\alpha\beta} \tag{425}$$

we then retrieve the following expression

$$E_{\Pi_i^{(\ell)}}^{(n,\ell)} = -(-1)^l \sum_{k=1}^{n} \frac{\sigma_w^2 \sigma_x^2 \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2} \cdot$$
$$\cdot \int \left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \frac{\sqrt{q^{(\ell)}}}{\sqrt{n}} \lambda_{\alpha\beta} \right) e^{-\frac{q^{(\ell)} tr \mathbf{\Sigma}}{2n} tr(\mathbf{\Lambda}\mathbf{\Lambda}^\top) - i tr \mathbf{\Lambda}^\top \mathbf{Z}} F(z) \mathcal{D}\lambda \mathcal{D}z$$
$$\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{426}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2\ell} \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2} \int \left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} \right)$$
$$e^{-\sum_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \left(\frac{tr(\mathbf{\Sigma}) \tilde{\lambda}_{\alpha\beta}^2}{2n} + i \frac{\tilde{\lambda}_{\alpha\beta} z_{\alpha\beta}}{\sqrt{q^{(\ell)}}} \right)} F(z) \left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} \frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi \sqrt{q^{(\ell)}}} \right)$$
$$\left(\prod_{(\alpha,\beta) \in \mathcal{Z}_{\Pi_j}} dz_{\alpha\beta} \right) \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{427}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2\ell} \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2} \cdot$$
$$\cdot \prod_{(\alpha\beta) \in \mathcal{Z}_{\Pi_j}} \left(\int \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} e^{-\frac{tr(\mathbf{\Sigma}) \tilde{\lambda}_{\alpha\beta}^2}{2n} - i \frac{\tilde{\lambda}_{\alpha\beta} z_{\alpha\beta}}{\sqrt{q^{(\ell)}}}} \phi(z_{\alpha\beta}) \frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi \sqrt{q^{(\ell)}}} dz_{\alpha\beta} \right)$$
$$\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right) \tag{428}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2 \sigma_w^{2\ell} \left(\prod_{j=1}^{\ell} \theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2}$$

$$\left(\int \frac{\tilde{\lambda}_{\alpha\beta}}{\sqrt{n}} e^{-\frac{1}{2}\left(\sqrt{\frac{tr(\mathbf{\Sigma})}{n}}\tilde{\lambda}_{\alpha\beta}+\frac{iz_{\alpha\beta}\sqrt{n}}{\sqrt{tr(\mathbf{\Sigma})}\sqrt{q^{(\ell)}}}\right)^2-\frac{z_{\alpha\beta}^2 n}{tr(\mathbf{\Sigma})2q^{(\ell)}}}\phi(z_{\alpha\beta})\frac{d\tilde{\lambda}_{\alpha\beta}}{2\pi\sqrt{q^{(\ell)}}}dz_{\alpha\beta}\right)^{2l}$$
$$\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{429}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2}$$
$$\left(\int \frac{\phi(z)}{\sqrt{2\pi}\sqrt{n}\sqrt{q^{(\ell)}}} e^{-\frac{z^2 n}{2q^{(\ell)} tr\mathbf{\Sigma}}} \left(\int \frac{\tilde{\lambda}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\sqrt{\frac{tr(\mathbf{\Sigma})}{n}}\tilde{\lambda}+\frac{iz\sqrt{n}}{\sqrt{tr(\mathbf{\Sigma})}\sqrt{q^{(\ell)}}}\right)^2} d\tilde{\lambda}\right) dz\right)^{2l}$$
$$\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{430}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2}$$
$$\left(\int \frac{\phi(z)}{\sqrt{2\pi}\sqrt{n}\sqrt{q^{(\ell)}}} e^{-\frac{z^2 n}{2tr(\mathbf{\Sigma})q^{(\ell)}}} \left(-\frac{iz\sqrt{n}}{\sqrt{tr(\mathbf{\Sigma})}\sqrt{q^{(\ell)}}}\right) dz\right)^{2l}$$
$$\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{431}$$
$$= -(-1)^l \sum_{k=1}^{n} \frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \mathbf{\Sigma}_{kk}^{(\ell)l}}{n^2}\left(\frac{-1}{n}\right)^l$$
$$\left(\int_{\mathbf{Z}} \frac{\phi(z)}{\sqrt{2\pi}} e^{-\frac{z^2 n}{2q^{(\ell)}tr(\mathbf{\Sigma})}} \frac{z\sqrt{n}}{\sqrt{tr(\mathbf{\Sigma})q^{(\ell)}}} dz\right)^{2l} \left(1+\mathcal{O}\left(\frac{1}{n}\right)\right). \tag{432}$$

Now the following change of variable is considered

$$\tilde{z} = \frac{z\sqrt{n}}{\sqrt{tr(\mathbf{\Sigma}^{(\ell)})}\sqrt{q^{(\ell)}}} \tag{433}$$

and then

$$E_{\Pi_j^{(\ell)}}^{(n,\ell)} = -\frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2}{n^{1+l}}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{\Sigma}_{kk}^{(\ell)l}\right)\left(\int_{\mathbf{Z}} \frac{\tilde{z}\phi(\sqrt{q^{(\ell)}}tr(\mathbf{\Sigma}^{(\ell)})/n\tilde{z})}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} d\tilde{z}\right)^{2l}$$
$$\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{434}$$
$$= -\frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2}{n^{1+l}}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{\Sigma}_{kk}^{(\ell)l}\right)\left(\sqrt{q^{(\ell)}}\int_{\mathbf{Z}} \frac{\phi'(\sqrt{q^{(\ell)}}tr(\mathbf{\Sigma}^{(\ell)})/n\tilde{z})}{\sqrt{2\pi}} e^{-\frac{\tilde{z}^2}{2}} d\tilde{z}\right)^{2l}$$
$$\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{435}$$
$$= -\frac{\sigma_x^2\sigma_w^{2\ell}\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2 \tilde{\theta}_2^{(n)l}}{n^{1+l}}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{\Sigma}_{kk}^{(\ell)l}\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right). \tag{436}$$

□

C.2.1 Proof of Lemma A.5

Proof. Lemma C.4 is now going to be used for the computation of E_{2k}. Assume that $k > 1$, and consider the contribution from an addend ω of (100) with n_ϕ and n_w factors of the type $\phi(\sum_q \mathbf{W}^{(\ell)}_{i_\xi,q}\mathbf{Y}^{(\ell)}_{q,\mu_\xi})\phi(\sum_q \mathbf{W}^{(\ell)}_{i_{\xi+1},q}\mathbf{Y}^{(\ell)}_{q,\mu_\xi})$ and $\sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2\mathfrak{W}^{(\ell)}_{i_\xi p}\mathfrak{W}^{(\ell)}_{i_{\xi+1}p}$ respectively, note that $k = n_w + n_\phi$. Then, given an index p

$$E_\omega^{(n,k,n_w,n_\phi,p)} = \int \prod_{\xi=1}^{n_w}\left(\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}^{(\ell)}_{i_\xi p}\left(\prod_{\ell=1}^{n_\phi^{(\xi)}}\phi(\sum_q \mathbf{W}^{(\ell)}_{i_{\xi+\ell},q}\mathbf{Y}^{(\ell)}_{q,\mu_{\xi+\ell}})\phi(\sum_q \mathbf{W}^{(\ell)}_{i_{\xi+\ell+1},q}Y^{(\ell)}_{q,\mu_{\xi+\ell}})\right)\right.$$
$$\left.\sigma_x\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)\mathfrak{W}^{(\ell)}_{i_{\xi+1}p}\right)\mathcal{D}\mathbf{W}\mathcal{D}\mathbf{X}\mathcal{D}\mathbf{\Sigma}^{(\ell)} \tag{437}$$

$$= \int\prod_{\xi=1}^{n_w} E_{\Pi_\xi}^{(n,n_\phi^{(\xi)})}\mathcal{D}\mathbf{\Sigma}^{(\ell)} \tag{438}$$

$$= \int\prod_{\xi=1}^{n_w} -\frac{\sigma_w^{2\ell}\sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2\tilde{\theta}_2^{(n)n_\phi^{(\xi)}}}{n^{1+n_w^{(\xi)}}}\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{\Sigma}_{kk}^{(\ell)n_\phi^{(\xi)}}\right)\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{439}$$

$$= \int\left(\frac{-\sigma_w^{2\ell}\sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2}{n}\right)^{n_w}\left(\frac{\tilde{\theta}_2^{(n)}}{n}\right)^{\sum_{\xi=1}^{n_w}n_\phi^{(\xi)}}\left(\prod_{\xi=1}\frac{1}{n}\sum_{k=1}^{n}\mathbf{\Sigma}_{kk}^{(\ell)n_\phi^{(\xi)}}\right)\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{440}$$

$$= \int\left(-\frac{\sigma_w^{2}\sigma_x^2\left(\prod_{j=1}^{\ell}\theta_3^{(j)}\right)^2}{n}\right)^{n_w}\left(\frac{\tilde{\theta}_2^{(n)}}{n}\right)^{k-n_w}\left(\prod_{\xi=1}\mu_{n_\phi^{(\xi)}}\right)\mathcal{D}\mathbf{\Sigma}^{(\ell)}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{441}$$

C.3 Lemma A.6

Considering the covariance as $\mathbf{\Sigma} = \mathbf{I}$, then all the moments are equal to one. Therefore,

$$E_\omega^{(k,n_w,n_\phi,p)} = \left(-\frac{\sigma_w^2\sigma_x^2\left(\prod_{l=1}^{\ell}\theta_3^{(l)}\right)^2}{n}\right)^{n_w}\left(\frac{\theta_2}{n}\right)^{k-n_w}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right). \tag{442}$$

Then the expected contribution for the $2k$ cycle is defined by considering the contribution $\tilde{E}_{2k}^{\phi} = n_0^{1-k}\theta_2^k\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$ of the addend ω whose $n_w = 0$ and of the addends for which $n_w \neq 0$

$$E_{2k} = \tilde{E}_{2k}^{\phi} + \sum_{n_w=1}^{k}\binom{k}{n_w}\sum_{p=1}^{n}E_\omega^{(k,n_w,n_\phi,p)} \tag{443}$$

$$= \left(n_0^{1-k}\theta_2^k + n_0^{1-k}\sum_{n_w=1}^{k}\binom{k}{n_w}\left(-\sigma_w^2\sigma_x^2\left(\prod_{l=1}^{\ell}\theta_3^{(l)}\right)^2\right)^{n_w}\theta_2^{k-n_w}\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{444}$$

$$= \left(n_0^{1-k}\left(\theta_2 - \sigma_w^{2\ell}\sigma_x^2\left(\prod_{l=1}^{\ell}\theta_3^{(l)}\right)^2\right)^k\right)\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right) \tag{445}$$

□