DEPARTAMENTO DE
MATEMÁTICA

RG Flow of Preactivations

Luiz Fernando Bossa
Universidade Federal de Santa Catarina

1 de junho de 2025

Recap

Deeper Layers: Accumulation of Non-Gaussianity

- Recursion
- Action
- Large-width expansion

Marginalization Rules

- Marginalization over samples
- Marginalization over neurons
- Running couplings with partial marginalizations

Recap

Deeper Layers: Accumulation of Non-Gaussianity

- Recursion
- Action
- Large-width expansion

Marginalization Rules

- Marginalization over samples
- Marginalization over neurons
- Running couplings with partial marginalizations

- Cálculo da distribuição condicional

$$p\left(z^{(2)}, z^{(1)} \middle| \mathcal{D}\right) = p\left(z^{(2)} \middle| z^{(1)}\right) p\left(z^{(1)} \middle| \mathcal{D}\right) \tag{4.32}$$

$$p\left(z^{(2)} \middle| z^{(1)}\right) = \frac{1}{\sqrt{\left|2\pi \hat{G}^{(2)}\right|^{n_2}}} \exp\left(-\frac{1}{2} \sum_{\alpha_1,\alpha_2 \in \mathcal{D}} \widehat{G}_{(2)}^{\alpha_1\alpha_2} z_{\alpha_1}^{(2)} \cdot z_{\alpha_2}^{(2)}\right) \tag{4.35}$$

- Métrica estocástica da 2ª camada

$$\widehat{G}^{(2)}_{\alpha_1\alpha_2} := C_b^{(2)} + C_W^{(2)} \frac{1}{n_1} \sum_{j=1}^{n_1} \sigma^{(1)}_{j;\alpha_1} \sigma^{(1)}_{j;\alpha_2} \tag{4.36}$$

- Média da métrica da 2ª camada

$$\begin{aligned} G^{(2)}_{\alpha_1\alpha_2} :&= \mathbb{E}\left[\widehat{G}^{(2)}_{\alpha_1\alpha_2}\right] = C_b^{(2)} + C_W^{(2)} \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbb{E}\left[\sigma^{(1)}_{j;\alpha_1} \sigma^{(1)}_{j;\alpha_2}\right] \\ &= C_b^{(2)} + C_W^{(2)} \left\langle \sigma_{\alpha_1} \sigma_{\alpha_2} \right\rangle_{G^{(1)}} \end{aligned} \tag{4.37}$$

- Flutuação da 2ª camada: desvio da média

$$\widehat{\Delta G}^{(2)}_{\alpha_1\alpha_2} := \widehat{G}^{(2)}_{\alpha_1\alpha_2} - G^{(2)}_{\alpha_1\alpha_2} \tag{4.38}$$

- Vértice de 4 pontos: tamanho médio da flutuação

$$\mathbb{E}\left[\widehat{G}^{(2)}_{\alpha_1\alpha_2}\widehat{G}^{(2)}_{\alpha_3\alpha_4}\right] = \frac{1}{n_1}\left(C_W^{(2)}\right)^2\left(\langle\sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4}\rangle_{G^{(1)}} - \langle\sigma_{\alpha_1}\sigma_{\alpha_2}\rangle_{G^{(1)}} \langle\sigma_{\alpha_3}\sigma_{\alpha_4}\rangle_{G^{(1)}} \right) =: V^{(2)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} \tag{4.40}$$

- Pré-ativação na camada $\ell+1$ é dada por

$$z_{i;\alpha}^{(\ell+1)} = b_i^{(\ell+1)} + \sum_{j=1}^{n_\ell} W_{ij}^{(\ell+1)} \sigma_{j;\alpha}^{(\ell)}$$

com

$$\sigma_{j;\alpha}^{(\ell)} := \sigma\left(z_{i;\alpha}^{(\ell)}\right)$$

$$p\left(z^{(\ell+1)}, z^{(\ell)} \middle| \mathcal{D}\right) = p\left(z^{(\ell+1)} \middle| z^{(\ell)}\right) p\left(z^{(\ell)} \middle| \mathcal{D}\right) \tag{4.67}$$

Distribuição condicional camada $\ell + 1$

$$p\left(z^{(\ell+1)} \middle| z^{(\ell)}\right) = \frac{1}{\sqrt{\left|2\pi \hat{G}^{(\ell+1)}\right|^{n_{\ell+1}}}} \exp\left(-\frac{1}{2} \sum_{\alpha_1, \alpha_2 \in \mathcal{D}} \widehat{G}^{\alpha_1 \alpha_2}_{(\ell+1)} z^{(\ell+1)}_{\alpha_1} \cdot z^{(\ell+1)}_{\alpha_2}\right) \tag{4.69}$$

Métrica estocástica da camada $\ell + 1$

$$\widehat{G}^{(\ell+1)}_{\alpha_1 \alpha_2} := C_b^{(\ell+1)} + C_W^{(\ell+1)} \frac{1}{n_1} \sum_{j=1}^{n_1} \sigma^{(\ell)}_{j;\alpha_1} \sigma^{(\ell)}_{j;\alpha_2} \tag{4.70}$$

Média da métrica estocástica da camada $\ell + 1$

$$G^{(\ell+1)}_{\alpha_1\alpha_2} := \mathbb{E}\left[\widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2}\right] = C_b^{(\ell+1)} + C_W^{(\ell+1)} \frac{1}{n_1} \sum_{j=1}^{n_\ell} \mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\right] \tag{4.72}$$

Essa média governa o correlator de dois pontos

$$\mathbb{E}\left[z^{(\ell+1)}_{i_1;\alpha_1} z^{(\ell+1)}_{i_2;\alpha_2}\right] = \delta_{i_1 i_2} G^{(\ell+1)}_{\alpha_1\alpha_2} \tag{4.73}$$

Flutuação da métrica

$$\widehat{\Delta G}^{(\ell+1)}_{\alpha_1\alpha_2} := \widehat{G}^{(\ell+1)}_{\alpha_1\alpha_2} - G^{(\ell+1)}_{\alpha_1\alpha_2} \tag{4.74}$$

Magnitude da flutuação

$$\frac{1}{n_\ell} V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} := \mathbb{E}\left[\widehat{\Delta G}^{(\ell+1)}_{\alpha_1\alpha_2} \widehat{\Delta G}^{(\ell+1)}_{\alpha_3\alpha_4}\right] \tag{4.76}$$

$$
\mathbb{E}\left[z^{(\ell+1)}_{i_1;\alpha_1} z^{(\ell+1)}_{i_2;\alpha_2} z^{(\ell+1)}_{i_3;\alpha_3} z^{(\ell+1)}_{i_4;\alpha_4}\right]\Big|_C =
$$

$$
= \frac{1}{n_\ell}\left(\delta_{i_1 i_2}\delta_{i_3 i_4} V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} + \delta_{i_1 i_3}\delta_{i_2 i_4} V^{(\ell+1)}_{(\alpha_1\alpha_3)(\alpha_2\alpha_4)} + \right.
$$

$$
\left. + \delta_{i_1 i_4}\delta_{i_2 i_3} V^{(\ell+1)}_{(\alpha_1\alpha_4)(\alpha_2\alpha_3)}\right) \quad (4.77)
$$

Podemos definir a distribuição na camada ℓ através da ação

$$p\left(z^{(\ell)}\middle|\mathcal{D}\right) = \frac{e^{-S(z^{(\ell)})}}{Z_\ell} \tag{4.78}$$

com

$$Z_\ell := \int \left[\prod_{i,\alpha} dz_{i;\alpha}^{(\ell)}\right] e^{-S(z^{(\ell)})} \tag{4.79}$$

sendo o termo de normalização.

Nosso modelo para a ação S será

$$S(z^{(\ell)}) := \frac{1}{2}\sum_{\alpha_1,\alpha_2} g_{(\ell)}^{\alpha_1\alpha_2} z_{\alpha_1}\cdot z_{\alpha_2} - \frac{1}{8}\sum_{\alpha_i\in\mathcal{D}}^{1\le i\le 4} v_{(\ell)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} z_{\alpha_1}\cdot z_{\alpha_2} z_{\alpha_3}\cdot z_{\alpha_4} + \dots \tag{4.80}$$

- Esse modelo funciona para a camada 1 com

$$g_{(1)}^{\alpha_1\alpha_2} = G_{(1)}^{\alpha_1\alpha_2}, \qquad v_{(1)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = 0.$$

- Funciona para a camada 2 com

$$g_{(2)}^{\alpha_1\alpha_2} = G_{(2)}^{\alpha_1\alpha_2} + O\left({}^{1}\!/_{n_1}\right), \quad v_{(2)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = \frac{1}{n_1} V_{(2)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} + O\left({}^{1}\!/_{n_1^2}\right)$$

Por analogia, temos

$$g_{(\ell)}^{\alpha_1\alpha_2} = G_{(\ell)}^{\alpha_1\alpha_2} + \mathcal{O}(v) \tag{4.81}$$

e

$$v_{(\ell)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = \frac{1}{n_{\ell-1}} V_{(\ell)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} + \mathcal{O}(v^2) \tag{4.82}$$

no qual o vértice invertido é dado por

$$V_{(\ell)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} := \sum_{\beta_i \in \mathcal{D}}^{1\le i\le 4} G_{(\ell)}^{\alpha_1\beta_1} G_{(\ell)}^{\alpha_2\beta_2} G_{(\ell)}^{\alpha_3\beta_3} G_{(\ell)}^{\alpha_4\beta_4} V^{(\ell)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} \tag{4.83}$$

- Simplificamos os cálculos fazendo

$$n_1, n_2, \ldots, n_L \sim n \gg 1 \tag{4.84}$$

Teorema

Se as métricas $G^{(\ell)}$ e $V^{(\ell)}$ são de ordem de grandeza constante $O(1)$, então $G^{(\ell+1)}$ e $V^{(\ell+1)}$ também são de ordem de grandeza constante.

Pela equação (4.72), temos que a métrica G da camada $\ell+1$ é dada por

$$G^{(\ell+1)}_{\alpha_1\alpha_2} = C_b^{(\ell+1)} + C_W^{(\ell+1)} \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} \mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\right]$$

Na sessão anterior, vimos a expressão para a esperança dentro do somatório.

A equação (4.61) calculada na sessão anterior nos dá a terrível fórmula

$$\mathbb{E}\left[\sigma_{j;\alpha_1}^{(\ell)}\sigma_{j;\alpha_2}^{(\ell)}\right] = \langle\sigma_{\alpha_1}\sigma_{\alpha_2}\rangle_{G^{(\ell)}} + \frac{1}{8}\sum_{\beta_i\in\mathcal{D}}^{1\leq i\leq 4} v_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)}\left(\text{𓀠}\right) + O(v^2) =$$

$$= \langle\sigma_{\alpha_1}\sigma_{\alpha_2}\rangle_{G^{(\ell)}} + \frac{1}{8}\sum_{\beta_i\in\mathcal{D}}^{1\leq i\leq 4} \frac{1}{n_\ell}V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)}\left(\text{𓀠}\right) + O(1/n_\ell^2) =$$

$$= \langle\sigma_{\alpha_1}\sigma_{\alpha_2}\rangle_{G^{(\ell)}} + \text{𓁶} + O(1/n_\ell^2)$$

em que o hieróglifo 𓀠 representa a exata sensação ao ver essa expressão.

- 𓀠 tem um termo de ordem n_ℓ, que se torna de ordem constante quando dividimos por n_ℓ. #Melholhar

Assim, a métrica da camada $\ell + 1$ é dada por

$$\begin{aligned} G^{(\ell+1)}_{\alpha_1\alpha_2} = C^{(\ell+1)}_b + C^{(\ell+1)}_W \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} \left[\langle \sigma_{\alpha_1}\sigma_{\alpha_2} \rangle_{G^{(\ell)}} + O(1/n) \right] \\ = C^{(\ell+1)}_b + C^{(\ell+1)}_W \langle \sigma_{\alpha_1}\sigma_{\alpha_2} \rangle_{G^{(\ell)}} + O\left(1/n\right) \end{aligned}$$

Pela hipótese de indução, essa expectativa em vermelho é de ordem constante. Segue que a métrica da camada $\ell + 1$ é de ordem constante.

Para o vértice de quatro pontos, temos

$$\frac{1}{n_\ell} V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = \left(\frac{C_W^{(\ell+1)}}{n_\ell}\right)^2 \sum_{j,k=1}^{n_\ell} \Big\{ \mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\sigma^{(\ell)}_{k;\alpha_3}\sigma^{(\ell)}_{k;\alpha_4}\right] - \mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\right]\mathbb{E}\left[\sigma^{(\ell)}_{k;\alpha_3}\sigma^{(\ell)}_{k;\alpha_4}\right]\Big\}$$

- Vamos dar um nome para a expressão entre chaves: $\Xi^{(\ell)}_{j;k}$

Para índices iguais, a equação (4.62) nos dá o seguinte resultado:

$$\mathbb{E}\left[\sigma_{j;\alpha_1}^{(\ell)}\sigma_{j;\alpha_2}^{(\ell)}\sigma_{j;\alpha_3}^{(\ell)}\sigma_{j;\alpha_4}^{(\ell)}\right] - \mathbb{E}\left[\sigma_{j;\alpha_1}^{(\ell)}\sigma_{j;\alpha_2}^{(\ell)}\right]\mathbb{E}\left[\sigma_{j;\alpha_3}^{(\ell)}\sigma_{j;\alpha_4}^{(\ell)}\right] = \left\langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4}\right\rangle_{G^{(\ell)}} - \left\langle \sigma_{\alpha_1}\sigma_{\alpha_2}\right\rangle_{G^{(\ell)}}\left\langle \sigma_{\alpha_3}\sigma_{\alpha_4}\right\rangle_{G^{(\ell)}} + O(1/n)$$

Para índices diferentes, a equação (4.63) nos dá o seguinte resultado:

$$\mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\sigma^{(\ell)}_{k;\alpha_3}\sigma^{(\ell)}_{k;\alpha_4}\right] - \mathbb{E}\left[\sigma^{(\ell)}_{j;\alpha_1}\sigma^{(\ell)}_{j;\alpha_2}\right]\mathbb{E}\left[\sigma^{(\ell)}_{k;\alpha_3}\sigma^{(\ell)}_{k;\alpha_4}\right] =$$

$$= \frac{1}{4}\sum_{\beta_i\in\mathcal{D}}^{1\leq i\leq 4} v_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)}\left(\text{𓄿}\right) + O(v^2) =$$

$$= \frac{1}{4}\sum_{\beta_i\in\mathcal{D}}^{1\leq i\leq 4} \frac{1}{n_\ell} V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)}\left(\text{𓄿}\right) + O(1/n_\ell^2)$$

- O termo 𓄿 é de ordem constante, pois só contém integrais gaussianas dependentes de $G^{(\ell)}$.

Voltando para nossa equação, separamos a soma de índices iguais e diferentes, e aplicamos $n_\ell = n$.

$$\frac{1}{n} V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = \left(\frac{C_W^{(\ell+1)}}{n}\right)^2 \left\{ \sum_{j=k}^{n} \Xi^{(\ell)}_{j;k} + \sum_{j\neq k}^{n} \Xi^{(\ell)}_{j;k} \right\} =$$

$$= \frac{{C_W^{(\ell+1)}}^2}{n^2} \left\{ \sum_{j=1}^{n} \langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4} \rangle_{G^{(\ell)}} - \langle \sigma_{\alpha_1}\sigma_{\alpha_2} \rangle_{G^{(\ell)}} \langle \sigma_{\alpha_3}\sigma_{\alpha_4} \rangle_{G^{(\ell)}} + O(1/n) \right.$$

$$\left. + \sum_{j\neq k}^{n} \left(\frac{1}{4n} \sum_{\beta_i \in \mathcal{D}}^{1\leq i\leq 4} V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)} \left(\text{𓅞} \right) + O(1/n^2) \right) \right\} =$$

$$
= \frac{C_W^{(\ell+1)^2}}{n^2} \Bigg\{ n \Big[\left\langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} - \left\langle \sigma_{\alpha_1}\sigma_{\alpha_2} \right\rangle_{G^{(\ell)}} \left\langle \sigma_{\alpha_3}\sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} + O(1/n) \Big]
$$

$$
+ (n^2 - n) \left[\frac{1}{4n} \sum_{\beta_i \in \mathcal{D}}^{1 \le i \le 4} V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)} \left(\text{𓅃} \right) + O(1/n^2) \right] \Bigg\} =
$$

$$
= C_W^{(\ell+1)^2} \Bigg\{ \frac{1}{n} \Big[\left\langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} - \left\langle \sigma_{\alpha_1}\sigma_{\alpha_2} \right\rangle_{G^{(\ell)}} \left\langle \sigma_{\alpha_3}\sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} \Big] + O(1/n^2)
$$

$$
+ \frac{1}{4n} \left[\sum_{\beta_i \in \mathcal{D}}^{1 \le i \le 4} V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)} (\text{𓅃}) \right] + O(1/n^2) \Bigg\}
$$

$$
C_W^{(\ell+1)^2} \left\{ \frac{1}{n} \Big[\left\langle \sigma_{\alpha_1} \sigma_{\alpha_2} \sigma_{\alpha_3} \sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} - \left\langle \sigma_{\alpha_1} \sigma_{\alpha_2} \right\rangle_{G^{(\ell)}} \left\langle \sigma_{\alpha_3} \sigma_{\alpha_4} \right\rangle_{G^{(\ell)}} \Big] \right.
$$

$$
\left. + \frac{1}{4n} \left[\sum_{\beta_i \in \mathcal{D}}^{1 \le i \le 4} V_{(\ell)}^{(\beta_1 \beta_2)(\beta_3 \beta_4)} \big(\text{𓄿} \big) \right] \right\} + O\left({}^1/_{n^2} \right)
$$

Por hipótese de indução, temos que as partes em azul são de ordem constante.

$$\frac{1}{n}V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = \frac{{C_W^{(\ell+1)}}^2}{n}\left\{ \left[\begin{array}{l} \langle \sigma_{\alpha_1}\sigma_{\alpha_2}\sigma_{\alpha_3}\sigma_{\alpha_4}\rangle_{G^{(\ell)}} \\ - \langle \sigma_{\alpha_1}\sigma_{\alpha_2}\rangle_{G^{(\ell)}} \langle \sigma_{\alpha_3}\sigma_{\alpha_4}\rangle_{G^{(\ell)}} \end{array} \right] + \frac{1}{4}\left[\sum_{\beta_i \in \mathcal{D}}^{1\leq i\leq 4} V_{(\ell)}^{(\beta_1\beta_2)(\beta_3\beta_4)}\left(\quad \right) \right] \right\} + O(1/n^2)$$

Logo, segue que

$$\frac{1}{n}V^{(\ell+1)}_{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} = O(1/n) \tag{4.91}$$

o que completa a indução. ■

Recap

Deeper Layers: Accumulation of Non-Gaussianity

Recursion

Action

Large-width expansion

Marginalization Rules

Marginalization over samples

Marginalization over neurons

Running couplings with partial marginalizations

- Queremos calcular o valor esperado de uma função $F(z_{\mathcal{I};\mathcal{A}})$, com $\mathcal{I} \subset \mathcal{N} =: \{1, \ldots, n_\ell\}$ e $\mathcal{A} \subset \mathcal{D}$.
- Para conjuntos $X \subset Y$, vamos usar a notação $\overline{X}$ para denotar o conjunto complementar de X em Y, notadamente $Y \setminus X$.
- Vamos separar as variáveis de integração em dois conjuntos: $\mathcal{I} \times \mathcal{A}$ que é de interesse e o seu complementar $\overline{\mathcal{I} \times \mathcal{A}}$.

$$
\begin{aligned}
\mathbb{E}[F(z_{\mathcal{I};\mathcal{A}})] &= \int \left[\prod_{(i,\alpha)\in\mathcal{N}\times\mathcal{D}} dz^{(\ell)}_{i;\alpha}\right] F(z_{\mathcal{I};\mathcal{A}})p(z_{\mathcal{N};\mathcal{D}} \mid \mathcal{D}) = \\
&= \int \left[\prod_{(i,\alpha)\in\mathcal{I}\times\mathcal{A}} dz^{(\ell)}_{i;\alpha} \prod_{(j,\beta)\in\overline{\mathcal{I}\times\mathcal{A}}} dz^{(\ell)}_{j;\beta}\right] F(z_{\mathcal{I};\mathcal{A}})p(z_{\mathcal{N};\mathcal{D}} \mid \mathcal{D}) = \\
&= \int \left[\prod_{(i,\alpha)\in\mathcal{I}\times\mathcal{A}} dz^{(\ell)}_{i;\alpha}\right] F(z_{\mathcal{I};\mathcal{A}}) \int \left[\prod_{(j,\beta)\in\overline{\mathcal{I}\times\mathcal{A}}} dz^{(\ell)}_{j;\beta}\right] p(z_{\mathcal{N};\mathcal{D}} \mid \mathcal{D}) = \\
&= \int \left[\prod_{(i,\alpha)\in\mathcal{I}\times\mathcal{A}} dz^{(\ell)}_{i;\alpha}\right] F(z_{\mathcal{I};\mathcal{A}})p(z_{\mathcal{I};\mathcal{A}} \mid \mathcal{A}) \qquad (4.92)
\end{aligned}
$$

$$p(z_{\mathcal{I};\mathcal{A}} \mid \mathcal{A}) := \int \left[\prod_{(j,\beta)\in\overline{\mathcal{I}\times\mathcal{A}}} dz^{(\ell)}_{j;\beta} \right] p(z_{\mathcal{N};\mathcal{D}} \mid \mathcal{D}) \tag{4.93}$$

- Ao invés de calcular $\langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(\ell)}}$ como uma integral $N_{\mathcal{D}}$-dimensional, podemos calcular como uma integral dupla.
- Da mesma forma, podemos calcular $\langle \sigma_{\alpha_1} \sigma_{\alpha_2} \sigma_{\alpha_3} \sigma_{\alpha_4} \rangle_{G^{(\ell)}}$ como uma integral em no máximo 4 variáveis.
- Para o cálculo de $V^{(\ell+1)}_{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}$, podemos usar (4.90) e somar somente sobre os índices $\mathcal{A} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, lembrando de ajustar a métrica inversa $V^{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}_{(\ell)}$

$$-\frac{1}{8}\sum_{i,j}^{n_\ell}\sum_{\alpha_i\in\mathcal{D}}^{1\le i\le 4} v_{(\ell)}^{(\alpha_1\alpha_2)(\alpha_3\alpha_4)} z_{i;\alpha_1}^{(\ell)} z_{i;\alpha_2}^{(\ell)} z_{j;\alpha_3}^{(\ell)} z_{j;\alpha_4}^{(\ell)} \sim O\left(\frac{n_\ell^2}{n_\ell}\right) = O(n) \tag{4.95}$$

$$\frac{1}{2}\sum_{i}^{n_\ell}\sum_{\alpha_i\in\mathcal{D}}^{1\le i\le 2} g_{(\ell)}^{\alpha_1\alpha_2} z_{i;\alpha_1}^{(\ell)} z_{i;\alpha_2}^{(\ell)} \sim O(n) \tag{4.96}$$

- Como utilizamos $m_\ell \ll n_\ell$, os somatórios reduzem drasticamente o número de termos.
- $(4.95) \sim O\left(\frac{m_\ell^2}{n_\ell}\right) = O\left(\frac{1}{n}\right)$
- $(4.96) \sim O(m_\ell) = O(1)$

- Como estamos calculando apenas sobre um subconjunto de neurônios e amostras, temos que ajustar g e v de acordo.
- Para simplificar, vamos considerar apenas um input x e vamos derrubar os índices de amostra.

$$
\begin{aligned}
&p(z_1^{(\ell)}, \ldots, z_{m_\ell}^{(\ell)}) \propto e^{-S(z_1^{(\ell)}, \ldots, z_{m_\ell}^{(\ell)})} \\
&\quad = \exp\left(-\frac{g_{(\ell),m_\ell}}{2} \sum_{i=1}^{m_\ell} z_i^{(\ell)} z_i^{(\ell)} + \frac{v_{(\ell),m_\ell}}{8} \sum_{j,k=1}^{m_\ell} z_j^{(\ell)} z_j^{(\ell)} z_k^{(\ell)} z_k^{(\ell)}\right)
\end{aligned}
\tag{4.97}
$$

- Vamos integrar sobre os últimos $n_\ell - m_\ell$ neurônios, ignorando as constantes de normalização.

$$e^{-S(z_1^{(\ell)},\dots,z_{m_\ell}^{(\ell)})} \propto p(z_1^{(\ell)},\dots,z_{m_\ell}^{(\ell)}) = \int \Big[\prod_{i=m_\ell+1}^{n_\ell} dz_i^{(\ell)}\Big] p(z_1^{(\ell)},\dots,z_{n_\ell}^{(\ell)})$$

$$\propto \int \Big[\prod_{i=m_\ell+1}^{n_\ell} dz_i^{(\ell)}\Big] \exp\left(-\frac{g_{(\ell),n_\ell}}{2}\sum_{i=1}^{n_\ell} z_i^{(\ell)} z_i^{(\ell)} + \frac{v_{(\ell)}}{8}\sum_{j,k=1}^{n_\ell} z_j^{(\ell)} z_j^{(\ell)} z_k^{(\ell)} z_k^{(\ell)}\right)$$

- Para simplificar, vamos sumir com os índices ℓ.
- Vamos modificar a notação

$$\int \left[\prod_{i=a}^{b} dz_i\right] = \int_{i=a}^{b} dz_i$$

- Vamos lembrar que $\exp(a+b) = \exp(a)\exp(b)$.
- Vamos separar o somatório duplo

$$\sum_{j,k=1}^{n} = \sum_{j,k=1}^{m} + \sum_{j=1}^{m}\sum_{k=m+1}^{n} + \sum_{j=m+1}^{n}\sum_{k=1}^{m} + \sum_{j,k=m+1}^{n}$$

$$
\begin{aligned}
p(z_1^{(\ell)}, \dots, z_{m_\ell}^{(\ell)}) \propto \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2} \sum_{i=1}^{n} z_i^2 + \frac{v}{8} \sum_{j,k=1}^{n} z_j^2 z_k^2 \right] = \\
= \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2} \sum_{i=1}^{n} z_i^2 \right] \exp\left[\frac{v}{8} \sum_{j,k=1}^{n} z_j^2 z_k^2 \right] = \\
= \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2} \sum_{i=1}^{m} z_i^2 - \frac{g}{2} \sum_{i=m+1}^{n} z_i^2 \right] \exp\left[\frac{v}{8} \sum_{j,k=1}^{n} z_j^2 z_k^2 \right] = \\
= \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2} \sum_{i=1}^{m} z_i^2 \right] \exp\left[-\frac{g}{2} \sum_{i=m+1}^{n} z_i^2 \right] \exp\left[\frac{v}{8} \sum_{j,k=1}^{n} z_j^2 z_k^2 \right] = \\
= \exp\left[-\frac{g}{2} \sum_{i=1}^{m} z_i^2 \right] \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2} \sum_{i=m+1}^{n} z_i^2 \right] \exp\left[\frac{v}{8} \sum_{j,k=1}^{n} z_j^2 z_k^2 \right] =
\end{aligned}
$$

MTM

$$= \exp\left[-\frac{g}{2}\sum_{i=1}^{m} z_i^2\right] \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2}\sum_{i=m+1}^{n} z_i^2\right] \exp\left[\frac{v}{8}\sum_{j,k=1}^{m} z_j^2 z_k^2 + \right.$$

$$\left. + \frac{v}{8}\sum_{j=1}^{m}\sum_{k=m+1}^{n} z_j^2 z_k^2 + \frac{v}{8}\sum_{j=m+1}^{n}\sum_{k=1}^{n} z_j^2 z_k^2 + \frac{v}{8}\sum_{j,k=m+1}^{n} z_j^2 z_k^2\right] =$$

$$= \exp\left[-\frac{g}{2}\sum_{i=1}^{m} z_i^2\right] \exp\left[\frac{v}{8}\sum_{j,k=1}^{m} z_j^2 z_k^2 + \right] \times$$

$$\times \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2}\sum_{i=m+1}^{n} z_i^2\right] \exp\left[\frac{v}{8}\left(2\sum_{j=1}^{m}\sum_{k=m+1}^{n} z_j^2 z_k^2 + \sum_{j,k=m+1}^{n} z_j^2 z_k^2\right)\right]$$

- Agora usamos $\exp(v\Sigma) \approx 1 + v\Sigma + O(v^2)$ para trocar a expoencial

$$\exp\left[\frac{v}{8}\left(2\sum_{j=1}^{m}\sum_{k=m+1}^{n} z_j^2 z_k^2 + \sum_{j,k=m+1}^{n} z_j^2 z_k^2\right)\right] \approx$$

$$\approx 1 + \frac{2v}{8}\sum_{j=1}^{m}\sum_{k=m+1}^{n} z_j^2 z_k^2 + \frac{v}{8}\sum_{j,k=m+1}^{n} z_j^2 z_k^2 + O(v^2)$$

MTM

$$p(z_1^{(\ell)}, \dots, z_{m_\ell}^{(\ell)}) \propto \exp\left[-\frac{g}{2}\sum_{i=1}^{m} z_i^2 + \frac{v}{8}\sum_{j,k=1}^{m} z_j^2 z_k^2 +\right] \times$$

$$\times \int_{i=m+1}^{n} dz_i \exp\left[-\frac{g}{2}\sum_{i=m+1}^{n} z_i^2\right] \left(1 + \frac{2v}{8}\sum_{j=1}^{m}\sum_{k=m+1}^{n} z_j^2 z_k^2 + \frac{v}{8}\sum_{j,k=m+1}^{n} z_j^2 z_k^2 + O(v^2)\right)$$

$$= \exp\left[-\frac{g}{2}\sum_{i=1}^{m} z_i^2 + \frac{v}{8}\sum_{j,k=1}^{m} z_j^2 z_k^2 +\right] \times$$

$$\times \left[1 + \frac{(n-m)}{4}\frac{v}{g}\left(\sum_{i=1}^{m} z_i^2\right) + \frac{v}{8g^2}\left[(n-m)^2 + 2(n-m)\right] + O(v^2)\right] (??)$$

Comparando a expressão acima com a equação (4.97), temos que

$$g_{(\ell),m_\ell} = g_{(\ell),n_\ell} - \frac{(n_\ell - m_\ell)}{4} \frac{v_{(\ell)}}{g_{(\ell),n_\ell}} \tag{4.100}$$

$$G^{(\ell)} = \frac{1}{g_{(\ell),m_\ell}} + \frac{(m_\ell + 2)}{2} \frac{v^{(\ell)}}{g^3_{(\ell),m_\ell}} + O\left(v^2\right) \tag{4.101}$$

$$\frac{1}{g_{(\ell),m_\ell}} = G^{(\ell)} - \frac{(m_\ell + 2)}{2} \frac{V^{(\ell)}}{n_{\ell-1} G^{(\ell)}} + O\left(\frac{1}{n^2}\right) \tag{4.102}$$

$$v^{(\ell)} = \frac{V^{(\ell)}}{n_{\ell-1} (G^{(\ell)})^4} + O\left(\frac{1}{n^2}\right)$$