



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

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

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

- Métrica estocástica da 2ª camada

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

- Média da métrica da 2ª camada

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

- Flutuação da 2ª camada: desvio da média

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

- Vértice de 4 pontos: tamanho médio da flutuação

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

## 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

- ▶ 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

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



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

Distribuição condicional camada  $\ell + 1$

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

Métrica estocástica da camada  $\ell + 1$

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

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

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

Essa média governa o correlator de dois pontos

$$\mathbb{E} \left[ z_{i_1; \alpha_1}^{(\ell+1)} z_{i_2; \alpha_2}^{(\ell+1)} \right] = \delta_{i_1 i_2} G_{\alpha_1 \alpha_2}^{(\ell+1)} \quad (4.73)$$

Flutuação da métrica

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

Magnitude da flutuação

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

$$\begin{aligned}
\mathbb{E} \left[ z_{i_1; \alpha_1}^{(\ell+1)} z_{i_2; \alpha_2}^{(\ell+1)} z_{i_3; \alpha_3}^{(\ell+1)} z_{i_4; \alpha_4}^{(\ell+1)} \right] \Big|_C &= \\
&= \frac{1}{n_\ell} \left( \delta_{i_1 i_2} \delta_{i_3 i_4} V_{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}^{(\ell+1)} + \delta_{i_1 i_3} \delta_{i_2 i_4} V_{(\alpha_1 \alpha_3)(\alpha_2 \alpha_4)}^{(\ell+1)} + \right. \\
&\quad \left. + \delta_{i_1 i_4} \delta_{i_2 i_3} V_{(\alpha_1 \alpha_4)(\alpha_2 \alpha_3)}^{(\ell+1)} \right) \quad (4.77)
\end{aligned}$$

Podemos definir a distribuição na camada  $\ell$  através da ação

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

com

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

sendo o termo de normalização.

Nosso modelo para a ação  $S$  será

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

- ▶ Esse modelo funciona para a camada 1 com

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

- ▶ Funciona para a camada 2 com

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

Por analogia, temos

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

e

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

no qual o vértice invertido é dado por

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

- Simplificamos os cálculos fazendo

$$n_1, n_2, \dots, n_L \sim n \gg 1 \quad (4.84)$$



### 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


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

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

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

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

-  tem um termo de ordem  $n_\ell$ , que se torna de ordem constante quando dividimos por  $n_\ell$ . #Melhorar

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

$$\begin{aligned} G_{\alpha_1 \alpha_2}^{(\ell+1)} &= C_b^{(\ell+1)} + C_W^{(\ell+1)} \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} \left[ \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(\ell)}} + O(1/n) \right] \\ &= C_b^{(\ell+1)} + C_W^{(\ell+1)} \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(\ell)}} + O(1/n) \end{aligned}$$

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

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

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

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

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

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

$$\begin{aligned}
 \mathbb{E} \left[ \sigma_{j;\alpha_1}^{(\ell)} \sigma_{j;\alpha_2}^{(\ell)} \sigma_{k;\alpha_3}^{(\ell)} \sigma_{k;\alpha_4}^{(\ell)} \right] - \mathbb{E} \left[ \sigma_{j;\alpha_1}^{(\ell)} \sigma_{j;\alpha_2}^{(\ell)} \right] \mathbb{E} \left[ \sigma_{k;\alpha_3}^{(\ell)} \sigma_{k;\alpha_4}^{(\ell)} \right] &= \\
 = \frac{1}{4} \sum_{\beta_i \in \mathcal{D}}^{1 \leq i \leq 4} v_{(\ell)}^{(\beta_1 \beta_2)(\beta_3 \beta_4)} \left( \mathbb{I} \right) + O(v^2) &= \\
 = \frac{1}{4} \sum_{\beta_i \in \mathcal{D}}^{1 \leq i \leq 4} \frac{1}{n_\ell} V_{(\ell)}^{(\beta_1 \beta_2)(\beta_3 \beta_4)} \left( \mathbb{I} \right) + O(1/n_\ell^2)
 \end{aligned}$$

- O termo  $\mathbb{I}$  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  $n_\ell = n$ .

$$\begin{aligned}
\frac{1}{n} V_{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}^{(\ell+1)} &= \left( \frac{C_W^{(\ell+1)}}{n} \right)^2 \left\{ \sum_{j=k}^n \Xi_{j;k}^{(\ell)} + \sum_{j \neq k}^n \Xi_{j;k}^{(\ell)} \right\} = \\
&= \frac{C_W^{(\ell+1)^2}}{n^2} \left\{ \sum_{j=1}^n \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \sigma_{\alpha_3} \sigma_{\alpha_4} \rangle_{G^{(\ell)}} - \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(\ell)}} \langle \sigma_{\alpha_3} \sigma_{\alpha_4} \rangle_{G^{(\ell)}} + O(1/n) \right. \\
&\quad \left. + \sum_{j \neq k}^n \left( \frac{1}{4n} \sum_{\substack{1 \leq i \leq 4 \\ \beta_i \in \mathcal{D}}} V_{(\ell)}^{(\beta_1 \beta_2)(\beta_3 \beta_4)} \left( \mathbb{I} \right) + O(1/n^2) \right) \right\} =
\end{aligned}$$



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

$$C_W^{(\ell+1)^2} \left\{ \frac{1}{n} \left[ \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \sigma_{\alpha_3} \sigma_{\alpha_4} \rangle_{G^{(\ell)}} - \langle \sigma_{\alpha_1} \sigma_{\alpha_2} \rangle_{G^{(\ell)}} \langle \sigma_{\alpha_3} \sigma_{\alpha_4} \rangle_{G^{(\ell)}} \right] \right. \\ \left. + \frac{1}{4n} \left[ \sum_{\beta_i \in \mathcal{D}}^{1 \leq i \leq 4} V_{(\ell)}^{(\beta_1 \beta_2)(\beta_3 \beta_4)}(\mathbb{I}) \right] \right\} + O(1/n^2)$$

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

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

Logo, segue que

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

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  $\mathcal{I} \subset \mathcal{N} =: \{1, \dots, n_\ell\}$  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  $\overline{\mathcal{I} \times \mathcal{A}}$ .

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

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

- ▶ 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  $V_{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}^{(\ell+1)}$ , 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  $V_{(\ell)}^{(\alpha_1 \alpha_2)(\alpha_3 \alpha_4)}$



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

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

- ▶ 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.

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

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

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

- ▶ Para simplificar, vamos sumir com os índices  $\ell$ .
- ▶ 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

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

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

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

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

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



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

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

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

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

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

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

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

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

$$\begin{aligned}
p(z_1^{(\ell)}, \dots, z_{m_\ell}^{(\ell)}) &\propto \exp \left[ -\frac{g}{2} \sum_{i=1}^m z_i^2 + \frac{v}{8} \sum_{j,k=1}^m z_j^2 z_k^2 + \right] \times \\
&\times \int_{i=m+1}^n dz_i \exp \left[ -\frac{g}{2} \sum_{i=m+1}^n z_i^2 \right] \left( 1 + \frac{2v}{8} \sum_{j=1}^m \sum_{k=m+1}^n z_j^2 z_k^2 + \frac{v}{8} \sum_{j,k=m+1}^n z_j^2 z_k^2 + O(v^2) \right) \\
&= \exp \left[ -\frac{g}{2} \sum_{i=1}^m z_i^2 + \frac{v}{8} \sum_{j,k=1}^m z_j^2 z_k^2 + \right] \times \\
&\times \left[ 1 + \frac{(n-m)}{4} \frac{v}{g} \left( \sum_{i=1}^m z_i^2 \right) + \frac{v}{8g^2} [(n-m)^2 + 2(n-m)] + O(v^2) \right] (??)
\end{aligned}$$



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

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

$$G^{(\ell)} = \frac{1}{g_{(\ell),m_\ell}} + \frac{(m_\ell + 2)}{2} \frac{v^{(\ell)}}{g_{(\ell),m_\ell}^3} + O(v^2) \quad (4.101)$$

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

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