



DEPARTAMENTO DE
MATEMÁTICA

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Luiz Fernando Bossa
Universidade Federal de Santa Catarina

24 de abril de 2025

Notações e Definições

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Redes Lineares Profundas

Criticalidade

Flutuacoes

Notações e Definições

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Redes Lineares Profundas

Criticalidade

Flutuações

Uma rede neural com L camadas, cada camada tendo n_ℓ neurônios e dados de entrada x_α é dada por:

$$\begin{aligned} z^{(1)} &= W^{(1)}x_\alpha + b^{(1)} \\ z^{(\ell+1)} &= W^{(\ell+1)}\sigma(z^{(\ell)}) + b^{(\ell)}, \quad \ell = 1, \dots, L-1 \end{aligned} \quad (2.5)$$

- ▶ $z^{(\ell)}$ é um vetor de tamanho n_ℓ
- ▶ $W^{(\ell)}$ é uma matriz de tamanho $n_\ell \times n_{\ell-1}$

Distribuição inicial: médias zero e variâncias dadas por

$$\mathbb{E} \left(b_i^{(\ell)} b_j^{(\ell)} \right) = \delta_{ij} C_b^{(\ell)} \quad (2.19)$$

$$\mathbb{E} \left(W_{ij}^{(\ell)} W_{kl}^{(\ell)} \right) = \delta_{ik} \delta_{jl} \frac{C_W^{(\ell)}}{n_{\ell-1}} \quad (2.20)$$

Estamos trabalhando com distribuições unidimensionais.

Para duas variáveis aleatórias X e Y com médias zero, temos

$$\text{Cov}(X, Y) = \mathbb{E}((X - 0)(Y - 0)) = \mathbb{E}(XY)$$

E em particular,

$$\text{Cov}(X, X) = \mathbb{E}(X^2) = \text{Var}(X)$$

Se A é uma matriz, utilizaremos a notação

- ▶ A_{ij} para o elemento da linha i e coluna j .
- ▶ A_{i*} para a linha i .
- ▶ A_{*j} para a coluna j .
- ▶ O produto interno dos vetores u e v será denotado por $u \cdot v$.

- ▶ Particularmente eu não gosto de salada de índice, não me cai bem.
- ▶ Fiz as seguintes transformações nos índices

Original	Minha notação	Índices
i_1, i_2	i, j	coordenada fixas
j_1, j_2, j	k, l, ν	coordenadas variáveis
α_1, α_2	α, β	dados de entrada

Assim, podemos escrever as equações (2.19) e (2.20) como

$$(2.19) = \begin{cases} \text{Cov} \left(b_i^{(\ell)}, b_j^{(\ell)} \right) = 0, & i \neq j \\ \text{Var} \left(b_i^{(\ell)} \right) = C_b^{(\ell)} \end{cases} \quad (2.19')$$

$$(2.20) = \begin{cases} \text{Cov} \left(W_{ij}^{(\ell)}, W_{kl}^{(\ell)} \right) = 0, & (i, j) \neq (k, l) \\ \text{Var} \left(W_{ij}^{(\ell)} \right) = \frac{C_W^{(\ell)}}{n_{\ell-1}} \end{cases} \quad (2.20')$$

Embora não valha para todas as distribuições¹, se X e Y são variáveis aleatórias gaussianas, então X e Y são independentes se e somente se $\text{Cov}(X, Y) = 0$.

Segue que as $b_i^{(\ell)}$ e $W_{ij}^{(\ell)}$ são variáveis gaussianas independentes, com médias zero e variâncias dadas por $C_b^{(\ell)}$ e $\frac{C_W^{(\ell)}}{n_{\ell-1}}$.

¹Independence of Normals

Notações e Definições

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Redes Lineares Profundas

Criticalidade

Flutuacoes

§3.1 Redes Lineares Profundas

§3.2 Criticalidade: cálculo do correlator de 2 pontos

§3.3 Flutuações: cálculo do correlator de 4 pontos

§3.4 ~~Caos: cálculo do correlator de 6 pontos~~

- ▶ São redes neurais com funções de ativação identidade $\sigma(x) = x$.
- ▶ Para simplificar a análise, zeramos os vieses $b^{(\ell)} \equiv \vec{0}$.
- ▶ A equação (2.5) se torna

$$\begin{aligned} z^{(1)} &= W^{(1)} x_{\alpha} \\ z^{(\ell+1)} &= W^{(\ell+1)} (z^{(\ell)}), \quad \ell = 1, \dots, L-1 \end{aligned}$$

$$z_{\alpha}^{(\ell)} = W^{(\ell)} W^{(\ell-1)} \dots W^{(1)} x_{\alpha} \quad (3.2)$$

Introduzimos a notação

$$\mathcal{W}^{(\ell)} = W^{(\ell)} W^{(\ell-1)} \dots W^{(1)} \quad (3.3)$$

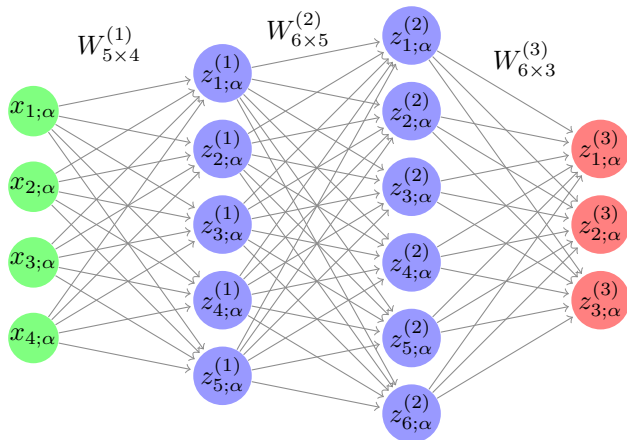
Fazemos todas as variâncias constantes e independentes da camada $C_W^{(\ell)} \equiv C_W$.

Entrada

Camada 1

Camada 2

Saída



Queremos calcular

$$p(z_{\alpha}^{(\ell)} \mid \mathcal{D})$$

- ▶ Uma distribuição é completamente determinada pelos seus momentos, que são dados por seus correlatores de M pontos.

- Note que pela equação (3.2), temos que

$$z_{\alpha}^{(\ell)} = W^{(\ell)} z_{\alpha}^{(\ell-1)} \quad (3.2')$$

- Podemos calcular a esperança de $z_{\alpha}^{(\ell)}$ componente a componente, lembrando que é o produto interno da i -ésima linha da matriz $W^{(\ell)}$ com o vetor $z_{\alpha}^{(\ell-1)}$.

$$\begin{aligned}\mathbb{E}(z_{i;\alpha}^{(\ell)}) &= \mathbb{E}\left(W_{i*}^{(\ell)} \cdot z_{\alpha}^{(\ell-1)}\right) \\ &= \mathbb{E}\left(\sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} z_{j;\alpha}^{(\ell-1)}\right) \\ &= \sum_{j=1}^{n_{\ell-1}} \mathbb{E}\left(W_{ij}^{(\ell)} z_{j;\alpha}^{(\ell-1)}\right) \\ &= \sum_{j=1}^{n_{\ell-1}} \underbrace{\mathbb{E}\left(W_{ij}^{(\ell)}\right)}_0 \mathbb{E}\left(z_{j;\alpha}^{(\ell-1)}\right) = 0\end{aligned}\tag{3.6}$$

- ▶ Os autores afirmam que, por um argumento similar, é possível mostrar que os momentos de ordem ímpar serão todos zerados.

Notações e Definições

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Redes Lineares Profundas

Criticalidade

Flutuacoes

- Vamos calcular o correlator de 2 pontos na primeira camada, coordenada a coordenada

$$\begin{aligned}\mathbb{E}(z_{i;\alpha}^{(1)} z_{j;\beta}^{(1)}) &= \mathbb{E}\left(W_{i*}^{(1)} \cdot x_{\alpha} W_{j*}^{(1)} \cdot x_{\beta}\right) \\ &= \mathbb{E}\left(\left(\sum_{k=1}^{n_0} W_{ik}^{(1)} x_{k;\alpha}\right) \left(\sum_{l=1}^{n_0} W_{jl}^{(1)} x_{l;\beta}\right)\right) \\ &= \mathbb{E}\left(\sum_{k=1}^{n_0} \sum_{l=1}^{n_0} W_{ik}^{(1)} x_{k;\alpha} W_{jl}^{(1)} x_{l;\beta}\right)\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_{k=1}^{n_0} \sum_{l=1}^{n_0} W_{ik}^{(1)} x_{k;\alpha} W_{il}^{(1)} x_{l;\beta} \right) = \sum_{k,l=1}^{n_0} \mathbb{E} \left(W_{ik}^{(1)} W_{jl}^{(1)} \right) x_{k;\alpha} x_{l;\beta} \\
&= \sum_{k,l=1}^{n_0} \delta_{ij} \delta_{kl} \frac{C_W}{n_0} x_{k;\alpha} x_{l;\beta} = \delta_{ij} \frac{C_W}{n_0} \sum_{k,l=1}^{n_0} \delta_{kl} x_{k;\alpha} x_{l;\beta} =^\dagger \\
&= \delta_{ij} \frac{C_W}{n_0} \sum_{\nu=1}^{n_0} x_{\nu;\alpha} x_{\nu;\beta} = \delta_{ij} \frac{C_W}{n_0} x_\alpha \cdot x_\beta \tag{3.8}
\end{aligned}$$

Na passagem \dagger , note que as parcelas somem quando $k \neq l$, então fazemos uma mudança de variáveis $\nu = k = l$.

Criamos a notação

$$G_{\alpha\beta}^{(0)} = \frac{1}{n_0} x_\alpha \cdot x_\beta \quad (3.9)$$

Assim

$$\mathbb{E}(z_{i;\alpha}^{(1)} z_{j;\beta}^{(1)}) = \delta_{ij} C_W G_{\alpha\beta}^{(0)} \quad (3.10)$$

- Note que no lado direito da equação acima, o único termo que depende das coordenadas i, j é δ_{ij} .

- Vamos calcular o correlator de 2 pontos na camada $\ell + 1$ de maneira recursiva, utilizando a equação (3.2')

$$z_{\alpha}^{(\ell+1)} = W^{(\ell+1)} z_{\alpha}^{(\ell)} \quad (3.2')$$

$$\begin{aligned}\mathbb{E}(z_{i;\alpha}^{(\ell+1)} z_{j;\beta}^{(\ell+1)}) &= \mathbb{E}\left(W_{i*}^{(\ell+1)} \cdot z_{\alpha}^{(\ell)} W_{j*}^{(\ell+1)} \cdot z_{\beta}^{(\ell)}\right) \\ &= \mathbb{E}\left(\left(\sum_{k=1}^{n_{\ell}} W_{ik}^{(\ell+1)} z_{k;\alpha}^{(\ell)}\right) \left(\sum_{l=1}^{n_{\ell}} W_{jl}^{(\ell+1)} z_{l;\beta}^{(\ell)}\right)\right) \\ &= \sum_{k,l=1}^{n_{\ell}} \mathbb{E}\left(W_{ik}^{(\ell+1)} W_{jl}^{(\ell+1)} z_{k;\alpha}^{(\ell)} z_{l;\beta}^{(\ell)}\right) \\ &= \sum_{k,l=1}^{n_{\ell}} \mathbb{E}\left(W_{ik}^{(\ell+1)} W_{jl}^{(\ell+1)}\right) \mathbb{E}\left(z_{k;\alpha}^{(\ell)} z_{l;\beta}^{(\ell)}\right)\end{aligned}$$

$$\begin{aligned}
&= \sum_{k,l=1}^{n_\ell} \mathbb{E} \left(W_{ik}^{(\ell+1)} W_{jl}^{(\ell+1)} \right) \mathbb{E} \left(z_{k;\alpha}^{(\ell)} z_{l;\beta}^{(\ell)} \right) \\
&= \sum_{k,l=1}^{n_\ell} \delta_{ij} \delta_{kl} \frac{C_W}{n_\ell} \mathbb{E} \left(z_{k;\alpha}^{(\ell)} z_{l;\beta}^{(\ell)} \right) = \delta_{ij} \frac{C_W}{n_\ell} \sum_{k,l=1}^{n_\ell} \delta_{kl} \mathbb{E} \left(z_{k;\alpha}^{(\ell)} z_{l;\beta}^{(\ell)} \right) \\
&= \delta_{ij} \frac{C_W}{n_\ell} \sum_{\nu=1}^{n_\ell} \mathbb{E} \left(z_{\nu;\alpha}^{(\ell)} z_{\nu;\beta}^{(\ell)} \right) \\
&= \delta_{ij} \frac{C_W}{n_\ell} \mathbb{E} \left(\sum_{\nu=1}^{n_\ell} z_{\nu;\alpha}^{(\ell)} z_{\nu;\beta}^{(\ell)} \right) = \delta_{ij} \frac{C_W}{n_\ell} \mathbb{E} (z_\alpha^{(\ell)} \cdot z_\beta^{(\ell)}) \tag{3.11}
\end{aligned}$$

- Em suma, a equação (3.11) vira

$$\mathbb{E}(z_{i;\alpha}^{(\ell+1)} z_{j;\beta}^{(\ell+1)}) = \delta_{ij} \frac{C_W}{n_\ell} \mathbb{E}(z_\alpha^{(\ell)} \cdot z_\beta^{(\ell)}) \quad (3.11)$$

- Em qualquer camada, o correlator das coordenadas i, j é sempre o delta de Kronecker vezes um número que não depende das coordenadas, permitindo assim introduzir a notação

$$\mathbb{E}(z_{i;\alpha}^{(\ell)} \cdot z_{j;\beta}^{(\ell)}) = \delta_{ij} G_{\alpha\beta}^{(\ell)} \quad (3.12)$$

- Para isolar $G_{\alpha\beta}^{(\ell)}$, vamos somar a equação (3.12) sobre todos os possíveis i e j .

$$\sum_{i,j=1}^{n_\ell} \mathbb{E}(z_{i;\alpha}^{(\ell)} z_{j;\beta}^{(\ell)}) = \sum_{i,j=1}^{n_\ell} \delta_{ij} G_{\alpha\beta}^{(\ell)}$$

$$\sum_{\nu=1}^{n_\ell} \mathbb{E}(z_{\nu;\alpha}^{(\ell)} z_{\nu;\beta}^{(\ell)}) = \sum_{\nu=1}^{n_\ell} \delta_{\nu\nu} G_{\alpha\beta}^{(\ell)}$$

$$\mathbb{E} \left(\sum_{\nu=1}^{n_\ell} z_{\nu;\alpha}^{(\ell)} z_{\nu;\beta}^{(\ell)} \right) = \sum_{\nu=1}^{n_\ell} G_{\alpha\beta}^{(\ell)}$$

$$\mathbb{E}(z_{\alpha}^{(\ell)} \cdot z_{\beta}^{(\ell)}) = n_\ell G_{\alpha\beta}^{(\ell)}$$

$$G_{\alpha\beta}^{(\ell)} = \frac{1}{n_\ell} \mathbb{E}(z_\alpha^{(\ell)} \cdot z_\beta^{(\ell)}) \quad (3.13)$$

Assim (3.11) se torna

$$\mathbb{E}(z_{i;\alpha}^{(\ell+1)} z_{j;\beta}^{(\ell+1)}) = \delta_{ij} C_W G_{\alpha\beta}^{(\ell)} \quad (3.11')$$

Usando (3.11'), podemos encontrar a recursão para $G_{\alpha\beta}^{(\ell+1)}$.

$$\begin{aligned} G_{\alpha\beta}^{(\ell+1)} &= \frac{1}{n_{\ell+1}} \mathbb{E} \left(z_{\alpha}^{(\ell+1)} \cdot z_{\beta}^{(\ell+1)} \right) \\ &= \frac{1}{n_{\ell+1}} \mathbb{E} \left(\sum_{\nu=1}^{n_{\ell+1}} z_{\nu;\alpha}^{(\ell+1)} z_{\nu;\beta}^{(\ell+1)} \right) \\ &= \frac{1}{n_{\ell+1}} \sum_{\nu=1}^{n_{\ell+1}} \mathbb{E} \left(z_{\nu;\alpha}^{(\ell+1)} z_{\nu;\beta}^{(\ell+1)} \right) \\ &= \frac{1}{n_{\ell+1}} \sum_{\nu=1}^{n_{\ell+1}} \delta_{\nu\nu} C_W G_{\alpha\beta}^{(\ell)} \\ &= \frac{C_W}{n_{\ell+1}} \sum_{\nu=1}^{n_{\ell+1}} G_{\alpha\beta}^{(\ell)} = \frac{C_W}{n_{\ell+1}} n_{\ell+1} G_{\alpha\beta}^{(\ell)} = C_W G_{\alpha\beta}^{(\ell)} \end{aligned} \tag{3.14}$$

Da equação (3.14) obtemos a recursão

$$G_{\alpha\beta}^{(\ell)} = (C_W)^\ell G_{\alpha\beta}^{(0)} \quad (3.15)$$

O observável $G_{\alpha\alpha}^{(L)}$ mede o tamanho médio do output da rede neural.

$$G_{\alpha\alpha}^{(L)} = \frac{1}{n_L} \mathbb{E} \left(z_{\alpha}^{(L)} \cdot z_{\alpha}^{(L)} \right) = \frac{1}{n_L} \mathbb{E} \left(\|z_{\alpha}^{(L)}\|^2 \right) \quad (3.16)$$

Por outro lado, note que

$$G_{\alpha\alpha}^{(L)} = (C_W)^L G_{\alpha\alpha}^{(0)}$$

Assim, dependendo do valor da variância C_W , podemos ter três cenários:

$$\lim_{L \rightarrow \infty} G_{\alpha\alpha}^{(L)} = \lim_{L \rightarrow \infty} (C_W)^L G_{\alpha\alpha}^{(0)} = \begin{cases} 0 & \text{se } C_W < 1 \\ G_{\alpha\alpha}^{(0)} & \text{se } C_W = 1 \\ \infty & \text{se } C_W > 1 \end{cases}$$

- ▶ Se $C_W < 1$, a rede neural não consegue aprender, pois o output tende a zero.
- ▶ Se $C_W > 1$, o valor do output diverge, o que significa instabilidade numérica.
- ▶ O único caso no qual a rede neural consegue aprender é quando $C_W = 1$.

Notações e Definições

Teoria Efetiva de Redes Lineares Profundas na Inicialização

Redes Lineares Profundas

Criticalidade

Flutuacoes

Novamente, para evitar subscrito duplo, vamos fazer as seguintes mudanças de notação:

- ▶ i_1, i_2, i_3, i_4 para $\hat{1}, \hat{2}, \hat{3}, \hat{4}$
- ▶ j_1, j_2, j_3, j_4 para $\underline{1}, \underline{2}, \underline{3}, \underline{4}$

- ▶ O correlator de 4 pontos na camada ℓ é dado por

$$\mathbb{E}(z_{\hat{1}}^{(\ell)} z_{\hat{2}}^{(\ell)} z_{\hat{3}}^{(\ell)} z_{\hat{4}}^{(\ell)})$$

- ▶ Vamos calcular o correlator de 4 pontos de maneira recursiva.
- ▶ Nessa sessão, vamos calcular as correlações em apenas uma entrada x_α , e portanto vamos abandonar o índice.

- ▶ Introduzimos a notação

$$G_2^{(\ell)} := G_{\alpha\alpha}^{(\ell)} = \frac{1}{n_\ell} \mathbb{E} \left(z_\alpha^{(\ell)} \cdot z_\alpha^{(\ell)} \right) \quad (3.17)$$

- ▶ Em particular, na camada 0,

$$G_2^{(0)} = \frac{1}{n_0} \mathbb{E} (x_\alpha \cdot x_\alpha) = \frac{1}{n_0} x \cdot x$$

Teorema de Wick

Para calcular momentos superiores de uma variável aleatória z , usamos a fórmula

$$\mathbb{E}(z_1 z_2 \dots z_{2m}) = \sum \mathbb{E}(z_{\widehat{k_1}} z_{\widehat{k_2}}) \mathbb{E}(z_{\widehat{k_3}} z_{\widehat{k_4}}) \dots \mathbb{E}(z_{\widehat{k_{2m-1}}} z_{\widehat{k_{2m}}})$$

em que a soma é feita sobre todos os pareamentos possíveis dos índices.

- O Teorema de Wick para 4 pontos nos diz que

$$\mathbb{E}(z_1 z_2 z_3 z_4) = \\ \mathbb{E}(z_1 z_2) \mathbb{E}(z_3 z_4) + \mathbb{E}(z_1 z_3) \mathbb{E}(z_2 z_4) + \mathbb{E}(z_1 z_4) \mathbb{E}(z_2 z_3)$$

$$\begin{aligned}
\mathbb{E}(z_{\hat{1}}^{(1)} z_{\hat{2}}^{(1)} z_{\hat{3}}^{(1)} z_{\hat{4}}^{(1)}) &= \mathbb{E}\left(W_{\hat{1}*}^{(1)} \cdot x W_{\hat{2}*}^{(1)} \cdot x W_{\hat{3}*}^{(1)} \cdot x W_{\hat{4}*}^{(1)} \cdot x\right) \\
&= \mathbb{E}\left(\sum_{\underline{1}=1}^{n_0} W_{\hat{1}\underline{1}}^{(1)} x_{\underline{1}} \sum_{\underline{2}=1}^{n_0} W_{\hat{2}\underline{2}}^{(1)} x_{\underline{2}} \sum_{\underline{3}=1}^{n_0} W_{\hat{3}\underline{3}}^{(1)} x_{\underline{3}} \sum_{\underline{4}=1}^{n_0} W_{\hat{4}\underline{4}}^{(1)} x_{\underline{4}}\right) \\
&= \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_0} \mathbb{E}\left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{2}\underline{2}}^{(1)} W_{\hat{3}\underline{3}}^{(1)} W_{\hat{4}\underline{4}}^{(1)}\right) x_{\underline{1}} x_{\underline{2}} x_{\underline{3}} x_{\underline{4}} \quad (3.18)
\end{aligned}$$

Aplicamos o Teorema de Wick para o termo com esperança, lembrando que $\mathbb{E}\left(W_{ij}^{(\ell)} W_{kl}^{(\ell)}\right) = \delta_{ik} \delta_{jl} \frac{C_W}{n_{\ell-1}}$.

$$\begin{aligned}
\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{2}\underline{2}}^{(1)} W_{\hat{3}\underline{3}}^{(1)} W_{\hat{4}\underline{4}}^{(1)} \right) = & \\
& \mathbb{E} \left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{2}\underline{2}}^{(1)} \right) \mathbb{E} \left(W_{\hat{3}\underline{3}}^{(1)} W_{\hat{4}\underline{4}}^{(1)} \right) + \\
& \mathbb{E} \left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{3}\underline{3}}^{(1)} \right) \mathbb{E} \left(W_{\hat{2}\underline{2}}^{(1)} W_{\hat{4}\underline{4}}^{(1)} \right) + \\
& \mathbb{E} \left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{4}\underline{4}}^{(1)} \right) \mathbb{E} \left(W_{\hat{2}\underline{2}}^{(1)} W_{\hat{3}\underline{3}}^{(1)} \right) = \\
& \delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \frac{C_W}{n_0} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} \frac{C_W}{n_0} + \\
& \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \frac{C_W}{n_0} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} \frac{C_W}{n_0} + \\
& \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \frac{C_W}{n_0} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}} \frac{C_W}{n_0} =
\end{aligned}$$

Agrupando os termos, obtemos

$$\begin{aligned}\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(1)} W_{\hat{2}\underline{2}}^{(1)} W_{\hat{3}\underline{3}}^{(1)} W_{\hat{4}\underline{4}}^{(1)} \right) &= \\ &= \frac{C_W^2}{n_0^2} (\delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} + \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} + \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}})\end{aligned}$$

Voltando para (3.18), obtemos

$$\mathbb{E}(z_{\hat{1}}^{(1)} z_{\hat{2}}^{(1)} z_{\hat{3}}^{(1)} z_{\hat{4}}^{(1)}) = \frac{C_W^2}{n_0^2} \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_0} (\delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} + \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} + \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}}) x_{\underline{1}} x_{\underline{2}} x_{\underline{3}} x_{\underline{4}}$$

Vamos nos atentar ao primeiro grupo de deltas, e os outros saem de maneira análoga.

Os índices ‘chapéu’ são fixos, então podemos retirar da soma

$$\sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_0} \delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} x_{\underline{1}} x_{\underline{2}} x_{\underline{3}} x_{\underline{4}} = \delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_0} \delta_{\underline{1}\underline{2}} \delta_{\underline{3}\underline{4}} x_{\underline{1}} x_{\underline{2}} x_{\underline{3}} x_{\underline{4}}$$

O primeiro delta só é diferente de zero quando $\underline{1} = \underline{2}$, assim fazemos a mudança de variáveis $\nu = \underline{1} = \underline{2}$. De modo análogo, $\mu = \underline{3} = \underline{4}$. Assim, obtemos

$$\begin{aligned} \delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} \sum_{\nu, \mu=1}^{n_0} \delta_{\nu\nu} \delta_{\mu\mu} x_{\nu} x_{\nu} x_{\mu} x_{\mu} &= \delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} \sum_{\nu=1}^{n_0} x_{\nu} x_{\nu} \sum_{\mu=1}^{n_0} x_{\mu} x_{\mu} \\ &= \delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} (x \cdot x)^2 \end{aligned}$$

Voltando para (3.18), obtemos

$$\begin{aligned}
\mathbb{E}(z_{\hat{1}}^{(1)} z_{\hat{2}}^{(1)} z_{\hat{3}}^{(1)} z_{\hat{4}}^{(1)}) &= \\
&= \frac{C_W^2}{n_0^2} (\delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} (x \cdot x)^2 + \delta_{\hat{1}\hat{3}} \delta_{\hat{2}\hat{4}} (x \cdot x)^2 + \delta_{\hat{1}\hat{4}} \delta_{\hat{2}\hat{3}} (x \cdot x)^2) \\
&= \frac{C_W^2}{n_0^2} (x \cdot x)^2 (\delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}} \delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}} \delta_{\hat{2}\hat{3}}) \\
&= C_W^2 \left(\frac{x \cdot x}{n_0} \right)^2 (\delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}} \delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}} \delta_{\hat{2}\hat{3}}) \\
&= C_W^2 \left(G_2^{(0)} \right)^2 (\delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}} \delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}} \delta_{\hat{2}\hat{3}}) \quad (3.18)
\end{aligned}$$

- O mesmo raciocínio pode ser aplicado para a camada ℓ , e obtemos

$$\begin{aligned}
 \mathbb{E}(z_{\hat{1}}^{(\ell+1)} z_{\hat{2}}^{(\ell+1)} z_{\hat{3}}^{(\ell+1)} z_{\hat{4}}^{(\ell+1)}) &= \\
 \mathbb{E}\left(W_{\hat{1}*}^{(\ell+1)} z_{\hat{1}}^{(\ell)} W_{\hat{2}*}^{(\ell+1)} z_{\hat{2}}^{(\ell)} W_{\hat{3}*}^{(\ell+1)} z_{\hat{3}}^{(\ell)} W_{\hat{4}*}^{(\ell+1)} z_{\hat{4}}^{(\ell)}\right) &= \\
 \mathbb{E}\left(\sum_{\underline{1}=1}^{n_\ell} W_{\hat{1}\underline{1}}^{(\ell+1)} z_{\underline{1}}^{(\ell)} \sum_{\underline{2}=1}^{n_\ell} W_{\hat{2}\underline{2}}^{(\ell+1)} z_{\underline{2}}^{(\ell)} \sum_{\underline{3}=1}^{n_\ell} W_{\hat{3}\underline{3}}^{(\ell+1)} z_{\underline{3}}^{(\ell)} \sum_{\underline{4}=1}^{n_\ell} W_{\hat{4}\underline{4}}^{(\ell+1)} z_{\underline{4}}^{(\ell)}\right) &= \\
 \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \mathbb{E}\left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)}\right) & \quad (3.20)
 \end{aligned}$$

O que acontece na camada $\ell + 1$ é independente do que acontece na camada ℓ , então a esperança do produto é o produto das esperanças.

$$\begin{aligned}
 \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)} \right) = \\
 \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} \right) \mathbb{E} \left(z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)} \right)
 \end{aligned} \tag{3.20}$$

Usando o Teorema de Wick, e fazendo o mesmo cálculo que fizemos para a camada 1, obtemos

$$\begin{aligned}
\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} \right) = \\
\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{2}\underline{2}}^{(\ell+1)} \right) \mathbb{E} \left(W_{\hat{3}\underline{3}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} \right) + \\
\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} \right) \mathbb{E} \left(W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} \right) + \\
\mathbb{E} \left(W_{\hat{1}\underline{1}}^{(\ell+1)} W_{\hat{4}\underline{4}}^{(\ell+1)} \right) \mathbb{E} \left(W_{\hat{2}\underline{2}}^{(\ell+1)} W_{\hat{3}\underline{3}}^{(\ell+1)} \right) = \\
\frac{C_W^2}{n_\ell^2} (\delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} + \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} + \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}})
\end{aligned}$$

Voltando para (3.20), obtemos

$$\begin{aligned}
& \mathbb{E}\left(z_{\hat{1}}^{(\ell+1)} z_{\hat{2}}^{(\ell+1)} z_{\hat{3}}^{(\ell+1)} z_{\hat{4}}^{(\ell+1)}\right) = \\
& = \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \frac{C_W^2}{n_\ell^2} \left(\begin{array}{c} \delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} \\ + \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} \\ + \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}} \end{array} \right) \mathbb{E}\left(z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)}\right)
\end{aligned} \tag{3.20}$$

Vamos nos concentrar no primeiro grupo de deltas

$$\begin{aligned}
\sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \delta_{\widehat{1}\widehat{2}} \delta_{\underline{1}\underline{2}} \delta_{\widehat{3}\widehat{4}} \delta_{\underline{3}\underline{4}} \mathbb{E} \left(z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)} \right) = \\
= \delta_{\widehat{1}\widehat{2}} \delta_{\widehat{3}\widehat{4}} \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \delta_{\underline{1}\underline{2}} \delta_{\underline{3}\underline{4}} \mathbb{E} \left(z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)} \right)
\end{aligned}$$

Novamente, fazendo $\nu = \underline{1} = \underline{2}$, $\mu = \underline{3} = \underline{4}$, temos

$$\begin{aligned}
\delta_{\widehat{1}\widehat{2}} \delta_{\widehat{3}\widehat{4}} \sum_{\nu, \mu=1}^{n_\ell} \delta_{\nu\nu} \delta_{\mu\mu} \mathbb{E} \left(z_{\nu}^{(\ell)} z_{\nu}^{(\ell)} z_{\mu}^{(\ell)} z_{\mu}^{(\ell)} \right) = \\
= \delta_{\widehat{1}\widehat{2}} \delta_{\widehat{3}\widehat{4}} \sum_{\nu, \mu=1}^{n_\ell} \mathbb{E} \left(z_{\nu}^{(\ell)} z_{\nu}^{(\ell)} z_{\mu}^{(\ell)} z_{\mu}^{(\ell)} \right)
\end{aligned}$$

Aplicando a ideia acima para todos os os grupos de deltas, obtemos

$$\begin{aligned}
& \mathbb{E}\left(z_{\hat{1}}^{(\ell+1)} z_{\hat{2}}^{(\ell+1)} z_{\hat{3}}^{(\ell+1)} z_{\hat{4}}^{(\ell+1)}\right) = \\
& \sum_{\underline{1}, \underline{2}, \underline{3}, \underline{4}=1}^{n_\ell} \frac{C_W^2}{n_\ell^2} \left(\begin{array}{c} \delta_{\hat{1}\hat{2}} \delta_{\underline{1}\underline{2}} \delta_{\hat{3}\hat{4}} \delta_{\underline{3}\underline{4}} \\ + \delta_{\hat{1}\hat{3}} \delta_{\underline{1}\underline{3}} \delta_{\hat{2}\hat{4}} \delta_{\underline{2}\underline{4}} \\ + \delta_{\hat{1}\hat{4}} \delta_{\underline{1}\underline{4}} \delta_{\hat{2}\hat{3}} \delta_{\underline{2}\underline{3}} \end{array} \right) \mathbb{E}\left(z_{\underline{1}}^{(\ell)} z_{\underline{2}}^{(\ell)} z_{\underline{3}}^{(\ell)} z_{\underline{4}}^{(\ell)}\right) = \\
& = \frac{C_W^2}{n_\ell^2} \left(\delta_{\hat{1}\hat{2}} \delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}} \delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}} \delta_{\hat{2}\hat{3}} \right) \sum_{\nu, \mu=1}^{n_\ell} \mathbb{E}\left(z_\nu^{(\ell)} z_\nu^{(\ell)} z_\mu^{(\ell)} z_\mu^{(\ell)}\right) \\
& \hspace{25em} (3.20)
\end{aligned}$$

- Novamente, podemos argumentar que o correlator de 4 pontos é proporcional ao fator

$$\delta_{\hat{1}\hat{2}}\delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}}\delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}}\delta_{\hat{2}\hat{3}}$$

que chamaremos de *fator Wick 4*.

- Chamando a constante de proporcionalidade de $G_4^{(\ell)}$, escrevemos a relação

$$\mathbb{E} \left(z_{\hat{1}}^{(\ell)} z_{\hat{2}}^{(\ell)} z_{\hat{3}}^{(\ell)} z_{\hat{4}}^{(\ell)} \right) = (\delta_{\hat{1}\hat{2}}\delta_{\hat{3}\hat{4}} + \delta_{\hat{1}\hat{3}}\delta_{\hat{2}\hat{4}} + \delta_{\hat{1}\hat{4}}\delta_{\hat{2}\hat{3}}) G_4^{(\ell)} \quad (3.21)$$

Comparando (3.21) com (3.18), obtemos a relação

$$G_4^{(1)} = C_W^2 \left(G_2^{(0)} \right)^2 \quad (3.22)$$

Aplicando (3.21) no somatório que aparece (3.20), obtemos

$$\begin{aligned}\sum_{\nu,\mu=1}^{n_\ell} \mathbb{E} \left(z_\nu^{(\ell)} z_\nu^{(\ell)} z_\mu^{(\ell)} z_\mu^{(\ell)} \right) &= \sum_{\nu,\mu=1}^{n_\ell} (\delta_{\nu\nu} \delta_{\mu\mu} + \delta_{\nu\mu} \delta_{\nu\mu} + \delta_{\nu\mu} \delta_{\nu\mu}) G_4^{(\ell)} = \\ &= \sum_{\nu,\mu=1}^{n_\ell} (\delta_{\nu\nu} \delta_{\mu\mu} + 2\delta_{\nu\mu} \delta_{\nu\mu}) G_4^{(\ell)} \quad (3.23)\end{aligned}$$

- ▶ O primeiro par de deltas é sempre 1, então essa primeira parte da soma é n_ℓ^2 .
- ▶ O segundo par de deltas só é diferente de zero quando $\nu = \mu$, então essa segunda parte da soma é n_ℓ .
- ▶ Portanto, a soma total é $(n_\ell^2 + 2n_\ell) G_4^{(\ell)}$.

Voltando para (3.20),

$$\begin{aligned}
& \mathbb{E}\left(z_{\widehat{1}}^{(\ell+1)} z_{\widehat{2}}^{(\ell+1)} z_{\widehat{3}}^{(\ell+1)} z_{\widehat{4}}^{(\ell+1)}\right) = \\
& = \frac{C_W^2}{n_\ell^2} \left(\delta_{\widehat{1}\widehat{2}}\delta_{\widehat{3}\widehat{4}} + \delta_{\widehat{1}\widehat{3}}\delta_{\widehat{2}\widehat{4}} + \delta_{\widehat{1}\widehat{4}}\delta_{\widehat{2}\widehat{3}}\right) \sum_{\nu,\mu=1}^{n_\ell} \mathbb{E}\left(z_\nu^{(\ell)} z_\nu^{(\ell)} z_\mu^{(\ell)} z_\mu^{(\ell)}\right) = \\
& = \frac{C_W^2}{n_\ell^2} \left(\delta_{\widehat{1}\widehat{2}}\delta_{\widehat{3}\widehat{4}} + \delta_{\widehat{1}\widehat{3}}\delta_{\widehat{2}\widehat{4}} + \delta_{\widehat{1}\widehat{4}}\delta_{\widehat{2}\widehat{3}}\right) (n_\ell^2 + 2n_\ell) G_4^{(\ell)} = \\
& = \left(\delta_{\widehat{1}\widehat{2}}\delta_{\widehat{3}\widehat{4}} + \delta_{\widehat{1}\widehat{3}}\delta_{\widehat{2}\widehat{4}} + \delta_{\widehat{1}\widehat{4}}\delta_{\widehat{2}\widehat{3}}\right) \left(1 + \frac{2}{n_\ell}\right) C_W^2 G_4^{(\ell)}
\end{aligned}$$

Por outro lado, definimos que o correlator de 4 pontos na camada $\ell + 1$ é dado pelo fator Wick 4 multiplicado pela constante de proporcionalidade $G_4^{(\ell+1)}$. Assim, podemos escrever a recorrência

$$G_4^{(\ell+1)} = C_W^2 \left(1 + \frac{2}{n_\ell} \right) G_4^{(\ell)} \quad (3.24)$$

Se abrirmos essa recursão da camada ℓ até a camada 1, e aplicamos (3.22), obtemos

$$\begin{aligned}
 G_4^{(\ell)} &= \left[\prod_{\widehat{\ell}=1}^{\ell-1} C_W^2 \left(1 + \frac{2}{n_\ell} \right) \right] G_4^{(1)} = (C_W^2)^{\ell-1} G_4^{(1)} \prod_{\widehat{\ell}=1}^{\ell-1} \left(1 + \frac{2}{n_\ell} \right) \\
 &= (C_W^2)^{\ell-1} C_W^2 \left(G_2^{(0)} \right)^2 \prod_{\widehat{\ell}=1}^{\ell-1} \left(1 + \frac{2}{n_\ell} \right) = \\
 &= \left(C_W^\ell G_2^{(0)} \right)^2 \prod_{\widehat{\ell}=1}^{\ell-1} \left(1 + \frac{2}{n_\ell} \right)
 \end{aligned}$$

Aplicando (3.15), o fator $C_W^\ell G_2^{(0)} = G_2^{(\ell)}$, e obtemos

$$G_4^{(\ell)} = \left(G_2^{(\ell)}\right)^2 \prod_{\widehat{\ell}=1}^{\ell-1} \left(1 + \frac{2}{n_\ell}\right) \quad (3.25)$$

que relaciona o correlator de 4 pontos com o correlator de 2 pontos.