Title: Lecture 2: Universality Classes of Nonlinear Networks

Speakers: Yonatan Kahn

Collection/Series: Special Seminars

Subject: Other

Date: October 10, 2025 - 5:00 PM

URL: https://pirsa.org/25100144

Universality classes of nonlinear networks

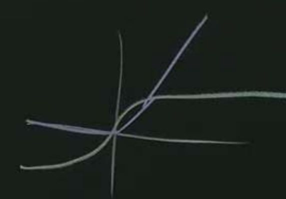
$$z_i^{(1)} = b_i^{(1)} + W_{ij}^{(1)} x_j, \quad z_i^{(\ell+1)} = b_i^{(\ell+1)} + W_{ij}^{(\ell+1)} \sigma(z_j^{(\ell)}) \qquad (MLP)$$

$$\mathbb{E}[W_{ij}] = 0, \quad \mathbb{E}[W_{i_1 j_1} W_{i_2 j_2}] = \frac{C_w}{n} \delta_{i_1 i_2} \delta_{j_1 j_2}$$

$$\mathbb{E}[b_i] = 0, \quad \mathbb{E}[b_i b_j] = C_b \delta_{ij}$$

Two common examples for $\sigma(z)$:

- ReLU $\sigma(z) = \max(0, z)$
- $\sigma(z) = \tanh(z)$



$-p)$

Start by computing 2-pt function at layer 1

$$\mathbb{E}[z^{(1)}_{i_1\alpha_1} z^{(1)}_{i_2\alpha_2}] = \mathbb{E}[(b^{(1)}_{i_1} + W^{(1)}_{i_1 j_1} x_{j_1\alpha_1})(b^{(1)}_{i_2} + W^{(1)}_{i_2 j_2} x_{j_2\alpha_2})]$$

$$= \underbrace{\left(C_b + \frac{C_W}{n} \vec{x}_{\alpha_1} \cdot \vec{x}_{\alpha_2}\right)}_{G^{(1)}_{\alpha_1\alpha_2}} \delta_{i_1 i_2}$$

$$\mathbb{E}[z^{(1)} z^{(1)} z^{(1)} z^{(1)}]_{\text{conn.}} = 0 + O\left(\frac{1}{n}\right)$$

$w^{(1)}_{i_2 j_2} x_{j_2 \alpha_2})\Big]$

$$p(z^{(1)} | \mathcal{D}) = \frac{1}{[\det(2\pi G^{(1)})]^{n/2}} \exp\left(-\frac{1}{2} G^{\alpha_1 \alpha_2}_{(1)} \vec{z}^{(1)}_{\alpha_1} \cdot \vec{z}_{\alpha_2}\right)$$

$$\left(\text{to } \mathcal{O}\left(\tfrac{1}{n}\right)\right)$$

$$p(z^{(\ell+1)} | z^{(\ell)}) = \frac{1}{\sqrt{\det(2\pi \hat{G}^{(\ell+1)})^{n}}} \exp\left(-\frac{1}{2} \hat{G}^{\alpha_1 \alpha_2}_{(\ell+1)} \vec{z}^{(\ell+1)}_{\alpha_1} \cdot \vec{z}^{(\ell+1)}_{\alpha_2}\right)$$

$$C_b + \frac{C_W}{n} \vec{\sigma}^{(\ell)}_{\alpha_1} \cdot \vec{\sigma}^{(\ell)}_{\alpha_2}$$

$$\sigma(z^{(\ell)}_{\alpha})$$

Universality classes of nonlinear networks

$$z_i^{(1)} = b_i^{(1)} + W_{ij}^{(1)} x_j, \quad z_i^{(\ell+1)} = b_i^{(\ell+1)} + W_{ij}^{(\ell+1)} \sigma(z_j^{(\ell)}) \qquad \text{(MLP)}$$

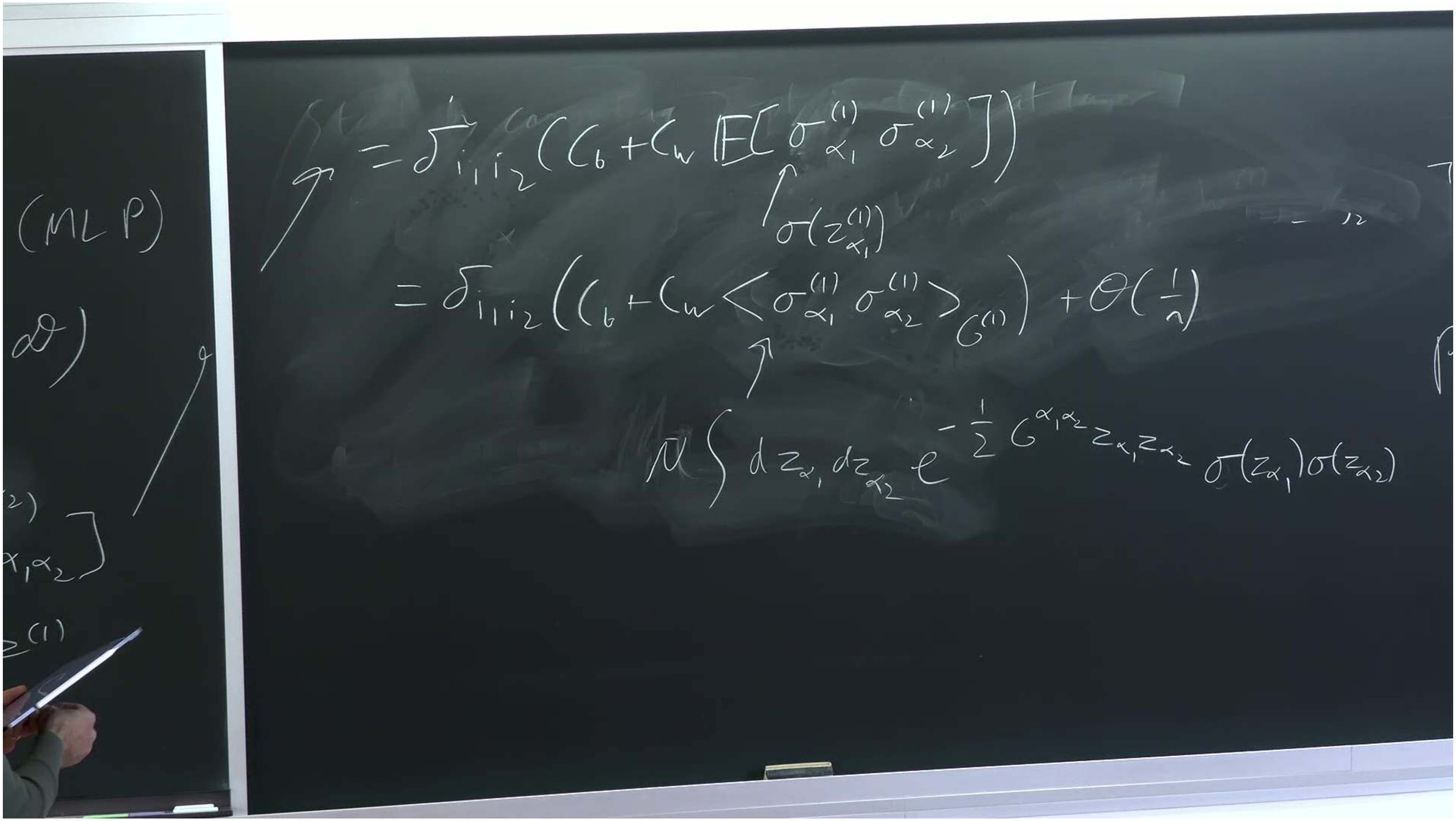
$$p(z^{(\ell+1)} | \mathcal{D}) = \int \prod_i dz_i^{(\ell)} \, p(z^{(\ell+1)} | z^{(\ell)}) \, p(z^{(\ell)} | \mathcal{D})$$

Build up from its moments

Start with $\ell = 2$: $\mathbb{E}[z^{(2)}_{i_1\alpha_1} z^{(2)}_{i_2\alpha_2}] = \delta_{i_1 i_2} \mathbb{E}[\hat{G}^{(2)}_{\alpha_1\alpha_2}]$

Over $z^{(1)}$ and $z^{(2)}$

Only over $z^{(1)}$


(MLP)
$= \delta_{i_1 i_2}\left(C_b + C_w \mathbb{E}\left[\sigma^{(1)}_{\alpha_1} \sigma^{(1)}_{\alpha_2}\right]\right)$
$\sigma(z^{(1)}_{\alpha_1})$
$= \delta_{i_1 i_2}\left(\left(C_b + C_w \langle \sigma^{(1)}_{\alpha_1} \sigma^{(1)}_{\alpha_2} \rangle_{G^{(1)}}\right) + O\left(\frac{1}{n}\right)\right)$
$N \int dz_{\alpha_1} dz_{\alpha_2}\, e^{-\frac{1}{2} G^{\alpha_1 \alpha_2} z_{\alpha_1} z_{\alpha_2}} \sigma(z_{\alpha_1}) \sigma(z_{\alpha_2})$

$p(f|\mathcal{D}\theta)$

Let $G^{(\ell)}_{\alpha_1\alpha_2} = \mathbb{E}[\hat{G}^{(\ell)}_{\alpha_1\alpha_2}]$

ansatz $\mathbb{E}[z^{(\ell)}_{i_1\alpha_1} z^{(\ell)}_{i_2\alpha_2}] = \delta_{i_1 i_2} G^{(\ell)}_{\alpha_1\alpha_2}$

$$\Rightarrow \boxed{G^{(\ell+1)}_{\alpha_1\alpha_2} = C_b + C_W \langle \sigma^{(\ell)}_{\alpha_1} \sigma^{(\ell)}_{\alpha_2} \rangle_{G^{(\ell)}} + \mathcal{O}\left(\frac{1}{n}\right)}$$

$$= \delta_{i_1 i_2}\left(C_b + C_w\, \mathbb{E}\left[\sigma^{(1)}_{\alpha_1}\sigma^{(1)}_{\alpha_2}\right]\right)$$

$$\sigma(z^{(1)}_{\alpha_1})$$

$$= \delta_{i_1 i_2}\left(\left(C_b + C_w \langle \sigma^{(1)}_{\alpha_1}\sigma^{(1)}_{\alpha_2}\rangle_{G^{(1)}}\right) + \mathcal{O}\left(\tfrac{1}{n}\right)\right)$$

let $\lim_{n\to\infty} G = K \implies K^{(\ell+1)}_{\alpha\beta} = C_b + C_w \langle \sigma_\alpha \sigma_\beta \rangle_{K^{(\ell)}}$

$\sigma(z) = z^2 \implies$

$$= C_b + C_w\left(K^{(\ell)}_{\alpha\alpha}K^{(\ell)}_{\beta\beta} + 2\left(K^{(\ell)}_{\alpha\beta}\right)^2\right)$$

Let G

ansatz $\mathbb{E}$

$\implies \boxed{G^{(\ell+}_{\alpha_1}}$

Let

$(\frac{1}{n})$

$C_b + C_w \langle \sigma_\alpha \sigma_\beta \rangle_{K^{(\ell)}}$ +

$C_b + C_w \left(K^{(\ell)}_{\alpha\alpha} K^{(\ell)}_{\beta\beta} + 2 (K^{(\ell)}_{\alpha\beta})^2 \right)$

Diagonal component, $\alpha = 0$

$$K_{00}^{(\ell+1)} = C_b + C_w \langle \sigma^2 \rangle_{K^{(\ell)}_{00}}$$

$$= C_b + C_w \underbrace{\left[\frac{1}{\sqrt{2\pi K^{(\ell)}}} \int_{-\infty}^{\infty} dz \, e^{-\frac{z^2}{2K^{(\ell)}}} \sigma(z)^2 \right]}_{g(K)}$$

Find fixed point by linearizing:

$$K_{00}^{(\ell)} = K^*_{00} + \Delta K_{00}^{(\ell)}$$

$\int dz\, e^{-\frac{z^2}{2K^{(\ell)}}} \sigma(z)^2 \Big]$

(K)

izing.

To first order in Δ,

$$\Delta K_{00}^{(\ell+1)} = \chi_{\parallel}(K_{00}^{*}) \times \Delta K_{00}^{(\ell)}$$

$\sigma(z) = z$

$g'(K) = 1$

$$C_W\, g'(K) = \frac{C_W}{2K^2} \langle \sigma(z)\sigma(z)(z^2 - K) \rangle_K$$

Stay at the fixed pt $\Rightarrow \chi_{\parallel}(K_{00}^{*}) = 1$

Let

Fixed pt for K_{00} ($\Leftrightarrow \chi_{\parallel} = 1$) ensures norms of preactivations don't change exponentially w/ depth

Second criticality condition on $K_{\alpha\beta}$

$$\Rightarrow \quad C_W \underbrace{\langle \sigma'(z)^2 \rangle_{K^*}}_{\chi_\perp(K^*)} = 1 \quad \Rightarrow \quad \left[\frac{2K^2 \langle \sigma'(z) \rangle_K}{\langle \sigma(z)^2 (z^2 - K) \rangle_K}\right]_{K=K^*} = 1$$

$$C_W = \frac{1}{\langle \sigma'(z)^2 \rangle_{K^*}}, \quad C_b = K^* - \frac{\langle \sigma(z)^2 \rangle_{K^*}}{\langle \sigma'(z)^2 \rangle_{K^*}}$$

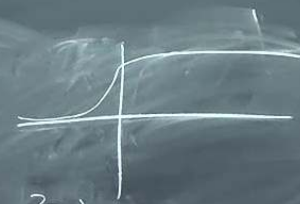
Let

th

$$\left[\frac{\ldots(z)\rangle_K}{\ldots(K)\rangle_K}\right]_{K=K^*} = 1$$

$$\ldots\rangle_{K^*}, \quad C_b = K^* - \frac{\langle \sigma(z)^2 \rangle_{K^*}}{\langle \sigma'(z)^2 \rangle_{K^*}}$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



$$\Rightarrow C_b = -\left(\frac{\sigma(0)^2}{\sigma'(0)^2}\right) < 0$$

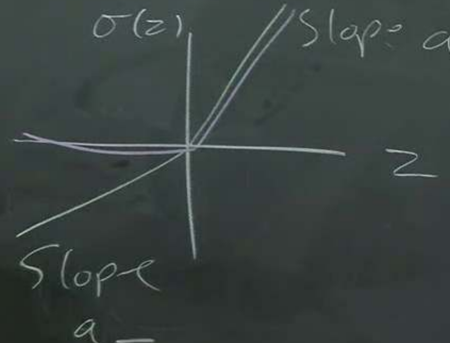
$$\sigma(0) \neq 0$$

If $\sigma(z)$ is smooth, want $\sigma(0) = 0$

Universality classes of nonlinear networks

Two universality classes depending on K^*:

- Scale-invariant activations (piecewise linear w/ $\sigma(0)=0$)



$$K^* = \frac{2}{a_+^2 + a_-^2}\left[\frac{1}{n}\vec{x}\cdot\vec{x}\right]$$

$$C_b = 0,\quad C_W = \frac{2}{a_+^2 + a_-^2} \quad (\text{ReLU: } C_W = 2)$$

- Smooth activations w/ $\sigma(0)=0$

$(K^* = 0)$

$$\sigma(z) = \sum_{p=0}^{\infty} \frac{\sigma_p}{p!} z^p$$

$$\langle \sigma(z)^2 \rangle_K = \sigma_0^2 + (\sigma_1^2 + 2\sigma_0\sigma_2) K + O(K^2)$$

$K^* = C_b + C_W \langle \sigma^2 \rangle_{K^*}$ has a sol'n

$K^* = 0$ iff $\sigma_0 = 0$, and $C_b = 0$

$$K^* - \frac{\langle \sigma(z)^2 \rangle_{K^*}}{\langle \sigma'(z)^2 \rangle_{K^*}}$$

To first order in Δ,

$$\Delta K_{00}^{(\ell+1)} = \chi_{\parallel} (K$$

$C_W\, g'(K) =$

Stay at the fixed pt

$) = 0$

$)K + \mathcal{O}(K^2)$

sol'n

$_6 = 0$

Linearize $\chi_\perp$ and $\chi_\parallel$ about $z=0$, expand in K:

$$\Rightarrow C_W = \frac{1}{\sigma_1^2} = \frac{1}{\sigma'(0)^2}$$

$$\Rightarrow K_{00}^{(\ell)} = \frac{1}{2\ell} + \mathcal{O}\left(\frac{\log \ell}{\ell^2}\right) + \mathcal{O}\left(\frac{1}{\ell^2}\right)$$

$$K_{\alpha\alpha}^{(1)} = \frac{1}{\sigma'(0)^2} \times \frac{\vec{x}_\alpha \cdot \vec{x}_\alpha}{n}$$

Let

$= 1$

$K = K^*$

$C_b = K^* - \frac{\langle \sigma(z)^2 \rangle_{K^*}}{\langle \sigma'(z)^2 \rangle_{K^*}}$

If instead we take $C_W^{(1)} = 1$,

$$C_W^{(\ell)} = \frac{1}{\langle \sigma'(z)^2 \rangle}$$

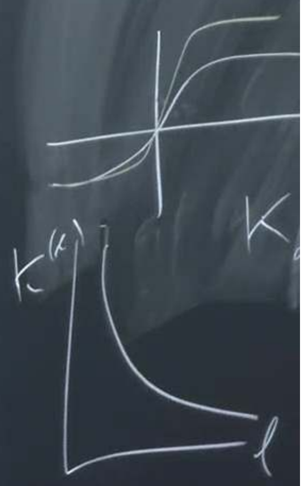
normalize data to $\frac{1}{n} \vec{x}_\alpha \cdot \vec{x}_\alpha = 1$

$$\Rightarrow K^{(\ell)}_{\alpha\alpha} = 1, \quad K^{(\ell)}_{\alpha\beta} \sim \exp(-\Omega \ell)$$

(2510.06527)

Linearize

$\Rightarrow C_W$



Universality classes of nonlinear networks

Can also derive recursions for 4-pt correlator.

non-Gaussianities from $\mathbb{E}[(\hat{G}-G)^2]$ and $\det(2\pi\hat{G})$

$$\mathbb{E}\left[z^{(\ell)}_{i_1} z^{(\ell)}_{i_2} z^{(\ell)}_{i_3} z^{(\ell)}_{i_4}\right]_{\text{conn.}} \sim \frac{V^{(\ell)}}{n}\left(\delta_{i_1 i_2}\delta_{i_3 i_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} + \delta_{i_1 i_4}\delta_{i_2 i_3}\right)$$

$$\Rightarrow \quad V^{(\ell+1)} = \chi_{\parallel}^2(K^{(\ell)})V^{(\ell)} + C_W^2\left[\langle\sigma^4\rangle_{K^{(\ell)}} - \langle\sigma^2\rangle^2_{K^{(\ell)}}\right]$$

Facts:

· tuning C_b and C_W to fix $K^{(\ell)}$

also cures exponential behavior in $V^{(\ell)}$

— Scale-inv't $\quad V^{(\ell)} = (\ell-1)(\#)(K^{*})^2$

— $K^{*} = 0$: $\quad V^{(\ell)} = (\#)\left(\frac{1}{\ell}\right)$

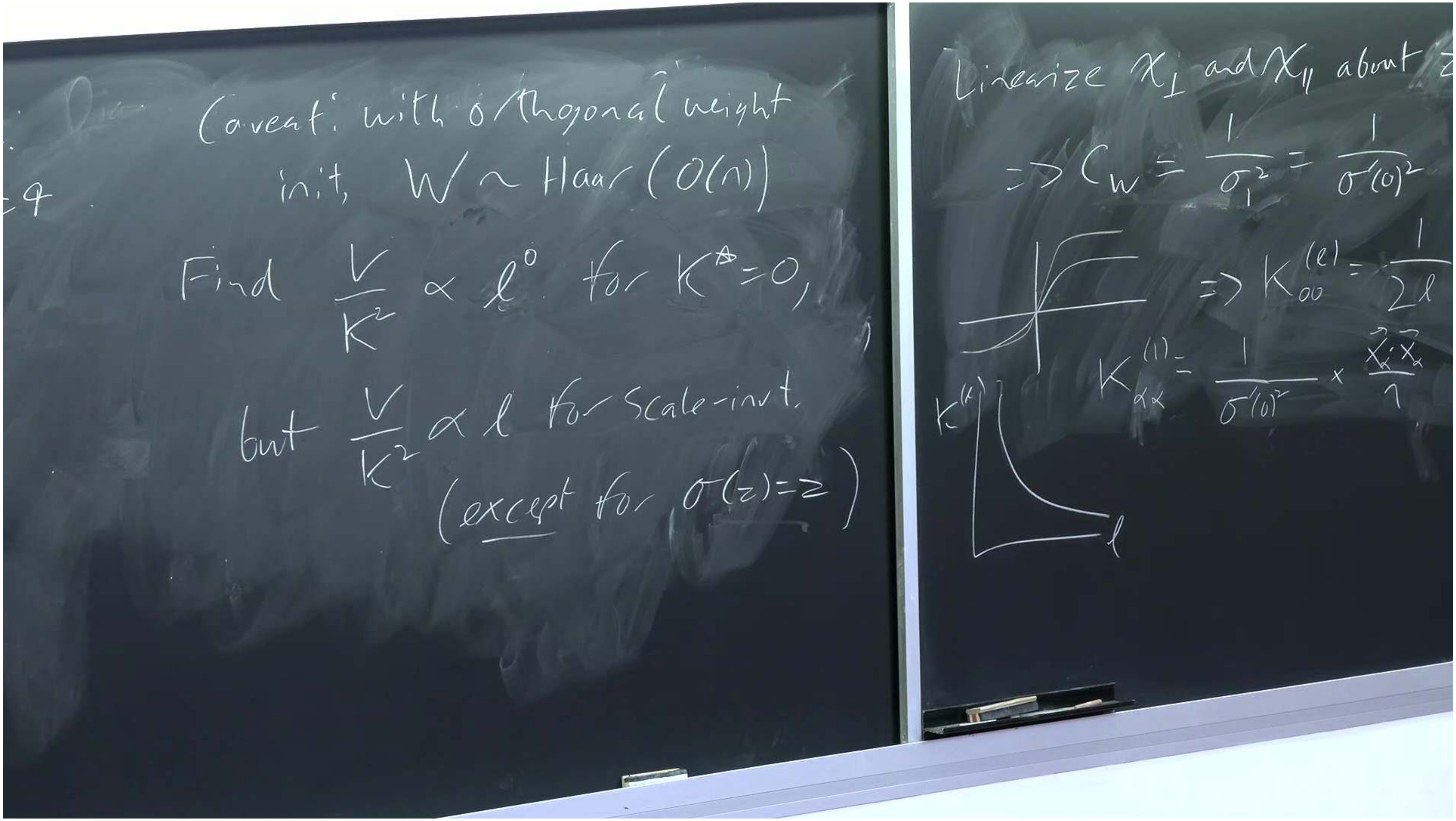
Assign a power-counting scheme:

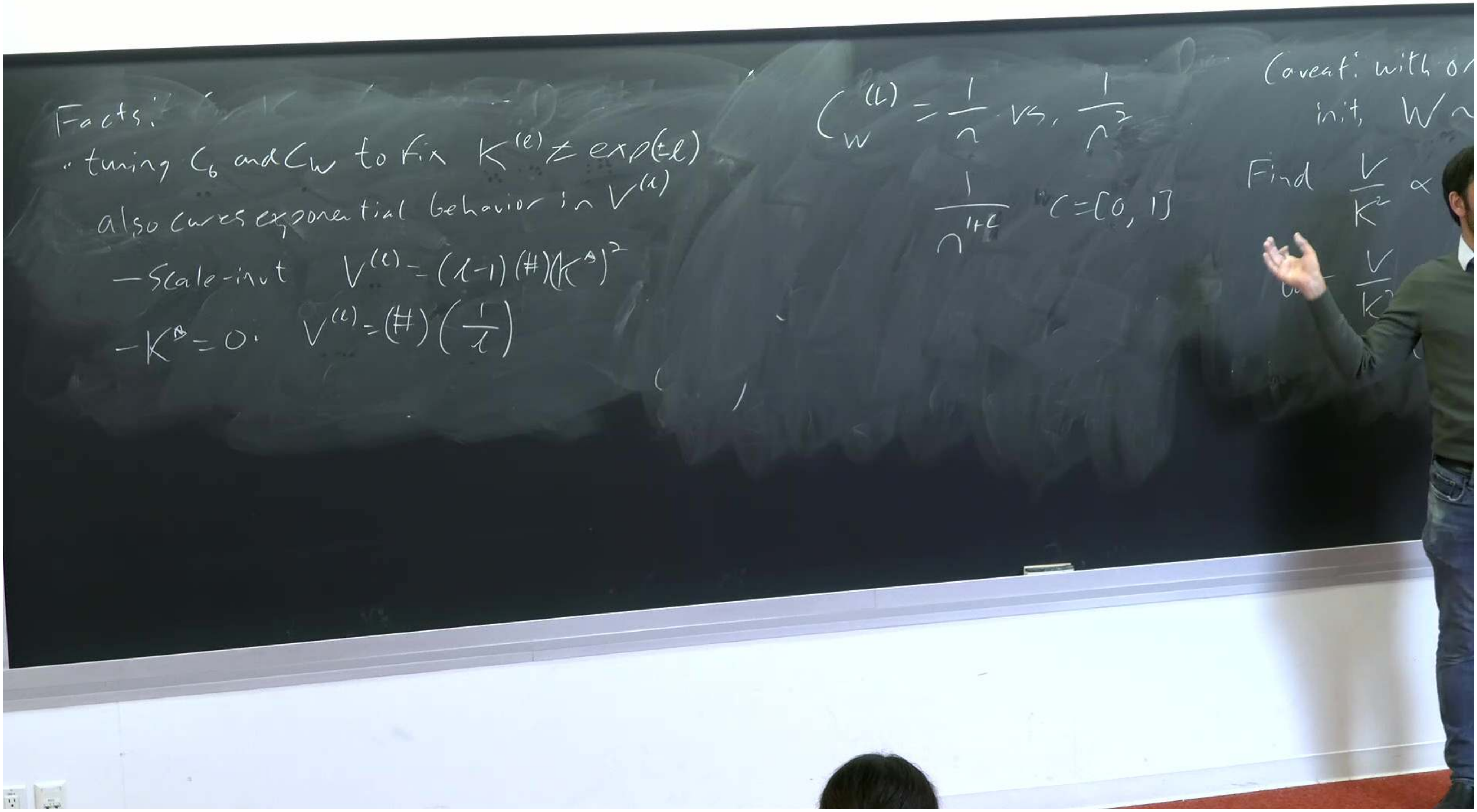
$[z] = 1 \Rightarrow [K] = 2,\ [V] = 4$ $\quad \ell =$

dimensionless 4-pt function is

$$\underbrace{\frac{V}{K^2}}_{O\left(\frac{1}{n}\right)} \sim \frac{\ell}{n} \text{ for } \underline{\text{both}}$$

universality classes


Caveat: with orthogonal weight
init, $W \sim \mathrm{Haar}(O(n))$
Find $\frac{V}{K^2} \propto \ell^0$ for $K^* = 0$,
but $\frac{V}{K^2} \propto \ell$ for scale-inv.
(except for $\sigma(z) = z$)
Linearize $\chi_\perp$ and $\chi_\parallel$ about
$\Rightarrow C_W = \frac{1}{\sigma_1^2} = \frac{1}{\sigma'(0)^2}$
$\Rightarrow K_{00}^{(\ell)} = \frac{1}{2\ell}$
$K_{\alpha\alpha}^{(1)} = \frac{1}{\sigma'(0)^2} \times \frac{\vec{x}_\alpha \cdot \vec{x}_\alpha}{n}$
$K^{(\ell)}$
ℓ


Facts:
· tuning C_b and C_w to fix $K^{(\ell)} \neq \exp(\pm\ell)$
also cures exponential behavior in $V^{(\ell)}$
— Scale-inv't $V^{(\ell)} = (\ell-1)(\#)(K^*)^2$
— $K^* = 0$: $V^{(\ell)} = (\#)\left(\frac{1}{\ell}\right)$
$C_w^{(\ell)} = \frac{1}{n}$ vs. $\frac{1}{n^2}$
$\frac{1}{n^{1+\epsilon}}$ $\epsilon = [0, 1]$
Caveat: with o
init, W ~
Find $\frac{V}{K^2} \propto$
$\frac{V}{K^2}$