Injectivity of ReLU networks

Perspectives from integral geometry and statistical physics

arXiv:2302.14112

Antoine Maillard

A. S. Bandeira, D. Belius, I. Dokmanic, S. Nakajima

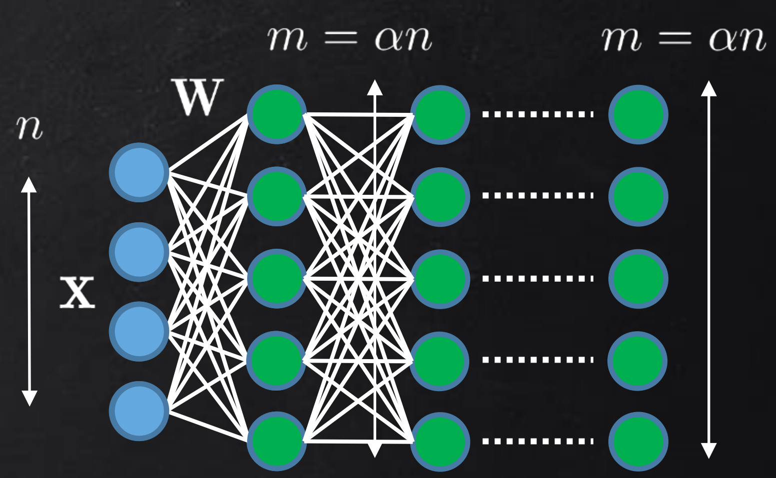
ETH zürich

High Dimensional Statistics and Random Matrices – June 15th 2023

Injectivity of ReLU networks

$\mathrm{ReLU}(u) = \max(0, u)$

When is a ReLU neural network an injective function ?



$\varphi_{\mathbf{W}}(\mathbf{x})_\mu = \mathrm{ReLU}(\mathbf{W}_\mu \cdot \mathbf{x})$

- General weights: $\alpha = 2$ is necessary and sufficient [Puthawala &al '22].
- Injectivity for "generic networks": random weights $\mathbf{W}_\mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n)$.

Layerwise injectivity

In the limit $m, n \to \infty$ with $\alpha = m/n = \Theta(1)$, let $p_{m,n} \equiv \mathbb{P}[\varphi_{\mathbf{W}} \text{ is injective}]$. Thresholds ?

$$\alpha_{\text{inj}}^- \equiv \sup\{\alpha : \lim_{n\to\infty} p_{m,n} = 0\} \leq \alpha_{\text{inj}}^+ \equiv \inf\{\alpha : \lim_{n\to\infty} p_{m,n} = 1\}$$

Sharp transition ?

Injectivity and random geometry

$\varphi_{\mathbf{W}}(\mathbf{x})_\mu = \mathrm{ReLU}(\mathbf{W}_\mu \cdot \mathbf{x})$

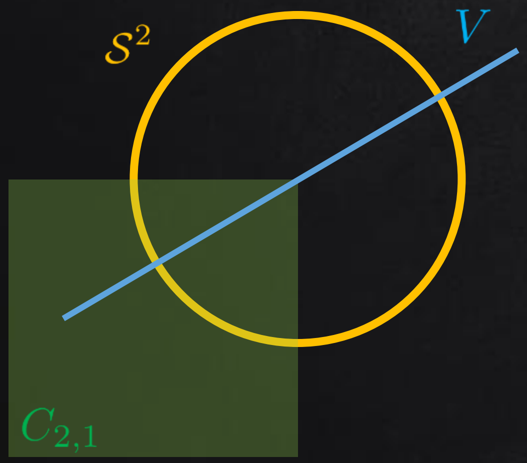
<u>Idea:</u> Quotient $\mathcal{R} \equiv \mathbb{R}^n / \sim$, with $\mathbf{x} \sim \mathbf{y} \Leftrightarrow \mathbf{W}\mathbf{x}$ and $\mathbf{W}\mathbf{y}$ are positive on the same set of coordinates.

→ $\varphi_{\mathbf{W}}$ is linear on each $R \in \mathcal{R}$

$\varphi_{\mathbf{W}}$ is injective ⟺ there is no $R \in \mathcal{R}$ with less than n positive coordinates. a.s.

→ **Theorem** : $p_{m,n} \equiv \mathbb{P}[\varphi_{\mathbf{W}} \text{ is injective}] = \mathbb{P}[V \cap C_{m,n} = \{0\}] = \mathbb{P}[V \cap C_{m,n} \cap \mathcal{S}^{m-1} = \emptyset]$

- $V = \mathbf{W}\mathbb{R}^n$ is a random n-dimensional subspace of $\mathbb{R}^m$.
- $C_{m,n}$ is the (nonconvex) cone of vectors in $\mathbb{R}^m$ with <u>less than n positive coordinates</u>.



$p_{2,1} = 1/4$

First bounds

$m/n \to \alpha$

- Cover's theorem [Cover '65]: $p_{m,n} \to 0$ if $\alpha < 3$.
- Union bound over the orthants in $C_{m,n}$: $p_{m,n} \to 1$ if $\alpha > 9.09$.
- Explicit certificate for non-injectivity: $p_{m,n} \to 0$ if $\alpha \lesssim 3.3$.

[Puthawala&al '22, Paleka '21]

Approach from integral geometry

Details in [Paleka '21, Clum '22]

V : random n-dimensional subspace of $\mathbb{R}^m$.

Kinematic Crofton formula [Amelunxen &al '13]: estimate $\mathbb{P}[V \cap C \neq \{0\}]$ for C a convex cone

Spherical cinematic formulas [Schneider & Weil '08]

i. C is a finite union of convex cones ✓

ii. $F(A \cup B) + F(A \cap B) = F(A) + F(B)$

→ $\mathbb{E}[F(V \cap C)]$

⚠ $C = C_{m,n}$ not convex

$$p_{m,n} = \mathbb{P}[V \cap C_{m,n} \cap \mathcal{S}^{m-1} = \emptyset] = 1 - \mathbb{E}[\mathbb{1}^S(V \cap C_{m,n})]$$

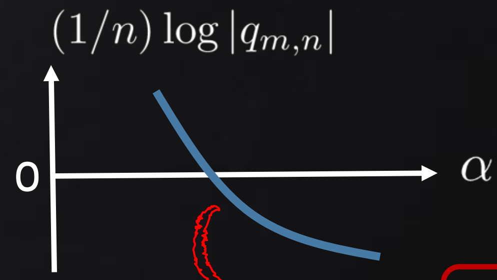
$$\mathbb{1}^S(A) \equiv \mathbb{1}\{A \cap S \neq \emptyset\}$$

Theorem: Only F that agrees with $\mathbb{1}^S$ on convex cones is $\chi^S(A) = \chi(A \cap \mathcal{S}^{m-1})$

Idea : Use $q_{m,n} = \mathbb{E}[\chi^S(V \cap C_{m,n})]$ as a surrogate for $1 - p_{m,n}$.

Excursion sets of Gaussian random fields [Adler&Taylor '07]

$$q_{m,n+1} = \frac{(-1)^n}{2^{m-n-1}} \sum_{i=0}^{\lfloor n/2 \rfloor} \sum_{l=0}^{n} \binom{m}{n-2i} \left(-\frac{1}{2}\right)^l \binom{n-2i}{l-2i} \sum_{j=0}^{l} \binom{m-n+l}{j}$$



$\alpha_{\mathrm{inj}}^{\mathrm{Euler}} \simeq 8.34$

Physics point of view

$\theta(z) = \mathbb{1}[z > 0]$

$$p_{m,n} = \mathbb{P}[\mathbf{W}\mathbb{R}^n \cap C_{m,n} \cap \mathcal{S}^{m-1} = \emptyset] = \mathbb{P}_{\mathbf{W}}\Big[\min_{\mathbf{x}\in\mathcal{S}^{n-1}} E_{\mathbf{W}}(\mathbf{x}) \geq n\Big]$$

$$E_{\mathbf{W}}(\mathbf{x}) \equiv \sum_{\mu=1}^{m} \theta[(\mathbf{W}\mathbf{x})_\mu]$$

"Gardner–Derrida" perceptron
cf. also [Franz & al '17], [Urbani '18]

Injectivity ⟷ Property of the GS of a spherical perceptron (F-RSB phase)

RSB strategy

Free entropy $\Phi(\beta) \equiv \lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathcal{S}^{n-1}} \mathrm{d}\mu_n(\mathbf{x})\, e^{-\beta E_{\mathbf{W}}(\mathbf{x})}$ → $\lim_{\beta\to\infty} -\frac{\Phi(\beta)}{\beta} \overset{?}{=} \text{p-lim}_{n\to\infty} \left\{\frac{1}{n} \min_{\mathbf{x}\in\mathcal{S}^{n-1}} E_{\mathbf{W}}(\mathbf{x})\right\}$

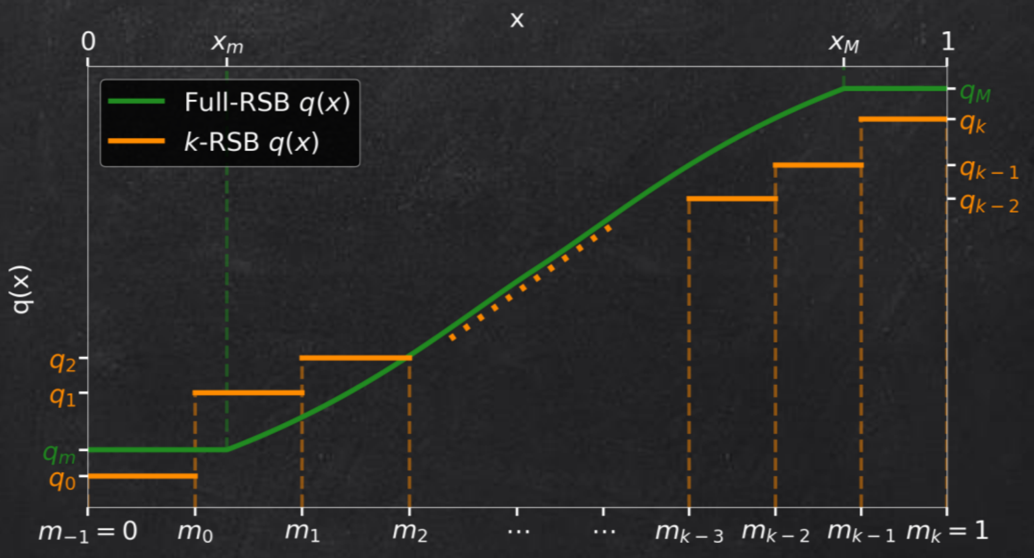
Ground-state energy

$f^\star_{k-\mathrm{RSB}}(\alpha) \equiv \lim_{\beta\to\infty} [-\Phi_{k-\mathrm{RSB}}(\beta)/\beta]$ → "Injectivity thresholds in the k-RSB approximation"

$$\alpha_{\mathrm{inj}} \overset{\mathrm{conj.}}{=} \alpha_{\mathrm{inj}}^{\mathrm{FRSB}} \leq \cdots \leq \alpha_{\mathrm{inj}}^{\mathrm{(k+1)-RSB}} \leq \alpha_{\mathrm{inj}}^{\mathrm{k-RSB}} \leq \cdots \leq \alpha_{\mathrm{inj}}^{\mathrm{1RSB}} \leq \alpha_{\mathrm{inj}}^{\mathrm{RS}}$$

REPLICA SYMMETRY BREAKING THEORY

$$\gamma_w(x) = e^{-x^2/2w}/\sqrt{2\pi w}$$



Finite-temperature Parisi formula

$$\Phi_{\rm FRSB}(\beta) = \inf_{\{q(x)\}} \left\{ \frac{1}{2}\ln(1-q(1)) + \frac{q(0)}{2(1-\langle q\rangle)} + \frac{1}{2}\int_0^1 du \frac{q'(u)}{\lambda(u)} + \alpha\gamma_{q(0)} \star \phi(x=0,h=0) \right\}$$

$$\begin{cases} \phi(1,h) &= \ln \gamma_{1-q(1)} \star e^{-\beta\theta(h)} \\ \partial_x \phi(x,h) &= -\frac{q'(x)}{2}\left[\partial_h^2\phi(x,h) + x\partial_h\phi(x,h)^2\right], \qquad x\in(0,1) \end{cases}$$

Parisi PDE

Zero-temperature scalings

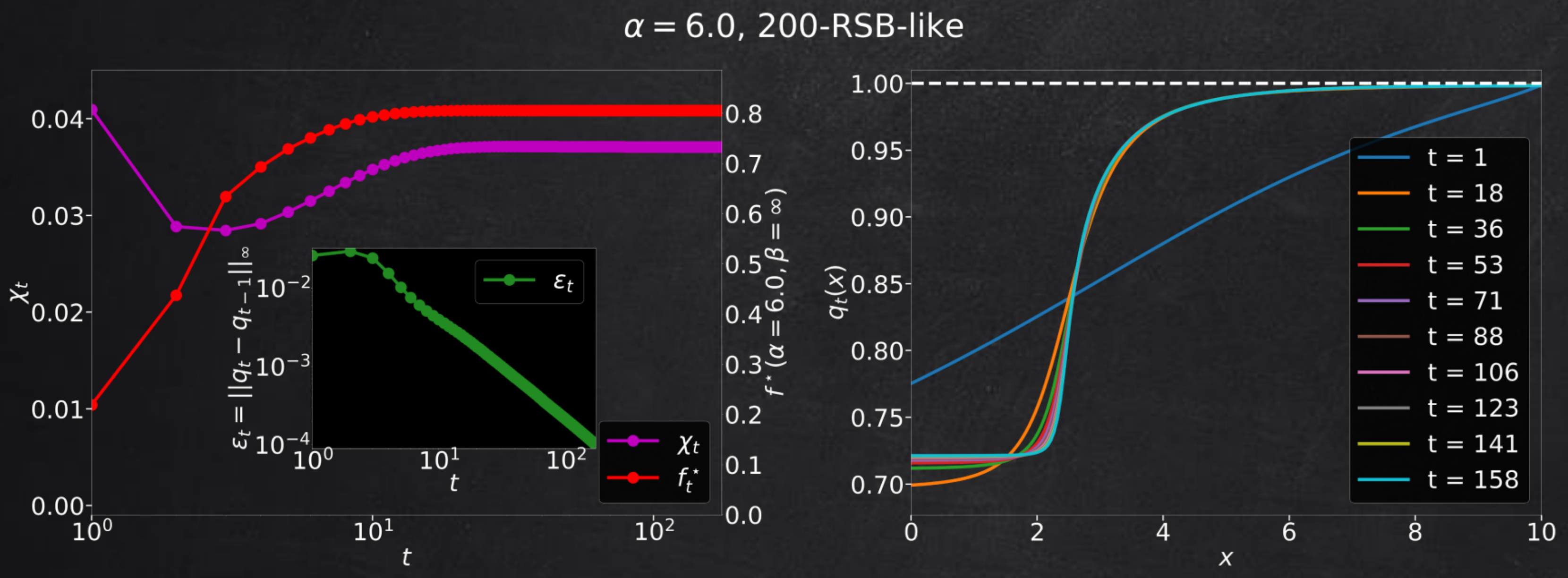
$$\begin{cases} q_M \simeq 1-\chi/\beta \\ \phi(q,h) \simeq \beta\phi_\infty(q,h) \\ \vdots \end{cases}$$

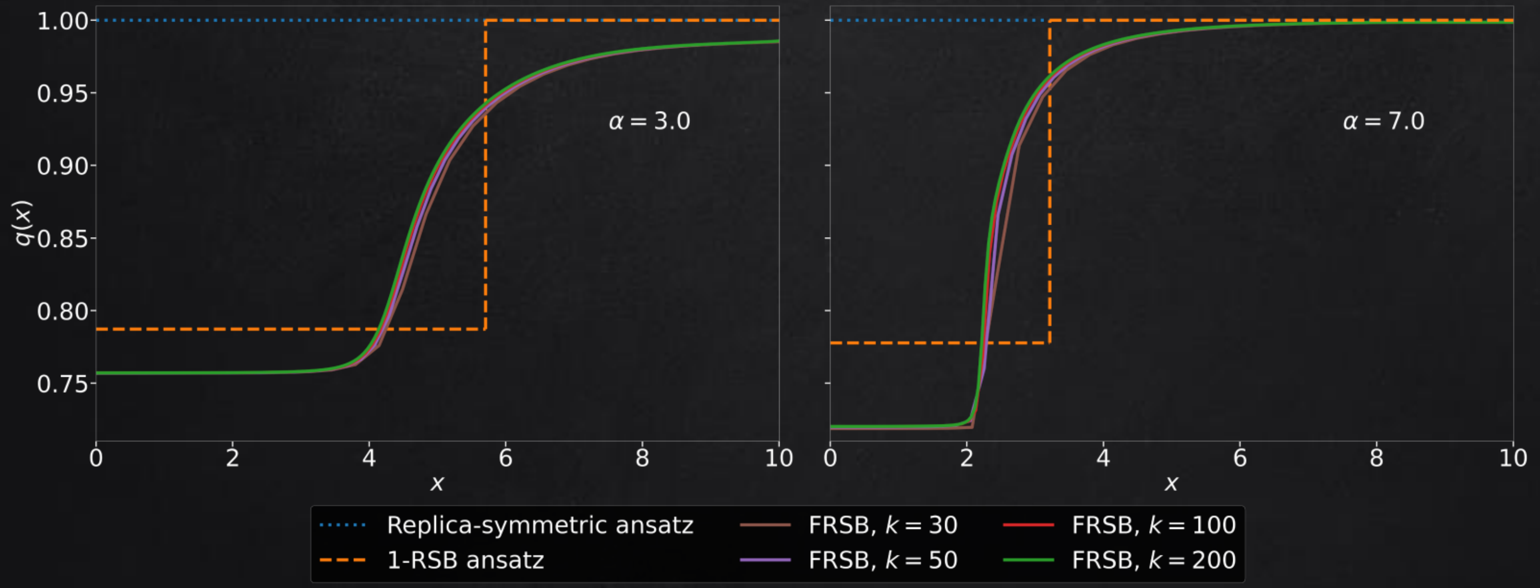
→ Zero-temperature Parisi PDE

↓

Discretization in a "k-RSB-like" ansatz

Algorithmic solution to F-RSB equations at T = 0





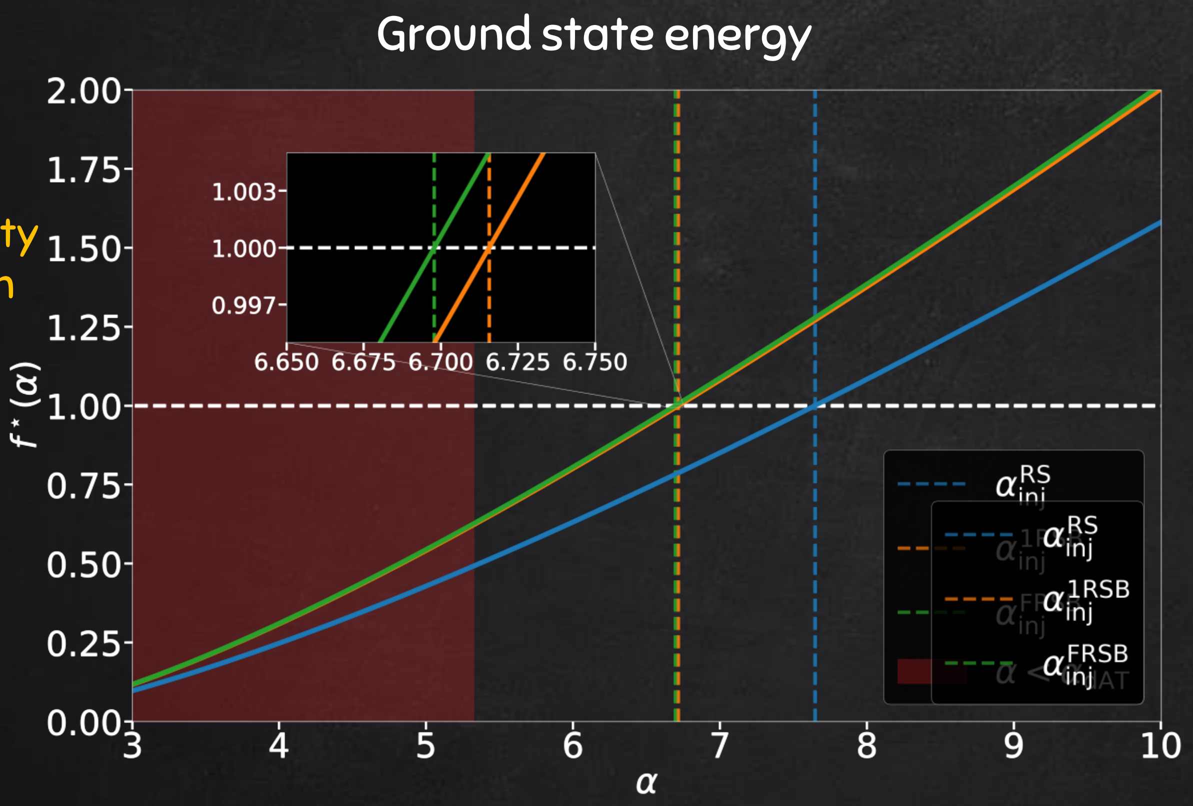
Discretization of the Parisi PDE in a "k-RSB-like" ansatz

- Fast gaussian convolutions using DFT and Shannon-Whittaker interpolation

$$\phi(x, h) \simeq \sum_{i=-N}^{N} \phi_i(x) \operatorname{sinc}\left(\frac{h - h_i}{\Delta h}\right)$$

- Scales very well: ~5 minutes with k = 200 on a standard desktop GPU.

Results of replica symmetry breaking theory



Injectivity criterion

Full-RSB conjecture

$$\alpha_{\text{inj}}^{+} = \alpha_{\text{inj}}^{-} = \alpha_{\text{inj}}^{\text{FRSB}} \in (6.6979, 6.6982)$$

Lower bound from RS solution at T > 0 : $\alpha_{\text{dAT}} \simeq 5.32$

Proof would require showing RS in the high-temperature phase

$$\min_{\mathbf{x} \in \mathcal{S}^{n-1}} \sum_{\mu=1}^{n} \mathbb{1}\{\mathbf{W}_\mu \cdot \mathbf{x} > 0\} = \min_{\substack{\mathbf{x} \in \mathcal{S}^{n-1} \\ \mathbf{z} \in \mathbb{R}^m}} \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m} \left\{ \boldsymbol{\lambda}^\intercal \mathbf{W} \mathbf{x} - \boldsymbol{\lambda}^\intercal \mathbf{z} + \sum_{\mu=1}^{n} \mathbb{1}\{z_\mu > 0\} \right\}$$

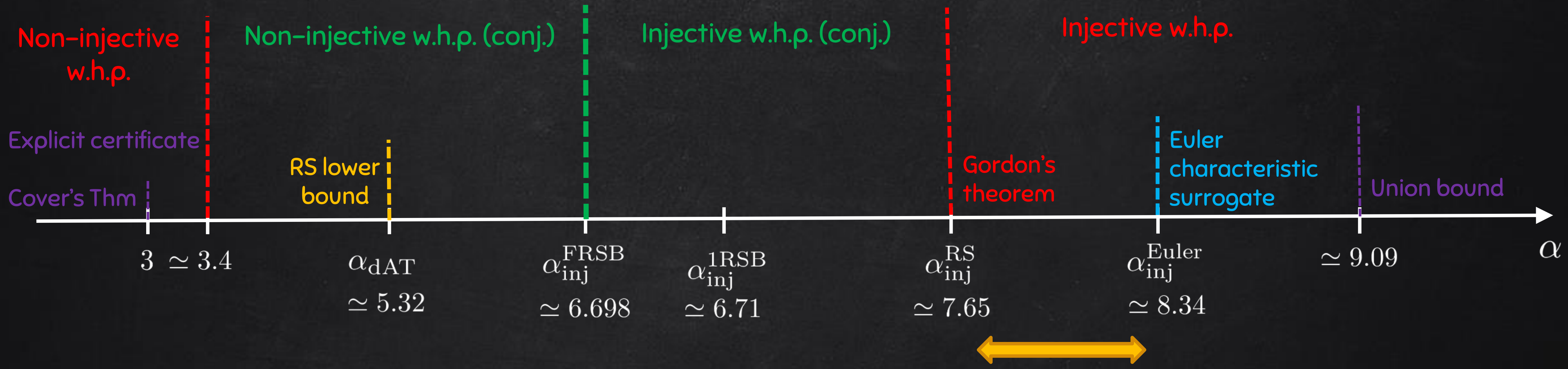
Gordon's minimax theorem [Gordon '85]

Theorem: $\alpha_{\text{inj}}^{+} \leq \alpha_{\text{inj}}^{\text{RS}} \simeq 7.65$

Similar to the proof of the RS upper bound for the Gardner capacity in the negative perceptron [Stojnic '13]

Summary & outlook

Expansivity thresholds for the injectivity of a random ReLU layer



Large deviations of sublevel sets of the perceptron ?

- Euler characteristic surrogate provably wrong !
- Stability : Lipschitz constant of $\varphi_{\mathrm{W}}^{-1}$ for $\alpha > \alpha_{\mathrm{inj}}^{\mathrm{FRSB}}$?
- Deep net: $\alpha \geq 2L \log L$ is enough for depth $L \gg 1$. [Paleka '21] Is this tight ? Geometry of the image of the network ?

THANK YOU !