Some results on average-case matrix discrepancy

Antoine Maillard

Inria

arXiv:2410.17887 + Work in progress

Spin glasses and related topics, IHP – January 28th 2025

Discrepancy theory

Cf. e.g. talks of Dan Spielman

"Divide a group of things into two similar groups"

Goal

Find a low-discrepancy coloring

$$\mathrm{disc}(\mathcal{S}) := \min_{\chi : [n] \to \{-1,1\}} \max_{S \in \mathcal{S}} \left| \sum_{j \in S} \chi(j) \right|$$

Set system $\mathcal{S} := \{S_1, \cdots, S_d\}$

$S_i \subseteq \{1, \cdots, n\} \Longleftrightarrow$ Characteristic $\#i$

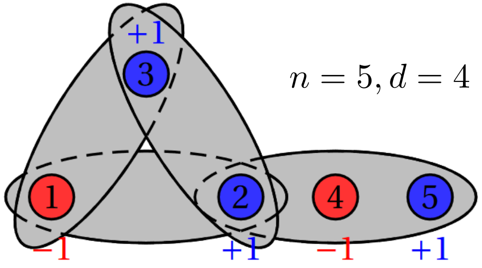
For a **coloring** $\chi \in \{-1,1\}^n$, we define its **discrepancy**

$$\mathrm{disc}(\chi) := \max_{i \in [d]} \left| \sum_{j \in S_i} \chi(j) \right|$$

Motivations / applications

Combinatorics, computational geometry, experimental design, theory of approximation algorithms, …

Matousek '09 ; Chen&al '14 ; …



$n = 5, d = 4$

$$\mathrm{disc}(\chi) = \max(0, 0, 1, 2) = 2$$

Discrepancy of a set system / a set of vectors

$$\mathrm{disc}(\mathcal{S}) = \min_{\varepsilon \in \{\pm 1\}^n} \max_{i \in [d]} |\langle \varepsilon, a_i \rangle| = \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^{n} \varepsilon_i u_i \right\|_\infty$$

$a_i := \mathbb{1}_{S_i} \in \{0,1\}^n$

$(u_i)_j := (a_j)_i$

Spencer's theorem

$a_i := \mathbb{1}_{S_i} \in \{0,1\}^n$

$(u_i)_j := (a_j)_i$

$$\mathrm{disc}(\mathcal{S}) = \min_{\varepsilon \in \{\pm 1\}^n} \max_{i \in [m]} |\langle \varepsilon, a_i \rangle| = \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^n \varepsilon_i u_i \right\|_\infty$$

Theorem (Spencer '85)

$\|u_i\|_\infty \leq 1 \quad \Rightarrow \quad \mathrm{disc}(u_1, \cdots, u_n) \leq 6\sqrt{n}$

- The scaling $\sqrt{n}$ is **optimal** (up to constants)
- **Bansal '10**: polynomial-time algorithms

Random signs $\varepsilon \sim \mathrm{Unif}(\{\pm 1\}^n)$

Hoeffding's inequality $\|a_i\|_\infty \leq 1$ ⇓ Union bound

$\mathbb{P}[\mathrm{disc}(\varepsilon) \leq \sqrt{2n \log(2d)}] > 0$

Can we beat random signings ?

Digression

Komlós conjecture (70s)

$\|u_i\|_2 \leq \sqrt{n}$ is sufficient for $\mathrm{disc}(\{u_i\}) \leq C\sqrt{n}$

Banaszczyk '98

Best-known result $\mathrm{disc}(\{u_i\}) \leq C\sqrt{n \log d}$

Hard result

~~Hoeffding's inequality~~

Random Spencer: the symmetric binary perceptron

$\|u_i\|_\infty \leq 1 \Rightarrow \mathrm{disc}(u_1, \cdots, u_n) \leq C\sqrt{n}$

What about **random vectors** ?

Given $g_1, \cdots, g_d \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_n)$, can we find $\varepsilon \in \{\pm 1\}^n$ such that $\max_{i \in [d]} |\langle g_i, \varepsilon \rangle| \leq \kappa\sqrt{n}$?

This is the **Symmetric Binary Perceptron (SBP)** Aubin, Perkins & Zdeborová '19

$(u_i)_j := (g_j)_i$

- $Z_\kappa := \# \left\{ \varepsilon \in \{\pm 1\}^n : \left\| \sum_{i=1}^n \varepsilon_i u_i \right\|_\infty \leq \kappa\sqrt{n} \right\}$
- $n/d \to \beta > 0$

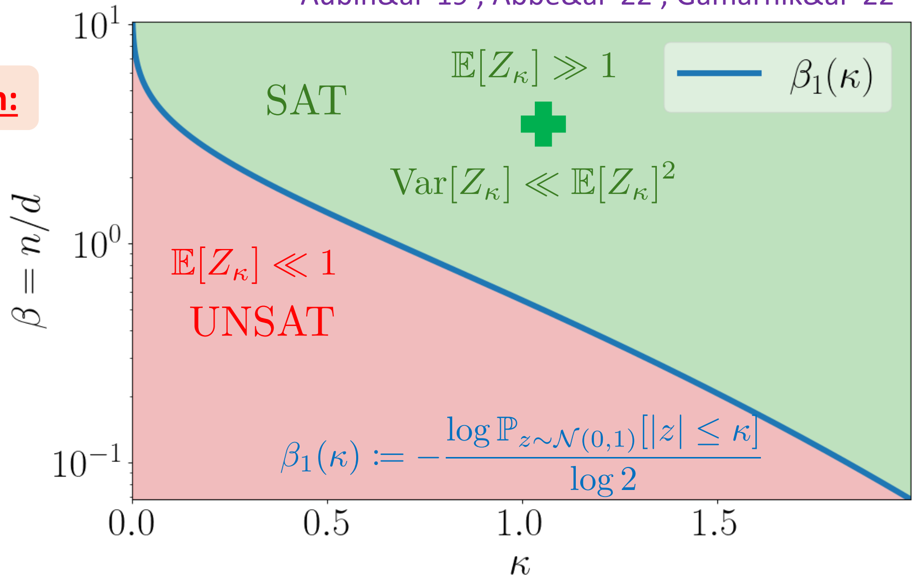
$\Rightarrow$ **Theorem:**

- There is a **sharp satisfiability threshold.**
- The **annealed** free energy is correct: $Z_\kappa \simeq \mathbb{E}[Z_\kappa]$
- Much more **detailed properties** are known:

  Structure of solution space, performance of solving algorithms, …

  Barbier&al '24 ; El Alaoui & Gamarnik '24 ; …

Aubin&al '19 ; Abbe&al '22 ; Gamarnik&al '22



Matrix discrepancy

$\|u_i\|_\infty \leq 1 \Rightarrow \mathrm{disc}(u_1, \cdots, u_n) \leq C\sqrt{n}$

What happens for **more complex discrepancy objectives** ?

Seminal example

$u_i \Rightarrow \mathbf{A}_i \in \mathbb{R}^{d \times d}$

$\| \cdot \|_\infty \Rightarrow \| \cdot \|_{\mathrm{op}}$ (Spectral norm)

Assume $\|\mathbf{A}_i\|_{\mathrm{op}} \leq 1$

Random signs $\varepsilon \sim \mathrm{Unif}(\{\pm 1\}^n)$

Non-commutative Khintchine inequality [Lust-Piquard & Pisier '91]

$$\mathrm{disc}(\mathbf{A}_1, \cdots, \mathbf{A}_n) \leq C\sqrt{n \log d}$$

Can we beat random signings ?

Matrix discrepancy

$$\mathrm{disc}(\mathbf{A}_1, \cdots, \mathbf{A}_n) := \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^n \varepsilon_i \mathbf{A}_i \right\|_{\mathrm{op}}$$

Applications in quantum random access codes, graph sparsification, …

Hopkins&al '22; Bansal&al '23; Batson&al '14; …

Conjecture ("Matrix Spencer")

$$\|\mathbf{A}_i\|_{\mathrm{op}} \leq 1 \Rightarrow \mathrm{disc}(\mathbf{A}_1, \cdots, \mathbf{A}_n) \leq C\sqrt{n}$$

Zouzias '12; Meka'14

Linked to the "ammount of commutativity" of $\{\mathbf{A}_i\}$, see talks of A. Bandeira.

$\{\mathbf{A}_i\}$ commute $\Rightarrow$ Spencer's theorem.

$\{\mathbf{A}_i\}$ "very non-commutative" $\Rightarrow$ $\mathbb{E}_\varepsilon \| \sum_i \varepsilon_i \mathbf{A}_i \|_{\mathrm{op}} \lesssim \sqrt{n}$

Bandeira, Boedihardjo & van Handel '23

Best result

Bansal&al '23 — Matrix Spencer holds if we further assume $\mathrm{rk}(\mathbf{A}_i) \lesssim n / \log^3 n$

Average-case matrix discrepancy: "Random Matrix Spencer"

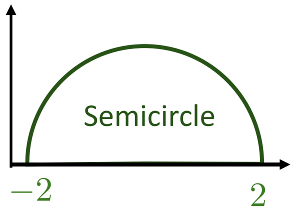
$W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, (1+\delta_{ij})/d)$ for $i \le j$.

What about **random matrices** ?

Given $\mathbf{W}_1, \cdots, \mathbf{W}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, can we find $\varepsilon \in \{\pm 1\}^n$ such that $\left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}} \le \kappa \sqrt{n}$?

Also introduced in [Kunisky & Zhang '23]

- Matrix analog of the SBP $(\mathbf{W}_i = \mathrm{Diag}(\mathbf{g}_i))$ recovers the Symmetric Binary Perceptron)
- Trivial bound: if $\kappa > 2$, $\mathbb{P}\left[\|\sum_{i=1}^n \mathbf{W}_i\|_{\mathrm{op}} \le \kappa\sqrt{n}\right] = \mathbb{P}_{\mathbf{W}\sim\mathrm{GOE}(d)}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa] = 1 - o(1)$



This talk

Goals

- Sharp **satisfiability transitions** ?
- Structure of solution space ?
- Polynomial-time solving algorithms ? [Kunisky & Zhang '23]
- Add **structure** to $\mathbf{W}_i$? To probe the **Matrix Spencer conjecture** ?

Results I: first moment asymptotics

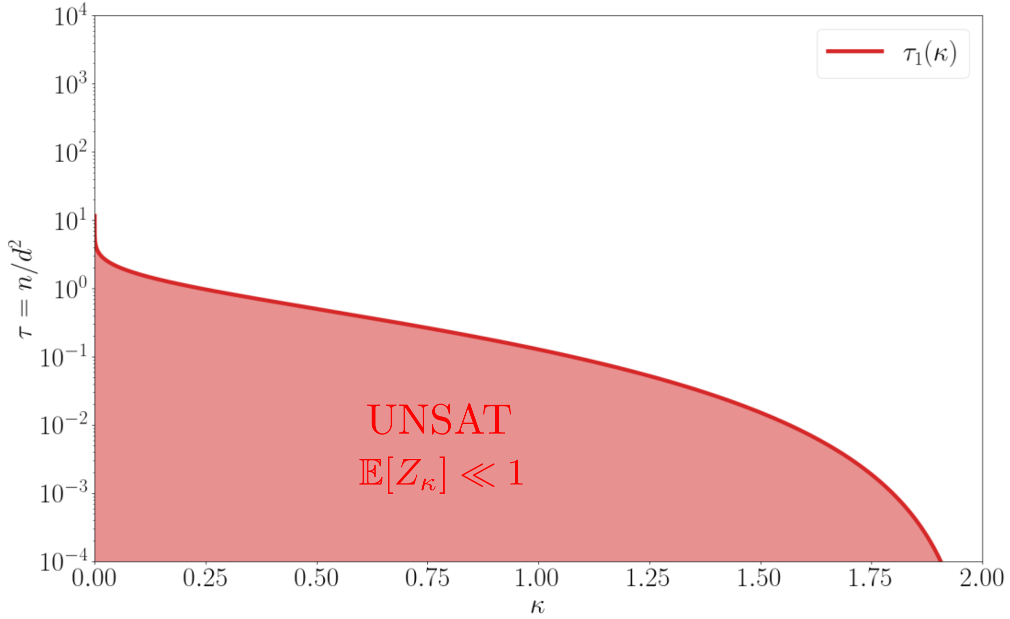
$n/d^2 \to \tau > 0$

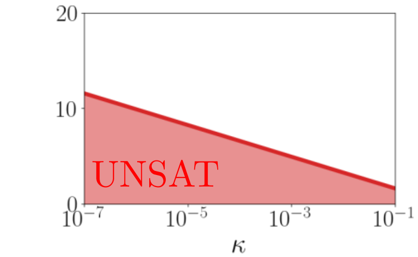
Number of solutions / Partition function

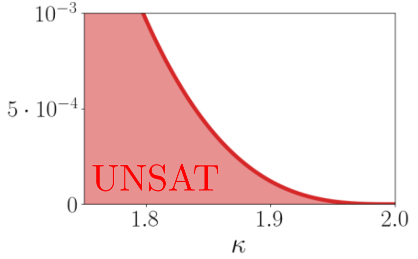
$$Z_\kappa := \# \left\{ \varepsilon \in \{\pm 1\}^n \text{ s.t. } \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}} \leq \kappa \sqrt{n} \right\}$$

Theorem: $$\lim_{d \to \infty} \frac{1}{d^2} \log \mathbb{E} Z_\kappa = (\tau - \tau_1(\kappa)) \log 2$$

$$\tau_1(\kappa) := \frac{1}{\log 2} \left[-\frac{\kappa^4}{128} + \frac{\kappa^2}{8} - \frac{1}{2} \log \frac{\kappa}{2} - \frac{3}{8} \right]$$





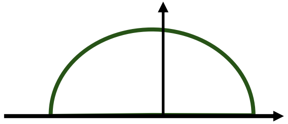


$\tau < \tau_1(\kappa)$

$\Downarrow$

$\mathbb{P}[Z_\kappa = 0] = 1 - o(1)$

First moment computation: a sketch



$\mathbf{W} \sim \mathrm{GOE}(d)$

$$\mathbb{E} Z_\kappa = \sum_{\varepsilon \in \{\pm 1\}^n} \mathbb{P}\left[\left\|\sum_{i=1}^n \varepsilon_i \mathbf{W}_i\right\|_{\mathrm{op}} \leq \kappa\sqrt{n}\right] = 2^n\, \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa]$$

→ **Left** ($\kappa < 2$) large deviations of $\|\mathbf{W}\|_{\mathrm{op}}$

$$I(\mu) := -\frac{1}{2}\int \mu(\mathrm{d}x)\mu(\mathrm{d}y) \log|x-y| + \frac{1}{4}\int \mu(\mathrm{d}x)\, x^2 - \frac{3}{8}$$

- **Intuition:** The events $\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa$ are driven by large deviations of the whole spectral density.

$$\mathbb{P}[\mu_{\mathbf{W}} \simeq \mu] \simeq \exp\{-d^2 I(\mu)\}$$

Ben Arous & Guionnet '97; Dean&Majumdar '06 '08;

- **Proof:** technical adaptations of the proof of [BAG '97]
  see also [Anderson, Guionnet & Zeitouni '10]

$$\Rightarrow \quad \lim \frac{1}{d^2} \log \mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \leq \kappa] = -\inf_{\mu \in \mathcal{M}([-\kappa,\kappa])} I(\mu)$$

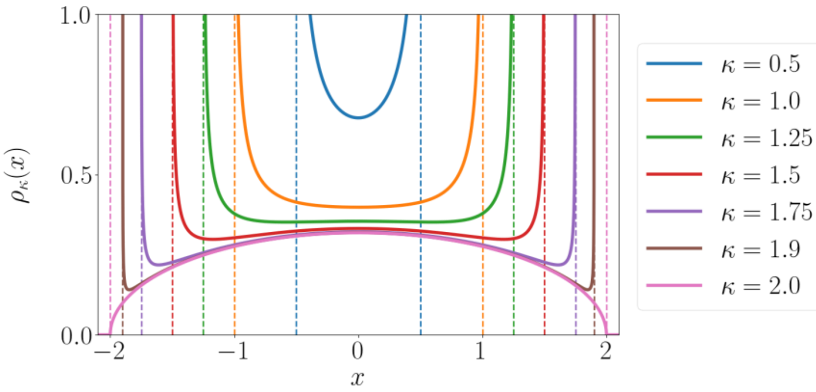
- Compute $\rho_\kappa(x)$ from **Tricomi's theorem**

Tricomi' 85; Dean&Majumdar '06 '08; Vivo&al '07, …

- Prove $\rho_\kappa = \arg\min_{\mu \in \mathcal{M}([-\kappa,\kappa])} I(\mu)$

Classical tools of logarithmic potential theory

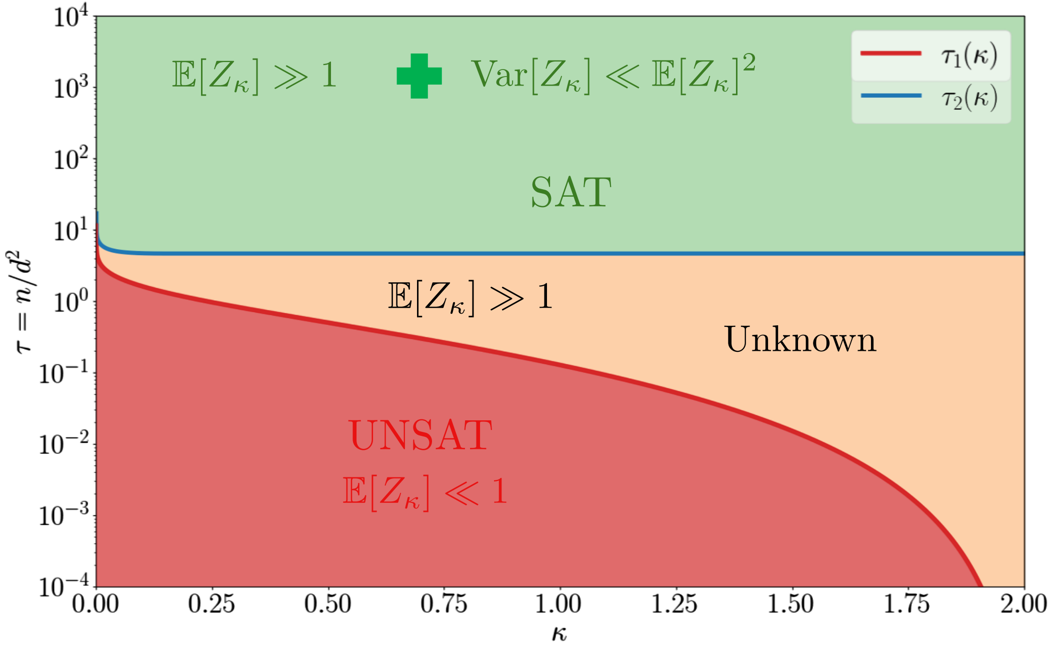
Saff&Totik'13; Ben Arous & Guionnet '97

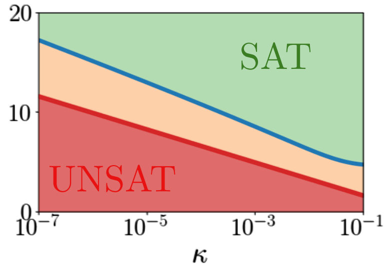


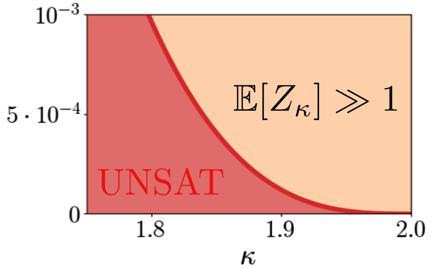
$$\rho_\kappa(x) = \frac{4 + \kappa^2 - 2x^2}{4\pi\sqrt{\kappa^2 - x^2}} \overset{?}{=} \arg\min_{\mu \in \mathcal{M}([-\kappa,\kappa])} I(\mu)$$

$n/d^2 \to \tau > 0$

Results II: Upper bounds via the second moment method







$$Z_\kappa := \# \left\{ \varepsilon \in \{\pm 1\}^n \text{ s.t. } \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}} \leq \kappa \sqrt{n} \right\}$$

Theorem I

$$\tau < \tau_1(\kappa) \Downarrow \mathbb{P}[Z_\kappa = 0] = 1 - o(1)$$

✚

Theorem II

$$\tau > \tau_2(\kappa) \Downarrow \mathbb{P}[Z_\kappa \geq 1] = 1 - o(1)$$

Second moment method: overview

$$\mathrm{disc}(\{\mathbf{W}_i\}) = \min_{\varepsilon \in \{\pm 1\}^n} \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}}$$

$n/d^2 \to \tau > 0$

Theorem

$$\tau > \tau_2(\kappa) \Rightarrow \mathbb{P}\left[Z_\kappa := \# \left\{ \varepsilon \in \{\pm 1\}^n \text{ s.t. } \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}} \leq \kappa \sqrt{n} \right\} \geq 1 \right] = 1 - o(1)$$

Explicit formula (non-optimal)

1 + 2 $\geq \delta$

3 $= 1 - o(1)$

Proof

1. **Sharp** 1[st] moment $(1/d^2) \log \mathbb{E} Z_\kappa \to \cdots$

2. Second moment **upper bound** $\dfrac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \lesssim \left[1 - \dfrac{\tau_2(\kappa)}{\tau} \right]^{-1/2}$

3. **Margin concentration** $\mathrm{Var}[\mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n)] \lesssim \dfrac{1}{d} \left(\mathbb{E}[\mathrm{disc}(\mathbf{W}_1, \cdots, \mathbf{W}_n)] \right)^2$

Results of [Altschuler'23], based on Talagrand's $L^1 - L^2$ inequality

Second moment upper bound: sketch

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \lesssim \left[1 - \frac{\tau_2(\kappa)}{\tau}\right]^{-1/2}$$

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} = \frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l}\exp\{nG_d(q_l)\},$$ where for $q \in [-1,1]$: $$G_d(q) := \frac{1}{n}\log\frac{\mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa \text{ and } \|q\mathbf{W} + \sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right]}{\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa]^2}$$

$q_l := 2(l/n) - 1$

$\mathbf{W}, \mathbf{Z} \sim \langle\cdot\rangle_{q,\kappa}$

Large deviations of the spectral norm of **correlated** $\mathrm{GOE}(d)$ matrices.

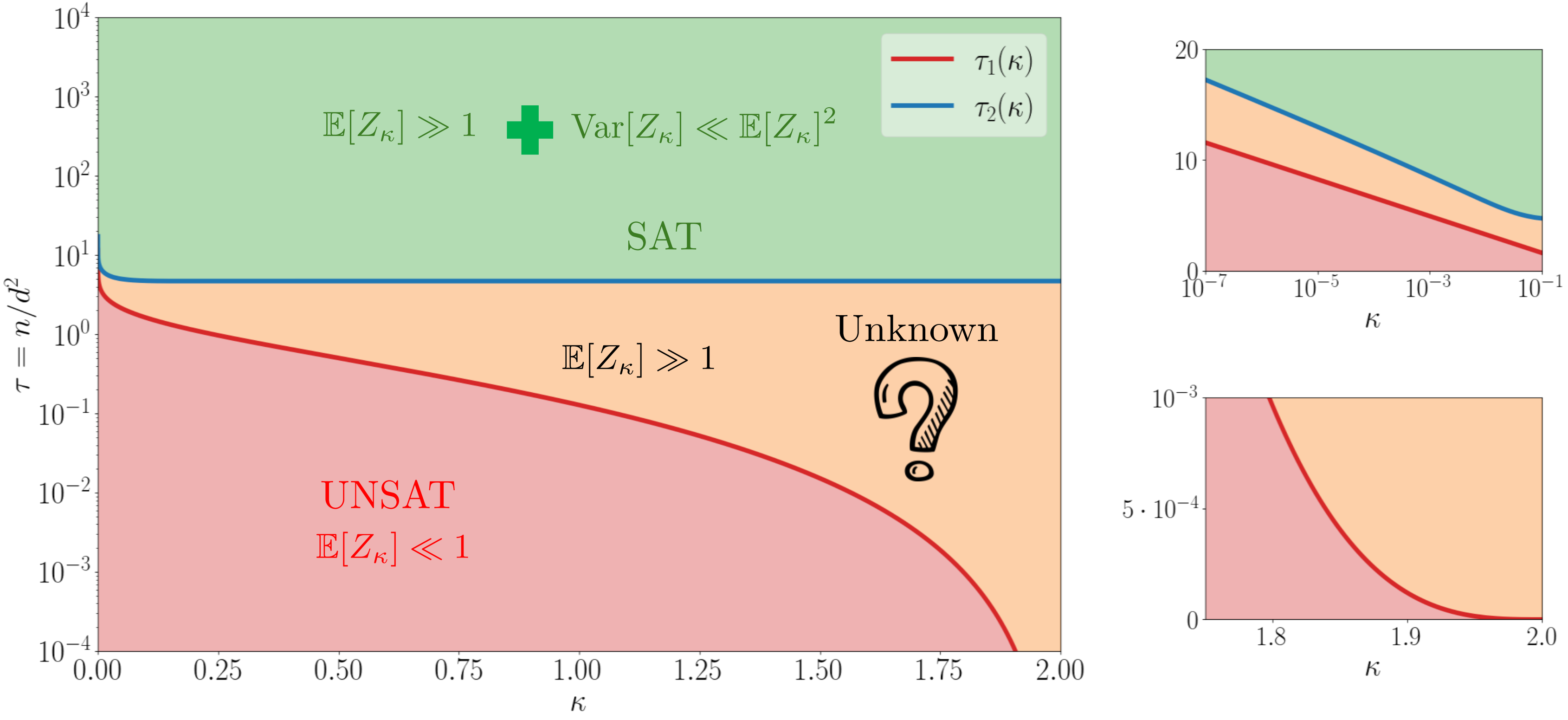
- Upper bound on $G_d(q)$

❖ Crude upper bound for q far from 0: $G_d(q) \le -\frac{1}{n}\log\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa]$.

❖ For small q, upper bounding $\sup_{|q|\le\varepsilon} G_d''(q) \le \frac{\tau_2(\kappa)}{\tau}$ ⇐

- Log-Sobolev inequality for $\langle\cdot\rangle_{q,\kappa}$
  - Approximation of $\mathbb{1}[|x| \le \kappa]$ by smooth functions
  - ✚ Bakry-Emery condition for smooth and strongly log-concave measures
- Concentration of moments $\mathrm{Tr}[\mathbf{W}^a\mathbf{Z}^b]$ under $\langle\cdot\rangle_{q,\kappa}$.

- **Discrete Laplace's method** over the overlap $q \in [-1,1]$ in $\frac{1}{2^n}\sum_{l=0}^{n}\binom{n}{l}\exp\{nG_d(q_l)\}$. ⇒ $$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} \lesssim \left[1 - \frac{\tau_2(\kappa)}{\tau}\right]^{-1/2}$$


$\mathbb{E}[Z_\kappa] \gg 1$
$\mathrm{Var}[Z_\kappa] \ll \mathbb{E}[Z_\kappa]^2$
SAT
Unknown
$\mathbb{E}[Z_\kappa] \gg 1$
UNSAT
$\mathbb{E}[Z_\kappa] \ll 1$
$\tau_1(\kappa)$
$\tau_2(\kappa)$
$\tau = n/d^2$
κ

Results III: failure of the second moment method

$$G_d(q) := \frac{1}{n} \log \frac{\mathbb{P}\left[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa \text{ and } \|q\mathbf{W} + \sqrt{1-q^2}\mathbf{Z}\|_{\mathrm{op}} \le \kappa\right]}{\mathbb{P}[\|\mathbf{W}\|_{\mathrm{op}} \le \kappa]^2}$$

$$\frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} = \frac{1}{2^n} \sum_{l=0}^{n} \binom{n}{l} \exp\{n G_d(q_l)\}$$ ✚
- Upper bound on $G_d(q), G_d''(q)$.
- Discrete Laplace's method.

➢ Explicit computation: $\lim_{d\to\infty} G_d''(0) = \frac{\tau_{\text{fail.}}(\kappa)}{\tau}$ $\tau_{\text{fail.}}(\kappa) := \frac{1}{2}\left(\frac{\kappa^2}{4} - 1\right)^4$

Theorem

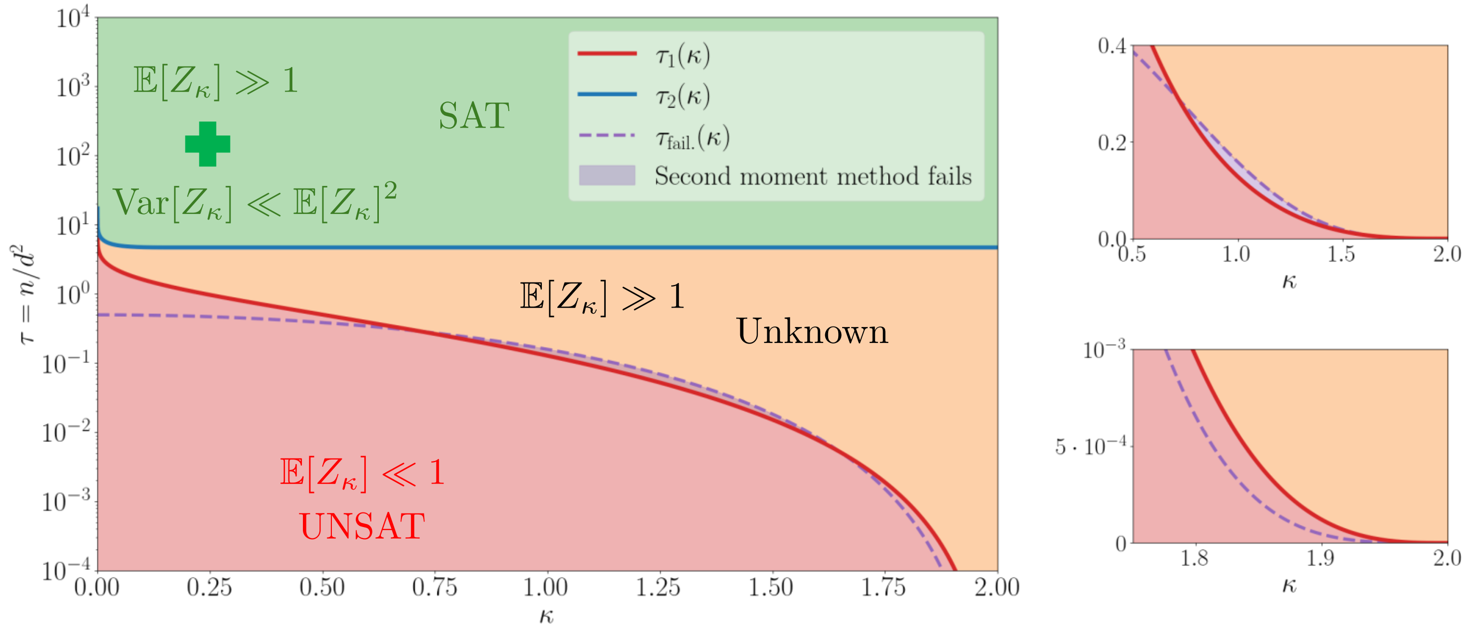
➢ Lower bound in Laplace's method: $n/d^2 \to \tau < \tau_{\text{fail.}}(\kappa) \Rightarrow \liminf_{d\to\infty} \frac{1}{d^2} \log \frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} > 0$

Technical assumption: uniform continuity of $G_d''(q)$ in $q = 0$ as $d \to \infty$

Non-concentration of Z_κ on $\mathbb{E}[Z_\kappa]$.

Failure of the second moment method

$$n/d^2 \to \tau < \tau_{\text{fail.}}(\kappa) \Rightarrow \liminf_{d\to\infty} \frac{1}{d^2} \log \frac{\mathbb{E}[Z_\kappa^2]}{\mathbb{E}[Z_\kappa]^2} > 0$$



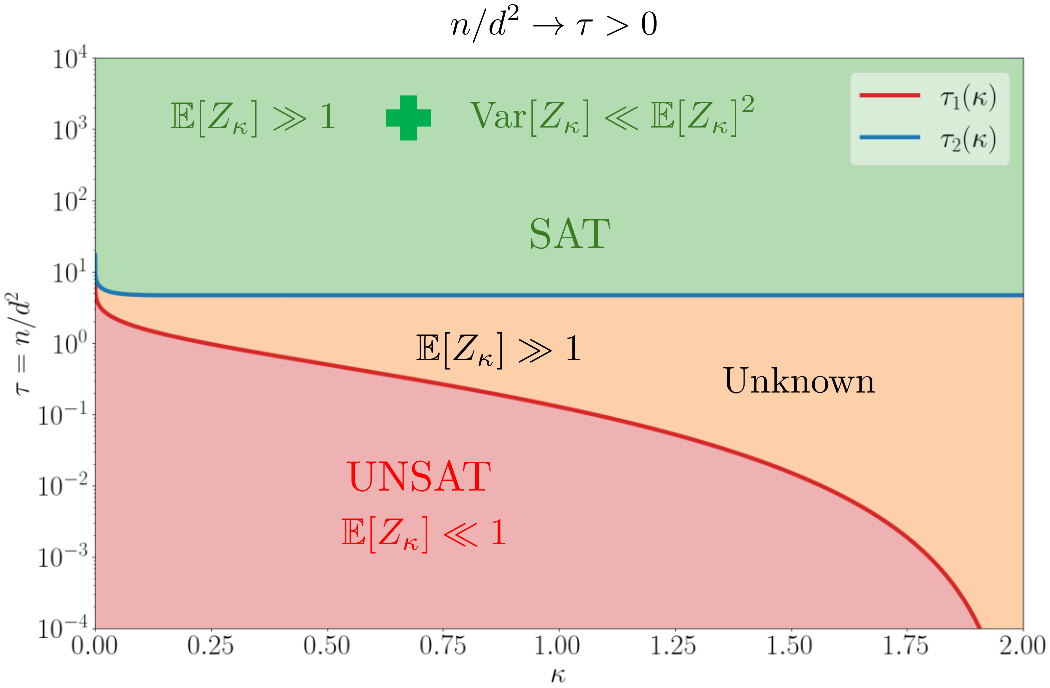
- **Purple region**: Z_κ provably does not concentrate on its average **+** $\mathbb{E}[Z_\kappa] \gg 1$

  Quenched $\neq$ annealed **at least** in the purple region

- The phase diagram is **more complex** than in the Symmetric Binary Perceptron !

Summary: average-case matrix discrepancy

$$Z_\kappa := \#\left\{ \varepsilon \in \{\pm 1\}^n \text{ s.t. } \left\| \sum_{i=1}^n \varepsilon_i \mathbf{W}_i \right\|_{\mathrm{op}} \le \kappa \sqrt{n} \right\}$$



THANK YOU !

Matrix analog of the SBP

Failure of the second moment method in part of the diagram ($\neq$ Symmetric Binary Perceptron)

What's next

- ❑ **Sharp** second moment
- ❑ **Replica** free energy (at least RS level)

- **Structure** of the solution space ?
  RS/RSB, …
- (Efficient) **algorithms** ?
- Applications to **non-GOE matrices** ?
  To interesting models for the **matrix Spencer conjecture** ?