Estimating Rank-One Matrices with Mismatched Prior and Noise: Universality and Large Deviations

Institut d'Études Scientifiques — Cargèse
Statistical Physics & Machine Learning Back Together Again

Alice Guionnet (ENS Lyon) **Justin Ko (ENS Lyon)**, Florent Krzakala (EPFL), Lenka Zdeborová (EPFL)

August 9, 2023

Table of Contents

Table of Contents

General Statistical Inference Problem

Signal: Want to recover a rank 1 signal $\boldsymbol{X}_0 \in \mathbb{R}^{N\times N}$ of the form

$$\boldsymbol{X}_0 = \boldsymbol{x}_0 \boldsymbol{x}_0^\intercal = (x_i^0 x_j^0)_{i,j\leq N} \quad x_i^0 x_j^0 = O(1),$$

$\boldsymbol{x}_0 \in \mathbb{R}^N$ is generated (independently) from the *signal measure* $\mathbb{P}_0$,

$$x_i^0 \sim \mathbb{P}_0 \,.$$

Observation: Observe the signal through some noisy data $\boldsymbol{Y} \in \mathbb{R}^{N\times N}$

$$\boldsymbol{Y} = (Y_{ij})_{i,j\leq N} \qquad Y_{ij} = Y_{ji} \qquad Y_{ij} = O(1),$$

$\boldsymbol{Y}$ is generated (independently) from the *output channel* $\mathbb{P}_{\rm out}(\cdot | \frac{1}{\sqrt{N}}\boldsymbol{X}_0)$,

$$Y_{ij} \sim \mathbb{P}_{out}\left(\cdot \;\middle|\; \frac{x_i^0 x_j^0}{\sqrt{N}} \right).$$

General Statistical Inference Problem

Signal: Want to recover a rank 1 signal $\boldsymbol{X}_0 \in \mathbb{R}^{N \times N}$ of the form

$$\boldsymbol{X}_0 = \boldsymbol{x}_0 \boldsymbol{x}_0^\mathsf{T} = (x_i^0 x_j^0)_{i,j \leq N} \quad x_i^0 x_j^0 = O(1),$$

$\boldsymbol{x}_0 \in \mathbb{R}^N$ is generated (independently) from the *signal measure* $\mathbb{P}_0$,

$$x_i^0 \sim \mathbb{P}_0 \,.$$

Observation: Observe the signal through some noisy data $\boldsymbol{Y} \in \mathbb{R}^{N \times N}$

$$\boldsymbol{Y} = (Y_{ij})_{i,j \leq N} \qquad Y_{ij} = Y_{ji} \qquad Y_{ij} = O(1),$$

$\boldsymbol{Y}$ is generated (independently) from the *output channel* $\mathbb{P}_{\mathrm{out}}(\cdot | \frac{1}{\sqrt{N}} \boldsymbol{X}_0)$,

$$Y_{ij} \sim \mathbb{P}_{out} \Big(\cdot \ \Big| \ \frac{x_i^0 x_j^0}{\sqrt{N}} \Big).$$

Example: The rank 1 spiked matrix, $Y_{ij} \sim N(\sqrt{\lambda} \frac{x_i^0 x_j^0}{\sqrt{N}}, 1)$

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

where $G_{ij} \sim N(0,1)$.

Minimal Mean Squared Error

Minimal Matrix Mean Squared Error

$$\mathrm{MMSE}(N) = \min_\theta \frac{2}{N(N-1)} \mathbb{E}\,\mathrm{tr}((\boldsymbol{X}_0 - \theta(\boldsymbol{Y}))^2$$

Minimal Mean Squared Error

Minimal Matrix Mean Squared Error

$$\begin{aligned}\mathrm{MMSE}(N) &= \min_\theta \frac{2}{N(N-1)} \mathbb{E}\,\mathrm{tr}((\boldsymbol{X}_0 - \theta(\boldsymbol{Y}))^2 \\ &= \frac{2}{N(N-1)} \sum_{i<j} \mathbb{E}(\boldsymbol{X}_0 - \mathbb{E}[\boldsymbol{X}_0 \mid \boldsymbol{Y}])^2.\end{aligned}$$

Minimal Mean Squared Error

Minimal Matrix Mean Squared Error

$$\begin{aligned}\mathrm{MMSE}(N) &= \min_\theta \frac{2}{N(N-1)} \mathbb{E}\,\mathrm{tr}((\boldsymbol{X}_0 - \theta(\boldsymbol{Y}))^2 \\ &= \frac{2}{N(N-1)} \sum_{i<j} \mathbb{E}(\boldsymbol{X}_0 - \mathbb{E}[\boldsymbol{X}_0 \mid \boldsymbol{Y}])^2.\end{aligned}$$

Optimal Estimator

$$\mathbb{E}[\boldsymbol{X}_0 | \boldsymbol{Y}]$$

Bayesian (Optimal) Inference

Posterior: The (optimal) statistician's best guess is

$$\mathbb{G}_N^{\mathrm{opt}}(\boldsymbol{X}) = \mathbb{P}_{\mathrm{opt}}(\boldsymbol{X}_0 = \boldsymbol{X} \mid \boldsymbol{Y}) = \frac{\mathbb{P}_{\mathrm{out}}(\boldsymbol{Y} \mid \frac{1}{\sqrt{N}}\boldsymbol{X})\,\mathbb{P}_0(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})}$$

$$= \prod_{i<j} \frac{\mathbb{P}_{\mathrm{out}}(Y_{ij} \mid \frac{1}{\sqrt{N}}x_i^0 x_j^0)\,\mathbb{P}_0(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})}$$

Overlap: A simple way to measure how good the estimate is the *overlap*

$$\boldsymbol{R}_{1,0} = \frac{1}{N}(\hat{\boldsymbol{x}} \cdot \boldsymbol{x}_0) \qquad \hat{\boldsymbol{x}} \sim \mathbb{G}_N^{\mathrm{opt}} = \mathbb{P}_{\mathrm{opt}}(\cdot \mid \boldsymbol{Y}).$$

The overlap is a fundamental object that is used to compute many interesting quantities (order parameters, free energy, MMSE, etc).

Spiked Matrix Example

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

Spiked Matrix Example

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}}\boldsymbol{X}_0$$

Posterior Distribution:

$$\left(Y_{ij} - \sqrt{\lambda}\frac{x_i^0 \cdot x_j^0}{\sqrt{N}}\right) = g_{ij} \implies \mathbb{P}_{\text{out}}(Y_{ij} \mid \boldsymbol{X}_0) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right)$$

$$\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{\mathbb{P}_{\text{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})}{\int \mathbb{P}_{\text{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})}$$

Spiked Matrix Example

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}}\boldsymbol{X}_0$$

Posterior Distribution:

$$\left(Y_{ij} - \sqrt{\lambda}\frac{x_i^0 \cdot x_j^0}{\sqrt{N}}\right) = g_{ij} \implies \mathbb{P}_{\mathrm{out}}(Y_{ij} \mid \boldsymbol{X}_0) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right)$$

$$\begin{aligned}\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) &= \frac{\mathbb{P}_{\mathrm{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})}{\int \mathbb{P}_{\mathrm{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})} \\ &= \frac{e^{-\sum_{i<j}\frac{1}{2}(Y_{ij} - \frac{\sqrt{\lambda}}{\sqrt{N}}x_i x_j)^2}d\,\mathbb{P}_X(\boldsymbol{X})}{\int e^{-\sum_{i<j}\frac{1}{2}(Y_{ij} - \frac{\sqrt{\lambda}}{\sqrt{N}}x_i x_j)^2}d\,\mathbb{P}_0(\boldsymbol{X})}.\end{aligned}$$

Spiked Matrix Example

Posterior Distribution:

$$\left(Y_{ij} - \sqrt{\lambda}\frac{x_i^0 \cdot x_j^0}{\sqrt{N}}\right) = g_{ij} \implies \mathbb{P}_{\text{out}}(Y_{ij} \mid \boldsymbol{X}_0) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right)$$

$$\begin{aligned}\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) &= \frac{\mathbb{P}_{\text{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})}{\int \mathbb{P}_{\text{out}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X})d\,\mathbb{P}_0(\boldsymbol{X})} \\ &= \frac{e^{H_N(x)+\cancel{C(Y)}}d\,\mathbb{P}_X(\boldsymbol{X})}{\underbrace{\int e^{H_N(x)+\cancel{C(Y)}}d\,\mathbb{P}_X(\boldsymbol{X})}_{Z_N(Y)}}.\end{aligned}$$

Hamiltonian

$$H_N(\boldsymbol{x}) = \sum_{i<j}\sqrt{\lambda}\frac{W_{ij}}{\sqrt{N}}x_ix_j + \frac{\lambda}{N}(x_ix_j)(x_i^0x_j^0) - \frac{\lambda}{2N}(x_ix_j)^2$$

Free Energy:

$$F_N(\lambda) = \frac{1}{N}\mathbb{E}\log Z_N(Y) = \frac{1}{N}\mathbb{E}\log\int e^{-\sum_{i<j}\frac{1}{2}\left(Y_{ij}-\frac{\sqrt{\lambda}}{\sqrt{N}}x_ix_j\right)^2}d\,\mathbb{P}_0(x).$$

The Limit of the Free Energy

Theorem 1 (Limiting Free Energy (Barbier et al, Lelarge - Miolane))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_q \varphi(q).$$

Replica Symmetric Functional:

$$\varphi(q) = -\frac{\lambda q^2}{4} + \mathbb{E}\ln\Big[\int \exp\Big(\sqrt{\lambda q}zx + \lambda x x_0 - \frac{\lambda x^2}{2}\Big)\, d\,\mathbb{P}_0(x)\Big].$$

The Limit of the Free Energy

Theorem 2 (Limiting Free Energy (Barbier et al, Lelarge - Miolane))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_q \varphi(q).$$

Replica Symmetric Functional:

$$\varphi(q) = -\frac{\lambda q^2}{4} + \mathbb{E} \ln \left[\int \exp\left(\sqrt{\lambda q} z x + \lambda x x_0 - \frac{\lambda x^2}{2} \right) d\,\mathbb{P}_0(x) \right].$$

Overlap Concentration (Barbier): Nishimimori identity

$$\mathbb{E}\langle (\boldsymbol{R}_{10} - \mathbb{E}\langle \boldsymbol{R}_{10}\rangle)^2 \rangle = \mathbb{E}\langle (\boldsymbol{R}_{12} - \mathbb{E}\langle \boldsymbol{R}_{12}\rangle)^2 \rangle \to 0$$

where $\langle \cdot \rangle$ is the average with respect to $\mathbb{G}_N^{\mathrm{opt}}$ and $\hat{\boldsymbol{x}} \sim \mathbb{G}_N^{\mathrm{opt}}$ and $\boldsymbol{R}_{12}$ is the overlap of two independent samples from $\mathbb{G}_N^{\mathrm{opt}}$.

The Limit of the Free Energy

Theorem 3 (Limiting Free Energy (Barbier et al, Lelarge - Miolane))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_q \varphi(q).$$

Replica Symmetric Functional:

$$\varphi(q) = -\frac{\lambda q^2}{4} + \mathbb{E}\ln\left[\int \exp\left(\sqrt{\lambda q}zx + \lambda x x_0 - \frac{\lambda x^2}{2}\right) d\,\mathbb{P}_0(x)\right].$$

Overlap Concentration (Barbier): Nishimimori identity

$$\mathbb{E}\langle(\boldsymbol{R}_{10} - \mathbb{E}\langle\boldsymbol{R}_{10}\rangle)^2\rangle = \mathbb{E}\langle(\boldsymbol{R}_{12} - \mathbb{E}\langle\boldsymbol{R}_{12}\rangle)^2\rangle \to 0$$

where $\langle\cdot\rangle$ is the average with respect to $\mathbb{G}_N^{\text{opt}}$ and $\hat{\boldsymbol{x}} \sim \mathbb{G}_N^{\text{opt}}$ and $\boldsymbol{R}_{12}$ is the overlap of two independent samples from $\mathbb{G}_N^{\text{opt}}$.

Order Parameter: The maximizing q satisfies

$$q = \mathbb{E}\langle\boldsymbol{R}_{1,0}\rangle$$

Bayesian (Mismatched) Inference

Posterior: The (realistic) statistician does not know how $\boldsymbol{Y}$ or $\boldsymbol{X}$ are generated, so they make their own model

$$x_i \sim \mathbb{P}_X \qquad Y_{ij} \sim \mathbb{P}_{\mathrm{mis}}(\cdot \mid \frac{1}{\sqrt{N}}\boldsymbol{X}_0).$$

The associated likelihood with this model is

$$\begin{aligned}\mathbb{P}_{\mathrm{mis}}(\boldsymbol{X}_0 = \boldsymbol{X} \mid \boldsymbol{Y}) &= \frac{\mathbb{P}_{\mathrm{mis}}(\boldsymbol{Y} \mid \frac{1}{\sqrt{N}}\boldsymbol{X})\,\mathbb{P}_X(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})} \\ &= \prod_{i<j} \frac{\mathbb{P}_{\mathrm{mis}}(Y_{ij} \mid \frac{1}{\sqrt{N}}x_i x_j)\,\mathbb{P}_X(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})}\end{aligned}$$

Overlap: A simple way to measure how good the estimate is the *overlap*

$$\boldsymbol{R}_{1,0} = \frac{1}{N}(\hat{\boldsymbol{x}} \cdot \boldsymbol{x}_0) \qquad \hat{\boldsymbol{x}} \sim \mathbb{P}_{\mathrm{mis}}(\cdot \mid \boldsymbol{Y}).$$

Bayesian (Mismatched) Inference

Posterior: The (realistic) statistician does not know how $\boldsymbol{Y}$ or $\boldsymbol{X}$ are generated, so they make their own model

$$x_i \sim \mathbb{P}_X \qquad Y_{ij} \sim \mathbb{P}_{\mathrm{mis}}(\cdot \mid \frac{1}{\sqrt{N}} \boldsymbol{X}_0).$$

The associated likelihood with this model is

$$\begin{aligned} \mathbb{P}_{\mathrm{mis}}(\boldsymbol{X}_0 = \boldsymbol{X} \mid \boldsymbol{Y}) &= \frac{\mathbb{P}_{\mathrm{mis}}(\boldsymbol{Y} \mid \frac{1}{\sqrt{N}} \boldsymbol{X}) \, \mathbb{P}_X(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})} \\ &= \prod_{i<j} \frac{\mathbb{P}_{\mathrm{mis}}(Y_{ij} \mid \frac{1}{\sqrt{N}} x_i x_j) \, \mathbb{P}_X(\boldsymbol{X})}{\mathbb{P}(\boldsymbol{Y})} \end{aligned}$$

Overlap: A simple way to measure how good the estimate is the *overlap*

$$\boldsymbol{R}_{1,0} = \frac{1}{N}(\hat{\boldsymbol{x}} \cdot \boldsymbol{x}_0) \qquad \hat{\boldsymbol{x}} \sim \mathbb{P}_{\mathrm{mis}}(\cdot \mid \boldsymbol{Y}).$$

Related Work: Pourkamali, Macris ’20, Camilli, Contucci, Mingione ’22, Barbier, Hou, Mondelli, Saenz ’22

Spiked Matrix Example with Mismatched Prior

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

Spiked Matrix Example with Mismatched Prior

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

Likelihood with Mismatch:

$$\mathbb{P}_{\text{mis}}(Y_{ij} \mid \boldsymbol{X}_0 = \boldsymbol{X}) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right) \qquad x_i \sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$$

$$\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{\mathbb{P}_{\text{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})}{\int \mathbb{P}_{\text{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})}$$

Spiked Matrix Example with Mismatched Prior

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

Likelihood with Mismatch:

$$\mathbb{P}_{\text{mis}}(Y_{ij} \mid \boldsymbol{X}_0 = \boldsymbol{X}) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right) \qquad x_i \sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$$

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) &= \frac{\mathbb{P}_{\text{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})}{\int \mathbb{P}_{\text{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})} \\
&= \frac{e^{-\sum_{i<j} \frac{1}{2}(Y_{ij} - \frac{\sqrt{\lambda}}{\sqrt{N}} x_i x_j)^2} d\,\mathbb{P}_X(\boldsymbol{X})}{\int e^{-\sum_{i<j} \frac{1}{2}(Y_{ij} - \frac{\sqrt{\lambda}}{\sqrt{N}} x_i x_j)^2} d\,\mathbb{P}_0(\boldsymbol{X})}.
\end{aligned}$$

Spiked Matrix Example with Mismatched Prior

$$\boldsymbol{Y} = \boldsymbol{G} + \frac{\sqrt{\lambda}}{\sqrt{N}} \boldsymbol{X}_0$$

Likelihood with Mismatch:

$$\mathbb{P}_{\mathrm{mis}}(Y_{ij} \mid \boldsymbol{X}_0 = \boldsymbol{X}) \sim N\left(\sqrt{\lambda}\frac{x_i \cdot x_j}{\sqrt{N}}, 1\right) \qquad x_i \sim \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$$

$$\mathbb{P}(\boldsymbol{X}_0 = d\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{\mathbb{P}_{\mathrm{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})}{\int \mathbb{P}_{\mathrm{mis}}(\boldsymbol{Y} \mid \boldsymbol{X}_0 = \boldsymbol{X}) d\,\mathbb{P}_X(\boldsymbol{X})}$$
$$= \frac{e^{H_N(x) + \cancel{C(Y)}} d\,\mathbb{P}_X(\boldsymbol{X})}{\underbrace{\int e^{H_N(x) + \cancel{C(Y)}} d\,\mathbb{P}_X(\boldsymbol{X})}_{Z_N(Y)}}.$$

Free Energy:

$$F_N(\lambda) = \frac{1}{N}\mathbb{E}_Y \log Z_N(Y) = \frac{1}{N}\mathbb{E}\log \int e^{-\sum_{i<j} \frac{1}{2}(Y_{ij} - \frac{\sqrt{\lambda}}{\sqrt{N}} x_i x_j)^2} d\,\mathbb{P}_X(x).$$

The Limit of the Free Energy

Theorem 4 (Limiting Free Energy (Camilli, Contucci, Mingione))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_x \phi(x).$$

Parisi Functional: Let $\mathcal{P}(\beta, h)$ be the Parisi functional for the SK model with temperature β and external field h,

$$\varphi(q) = -\frac{\lambda x^2}{2} + \mathcal{P}(\sqrt{\lambda}, \lambda x)$$

The Limit of the Free Energy

Theorem 5 (Limiting Free Energy (Camilli, Contucci, Mingione))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_x \phi(x).$$

Parisi Functional: Let $\mathcal{P}(\beta, h)$ be the Parisi functional for the SK model with temperature β and external field h,

$$\varphi(q) = -\frac{\lambda x^2}{2} + \mathcal{P}(\sqrt{\lambda}, \lambda x)$$

Challenges:

1. How is this related to the general inference problem we described before?

The Limit of the Free Energy

Theorem 6 (Limiting Free Energy (Camilli, Contucci, Mingione))

$$\lim_{N\to\infty} F_N(\lambda) = \sup_x \phi(x).$$

Parisi Functional: Let $\mathcal{P}(\beta, h)$ be the Parisi functional for the SK model with temperature β and external field h,

$$\varphi(q) = -\frac{\lambda x^2}{2} + \mathcal{P}(\sqrt{\lambda}, \lambda x)$$

Challenges:

1. How is this related to the general inference problem we described before?
    - Universality of Overlaps

Table of Contents

Exercise: Cramer's Theorem

$X_1, \ldots, X_N$ i.i.d. with sample mean $S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$.

Exercise: Cramer's Theorem

$X_1, \ldots, X_N$ i.i.d. with sample mean $S_N = \frac{1}{N}\sum_{i=1}^N X_i$.
Logarithmic moment generating function

$$\Lambda(\lambda) = \log \mathbb{E} e^{\lambda X_1}$$

Fenchel–Legendre transform

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \Lambda(\lambda))$$

Theorem 7

S_N satisfies the LDP with rate function Λ^* and speed N,

1. *For any closed $F \subset \mathbb{R}$,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(S_N \in F) \leq -\inf_{x\in F} \Lambda^*(x)$$

2. *For any open $G \subset \mathbb{R}$,*

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(S_N \in G) \geq -\inf_{x\in G} \Lambda^*(x).$$

Example: Perfect Information Case

Suppose we have perfect information and $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$. In this case,

$$R_{10} = \frac{1}{N} \sum_{i=1}^{N} (x_i^0)^2.$$

Cramer's theorem implies that for

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - \log \mathbb{E} e^{\lambda x_0^2})$$

and any set A

$$\begin{aligned} -\inf_{x \in A^\circ} \Lambda^*(x) &\leq \liminf_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\boldsymbol{R}_{10} \in A) \\ &\leq \limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\boldsymbol{R}_{10} \in A) \leq -\inf_{x \in \bar{A}} \Lambda^*(x) \end{aligned}$$

Example: Perfect Information Case

Suppose we have perfect information and $\hat{\boldsymbol{x}} = \boldsymbol{x}_0$. In this case,

$$R_{10} = \frac{1}{N}\sum_{i=1}^{N}(x_i^0)^2.$$

Cramer's theorem implies that for

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \log \mathbb{E}e^{\lambda x_0^2})$$

and any set A

$$\begin{aligned} -\inf_{x\in A^\circ}\Lambda^*(x) &\leq \liminf_{N\to\infty}\frac{1}{N}\log\mathbb{P}(\boldsymbol{R}_{10}\in A) \\ &\leq \limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}(\boldsymbol{R}_{10}\in A) \leq -\inf_{x\in\bar{A}}\Lambda^*(x) \end{aligned}$$

Question: Can we generalize this result to the non-trivial cases?

Universality

Consider the log likelihood functions

$$g^0(Y, w) := \ln \frac{d\,\mathbb{P}_{\mathrm{out}}(Y \mid w)}{dY}$$

$$g(Y, w) := \ln \frac{d\,\mathbb{P}_{\mathrm{mis}}(Y \mid w)}{dY}$$

and the corresponding Fisher score parameters

$$\beta = \left[\mathbb{E}_{\mathbb{P}_{\mathrm{out}}(Y|0)}\left[(\partial_w g(Y, 0))^2\right]\right]^{\frac{1}{2}}$$

$$\beta_{SNR} = \mathbb{E}_{\mathbb{P}_{\mathrm{out}}(Y|0)}\left[\partial_w g(Y, 0)\partial_w g^0(Y, 0)\right]$$

$$\beta_S = \mathbb{E}_{\mathbb{P}_{\mathrm{out}}(Y|0)}\left[\partial_w^2 g(Y, 0)\right].$$

If $g^0(Y, w) = g(Y, w)$ then $\beta^2 = \beta_{SNR} = \lambda$ and $\beta_S = -\lambda$

Universality

Likelihood:

$$Z_N^Y(A) = \int \mathbb{1}(\boldsymbol{x} \in A) e^{\sum_{ij} g(Y_{ij}, \frac{x_i x_j}{\sqrt{N}}) - g(Y_{ij}, 0)} \, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

$$\mathbb{G}_N^Y(A) = \frac{Z_N^Y(A)}{Z_N^Y(\mathbb{R})}$$

Gibbs Measure: Let $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$, $W_{ij} \sim N(0,1)$ i.i.d.

$$H_N^{\bar{\beta}}(\boldsymbol{x}) = \sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N}(x_i x_j)(x_i^0 x_j^0) + \frac{\beta_S}{2N}(x_i x_j)^2 \tag{1}$$

$$Z_N^{\bar{\beta}}(A) = \int \mathbb{1}(\boldsymbol{x} \in A) \exp(H_N^{\bar{\beta}}(\boldsymbol{x})) \, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

$$\mathbb{G}_N^{\bar{\beta}}(A) = \frac{Z_N^{\bar{\beta}}(A)}{Z_N^{\bar{\beta}}(\mathbb{R})}$$

Universality

Assumptions:

1. Compact support: $\mathbb{P}_0$ and $\mathbb{P}_X$ are compactly supported
2. Regularity: g, g^0 are three times differentiable (plus some bounds on the derivatives)
3. Consistent estimator: $\mathbb{E}_{Y|0}[\partial_w g(Y, 0)] = 0$

Overlaps: Let $\hat{\boldsymbol{x}}$ be a sample from a Gibbs measure (either $\mathbb{G}_N^Y$ or $\mathbb{G}_N^{\bar{\beta}}$),

$$R_{10} = \frac{\hat{\boldsymbol{x}} \cdot \boldsymbol{x}_0}{N} \qquad R_{11} = \frac{\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}}{N}$$

Theorem 8 (Universality of the Overlaps)

If $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$ corresponds to the Fisher score parameters, then the joint law of the overlaps (R_{10}, R_{11}) under $\mathbb{G}_N^Y$ and $\mathbb{G}_N^{\bar{\beta}}$ satisfy the same almost sure large deviations principle. Furthermore,

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E} \log Z_N^Y = \lim_{N\to\infty} \frac{1}{N}\mathbb{E} \log Z_N^{\bar{\beta}}.$$

Parisi Type Functional

- Let $\zeta(t)$ be a c.d.f.
- Let $\Phi_\zeta(t, y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t, y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1, y) = \log \int e^{yx + \lambda x + \mu x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

Define the functional

$$\varphi_{\bar{\beta}}(S, M) = \inf_{\mu,\lambda,\zeta} \Bigg(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \frac{\beta^2}{2} \int t\zeta(t)\, dt - \mu S - \lambda M + \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4} \Bigg).$$

Almost Sure Large Deviations Principle

Domain of Overlaps: For any ρ, t

$$\mathbb{E}_{x^0}[\mathrm{essinf}_x\{\rho x^2 + t x x^0\}] \le \rho \boldsymbol{R}_{11} + t \boldsymbol{R}_{10} \le \mathbb{E}_{x^0}[\mathrm{esssup}_x\{\rho x^2 + t x x^0\}]$$

Define

$$\mathcal{C} = \bigcap_{\rho,t\in[-1,1]^2} \Big\{(S, M) : \mathbb{E}_{x^0}[\mathrm{essinf}_x\{\rho x^2 + t x x^0\}] \le$$
$$\le \rho S + t M \le \mathbb{E}_{x^0}[\mathrm{esssup}_x\{\rho x^2 + t x x^0\}]\Big\}.$$

Almost Sure Large Deviations Principle

Theorem 9 (Almost Sure LDP)

For all real numbers $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$, the law of $(R_{1,1}, R_{1,0})$ under $\mathbb{G}_N^{\bar{\beta}}$ satisfies an almost sure large deviation principle with speed N and good rate function $I_{\bar{\beta}}^{FP}$ which is infinite if (S, M) do not belong to $\mathcal{C}$ and otherwise is given by

$$I_{\bar{\beta}}^{FP}(S, M) = -\varphi_{\bar{\beta}}(S, M) + \sup_{(s,m)\in\mathcal{C}} \varphi_{\bar{\beta}}(s, m)\,.$$

In other words,

- *for any closed subset F of $\mathbb{R}^2$, for almost all $(W, \boldsymbol{x}^0)$,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{G}_N^{\bar{\beta}}((R_{1,1}, R_{1,0}) \in F) \leq - \inf_{(S,M)\in F} I_{\bar{\beta}}^{FP}(S, M)$$

- *for any open subset O of $\mathbb{R}^2$, for almost all $(W, \boldsymbol{x}^0)$,*

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{G}_N^{\bar{\beta}}((R_{1,1}, R_{1,0}) \in O) \geq - \inf_{(S,M)\in O} I_{\bar{\beta}}^{FP}(S, M)\,.$$

Corollary 1 (LDP for the Overlaps)

For Fisher parameters $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$, the law of $R_{1,0}$ under $\mathbb{G}_N^Y$ satisfies an almost sure large deviation principle with speed N and good rate function $I_{\bar{\beta}}^{FP}$ which is infinite if (S, M) do not belong to $\mathcal{C}$ and otherwise is given by

$$I_{\bar{\beta}}^{FP}(S, M) = -\varphi_{\bar{\beta}}(S, M) + \sup_{(s,m)\in\mathcal{C}} \varphi_{\bar{\beta}}(s, m)\,.$$

In other words,

- *for any closed subset F of $\mathbb{R}$, for almost all $\boldsymbol{Y}$,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{G}_N^Y(R_{1,0} \in F) \leq -\inf_{M\in F} \inf_{S} I_{\bar{\beta}}^{FP}(S, M)$$

- *for any open subset O of $\mathbb{R}$, for almost all $\boldsymbol{Y}$,*

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{G}_N^Y(R_{1,0} \in O) \geq -\inf_{M\in O} \inf_{S} I_{\bar{\beta}}^{FP}(S, M)\,.$$

Limit of the Free Energy and Overlap Concentration

Corollary 2 (Limit of the Free Energy)

For any real numbers $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$*,*

$$\lim_{N\to\infty} F_N(\bar{\beta}) = \sup_{(s,m)\in\mathcal{C}} \varphi_{\bar{\beta}}(s,m).$$

For Fisher parameters $\bar{\beta} = (\beta, \beta_{SNR}, \beta_S)$

$$\lim_{N\to\infty} F_N(Y) = \sup_{(s,m)\in\mathcal{C}} \varphi_{\bar{\beta}}(s,m).$$

Corollary 3 (Overlap Concentration)

If $I_{\bar{\beta}}^{FP}$ *has a unique minimizer* $(S_{\bar{\beta}}, M_{\bar{\beta}})$ *then* (R_{11}, R_{10}) *converges to* $(S_{\bar{\beta}}, M_{\bar{\beta}})$ *almost surely.*

Table of Contents

Recall:

$$H_N^{\bar{\beta}}(\boldsymbol{x}) = \sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N}(x_i x_j)(x_i^0 x_j^0) + \frac{\beta_S}{2N}(x_i x_j)^2 \tag{2}$$

$$Z_N^{\bar{\beta}}(A) = \int \mathbb{1}((R_{11}, R_{10}) \in A) \exp(H_N^{\bar{\beta}}(\boldsymbol{x}))\, d\, \mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

$$\mathbb{G}_N^{\bar{\beta}}((R_{11}, R_{10}) \in A) = \frac{Z_N^{\bar{\beta}}(A)}{Z_N^{\bar{\beta}}(\mathbb{R})}$$

Large Deviations and Free Energies

$$\frac{1}{N} \log \mathbb{G}_N^{\bar{\beta}}((R_{11}, R_{10}) \in A) = \frac{1}{N} \log Z_N^{\bar{\beta}}(A) - \frac{1}{N} \log Z_N^{\bar{\beta}}(\mathbb{R})$$

Large Deviations Principle and Spin Glasses

Recall:

$$H_N^{\bar\beta}(\boldsymbol{x}) = \sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N}(x_i x_j)(x_i^0 x_j^0) + \frac{\beta_S}{2N}(x_i x_j)^2 \tag{2}$$

$$Z_N^{\bar\beta}(A) = \int \mathbb{1}((R_{11}, R_{10}) \in A) \exp(H_N^{\bar\beta}(\boldsymbol{x}))\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

$$\mathbb{G}_N^{\bar\beta}((R_{11}, R_{10}) \in A) = \frac{Z_N^{\bar\beta}(A)}{Z_N^{\bar\beta}(\mathbb{R})}$$

Large Deviations and Free Energies

$$\frac{1}{N}\log \mathbb{G}_N^{\bar\beta}((R_{11}, R_{10}) \in A) = \frac{1}{N}\log Z_N^{\bar\beta}(A) - \frac{1}{N}\log Z_N^{\bar\beta}(\mathbb{R})$$

Comments:

- By universality, the LDP for $\mathbb{G}_N^{\bar\beta}$ will extend to $\mathbb{G}_N^{Y}$.

Large Deviations Principle and Spin Glasses

Recall:

$$H_N^{\bar{\beta}}(\boldsymbol{x}) = \sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N}(x_i x_j)(x_i^0 x_j^0) + \frac{\beta_S}{2N}(x_i x_j)^2 \qquad (2)$$

$$Z_N^{\bar{\beta}}(A) = \int \mathbb{1}((R_{11}, R_{10}) \in A) \exp(H_N^{\bar{\beta}}(\boldsymbol{x}))\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

$$\mathbb{G}_N^{\bar{\beta}}((R_{11}, R_{10}) \in A) = \frac{Z_N^{\bar{\beta}}(A)}{Z_N^{\bar{\beta}}(\mathbb{R})}$$

Large Deviations and Free Energies

$$\frac{1}{N}\mathbb{E}\log \mathbb{G}_N^{\bar{\beta}}((R_{11}, R_{10}) \in A) = \frac{1}{N}\mathbb{E}\log Z_N^{\bar{\beta}}(A) - \frac{1}{N}\mathbb{E}\log Z_N^{\bar{\beta}}(\mathbb{R})$$

Comments:

- By universality, the LDP for $\mathbb{G}_N^{\bar{\beta}}$ will extend to $\mathbb{G}_N^{Y}$.
- The free energies concentrate on its expected value*

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx+\mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases} .$$

- Limit of the Franz–Parisi Potential:

$$\varphi_{\bar\beta}(S,M) = \inf_{\mu,\lambda,\zeta} \Bigg(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \frac{\beta^2}{2} \int t\zeta(t)\,dt - \mu S - \lambda M + \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4} \Bigg).$$

- Rate Function:

$$I_{\bar\beta}^{FP}(S,M) = -\varphi_{\bar\beta}(S,M) + \sup_{(s,m)\in\mathcal{C}} \varphi_{\bar\beta}(s,m)\,.$$

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx + \mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

- Limit of the Franz–Parisi Potential:

$$\begin{aligned} \varphi_{\bar\beta}(S,M) &= \inf_{\mu,\lambda,\zeta} \left(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \frac{\beta^2}{2} \int t\zeta(t)\,dt - \mu S - \lambda M \right. \\ &\qquad \left. + \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4} \right) \\ &= \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z_N^{\bar\beta}(A). \end{aligned}$$

- Rate Function:

$$I_{\bar\beta}^{FP}(S,M) = -\underbrace{\varphi_{\bar\beta}(S,M)}_{\frac{1}{N}\mathbb{E}\log Z_N^{\bar\beta}(A)} + \underbrace{\sup_{(s,m)\in\mathcal{C}} \varphi_{\bar\beta}(s,m)}_{-\frac{1}{N}\mathbb{E}\log Z_N^{\bar\beta}(\mathbb{R})}.$$

Varadhan's Lemma

- By Varadhan's Lemma, it suffices to take $\beta_{SNR}, \beta_S = 0$.
- Define

$$Z_N(A) = \int \mathbb{1}(B_\epsilon(S, M)) \exp\left(\sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j \right) d\, \mathbb{P}_X^{\otimes N}(\boldsymbol{x}).$$

where $\{(R_{11}, R_{10}) \in B_\epsilon(S, M)\} = \{|R_{11} - S| \leq \epsilon, |R_{10} - M| \leq \epsilon\}$

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx + \mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

Limit of the Franz–Parisi Potential:

$$\frac{1}{N}\mathbb{E} \log Z_N^{\bar{\beta}}(A) \to \inf_{\mu,\lambda,\zeta} \left(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \frac{\beta^2}{2} \int t\zeta(t)\, dt - \mu S - \lambda M + \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4} \right).$$

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx + \mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

Limit of the Franz–Parisi Potential:

$$\frac{1}{N}\mathbb{E}\log Z_N^{\bar{\beta}}(A) \to \inf_{\mu,\lambda,\zeta}\left(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \frac{\beta^2}{2}\int t\zeta(t)\,dt - \mu S - \lambda M \underbrace{+\frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4}}_{\text{Varadhan's Lemma}}\right).$$

Remaining terms

$$\frac{1}{N}\mathbb{E}\log Z_N(B_\epsilon(S,M))$$

Cavity Method

Step 1: Cavity Method / Aizenman–Sims–Starr Scheme

$$
\begin{aligned}
&\frac{1}{N}\mathbb{E}\log Z_N(B_\epsilon(S,M)) \\
&\approx \frac{1}{n}\left(\mathbb{E}\log Z_{N+n}(B_\epsilon^{N,n}(S,M)) - \mathbb{E}\log Z_N(B_\epsilon(S,M))\right)
\end{aligned}
$$

Cavity Method

Step 2: Cavity Fields $(x, y) \in \mathbb{R}^{N+n}$

$$\frac{1}{N}\mathbb{E}\log Z_N(B_\epsilon(S,M))$$
$$\approx \frac{1}{n}\mathbb{E}\log\left\langle \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(x)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y)\right\rangle' - \frac{1}{n}\mathbb{E}\log\left\langle e^{\beta\sqrt{n}Y(x)}\right\rangle'$$

where $(Z_i)_{i\le N}$ and Y are independent Gaussian processes with covariance

$$\mathbb{E}Z_i(x^1)Z_j(x^2) = \delta_{i=j}R_{1,2} \qquad \mathbb{E}Y_i(x^1)Y_i(x^2) = \frac{1}{2}R_{12}^2$$

and $\langle\cdot\rangle'$ are the averages with respect to the restricted Gibbs measure

$$G_N'(x) \propto \int_{B_\epsilon(S,M)} e^{\beta H_N'(x)}\, d\,\mathbb{P}_X^{\otimes N}(x)$$

Cavity Method

Step 3: Regularization + Ghirlanda–Guerra identities + ultrametricity

$$\frac{1}{N}\mathbb{E}\log Z_N(B_\epsilon(S,M))$$
$$\approx \frac{1}{n}\mathbb{E}\log\sum_\alpha v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(\alpha)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y) - \frac{1}{n}\mathbb{E}\log\sum_\alpha v_\alpha e^{\beta\sqrt{n}Y(\alpha)}$$

where $(Z_i)_{i\le N}$ and Y are independent Gaussian processes with covariance

$$\mathbb{E}Z_i(\alpha^1)Z_j(\alpha^2) = \delta_{i=j}Q_{\alpha^1\wedge\alpha^2} \qquad \mathbb{E}Y_i(\alpha^1)Y_i(\alpha^2) = \frac{1}{2}Q^2_{\alpha^1\wedge\alpha^2}$$

and v_α are the weights of the Ruelle–Probability–Cascades encoding the limiting law of the overlaps encoded by the c.d.f $\zeta(t)\approx\mathbb{E}\langle\mathbb{1}(R_{12}\le t)\rangle'$.

Cavity Method

Step 4: Compute using Ruelle–Probability–Cascades

$$\frac{1}{N}\mathbb{E}\log Z_N(B_\epsilon(S,M))$$
$$\approx \frac{1}{n}\mathbb{E}\log\sum_\alpha v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(\alpha)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y) - \frac{\beta^2}{2}\int t\zeta(t)\,dt$$

where the ζ is the c.d.f. of the limiting overlap $\zeta(t)\approx\mathbb{E}\langle\mathbb{1}(R_{12}\le t)\rangle'$.

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx + \mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

Limit of the Franz–Parisi Potential:

$$\frac{1}{N}\mathbb{E}\log Z_N^{\bar\beta}(A) \to \inf_{\mu,\lambda,\zeta}\Bigg(\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)] - \underbrace{\frac{\beta^2}{2}\int t\zeta(t)\,dt}_{\text{cavity computation II}} - \mu S - \lambda M$$

$$\underbrace{+\frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4}}_{\text{Varadhan's Lemma}}\Bigg).$$

Remaining Terms:

$$\frac{1}{n}\mathbb{E}\log \sum_\alpha v_\alpha \int_{B_\epsilon(S,M)} e^{\beta \sum_{i\le n} Z_i(\alpha) y_i} \, d\,\mathbb{P}_X^{\otimes n}(y)$$

Cramer's Tilting Argument

Lemma 1 (Large Deviations of Cavity I)

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\log\sum_{\alpha} v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(\alpha)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y)$$
$$= \inf_{\mu,\lambda}\left(\mathbb{E}\log\sum_{\alpha} v_\alpha \int e^{\beta Z(\alpha)y+\mu x x_0+\lambda x^2}\, d\,\mathbb{P}_X(y) - \mu S - \lambda M\right)$$

Cramer's Tilting Argument

Lemma 2 (Large Deviations of Cavity I)

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\log\sum_{\alpha} v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(\alpha)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y)$$

$$= \inf_{\mu,\lambda}\left(\mathbb{E}\log\sum_{\alpha} v_\alpha \int e^{\beta Z(\alpha)y+\mu x x_0+\lambda x^2}\, d\,\mathbb{P}_X(y) - \mu S - \lambda M\right)$$

Intuition: Cramer's Theorem

$$\frac{1}{n}\mathbb{E}\log\int \mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i = C\right) d\,\mathbb{P}_X^{\otimes n}(x) = \log\int e^{\lambda x}\, d\,\mathbb{P}_X(x) - \lambda C$$

Cramer's Tilting Argument

Lemma 3 (Large Deviations of Cavity I)

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\log\sum_{\alpha} v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le n} Z_i(\alpha)y_i}\, d\,\mathbb{P}_X^{\otimes n}(y)$$
$$= \inf_{\mu,\lambda}\left(\underbrace{\mathbb{E}\log\sum_{\alpha} v_\alpha \int e^{\beta Z(\alpha)y+\mu x x_0+\lambda x^2}\, d\,\mathbb{P}_X(y)}_{\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)]} - \mu S - \lambda M\right)$$

Intuition: Cramer's Theorem

$$\frac{1}{n}\mathbb{E}\log\int \mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i = C\right) d\,\mathbb{P}_X^{\otimes n}(x) = \log\int e^{\lambda x}\, d\,\mathbb{P}_X(x) - \lambda C$$

Recall the Franz–Parisi Functional

- Order Parameter: $\zeta : [0,1] \mapsto [0,1]$ c.d.f.
- Parisi PDE: Let $\Phi_\zeta(t,y)$ is the solution to Parisi's PDE

$$\begin{cases} \partial_t \Phi_\zeta = -\frac{1}{4}(\partial_y^2 \Phi_\zeta + \zeta([0,t])(\partial_y \Phi_\zeta)^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi_\zeta(1,y) = \log \int e^{yx + \mu x x_0 + \lambda x^2} \, d\,\mathbb{P}_X(x) \end{cases}.$$

Limit of the Franz–Parisi Potential:

$$\frac{1}{N} \mathbb{E} \log Z_N^{\bar{\beta}}(A) \to \inf_{\mu,\lambda,\zeta} \Bigg(\underbrace{\mathbb{E}_0[\Phi_{\lambda,\mu,\zeta}(0,0)]}_{\text{Cramer's Tilting}} - \underbrace{\frac{\beta^2}{2} \int t\zeta(t)\, dt}_{\text{cavity computation II}} - \underbrace{\mu S - \lambda M}_{\text{Cramer's Tilting}}$$

$$\underbrace{+ \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4}}_{\text{Varadhan's Lemma}} \Bigg).$$

Large Deviations Lower Bound: Main Challenges

1. We don't know the limiting law of the overlaps, so we have to express it in terms of a variational formula.

Large Deviations Lower Bound: Main Challenges

1. We don't know the limiting law of the overlaps, so we have to express it in terms of a variational formula.
2. Unlikely values of x^0 can have large influences on the entropy $|\{(R_{10}, R_{11}) \in A\}|$

Large Deviations Lower Bound: Main Challenges

1. We don't know the limiting law of the overlaps, so we have to express it in terms of a variational formula.
2. Unlikely values of x^0 can have large influences on the entropy $|\{(R_{10}, R_{11}) \in A\}|$
3. The region of integration $\mathbb{1}((R_{10}, R_{11}) \in A)$ is not smooth with respect to the signal x^0

Large Deviations Lower Bound: Main Challenges

1. We don't know the limiting law of the overlaps, so we have to express it in terms of a variational formula.
2. Unlikely values of x^0 can have large influences on the entropy $|\{(R_{10}, R_{11}) \in A\}|$
3. The region of integration $\mathbb{1}((R_{10}, R_{11}) \in A)$ is not smooth with respect to the signal x^0
4. The Nishimori identity fails, so we also need to know the law of $\boldsymbol{R}_{12}$.

Large Deviations Lower Bound: Main Challenges

1. We don't know the limiting law of the overlaps, so we have to express it in terms of a variational formula.
2. Unlikely values of x^0 can have large influences on the entropy $|\{(R_{10}, R_{11}) \in A\}|$
3. The region of integration $\mathbb{1}((R_{10}, R_{11}) \in A)$ is not smooth with respect to the signal x^0
4. The Nishimori identity fails, so we also need to know the law of $\boldsymbol{R}_{12}$.
5. Need uniform bounds to extend to an almost sure result (difficulty on the boundary of feasible values $\partial\mathcal{C}$)

Large Deviations: Lower Bound

Strategy: Based on LDP for the overlaps in Panchenko '15

1. Regularize the free energy by smoothing the indicator and localizing around typical values of $\boldsymbol{x}^0$.

Large Deviations: Lower Bound

Strategy: Based on LDP for the overlaps in Panchenko '15

1. Regularize the free energy by smoothing the indicator and localizing around typical values of $\boldsymbol{x}^0$.
2. Show that we can add a perturbation to the free energy to enforce ultrametricity

Large Deviations: Lower Bound

Strategy: Based on LDP for the overlaps in Panchenko '15

1. Regularize the free energy by smoothing the indicator and localizing around typical values of $\boldsymbol{x}^0$.
2. Show that we can add a perturbation to the free energy to enforce ultrametricity
3. Use interpolation to prove an upper bound and use the cavity computations to prove a matching lower bound (we don't know ζ)

Large Deviations: Lower Bound

Strategy: Based on LDP for the overlaps in Panchenko '15

1. Regularize the free energy by smoothing the indicator and localizing around typical values of $\boldsymbol{x}^0$.
2. Show that we can add a perturbation to the free energy to enforce ultrametricity
3. Use interpolation to prove an upper bound and use the cavity computations to prove a matching lower bound (we don't know ζ)
4. Use uniform bounds to go from a quenched LDP to an almost sure LDP

(Quenched) Universality

Compare

$$F_N(Y,A) = \frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} g(Y_{ij},\frac{x_i x_j}{\sqrt{N}}) - g(Y_{ij},0)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

and

$$F_N(\bar{\beta},A) = \frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{H_N^{\bar{\beta}}(x)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right).$$

(Quenched) Universality

$$F_N(Y,A) = \frac{1}{N}\left(\mathbb{E}\Big(\log\int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} g(Y_{ij},\frac{x_i x_j}{\sqrt{N}})-g(Y_{ij},0)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

$$\downarrow$$

$$\frac{1}{N}\left(\mathbb{E}\Big(\log\int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} \partial_w g(Y_{ij},0)\frac{x_i x_j}{\sqrt{N}}+\partial_w^2 g(Y_{ij},0)\frac{(x_i x_j)^2}{2N}}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

Step 1: Expand g in the second variable

$$g\Big(Y_{ij},\frac{x_i x_j}{\sqrt{N}}\Big)-g\Big(Y_{ij},0\Big) = \partial_w g(Y_{ij},0)\frac{x_i x_j}{\sqrt{N}}+\partial_w^2 g(Y_{ij},0)\frac{(x_i x_j)^2}{2N}+o(N^{-1})$$

(Quenched) Universality

$$F_N(Y,A) = \frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} g(Y_{ij},\frac{x_i x_j}{\sqrt{N}}) - g(Y_{ij},0)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

$$\downarrow$$

$$\frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} \partial_w g(Y_{ij},0)\frac{x_i x_j}{\sqrt{N}} + \beta_S \frac{(x_i x_j)^2}{2N}}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

Step 2: By concentration,

$$g\Big(Y_{ij},\frac{x_i x_j}{\sqrt{N}}\Big) - g\Big(Y_{ij},0\Big) = \partial_w g(Y_{ij},0)\frac{x_i x_j}{\sqrt{N}} + \underbrace{\mathbb{E}[\partial_w^2 g(Y_{ij},0)]}_{\beta_S}\frac{(x_i x_j)^2}{2N} + o(N^{-1})$$

(Quenched) Universality

$$F_N(Y, A) = \frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11}, R_{10}) \in A} e^{\sum_{i<j} g(Y_{ij}, \frac{x_i x_j}{\sqrt{N}}) - g(Y_{ij}, 0)} \, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

$$\downarrow$$

$$\frac{1}{N}\left(\mathbb{E}\Big(\log \int_{(R_{11}, R_{10}) \in A} e^{\sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N}(x_i x_j)(x_i^0 x_j^0) + \beta_S \frac{(x_i x_j)^2}{2N}} \, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\right)$$

Step 3: Under the consistent estimate assumption

$$\mathbb{E}[\partial_w g(Y_{ij}, 0)] = \frac{\beta_{SNR}}{\sqrt{N}} x_i^0 x_j^0 + O(N^{-1}) \quad \mathrm{Var}[\partial_w g(Y_{ij}, 0)] = \beta^2 + O(N^{-1/2}).$$

Use universality of spin glasses to replace $\partial_w g(Y_{ij}, 0)$ with a Gaussian with variance β^2 and mean $\frac{\beta_{SNR}}{\sqrt{N}} x_i^0 x_j^0$.

(Quenched) Universality

$$F_N(Y,A) = \frac{1}{N}\Big(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{\sum_{i<j} g(Y_{ij},\frac{x_i x_j}{\sqrt{N}}) - g(Y_{ij},0)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\Big)$$

$$\downarrow$$

$$F_N(\bar\beta,A) = \frac{1}{N}\Big(\mathbb{E}\Big(\log \int_{(R_{11},R_{10})\in A} e^{H_N^{\bar\beta}(x)}\, d\,\mathbb{P}_X^{\otimes N}(\boldsymbol{x})\Big)\Big).$$

Step 3: Under the consistent estimator assumption

$$\mathbb{E}[\partial_w g(Y_{ij},0)] = \frac{\beta_{SNR}}{\sqrt{N}} x_i^0 x_j^0 + O(N^{-1}) \quad \mathrm{Var}[\partial_w g(Y_{ij},0)] = \beta^2 + O(N^{-1/2}).$$

Use universality in disorder for spin glasses to replace $\partial_w g(Y_{ij},0)$ with a Gaussian with variance β^2 and mean $\frac{\beta_{SNR}}{\sqrt{N}} x_i^0 x_j^0$.

Table of Contents

Open Problems

1. Generalization to higher rank models or rectangular models. (Miolane et al)
2. Generalization to higher tensor estimation. (Barbier et al)
3. Generalization to entry wise dependent output channels (Reeves et al, Camilli et al)
4. Generalizations to temperature chaos (Chen, Panchenko, Subag)
5. Studying phase transitions for these models (Auffinger et al, Jagannath et al)

Thank you!