About non-Bayesian optimal inference

Alice Guionnet

CNRS and Ens de Lyon

Elisabeth's birthday

May 31, 2023

Joint work with Justin Ko, Lenka Zdeborova and Florent Krzakala

Matrix Estimation

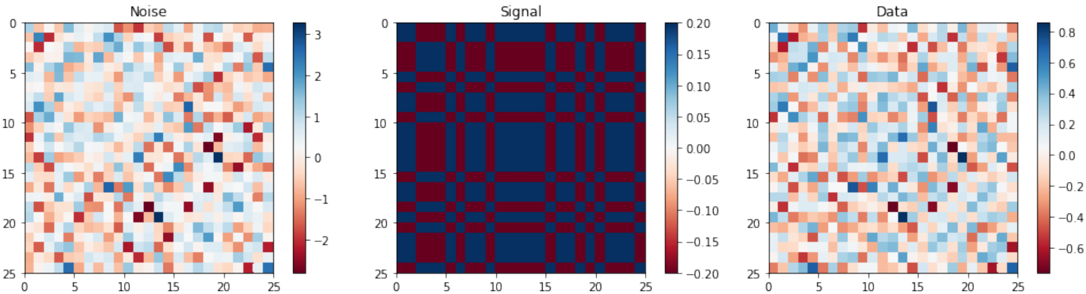
Observe a $N \times N$ symmetric matrix Y given by

$$\mathbf{Y} = \mathbf{G} + \mathbf{X}$$

where $\mathbf{G}$ is a noise. $\mathbf{X}$ is the signal that we will assume one dimensional

$$\mathbf{X} = \rho u u^T$$

How can we estimate the signal $\mathbf{X}$ when the dimension N goes to infinity ?



Asymptotic of the spectrum of the noise matrix

Take **G** to be symmetric, with centered independent entries with covariance $1/N$ and let $(\lambda_i)_{1\leq i\leq N}$ be its eigenvalues.

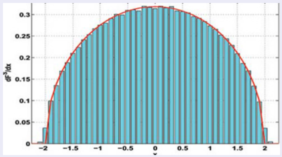
Theorem (Wigner '56 , Komlos-Furedi 81')

Almost surely, for any $a < b$

$$\lim_{N\to\infty} \frac{1}{N}\#\{i : \lambda_i \in [a,b]\} = \sigma([a,b])$$

with σ the semi-circle law :

$$\sigma(dx) = \frac{1}{2\pi}\sqrt{4-x^2}dx$$



If $E[|X_{ij}|^{4+\epsilon}] < \infty$, the eigenvalues stick to the bulk :

$$\lim_{N\to\infty} \max_{1\leq i\leq N} \lambda_i = 2 \qquad a.s$$

The Baik-Ben Arous-Péché phase transition '05

Consider a $N \times N$ random matrix with independent centered entries with covariance

$$\mathbf{Y} = \mathbf{G} + \rho u u^T$$

where u is a unit vector and $\rho > 0$.

- If $\rho < 1$, the largest eigenvalue converges almost surely towards 2, as when $\rho = 0$
- If $\rho > 1$, the largest eigenvalue converges almost surely towards $\rho + \rho^{-1}$.

The Baik-Ben Arous-Péché phase transition '05

Consider a $N \times N$ random matrix with independent centered entries with covariance

$$\mathbf{Y} = \mathbf{G} + \rho u u^T$$

where u is a unit vector and $\rho > 0$.

- If $\rho < 1$, the largest eigenvalue converges almost surely towards 2, as when $\rho = 0$
- If $\rho > 1$, the largest eigenvalue converges almost surely towards $\rho + \rho^{-1}$.

If $\rho > 1$, the signal can be detected by the largest eigenvalue and moreover Benaych-George-Rao '12 showed that it can be weakly recovered in the sense that the eigenvector v corresponding to the largest eigenvalue is such that

$$\langle u, v \rangle^2 \to c(\rho) > 0\,.$$

Detection

$$\mathbf{Y} = \mathbf{G} + \rho u u^T$$

- If $\rho > 1$, signal can be detected and recovered by Principal Component Analysis (cf Test by P. Bianchi, M. Debbah, M. Maida, J.Najim '11)
- If $\rho < 1$ and the noise is Gaussian, %beginitemize No test based on the eigenvalues can reliably detect the signal (Montanari, Reichman, Zeitouni '17)

Minimal Mean Squared Error

$$\mathbf{Y} = \mathbf{G} + \mathbf{X}_0$$

Beyond PCA, a natural approach to estimate $\mathbf{X}$ is to minimize the mean square error

$$\begin{aligned}\mathrm{MMSE_N} &= \min_\theta \frac{1}{N^2}\mathbb{E}\mathrm{Tr}(\mathbf{X} - \theta(\mathbf{Y}))^2 \\ &= \frac{1}{N^2}\mathbb{E}\mathrm{Tr}(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^2.\end{aligned}$$

because the minimum is achieved at $\theta(\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$

How can we estimate $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ and MMSE for N large ?

Estimating $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$

Assume G_{ij} is an array of independent variables with law $e^{g(y)}dy/Z$. Then with $X_{ij} = \sqrt{N}\rho u_i u_j$,

$$Y_{ij} - \omega_{ij} = G_{ij} \Rightarrow \mathbb{P}(Y_{ij}|\mathbf{X}) \sim e^{g(y-X_{ij})}dy$$

By Bayes Theorem, if the law of $\mathbf{X}$ is known and equal to $\mathbb{P}_X$

$$\mathbb{P}(\mathbf{X}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{X})\mathbb{P}_X(\mathbf{X})}{\int \mathbb{P}(\mathbf{Y}|\mathbf{X})d\mathbb{P}_X(\mathbf{X})}$$

$$= \frac{e^{\sum_{i<j} g(Y_{ij}-X_{ij})}d\mathbb{P}_X(\mathbf{X})}{\underbrace{\int e^{\sum_{i<j} g(Y_{ij}-X_{ij})}d\mathbb{P}_X(\mathbf{X})}_{Z_N(Y)}}.$$

Estimating $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$

Assume G_{ij} is an array of independent variables with law $e^{g(y)}dy/Z$. Then with $X_{ij} = \sqrt{N}\rho u_i u_j$,

$$Y_{ij} - \omega_{ij} = G_{ij} \Rightarrow \mathbb{P}(Y_{ij}|\mathbf{X}) \sim e^{g(y - X_{ij})}dy$$

By Bayes Theorem, if the law of $\mathbf{X}$ is known and equal to $\mathbb{P}_X$

$$\mathbb{P}(\mathbf{X}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{X})\mathbb{P}_X(\mathbf{X})}{\int \mathbb{P}(\mathbf{Y}|\mathbf{X})d\mathbb{P}_X(\mathbf{X})}$$
$$= \frac{e^{\sum_{i<j} g(Y_{ij} - X_{ij})}d\mathbb{P}_X(\mathbf{X})}{\underbrace{\int e^{\sum_{i<j} g(Y_{ij} - X_{ij})}d\mathbb{P}_X(\mathbf{X})}_{Z_N(Y)}}.$$

If $g(x) = -x^2/2$,

$$\partial_{Y_{ij}} \ln Z_N(\mathbf{Y}) = (\mathbb{E}[X_{ij}|\mathbf{Y}] - Y_{ij})$$

and one can retrieve the MMSE from the typical behavior of $\mathbf{X}$ under the above measure. One can also retrieve the mutual information from the free energy $\mathbb{E}_Y \log Z_N(Y)$.

The free energy

Assume G_{ij} is an array of independent variables with law $e^{g(y)}dy/Z$ and set $X_{ij} = \sqrt{N}\rho u_i u_j$, One wants to estimate

$$F_N = \frac{1}{N}\mathbb{E}_Y \log Z_N(Y) = \frac{1}{N}\mathbb{E}_{\mathbf{Y}} \log \int e^{\sum_{i,j} g(Y_{ij}-X_{ij})^2} d\mathbb{P}_X(\mathbf{X}).$$

Y is assumed to be distributed according to $\mathbf{G}_0 + \rho N^{-1/2} x_0 x_0^T$.

The free energy

Assume G_{ij} is an array of independent variables with law $e^{g(y)}dy/Z$ and set $X_{ij} = \sqrt{N}\rho u_i u_j$, One wants to estimate

$$F_N = \frac{1}{N}\mathbb{E}_Y \log Z_N(Y) = \frac{1}{N}\mathbb{E}_{\mathbf{Y}} \log \int e^{\sum_{i,j} g(Y_{ij}-X_{ij})^2} d\mathbb{P}_X(\mathbf{X}).$$

$\mathbf{Y}$ is assumed to be distributed according to $\mathbf{G}_0 + \rho N^{-1/2} x_0 x_0^T$.

If the signal has rank one, and $g(x) = -x^2/2$ then F_N ressembles the free energy of spin glasses that can be estimated by Guerra-Talagrand's techniques :
if $X_{ij} = N^{1/2}\rho u_i u_j$, $u_i = N^{-1/2}x_i$, x_i iid law $\mathbb{P}_X$,

$$F_N = \frac{1}{N}\mathbb{E}_{\mathbf{Y}} \log \int e^{\rho N^{-1/2}\langle \mathbf{x}, \mathbf{Y}\mathbf{x}\rangle} d\mathbb{P}_X^{\otimes N}(\mathbf{x}) + C$$

Mutual information in the homogeneous low rank case

Consider the Bayes optimal setting

$$\mathbf{Y} = \mathbf{G}_0 + \frac{\rho}{N}\mathbf{x}_0\mathbf{x}_0^T$$

where $\mathbf{G}_0$ follows, as $\mathbf{G}$, the Gaussian law ($g(x) = -x^2/2$) and the $x_0(i)$ iid with law $\mathbb{P}_0 = \mathbb{P}_X$ centered.

Theorem (Lelarge-Miolane '17)

$$\lim_{N\to\infty} F_N = \frac{\rho}{4}\mathbb{E}_{\mathbb{P}_0}[x^2] - \sup_{q\geq 0}\mathcal{F}(\rho, q)$$

The supremum in q is achieved at $q^*(\rho) = \max(0, 1-1/\rho)$ and $\mathcal{F}(\rho, 0) = 0$. Moreover :

$$\lim_{N\to\infty} MMSE_N(\rho) = E_{P_0}[X^2]^2 - q^*(\rho)^2 .$$

As for the BBP transition, the transition occurs at $\rho = 1$.

Here $\mathcal{F}(s, q)$ is the Parisi functional :

$$\mathcal{F}(s, q) = -\frac{\rho^2}{4}q^2 + \mathbb{E}_{\substack{Z\simeq N(0,1)\\ X\simeq\mathbb{P}_0}}[\log\int d\mathbb{P}_0(X)\exp\{\rho\sqrt{q}Zx + \rho qxX - \frac{\rho}{2}qx^2\}]$$

Inhomogeneous Low Rank estimation $\mathbf{Y} = \Delta \odot \mathbf{G} + \mathbf{X}, \mathbf{X} = xx^T$

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

Theorem (Barbier-Reeves'20, Behne-Reeves '22, Ko-G-Zdeborova-Krzakala '22)

Assume $\Delta_{ij} = \Delta_{st}$ for $i, j \in I_s \times I_t$, $|I_s|/N \mapsto \alpha_s, 1 \le i, j \le n$ so that $(\Delta_{s,t}^{-1})_{s,t} \ge 0$ and $\mathbb{P}_0 = \mathbb{P}_X$. Then,

$$\lim_{N\to\infty} F_N(\Delta) = \sup_Q \varphi_\Delta(\mathbf{Q}).$$

Inhomogeneous Low Rank estimation $\mathbf{Y} = \Delta \odot \mathbf{G} + \mathbf{X}, \mathbf{X} = xx^T$

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

Theorem (Barbier-Reeves'20, Behne-Reeves '22, Ko-G-Zdeborova-Krzakala '22)

Assume $\Delta_{ij} = \Delta_{st}$ for $i,j \in I_s \times I_t$, $|I_s|/N \mapsto \alpha_s, 1 \leq i,j \leq n$ so that $(\Delta_{s,t}^{-1})_{s,t} \geq 0$ and $\mathbb{P}_0 = \mathbb{P}_X$. Then,

$$\lim_{N\to\infty} F_N(\Delta) = \sup_Q \varphi_\Delta(\mathbf{Q}).$$

- *If* $\left\|\sqrt{\alpha}\frac{1}{\Delta^2}\sqrt{\alpha}\right\|_{op} < \frac{1}{9d^4C^6}$ *then* $\lim_{N\to\infty}\mathrm{MMSE} = \mathbb{E}_X\|xx^T\|_2^2.$

Inhomogeneous Low Rank estimation $\mathbf{Y} = \Delta \odot \mathbf{G} + \mathbf{X}, \mathbf{X} = xx^T$

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

Theorem (Barbier-Reeves'20, Behne-Reeves '22, Ko-G-Zdeborova-Krzakala '22)

Assume $\Delta_{ij} = \Delta_{st}$ for $i, j \in I_s \times I_t$, $|I_s|/N \mapsto \alpha_s, 1 \leq i, j \leq n$ so that $(\Delta_{s,t}^{-1})_{s,t} \geq 0$ and $\mathbb{P}_0 = \mathbb{P}_X$. Then,

$$\lim_{N\to\infty} F_N(\Delta) = \sup_Q \varphi_\Delta(\mathbf{Q}).$$

- *If* $\left\| \sqrt{\alpha} \frac{1}{\Delta^2} \sqrt{\alpha} \right\|_{op} < \frac{1}{9d^4C^6}$ *then* $\lim_{N\to\infty} \mathrm{MMSE} = \mathbb{E}_X \|xx^T\|_2^2$.
- *If* $\left\| \sqrt{\alpha} \frac{1}{\Delta^2} \sqrt{\alpha} \right\|_{op} > \frac{1}{\mathbb{E}_X \|xx^T\|_2^2}$ *then* $\lim_{N\to\infty} \mathrm{MMSE} < \mathbb{E}_X \|xx^T\|_2^2$. *Sharp if $\mathbb{P}_X$ is Gaussian.*

Inhomogeneous Low Rank estimation $\mathbf{Y} = \Delta \odot \mathbf{G} + \mathbf{X}, \mathbf{X} = xx^T$

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

Theorem (Barbier-Reeves'20, Behne-Reeves '22, Ko-G-Zdeborova-Krzakala '22)

Assume $\Delta_{ij} = \Delta_{st}$ for $i,j \in I_s \times I_t$, $|I_s|/N \mapsto \alpha_s, 1 \le i,j \le n$ so that $(\Delta_{s,t}^{-1})_{s,t} \ge 0$ and $\mathbb{P}_0 = \mathbb{P}_X$. Then,

$$\lim_{N\to\infty} F_N(\Delta) = \sup_Q \varphi_\Delta(\mathbf{Q}).$$

- *If $\left\|\sqrt{\alpha}\frac{1}{\Delta^2}\sqrt{\alpha}\right\|_{op} < \frac{1}{9d^4C^6}$ then $\lim_{N\to\infty}\mathrm{MMSE} = \mathbb{E}_X\|xx^T\|_2^2$.*
- *If $\left\|\sqrt{\alpha}\frac{1}{\Delta^2}\sqrt{\alpha}\right\|_{op} > \frac{1}{\mathbb{E}_X\|xx^T\|_2^2}$ then $\lim_{N\to\infty}\mathrm{MMSE} < \mathbb{E}_X\|xx^T\|_2^2$. Sharp if $\mathbb{P}_X$ is Gaussian.*
- *same transition as BBP for $\Delta^{-2}\odot\mathbf{Y} - \frac{1}{N}diag(\Delta^{-2}1)$ (WIP Ko, Mergny, Pak)*

Universality Questions

- Universality with respect to the law of the noise $\mathbf{G}$,
- Universality with respect to the noise which may not be additive,
- Universality with respect to the distribution of $u = x/\sqrt{N}$ and $u_0 = x_0/\sqrt{N}$ and $\mathbf{G}, \mathbf{G}_0$: in non-Bayesian (or mismatched setting), we may have

$$\mathbb{P}_0 \neq \mathbb{P}_X, \qquad \text{Law of}(\mathbf{G}_0|u_0) \neq \text{ Law of}(\mathbf{G}|u)$$

  This is a technical challenge as this destroys symmetry between replicas (we are not on the Nishimori line anymore) so that classical spin glass techniques do not apply.

Universality Heuristics

Lesieur, Krzakala, Zdeborova '17 : Study for general g the conditional law

$$d\mathbf{G}_N^Y(\mathbf{x}) = \frac{1}{Z_X^g(\mathbf{Y})} \prod_{1\le i<j\le N} e^{g(Y_{ij}, \frac{x_i x_j}{\sqrt{N}})} \prod_{1\le i\le N} d\mathbb{P}_X(x_i)$$

where

$$\text{Law of}(\mathbf{Y}) = \frac{1}{Z_0} e^{\sum_{ij} g^0(y_{ij}, \rho N^{-1/2} x_i^0 x_j^0)} d\mathbb{P}_0^{\otimes N}(x_0) \prod dy_{ij}\,,$$

with $g \neq g^0$ and $\mathbb{P}_X \neq \mathbb{P}_0$.
Predict by the so-called replica approach the limit of

$$F_N(g) = \mathbb{E}_{\mathbf{Y}}[\frac{1}{N} \log Z_X^g(\mathbf{Y}) - \frac{1}{N} \sum_{i<j} g(y_{ij}, 0)]\,.$$

$-\!>$ The goal of our recent research is to prove these results.

Universality

- Assume that $\mathbb{P}_0$ and $\mathbb{P}_X$ are compactly supported,
- The functions $g(Y,w), g^0(Y,w)$ are three times differentiable in w,
- Consistent estimator : $\int \partial_w g(y,0) e^{g^0(y,0)} dy = 0$.

Theorem (Ko-G-Zdeborova-Krzakala '23)

Let

$$H_N^{\bar\beta}(\mathbf{x}:\mathbf{x}^0, W) = \sum_{i<j} \left(\beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N} x_i x_j x_i^0 x_j^0 + \frac{\beta_S}{2N} (x_i x_j)^2 \right)$$

and if W_{ij} are iid $N(0,1)$, x_0 iid law $\mathbb{P}_0$,

$$F_N(\bar\beta) = \mathbb{E}_{W,\mathbf{x}^0}[\frac{1}{N} \log \int e^{H_N^{\bar\beta}(\mathbf{x}:\mathbf{x}^0, W)} d\mathbb{P}_X^{\otimes N}(\mathbf{x})]$$

Then

$$\left| F_N(g) - F_N(\bar\beta) \right| = O(N^{-1/2})$$

where, if $\mathbb{P}^0_{\text{out}} \simeq e^{g^0(y,0)} dy$, $\beta = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[(\partial_w g(y,0))^2]^{1/2}$, $\beta_{SNR} = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[\partial_w g(y,0) \partial_w g^0(y,0)]$, $\beta_S = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[\partial_w^2 g(y,0)]$.

Limiting free energy

Theorem (Ko-G-Zdeborova-Krzakala '23)

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

$$\lim_{N\to\infty} F_N(\bar{\beta}) = \sup_{S,M}\{\varphi_{\bar{\beta}}(S,M)\}, \quad \varphi_{\bar{\beta}}(S,M) = \varphi_\beta(S,M) + \frac{\beta_{SNR} M^2}{2} + \frac{\beta_S S^2}{4}$$

The limit is given by

$$\varphi_\beta(S,M) = \inf_{\mu,\lambda,\zeta,Q}\left(\mathbb{E}_0[X_0(\lambda,\mu,Q,\zeta)] - \mu S - \lambda M - \frac{\beta^2}{4}\sum_{k=0}^{r-1}\zeta_k(Q_{k+1}^2 - Q_k^2)\right)$$

where $\lambda, \mu \in \mathbb{R}^2$ and for $\zeta_{-1} = 0 < \zeta_0 < \cdots < \zeta_{r-1} < 1$ and $0 = Q_0 \leq Q_1 \leq \cdots \leq Q_{r-1} \leq Q_r = S$ we defined recursively the random variables $X_r, X_{r-1}, \ldots, X_0$ by

$$X_r = \log \int e^{\beta \sum_{j=1}^r z_i x + \lambda x^2 + \mu x x^0}\, d\mathbb{P}_X(x), X_j = \frac{1}{\zeta_j} \log \mathbb{E}_{z_{j+1}} e^{\zeta_j X_{j+1}}.$$

where z_j are Gaussian random variables with variance $Q_j - Q_{j-1}$ and x^0 is an independent random variable with distribution $\mathbb{P}_0$.

Quenched Large deviations for the overlaps

Recall $H_N^{\bar\beta}(\mathbf{x}:\mathbf{x}^0,W)=\sum_{i<j}\beta\frac{W_{ij}}{\sqrt{N}}x_ix_j+\frac{\beta_{SNR}}{2}NR_{1,0}^2+\frac{\beta_S}{4}NR_{1,1}^2$ and let

$$d\mathbf{G}_{\bar\beta}^N(\mathbf{x})=\frac{1}{Z_N(\mathbf{x}^0,W)}e^{H_N^{\bar\beta}(\mathbf{x}:\mathbf{x}^0,W)}d\mathbb{P}_X^{\otimes N}(\mathbf{x})$$

where $R_{.,*}$ are the overlaps $R_{1,1}:=\frac{1}{N}\sum_{i=1}^N x_i^2,\quad R_{1,0}:=\frac{1}{N}\sum_{i=1}^N x_ix_i^0$.

Theorem

For every $\bar\beta=(\beta,\beta_{SNR},\beta_S)\in\mathbb{R}^3$, the law of $(R_{1,1},R_{1,0})$ under $\mathbf{G}_{\bar\beta}^N$ satisfies an almost sure LDP with good rate function $I_{\bar\beta}^{FP}$ given by

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

In other words, for any measurable subset B of $\mathbb{R}^2$, for almost all $(W,\mathbf{x}^0)$,

$$-\inf_{(S,M)\in B^o}I_{\bar\beta}^{FP}(S,M)\le\liminf_{N\to\infty}\frac{1}{N}\log\mathbf{G}_{\bar\beta}^N((R_{1,1},R_{1,0})\in O)\le$$

$$\le\limsup_{N\to\infty}\frac{1}{N}\log\mathbf{G}_{\bar\beta}^N((R_{1,1},R_{1,0})\in B)\le-\inf_{(S,M)\in\bar B}I_{\bar\beta}^{FP}(S,M)$$

Comments on large deviations

- The large deviation for $R_{1,1}$ under $\mathbf{G}^N_{\beta,0,0}$ was proven by Panchenko '15, The main point of our work is to extend it to $R_{1,0}$.

Comments on large deviations

- The large deviation for $R_{1,1}$ under $\mathbf{G}^N_{\beta,0,0}$ was proven by Panchenko '15, The main point of our work is to extend it to $R_{1,0}$.
- It is enough to prove the LDP when $\beta_{SNR} = \beta_S = 0$ by Varadhan's Lemma,

Comments on large deviations

- The large deviation for $R_{1,1}$ under $\mathbf{G}^N_{\beta,0,0}$ was proven by Panchenko '15, The main point of our work is to extend it to $R_{1,0}$.
- It is enough to prove the LDP when $\beta_{SNR} = \beta_S = 0$ by Varadhan's Lemma,
- If $I^{FP}_{\bar{\beta}}$ has a unique minimizer (S^*, M^*), they are the almost sure limit of the overlaps under $\mathbf{G}^N_{\bar{\beta}}$.

Comments on large deviations

- The large deviation for $R_{1,1}$ under $\mathbf{G}^N_{\beta,0,0}$ was proven by Panchenko '15, The main point of our work is to extend it to $R_{1,0}$.
- It is enough to prove the LDP when $\beta_{SNR} = \beta_S = 0$ by Varadhan's Lemma,
- If $I^{FP}_{\bar\beta}$ has a unique minimizer (S^*, M^*), they are the almost sure limit of the overlaps under $\mathbf{G}^N_{\bar\beta}$.
- The large deviations can be extended to the original Gibbs measures

$$d\mathbf{G}^N_Y(\mathbf{x} \in A) = \frac{1}{Z^N_Y} e^{\sum_{ij} g(Y_{ij}, \frac{x_i x_j}{\sqrt{N}})} d\mathbb{P}^{\otimes N}_X(\mathbf{x})$$

  and our results show they are universal given the parameters $\beta = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[(\partial_w g(y,0))^2]^{1/2}$, $\beta_{SNR} = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[\partial_w g(y,0)\partial_w g^0(y,0)]$, $\beta_S = \mathbb{E}_{\mathbb{P}^0_{\text{out}}}[\partial^2_w g(y,0)]$. This shows universality of likelihood in a large class of problems.

Comments on phase transitions

- Phase transition are complicated to establish in general : it is open.

Comments on phase transitions

- Phase transition are complicated to establish in general : it is open.
- Particular cases of mismatched studied by Pourkamali-Macris '20, Camilli, Contucci, Mingione '22, Barbier, Hou, Mondelli, Saenz '22.

Comments on phase transitions

- Phase transition are complicated to establish in general : it is open.
- Particular cases of mismatched studied by Pourkamali-Macris '20, Camilli, Contucci, Mingione '22, Barbier, Hou, Mondelli, Saenz '22.
- When $\beta = L\beta', \beta_{SNR} = L\beta_{SNR}, \beta_S = L\beta'_S$, and $\mathbb{P}_X$ is invariant under rotation, we find again the BBP transition since :

$$H_N^{\bar{\beta}}(\mathbf{x} : \mathbf{x}^0, W) = \sum_{i<j} \left(\beta \frac{W_{ij}}{\sqrt{N}} x_i x_j + \frac{\beta_{SNR}}{N} x_i x_j x_i^0 x_j^0 + \frac{\beta_S}{2N} (x_i x_j)^2 \right)$$

so that for L large

$$\frac{1}{L} F_N(\bar{\beta}) = E_{W,\mathbf{x}^0}[\frac{1}{NL} \log \int e^{H_N^{\bar{\beta}}(\mathbf{x}:\mathbf{x}^0, W)} d\mathbb{P}_X^{\otimes N}(\mathbf{x})]$$

$$\simeq \frac{1}{2} \text{essup}_{x \in \mathbb{R}^N} \{ \langle x, (\beta' W + \beta'_{SNR} x_0 x_0^T) x \rangle + \frac{\beta'_S}{2} \|x\|_2^4 \}$$

Idea of the proof : Universality

We prove that for A a measurable subset of $\mathbb{R}^2$

$$F_N(g,A) = \mathbb{E}[\frac{1}{N}\log\int 1_{(R_{1,1},R_{1,0})\in A} e^{\sum_{ij}(g(Y_{ij},\frac{x_ix_j}{\sqrt{N}})-g(Y_{ij},0))}\, d\mathbb{P}_X^{\otimes N}(\mathbf{x})]$$

is close to

$$F_N(\bar\beta,A) = \mathbb{E}\frac{1}{N}\log\int 1_{(R_{1,1},R_{1,0})\in A} e^{\sum_{i<j}\left(\beta\frac{W_{ij}}{\sqrt{N}}x_ix_j+\frac{\beta_{SNR}}{N}x_ix_jx_i^0x_j^0+\frac{\beta_S}{2N}(x_ix_j)^2\right)}\, d\mathbb{P}_X^{\otimes N}(\mathbf{x}) :$$

- by expanding g with respect to its second variable (bounded by $1/\sqrt{N}$) :

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

- Conditionally to $\mathbf{x}^0$, $(\partial_w g(Y_{ij},0))_{i<j}$ are independent variables with, by the consistent estimator hypothesis $\mathbb{E}_{P^0_{\rm out}}\partial_w g(Y,0)=0$

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

- The usual universality techniques for spin glasses can be generalized to our setting.

Idea of the proof of the Large deviations principle

- We show that

$$\mathbf{x}^0, W \to F_N^{SK}(A) = \frac{1}{N} \log \int 1_{(R_{1,1},R_{1,0})\in A} e^{\sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j} d\mathbb{P}_X^{\otimes N}(\mathbf{x})$$

self-averages by concentration of measure, cf Talagrand (the difficulty being that it is not smooth in $\mathbf{x}^0$)

Idea of the proof of the Large deviations principle

- We show that

$$\mathbf{x}^0, W \to F_N^{SK}(A) = \frac{1}{N} \log \int 1_{(R_{1,1},R_{1,0})\in A} e^{\sum_{i<j} \beta \frac{W_{ij}}{\sqrt{N}} x_i x_j} d\mathbb{P}_X^{\otimes N}(\mathbf{x})$$

self-averages by concentration of measure, cf Talagrand (the difficulty being that it is not smooth in $\mathbf{x}^0$)
- We use the interpolation trick to show that, for any $0 < \zeta_0 < \cdots < \zeta_{r-1} < 1$ and $0 = Q_0 \leq Q_1 \leq \cdots \leq Q_{r-1} \leq Q_r = S$, even though we are not on the Nishimori line, because the overlaps are fixed in $B_\epsilon(S, M) = \{|R_{1,1} - S| \leq \epsilon\} \cap \{|R_{1,0} - M| \leq \epsilon\}$:

$$\mathbb{E}[F_N^{SK}(B_\epsilon(S, M))] \leq \frac{1}{N}\mathbb{E} \log \sum_\alpha v_\alpha \int_{B_\epsilon(S,M)} e^{\beta \sum_{i\leq N} Z_i(\alpha) x_i} d\mathbb{P}_X^{\otimes N}(\mathbf{x}) - \frac{1}{N}\mathbb{E} \log \sum_\alpha v_\alpha e^{\sqrt{N}\beta Y(\alpha)} + o_{\epsilon,N}(1)$$

where v_α are Ruelle probability cascades, and $Z(\alpha)$ and $Y(\alpha)$ centered Gaussian processes

$$\mathbb{E} Z(\alpha^1) Z(\alpha^2) = Q_{\alpha^1 \wedge \alpha^2} \quad \mathbb{E} Y(\alpha^1) Y(\alpha^2) = \frac{1}{2} Q^2_{\alpha^1 \wedge \alpha^2}.$$

Idea of the proof : Large deviations

- We use Cramer's tilting argument to bound the first term by

$$
\begin{aligned}
&\frac{1}{N}\mathbb{E}\log\sum_{\alpha} v_\alpha \int_{B_\epsilon(S,M)} e^{\beta\sum_{i\le N} Z_i(\alpha)x_i}\, d\mathbb{P}_X^{\otimes N}(\mathbf{x}) \\
&\le -\mu S - \lambda M + \frac{1}{N}\mathbb{E}\log\sum_{\alpha} v_\alpha \int e^{\sum_{i\le N}\left\{\beta Z_i(\alpha)x_i + \lambda x_i^2 + \mu x_i x_i^0\right\}}\, d\mathbb{P}_X^{\otimes N}(\mathbf{x}) + o(1) \\
&\le -\mu S - \lambda M + \mathbb{E}_0 X_0 - \frac{\beta^2}{4}\sum_{k=0}^{r-1}\zeta_k(Q_{k+1}^2 - Q_k^2) + o(1).
\end{aligned}
$$

Idea of the proof : Large deviations

- We use Cramer's tilting argument to bound the first term by

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

$$\le -\mu S - \lambda M + \frac{1}{N}\mathbb{E}\log\sum_{\alpha} v_\alpha \int e^{\sum_{i\le N}\left\{\beta Z_i(\alpha)x_i + \lambda x_i^2 + \mu x_i x_i^0\right\}}\, d\mathbb{P}_X^{\otimes N}(\mathbf{x}) + o(1$$

$$\le -\mu S - \lambda M + \mathbb{E}_0 X_0 - \frac{\beta^2}{4}\sum_{k=0}^{r-1}\zeta_k(Q_{k+1}^2 - Q_k^2) + o(1).$$

- To prove the lower bound, we use the cavity computations and the standard procedure of the Aizenman–Sims–Starr scheme (but now we may have symmetry breaking and use Ruelle probability cascades). We then remove the indicator function of $B_\epsilon(S,M)$ by showing that tilting is optimal for some choice of μ,λ as in Cramer's proof. A difficulty is to deal with atypical $\mathbf{x}^0$.

Conclusion

- The problem of estimating a vector from its noisy observation leads to exciting problems in random matrix theory and spin glasses theory,
- Studying transition and detectability from the formulas is not obvious.
- Constructing optimal algorithms is a natural question, see Krzakala, Ko, Pak '23.

Joyeux anniversaire Elisabeth !