STAT260 Mean Field Asymptotics in Statistical Learning — Lecture 10 - 02/24/2021

Lecture 10: Replica method III: $\mathbb{Z}_2$ synchronization

Lecturer: Song Mei *Scriber: Zitong Yang* *Proof reader: Zitong Yang*

In this lecture, we consider the $\mathbb{Z}_2$ synchronization problem. We use the free energy trick and the replica trick to derive the asymptotic expressions for some desired observables.

1 $\mathbb{Z}_2$ synchronization problem

We first recall the statement of $\mathbb{Z}_2$ synchronization problem introduced in Lecture 4. The goal of $\mathbb{Z}_2$ synchronization problem is to estimate the signal $\boldsymbol{\theta} \in \{-1, +1\}^n$ from observation

$$\boldsymbol{Y} = \frac{\lambda}{n}\boldsymbol{\theta}\boldsymbol{\theta}^\top + \boldsymbol{W} \in \mathbb{R}^{n\times n},$$

where $\boldsymbol{W} \sim \mathrm{GOE}(n)$ are noise associated with the observation and $\lambda > 0$ is a parameter that controls the signal-to-noise ratio. We assume a Bayesian prior of $\boldsymbol{\theta}_i \sim_{\text{i.i.d.}} \mathrm{Unif}(\{\pm 1\})$ on $\boldsymbol{\theta}$. As we have derived in Lecture 4, the log-likelihood for a given parameter $\boldsymbol{\sigma} \in \{-1, +1\}^n$ of the problem is given by (up to a constant)

$$-n\left\|\boldsymbol{Y} - \frac{\lambda}{n}\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\right\|_F^2 \propto_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma}, \boldsymbol{Y}\boldsymbol{\sigma}\rangle. \tag{1}$$

This leads to the Gibbs measure on $\{-1, 1\}^n$

$$\mathbb{P}_{\beta,\lambda}(\boldsymbol{\sigma}) \propto \exp\left\{\beta\langle \boldsymbol{\sigma}, \boldsymbol{Y}\boldsymbol{\sigma}\rangle\right\}.$$

We will consider the performance of the average estimator

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Y}) = \langle \boldsymbol{\sigma}\rangle_{\beta,\lambda} = \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^n} \boldsymbol{\sigma}\cdot \mathbb{P}_{\beta,\lambda}(\boldsymbol{\sigma}).$$

We have shown that when $\beta = \infty$, $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ is the maximum likelihood estimator, and when $\beta = \lambda/2$, $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ reduces to the Bayes estimator.

2 Overview of the Main Results

The goal of this lecture is to initiate the derivation of the following observable in the thermal dynamic limit

$$m_*(\beta,\lambda) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \langle \psi(\boldsymbol{\sigma}_i, \boldsymbol{\theta}_i)\rangle_{\beta,\lambda}\right],$$
$$s_*(\beta,\lambda) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \psi(\langle \boldsymbol{\sigma}_i\rangle_{\beta,\lambda}, \boldsymbol{\theta}_i)\right],$$

where $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a test function.

2.1 Formalism

Our first result states that

$$m_*(\beta,\lambda) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} \langle \psi(\boldsymbol{\sigma}_i, \boldsymbol{\theta}_i)\rangle_{\beta,\lambda}\right] = \mathbb{E}_{G,\theta}\left\{\mathbb{E}_{\bar{\boldsymbol{\sigma}}\sim D(\tanh(2\beta\lambda\mu_*\theta+2\beta\sqrt{q_*}G))}\left[\psi(\bar{\boldsymbol{\sigma}},\theta)\right]\right\}, \tag{2}$$

where $G \sim \mathcal{N}(0,1)$, $\theta \sim \text{Unif}(\{\pm 1\})$, and $\bar{\boldsymbol{\sigma}} \sim D(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G)$ represents the distribution supported on $\{\pm 1\}$ uniquely defined the expectation

$$\mathbb{E}\left[\bar{\boldsymbol{\sigma}}\right] = \tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G).$$

Here q_*, μ_* are two real numbers that can be computed as solutions to the following self-consistent equations:

$$\begin{aligned} q_* &= \mathbb{E}_{G,\theta}\left[\tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G)^2\right], \\ \mu_* &= \mathbb{E}_{G,\theta}\left[\theta\tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G)\right], \end{aligned}$$

Although we will not show in this lecture series, a more general results suggests that for $\boldsymbol{\sigma} \sim \mathbb{P}_{\beta,\lambda}$ and $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_i \overset{i.i.d.}{\sim} \text{Unif}(\{\pm 1\})$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(\boldsymbol{\sigma}_i, \theta_i) = \mathbb{E}_{G,\theta}\left\{\mathbb{E}_{\bar{\boldsymbol{\sigma}}\sim D(\tanh(2\beta\lambda\mu_*\theta+2\beta\sqrt{q_*}G))}\left[\psi(\bar{\boldsymbol{\sigma}},\theta)\right]\right\}.$$

The interpretation of this more general result is that the empirical law $\frac{1}{n}\sum_{i=1}^{n}\delta_{(\boldsymbol{\sigma}_i,\theta_i)}$ converges to the law of $(\bar{\boldsymbol{\sigma}}, \theta)$ from equation (2).

Our second result states that

$$s_*(\beta,\lambda) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} \psi(\langle \boldsymbol{\sigma}_i\rangle_{\beta,\lambda}, \boldsymbol{\theta}_i)\right] = \mathbb{E}_{G,\theta}\left[\psi(\tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G),\theta)\right]. \tag{3}$$

Similarly, we can interpret the result as the convergence of empirical law

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(\langle\boldsymbol{\sigma}_i\rangle_{\beta,\lambda},\theta_i)} \to \text{Law of } (m,\theta),$$

where

$$\theta \sim \text{Unif}(\{\pm 1\}), \quad G \sim \mathcal{N}(0,1), \quad m = \tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G).$$

2.2 Illustrations of the formalism

In this subsection, we provide several example to illustrate how our formalism from the previous section could be applied.

Square loss. For the square loss $\psi(a,\theta) = (a-\theta)^2$, equation (3) translates to

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\|\langle\boldsymbol{\sigma}\rangle_{\beta,\lambda} - \boldsymbol{\theta}\|_2^2 = \mathbb{E}_{G,\theta}\left[(\tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G) - \theta)^2\right].$$

Signed square loss. Since $\boldsymbol{\theta} \in \{-1,+1\}^n$, a natural loss to consider is $\psi(a,\theta) = (\text{sign}(a) - \theta)^2$. Equation (3) suggests that

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\|\text{sign}\langle\boldsymbol{\sigma}\rangle_{\beta,\lambda} - \boldsymbol{\theta}\|_2^2 = \mathbb{E}_{G,\theta}\left[(\text{sign}(\tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G)) - \theta)^2\right].$$

Alignment loss. Another loss to consider is to count how many $\boldsymbol{\sigma}_i$'s are aligned with its $\boldsymbol{\theta}_i$. This leads to the loss $\psi(a, \theta) = a \cdot \theta$. Then equation (3) implies Then

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\sum_{i=1}^{n} \boldsymbol{\theta}_i \langle \boldsymbol{\sigma}_i \rangle_{\beta,\lambda} = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\langle \boldsymbol{\theta}, \langle \boldsymbol{\sigma} \rangle_{\beta,\lambda}\rangle = \mathbb{E}_{G,\theta}\left[\theta \cdot \text{sign}(\tanh(2\beta\lambda\mu_* \theta + 2\beta\sqrt{q_*}G))\right] = \mu_\star.$$

3 Derivation of the main result

In this section, we will derive the expression for $m_*(\beta, \lambda)$ using the free energy trick and the replica trick.

3.1 Formulation

We first formulate the problem into statistical physics language. The configuration space would be $\Omega = \{-1, +1\}^n$ with base measure ν_0 = Uniform measure over Ω (Bayesian prior). The Hamiltonian for the original system is the log likelihood (1). In order to compute the observable of associated with $m_\star(\beta, \lambda)$ quantity, we define the perturbed Hamiltonian

$$H_{\lambda,h}(\boldsymbol{\sigma}) = -\langle \boldsymbol{\sigma}, \boldsymbol{W}\boldsymbol{\sigma}\rangle - \lambda\langle \boldsymbol{\sigma}, \theta\rangle^2 - h\sum_{i=1}^{n}\psi(\boldsymbol{\sigma}_i, \theta_i).$$

The free energy of the system is

$$Z_n(\beta, \lambda, h) = \int_\Omega \exp\left\{-\beta H_{\lambda,h}(\boldsymbol{\sigma})\right\} \boldsymbol{\nu}_0(d\boldsymbol{\sigma}).$$

Finally, we can define the free energy density and compute the limiting observable by taking derivative w.r.t. the perturbation strength h.

$$\varphi(\beta, \lambda, h) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\log Z_n(\beta, \lambda, h)\right],$$
$$m_*(\beta, \lambda) = \frac{1}{\beta}\partial_h \varphi(\beta, \lambda, h)\Big|_{h=0}.$$

Given the aforementioned setup, the following claims gives an implicit expression for the free energy density $\varphi(\beta, \lambda, h)$.

Claim 1. *For some values of (β, λ)[1], the free energy density for $\mathbb{Z}_2$ synchronization problem can be written as*

$$\varphi(\beta, \lambda, h) = \underset{\mu,q}{ext}\, u(\mu, q; \beta, \lambda, h)$$

where

$$u(\mu, q; \beta, \lambda, h) = -\beta\lambda\mu^2 + \beta^2(1-q)^2 + \mathbb{E}_{G,\theta}\left\{\log\left[\mathbb{E}_\sigma \exp\left\{2\beta\lambda\mu\sigma\theta + 2\beta\sqrt{q}G\sigma + \beta h\psi(\sigma, \theta)\right\}\right]\right\}.$$

The expectation is taken over the randomness $(G, \theta, \sigma) \sim \mathcal{N}(0, 1) \times Unif(\{\pm 1\}) \times Unif(\{\pm 1\})$.

We will derive the claim above using the replica trick from the last lecture to compute the free energy integral:

Lemma 2 (Replica Trick)**.** *For a given random variable Z*

$$\mathbb{E}[\log Z] = \lim_{k\to 0} \frac{1}{k}\log \mathbb{E}[Z^k].$$

[1]This is a result of our heuristic approach. The precise region of (β, λ) where the claim holds is beyond the scope of this lecture.

This suggests us to divide the derivation into three parts

$$
\begin{aligned}
&(a). \quad S(k,\beta,\lambda,h) \equiv \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[Z_n(\beta,\lambda,h)^k\right], \quad \textit{(The } n \textit{ limit.)} \\
&(b). \quad \varphi(\beta,\lambda,h) \equiv \lim_{k\to 0} \frac{1}{k} S(k,\beta,\lambda,h), \quad \textit{(The } k \textit{ limit.)} \\
&(c). \quad m_*(\beta,\lambda) \equiv \frac{1}{\beta}\partial_h \varphi(\beta,\lambda,h)\Big|_{h=0}. \quad \textit{(The } h \textit{ derivative.)}
\end{aligned}
$$

In the following subsections, we will go through $(a)-(c)$.

3.2 The $n \to \infty$ limit

We will start by computing the free energy integral

$$
\begin{aligned}
\mathbb{E}[Z_n(\beta,\lambda,h)^k] &= \mathbb{E}\left[\int_\Omega \exp(-\beta H_{\lambda,h}(\sigma)\nu_0(d\sigma))\right]^k \\
&= \mathbb{E}\left[\int_{\Omega^{\otimes k}} \exp\left(-\beta \sum_{a=1}^k H_{\lambda,h}(\sigma^a)\right) \prod_{a=1}^k \nu_0(d\sigma^a)\right] \\
&= \int_{\Omega^{\otimes k}} \exp\left\{\beta\left(\sum_{a=1}^k \lambda \frac{\langle \boldsymbol{\theta}, \sigma^a\rangle^2}{n} + h\sum_{a=1}^k\sum_{i=1}^n \psi(\boldsymbol{\sigma}_i^a, \boldsymbol{\theta}_i)\right)\right\} \underbrace{\mathbb{E}\left[\exp\left(\beta\sum_{a=1}^k \langle \boldsymbol{\sigma}^a, \boldsymbol{W}\boldsymbol{\sigma}^a\rangle\right)\right]}_{\text{Moment Generating Function}} \prod_{a=1}^k \nu_0(d\boldsymbol{\sigma}^a)
\end{aligned}
$$

Here the first equation follows by writing out the exponent k as the integral over k replicas, and the second line follows from the definition of $H_{\lambda,h}$. Since $W \sim \mathsf{GOE}(n)$, the expectation over $\boldsymbol{W}$ can be computed using the formula for the moment generating function of Gaussian random variable. This gives

$$
\mathbb{E}\left[\exp\left(\beta\sum_{a=1}^k \langle \boldsymbol{\sigma}^a, \boldsymbol{W}\boldsymbol{\sigma}^a\rangle\right)\right] = \exp\left[\beta^2 \sum_{a,b=1}^k \langle \sigma^a, \sigma^b\rangle^2/n\right].
$$

Plug this into $\mathbb{E}[Z_n(\beta,\lambda,h)^k]$, we get

$$
\mathbb{E}[Z_n(\beta,\lambda,h)^k] = \int_{\Omega^{\otimes k}} \prod_{a=1}^k \nu_0(d\boldsymbol{\sigma}^a) \exp\left\{\beta\sum_{a=1}^k \lambda\frac{\langle \boldsymbol{\theta}, \boldsymbol{\sigma}^a\rangle^2}{n} + \beta^2 \sum_{a,b=1}^k \frac{\langle \boldsymbol{\sigma}^a . \boldsymbol{\sigma}^b\rangle^2}{n} + \beta h \sum_{a=1}^k\sum_{i=1}^n \psi(\boldsymbol{\sigma}_i^a, \boldsymbol{\theta}_i)\right\}
$$

Same as the last lecture, we introduce the delta identity function to factor out the first two terms in the exponential from the integral

$$
1 = \int \prod_{a=1}^k dq_{0a} \prod_{a,b=1}^k dq_{ab} \prod_{a=1}^k \delta(\langle \boldsymbol{\theta}, \boldsymbol{\sigma}^a\rangle - nq_{0a}) \prod_{a,b=1}^k \delta(\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle - nq_{ab}).
$$

Then

$$
\mathbb{E}[Z_n(\beta,\lambda,h)^k] = \exp\left(\beta\lambda n \sum_{a=1}^k q_{0a}^2 + \beta^2 n \sum_{a,b=1}^k q_{ab}^2\right) \times \mathsf{Ent},
$$

where

$$
\mathsf{Ent} = \int_{\Omega^{\otimes k}} \prod_{a=1}^k \nu_0(d\boldsymbol{\sigma}^a) \prod_{a=1}^k \delta(\langle \boldsymbol{\theta}, \boldsymbol{\sigma}^a\rangle - nq_{0a}) \prod_{a,b=1}^k \delta(\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b\rangle - nq_{ab}) \exp\left[\beta h \sum_{a=1}^k\sum_{i=1}^n \psi(\boldsymbol{\sigma}_i^a, \boldsymbol{\theta}_i)\right].
$$

The entropy term Ent can be computed using the delta identity formula and the saddle point approximation, which gives

$$\lim_{n\to\infty}\frac{1}{n}\log\mathsf{Ent} = \inf_{\boldsymbol{\Lambda}\in\mathbb{R}^{(k+1)\times(k+1)}}\langle\boldsymbol{Q},\boldsymbol{\Lambda}\rangle/2+\log\mathbb{E}_{\boldsymbol{\sigma}}\left[\exp\left(-\sum_{a,b=0}^{k}\lambda_{ab}\frac{\sigma^a\sigma^b}{2}+\beta h\sum_{a=1}^{k}\psi(\sigma^a,\sigma^0)\right)\right],$$

where $\boldsymbol{\sigma}=(\sigma^0,\sigma^1,\dots,\sigma^a)\overset{i.i.d.}{\sim}\text{Unif}\{-1,1\}$. By another use of saddle point approximation, we get

$$S(k,\beta,\lambda,h)=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}Z_n(\beta,\lambda,h)^k=\sup_{\boldsymbol{Q}}\inf_{\boldsymbol{\Lambda}}U(\boldsymbol{Q},\boldsymbol{\Lambda}),$$

where

$$U(\boldsymbol{Q},\boldsymbol{\Lambda})=\beta\lambda\sum_{a=1}^{k}q_{0a}^2+\beta^2\sum_{a,b=1}^{k}q_{ab}^2+\frac{\langle\boldsymbol{Q},\boldsymbol{\Lambda}\rangle}{2}+\log\mathbb{E}_{\boldsymbol{\sigma}}\left[\exp\left(-\sum_{a,b=0}^{k}\lambda_{ab}\frac{\sigma^a\sigma^b}{2}+\beta h\sum_{a=1}^{k}\psi(\sigma^a,\sigma^0)\right)\right].$$

3.3 Heuristic solution to the minimax problem

The distinction between sup v.s. inf isn't crucial in this setting. Ultimately we want to use the first order stationary condition to figure out the $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ that achieves the supremum/infimum. On a high level, we want to solve the stationary equation

$$\partial_{\boldsymbol{Q}}U(\boldsymbol{Q},\boldsymbol{\Lambda})=0,\quad \partial_{\boldsymbol{\Lambda}}U(\boldsymbol{Q},\boldsymbol{\Lambda})=0.$$

Since $U(\boldsymbol{Q},\boldsymbol{\Lambda})$ is quadratic in $\boldsymbol{Q}$, we can start by

$$\begin{aligned}\partial_{q_{0a}}U(\boldsymbol{Q},\boldsymbol{\Lambda})=0&\Rightarrow\lambda_{0a}=-2\beta\lambda q_{0a},\ \ 1\le a\le k,\\ \partial_{q_{ab}}U(\boldsymbol{Q},\boldsymbol{\Lambda})=0&\Rightarrow\ \lambda_{ab}=-4\beta^2 q_{ab},\ \ 1\le a\ne b\le k.\end{aligned}$$

Plug this back into the formula for $U=U(Q)$ and with a slight abuse of notation, we get

$$U(Q)=-\beta\lambda\sum_{a=1}^{k}q_{0a}^2-\beta^2\sum_{a,b=1}^{k}q_{ab}^2+\log\mathbb{E}_{\boldsymbol{\sigma}}\left[\exp\left(2\beta\lambda\sum_{a=1}^{k}q_{0a}\sigma^a\sigma^0+2\beta^2\sum_{a,b=1}^{k}q_{ab}\sigma^a\sigma^b+\beta h\sum_{a=1}^{k}\psi(\sigma^a,\sigma^0)\right)\right].$$

Then, the S quantity is extrema of $U: S(k,\beta,\lambda,h)=\text{ext}_{\boldsymbol{Q}}U(\boldsymbol{Q})$. However, after we plug in the stationary condition to get rid of λ_{ab}, the equation $\partial_{\boldsymbol{Q}}U(\boldsymbol{Q})=0$ becomes highly non-linear and no closed form solution can be found. The core difficulty here is the high-dimension nature of the stationary equation. We can reduce the high dimensional problem to a low dimensional one by exploiting the symmetry of the objective function. More concretely,

$$U(\Pi\boldsymbol{Q}\Pi^{\mathsf{T}})=U(\boldsymbol{Q}),\ \text{where}\ \Pi=\begin{bmatrix}1&0\\0&\overline{\Pi}\end{bmatrix},\ \overline{\Pi}\in\mathbb{R}^{k\times k}\ \text{a permutation matrix.}$$

This leads to the replica symmetric ansatz

$$\boldsymbol{Q}=\begin{bmatrix}1&\mu&\dots&\mu\\ \mu&1&&q\\ \vdots&&1&\\ \mu&q&&1\end{bmatrix},\mu,q\in\mathbb{R}.$$

In another word, we guess that the extrema is achieved at $\boldsymbol{Q}$ that satisfies the aforementioned functional form. In this ansatz, the diagonal entries are all 1 because for $\boldsymbol{\sigma}^a \in \Omega = \{-1, +1\}^n$, $\langle \boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b \rangle / n = 1$. The ansatz gives a simplified expression

$$\begin{aligned} U(k, \mu, q) = &- \beta\lambda k\mu^2 - \beta^2 \left(k + k(k-1)q^2\right) + 2\beta^2(1-q)k \\ &+ \log \mathbb{E}_{\boldsymbol{\sigma}} \left[\exp\left(2\beta\lambda\mu \sum_{a=1}^{k} \sigma^a \sigma^0 + 2\beta^2 q \sum_{a,b=1}^{k} \sigma^a \sigma^b + \beta h \sum_{a=1}^{k} \psi(\sigma^a, \sigma^0) \right) \right], \end{aligned}$$

which makes computing the $k \to 0$ limit tractable.

3.4 The $k \to 0$ limit

The goal of this section is to compute

$$\varphi(\beta, \lambda, h) = \lim_{k \to 0} \frac{1}{k} S(k, \beta, \lambda, h)$$

for $S(k, \beta, \lambda, h) = \mathrm{ext}_{\boldsymbol{Q}} U(\boldsymbol{Q}) = \mathrm{ext}_{\mu,q} U(k, \mu, q)$. Here we assume the lim and ext operation can be exchanged, and we get

$$\varphi(\beta, \lambda, h) = \operatorname*{ext}_{\mu,q} \lim_{k \to 0} \frac{1}{k} U(k, \mu, q) = \operatorname*{ext}_{\mu,q} u(\mu, q; \beta, \lambda, h).$$

To compute $u(\mu, q; \beta, \lambda, h)$, we want to further simplify $U(k, \mu, q)$ first. We can remove the cross term $\langle \sigma^a, \sigma^b \rangle$ in $U(k, \mu, q)$ using the following expression derived from Gaussian moment generating function.

Lemma 3 (Gaussian Moment Generating Function)**.** *For $G \sim \mathcal{N}(0, 1)$,*

$$\mathbb{E}\left[e^{\lambda G \sum_{a=1}^{k} \sigma^a} \right] = \exp\left[\frac{\lambda^2}{2} \sum_{a,b=1}^{k} \sigma^a \sigma^b \right]$$

Direct application of the above lemma gives

$$\begin{aligned} U(k, \mu, q) &= -\beta\lambda k\mu^2 - \beta^2 \left(k + k(k-1)q^2\right) + 2\beta^2(1-q)k \\ &\quad + \log \mathbb{E}_{G, \boldsymbol{\sigma}} \left[\exp\left\{ 2\beta\lambda\mu \sum_{a=1}^{k} \sigma^a \sigma^0 + 2\beta\sqrt{q} G \sum_{a=1}^{k} \sigma^a + \beta h \sum_{a=1}^{k} \psi(\sigma^a, \sigma^0) \right\} \right] \\ &= -\beta\lambda k\mu^2 - \beta^2 \left(k + k(k-1)q^2\right) + 2\beta^2(1-q)k + \log \mathbb{E}_{G,\theta} \left[\left(\mathbb{E}_\sigma \exp\left\{ 2\beta\lambda\mu\sigma\theta + 2\beta\sqrt{q} G\sigma + \beta h \psi(\sigma, \theta) \right\} \right)^k \right], \end{aligned}$$

where

$$G \sim \mathcal{N}(0, 1), \theta \sim \mathrm{Unif}(\{-1, 1\}), \sigma \sim \mathrm{Unif}(\{-1, 1\}).$$

To deal with the term involving expectation to the k-th moment, we can use the inverse replica trick:

Lemma 4 (Inverse Replica Trick)**.** *For a random variable T,*

$$\lim_{k \to 0} \frac{1}{k} \log \mathbb{E}\, T^k = \mathbb{E} \log T.$$

Setting $T = \mathbb{E}_\sigma \exp\left\{ 2\beta\lambda\mu\sigma\theta + 2\beta\sqrt{q} G\sigma + \beta h \psi(\sigma, \theta) \right\}$ in the above lemma, we get

$$\begin{aligned} u(\mu, q; \beta, \lambda, h) &\equiv \lim_{k \to 0} \frac{1}{k} U(k, \mu, q) \\ &= -\beta\lambda\mu^2 - \beta^2(1-q^2) + 2\beta^2(1-q) + \mathbb{E}_{G,\theta} \left\{ \log\left[\mathbb{E}_\sigma \exp\left\{ 2\beta\lambda\mu\sigma\theta + 2\beta\sqrt{q} G\sigma + \beta h \psi(\sigma, \theta) \right\} \right] \right\} \\ &= -\beta\lambda\mu^2 + \beta^2(1-q)^2 + + \mathbb{E}_{G,\theta} \left\{ \log\left[\mathbb{E}_\sigma \exp\left\{ 2\beta\lambda\mu\sigma\theta + 2\beta\sqrt{q} G\sigma + \beta h \psi(\sigma, \theta) \right\} \right] \right\}. \end{aligned}$$

This concludes the derivation of Claim 1.

3.5 The h derivative

Finally, we need to compute $\frac{1}{\beta}\partial_h\varphi(\beta,\lambda,h)\Big|_{h=0}$ for $\varphi(\beta,\lambda,h) = \operatorname*{ext}_{\mu,q} u(\mu,q;\beta,\lambda,h)$. By implicit differentiation theorem

$$\frac{1}{\beta}\partial_h\varphi(\beta,\lambda,h)\Big|_{h=0} = \frac{1}{\beta}\partial_h u(\mu,q;\beta,\lambda,h)\Big|_{(q=q_*,\mu=\mu_*,h=0)},$$

where $(\mu_*,q_*) = \arg\operatorname*{ext}_{\mu,q} u(\mu,q;\beta,\lambda,h)\Big|_{h=0}$. Direction calculation using the first order stationary condition gives

$$\begin{cases} \mu_* = \mathbb{E}_{G,\theta}\left[\theta\cdot\tanh\left(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*G}\right)\right], \\ q_* = \mathbb{E}_{G,\theta}\left[\tanh\left(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*G}\right)^2\right]. \end{cases}$$

Plugging in $(\mu_\star,q_\star)$ gives

$$\begin{aligned} m_\star(\beta,\lambda) &= \frac{1}{\beta}\partial_h\varphi(\beta,\lambda,h)\Big|_{h=0} \\ &= \mathbb{E}_{G,\theta}\left[\frac{\mathbb{E}_\sigma\left[\exp\left\{2\beta\lambda\mu_*\sigma\theta + 2\beta\sqrt{q_*G\sigma}\right\}\psi(\sigma,\theta)\right]}{\mathbb{E}_\sigma\left[\exp\left\{2\beta\lambda\mu_*\sigma\theta + 2\beta\sqrt{q_*G\sigma}\right\}\right]}\right] \\ &= \mathbb{E}_{G,\theta}\left[\mathbb{E}_{\bar\sigma\sim D}\left[\psi(\bar\sigma,\theta)\right]\right], \end{aligned}$$

where $\bar\sigma$ is a random variable supported on $\{-1,1\}$ uniquely defined by the condition

$$\mathbb{E}[\bar\sigma] = \tanh(2\beta\lambda\mu_*\theta + 2\beta\sqrt{q_*}G).$$

This completes the derivation of the formalism (2) in 2.1.