Lecture-19: Free energy approach

1 $\mathbb{Z}_2$ synchronization model

We define $\mathbb{Z}_2 \triangleq \{-1,1\}$ and $\Theta \triangleq \mathbb{Z}_2^N$. We assume that $\theta : \Omega \to \mathbb{Z}_2^N$ is an *i.i.d.* random vector with $\mathbb{E}\theta_i = 0$ for all $i \in [N]$. Let $W : \Omega \to \mathbb{R}^{N\times N}$ be a random matrix from Gaussian orthogonal ensemble independent of θ, so that the observation $Y : \Omega \to \mathbb{R}^{N\times N}$ is

$$Y = \frac{\lambda}{N}\theta \otimes \theta + W.$$

Let $f : \mathbb{R}^{N\times N} \to \mathcal{A} \triangleq \Theta \otimes \Theta$ be an estimator of $\theta \otimes \theta$, and $L : \mathcal{A} \times \Theta \to \mathbb{R}$ a matrix square loss function defined for all $(A,\theta) \in \mathcal{A} \times \Theta$, as

$$L(A,\theta) \triangleq \frac{1}{N^2}\left\|A - \theta \otimes \theta\right\|_F^2.$$

The expected risk for the estimator f for the loss function L under uniform distribution $Q \in \mathcal{M}(\Theta)$ for parameter θ and parametrized observation distribution $P_\theta \in \mathcal{M}(\mathbb{R}^{N\times N})$, is

$$R(Q,f,L) = \int_\Theta dQ(\theta)\int_Y dP_\theta(Y)L(f(Y),\theta).$$

2 Stochastic block model

Let $\theta : \Omega \to \Theta \triangleq \mathbb{Z}_2^N$ be an *i.i.d.* random vector with the common distribution uniform over $\mathbb{Z}_2$.

Definition 2.1. Consider a random graph $G = (V,E)$ where the vertex set $V = [N]$ and $E \subseteq V \times V$ is randomly generated. Let $\theta : \Omega \to \Theta \triangleq \mathbb{Z}_2^V$ denote the configuration of the graph and define the sets of vertices

$$V_+ \triangleq \{v \in V : \theta_v = 1\}, \qquad V_- \triangleq \{v \in V : \theta_v = -1\}. \tag{1}$$

We define a random symmetric matrix $A \in \{0,1\}^{V\times V}$ such that $A_{vv} = 0$ for all $v \in V$ and A_{vw} is independent and random for all $v < w$. For $a > b$, we define $\mathbb{E}A_{v,w} \triangleq \Big(\frac{a}{N}\mathbb{1}_{\{\theta_v=\theta_w\}} + \frac{b}{N}\mathbb{1}_{\{\theta_v\neq\theta_w\}}\Big)\mathbb{1}_{\{w\neq v\}}$. The edge set $E \triangleq \{(v,w) \in E : A_{v,w} = 1\}$.

Remark 1. We can write the conditional expectation of the adjacency matrix as

$$\mathbb{E}[A \mid \theta] = \frac{1}{N}\sum_{v,w} e_v^T e_w\Big(a\frac{(1+\theta_v\theta_w)}{2} + b\frac{(1-\theta_v\theta_w)}{2}\Big)\mathbb{1}_{\{v\neq w\}} = \frac{(a+b)}{2N}1^T1 - \frac{a}{2N}I + \frac{(a-b)}{2N}\theta^T\theta.$$

Recall that the adjacency matrix depends on the configuration θ and hence $\frac{(a-b)}{2N}\theta^T\theta$ is considered the signal part of the conditional expectation.

Definition 2.2. We define the zero-mean noise in the adjacency matrix as $W \triangleq A - \mathbb{E}[A \mid \theta]$, and the observation as the de-noised adjacency matrix, i.e. $Y \triangleq A - \Big(\mathbb{E}[A \mid \theta] - \frac{(a-b)}{2N}\theta^T\theta\Big) = \frac{(a-b)}{2N}\theta^T\theta + W$.

Remark 2. From the definition of W, we can find the variance of its entries $v \neq w$ as

$$\begin{aligned}\mathrm{Var}(W_{vw}) &= \mathbb{E}[(A_{v,w} - \mathbb{E}[A_{v,w} \mid \theta])^2] = \mathbb{E}[(A_{v,w} - \frac{(a+b)}{2N} - \frac{(a-b)}{2N}\theta_v\theta_w)^2] \\ &= \mathbb{E}A_{v,w}^2 + \frac{(a+b)^2}{4N^2} + \frac{(a-b)^2\mathbb{E}\theta_v^2\theta_w^2}{4N^2} - \frac{(a+b)\mathbb{E}A_{v,w}}{N} - \frac{(a-b)\mathbb{E}A_{v,w}\theta_v\theta_w}{N} + \frac{(a^2-b^2)\mathbb{E}\theta_v\theta_w}{2N^2}.\end{aligned}$$

Recall that θ is *i.i.d.* zero mean for $v \neq w$, and hence $\mathbb{E}\theta_v\theta_w = 0$ for $v \neq w$. Further, $\theta_v^2 = 1$ for all $v \in V$. In addition, we observe that $A_{v,w}^2 = A_{v,w}$ and from the tower property of conditional expectation $\mathbb{E}A_{v,w} = \mathbb{E}[\mathbb{E}[A_{v,w} \mid \theta]]$. Therefore, for $v \neq w$, we obtain

$$\mathbb{E}A_{v,w}^2 = \mathbb{E}A_{v,w} = \frac{1}{2N}\mathbb{E}[(a+b) + (a-b)\theta_v\theta_w] = \frac{(a+b)}{2N}, \quad \mathbb{E}[A_{v,w}\theta_v\theta_w] = \mathbb{E}[\mathbb{E}[A_{v,w}\theta_v\theta_w \mid \theta]] = \frac{(a-b)}{2N}.$$

Combining these results, we obtain

$$\mathrm{Var}(W_{vw}) = \frac{(a+b)}{2N} - \frac{a^2+b^2}{2N^2} = \frac{a}{2N}(1-\frac{a}{N}) + \frac{b}{2N}(1-\frac{b}{N}) \approx \frac{(a+b)}{2N}.$$

Rescaling by $\sqrt{\frac{(a+b)}{2}}$ on both sides and defining $W' \triangleq \frac{W}{\sqrt{\frac{(a+b)}{2}}}, Y' \triangleq \frac{Y}{\sqrt{\frac{(a+b)}{2}}}$, and $\lambda \triangleq \frac{(a-b)}{\sqrt{2(a+b)}}$, we obtain

$$Y' = \frac{\lambda}{N}\theta^T\theta + W'.$$

We note that $\mathrm{Var}(W'_{vw}) = \frac{1}{N}$, similar to the variance of noise in the Sherrington-Kirkpatrick model.

3 Estimators for $\mathbb{Z}_2$ synchronization

Recall that we have derived the maximum likelihood and the Bayes estimator for the $\mathbb{Z}_2$ synchronization problem. Maximum likelihood estimate (MLE) is NP-hard to compute for this case, and hence, we introduce two other estimators, which are relaxed versions of MLE and are easier to compute. Recall that MLE is given by

$$f_{\mathrm{ML}}(Y) = \arg\max\left\{\left\langle\theta, \theta Y^T\right\rangle : \theta \in \Theta\right\}.$$

Definition 3.1. We define the N-dimensional sphere of radius $\sqrt{N-1}$ defined as $\mathbb{S}^{N-1}(\sqrt{N}) \triangleq \left\{\theta \in \mathbb{R}^N : \|\theta\| = \sqrt{N}\right\}$. The **spectral estimator** is denoted by $f_{\mathrm{sp}} : \mathbb{R}^{N\times N} \to \mathbb{S}^{N-1}(\sqrt{N})$, and defined as

$$f_{\mathrm{sp}}(Y) \triangleq \arg\max\left\{\left\langle\theta, \theta Y^T\right\rangle : \theta \in \mathbb{S}^{N-1}(\sqrt{N})\right\}.$$

Remark 3. We observe that $\theta \in \mathbb{R}^N$ and $\|\theta\| = \sqrt{N}$ for all $\theta \in \Theta$. It follows that $\Theta \subseteq \mathbb{S}^{N-1}(\sqrt{N})$ and the spectral estimator is a relaxed maximum likelihood estimator over N-dimensional spheres of norm $\sqrt{N}$.

Definition 3.2. We define the space of all positive semi-definite $N \times N$ matrices with unit diagonal entries as

$$\mathcal{X} \triangleq \left\{X \in \mathbb{R}^{N\times N} : X_{ii} = 1 \text{ for all } i \in [N], X \succeq 0\right\}.$$

The **semi-definite program estimator** is denoted by $f_{\mathrm{SDP}} : \mathbb{R}^{N\times N} \to \mathcal{X}$, and defined as

$$f_{\mathrm{SDP}}(Y) \triangleq \arg\max\left\{\langle X, Y\rangle_F : X \in \mathcal{X}\right\}.$$

Remark 4. For all $\theta \in \Theta$, we observe that $\theta^T\theta \in \mathbb{R}^{N\times\mathbb{N}}$ is positive semi-definite and $\theta_i^2 = 1$ for all $i \in [N]$. Therefore, $\Theta\otimes\Theta \subseteq \mathcal{X}$ and the SDP estimator is a relaxed maximum likelihood estimator. We further observe that if all elements of $\mathcal{X}$ have an additional constraint of unit rank, then $\mathcal{X} = \Theta \otimes \Theta$.

We also recall that the Bayes estimator is given by $f_B : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ defined in terms of distribution $P(\theta \mid Y) \propto e^{\frac{\lambda}{2}\left\langle\theta, \theta Y^T\right\rangle}$ for all $\theta \in \Theta$, as

$$f_B(Y) \triangleq \mathbb{E}[\theta^T\theta \mid Y].$$

3.1 Asymptotic risk for different estimators for $\mathbb{Z}_2$ synchronization

Proposition 3.3 (BBP phase transition). *The following result holds true for spectral estimator.*

$$\left(1 - \frac{1}{\lambda^2}\right)\mathbb{1}_{\{\lambda>1\}} = \left(1 - \lim_{N\to\infty}\frac{1}{2N^2}\left\|f_{\mathrm{sp}}(Y)\otimes f_{\mathrm{sp}}(Y) - \theta\otimes\theta\right\|_F^2\right) = \lim_{N\to\infty}\frac{1}{N^2}\left\langle f_{\mathrm{sp}}(Y), \theta\right\rangle^2 \text{ almost surely.}$$

Proposition 3.4 (Bayes phase transition). *The following result holds true for Bayes estimator.*

$$q_*(\lambda)^2 \mathbb{1}_{\{\lambda>1\}} = \Big(1 - \lim_{N\to\infty} \frac{1}{N^2} \|f_B(Y) - \theta\otimes\theta\|_F^2\Big) = \lim_{N\to\infty}\frac{1}{N^2}\left\langle \theta, \theta f_B(Y)^T\right\rangle^2 \text{ almost surely,}$$

where N is a zero-mean unit-variance Gaussian and $q_*(\lambda)^2$ is the unique non-negative solution of

$$q = \mathbb{E}[\tanh(\lambda^2 q + \lambda\sqrt{q}N)^2].$$

Remark 5. The expected risk $R(Q,f,L) = \frac{1}{N}\mathbb{E}\left\|f(Y) - \theta^T\theta\right\|_F^2$ for matrix square loss function L, uniform distribution $Q \in \mathcal{M}(\Theta)$ and different estimators f can be computed as a function of signal-to-noise ratio λ. It turns out that $\lim_{\lambda\to\infty} R = 0$ for all estimators discussed above. Further, the expected risk R doesn't reduce for $\lambda < 1$; thus, $\lambda = 1$ is referred to as the **information-theoretic threshold**, at which a phase transition occurs. For $\lambda < 1$, the expected risk of the Bayes estimator and the zero estimator $f(Y) = 0$ are identical. For this model, the Bayes estimator can be computed efficiently; hence, there is no statistical-computational gap.

Proposition 3.5 (Asymptotic mean Bayesian risk). *For the Bayesian estimator, we have*

$$\lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}\,\|f_B(Y) - \theta\otimes\theta\|_F^2 = 1 - 2a_* + c_*,$$

where the two constants are defined as

$$a_* \triangleq \lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}\left\langle\theta, \theta f_B(Y)^T\right\rangle, \qquad c_* \triangleq \lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}\,\|f_B(Y)\|_F^2. \tag{2}$$

Proof. Recall that $\theta\in\Theta$ and hence $\|\theta\otimes\theta\|_F^2 = \operatorname{tr}\theta^T\theta\theta^T\theta = \langle\theta,\theta\rangle^2 = N^2$ and $\left\langle f_B(Y), \theta^T\theta\right\rangle_F = \operatorname{tr} f_B(Y)\theta^T\theta = \left\langle\theta, \theta f_B(Y)^T\right\rangle$. Thus, we can write the Frobenius norm of the difference as

$$\|f_B(Y) - \theta\otimes\theta\|_F^2 = \|f_B(Y)\|_F^2 + N^2 - 2\left\langle\theta, \theta f_B(Y)^T\right\rangle.$$

The result follows from taking expectation on both sides, dividing by N^2, and the definitions of a_*, c_*. ☐

3.2 Computing the asymptotic mean of an observable

Proposition 3.6. *Consider an N-particle interacting system with state space $\mathcal{X} \triangleq \mathcal{Z}^N$ and aggregate parametrized energy $E_\lambda : \mathcal{X} \to \mathbb{R}$ such that for any configuration $x \in \mathcal{X}$ the Boltzmann distribution is $\mu_{N,\beta,\lambda}(x) \propto e^{-\beta E_\lambda(x)}$. Let $m_N \triangleq \frac{\partial E_\lambda}{\partial\lambda}\big|_{\lambda=0}$, then*

$$m_* \triangleq \lim_{N\to\infty}\frac{1}{N}\langle m_N\rangle_{\beta,\lambda} = \frac{\partial f(\beta,\lambda)}{\partial\lambda}.$$

Proof. From Taylor series expansion of E_λ around $\lambda = 0$ for any configuration $x\in\mathcal{X}$, we get

$$E_\lambda(x) = E_0(x) + \lambda m_N(x) + \Theta(x^2).$$

Recall that the free energy is $F_N(\beta,\lambda) = -\frac{1}{\beta}\ln Z_N(\beta,\lambda) = -\frac{1}{\beta}\ln\int_{\mathcal{X}} e^{-\beta E_\lambda(x)}d\nu_0(x)$. Therefore, we obtain

$$\frac{\partial F_N(\beta,\lambda)}{\partial\lambda} = -\frac{1}{\beta}\frac{1}{Z_N(\beta,\lambda)}\int_{\mathcal{X}} e^{-\beta E_\lambda(x)}(-\beta m_N(x))d\nu_0(x) = \int_{\mathcal{X}} m_N(x)\mu_{N,\beta,\lambda}(x)\nu_0(x) = \langle m_N\rangle_{\beta,\lambda}.$$

The result follows from dividing by N on both sides and interchanging limit and derivative. ☐

Corollary 3.7. *Consider observation $Y = \frac{\lambda}{N}\theta^T\theta + W$ for $\mathbb{Z}_2 \triangleq \{-1,1\}$ synchronization model, and an N particle system with state space $\Theta \triangleq \mathbb{Z}_2^N$ with parametrized aggregate energy function $E_\lambda : \Theta\to\mathbb{R}$ defined for all configurations $x\in\Theta$, as*

$$E_\lambda(x) \triangleq -\frac{1}{2}\langle xW, x\rangle - \frac{\lambda}{2N}\langle x,\theta\rangle^2.$$

The Boltzmann distribution for this system is denoted by $\mu_{\beta,\lambda}\in\mathcal{M}(\Theta)$ and free energy density by $f(\beta,\lambda)$. Then, the constant a_* defined in (2) for the asymptotic Bayesian risk of the $\mathbb{Z}_2$ synchronization problem is

$$a_* = -2\frac{\partial f(\beta,\lambda)}{\partial\lambda}\bigg|_{\beta=\lambda}.$$

Proof. Recall that Bayesian estimator $f_B(Y) = \sum_{\sigma\in\Theta} \sigma^T\sigma P(\sigma \mid Y)$ where the conditional distribution $P(\sigma \mid Y) = \mu_{\beta,\lambda}(\sigma)|_{\beta=\lambda}$. Further, we observe that $\frac{\partial E_\lambda(\sigma)}{\partial\lambda} = -\frac{1}{2N}\langle\sigma,\theta\rangle^2$.

From the definition of conditional distribution $P(\sigma \mid Y)$ and Bayesian estimator $f_B(Y)$, we observe that

$$\frac{1}{N^2}\left\langle\theta,\theta f_B(Y)^T\right\rangle = \theta\sum_{\sigma\in\Theta}\sigma^T\sigma P(\sigma \mid Y)\theta^T = \sum_{\sigma\in\Theta}\frac{\langle\sigma,\theta\rangle^2}{N^2}P(\sigma \mid Y) = -\frac{2}{N}\left\langle\frac{\partial E_\lambda(\sigma)}{\partial\lambda}\right\rangle_{\beta,\lambda}\Bigg|_{\beta=\lambda}.$$

The result follows from taking limit $N\to\infty$ on both sides of the above equation, and applying Proposition 3.6 to the right hand side. □

Corollary 3.8. *Consider observation $Y = \frac{\lambda}{N}\theta^T\theta + W$ for $\mathbb{Z}_2 \triangleq \{-1,1\}$ synchronization model, and an N particle system with state space $\mathcal{X} \triangleq \Theta\times\Theta$ with parametrized aggregate energy function $E_{\lambda,h}:\mathcal{X}\to\mathbb{R}$ defined for all configurations $(x,z)\in\mathcal{X}$ as*

$$E_{\lambda,h}(x,z) \triangleq -\frac{1}{2}xWx^T - \frac{\lambda}{2N}\langle x,\theta\rangle^2 - \frac{1}{2}zWz^T - \frac{\lambda}{2N}\langle z,\theta\rangle^2 + \frac{h}{N}\langle x,z\rangle^2.$$

The Boltzmann distribution for this system is denoted by $\mu_{\beta,\lambda,h}\in\mathcal{M}(\mathcal{X})$ and free energy density by $f(\beta,\lambda,h)$. Then, the constant c_* defined in (2) *for the asymptotic Bayesian risk of the $\mathbb{Z}_2$ synchronization problem is*

$$c_* = \frac{\partial f(\beta,\lambda,h)}{\partial h}\bigg|_{\beta=\lambda,h=0}.$$

Proof. From the definition of Bayesian estimator $f_B(Y) = \sum_{\sigma\in\Theta}\sigma^T\sigma P(\sigma \mid Y)$, we can write

$$\|f_B(Y)\|_F^2 = \operatorname{tr} f_B(Y)f_B(Y)^T = \operatorname{tr}\sum_{x,z\in\Theta}x^TxP(x \mid Y)z^TzP(z \mid Y) = \sum_{x,z\in\Theta}\langle x,z\rangle^2 P(x \mid Y)P(z \mid Y).$$

From the definition of conditional distribution $P(x \mid Y)$, observation Y, and the definition of parametrized energy function $E_{\lambda,h}(x,z)$, it follows that $P(x \mid Y)P(z \mid Y) = \mu_{\beta,\lambda,h}(x,z)|_{\beta=\lambda,h=0}$ and $\frac{\partial E_{\lambda,h}(x,z)}{\partial h} = \frac{1}{N}\langle x,z\rangle^2$. Therefore,

$$\frac{1}{N^2}\|f_B(Y)\|_F^2 = \frac{1}{N}\left\langle\frac{\partial E_{\lambda,h}(x,z)}{\partial h}\right\rangle_{\beta,\lambda,h}\Bigg|_{\beta=\lambda,h=0}.$$

The result follows from taking limit $N\to\infty$ on both sides of the above equation, and applying Proposition 3.6 to the right hand side. □

Corollary 3.9. *Consider observation $Y = \frac{\lambda}{N}\theta^T\theta + W$ for $\mathbb{Z}_2 \triangleq \{-1,1\}$ synchronization model, and an N particle system with state space $\mathcal{X} \triangleq \mathbb{S}^{N-1}(\sqrt{N})$ with parametrized aggregate energy function $E_{\lambda,h}:\mathcal{X}\to\mathbb{R}$ defined for all configurations $(x,z)\in\mathcal{X}$ as*

$$E_{\lambda,h}(x) \triangleq -\frac{1}{2}\langle xY,x\rangle + \frac{h}{N}\langle x,\theta\rangle^2 = -\frac{1}{2}\langle xW,x\rangle - \frac{\lambda}{2N}\langle x,\theta\rangle^2 + \frac{h}{N}\langle x,\theta\rangle^2.$$

The Boltzmann distribution for this system is denoted by $\mu_{\beta,\lambda,h}\in\mathcal{M}(\mathcal{X})$ and free energy density by $f(\beta,\lambda,h)$. Then, we have

$$\lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}\langle f_{\mathrm{sp}}(Y),\theta\rangle^2 = \lim_{\beta\to\infty}\frac{\partial f(\beta,\lambda,h)}{\partial h}\bigg|_{h=0}.$$

Proof. We observe that $\frac{\partial E_{\lambda,h}(x)}{\partial h} = \frac{1}{N}\langle x,\theta\rangle^2$, and hence

$$\frac{1}{N^2}\langle f_{\mathrm{sp}}(Y),\theta\rangle^2 = \frac{1}{N}\frac{\partial E_{\lambda,h}(f_{\mathrm{sp}}(Y))}{\partial h}.$$

Recall that the spectral estimator for observation Y is

$$f_{\mathrm{sp}}(Y) = \arg\max\{\langle xY,x\rangle : x\in\mathcal{X}\} = \arg\max\{\mu_{\beta,\lambda,0}(x) : x\in\mathcal{X}\}.$$

Since the probability concentrates uniformly at the lowest energy states at $\beta = \infty$, we observe that

$$\frac{1}{N} \frac{\partial E_{\lambda,h}(f_{\mathrm{sp}}(Y))}{\partial h} = \frac{1}{N} \lim_{\beta \to \infty} \left\langle \frac{\partial \mathbb{E}_{\lambda,h}}{\partial h} \right\rangle_{\beta,\lambda,h} \Bigg|_{h=0} .$$

The result follows from taking limit $N \to \infty$ on both sides of the above equation, and applying Proposition 3.6 to the right hand side. □