Variational inference, spin glasses, and TAP free energy

Song Mei

April 20, 2021

Joint work with Zhou Fan and Andrea Montanari

1/27

Song Mei TAP free energy April 20, 2021

General motivation

- ▶ Bayesian inference: topic modeling, Bayesian GLM, Bayesian NN... High dim. integration is hard!
- ▶ Variational inference: integration \rightarrow optimization.
- Popular objective function: "mean field free energy".
- ... but not optimal for moderate SNR.
- ► Today: optimal objective "TAP free energy".

\mathbb{Z}_2 synchronization

► Signal:

$$oldsymbol{x} = [x_1, \dots, x_n]^\mathsf{T} \in \mathbb{Z}_2^n, \quad x_i \overset{i.i.d.}{\sim} \mathrm{Unif}(\mathbb{Z}_2), \quad \mathbb{Z}_2 = \{+1, -1\}.$$

▶ Observation $Y \in \mathbb{R}^{n \times n}$:

$$Y = \frac{\lambda}{n} x x^{\mathsf{T}} + W.$$

- ▶ Noise: $W \sim GOE(n)$.
- ▶ SNR $\lambda \in [0, \infty)$ fixed, dimension $n \to \infty$.
- ▶ Task: given $Y = (Y_{ij})$, estimate x (or say $X = xx^{\top}$).

Bayes estimation in \mathbb{Z}_2 synchronization

▶ Estimate $X = xx^{\mathsf{T}}$ with loss:

$$\ell(\boldsymbol{X},\widehat{\boldsymbol{X}}) = (1/n^2)\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\|_F^2.$$

- $lacksquare \operatorname{MSE}(\widehat{oldsymbol{X}}) = \mathbb{E}[(1/n^2)\|oldsymbol{X} \widehat{oldsymbol{X}}\|_F^2].$
- For $\lambda < 1$, impossible.
- For $\lambda > 1$, possible and efficient. BBP.
- The optimal estimator is the Bayes estimator:

$$\widehat{m{X}}_{ ext{Bayes}} = \mathbb{E}[m{x}m{x}^{\mathsf{T}}|m{Y}].$$

Song Mei

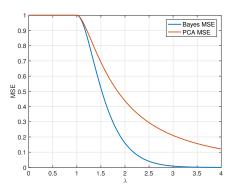
Bayes estimation in \mathbb{Z}_2 synchronization

Settings:

$$oldsymbol{x} \sim ext{Unif}(\mathbb{Z}_2^n), \qquad oldsymbol{Y} = (rac{\lambda}{n}) x x^\mathsf{T} + oldsymbol{W}.$$

► Risk:

$$ext{MSE}_{oldsymbol{\lambda}}(\widehat{oldsymbol{X}}) = (1/n^2) \mathbb{E}[\|oldsymbol{x} oldsymbol{x}^{\mathsf{T}} - \widehat{oldsymbol{X}}\|_F^2].$$



Compute the Bayesian estimator

► The Bayesian estimator:

$$\widehat{m{X}}_{ ext{Bayes}} = \mathbb{E}[m{x}m{x}^{\mathsf{T}}|m{Y}] = \sum_{m{\sigma} \in \mathbb{Z}_2^n} m{\sigma}m{\sigma}^{\mathsf{T}} p(m{\sigma}|m{Y}).$$

► The posterior distribution:

$$p(oldsymbol{\sigma}|oldsymbol{Y}) = rac{1}{Z} \exp\{rac{oldsymbol{\lambda}}{\langle oldsymbol{\sigma}, oldsymbol{Y}oldsymbol{\sigma}
angle}/2\}.$$

6 / 27

Song Mei TAP free energy April 20, 2021

► The posterior distribution:

$$p(\sigma|Y) = \frac{1}{Z} \exp\{\frac{\lambda}{\langle \sigma, Y\sigma \rangle/2}\}.$$

• Approximate $p(\sigma|Y)$ by $q \in \mathcal{P}_{\mathrm{MF}}$:

$$\mathcal{P}_{\mathrm{MF}} = \left\{ q(\boldsymbol{\sigma}) = \prod_{i=1}^n q_i(\sigma_i) : q_i \in \mathcal{P}(\mathbb{Z}_2) \right\} \cong [-1, 1]^n.$$

▶ Minimize the relative entropy between q and $p(\sigma|Y)$:

$$\min_{q \in \mathcal{P}_{\mathrm{MF}}} \mathsf{D}_{\mathrm{kl}}(q \| p(\boldsymbol{\sigma} | \boldsymbol{Y})).$$

ightharpoonup Equivalently minimizing $\min_{m{m}\in[-1,1]^n}\mathcal{F}_{\mathrm{MF}}(m{m})$

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - m{\lambda} \langle m{m}, m{Y}m{m}
angle / 2 \geq -\log Z,$$

where
$$h(m) = -\frac{1-m}{2}\log(\frac{1-m}{2}) - \frac{1+m}{2}\log(\frac{1+m}{2})$$

► The posterior distribution:

$$p(\boldsymbol{\sigma}|\boldsymbol{Y}) = rac{1}{Z} \exp\{ rac{\lambda}{\langle \boldsymbol{\sigma}, \boldsymbol{Y} \boldsymbol{\sigma}
angle / 2 \}.$$

▶ Approximate $p(\sigma|Y)$ by $q \in \mathcal{P}_{MF}$:

$$\mathcal{P}_{ ext{MF}} = \Bigr\{ q(oldsymbol{\sigma}) = \prod_{i=1}^n q_i(\sigma_i) : q_i \in \mathcal{P}(\mathbb{Z}_2) \Bigr\} \cong [-1,1]^n.$$

Minimize the relative entropy between q and $p(\sigma|Y)$:

$$\min_{q \in \mathcal{P}_{\mathrm{MF}}} \mathsf{D}_{\mathrm{kl}}(q \| p(\boldsymbol{\sigma} | \boldsymbol{Y})).$$

ightharpoonup Equivalently minimizing $\min_{m{m}\in[-1,1]^n}\mathcal{F}_{\mathrm{MF}}(m{m})$

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - m{\lambda} \langle m{m}, m{Y}m{m}
angle / 2 \geq -\log Z,$$

where
$$h(m) = -\frac{1-m}{2}\log(\frac{1-m}{2}) - \frac{1+m}{2}\log(\frac{1+m}{2})$$
.

► The posterior distribution:

$$p(\boldsymbol{\sigma}|\boldsymbol{Y}) = rac{1}{Z} \exp\{ rac{\lambda}{\langle \boldsymbol{\sigma}, \boldsymbol{Y} \boldsymbol{\sigma}
angle / 2 \}.$$

▶ Approximate $p(\sigma|Y)$ by $q \in \mathcal{P}_{MF}$:

$$\mathcal{P}_{\mathrm{MF}} = \Bigr\{ q(oldsymbol{\sigma}) = \prod_{i=1}^n q_i(\sigma_i) : q_i \in \mathcal{P}(\mathbb{Z}_2) \Bigr\} \cong [-1,1]^n.$$

▶ Minimize the relative entropy between q and $p(\sigma|Y)$:

$$\min_{q \in \mathcal{P}_{\mathrm{MF}}} \mathsf{D}_{\mathrm{kl}}(q \| p(oldsymbol{\sigma} | oldsymbol{Y})).$$

ightharpoonup Equivalently minimizing $\min_{m{m}\in[-1,1]^n}\mathcal{F}_{\mathrm{MF}}(m{m})$

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - m{\lambda} \langle m{m}, m{Y}m{m}
angle / 2 \geq -\log Z,$$

where $h(m) = -\frac{1-m}{2} \log(\frac{1-m}{2}) - \frac{1+m}{2} \log(\frac{1+m}{2})$

► The posterior distribution:

$$p(\boldsymbol{\sigma}|\boldsymbol{Y}) = rac{1}{Z} \exp\{ rac{\lambda}{\langle \boldsymbol{\sigma}, \boldsymbol{Y} \boldsymbol{\sigma}
angle / 2 \}.$$

▶ Approximate $p(\sigma|Y)$ by $q \in \mathcal{P}_{MF}$:

$$\mathcal{P}_{\mathrm{MF}} = \Bigr\{ q(oldsymbol{\sigma}) = \prod_{i=1}^n q_i(\sigma_i) : q_i \in \mathcal{P}(\mathbb{Z}_2) \Bigr\} \cong [-1,1]^n.$$

▶ Minimize the relative entropy between q and $p(\sigma|Y)$:

$$\min_{q \in \mathcal{P}_{\mathrm{MF}}} \mathsf{D}_{\mathrm{kl}}(q \| p(oldsymbol{\sigma} | oldsymbol{Y})).$$

• Equivalently minimizing $\min_{m \in [-1,1]^n} \mathcal{F}_{\mathrm{MF}}(m)$

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{m{\lambda}} \langle m{m}, m{Y}m{m}
angle / 2 \geq -\log Z,$$

where
$$\mathsf{h}(m) = -\frac{1-m}{2}\log(\frac{1-m}{2}) - \frac{1+m}{2}\log(\frac{1+m}{2})$$
.

Mean field free energy:

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{m{\lambda}} \langle m{m}, m{Y}m{m}
angle / 2.$$

lacksquare For $m_\star = rg \min_{m{m}} \mathcal{F}_{\mathrm{MF}}(m{m})$, we hope

$$oldsymbol{m}_{\star}oldsymbol{m}_{\star}^{\mathsf{T}}pprox\widehat{oldsymbol{X}}_{\mathsf{Bayes}}=\mathbb{E}[oldsymbol{x}oldsymbol{x}^{\mathsf{T}}|oldsymbol{Y}].$$

- lt was shown that $m_{\star}m_{\star}^{\top} \not\approx \mathbb{E}[xx^{\top}|Y]$ [Ghorbani, Javadi, and Montanari, 2017].
- ► The assumption that posterior distribution can be approximately factorized into the product of marginals is wrong!

◆ロト ◆個ト ◆重ト ◆重ト 重 めなべ

► Mean field free energy:

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{m{\lambda}} \langle m{m}, m{Y}m{m}
angle / 2.$$

lacktriangle For $m_\star = \mathop{\mathrm{arg\,min}}_m \mathcal{F}_{\mathrm{MF}}(m)$, we hope

$$oldsymbol{m}_{\star}oldsymbol{m}_{\star}^{\mathsf{T}}pprox\widehat{oldsymbol{X}}_{\mathsf{Bayes}}=\mathbb{E}[oldsymbol{x}oldsymbol{x}^{\mathsf{T}}|oldsymbol{Y}].$$

- lt was shown that $m_{\star}m_{\star}^{\top} \approx \mathbb{E}[xx^{\top}|Y]$ [Ghorbani, Javadi, and Montanari, 2017].
- ► The assumption that posterior distribution can be approximately factorized into the product of marginals is wrong!

◆ロト ◆問 ト ◆ 恵 ト ◆ 恵 ・ り へ ○

► Mean field free energy:

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{m{\lambda}} \langle m{m}, m{Y}m{m}
angle / 2.$$

lacktriangle For $m_\star = rg \min_{m{m}} \mathcal{F}_{\mathrm{MF}}(m{m})$, we hope

$$oldsymbol{m}_{\star}oldsymbol{m}_{\star}^{\mathsf{T}}pprox\widehat{oldsymbol{X}}_{\mathsf{Bayes}}=\mathbb{E}[oldsymbol{x}oldsymbol{x}^{\mathsf{T}}|oldsymbol{Y}].$$

- ▶ It was shown that $m_{\star}m_{\star}^{\mathsf{T}} \not\approx \mathbb{E}[xx^{\mathsf{T}}|Y]$ [Ghorbani, Javadi, and Montanari, 2017].
- ► The assumption that posterior distribution can be approximately factorized into the product of marginals is wrong!

◆ロト ◆部ト ◆恵ト ◆恵ト ・恵 ・ からで

Mean field free energy:

$$\mathcal{F}_{ ext{MF}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{m{\lambda}} \langle m{m}, m{Y}m{m}
angle / 2.$$

• For $m_{\star} = \operatorname{arg\,min}_{m} \mathcal{F}_{\mathrm{MF}}(m)$, we hope

$$m{m}_{\star}m{m}_{\star}^{\mathsf{T}}pprox\widehat{m{X}}_{\mathsf{Bayes}}=\mathbb{E}[m{x}m{x}^{\mathsf{T}}|m{Y}].$$

- ▶ It was shown that $m_{\star}m_{\star}^{\mathsf{T}} \not\approx \mathbb{E}[xx^{\mathsf{T}}|Y]$ [Ghorbani, Javadi, and Montanari, 2017].
- The assumption that posterior distribution can be approximately factorized into the product of marginals is wrong!

Thouless, Anderson, and Palmer (1977) proposed the TAP free energy when they study the Sherrington-Kirkpatrick model, whose Gibbs measure gives

$$G_{oldsymbol{eta},\lambda}(oldsymbol{\sigma}) = rac{1}{Z_{oldsymbol{eta},\lambda}} \exp\{oldsymbol{eta}\langle oldsymbol{\sigma}, oldsymbol{Y}oldsymbol{\sigma}
angle\}.$$

where $Y_{ij} \sim \mathcal{N}(\lambda/n, 1/n)$.

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{\lambda}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{n\lambda^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

April 20, 2021

► Thouless, Anderson, and Palmer (1977) proposed the TAP free energy when they study the Sherrington-Kirkpatrick model, whose Gibbs measure gives

$$G_{oldsymbol{eta},\lambda}(oldsymbol{\sigma}) = rac{1}{Z_{oldsymbol{eta},\lambda}} \exp\{oldsymbol{eta}\langle oldsymbol{\sigma}, oldsymbol{Y}oldsymbol{\sigma}
angle\}.$$

where $Y_{ij} \sim \mathcal{N}(\lambda/n, 1/n)$.

- ▶ When $\beta = \lambda$, the Gibbs measure of SK model is the same as the posterior of \mathbb{Z}_2 synchronization
- ► The TAP free energy (when $\beta = \lambda$) gives

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{\lambda}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{n\lambda^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

◆ロト ◆問ト ◆言ト ◆言ト ・言・ かへで

► Thouless, Anderson, and Palmer (1977) proposed the TAP free energy when they study the Sherrington-Kirkpatrick model, whose Gibbs measure gives

$$G_{oldsymbol{eta},\lambda}(oldsymbol{\sigma}) = rac{1}{Z_{oldsymbol{eta},\lambda}} \exp\{oldsymbol{eta}\langle oldsymbol{\sigma}, oldsymbol{Y}oldsymbol{\sigma}
angle\}.$$

where $Y_{ij} \sim \mathcal{N}(\frac{\lambda}{n}, 1/n)$.

- ▶ When $\beta = \lambda$, the Gibbs measure of SK model is the same as the posterior of \mathbb{Z}_2 synchronization
- ► The TAP free energy (when $\beta = \lambda$) gives

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{m{n}m{\lambda}^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

► The TAP free energy

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{m{n}m{\lambda}^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

lacktriangleright For $m_\star = rg \min_m \mathcal{F}_{ exttt{TAP}}(m)$, we hope

$$oldsymbol{m}_{\star}oldsymbol{m}_{\star}^{\mathsf{T}}pprox\widehat{oldsymbol{X}}_{\mathsf{Bayes}}=\mathbb{E}[oldsymbol{x}oldsymbol{x}^{\mathsf{T}}|oldsymbol{Y}].$$

▶ Our main theorem shows that this is correct.

| ◀□▶ ◀圖▶ ◀圖▶ | 圖 | 釣९♡

10 / 27

Song Mei TAP free energy April 20, 2021

► The TAP free energy

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{m{n}m{\lambda}^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

lacktriangleq For $m_\star = rg \min_m \mathcal{F}_{ exttt{TAP}}(m)$, we hope

$$m{m}_{\star}m{m}_{\star}^{\mathsf{T}}pprox\widehat{m{X}}_{\mathsf{Bayes}}=\mathbb{E}[m{x}m{x}^{\mathsf{T}}|m{Y}].$$

▶ Our main theorem shows that this is correct.

◆ロト ◆御 ト ◆ 恵 ト ◆ 恵 ・ り へ ○

Song Mei

► The TAP free energy

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv \underbrace{-\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{m{\lambda}}{2} \langle m{m}, m{Y}m{m}
angle}_{\mathcal{F}_{ ext{MF}}} \underbrace{-rac{m{n}m{\lambda}^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2}_{ ext{Onsager's correction term}}.$$

lacktriangleright For $m_\star = rg \min_m \mathcal{F}_{ exttt{TAP}}(m)$, we hope

$$m{m}_{\star}m{m}_{\star}^{\mathsf{T}}pprox\widehat{m{X}}_{\mathsf{Bayes}}=\mathbb{E}[m{x}m{x}^{\mathsf{T}}|m{Y}].$$

▶ Our main theorem shows that this is correct.

◆ロト ◆部ト ◆恵ト ◆恵ト ・恵 ・ からぐ

10 / 27

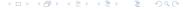
Song Mei TAP free energy April 20, 2021

Remarks

- Derivation of TAP can be obtained by expectation consistency/Plefka's expansion.
- ▶ The stationery equation of TAP $\nabla \mathcal{F}_{TAP}(m) = 0$ is the fixed point equation for AMP algorithm

$$m^{k+1} = AMP(m^k, m^{k-1}), \qquad m^{k+1} = m^k = m^{k-1}.$$

▶ AMP is not a descent algorithm on TAP free energy.



11 / 27

ng Mei TAP free energy April 20, 2021

Main theorem

Theorem (Fan, M., Montanari, 2018)

Denote $C_{\lambda,n} = \{ m \in [-1,1]^n : \nabla \mathcal{F}_{TAP}(m) = 0, \mathcal{F}_{TAP}(m) \leq -\lambda^2/3 \}$. There exists $\lambda_0 > 0$, such that for any $\lambda > \lambda_0$, we have

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{\boldsymbol{m}\in\mathcal{C}_{\lambda,n}} \frac{1}{n^2} \|\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} - \widehat{\boldsymbol{X}}_{\mathrm{Bayes}}\|_F^2 \wedge 1\Big] = 0. \tag{1}$$

All the critical points (below a threshold) are close to the Bayesian estimator.

Main theorem

Theorem (Fan, M., Montanari, 2018)

Denote $C_{\lambda,n} = \{ m \in [-1,1]^n : \nabla \mathcal{F}_{TAP}(m) = 0, \mathcal{F}_{TAP}(m) \leq -\lambda^2/3 \}$. There exists $\lambda_0 > 0$, such that for any $\lambda > \lambda_0$, we have

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{\boldsymbol{m}\in\mathcal{C}_{\lambda,n}} \frac{1}{n^2} \|\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} - \widehat{\boldsymbol{X}}_{\mathrm{Bayes}}\|_F^2 \wedge 1\Big] = 0. \tag{1}$$

All the critical points (below a threshold) are close to the Bayesian estimator.

◆ロト ◆部ト ◆恵ト ◆恵ト ・恵 ・ からぐ

Related literatures in spin glass theory

TAP free energy in unbiased SK.

- ► TAP equations: [Talagrand, 2004], [Chatterjee, 2009], [Chen, 2011], [Auffinger and Jagannath, 2016], Posterior means/Pure states satisfy TAP equations.
- ► TAP free energy: [Chen and Panchenko, 2017], constrained TAP minimum are exact.

Calculating the complexity.

► [Auffinger, Ben Arous, and Cerny, 2010], [Subag, 2016].

And a few more after 2018

Main theorem

Theorem (Fan, M., Montanari, 2018)

Denote $C_{\lambda,n} = \{ m \in [-1,1]^n : \nabla \mathcal{F}_{TAP}(m) = 0, \mathcal{F}_{TAP}(m) \leq -\lambda^2/3 \}$. There exists $\lambda_0 > 0$, such that for any $\lambda > \lambda_0$, we have

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{\boldsymbol{m}\in\mathcal{C}_{\lambda,n}} \frac{1}{n^2} \|\boldsymbol{m}\boldsymbol{m}^\mathsf{T} - \widehat{\boldsymbol{X}}_{\mathrm{Bayes}}\|_F^2 \wedge 1\Big] = 0. \tag{1}$$

All the critical points (below a threshold) are close to the Bayesian estimator.

◆ロト ◆部ト ◆恵ト ◆恵ト ・恵 ・ からぐ

14 / 27

ong Mei TAP free energy April 20, 2021

► Recall

$$\mathcal{F}_{ ext{TAP}}(m{m}) \equiv -\sum_{i=1}^n \mathsf{h}(m{m}_i) - rac{\pmb{\lambda}}{2} \langle m{m}, m{Y}m{m}
angle - rac{m{n} \pmb{\lambda}^2}{4} \Big[1 - rac{\|m{m}\|_2^2}{n}\Big]^2.$$

ightharpoonup Define some important statistics of m:

$$E(oldsymbol{m}) = \mathcal{F}_{ ext{TAP}}(oldsymbol{m})/n, \quad Q(oldsymbol{m}) = \|oldsymbol{m}\|_2^2/n, \quad M(oldsymbol{m}) = \langle oldsymbol{m}, oldsymbol{x}
angle/n.$$

▶ For any $U \subseteq \mathbb{R}^3$, define

$$\operatorname{Crit}_n(U) \equiv \#\{m: \nabla E(m) = 0, (Q(m), M(m), E(m)) \in U\}.$$
 (2)

◆ロト ◆問ト ◆見ト ◆見ト ・ 更 ・ 釣へ○・

Calculating the Crit: Kac-Rice formula

Lemma (Kac-Rice formula, c.f. [Adler and Taylor, 2007)

] Let $f: \mathbb{R}^d \to \mathbb{R}$ be a "sufficiently regular" random morse function. Let $p_{\boldsymbol{m}}(\boldsymbol{z})$ be the density of $\nabla f(\boldsymbol{m})$ at \boldsymbol{z} . For any Borel measurable set $T \subseteq \mathbb{R}^d$, denote

$$Crit(T) = \#\{\boldsymbol{m} \in T : \nabla f(\boldsymbol{m}) = \boldsymbol{0}\}.$$

Then

$$egin{aligned} \mathbb{E}[\operatorname{Crit}(T)] = & \mathbb{E}\Big[\int_{T} ig| \det
abla^2 f(oldsymbol{m}) ig| \cdot \delta(
abla f(oldsymbol{m})) \cdot doldsymbol{m} \Big] \ = & \int_{T} \mathbb{E}\Big[ig| \det
abla^2 f(oldsymbol{m}) ig| ig|
abla f(oldsymbol{m}) = oldsymbol{0} ig| p_{oldsymbol{m}}(oldsymbol{0}) doldsymbol{m}. \end{aligned}$$

▶ $|\det \nabla^2 f(m)|$ is the correct weight function so that each critical point count exactly once.

- 4 ロ ト 4 昼 ト 4 昼 ト - 夏 - か 9 0 0

Proposition

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e) + o(n)\Big\}.$$

$$S_{\star}(q,arphi,e) = \sup_{a \in \mathbb{R}} \inf_{(\mu,
u, au, au,\gamma) \in \mathbb{R}^4} S(q,arphi,a,e;\mu,
u, au,\gamma),$$

where

$$egin{aligned} S(q,arphi,a,e;\mu,
u, au,\gamma) = &rac{1}{4eta^2} \Big[rac{a}{q} - rac{eta\lambdaarphi^2}{q} - eta^2(1-q)\Big]^2 \ &-q\mu - arphi
u - a au - \Big[-rac{eta^2}{4}(1-q^2) + rac{a}{2} - e\Big]\gamma + \log I, \end{aligned}$$

and

$$I = \int_{-\infty}^{\infty} rac{1}{(2\pieta^2q)^{1/2}} \exp\Big\{-rac{(x-eta\lambdaarphi)^2}{2eta^2q} \ + \mu anh^2(x) +
u anh(x) + anh(x) + anh(x) + \gamma \log[2\cosh(x)]\Big\} \mathrm{d}x.$$

▶ Key proposition: for $U \subseteq \mathbb{R}^3$,

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\overbrace{\sup_{(q,arphi,e)\in U} S_{\star}(q,arphi,e)}^{T(U)} + o(n)\Big\},$$

- For any U such that T(U) > 0, there could potentially be critical points of \mathcal{F}_{TAP} in U.
- For any U such that T(U) < 0, there is no critical points of \mathcal{F}_{TAP} in U, with high probability.
- If we admit the key proposition, suffice to show that T(U) < 0 unless U contains a neighborhood of the Bayes estimator.

▶ Key proposition: for $U \subseteq \mathbb{R}^3$,

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\overbrace{\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e)}^{T(U)} + o(n)\Big\},$$

- For any U such that T(U) > 0, there could potentially be critical points of \mathcal{F}_{TAP} in U.
- For any U such that T(U) < 0, there is no critical points of \mathcal{F}_{TAP} in U, with high probability.
- If we admit the key proposition, suffice to show that T(U) < 0 unless U contains a neighborhood of the Bayes estimator.

▶ Key proposition: for $U \subseteq \mathbb{R}^3$,

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\overbrace{\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e)}^{T(U)} + o(n)\Big\},$$

- For any U such that T(U) > 0, there could potentially be critical points of \mathcal{F}_{TAP} in U.
- For any U such that T(U) < 0, there is no critical points of \mathcal{F}_{TAP} in U, with high probability.
- If we admit the key proposition, suffice to show that T(U) < 0 unless U contains a neighborhood of the Bayes estimator.

18 / 27

ng Mei TAP free energy April 20, 2021

▶ Key proposition: for $U \subseteq \mathbb{R}^3$,

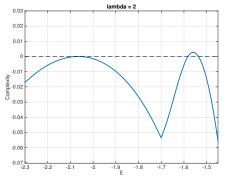
$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\overbrace{\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e)}^{T(U)} + o(n)\Big\},$$

- For any U such that T(U) > 0, there could potentially be critical points of \mathcal{F}_{TAP} in U.
- For any U such that T(U) < 0, there is no critical points of \mathcal{F}_{TAP} in U, with high probability.
- If we admit the key proposition, suffice to show that T(U) < 0 unless U contains a neighborhood of the Bayes estimator.

18 / 27

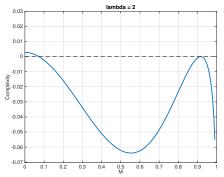
ng Mei TAP free energy April 20, 2021

 $ightharpoonup S_{\star}(e) = \sup_{q, arphi} S_{\star}(q, arphi, e).$



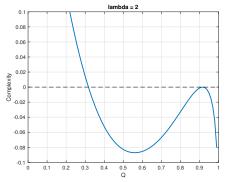
Song Mei

 $ightharpoonup S_{\star}(arphi) = \sup_{q,e} S_{\star}(q,arphi,e).$



 $\blacktriangleright \text{ At } \varphi_{\star}, \ S_{\star}(\varphi_{\star}) = 0.$

 $ightharpoonup S_{\star}(q) = \sup_{\varphi,e} S_{\star}(q,\varphi,e).$



 $\blacktriangleright \text{ At } q_{\star}, \ S_{\star}(q_{\star}) = 0.$

Song Mei

There exists λ_0 , for $\lambda \geq \lambda_0$,

- $\begin{array}{l} \blacktriangleright \;\; S_{\star}(q_{\star},\varphi_{\star},e_{\star}) = \mathsf{0}, \; \text{where} \; (q_{\star},\varphi_{\star},e_{\star}) \approx (Q(m_{\star}),M(m_{\star}),E(m_{\star})) \\ \text{for} \;\; \widehat{\boldsymbol{X}}_{\text{Bayes}} \approx m_{\star}m_{\star}^{\mathsf{T}}. \end{array}$
- $ightharpoonup S_{\star}(q, \varphi, e) < 0 ext{ for any } e \leq -\lambda^2/3 ext{ and } (q, \varphi, e)
 eq (q_{\star}, \varphi_{\star}, e_{\star}).$

The proof of these two properties is more than calculus. It requires bounds using concentration inequalities.

Combining with the key inequality it is easy to show the main theorem

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\sup_{(q,arphi,e)\in U} S_{\star}(q,arphi,e) + o(n)\Big\}.$$

Now suffice to show the key inequality.

- 4 ロ b - 4 個 b - 4 種 b - 4 種 b - 9 9 9 0 0 0 0

There exists λ_0 , for $\lambda \geq \lambda_0$,

- $\begin{array}{l} \blacktriangleright \;\; S_{\star}(q_{\star},\varphi_{\star},e_{\star}) = \mathsf{0}, \; \text{where} \; (q_{\star},\varphi_{\star},e_{\star}) \approx (Q(m_{\star}),M(m_{\star}),E(m_{\star})) \\ \text{for} \;\; \widehat{\boldsymbol{X}}_{\text{Bayes}} \approx m_{\star}m_{\star}^{\mathsf{T}}. \end{array}$
- $ightharpoonup S_{\star}(q, \varphi, e) < 0 ext{ for any } e \leq -\lambda^2/3 ext{ and } (q, \varphi, e)
 eq (q_{\star}, \varphi_{\star}, e_{\star}).$

The proof of these two properties is more than calculus. It requires bounds using concentration inequalities.

Combining with the key inequality it is easy to show the main theorem

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e) + o(n)\Big\}.$$

Now suffice to show the key inequality.

- ◀ □ ▶ ◀ 圖 ▶ ◀ 圖 ▶ ¶ ♀ ♡ Q (~)

There exists λ_0 , for $\lambda \geq \lambda_0$,

- $\begin{array}{l} \blacktriangleright \;\; S_{\star}(q_{\star},\varphi_{\star},e_{\star}) = 0, \; \text{where} \; (q_{\star},\varphi_{\star},e_{\star}) \approx (Q(m_{\star}),M(m_{\star}),E(m_{\star})) \\ \text{for} \;\; \widehat{\boldsymbol{X}}_{\text{Bayes}} \approx m_{\star}m_{\star}^{\mathsf{T}}. \end{array}$
- $ightharpoonup S_{\star}(q, \varphi, e) < 0 ext{ for any } e \leq -\lambda^2/3 ext{ and } (q, \varphi, e)
 eq (q_{\star}, \varphi_{\star}, e_{\star}).$

The proof of these two properties is more than calculus. It requires bounds using concentration inequalities.

Combining with the key inequality it is easy to show the main theorem.

$$\mathbb{E}[\operatorname{Crit}_n(U)] \leq \exp\Big\{n\sup_{(q,arphi,e)\in U} S_\star(q,arphi,e) + o(n)\Big\}.$$

Now suffice to show the key inequality.

- 4 □ ▶ 4 圖 ▶ 4 圖 ▶ 9 9 0 0

Dealing with determinant of Hessian

▶ The conditional Hessian is distributed as (up to some scaling)

$$[
abla^2 \mathcal{F}_{\mathrm{TAP}}(m{m}) |
abla \mathcal{F}_{\mathrm{TAP}}(m{m}) = m{0}] \stackrel{d}{=} m{D} + m{W} + \mathrm{low} \; \mathrm{rank} \; \mathrm{perturbation},$$
 where $m{D} = \mathrm{diag}(m{d}_i)$, and $m{W} \sim \mathrm{GOE}(m{n})$.

▶ The low rank perturbation has vanishing effects. Therefore, we just need to calculate $\mathbb{E}[|\det(\boldsymbol{H})|]$, with

$$H = D + W$$
.

Song Mei TAP fre

23 / 27

Determinant of Hessian: "replica trick"

$$\boldsymbol{H} = \boldsymbol{D} + \boldsymbol{W} = \operatorname{diagonal} + \operatorname{GOE}.$$

We have

$$egin{aligned} \mathbb{E}[|\det(oldsymbol{H})|] &pprox \mathbb{E}[|\det(oldsymbol{H})|^{-1/2}]^2 \ &= \mathbb{E}\Big[\int_{\mathbb{R}^d} \exp\{-\langle oldsymbol{x}, oldsymbol{H}oldsymbol{x}
angle/2\} \mathrm{d}oldsymbol{x}\Big]^2 \ &= \mathbb{E}\Big[\int_{\mathbb{R}^d} \exp\{-\langle oldsymbol{x}, oldsymbol{W}oldsymbol{x}
angle/2 - \langle oldsymbol{x}, oldsymbol{D}oldsymbol{x}
angle\Big]^2 \end{aligned}$$

Determinant of Hessian: Stieltjes transform

$$m{H} = m{D} + m{W} = ext{diagonal} + ext{GOE}.$$
 $rac{1}{n}\log|\det(m{H})| = rac{1}{n}\log\prod_{i=1}^n|\lambda_i(m{H})| = \int_{\mathbb{R}}\log|x|\cdot\mu_{m{H}}(\mathrm{d}x),$ where $\mu_{m{H}} = (1/n)\sum_{i=1}^n\delta(\lambda_i(m{H})).$

- The Stieltjes transform of μ_H can be approximately calculated using free probability theory.
- ightharpoonup Once the Stieltjes transform of μ_H is known, the quantity $\mathbb{E}\left[\int_{\mathbb{R}}(\log|x|)\mu_H(\mathrm{d}x)\right]$ can be computed.

◆ロト ◆部ト ◆恵ト ◆恵ト ・恵 ・ 釣へ○

Determinant of Hessian: Stieltjes transform

$$m{H} = m{D} + m{W} = ext{diagonal} + ext{GOE}.$$
 $rac{1}{n}\log|\det(m{H})| = rac{1}{n}\log\prod_{i=1}^n|\lambda_i(m{H})| = \int_{\mathbb{R}}\log|x|\cdot \mu_{m{H}}(\mathrm{d}x),$ where $\mu_{m{H}} = (1/n)\sum_{i=1}^n\delta(\lambda_i(m{H})).$

- \triangleright The Stieltjes transform of μ_H can be approximately calculated using free probability theory.
- \triangleright Once the Stieltjes transform of μ_H is known, the quantity $\mathbb{E}\left[\int_{\mathbb{R}}(\log|x|)\mu_{H}(\mathrm{d}x)\right]$ can be computed.

Free convolution of two distribution

Let $A \in \mathbb{R}^{n \times n}$, and $\mu_A = (1/n) \sum_{i=1}^n \delta(\lambda_i(A))$. For any $z \in \mathbb{C}_+$, the Stieltjes transform of μ_A is defined as

$$g_{oldsymbol{A}}(z) = \int_{\mathbb{R}} rac{1}{x-z} \mu_{oldsymbol{A}}(\mathrm{d}x) = rac{1}{n} \sum_{i=1}^n rac{1}{\lambda_i(oldsymbol{A})-z}.$$

Lemma (Free probability theory heuristics/Leave one out)

Let $oldsymbol{D} = ext{diag}(d_i)$ be a diagonal matrix, and let $oldsymbol{H} = oldsymbol{D} + oldsymbol{W}$. Then

$$\mathbb{E}g_{H}(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_{i} - z - \mathbb{E}g_{H}(z)} + o_{n}(1).$$
 (3)

Free convolution of two distribution

Let $A \in \mathbb{R}^{n \times n}$, and $\mu_A = (1/n) \sum_{i=1}^n \delta(\lambda_i(A))$. For any $z \in \mathbb{C}_+$, the Stieltjes transform of μ_A is defined as

$$g_{oldsymbol{A}}(z) = \int_{\mathbb{R}} rac{1}{x-z} \mu_{oldsymbol{A}}(\mathrm{d}x) = rac{1}{n} \sum_{i=1}^n rac{1}{\lambda_i(oldsymbol{A})-z}.$$

Lemma (Free probability theory heuristics/Leave one out)

Let $oldsymbol{D} = \operatorname{diag}(d_i)$ be a diagonal matrix, and let $oldsymbol{H} = oldsymbol{D} + oldsymbol{W}$. Then

$$\mathbb{E}g_{H}(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{d_{i} - z - \mathbb{E}g_{H}(z)} + o_{n}(1). \tag{3}$$

Summary

- ▶ TAP free energy is accurate for \mathbb{Z}_2 synchronization.
- Can be generalized to topic modeling, low rank matrix estimation, compressed sensing, etc...
- More sophiscated variational inference beyond naive mean field.

Summary

- ▶ TAP free energy is accurate for \mathbb{Z}_2 synchronization.
- ► Can be generalized to topic modeling, low rank matrix estimation, compressed sensing, etc...
- More sophiscated variational inference beyond naive mean field.

Summary

- ▶ TAP free energy is accurate for \mathbb{Z}_2 synchronization.
- Can be generalized to topic modeling, low rank matrix estimation, compressed sensing, etc...
- ▶ More sophiscated variational inference beyond naive mean field.