arXiv:2304.04031v1 [math.PR] 8 Apr 2023

受约束重叠多球面Sherrington-Kirkpatrick模型的TAP变分原理

David Belius*,**, Leon Fr¨ober*, Justin Ko†

Abstract. 涉及多个具有受约束重叠的副本的自旋玻璃模型已在[FPV92；PT07；Pan18a]中得到研究。对于这些模型的球面版本， [Ko19；Ko20]证明了，极限自由能由Parisi型极小化给出。在本文中，我们证明，对于Sherrington-Kirkpatrick（即2-自旋）相互作用，它也可以用Thouless-Andersson-Palmer（TAP）变分原理表示。这是第二个在所有温度和外场下实现自由能严格TAP计算的自旋玻璃模型。我们在此推导出的变分公式还证实，该模型是副本对称的；这一事实虽然自然，却无法从其Parisi公式中显然推出。

1. Introduction

我们研究了[FPV92；PT07；Pan18a], 受约束的多副本自旋玻璃模型的自由能，也称为向量自旋模型。在物理学中，这一自由能被称为Franz–Parisi势[FPV92]。该模型包含具有受约束重叠的多个副本，最初是为了研究标准单副本自旋玻璃的亚稳态而提出的[FPV92], ，此后还被用于研究单副本模型的其他一些性质[Pan16；CP17；BJ18；Jag19；AJ21；AK18；FR20；JLM20]。

我们通过将[BK19]的Thouless-Andersson-Palmer（TAP）方法应用于该模型的球面Sherrington-Kirkpatrick（SK；即2-自旋）版本，提出了一种研究该模型的新方法。我们证明了一个以TAP自由能表示其自由能的变分公式，并计算出了最大TAP自由能的公式，从而为原始自由能得到了一个具体公式。在[BK19]之后，这是第二个在所有温度和外场下，使用严格数学意义上的TAP方法计算出自旋玻璃模型自由能的情形。

现在我们正式介绍该模型。2-自旋SK哈密顿量是如下形式的高斯过程

$$H_N(\sigma) = \sqrt{N} \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j \tag{1.1}$$

由 $\sigma \in \mathbb{R}^N$ 索引，其中 J_{ij} 是独立同分布的标准高斯随机变量。对于 $n \geq 1$，我们考虑由矩阵表示的 n 个副本的多自旋构型。

$$\boldsymbol{\sigma} = (\sigma^1, \dots, \sigma^n) \in \mathbb{R}^{n\times N},$$

电子邮箱地址：david.belius@cantab.net、leon.froeber@unibas.ch、justin.ko@ens-lyon.fr。

* 由SNSF资助项目176918支持。** 由SNSF资助项目206148支持。

† 由ERC项目LDRAM支持：ERC-2019-ADG。

其中每个 $\sigma^k \in \mathbb{R}^N$ 表示 $\boldsymbol{\sigma}$ 的第 k行，而 σ_i^k 表示第 i 列第 k 行中的元素。令

$$\mathcal{S}_{N-1} = \{\sigma \in \mathbb{R}^N : \sigma_1^2 + ... + \sigma_N^2 = 1\}$$

表示 $\mathbb{R}^N$ 中的单位球面。令 $\boldsymbol{h} = (h^1、...、h^n) \in \mathbb{R}^{n\times N}$和 $\beta = (\beta_1、...、\beta_n)$ 分别表示每个副本的外场和逆温度。此外，假设 $|h^k| = h_k \in \mathbb{R}$ 并令 $h = (h_1、...、h_n) \in \mathbb{R}^n$。令 $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ 为一个对角线上全为 1 的半正定矩阵，用于约束副本之间的重叠。对于矩阵 $\boldsymbol{A}$，令 $\|\boldsymbol{A}\|_\infty$表示上确界范数 $\max_{k,\ l}|A_{k,\ l}|$，而对于 $\varepsilon > 0$，令

$$\boldsymbol{Q}_\varepsilon = \{\boldsymbol{\sigma} : \|\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal - \boldsymbol{Q}\|_\infty \le \varepsilon\}, \tag{1.2}$$

表示重叠接近 $\boldsymbol{Q}$ 的副本集合。

我们的目标是计算副本约束自由能的极限

$$F_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + N h^k\cdot\sigma^k} d\boldsymbol{\sigma} \tag{1.3}$$

对于固定的模型参数 $\boldsymbol{Q}, \boldsymbol{h}, \beta$，其中 $d\boldsymbol{\sigma} = (d\sigma)^{\otimes n}$ 是球面 $\mathcal{S}_{N-1}$ 上均匀测度 $d\sigma$ 的乘积。注意，由于副本重叠被约束为相应的 $\boldsymbol{Q}$ 值，无法利用 Fubini定理将该积分简单地化为单副本积分。还要注意，每个副本共享相同的无序 J_{ij}，但可以具有不同的逆温度 β_k 和外场 h^k。

我们为该模型推导出的TAP自由能为

$$F_{\mathrm{TAP}}(\boldsymbol{m}) = \frac{N}{2}\log|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal| + \sum_{k=1}^n \beta_k H_N(m^k) + N\sum_{k=1}^n h^k\cdot m^k + \frac{N}{2}\beta^\intercal(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal)^{\odot 2}\beta, \tag{1.4}$$

其中 $\boldsymbol{m} = (m^1, \ldots, m^n) \in \mathbb{R}^{n\times N}$为磁化向量，$|\cdot|$表示行列式，而$\boldsymbol{A}^{\odot 2} = \boldsymbol{A}\odot\boldsymbol{A} = (A_{k,l}^2)_{k,l=1,\ldots,n}$表示 $\boldsymbol{A}$各元素的Hadamard平方。我们进一步为向量自旋模型引入Plefka条件 [TAP77; Ple82]，其定义为 $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta)$，其中

$$\begin{aligned} \mathrm{Plef}_n(\boldsymbol{Q}, \beta) &= \left\{\tilde{\boldsymbol{Q}} \in [-1,1]^{n\times n} : \boldsymbol{0} \le \tilde{\boldsymbol{Q}} < \boldsymbol{Q},\ \|\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q} - \tilde{\boldsymbol{Q}})\boldsymbol{\beta}^{\frac{1}{2}}\|_2 \le \tfrac{1}{\sqrt{2}}\right\}, \\ \mathrm{Plef}_N(\boldsymbol{Q}, \beta) &= \left\{\boldsymbol{m} \in \mathbb{R}^{n\times N} : \boldsymbol{m}\boldsymbol{m}^\intercal \in \mathrm{Plef}_n(\boldsymbol{Q}, \beta)\right\}, \end{aligned} \tag{1.5}$$

其中 $\boldsymbol{\beta} = \mathrm{diag}(\beta) \in \mathbb{R}^{n\times n}$, $\|\cdot\|_2$ 表示谱范数（对称半正定矩阵的最大特征值），而 $\le$ 表示矩阵上的Loewner偏序（即对于 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$，若满足 $\boldsymbol{A} \ge 0$，则 $\boldsymbol{A}$ 为半正定矩阵）。

我们的主要定理是一个TAP变分原理，它将模型的极限自由能表示为在 $\boldsymbol{m} \in \mathrm{Plef}_n(\boldsymbol{Q}, \beta)$ 上的上确界。

定理 1.1（TAP变分原理）。设 $n \ge 1$ 和 $\boldsymbol{Q} \in [-1,1]^{n\times n}$ 为正定，且 $Q_{k,k} = 1$ 对 $k = 1, \ldots, n$成立。则有

$$\lim_{\varepsilon\to 0}\limsup_{N\to\infty}|F_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q}) - \sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)}\frac{1}{N}F_{\mathrm{TAP}}(\boldsymbol{m})| = 0, \tag{1.6}$$

其中极限依概率收敛。

If $\boldsymbol{Q}$ 不是正定的，则 $\lim_{\varepsilon\to 0}\lim_{N\to\infty} F_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q}) = -\infty$ (见 (3.12))

我们还计算了（1.6）中的上确界，此时 h^1，.....，h^n 是同一个向量的倍数。为此，我们令任意 β、$h \in \mathbb{R}^n$ 和正定 $n\times n$ 约束矩阵 $\boldsymbol{Q}$

$$\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) = \sqrt{2}\mathrm{Tr}\left(\sqrt{\left(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\left(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}}\right). \tag{1.7}$$

注意，右侧的迹是 $(\frac{1}{2}hh^\intercal + \beta\tilde{Q})^{1/2}\tilde{Q}^{1/2}$ 的奇异值之和。具有受限重叠的磁化 $\boldsymbol{m}$ 上能量的基态收敛于此极限：

定理 1.2（基态能量）。假设 $h^i = h_i u$ 对于一列单位向量$u \in \mathbb{R}^N$对于 $i = 1, \ldots\ldots, n$。对于所有 β、h 以及正定 $\tilde{\boldsymbol{Q}}$

$$\lim_{N\to\infty} \sup_{\boldsymbol{m}\boldsymbol{m}^\intercal = \tilde{\boldsymbol{Q}}} \left(\sum_{k=1}^{n} \frac{1}{N}\beta_k H_N(m^k) + \sum_{k=1}^{n} m^k \cdot h^k \right) = \mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}), \tag{1.8}$$

其中，极限依概率成立。

下一个定理将极限最大 TAP自由能表示为一个低维优化问题，即表示为 $n \times n$（因此有界于 N）而不是 $n \times N$维空间中的一个优化问题。这直接由（1.4）和定理 1.2得出。

推论1.3（低维变分原理）。假设 $h^i = h_i u$ 对于一列单位向量 $u \in \mathbb{R}^N$对于 $i = 1, \ldots\ldots, n$。对于所有 β、h 以及正定 $\boldsymbol{Q}$，均有1.9

$$\lim_{N\to\infty} \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta)} \frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m}) = \sup_{\tilde{\boldsymbol{Q}}\in \mathrm{Plef}_n(\boldsymbol{Q},\beta)} \left(\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) + \frac{1}{2}\log|\boldsymbol{Q} - \tilde{\boldsymbol{Q}}| + \frac{1}{2}\beta^\intercal (\boldsymbol{Q} - \tilde{\boldsymbol{Q}})^{\odot 2}\beta \right),$$

其中，极限依概率成立。

由定理 1.1 和推论1.3 可立即得出，极限自由能也由相同的低维优化问题给出。

推论1.4. 假设对于一列单位向量 $h^i = h_i u$， $u \in \mathbb{R}^N$ 对于 $i = 1, \ldots\ldots, n$。对于所有 β、h 和正定的 $\boldsymbol{Q}$，自由能的极限为

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) = \sup_{\tilde{\boldsymbol{Q}}\in \mathrm{Plef}_n(\boldsymbol{Q},\beta)} \left(\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) + \frac{1}{2}\log|\boldsymbol{Q} - \tilde{\boldsymbol{Q}}| + \frac{1}{2}\beta^\intercal (\boldsymbol{Q} - \tilde{\boldsymbol{Q}})^{\odot 2}\beta \right), \tag{1.10}$$

其中的极限依概率成立。

注记 1.5. 当 $n = 1$ 我们得到了 [BK19]的结果。事实上，唯一有效的约束是 $\boldsymbol{Q} = 1$，在此约束下 F_{TAP} 与 H_{TAP} 中的 [BK19], 一致，而定理 1.1与 [BK19中的定理 1]一致。泛函（1.7）是

$$\mathrm{GSE}(\beta, h, \tilde{q}) = \sqrt{2\beta^2\tilde{q}^2 + h^2\tilde{q}}, \tag{1.11}$$

参见 [BK19, （1.6）以及引理 20]。推论 1.3 表明

$$\lim_{N\to\infty} F_N^{\varepsilon}(\beta, h, 1) = \sup_{m:\beta(1-\tilde{q})\le\frac{1}{\sqrt{2}}} \left(\sqrt{2\beta^2\tilde{q}^2 + h^2\tilde{q}} + \frac{\beta^2}{2}(1-\tilde{q})^2 + \frac{1}{2}\log|1-\tilde{q}| \right),$$

对所有 $\varepsilon > 0$，参见 [BK19, 引理 2]。

1.1. 讨论。 关于单副本（$n = 1$）自旋玻璃模型[SK75][1] 最重要的结果是 Parisi 公式 [Par80; Par79; MPV87] 关于极限自由能，该结果已通过 Guerra、Aizenman–Sims–Starr、Talagrand 和 Panchenko 的方法得到严格证明[Gue03; ASS03; Che13; Tal06b; Tal06a; Pan13; Pan14]。TAP方法是一个

[1]参见 [KTJ76; Der80; GM84; CS92; Tal00; CL04; Tal06a] 关于原始 Ising 型 2 自旋 SK 模型的各种推广。

有吸引力的方案 [TAP77] 一种用于计算自由能的替代框架，目前正处于积极研究之中，至少有三个项目正在开展，以对其进行严格的数学实现（[Bol14； Bol19； BY22], [Sub17； Sub20； CPS22； Sub21], [BK19； Bel22]）。

关于受约束多重自旋玻璃模型（$n \geq 1$; [FPV92；PT07；Pan18a]），[PT07] 利用Guerra插值方案证明了球面模型自由能的一个上界。针对该模型的匹配下界，则由 [Ko20；Ko19]通过改造Panchenko在[Pan18a；Pan18b] 中为受约束多重自旋模型相对于乘积测度推导出的同步性质，并结合Aizenman–Sims–Starr方案予以证明。本文采用[BK19]的TAP方法研究受约束多重球面二自旋模型（另见 [Bel22]），并推导出极限自由能的新变分表达式（1.10）。该变分公式表示为定义在 $n \times n$ 矩阵上的一个泛函的最大值。它比[Ko19] 提出的、以矩阵路径表示的二自旋版Parisi变分公式 [Ko19, Theorem 1 and Theorem 3]简单得多。继[BK19] 之后，我们的结果仅是第二个案例：在该案例中，利用数学上严谨的TAP方法，在所有温度和外场下计算出了自旋玻璃模型的自由能（[Sub21] 采用另一种TAP方法，在所有温度下计算无外场纯 p-自旋球形自旋玻璃的自由能）。我们希望未来能进一步改进本文的TAP方法，并将其推广到更广泛的自旋玻璃模型。

经典（$n = 1$）球面二自旋模型的一个众所周知的性质是，在任意逆温度和外场下它都是副本对称的，这可以通过研究该模型的 Parisi 公式来验证 [Tal06a，第 2]节。对于受约束的多自旋模型， [AZ22]给出了基态的零温 Parisi 公式，并证明在二自旋相互作用的情形下，极小化者是副本对称的 [AZ22，命题 7]。正温下的类似计算似乎不可行，因此目前无法从其 Parisi 公式推断受约束多二自旋模型的自由能是副本对称的。由于我们采用 TAP 方法，并不直接研究该模型的 Parisi 公式，而是得到不同的公式（1.10）。然而，公式（1.10）表明自由能是副本~对称的，因为最大化仅针对一个矩阵 $\boldsymbol{Q}$。

1.2. 证明概要。证明的起点是在没有外场的情况下计算高温时的自由能。当 $n = 1$ （以唯一可能的$\boldsymbol{Q} = 1$ 作为约束）时，退火自由能为 $\frac{1}{2}\beta^2$；如果哈密顿量处于高温状态，这也就是淬火自由能，而当 $\beta \leq \frac{1}{\sqrt{2}}$时正是如此。当 $n \geq 2$且带有约束 $\boldsymbol{Q}$ 时，退火自由能结果为 $\frac{1}{2}\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta$ （减去归一化因子1 2log $|\boldsymbol{Q}|$ 之后，该因子对应于满足约束的自旋向量的对数尺度体积；见引理3.2、3.3）。类似地，如果哈密顿量处于高温状态，这也就是淬火自由能，而事实证明当 $\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{2}}$时正是如此。众所周知，使用二阶矩方法 [Tal03, 第 2 节2]可以验证具有 $n = 1$ 的模型的这些性质。在本文中，我们发现二阶矩计算同样给出了具有 $n \geq 2$的模型的上述性质，尽管二阶矩计算更具挑战性（见引理 3.3、3.4和命题3.6、3.10）。顺便指出，关于淬火自由能的上述断言是 (1.10)在 $h = 0$和 $\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{2}}$情况下的特例，其中可以看出最大化点为 $\tilde{\boldsymbol{Q}} = 0$。

掌握了高温相的相关知识后，定理 1.1 的证明分为关于 $F_N^{\varepsilon} = F_N^{\varepsilon}(\beta$、 $\boldsymbol{h}$、 $\boldsymbol{Q}$的下界和上界，这两部分都通过估计

配分函数积分限制在 $\mathcal{S}_{N-1}^n$ 的某些子集上，这些子集是磁化向量 $\boldsymbol{m}$的邻域。也就是说，对于每个这样的子集 $A(\boldsymbol{m}) \subset \mathcal{S}_{N-1}^n$，我们估计 $\int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} e^{f(\boldsymbol{m})} d\boldsymbol{\sigma}$，其中 $f(\boldsymbol{\sigma}) = \sum_{k=1}^n \left(\beta_k H_N\left(\sigma^k\right) + Nh^k \cdot \sigma^k\right)$。我们对积分进行归一化，减去中心化项 $f(\boldsymbol{m})$，并取对数，得到

$$\text{(I)} \qquad \log \int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} e^{f(\sigma)} d\boldsymbol{\sigma} =$$

$$\overbrace{\underbrace{\log \int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma}}_{\text{对数}} + f(\boldsymbol{m}) + \log \frac{\int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} e^{f(\boldsymbol{\sigma})-f(\boldsymbol{m})}\, d\boldsymbol{\sigma}}{\int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma}}}$$

这些定 ms 各自通过该近似产生 $F_{\mathrm{TAP}}(\boldsymbol{m})$ 中的一项 s

$$\overbrace{\underbrace{\frac{N}{2}\log\left|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^T\right|}_{\wr} + \underbrace{\sum_{k=1}^n \left(\beta_k H_N\left(m^k\right) + Nh^k \cdot m^k\right)}_{\|} + \underbrace{\frac{N}{2}\beta^{\mathsf{T}}\left(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^T\right)^{\odot 2}\beta}_{\wr}}$$

$$\text{(II)} \qquad = F_{\mathrm{TAP}}(\boldsymbol{m})$$

此外，每一项都有其自然的解释

$$\text{Entropy} \quad + \quad \text{Local mean energy} \quad + \quad \begin{matrix}\text{Local free energy}\\ \text{(Onsager term)}\end{matrix}$$

如下所述。

事实上，(I) 中的第一项恰好是满足 $\boldsymbol{Q}$ 约束、且位于 $A(\boldsymbol{m})$ 中的 $\boldsymbol{\sigma}$ 的对数体积，因此它是一个熵。邻域 $A(\boldsymbol{m})$ 的选择，本质上是取经过 $\boldsymbol{m}$ 的“切片”中的一个子集；该切片是指法向量为 $\boldsymbol{m}$ 、经过 $\boldsymbol{m}$ 的超平面与 $\mathcal{S}_{N-1}^n$ 的交集。事实证明，这样的切片的对数体积近似为$\frac{N}{2}\log\left|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\right|$（即 (II) 中的第一项），而我们所选的子集保留了切片足够多的体积，因此其对数体积近似相同。

(I)、(II) 中的中心化项 $f(\boldsymbol{m}) = \sum_{k=1}^n \left(\beta H_N\left(m^k\right) + Nh^k \cdot m^k\right)$ 表示 $A(\boldsymbol{m})$ 上的“局部”平均能量。

对于 (I) 的最后一项，我们使用本小节第一段中关于高温相的知识。恒等式

$$H_N(\sigma) = H_N(m) + \nabla H_N(m)\cdot(\sigma - m) + H_N(\sigma - m)$$

对所有 m、$\sigma \in \mathbb{R}^N$ 均成立意味着

$$f(\boldsymbol{\sigma}) - f(\boldsymbol{m}) = \sum_{k=1}^N \underbrace{\left(\beta_k \nabla H_N(m^k) + h^k\right)}_{\text{effective external fields}} \cdot (\sigma^k - m^k) + \sum_{k=1}^N \underbrace{\beta_k H_N(\sigma^k - m^k)}_{\text{effective Hamiltonian}}.$$

由此可见，(I) 中的最后一项可解释为自旋构型空间 $A(\boldsymbol{m}) \cap \boldsymbol{Q}_\varepsilon$ 上、受有效外场作用的有效哈密顿量的自由能。在证明中，我们构造集合 $A(\boldsymbol{m})$，使得对于 $\sigma \in A(\boldsymbol{m})$，有效外场项消失（最简单的情形是，将切片与法向量由有效外场给出的超平面相交）。此外，在对 $\sigma^k - m^k$ 归一化后，它

会发现，重新中心化哈密顿量实质上是原哈密顿量加上一个有效约束 $\hat{}\, Q(\boldsymbol{m})^{ij} = (\boldsymbol{Q} - (\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}))_{ij}/(\sqrt{1-|m^i|^2}\sqrt{1-|m^j|^2})$，并处于有效温度 $\boldsymbol{\beta}_{\boldsymbol{m}} = (\beta_1(1-|m^1|^2)\ldots, \beta_n(1-|m^n|^2))$ 下。因此，应用本小节第一段中关于高温自由能的近似可得：若 $\|\boldsymbol{\beta}_{\boldsymbol{m}}\, \hat{}\, \boldsymbol{Q}(\boldsymbol{m})\boldsymbol{\beta}_{\boldsymbol{m}}\|_2 \le \frac{1}{\sqrt{2}}$，则 (I) 的第三项可近似为 $\frac{N}{2}\beta_{\boldsymbol{m}}^{\mathsf{T}}\, \hat{}\, \boldsymbol{Q}(\boldsymbol{m})^{\odot 2}\beta_{\boldsymbol{m}}$。由于 $\|\boldsymbol{\beta}_{\boldsymbol{m}}\, \hat{}\, \boldsymbol{Q}(\boldsymbol{m})\boldsymbol{\beta}_{\boldsymbol{m}}\|_2 = \|\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\tilde{\boldsymbol{Q}})\boldsymbol{\beta}^{\frac{1}{2}}\|_2$，前一个条件恰好就是Plefka条件；又由于 $\frac{N}{2}\beta_{\boldsymbol{m}}^{\mathsf{T}}\, \hat{}\, \boldsymbol{Q}(\boldsymbol{m})^{\odot 2}\beta_{\boldsymbol{m}} = \frac{N}{2}\beta^{\mathsf{T}}\left(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{T}\right)^{\odot 2}\beta$，后一个条件恰好对应于将 (I) 的最后一项近似为 (II) 中的Onsager项。

只要 $\boldsymbol{m}$满足Plefka条件，这就证明了对数$\int_{A(\boldsymbol{m})\cap Q_\varepsilon} e^{f(\boldsymbol{\sigma})}d\boldsymbol{\sigma} \approx F_{\mathrm{TAP}}(\boldsymbol{m})$的近似是合理的。

最后，事实证明，只有满足Plefka条件的 $\boldsymbol{m}$ 才与问题相关。事实上，在第3 节中，我们只考虑满足Plefka条件的 $\boldsymbol{m}$ ，从而证明了 F_N^ε 的下界，并由此推出 F_N^ε 的下界为 $F_{\mathrm{TAP}}(\boldsymbol{m})$，对任何 $\boldsymbol{m}$ 满足该条件的情形均成立。

第 F_N^ε 节中证明上界的核心难点4 在于，我们无法事先忽略不满足Plefka条件的 $\boldsymbol{m}$。因此，我们用一个在某种意义上始终处于高温状态的哈密顿量来近似重新中心化哈密顿量，即使Plefka条件不成立时也是如此。由此，我们得到 F_N^ε 的上界，该上界由一个修正TAP自由能给出，其Onsager项有所不同。随后我们证明，这个修正TAP自由能的任何最大化点实际上都必须满足Plefka条件；在这种情况下，其Onsager项与通常的Onsager项十分接近。

上述内容构成了一个多维（$n \ge 2$）方法的改编（针对 $n = 1$），见[BK19]（上述思想中的一些元素也被 TAP 使用，参见 [TAP77], Bolthausen [Bol14；Bol19]以及 Subag 及其合作者 [Sub17；Sub20；CPS22；Sub21]）。

最后，在第 5节中，我们使用拉格朗日乘子法，将哈密顿量的基态表示为一个关于半正定矩阵的有限维变分问题。所得变分问题可以显式求解，从而得到定理1.2中的闭式表示。

2. Preliminaries

我们用 c 表示常数，其值可能在不同的行之间甚至在同一表达式中发生变化。除非另有说明，它们可能依赖于副本数 n，但与所有其他参数无关。

在证明的某些步骤中，我们将使用以下标准事实：

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{m\in\mathbb{R}^N:|m|\le 1} |H_N(m)| \le cN, \sup_{m\in\mathbb{R}^N:|m|\le 1} |\nabla H_N(m)| \le cN\right) = 1. \tag{2.1}$$

例如，可以写成 $\boldsymbol{J} = (J_{i,\,j})_{i,\,j=1,\,\ldots\ldots,\,N}$，使得 $\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}$ 是一个 GOE 随机矩阵，且 $H_N(m) = m^{\mathsf{T}}\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}m$，并注意到 $\nabla H_N(m) = (\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}})m$ 且

$$\lim_{N\to\infty} \mathbb{P}\left(\|\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}\|_2 \le cN\right) = 1. \tag{2.2}$$

我们还将用到以下事实：将 $N\lambda_1 < \ldots < N\lambda_N$记为 $\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}$的特征值，我们有

$$\max_{i=1,\ldots,N-1} |\lambda_{i+1}-\lambda_i| \xrightarrow{\mathbb{P}} 0 \text{ as } N\to\infty. \tag{2.3}$$

最后，对于上界，我们将用到

$$\max_{i=1,\dots,N} |\lambda_i - \theta_{i/N}| \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty, \tag{2.4}$$

（见 [EYY12，定理 2。2]），其中 $\theta_{i/N}$ 是半圆分布定义的经典位置

$$\theta_{i/N} = \inf\left\{\theta : \int_{-\sqrt{2}}^{\theta} d\mu_{\mathrm{sc}}(x) = \frac{i}{N}\right\}, \tag{2.5}$$

根据半圆分布定义

$$d\mu_{\mathrm{sc}}(x) = \frac{1}{\pi}\sqrt{2-x^2}\,\mathbb{I}_{[-\sqrt{2},\sqrt{2}]}(x)dx. \tag{2.6}$$

由（2.5）可得

$$\lim_{N\to\infty}\lim_{\varepsilon\to 0}\sup_{|i-j|\le \varepsilon N}\left|\theta_{i/N} - \theta_{j/N}\right| = 0. \tag{2.7}$$

注意，（2.2）和（2.3）是（2.4）—（2.7）的推论。

3. Lower bound

本节将证明自由能的以下下界。

命题3.1（TAP下界）。设 $n \ge 1$ 和 $\boldsymbol{Q} \in [-1,1]^{n\times n}$ 为正定，且$Q_{k,\,k} = 1$对 $k = 1$、……、n成立。设 h_1、……、$h_n \in [0,\infty)$）和 h^1、……、h^n）为一列向量，满足 $h^k \in \mathbb{R}^N$和 $|h^k| = h_k$。那么存在一个 $c = c$（n、h_1、……、h_n、$\boldsymbol{Q}$）> 0，使得对所有$\varepsilon > \in(0, c^{-1})$

$$\lim_{N\to\infty}\mathbb{P}\left(F_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q}) \ge \frac{1}{N}\sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta)} F_{\mathrm{TAP}}(\boldsymbol{m}) - c\sqrt{\varepsilon}\right) = 1. \tag{3.1}$$

为证明这一点，我们首先在3.1小节中使用二阶矩方法计算无外场时的高温自由能。随后，在3.2小节中，我们考虑任意温度下带外场的模型，并按照1.2小节中的描述，先固定一个满足 Plefka条件的 $\boldsymbol{m}$，构造一个集合 $A(\boldsymbol{m})$（见（3.51））， 该集合“以 $\boldsymbol{m}$为中心”；再将哈密顿量重新以这个 $\boldsymbol{m}$ 为中心（见（3.48）），并估计该重新中心化的哈密顿量在集合 $A(\boldsymbol{m})$ 上的自由能（见（3.67）。

3.1. 无外场时的自由能。定义

$$Z_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q}) = \int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + N h^k \sigma^k}\, d\boldsymbol{\sigma} \overset{(1.3)}{=} e^{N F_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q})}. \tag{3.2}$$

本小节的目标是对 $Z_N^{\varepsilon}(\beta, 0, \boldsymbol{Q})$ 使用二阶矩方法，以证明它在 $N \to \infty$ 时集中。本节的第一个引理将证明，$\boldsymbol{Q}$-约束的 n 重球面的体积在指数尺度上约为 $\frac{1}{2}\log|\boldsymbol{Q}|$，据此我们可以计算 $Z_N^{\varepsilon}(\beta, 0, \boldsymbol{Q})$ 的矩。

回想一下，$\|\boldsymbol{A}\|_2$ 表示 $\boldsymbol{A}$ 的谱范数。

引理3.2（受限体积）。设 $n \ge 1$。存在常数 $c > 0$，使得对于所有对角线上元素均为1的对称半正定 $\boldsymbol{Q} \in [-1,1]^{n\times n}$，以及所有 $\varepsilon \in (0, c^{-1})$ 和 $N \ge c\varepsilon^{-1}$，我们有

$$\left|\frac{1}{N}\log\int \mathbb{I}_{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}} - \boldsymbol{Q}\|_\infty \le \varepsilon\}}\, d\boldsymbol{\sigma} - \frac{1}{2}\log|\boldsymbol{Q}|\right| \le c(1 + \|\boldsymbol{Q}^{-1}\|_2)\varepsilon, \tag{3.3}$$

and

$$\frac{1}{N}\log\int_{\mathbb{I}^{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^{\intercal}-\boldsymbol{Q}\|_{\infty}\leq\varepsilon\}}}d\boldsymbol{\sigma}\leq\frac{1}{2}\log|\varepsilon\boldsymbol{I}+\boldsymbol{Q}|+c. \tag{3.4}$$

证明。令 $u_{i,k},k=1,\ldots,n,i=1,\ldots,N$为独立同分布的标准正态随机变量，并令$u^{k}=(u_{1,k},\ldots,u_{n,k})\in\mathbb{R}^{N}$ 以及 $u_{i}=(u_{i,1},\ldots,u_{i,n})\in\mathbb{R}^{n}$。注意，在事件 $|\frac{|u^{k}|}{\sqrt{N}}-1|<\varepsilon$上的 $u^{k}/\left|u^{k}\right|$ 的条件分布与 σ_{k}在 $k=1,\ldots,n$时的分布相同，所以

$$\begin{aligned}&\frac{1}{N}\log\int_{\mathbb{I}^{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^{\intercal}-\boldsymbol{Q}\|_{\infty}\leq\varepsilon\}}}d\boldsymbol{\sigma}\\&=\frac{1}{N}\log\mathbb{P}\left(\max_{k,l}\left|\frac{u^{k}\cdot u^{l}}{|u^{k}|\,|u^{l}|}-Q_{k,l}\right|\leq\varepsilon\;\middle|\;\sup_{k}\left|\frac{|u^{k}|}{\sqrt{N}}-1\right|<\varepsilon\right)\\&=\frac{1}{N}\log\mathbb{P}\left(\max_{k,l}\left|\frac{u^{k}\cdot u^{l}}{|u^{k}|\,|u^{l}|}-Q_{k,l}\right|\leq\varepsilon,\sup_{k}\left|\frac{|u^{k}|}{\sqrt{N}}-1\right|<\varepsilon\right)-\frac{1}{N}\log\mathbb{P}\left(\sup_{k}\left|\frac{|u^{k}|}{\sqrt{N}}-1\right|<\varepsilon\right).\end{aligned}$$

根据切比雪夫不等式 $\mathbb{P}\left(\left|\frac{|u^{k}|}{\sqrt{N}}-1\right|>\varepsilon\right)\leq\frac{c}{\varepsilon^{2}N^{2}}$，这意味着只要 c 足够大且 $N\geq c\varepsilon^{-1}$，最后一项就可由 $c\varepsilon$ 控制，因此只需控制该事件的概率

$$A=\left\{\max_{k,l}\left|\frac{u^{k}\cdot u^{l}}{|u^{k}|\,|u^{l}|}-Q_{k,l}\right|\leq\varepsilon,\sup_{k}\left|\frac{|u^{k}|}{\sqrt{N}}-1\right|<\varepsilon\right\}.$$

我们将Cramér定理的标准证明应用于独立同分布的向量 $u_{1},\ldots,u_{N}$，同时注意获得关于 $\boldsymbol{Q}$均匀分布的界。存在常数 c_{1},c_{2}，使得对于所有小于某个常数的 ε 以及所有满足 $\|\boldsymbol{Q}\|_{\infty}\leq1$的 $\boldsymbol{Q}$

$$\left\{\max_{k,l}\left|u^{k}\cdot u^{l}-NQ_{k,l}\right|\leq c_{1}\varepsilon N\right\}\subseteq A\subseteq\left\{\max_{k,l}\left|u^{k}\cdot u^{l}-NQ_{k,l}\right|\leq c_{2}\varepsilon N\right\}.$$

我们从上界开始。对于所有对称的 $n\times n$ 矩阵 $\boldsymbol{\Lambda}$

$$\mathbb{P}\left(\max_{k,l}\left|u^{k}\cdot u^{l}-NQ_{k,l}\right|\leq c_{2}\varepsilon N\right)\leq\mathbb{E}\left[\exp\left(\sum_{k,l=1}^{n}\Lambda_{k,l}u^{k}\cdot u^{l}\right)\right]e^{-N\sum_{k,l=1}^{n}\Lambda_{k,l}Q_{k,l}+c_{2}\varepsilon n^{2}\|\boldsymbol{\Lambda}\|_{\infty}N}.$$

如果 $2\boldsymbol{\Lambda}<\boldsymbol{I}$，则成立

$$\mathbb{E}\left[\exp\left(\sum_{k,l=1}^{n}\Lambda_{k,l}u^{k}\cdot u^{l}\right)\right]=\mathbb{E}\left[\exp\left(\sum_{i=1}^{N}(u_{i})^{\intercal}\boldsymbol{\Lambda}u_{i}\right)\right]=\left(\mathbb{E}\left[\exp\left((u_{1})^{\intercal}\boldsymbol{\Lambda}u_{1}\right)\right]\right)^{N}=|\boldsymbol{I}-2\boldsymbol{\Lambda}|^{-\frac{N}{2}}.$$

因此，对于所有这样的 $\boldsymbol{\Lambda}$

$$\mathbb{P}\left(\max_{k,l}\left|u^{k}\cdot u^{l}-NQ_{k,l}\right|\leq c_{2}\varepsilon N\right)\leq\exp\left(-\frac{N}{2}\log|\boldsymbol{I}-2\boldsymbol{\Lambda}|-N\sum_{kl}\Lambda_{k,l}Q_{k,l}+c_{2}\varepsilon n^{2}\|\boldsymbol{\Lambda}\|_{\infty}N\right). \tag{3.5}$$

右侧的非误差项通过选择 $\boldsymbol{\Lambda}=\frac{\boldsymbol{I}-\boldsymbol{Q}^{-1}}{2}$ 取最小值，此时 $-\frac{1}{2}\log\ |\boldsymbol{I}-2\boldsymbol{\Lambda}|=_{1\ 2}\log|\boldsymbol{Q}|$ 并且

–

$$\sum_{kl}\Lambda_{k,l}Q_{k,l}=\mathrm{Tr}\left(\boldsymbol{\Lambda}\boldsymbol{Q}\right)=\mathrm{Tr}\left(\frac{\boldsymbol{Q}-\boldsymbol{I}}{2}\right)=0. \tag{3.6}$$

因此，我们有♀♀♀♀♀♀

$$\mathbb{P}\left(\max_{k,l}\left|u^{k}\cdot u^{l}-NQ_{k,l}\right|\leq c_{2}\varepsilon N\right)\leq\exp\left(\frac{N}{2}\log|\boldsymbol{Q}|+c(1+\|\boldsymbol{Q}^{-1}\|_{2})\varepsilon N\right),$$

由于 $c_2 n^2 \|\mathbf{\Lambda}\|_\infty \le c(1+\|\boldsymbol{Q}^{-1}\|_2)$，对足够大的 c（仅取决于 n）。这证明了 (3.3) 的上界。

为得到(3.4)，令 $\mathbf{\Lambda} = \underline{\boldsymbol{I} - (\boldsymbol{Q}+\varepsilon \boldsymbol{I})_2}^{-1}$，并注意到此时 $\sum_{kl} \Lambda_{k,l} Q_{k,l} = \mathrm{Tr}\,(\mathbf{\Lambda}\boldsymbol{Q}) = \mathrm{Tr}\left(\frac{\boldsymbol{Q}-(\boldsymbol{Q}+\varepsilon \boldsymbol{I})^{-1}\boldsymbol{Q}}{2}\right) \ge -\frac{n}{2}$ 以及 $\|\mathbf{\Lambda}\|_\infty \le c\varepsilon^{-1}$。

对于（的下界3.3）我们使用测度变换

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{i=1}^N \frac{\exp\left(\sum_{kl} \Lambda_{k,l} u^k \cdot u^l\right)}{|\boldsymbol{I} - 2\mathbf{\Lambda}|^{-1/2}} = \prod_{i=1}^N \frac{\exp\left(\sum_i (u_i)^\top \mathbf{\Lambda} u_i\right)}{|\boldsymbol{I} - 2\mathbf{\Lambda}|^{-1/2}}$$

对于 $\mathbf{\Lambda} = \frac{\boldsymbol{I} - \boldsymbol{Q}^{-1}}{2}$。在测度 $\mathbb{Q}$ 下，u_i 是 $\mathbb{R}^n$ 中均值为零的独立同分布高斯向量，其协方差为 $\boldsymbol{Q}$。我们有

$$\begin{aligned}
&\mathbb{P}\left(\left|u^k \cdot u^l - N Q_{k,l}\right| \le c_1 \varepsilon N\right) \\
&= \mathbb{Q}\left(\mathbb{I}_{\{\left|u^k \cdot u^l - N Q_{k,l}\right| \le c_1 \varepsilon N\}} \frac{d\mathbb{P}}{d\mathbb{Q}}\right) \\
&\ge \mathbb{Q}\left(\max_{k,l} \left|u^k \cdot u^l - N Q_{k,l}\right| \le c_1 \varepsilon N\right) \frac{\exp\left(-\sum_{kl} \Lambda_{k,l} Q_{k,l} - c_1 n^2 \varepsilon \|\mathbf{\Lambda}\|_\infty N\right)}{|\boldsymbol{I} - 2\mathbf{\Lambda}|^{N/2}} \\
&\overset{(3.6)}{\ge} \mathbb{Q}\left(\max_{k,l} \left|\tfrac{1}{N}\sum_i u_i^k u_i^l - Q_{k,l}\right| \le c_1 \varepsilon\right) \exp\left(\tfrac{N}{2} \log |\boldsymbol{Q}| - c(1+\|\boldsymbol{Q}^{-1}\|_2)\varepsilon N\right).
\end{aligned} \tag{3.7}$$

利用并集界和切比雪夫不等式（回顾$\mathbb{Q}\left(u_i^k u_i^l\right) = Q_{k,l}$）我们得到

$$\begin{aligned}
\mathbb{Q}\left(\left\{\max_{k,l} \left|\tfrac{1}{N}\sum_i u_i^\mathsf{T} u_i - Q_{k,l}\right| \le c_1\varepsilon\right\}^c\right) &\le \sum_{k,l} \mathbb{Q}\left(\left|\tfrac{1}{N}\sum_i u_i^k u_i^l - Q_{k,l}\right| \ge c_1\varepsilon\right) \\
&\le \sum_{k,l} \frac{\mathrm{Var}_\mathbb{Q}\left(\frac{1}{N}\sum_i u_i^k u_i^l\right)}{c_1 2\varepsilon^2} \\
&= \sum_{k,l} \frac{\mathrm{Var}_\mathbb{Q}\left(u_1^k u_1^l\right)}{c_1^2 \varepsilon^2 N}.
\end{aligned} \tag{3.8}$$

现在进行粗略估计

$$\mathrm{Var}_\mathbb{Q}\left(u_1^k u_1^l\right) \le \mathbb{Q}\left(\left(u_1^k u_1^l\right)^2\right) \le \sqrt{\mathbb{Q}\left(\left(u_1^k\right)^4\right)}\sqrt{\mathbb{Q}\left(\left(u_1^l\right)^4\right)} \le c,$$

其中最后一个常数与 $\boldsymbol{Q}$ 无关，因为 u_1^k 是方差为 $Q_{kk} = 1$ 下的高斯变量$\mathbb{Q}$，对所有 k。因此，只要 ε小于某个仅依赖于 n，则（3.8）的右端至多为$_{1\ 2}$，所以由（3.7）可知

$$\mathbb{P}\left(\left|u^k \cdot u^l - N Q_{k,l}\right| \le c_1 \varepsilon N\right) \ge \exp\left(\frac{N}{2}\log|\boldsymbol{Q}| - c(1+\|\boldsymbol{Q}^{-1}\|_2)\varepsilon N - c\right),$$

给出（3.3）的下界。 □

现在我们计算一阶矩，或等价地计算退火自由能。回顾一下，$\boldsymbol{A} \ge \delta \boldsymbol{I}$ 表示 $\boldsymbol{A}$ 的所有特征值都大于 δ。我们还用记号 $\boldsymbol{A}^{\odot 2} = \boldsymbol{A} \odot \boldsymbol{A} = (A_{k,l}^2)_{k,l=1,\dots,n}$ 表示 $\boldsymbol{A}$各元素的Hadamard平方。

引理 3.3（第一矩；无外场时的退火自由能）。令 $n \ge 1$。对于所有 $\delta \in (0,1)$ $C > 0$，存在常数 $c = c(\delta, C) > 0$，使得对于所有小于一个普适常数的 ε、$|\beta| \le C$ 以及 $N \ge c(\delta, \varepsilon)$，我们有

$$\sup_{\boldsymbol{Q} \ge \delta \boldsymbol{I}} \left| \frac{1}{N} \log \mathbb{E}[Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})] - \left(\frac{1}{2}\beta^\mathsf{T} \boldsymbol{Q}^{\odot 2} \beta + \frac{1}{2}\log|\boldsymbol{Q}|\right)\right| \le c\varepsilon, \tag{3.9}$$

其中上确界取遍所有对角线上为 1 的对称 $\boldsymbol{Q} \in [-1,1]^{n\times n}$。

证明。我们有

$$\mathbb{E}[Z_N^{\varepsilon}(\beta,0,\boldsymbol{Q})] \overset{(3.2)}{=} \mathbb{E}\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} d\boldsymbol{\sigma} = \int_{\boldsymbol{Q}_\varepsilon} \mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] d\boldsymbol{\sigma}. \tag{3.10}$$

由于对于固定的 $\boldsymbol{\sigma}$，哈密顿量是高斯变量之和，因此对于 $\boldsymbol{\sigma} \in \boldsymbol{Q}_\varepsilon$，我们有

$$\mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] = \exp\left(\frac{1}{2}\operatorname{Var}\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right). \tag{3.11}$$

由于 $\mathbb{E}[H_N(\sigma)H_N(\sigma')] = N(\sigma\cdot\sigma')^2$，我们有

$$\operatorname{Var}\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right) = \sum_{k,\ell=1}^n \beta_k\beta_\ell(\sigma_k\cdot\sigma_l)^2. \tag{3.12}$$

对于 $\boldsymbol{\sigma} \in Q_\varepsilon$ 我们有 $|\sum_{k,\ell=1}^n \beta_k\beta_\ell(\sigma_k\cdot\sigma_l)^2 - \sum_{k,\ell=1}^n \beta_k\beta_\ell Q_{k,l}^2| \le c\varepsilon N$ 对于一个仅依赖于 C 和 n的常数 c，以及 $\sum_{k,\ell=1}^n \beta_k\beta_\ell Q_{k,l}^2 = \beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta$因此对于所有 $\boldsymbol{\sigma} \in \boldsymbol{Q}_\varepsilon$

$$\left|\log\mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] - \frac{1}{2}\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta\right| \le c\varepsilon N.$$

引理 3.2 意味着对于 $\boldsymbol{Q} \ge \delta\boldsymbol{I}$

$$\left|\log\int_{\boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma} - \frac{N}{2}\log|\boldsymbol{Q}|\right| \le c\varepsilon N,$$

这就完成了证明。 □

接下来我们计算二阶矩。

引理 3.4（二阶矩）。设 $n \ge 1$。对于所有 $\delta, C > 0$，存在常数 $c = c(\delta, C) > 0$，使得对于所有 $\varepsilon \in (0, c^{-1})$ $|\beta| \le C$ 以及所有 $N \ge c(\varepsilon,\delta)$，均有

$$\sup_{\boldsymbol{Q}\ge\delta\boldsymbol{I}}\left(\frac{1}{N}\log\mathbb{E}[Z_N^{\varepsilon}(\beta,0,\boldsymbol{Q})^2] - \left(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta + \sup_{\boldsymbol{A}} V(\boldsymbol{A})\right)\right) \le c\varepsilon, \tag{3.13}$$

其中，上确界取遍所有对角线元素为 1 的对称 $\boldsymbol{Q} \in [-1,1]^{n\times n}$，并且

$$V(\boldsymbol{A}) = \beta^{\mathsf{T}}\boldsymbol{A}^{\odot 2}\beta + \frac{1}{2}\log\begin{vmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{vmatrix}. \tag{3.14}$$

证明。我们有

$$\begin{aligned}\mathbb{E}[Z_N^{\varepsilon}(\beta,0,\boldsymbol{Q})^2] &= \mathbb{E}\left[\int_{\boldsymbol{Q}_\varepsilon}\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} e^{\sum_{k=1}^n \beta_\ell H_N(\tau^\ell)} d\boldsymbol{\sigma} d\boldsymbol{\tau}\right]\\ &= \int_{\boldsymbol{Q}_\varepsilon}\int_{\boldsymbol{Q}_\varepsilon} \mathbb{E}\left[e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + \sum_{\ell=1}^n \beta_\ell H_N(\tau^\ell)}\right] d\boldsymbol{\sigma} d\boldsymbol{\tau}.\end{aligned}$$

与前一个引理的证明类似，内层期望是一个满足以下条件的高斯指数矩：

$$\left|\log\mathbb{E}\left[e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + \sum_{\ell=1}^n \beta_\ell H_N(\tau^\ell)}\right] - N\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta - N\sum_{k,\ell=1}^n \beta_k\beta_\ell\left(\sigma^k\cdot\tau^\ell\right)^2\right| \le Nc\varepsilon,$$

对于所有 σ、$\tau \in \boldsymbol{Q}_\varepsilon$，其中 c 仅取决于 C、n。

还需证明

$$\int_{\boldsymbol{Q}_\varepsilon}\int_{\boldsymbol{Q}_\varepsilon} e^{N\sum_{k,\ell=1}^n \beta_k\beta_\ell\left(\sigma^k\cdot\tau^\ell\right)^2} d\boldsymbol{\sigma}d\boldsymbol{\tau} \le \exp\left(N\sup_{\boldsymbol{A}} V(\boldsymbol{A}) + Nc\varepsilon\right). \tag{3.15}$$

将空间 $[-1,1]^{n\times n}$ 划分为至多 $\lceil\frac{1}{2\varepsilon}\rceil^{n^2}$ 个直径为 ε 量级的子集，可得对所有 $\varepsilon>0$

$$\begin{aligned}&\int_{\boldsymbol{Q}_\varepsilon}\int_{\boldsymbol{Q}_\varepsilon} e^{N\sum_{k,\ell=1}^n \beta_k\beta_\ell\left(\sigma^k\cdot\tau^\ell\right)^2} d\boldsymbol{\sigma}d\boldsymbol{\tau}\\ &\le \exp\left(N\sup_{A\in[-1,1]^{n\times n}}\left\{\beta^{\mathsf{T}}\boldsymbol{A}^{\odot 2}\beta + \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon}\int_{\boldsymbol{Q}_\varepsilon} \mathbb{I}_{\{\|\boldsymbol{\sigma}\boldsymbol{\tau}^{\mathsf{T}}-\boldsymbol{A}\|_\infty\le\varepsilon\}}\, d\boldsymbol{\sigma}d\boldsymbol{\tau}\right\} + Nc\varepsilon\right),\end{aligned} \tag{3.16}$$

存在一个依赖于 C 的 c，使得对所有 $N\ge c(\varepsilon)$。注意，上确界中的第二项等于

$$\frac{1}{N}\log\int \mathbb{I}_{\left\{\left\|\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}-\begin{pmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{pmatrix}\right\|_\infty\le\varepsilon\right\}} d\boldsymbol{\nu},$$

其中积分是在 $\boldsymbol{\nu}\in\mathcal{S}_{N-1}^{2n}$ 上进行的。

注意，对于任意满足 $\|\boldsymbol{B}\|_\infty\le 1$ 的矩阵 $\boldsymbol{B}\in\mathbb{R}^{n\times n}$，我们有

$$\|\boldsymbol{B}\|_2\le n \text{ and } \delta^n\le|\boldsymbol{B}|\le\delta n^{n-1} \text{ if } \delta\ge 0 \text{ is } \boldsymbol{B}\text{'s smallest eigenvalue.} \tag{3.17}$$

固定一个仅依赖于 δ、n、C 且足够小的 $\tilde{\delta}\in\left(0,\frac{1}{2}\right)$，使得

$$\frac{1}{4}\log(2\tilde{\delta}) + \frac{1}{4}\log\left(4n\right)^{2n-1} + n^2\max\beta_i^2 \le n\log\delta \overset{(3.17)}{\le} V(0).$$

那么，如果 $\boldsymbol{A}$ 满足 $\begin{pmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{pmatrix}$ 的某个特征值小于或等于 $\tilde{\delta}$ 且 $\varepsilon\in(0,\tilde{\delta})$，则由（3.4）和（3.17）可得

$$\frac{1}{N}\log\int \mathbb{I}_{\left\{\left\|\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}-\begin{pmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{pmatrix}\right\|_\infty\le\varepsilon\right\}} d\boldsymbol{\nu} \le V(0) - N\beta^{\mathsf{T}}\boldsymbol{A}^{\odot 2}\beta. \tag{3.18}$$

$\tilde{\delta}$（必要时进一步减小 δ，具体取决于（3.4）中的常数）。另一方面，如果 $\begin{pmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{pmatrix} > \delta'\boldsymbol{I}$，那么（3.3）意味着

$$\frac{1}{N}\log\int \mathbb{I}_{\left\{\left\|\boldsymbol{\nu}\boldsymbol{\nu}^{\mathsf{T}}-\begin{pmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{pmatrix}\right\|_\infty\le\varepsilon\right\}} d\boldsymbol{\nu} \le \frac{1}{2}\log\begin{vmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{vmatrix} + c\tilde{\delta}^{-1}\varepsilon. \tag{3.19}$$

界（3.16）、（3.18）和（3.19）推出（3.15）。

□

我们将证明，对于落在以下范围内的 β，$V(\boldsymbol{A})$ 在零处取得最大值

$$\mathrm{HT}(\boldsymbol{Q}) := \left\{\beta\in\mathbb{R}^n : \|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2\le\frac{1}{\sqrt{2}}\right\}, \tag{3.20}$$

这是模型的高温区域（回忆一下，$\boldsymbol{\beta}=$对角矩阵$(\beta_1,\ .....,\ \beta_n)$）。结合引理 3.4即可推出

$$\frac{1}{N}\log\mathbb{E}[Z_N^\varepsilon(\beta,0,\boldsymbol{Q})^2] - \frac{1}{N}\log\mathbb{E}[Z_N^\varepsilon(\beta,0,\boldsymbol{Q})]^2 \le c\varepsilon$$

对于 $\beta\in\mathbf{HT}(\boldsymbol{Q})$，借助它我们可以使用二阶矩方法证明此类 β 的 $Z_N^\varepsilon(\beta,0,\boldsymbol{Q})$ 集中。

在证明 $V(\boldsymbol{A})$在零点处取得最大值的计算中，自然出现了高温条件的另一种形式。下一个引理表明，这种形式等价于（3.20）中的条件。

引理3.5（高温条件两种形式的等价性）。对于任意正定对称矩阵 $\boldsymbol{Q}$,

$$\mathrm{HT}(\boldsymbol{Q})=\left\{\beta\in\mathbb{R}^n:\ \sup_{\|\boldsymbol{B}\|_F=1}\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\beta}^{\frac{1}{2}}\|_F\le\frac{1}{\sqrt{2}}\right\},\tag{3.21}$$

其中 $\|\cdot\|_F$ 表示弗罗贝尼乌斯范数。

证明。一旦我们证明了该断言所述内容，结论即得。

$$\sup_{\|\boldsymbol{B}\|_F^2=1}\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}^{1/2}\|_F=\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2.\tag{3.22}$$

为此，注意到

$$\begin{aligned}\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}^{1/2}\|_F^2 &= \mathrm{Tr}\left(\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}^{1/2}\left(\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}^{1/2}\right)^{\mathsf{T}}\right)\\ &= \mathrm{Tr}\left(\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\right)\\ &= \|\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\|_F^2.\end{aligned}$$

令 $\tilde{\boldsymbol{\beta}}$ 为 $\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}$的特征值对角矩阵，并令 $\tilde{\boldsymbol{B}}$ 表示 $\boldsymbol{B}$ 在 $\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}$的（正交）对角化基中。则 $\|\tilde{\boldsymbol{B}}\|_F^2=\|\boldsymbol{B}\|_F^2$和 $\|\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\|_F^2=\|\tilde{\boldsymbol{B}}\tilde{\boldsymbol{\beta}}\|_F^2$，因此

$$\sup_{\|\boldsymbol{B}\|_F^2=1}\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{1/2}\boldsymbol{B}\boldsymbol{Q}^{1/2}\boldsymbol{\beta}^{1/2}\|_F^2=\sup_{\|\tilde{\boldsymbol{B}}\|_F^2=1}\|\tilde{\boldsymbol{B}}\tilde{\boldsymbol{\beta}}\|_F^2.$$

由于 $\|\tilde{\boldsymbol{B}}\tilde{\boldsymbol{\beta}}\|_F^2=\sum_i\left(\sum_j\tilde{B}_{i,j}^2\right)\tilde{\beta}_{i,i}^2$右端显然等于$\max_i\tilde{\beta}_{i,i}^2$。由于 AB 和 BA 对于任意方阵 A, B具有相同的特征值，也

$$\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\ \text{ and }\ \boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}\ \text{ have the same eigenvalues,}\tag{3.23}$$

这就证明了（3.22）。 □

现在我们可以证明，当 $\beta\in\mathrm{HT}(\boldsymbol{Q})$ 时， $V(\boldsymbol{A})$ 在 $A=0$ 处取得最大值 .

命题3.6. 对于任意正定的 $\boldsymbol{Q}$ 和 $\beta\in\mathrm{HT}(\boldsymbol{Q})$，有

$$\sup_{\boldsymbol{A}}V(A)=V(0).\tag{3.24}$$

证明。使用舒尔补公式

$$\begin{vmatrix}\boldsymbol{Q} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathsf{T}} & \boldsymbol{Q}\end{vmatrix}=|\boldsymbol{Q}||\boldsymbol{Q}-\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-1}\boldsymbol{A}|.$$

我们有

$$|\boldsymbol{Q}-\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-1}\boldsymbol{A}|=|\boldsymbol{Q}-\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}|=|\boldsymbol{Q}-(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})^{\mathsf{T}}(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})|.$$

根据矩阵行列式引理，这等于

$$|\boldsymbol{Q}||\boldsymbol{I}-(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})\boldsymbol{Q}^{-1}(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})^{\mathsf{T}}|=|\boldsymbol{Q}||\boldsymbol{I}-(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{Q}^{-\frac{1}{2}})(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-\frac{1}{2}})^{\mathsf{T}}|.\tag{3.25}$$

Thus

$$V(\boldsymbol{A})=\beta^{\mathsf{T}}\boldsymbol{A}^{\odot 2}\beta+\log|\boldsymbol{Q}|+\frac{1}{2}\log|\boldsymbol{I}-(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{Q}^{-\frac{1}{2}})(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-\frac{1}{2}})^{\mathsf{T}}|.$$

现在进行变量替换 $\boldsymbol{B}=\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Q}^{-\frac{1}{2}} \Leftrightarrow \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{\frac{1}{2}}=\boldsymbol{A}$，得到

$$V(\boldsymbol{A})=\beta^{\mathsf{T}}(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta+\log|\boldsymbol{Q}|+\frac{1}{2}\log|\boldsymbol{I}-\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}|. \tag{3.26}$$

因此，只需证明右端在 $\boldsymbol{B}=0$ 处取得最大值。

为此，我们先固定 $\boldsymbol{B}$，沿射线进行优化，并考虑

$$v(t)=V(\sqrt{t}\boldsymbol{B})=t\beta^{\mathsf{T}}(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta+\log|\boldsymbol{Q}|+\frac{1}{2}\log|\boldsymbol{I}-t\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}|, t\geq 0.$$

泛函 $v(t)$ 关于 $[0,\infty)$ 显然是凹的，因为第一项关于 t 是线性的，第二项是常数，最后一项关于 t 是凹的（例如，可通过对角化 $\boldsymbol{B}\boldsymbol{B}^{\mathsf{T}}$ 得出）。因此，要证明 $v(t)$ 在 0 处取得全局最大值，只需证明 $v'(0)\leq 0$。

我们有

$$v'(0)=\beta^{\mathsf{T}}(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta+\frac{1}{2}\frac{d}{dt}\log|\boldsymbol{I}-t\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}|\Big|_{t=0}.$$

由于

$$\frac{d}{dt}\log|\boldsymbol{I}-t\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}|\Big|_{t=0}=-\mathrm{Tr}(\boldsymbol{I}-t\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})\Big|_{t=0}=-\mathrm{Tr}(\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B})=-\|\boldsymbol{B}\|_F^2,$$

和 $w^{\mathsf{T}}\boldsymbol{A}^{\odot 2}w=\|$对角矩阵$(w)^{\frac{1}{2}}\boldsymbol{A}$对角矩阵$(w)^{\frac{1}{2}}\|_F^2$对于任意向量 w和矩阵 $\boldsymbol{A}$，我们得到

$$v'(0)=\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\beta}^{1/2}\|_F^2-\frac{1}{2}\|\boldsymbol{B}\|_F^2=\|\boldsymbol{B}\|_F^2\left(\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{\frac{1}{2}}\hat{\boldsymbol{B}}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\beta}^{1/2}\|_F^2-\frac{1}{2}\right),$$

ˆ其中 $\boldsymbol{B}=\boldsymbol{B}/\|\boldsymbol{B}\|_F$。若 β 满足高温条件（3.21），则右端非负，因此 $v'(0)\leq 0$，且事实上 $v(t)$ $t\in[0,\infty)$ 在 $t=0$处取得最大值。

但由于这对任意 $\boldsymbol{B}$都成立，因此 (3.26) 的右端必在 $\boldsymbol{B}=0$时取得最大值，所以 $\boldsymbol{A}=\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^{\mathsf{T}}\boldsymbol{Q}^{\frac{1}{2}}=0$是泛函 $V(\boldsymbol{A})$ 的全局最大化点。 □

注 3.7。 注意，对于 v 在前面的证明中 $v'(0)\leq 0$ 对所有 $\boldsymbol{B}$ 当且仅当 条件（3.20）得到满足时，因此命题的逆蕴含也成立（尽管我们不需要这一事实）。

利用二阶矩方法证明自由能的下界，需要标准的指数集中不等式。

引理 3.8（自由能的指数集中性）。设 $n\geq 1$ 和 $C>0$。存在一个 $c=c(C)>0$，使得对于所有 $\varepsilon>0$、$|\beta|\leq C$ 和 $\boldsymbol{Q}>0$

$$\mathbb{P}\left(\left|F_N^{\varepsilon}(\beta,0,\boldsymbol{Q})-\mathbb{M}\left[F_N^{\varepsilon}(\beta,0,\boldsymbol{Q})\right]\right|\geq t\right)\leq\exp\left(-ct^2N\right), \tag{3.27}$$

其中 $\mathbb{M}$ 表示中位数。

证明。 这由高斯集中性可得 [BLM13，定理 10。17], 因为对于所有 i， j， β， $\boldsymbol{Q}$

$$\partial_{J_{ij}}F_N^{\varepsilon}\left(\beta,0,\boldsymbol{Q}\right)\overset{(1.3)}{=}\frac{1}{\sqrt{N}}\sum_{k=1}^{n}\beta_k\left\langle\sigma_i^k\sigma_j^k\right\rangle,$$

其中，$\langle\cdot\rangle=\int_{\boldsymbol{Q}_\varepsilon}\cdot\, e^{\sum_{k=1}^n\beta_k H_N(\sigma^k)}d\boldsymbol{\sigma}/\int_{\boldsymbol{Q}_\varepsilon}e^{\sum_{k=1}^n\beta_k H_N\sigma^k(d\boldsymbol{\sigma})}$表示关于吉布斯测度的期望，因此当 $|\beta|\leq C$

$$\left|\nabla_J F_N^{\varepsilon}\left(\beta,0,\boldsymbol{Q}\right)\right|^2\leq\frac{C^2}{N}\sum_{i,j}\sum_{k=1}^{n}\left\langle\sigma_i^k\sigma_j^k\right\rangle^2\leq\frac{C^2}{N}\sum_{i,j}\sum_{k=1}^{n}\left\langle\left(\sigma_i^k\right)^2\left(\sigma_j^k\right)^2\right\rangle=\frac{C^2n}{N},$$

这意味着映射 $J\to F_N^{\varepsilon}(\beta,0,\boldsymbol{Q})$ 是利普希茨连续的，利普希茨常数为 $N^{-1/2}Cn^{1/2}$。 □

为了对 β 和 $\boldsymbol{Q}$ 上的自由能作出均匀估计（见（3.29）），我们将使用以下结果。

引理 3.9（自由能的利普希茨性质）。设 $n \geq 1, C > 0, \varepsilon > 0$。存在一个$L = L(C) > 0$ 使得

$$\lim_{N\to\infty} \mathbb{P}\left(\forall \boldsymbol{Q} > 0 \text{ and } |\beta^1|, |\beta^2| \leq C : \left|F_N^{\varepsilon}(\beta^1, 0, \boldsymbol{Q}) - F_N^{\varepsilon}(\beta^2, 0, \boldsymbol{Q})\right| \leq L\left|\beta^1 - \beta^2\right|\right) = 1.$$

证明。对于任意 $k \in \{1, ..., n\}$，有

$$\frac{\partial}{\partial \beta^k} F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) \overset{(1.3)}{=} \langle \tfrac{1}{N}\beta_k H_N(\sigma^k)\rangle \tag{3.28}$$

其中 $\langle\cdot\rangle$ 表示如前一个引理中所述，对吉布斯测度取期望。因此，由（2.1），当 $L = L(C$足够大时

$$\mathbb{P}\left(\sup_{\boldsymbol{Q}, |\beta|\leq C}\left|\frac{\partial}{\partial \beta^k} F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q})\right| \leq L\right) \to 1.$$

这意味着 $F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q})$ 关于 β是利普希茨连续的，且该结论以趋于 1 的概率成立。 □

接下来的命题将结合此前所有论证，证明当外场 $\boldsymbol{h}$ 为零且 β 处于高温区域时，F_N^{ε} 会随着 $N \to \infty$ 而集中。

命题 3.10（高温下的自由能）。设 $n \geq 1$。令 δ、$C > 0$ 为一些常数。存在一个 $c = c(\delta, C) > 0$，使得对所有 $\varepsilon \in (0, c^{-1})$

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{\boldsymbol{Q}\geq \delta \boldsymbol{I}} \sup_{\substack{\beta \in \mathrm{HT}(\boldsymbol{Q}) \\ |\beta|\leq C}} \left|F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) - \frac{1}{2}(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta + \log|\boldsymbol{Q}|)\right| \leq c\sqrt{\varepsilon}\right) = 1, \tag{3.29}$$

其中，上确界取遍所有对称的 $\boldsymbol{Q} \in [-1, 1]^{n\times n}$，且对角线上元素均为 1。

证明。令 $\boldsymbol{A}^i \in [-1, 1]^{n\times n}, i = 1, \ldots, M$，其中 $M \leq \lceil \frac{1}{2\varepsilon}\rceil^{n^2}$，使得对所有 $\boldsymbol{Q} \in [-1, 1]^{n\times n}$，都存在一个 i，满足

$$\|\boldsymbol{A}^i - \boldsymbol{Q}\|_{\infty} \leq \varepsilon, \quad \|(\boldsymbol{A}^i)^{\odot 2} - \boldsymbol{Q}^{\odot 2}\|_{\infty} \leq 2\varepsilon, \quad \left|\left|\boldsymbol{A}^i\right| - |\boldsymbol{Q}|\right| \leq n^{n+1}\varepsilon. \tag{3.30}$$

对于这个 i，我们有 $\boldsymbol{A}^i{}_{\frac{1}{2}\varepsilon} \subset \boldsymbol{Q}_{\varepsilon} \subset \boldsymbol{A}^i{}_{2\varepsilon}$，因此对所有 β

$$F_N^{\frac{1}{2}\varepsilon}(\beta, 0, \boldsymbol{A}^i) \leq F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) \leq F_N^{2\varepsilon}(\beta, 0, \boldsymbol{A}^i). \tag{3.31}$$

我们还构造一个有限序列 $\beta^1, ..., \beta^L$，满足 $L \leq \lceil \frac{C}{\varepsilon}\rceil^n$，使得对每个 $\beta \in \{b \in \mathbb{R}^n : |b| \leq C\}$ 都存在一个 $j \leq L$，满足 $|\beta - \beta^j| \leq \varepsilon$。接着，根据引理3.9，以趋于 1 的概率，对每个 β，都有 $|\beta| \leq C$ 某个 j 满足

$$\left|F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) - F_N^{\varepsilon}(\beta^j, 0, \boldsymbol{Q})\right| \leq c(C)\varepsilon \text{ for all } \boldsymbol{Q}, \tag{3.32}$$

并且对每个 $\boldsymbol{Q}$ 满足 $\boldsymbol{Q} > \delta \boldsymbol{I}$ 且 i 使得（3.31）成立

$$\left|\beta^{\top}\boldsymbol{Q}^{\odot 2}\beta + \log|\boldsymbol{Q}| - (\beta^j)^{\top}\boldsymbol{A}^{i\odot 2}\beta^j - \log\left|\boldsymbol{A}^i\right|\right| \leq c(\delta, n, C)\varepsilon, \tag{3.33}$$

前提是 ε 足够小，具体取决于 δ。

上界：这意味着，如果常数 c 取得足够大，具体取决于 δ、C、n，那么

$$\mathbb{P}\left(\forall \boldsymbol{Q} \geq \delta \boldsymbol{I},\ |\beta| \leq C:\ Z_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) > \exp\left(\frac{N}{2}\left(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta + \log|\boldsymbol{Q}| + c\varepsilon\right)\right)\right)$$
$$\leq \mathbb{P}\left(\exists i = 1, ..., M,\ j = 1, ..., L: Z_N^{2\varepsilon}(\beta^j, 0, \boldsymbol{A}^i) > \exp\left(\frac{N}{2}\left((\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \log|\boldsymbol{A}^i| + \frac{c}{2}\varepsilon\right)\right)\right),$$

由马尔可夫不等式和引理 3.3 可知，右端有界于

$$\sum_{i=1}^{M}\sum_{j=1}^{L}\frac{\mathbb{E}\left[Z_N^{2\varepsilon}(\beta^j, 0, \boldsymbol{A}^i)\right]}{\exp\left(\frac{N}{2}\left((\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \log|\boldsymbol{A}^i| + \frac{c}{2}\varepsilon\right)\right)} \leq \exp\left(-N\frac{c}{4}\varepsilon\right),$$

因此

$$\mathbb{P}\left(\forall \boldsymbol{Q} \geq \delta \boldsymbol{I},\ |\beta| \leq C:\ F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) \leq \frac{1}{2}(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta) + \frac{1}{2}\log|\boldsymbol{Q}| + c\varepsilon\right) \to 1.$$

下界： 由 Paley-Zygmund 不等式以及引理 3.3 和 3.4 可知，对于任意足够大的 c（取决于 δ、C、n），对所有 $\varepsilon \in (0, c^{-1})$ 以及 $N \geq c(\varepsilon, \delta)$ 和 i, j

$$\begin{aligned} & \mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) > \tfrac{1}{2}(\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \tfrac{1}{2}\log|\boldsymbol{A}^i| - \tfrac{c}{8}\varepsilon\right) \\ \geq\ & \mathbb{P}\left(Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) > \tfrac{1}{2}\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i)]\right) \\ \geq\ & \tfrac{1}{4}\frac{\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i)]^2}{\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i)^2]} \\ \geq\ & \tfrac{1}{4}e^{-c\varepsilon N}. \end{aligned} \tag{3.34}$$

否则，根据引理 3.8 （在可能增大 c之后）可得矛盾，这意味着 $\mathbb{M}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j、0、\boldsymbol{A}^i)\right) \geq \frac{1}{2}(\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{4}\sqrt{\varepsilon}$对所有 $\varepsilon \in (0, c^{-1})$ 和 $N \geq N(\varepsilon, \delta)$ 成立；随后再次应用引理 3.8 可得

$$\lim_{N\to\infty}\mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) < \frac{1}{2}(\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{2}\sqrt{\varepsilon}\right) = 0 \text{ for all } \varepsilon \in (0, c^{-1}), i, j.$$

于是（可能再次增大 c）我们有

$$\mathbb{P}\left(\exists \boldsymbol{Q} \geq \delta \boldsymbol{I},\ |\beta| \leq C:\ F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) < \frac{1}{2}(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta) + \frac{1}{2}\log|\boldsymbol{Q}| - c\sqrt{\varepsilon}\right) \tag{3.35}$$

$$\leq \sum_{i=1}^{M}\sum_{j=1}^{L}\mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) < \exp\left(\frac{1}{2}(\beta^j)^{\mathsf{T}}(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{2}\sqrt{\varepsilon}\right)\right) \to 0 \tag{3.36}$$

对于所有 $\varepsilon \in (0, c^{-1})$，这给出了下界。

□

3.2. 含外场。现在我们证明存在外场时所有温度下的下界。我们将遵循 [BK19，引理 5]。我们首先说明，如果将配分函数中的积分限制在单位球面乘积与高维超平面的交集上，引理 3.4仍然成立。

在下文中，将在球面 $\int \cdot d\sigma$ 上的积分 $\mathcal{S}_{N-1}$ 以及在 $\mathcal{S}_{N-1}^n$ 上的积分 $\int \cdot d\boldsymbol{\sigma}$ 记为 $E[\cdot]$。对于子空间 $U \subset \mathbb{R}^{n\times N}$，我们记 E^U 表示关于 $\boldsymbol{\sigma}$、在给定 $\boldsymbol{\sigma} \in U$ 条件下的期望/积分。

引理3.11. 对所有 $\delta>0$，均有

(3.37)

$$\mathbb{P}\left(\sup_{\boldsymbol{Q}\geq\delta\boldsymbol{I}}\sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})}\sup_{U}\left|\frac{1}{N}\log E^{U^\perp}\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}e^{\sum_{k=1}^n\beta_k H_N(\sigma^k)}\right]-\left(\frac{\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta}{2}+\frac{1}{2}\log|\boldsymbol{Q}|\right)\right|\leq c\sqrt{\varepsilon}\right)\to 1,$$

表示为 $N\to\infty$，其中最内层上确界取遍所有维数为 $N-2n$ 的子空间，最外层上确界取遍所有对称 $\boldsymbol{Q}\in[-1,1]^{n\times n}$，其对角线上元素均为 1。

证明。定义一个标准正交基 $w_1,...,w_N$，属于 $\mathbb{R}^N$，使得

$$U=\langle w_{N-2n},\ldots,w_N\rangle.$$

设 $\boldsymbol{A}$ 是 $\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}$ 在基 $w_1,...,w_N$ 下表示时左上角的 $(N-2n)\times(N-2n)$ 子式。对于 $\sigma\in U^\perp$，我们有 $H_N(\sigma)=\sum_{i,j=1}^{N-2n}\tilde{\sigma}_i\tilde{\sigma}_jA_{ij}=N\sum_{i=1}^{N-2n}a_i(\sigma_i)^2$，其中 $\tilde{\sigma}$ 是在基 $w_1,\ldots,w_N$ 下的 σ，而 $Na_1<...<Na_{N-2n}$ 是 $\boldsymbol{A}$ 的特征值。因此

$$E^{U^\perp}\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_kH_N(\sigma^k)\right)\right]=E^{N-2n}\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}\exp\left(N\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}a_i(\sigma_i^k)^2\right)\right],\tag{3.38}$$

其中 E^{N-2n} 是关于 $\boldsymbol{\sigma}$ 在 $\mathcal{S}^n_{N-2n-1}$ 上均匀分布的期望。令 $\boldsymbol{B}$ 是左上角 $(N-2n)\times(N-2n)$ 阶子式，它是 $\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}$ 在标准基下表示时的子式，并令 $Nb_1<...<Nb_{N-2n}$ 为其特征值。回忆 $N\lambda_1<\ldots<N\lambda_N$ 是 $\frac{\boldsymbol{J}+\boldsymbol{J}^{\mathsf{T}}}{2}$ 的特征值，因此根据柯西特征值交错不等式（见 [Par87，定理 10.1。1]）有 $\lambda_i<a_i$，$b_i<\lambda_{i+2n}$。因此（2.3）可得

$$\sup_U\sup_{\boldsymbol{\sigma}\in\mathcal{S}^n_{N-2n-1}}\left|\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}a_i(\sigma_i^k)^2-\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}\frac{\sqrt{N-2n}}{\sqrt{N}}b_i(\sigma_i^k)^2\right|\xrightarrow{\mathbb{P}}0\text{ as }N\to\infty,\tag{3.39}$$

因此

(3.40)

$$E^{U^\perp}\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_kH_N(\sigma^k)\right)\right]=E^{N-2n}\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k(\sigma^k)^{\mathsf{T}}\frac{\sqrt{N-2n}}{\sqrt{N}}\boldsymbol{B}\sigma^k\right)\right]e^{o(N)},$$

在 U 上一致地。应用命题 3.10，以 $N-2n$ 替代 N，右端等于

$$e^{\frac{N-2n}{2}(\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta+\log|\boldsymbol{Q}|)(1+\mathcal{O}(\sqrt{\varepsilon}))},\tag{3.41}$$

对所有 $\boldsymbol{Q}\geq\delta\boldsymbol{I}$ 以及 $\beta\in\mathrm{HT}(\boldsymbol{Q})$（注意，$\mathrm{HT}(\boldsymbol{Q})$ 是一个有界集），从而证明该结论 □

在证明 $\boldsymbol{m}$ 命题3.1 中围绕“有效约束矩阵” $\boldsymbol{Q}(\boldsymbol{m})\in\mathbb{R}^{n\times n}$ 给出如下

$$\hat{Q}_{k,\ell}=\frac{Q_{k,\ell}-m^k\cdot m^\ell}{\sqrt{1-|m^k|^2}\sqrt{1-|m^\ell|^2}},\tag{3.42}$$

将会出现（见（3.56）–（3.58））。下一个引理将在命题 3.1 的证明中使用，以排除 $\boldsymbol{m}$ 使得 $\hat{}\,\boldsymbol{Q}(\boldsymbol{m})$ 有一个较小的特征值，或者 $\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}$ 有一个接近 1 的条目。

引理 3.12。对于任意的 β、$(h_1,\ldots,h_n)$ 以及对角线上为 $1's$ 的正定对称 $\boldsymbol{Q}\in[-1,1]^{n\times n}$，存在一个 $\delta\in(0,1)$，使得

$$\mathbb{P}\left(\sup_{\substack{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})>0,\hat{\boldsymbol{Q}}(\boldsymbol{m})\text{ has eval. }<\delta\\ \text{or }\|\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\|_\infty>1-\delta}}F_{\mathrm{TAP}}(\boldsymbol{m})\leq\sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\geq\delta\boldsymbol{I},\|\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\|_\infty\leq 1-\delta}F_{\mathrm{TAP}}(\boldsymbol{m})\right)\to 1.$$

证明。取满足 $\eta \le (\sqrt{2}\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2)^{-1}$ 的 $\eta \in (0,1)$，并令 $\tilde{\boldsymbol{Q}} = (1-\eta)\boldsymbol{Q}$，使得依据 (1.5)，$\tilde{\boldsymbol{Q}} \in \mathrm{Plef}_n(\boldsymbol{Q},\beta)$。令 $\tilde{m}_i^k = (\tilde{\boldsymbol{Q}}^{1/2})_{k,i}$（$k,i = 1,\dots,n$ 时），其他情况下为 $\tilde{m}_i^k = 0$，并令 $\tilde{\boldsymbol{m}} = (\tilde{m}^1,\dots,\tilde{m}^n)$，使得 $\tilde{\boldsymbol{m}}\tilde{\boldsymbol{m}}^\intercal = \tilde{\boldsymbol{Q}}$，因而 $\tilde{\boldsymbol{m}} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$。于是 $\|\tilde{\boldsymbol{m}}\tilde{\boldsymbol{m}}^\intercal\|_\infty = 1-\eta$。令 $\tilde{\delta}$ 为 $\eta > 0$ 与 $\boldsymbol{Q}(\tilde{\boldsymbol{m}})$ 的最小特征值中的较小者。于是我们有

$$\begin{aligned}\frac{1}{N}\sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\ge\tilde{\delta}\boldsymbol{I},\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty\le 1-\tilde{\delta}} & F_{\mathrm{TAP}}(\boldsymbol{m}) \\ \overset{(1.4),(2.1)}{\ge} & -cn\max_i(|\beta_i|+h_i) + \tfrac{\eta}{2}\beta^\intercal\boldsymbol{Q}^{\odot 2}\beta + \tfrac{n}{2}\log\eta + \tfrac{1}{2}\log|\boldsymbol{Q}|\end{aligned} \tag{3.43}$$

其概率趋于 1。

另一方面，假设 $\delta \in (0,1)$ 和 $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$ 以及 $\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \ge 1-\delta$。那么 $(\boldsymbol{m}\boldsymbol{m}^\intercal)_{k,k} \ge 1-\delta$ 对于某个 k，并且由于 $|\boldsymbol{A}| \le \prod_k A_{k,k}$ 对于任意半正定 $\boldsymbol{A}$，我们有 $|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal| \le \delta$ 和

$$\frac{1}{N}F_{\mathrm{TAP}}(\boldsymbol{m}) \overset{(1.4),(2.1)}{\le} cn\max_i(|\beta_i|+h_i) + n^2\max_i\beta_i^2 + \frac{1}{2}\log\delta. \tag{3.44}$$

现在反过来假设 $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$，且 $\boldsymbol{Q}(\boldsymbol{m})$ 的最小特征值小于δ。令 $\boldsymbol{S} = \mathrm{Diag}((1-|m^1|^2)^{-1/2},\dots,(1-|m^n|^2)^{-1/2})$。由于 $\hat{\boldsymbol{Q}}(\boldsymbol{m}) = \boldsymbol{S}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal)\boldsymbol{S}$ 和 $\boldsymbol{S}_{ii} > 1$，$\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal$ 的最小特征值不大于 $\hat{\boldsymbol{Q}}(\boldsymbol{m})$ 的最小特征值。因此，若 $\boldsymbol{m}$ 满足 $\boldsymbol{Q}(\boldsymbol{m})$ 的最小特征值小于 δ，那么 $\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal$ 的最小特征值也小于 δ，所以以趋于 1 的概率，所有这样的 $\boldsymbol{m}$ 都满足

$$\frac{1}{N}F_{\mathrm{TAP}}(\boldsymbol{m}) \overset{(1.4),(2.1),(3.17)}{\le} cn\max_i(|\beta_i|+h_i) + n^2\max_i\beta_i^2 + \frac{1}{2}(n-1)\log n + \frac{1}{2}\log\delta. \tag{3.45}$$

如果 δ 取得足够小（具体取值依赖于 n，β，$\boldsymbol{Q}$，$(h_1,\dots,h_n)$），那么 (3.44) 和 (3.45) 右侧的值都小于 (3.43) 的最后一行；如果我们还确保 $\delta \le \tilde{\delta}$，就证明了这一结论。□

现在我们可以证明 TAP 下界命题 3.1 了。证明的思路是将 $\boldsymbol{\sigma}$ 以某个向量 $\boldsymbol{m}$ 为中心重新定心，然后将配分函数积分限制在外场贡献可以忽略的集合上，这样便能使用上一小节关于无外场时自由能的结果。

命题 *3.1* 的证明。由引理 3.12 可知，存在一个 $\delta \in (0,1)$，使得只要证明以下结论，式 (3.1) 就成立：

$$\lim_{N\to\infty}\mathbb{P}\left(F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) \ge \frac{1}{N}\sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\ge\delta\boldsymbol{I},\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty\le 1-\delta} F_{\mathrm{TAP}}(\boldsymbol{m}) - c\sqrt{\varepsilon}\right) = 1, \tag{3.46}$$

对所有 $\varepsilon \in (0, c^{-1})$ 均成立。

固定一些 $m^1,\dots,m^n \in \mathbb{R}^N$ 满足 $\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \le 1-\delta$，$\hat{\boldsymbol{Q}}(\boldsymbol{m}) \ge \delta\boldsymbol{I}$ 以及 $\boldsymbol{m}\boldsymbol{m}^\intercal \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$。根据定义（1.1），可知对所有 σ，$m \in \mathbb{R}^N$

$$H_N(\sigma) = H_N(m) + \nabla H_N(m)\cdot(\sigma-m) + H_N(\sigma-m), \tag{3.47}$$

从而对于所有 k

$$\beta_k H_N(\sigma^k) + Nh^k\cdot\sigma^k = \beta_k H_N(m^k) + Nh^k\cdot m^k + Nh^{m,k}\cdot(\sigma^k-m^k) + \beta_k H_N(\sigma^k-m^k), \tag{3.48}$$

where

$$h^{m,k} = \frac{\beta_k}{N}\nabla H_N(m^k) + h^k,\ k = 1,\dots,n, \tag{3.49}$$

是有效外场。利用这一点，我们得到

$$Z_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q}) = e^{\sum_{k=1}^n\left(\beta_k H_N(m^k)+Nh^k\cdot m^k\right)}\ E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon}\, e^{\sum_{k=1}^n\left(\beta_k H_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right]. \tag{3.50}$$

令 U 为一个 $2n$维空间，其张成空间包含 m^k、$h^{m,\,k}$（$k=1$，……，n）。现在，我们将通过插入另一个由下式给出的指示函数$\mathbb{I}^A$，对右端的期望进行下界估计

$$A := \left\{\boldsymbol{\sigma} : |P^U(\sigma^k - m^k)| \le \frac{\varepsilon}{4} \text{ for } k=1,\dots,n\right\}. \tag{3.51}$$

注意，在事件（2.1）上，对于 $\boldsymbol{\sigma}\in A$

$$|(\sigma^k-m^k)\cdot m^\ell| \le \frac{\varepsilon}{4} \quad \text{and} \quad |(\sigma^k-m^k)\cdot h^{m,\ell}| \overset{(2.1)}{\le} c\varepsilon, \tag{3.52}$$

对所有 k、$\ell\in\{1,...,n\}$均成立。因此我们得到

$$E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\, e^{\sum_{k=1}^n\left(\beta_k H_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right] \ge e^{-c\varepsilon N} E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\, e^{\sum_{k=1}^n \beta_k H_N(\sigma^k-m^k)}\right]. \tag{3.53}$$

定义 σ^k-m^k 在 $U^\perp$ 上的归一化投影为

$$\hat{\sigma}^k = \frac{P^{U^\perp}\left(\sigma^k-m^k\right)}{\left|P^{U^\perp}\left(\sigma^k-m^k\right)\right|}, k=1,\dots,n.$$

对于 $\boldsymbol{\sigma}\in A$ 和 k，有

$$|\sigma-m^k|^2 = |\sigma|^2-|m^k|^2-2(\sigma^k-m^k)\cdot m^k = 1-|m^k|^2+\mathcal{O}(\varepsilon) \tag{3.54}$$

因此

$$\left||P^{U^\perp}\left(\sigma-m^k\right)|^2-\left(1-\left|m^k\right|^2\right)\right| \le c\varepsilon.$$

因此，利用 H_N 是二次齐次的这一性质，我们得到，在事件（2.1）

$$H_N\left(\sigma-m^k\right) \ge (1-\left|m^k\right|^2)H_N\left(\hat{\sigma}\right)-c\varepsilon,$$

对所有 k，并且

$$E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\, e^{\sum_{k=1}^n \beta_k H_N(\sigma^k-m^k)}\right] \ge e^{-c\varepsilon N} E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\, e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\hat{\sigma}^k)}\right]. \tag{3.55}$$

为了将 $\mathbb{I}^{\boldsymbol{Q}_\varepsilon}$ 替换为仅依赖于 $\hat{}\boldsymbol{\sigma}=(\hat{}\sigma^1,...,\hat{}\sigma^n)$ 的指示函数，注意到如果

$$\left|\hat{\sigma}^k\cdot\hat{\sigma}^\ell-\hat{Q}_{k,\ell}\right| \overset{(3.42)}{=} \left|\hat{\sigma}^k\cdot\hat{\sigma}^\ell-\frac{Q_{k,\ell}-m^k\cdot m^\ell}{\sqrt{1-|m^k|^2}\sqrt{1-|m^\ell|^2}}\right| \le \frac{\varepsilon}{2},\quad \forall k,\ell=1,...,n \tag{3.56}$$

then

$$\begin{aligned} |\sigma^k\cdot\sigma^\ell-Q_{k,\ell}| &\le |(\sigma^k-m^k)\cdot(\sigma^\ell-m^\ell)-(Q_{k,\ell}-m^k\cdot m^\ell)|+|m^k\cdot(\sigma^\ell-m^\ell)+m^\ell\cdot(\sigma^k-m^k)| \\ &\overset{(3.52)}{\le} \sqrt{1-|m^k|^2}\sqrt{1-|m^\ell|^2}\tfrac{\varepsilon}{2}+\tfrac{\varepsilon}{2}\le\varepsilon. \end{aligned} \tag{3.57}$$

因此我们得到

$$\{\boldsymbol{\sigma}:\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\}\subset\{\boldsymbol{\sigma}:\boldsymbol{\sigma}\in\boldsymbol{Q}_\varepsilon\}$$

and

$$E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\, e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\hat{\sigma}^k)}\right] \ge E\left[\mathbb{I}^A\,\mathbb{I}^{\{\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\}}\, e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\hat{\sigma}^k)}\right]. \tag{3.58}$$

令 $\mathcal{A}$ 为由 $P^U\sigma^k$ 生成的 σ-代数，适用于 $k=1,...,n$。注意，在 $P[\cdot|\mathcal{A}]$下，$\hat{}\sigma^k$ 彼此独立，且在 $\mathcal{S}_{N-1}\cap U^\perp$ 上均匀分布。因此

$$\begin{aligned} &E\left[\mathbb{I}^A\,\mathbb{I}\{\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\}\,e^{\sum_{k=1}^n\beta_k(1-|m^k|^2)H_N(\hat{\sigma}^k)}\right]\\ &=E\left[\mathbb{I}^A\,E\left[\mathbb{I}\{\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\}\,e^{\sum_{k=1}^n\beta_k(1-|m^k|^2)H_N(\hat{\sigma}^k)}\middle|\mathcal{A}\right]\right]\\ &=E\left[\mathbb{I}^A\,E^{U^\perp}\left[\mathbb{I}_{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}}\,e^{\sum_{k=1}^n\beta_k(1-|m^k|^2)H_N(\sigma)}\right]\right]\end{aligned}\tag{3.59}$$

注意，通过令

$$(\beta_{\boldsymbol{m}})_k=\beta_k(1-|m^k|^2),k=1,...,n\quad\text{and}\quad\boldsymbol{\beta}_m=\operatorname{diag}(\beta_{\boldsymbol{m}})\in\mathbb{R}^{n\times n},\tag{3.60}$$

我们有

$$\|\boldsymbol{\beta}^{\frac12}\boldsymbol{Q}\boldsymbol{\beta}^{\frac12}\|_2=\|\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac12}\hat{\boldsymbol{Q}}(\boldsymbol{m})\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac12}\|_2,$$

从而

$$\begin{aligned}\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\in\operatorname{Plef}_n(\boldsymbol{Q},\beta)\overset{(1.5)}{\Leftrightarrow}\|\boldsymbol{\beta}^{\frac12}\boldsymbol{Q}\boldsymbol{\beta}^{\frac12}\|_2\le\tfrac{1}{\sqrt2}\quad&\Leftrightarrow\quad\|\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac12}\hat{\boldsymbol{Q}}(\boldsymbol{m})\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac12}\|_2\le\tfrac{1}{\sqrt2}\\ &\overset{(3.20)}{\Leftrightarrow}\quad\beta_{\boldsymbol{m}}\in\operatorname{HT}(\hat{\boldsymbol{Q}}(\boldsymbol{m})).\end{aligned}\tag{3.61}$$

因此　　引理 3.11 意味着在（3.37）中的事件上，量（3.59）等于 t　　o

$$\exp\left(\frac{\beta_{\boldsymbol{m}}{}^{\mathsf{T}}\hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot2}\beta_{\boldsymbol{m}}}{2}+\frac12\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|+O(\sqrt{\varepsilon})\right).\tag{3.62}$$

因此，在该事件上，(3.59)有如下下界

$$E\left[\mathbb{I}^A\right]\,\exp\left(N\left(\frac{\beta_{\boldsymbol{m}}{}^{\mathsf{T}}\hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot2}\beta_{\boldsymbol{m}}}{2}+\frac12\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\right)-Nc\sqrt{\varepsilon}\right).\tag{3.63}$$

我们还有

$$E\left[\mathbb{I}^A\right]\ge\prod_{k=1}^n\left(c\varepsilon^{2n}(1-|m^k|^2-c\varepsilon)^{\frac{N-2n-2}{2}}\right)\ge\exp\left(\frac{N}{2}\sum_{k=1}^n\log(1-|m^k|^2)-c\varepsilon N\right),\tag{3.64}$$

其中常数 c 取决于 δ 和 $N\ge c$，因为 $|m^k|^2=(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_{k,\,k}\le1-\delta$ (参见 [BK19,(2.9)])。结合 (3.53)、(3.55)、(3.58)、(3.59)、(3.63)、(3.64)，可得

$$\begin{aligned}&E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\,e^{\sum_{k=1}^n\left(\beta_kH_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right]\\ &\ge\exp\left(\sum_{k=1}^n\tfrac{N}{2}\log\left(1-|m^k|^2\right)+N\left(\tfrac{\beta_{\boldsymbol{m}}{}^{\mathsf{T}}\hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot2}\beta_{\boldsymbol{m}}}{2}+\tfrac12\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\right)-Nc\sqrt{\varepsilon}\right)\end{aligned}$$

回忆一下 (3.42) 和 (3.60)，由此可知

$$\sum_{k=1}^n\log\left(1-|m^k|^2\right)+\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|=\log\left(|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\,\prod_{k=1}^n\left(1-|m^k|^2\right)\right)=\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|,\tag{3.65}$$

and

$$\beta_{\boldsymbol{m}}{}^{\mathsf{T}}\hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot2}\beta_{\boldsymbol{m}}=\beta^{\mathsf{T}}\left(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\right)^{\odot2}\beta.\tag{3.66}$$

这意味着

$$\begin{aligned}&E\left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon\cap A}\,e^{\sum_{k=1}^n\left(\beta_kH_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right]\\ &\ge\exp\left(\tfrac{N}{2}\beta^{\mathsf{T}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})^{\odot2}\beta+\tfrac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|-Nc\sqrt{\varepsilon}\right).\end{aligned}\tag{3.67}$$

将此与（3.50）结合，我们得到

$$Z_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) \geq e^{\sum_{k=1}^n \left(\beta_k H_N(m^k)+Nh^k\cdot m^k\right)+\frac{N}{2}\beta^{\intercal}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\intercal})^{\odot 2}\beta+\frac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\intercal}|-Nc\sqrt{\varepsilon}}, \tag{3.68}$$

对所有 $\boldsymbol{m}$ 满足 $\|\boldsymbol{m}\boldsymbol{m}^{\intercal}\| \leq 1-\delta$，$\hat{\boldsymbol{Q}}(\boldsymbol{m}) \geq \delta \boldsymbol{I}$ 以及 $\boldsymbol{m}\boldsymbol{m}^{\intercal} \in \mathrm{Plef}_N(\boldsymbol{Q},\ \beta)$，其概率趋于1。回顾 (1.4)，我们看到 (3.68) 等价于

$$Z_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) \geq e^{F_{\mathrm{TAP}}(\boldsymbol{m})-Nc\sqrt{\varepsilon}}.$$

这证明了（3.46），从而完成了命题 3.1的证明。

□

4. Upper bound

在本节中，我们证明以下自由能上界。

命题4.1 (TAP上界)。设 $\boldsymbol{Q}$，β，h 如命题 *3.1*所述。对于任意$\eta>0$ 存在一个 $c=c(\beta,\ \boldsymbol{h},\ \boldsymbol{Q},\ \eta)$ 使得对所有 $\varepsilon\in(0,\ c)$

$$\lim_{N\to\infty}\mathbb{P}\left(F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) \leq \frac{1}{N}\sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)} F_{\mathrm{TAP}}(\boldsymbol{m})+\eta\right)=1. \tag{4.1}$$

该证明涉及在第4.2小节中构造一个磁化的低维子空间 $\mathcal{M}_N^n$，其性质是：围绕 $\boldsymbol{m}\in\mathcal{M}_N^n$ 重新中心化后，有效外场再次几乎完全包含于 $\mathcal{M}_N^n$ 中。第1.2小节所描述的集合 $A(\boldsymbol{m})$在这里本质上是超平面 $\boldsymbol{m}+(\mathcal{M}_N^n)^\perp$ 与笛卡尔积 $\mathcal{S}_{N-1}^n$ 的交集，其中 $(\mathcal{M}_N^n)^\perp$ 是正交空间。我们使用Fubini定理，将 $F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q})$ 中的积分写成二重积分，首先对 $\mathcal{M}_N^n$ 积分，然后对正交空间 $(\mathcal{M}_N^n)^\perp$ 积分，从而内层积分是在集合 $A(\boldsymbol{m})$ 上对重新中心化哈密顿量的积分。后者不含外场，并且具有比原模型更高的有效温度（只要 $\boldsymbol{m}\neq 0$）。然而，与下界的证明不同，对于某些 $\boldsymbol{m}$，Plefka条件可能不满足，这意味着该重新中心化哈密顿量并不处于高温状态。

因此，我们将有效哈密顿量替换为一种近似，其配分函数本质上是一个低秩Harish-Chandra-Itzykson-Zuber（HCIZ积分），并且从某种意义上始终处于高温状态。在第4.1节中，我们估计此类积分。利用第4.2节中的这些估计，我们对内层积分进行积分消去，使剩余的外层积分现在成为修正TAP自由能的积分，其中 Onsager项$\frac{N}{2}\beta^{\intercal}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\intercal})^{\odot 2}\beta$ 被 HCIZ积分的渐近式所替代。式 $F_N^\varepsilon(\beta、\ \boldsymbol{h}、\ \boldsymbol{Q})$ N 中的积分于是化为在低维空间 $\mathcal{M}_N^n$上对指数函数进行积分，其中指数为修正TAP自由能的 $\boldsymbol{m}$倍；通过拉普拉斯方法，该积分的对数转化为在所有 $\boldsymbol{m}$上的修正TAP自由能最大化点。

在小节4.3 中，我们将证明，若修正TAP自由能在临界点处的海森矩阵为负半定（最大化点处必然如此），则 $\boldsymbol{m}$ 满足Plefka条件。此外，我们还将证明，修正TAP自由能与原始TAP自由能中的Onsager项 $F_{\mathrm{TAP}}(\boldsymbol{m})$彼此接近，因此，基于修正TAP自由能得到的自由能上界$F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q})$可推出基于原始TAP自由能的上界（4.1）。

为了实施上述策略，我们必须比第3节更加依赖随机矩阵计算。将哈密顿量的一个确定性版本定义如下：

$$\tilde{H}_N(\sigma) = N\sum_{i=1}^{N}\theta_{i/N}\sigma_i^2, \tag{4.2}$$

其中 $\theta_{i/N}$ 是（2.4）中的经典位置。如果 $\boldsymbol{U}$ 是将 $\boldsymbol{J}+\boldsymbol{J}^\intercal$ 对角线化的基变换矩阵，那么由（2.4）

$$\lim_{N\to\infty}\sup_{\sigma\in\mathcal{S}_{N-1}}\left|H_N(\sigma)-\tilde{H}_N(\boldsymbol{U}\sigma)\right| = 0 \text{ in probability}, \tag{4.3}$$

因此，只需证明确定性哈密顿量的自由能上界即可。

$$\tilde{F}_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k\tilde{H}_N(\sigma^k)+\tilde{h}^k\cdot\sigma^k}\,d\boldsymbol{\sigma}, \tag{4.4}$$

其中，$\tilde{h}^k$ 是无序矩阵 $\boldsymbol{J}+\boldsymbol{J}^\intercal$的对角化基中的外场 h^k。上界将以相应的TAP自由能表示。

$$\tilde{F}_{\mathrm{TAP}}(\boldsymbol{m}) = \sum_{k=1}^n \beta_k\tilde{H}_N(m^k) + N\sum_{k=1}^n \tilde{h}^k\cdot m^k + \frac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal| + \frac{N}{2}\beta^\intercal(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal)^{\odot 2}\beta, \tag{4.5}$$

这就是简单的 F_{TAP} 由(1.4)得到，其中原始哈密顿量和外场被确定性的对角线哈密顿量与旋转后的外场所替代。

我们进一步离散化确定性哈密顿量 $[-\sqrt{2},\sqrt{2}]$ 中的 $\tilde{K}$ 个等距数

$$-\sqrt{2} = x_1 < x_2 < \cdots < x_K = \sqrt{2} - \frac{2\sqrt{2}}{K} \quad \text{and} \quad x_{k+1}-x_k = \frac{2\sqrt{2}}{K} \tag{4.6}$$

以及由以下各项给出的 $\{1,\dots,N\}$ 的相应划分 I_1、...、I_K

$$I_k = \{i : x_k \le \theta_{i/N} < x_{k+1}\} \quad \text{and} \quad I_K = \{i : x_K \le \theta_{i/N}\}. \tag{4.7}$$

考虑“分箱”哈密顿量

$$\tilde{H}_N^K(\sigma) = N\sum_{k=1}^K\sum_{i\in I_k} x_k\sigma_i^2, \tag{4.8}$$

其中，特征值 $\theta_{i/N}$ 被替换为其所属“箱”的左端点。我们将计算分箱哈密顿量的自由能上界

$$\tilde{F}_{N,K}^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k\tilde{H}_N^K(\sigma^k)+\tilde{h}^k\cdot\sigma^k}\,d\boldsymbol{\sigma}, \tag{4.9}$$

并取 $K\to\infty$ 得到（4.4）的上界。

我们首先证明自由能的上界 $F_{N,\,K}$ 在不存在外场的情况下，即 $\boldsymbol{h}=0$。为此我们使用了 [GH21] 关于 HCIZ [Har56；IZ80]秩为 n （或 n 按照 [GH21的术语，维数为 GM05；HK22]）球积分渐近性的结果。

4.1.无外场下的分箱哈密顿量。本小节中，我们计算无外场下分箱哈密顿量的自由能。

我们首先使用 [GH21] 来计算带有单位约束的自由能 $\boldsymbol{Q} = \boldsymbol{I}$，这本质上是一个秩为 n的 HCIZ 积分。现在回顾 [GH21] 的极限公式（由于这里不存在离群特征值，该公式得到了简化）。给定任意测度 ν 令 G_ν表示其定义在 $\mathbb{C} \setminus \operatorname{supp}(\nu)$ 上的 Stieltjes 变换,

$$G_\nu(z) = \int (z - x)^{-1} \, d\nu(x), \tag{4.10}$$

并且如果 λ^* 是 ν支撑集中的最右端点，我们按照 [GH21命题 1] 定义函数

$$J_\nu(z) = \lambda^* z + (v_\nu(z) - \lambda^*) G_\nu(v_\nu(z)) - \log z - \int \log |v_\nu(z) - x| \, d\nu(x) - 1 \text{ for } z > 0$$

其中

$$v_\nu(z) = \begin{cases} \lambda^* & \text{if } G_\nu(\lambda^*) \le z \\ G_\nu^{-1}(z), & \text{if } G_\nu(\lambda^*) > z. \end{cases} \tag{4.11}$$

令 E_{Haar} 表示如下概率测度：其中 $(\sigma^1, \ldots, \sigma^n)$ 是均匀采样得到的正交向量（即 Haar 分布正交随机矩阵的前 k 行）。此外，假设 $\boldsymbol{X}_N$ 是一个经验谱分布为 ν 的矩阵，并假设 $\boldsymbol{X}_N$ 的极端特征值收敛到ν支撑集中的相应最小点和最大点。结果 [GH21，命题 1] 蕴含</style}

$$\lim_{N\to\infty} \frac{1}{N} \log E_{\text{Haar}} \left[\exp\left(N \sum_{k=1}^{n} \beta_k (\sigma^k)^\intercal \boldsymbol{X}_N \sigma^k \right) \right] = \frac{1}{2} \sum_{k=1}^{n} J_\nu(2\beta_k). \tag{4.12}$$

(1.1) 中的哈密顿量可写为

$$\sum_{k=1}^{n} \beta_k \tilde{H}_N^K(\sigma^k) = N \sum_{k=1}^{n} \beta_k (\sigma^k)^\intercal \boldsymbol{X}_K \sigma^k, \tag{4.13}$$

for

$$\boldsymbol{X}_K = \operatorname{diag}\Big(\underbrace{x_1, \ldots, x_1}_{|I_1|}, \underbrace{x_2, \ldots, x_2}_{|I_2|}, \ldots, \underbrace{x_K, \ldots, x_K}_{|I_K|} \Big), \tag{4.14}$$

并且 $\boldsymbol{X}_K$ 的极限谱分布等于

$$\mu_K = \sum_{k=1}^{K} \rho_k \delta_{x_k} \quad \text{where} \quad \rho_k = \lim_{N\to\infty} \frac{|I_k|}{N} = \int_{x_k}^{x_{k+1}} d\mu_{\text{sc}}(x), \tag{4.15}$$

(回忆（4.6）-（4.7））。因此定义

$$\mathcal{F}_K(\beta) = \frac{1}{2} J_{\mu_K}(2\beta) \text{ for } \beta > 0, \tag{4.16}$$

由（4.12）可得

$$\lim_{N\to\infty} \frac{1}{N} \log E_{\text{Haar}} \left[\exp\left(N \sum_{k=1}^{n} \beta_k (\sigma^k)^\intercal \boldsymbol{X}_K \sigma^k \right) \right] = \sum_{k=1}^{n} \mathcal{F}_K(\beta_k). \tag{4.17}$$

分箱哈密顿量的自由能上界（4.9）将用修正TAP自由能表示，其中（ β_k 4.12）右侧的4.12将替换为特征值 $\tilde{\beta}_1(\boldsymbol{m})\ldots,\tilde{\beta}_n(\boldsymbol{m})$ ，它们属于 $\boldsymbol{\beta}^{1/2}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{1/2}$，即

$$\tilde{F}^{K}_{\mathrm{TAP}}(\boldsymbol{m})=\sum_{k=1}^{n}\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)+\frac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|+N\sum_{k=1}^{n}\mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})), \tag{4.18}$$

对于 $\boldsymbol{m}=(m^1,\ldots,m^n)\in\mathbb{R}^{n\times N}$，参见 (1.4), (4.5).

对于 $\nu=\mu_K$，我们有 $\lambda^*=\sqrt{2}$ 和 $G_{\mu_K}(\lambda^*)=\infty$，因此

$$v_{\mu_K}(z)=G^{-1}_{\mu_K}(z), \tag{4.19}$$

它在 $(0,\infty)$ 上是光滑的。因此， J_{μ_K} 是 $(0,\infty)$ 上的光滑函数， $\mathcal{F}_K$ 也是如此。注意，

$$\mathcal{F}_K \text{ coincides with the } \mathcal{F}_K \text{ from [BK19, (4.15)]}, \tag{4.20}$$

这可通过比较 [BK19,(4.13)-(4.15)]和(4.10)-(4.16)得到验证，其中根据 (4.19)， [BK19]的 $\lambda_K(\beta)$ 与 $v_{\mu_K}(2\beta)$相同。 [BK19 的表示、引理 10],或下方 (4.22) 的情形 $n=1$ 表明， $\mathcal{F}_K(\beta)\to 0$ 当 $\beta\to 0$ 时，因此令 $\mathcal{F}_K(0)=0$ 可将该函数连续延拓到 $[0,\infty)$（事实上， $\mathcal{F}_K$ 在 $[0,\infty$ 上是光滑的，但我们不再证明或使用这一事实）。此外，[BK19的表示、引理 10]（或 (4.22) 的情形 $n=1$ ）表明 $\mathcal{F}_K$ 是凸的，从而

$$\mathcal{F}_K(x) \text{ is Lipschitz on compact subsets of } [0,\infty). \tag{4.21}$$

我们现在使用逼近论证来推导极限自由能的公式

$$\tilde{F}^{\varepsilon}_{N,K}(\beta,0,\boldsymbol{I})=\frac{1}{N}\log\int_{\boldsymbol{I}_\varepsilon}\exp\left(\sum_{k=1}^{n}\beta_k\tilde{H}^K_N(\sigma^k)\right)$$

带有单位约束的分箱哈密顿量，来自（4.17）。由于（4.22）的右侧是光滑的，我们看到，对于所有有限的 K ，右侧的配分函数对所有 β均处于高温状态。

引理4.2。对于任意 $\beta\in[0,\infty)^n$ 和 $K\geq 1$。

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\frac{1}{N}\log\int_{\boldsymbol{I}_\varepsilon}\exp\left(\sum_{k=1}^{n}\beta_k\tilde{H}^K_N(\sigma^k)\right)d\boldsymbol{\sigma}=\sum_{k=1}^{n}\mathcal{F}_K(\beta_k), \tag{4.22}$$

其中积分区域 $\boldsymbol{I}_\varepsilon$ 是单位矩阵 $\boldsymbol{I}$ 的邻域，定义见(1.2)。

证明。 我们用（4.22）左端的积分来近似（4.17）中的期望。

为此，令 E 表示如下测度：在该测度下，$(\sigma^1,\ldots,\sigma^n)$ 是球面上的相互独立的均匀分布单位向量。定义 $\tilde{\boldsymbol{\sigma}}=(\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}})^{-\frac{1}{2}}\boldsymbol{\sigma}$，其几乎必然在 E下存在。根据构造，矩阵 $\tilde{\boldsymbol{\sigma}}$ 具有正交行。此外，如果 $\boldsymbol{O}$ 是任意的 $N\times N$ 正交矩阵，那么由于乘积球面上的均匀测度具有旋转对称性，在 E 下 $\boldsymbol{\sigma}\overset{d}{=}\boldsymbol{\sigma}\boldsymbol{O}$ 。由此可知，在 E 下

$$\tilde{\boldsymbol{\sigma}}\overset{d}{=}(\boldsymbol{\sigma}\boldsymbol{O}(\boldsymbol{\sigma}\boldsymbol{O})^{\mathsf{T}})^{-\frac{1}{2}}\boldsymbol{\sigma}\boldsymbol{O}=(\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}})^{-\frac{1}{2}}\boldsymbol{\sigma}\boldsymbol{O}=\tilde{\boldsymbol{\sigma}}\boldsymbol{O},$$

因此， $\tilde{\boldsymbol{\sigma}}$ 的 E 分布为 E_{Haar}。由于对于任意正交的 $\boldsymbol{O}$，$(\boldsymbol{\sigma}\boldsymbol{O})(\boldsymbol{\sigma}\boldsymbol{O})^{\mathsf{T}}=\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}}$ ，因此对于任意可测集 A，$P(\boldsymbol{\sigma}\in A,\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon)=P(\boldsymbol{\sigma}\boldsymbol{O}\in A,\boldsymbol{\sigma}\boldsymbol{O}\in\boldsymbol{I}_\varepsilon)=P(\boldsymbol{\sigma}\boldsymbol{O}\in A,\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon)$

同样， $E[\cdot|\boldsymbol{I}_\varepsilon]$ 的分布 $\tilde{\sigma}$ 为 E_{Haar}。引理3.2推出 $\lim_{N\to\infty}\frac{1}{N}\log P(\boldsymbol{I}_\varepsilon)=\frac{1}{2}\log|\boldsymbol{I}|=0$对于所有 $\varepsilon>0$。因此，对于所有 $\varepsilon>0$ 由式（4.17）可得

$$\lim_{N\to\infty}\frac{1}{N}\log E\left[\mathbb{I}^{\boldsymbol{I}_\varepsilon}\exp\left(N\sum_{k=1}^{n}\beta_k(\tilde{\sigma}^k)^{\mathsf{T}}\boldsymbol{X}_K\tilde{\sigma}^k\right)\right]=\sum_{k=1}^{n}\mathcal{F}_K(\beta_k). \tag{4.23}$$

由于 $(\boldsymbol{I}+\boldsymbol{A})^{-1/2}=I+O(\|\boldsymbol{A}\|_2)$，若 $\|\boldsymbol{A}\|_2\le\frac{1}{2}$，则对于 $\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon$，有

$$\|\tilde{\boldsymbol{\sigma}}-\boldsymbol{\sigma}\|_F=\|\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}}\right)^{-1/2}\boldsymbol{\sigma}-\boldsymbol{\sigma}\|_F\le\sqrt{n}\|\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}}\right)^{-1/2}-\boldsymbol{I}\|_2\le c\|\boldsymbol{I}-\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}}\|_2\le c\|\boldsymbol{I}-\boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}}\|_\infty\le c\varepsilon,$$

and so for $\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon$

$$\left|\left(\tilde{\sigma}^k\right)^{\mathsf{T}}\boldsymbol{X}_k\tilde{\sigma}^k-\left(\sigma^k\right)^{\mathsf{T}}\boldsymbol{X}_k\sigma^k\right|\le c\varepsilon.$$

因此，（4.22）由（4.23）和（4.13）得出。

□

我们将把这个公式推广到一般的正定约束 $\boldsymbol{Q}>0$。为此，我们需要下面的引理，该引理使用变量替换，将带有一般约束的积分估计为带有单位约束的积分。令 $\mathcal{B}_N(r)\subset\mathbb{R}^N$ 表示半径为 r 的闭球。

引理4.3。对于任意 $\delta\in(0,1)$，存在常数 $c=c(\delta)$，使得对于任意在对角线上的元素为 1 且 $\boldsymbol{Q}>\delta\boldsymbol{I}$ 的对称 $\boldsymbol{Q}\in[-1,1]^{n\times n}$，任意 $\varepsilon>(0,c^{-1})$，以及任意 Lipschitz $f:(\mathcal{B}_N(1+c^{-1}\varepsilon))^n\to\mathbb{R}$（其 Lipschitz 常数为 L），我们有对于 $N\ge c(\varepsilon)$

$$\begin{aligned}&\frac{1}{N}\log\int\mathbb{I}^{\boldsymbol{I}_{c^{-1}\varepsilon}}\exp\left(f\left(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma}+\frac{1}{2}\log|\boldsymbol{Q}|-c\left(1+\frac{L}{N}\right)\varepsilon\\&\le\frac{1}{N}\log\int\mathbb{I}^{\boldsymbol{Q}_\varepsilon}\exp\left(f\left(\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma}\\&\le\frac{1}{N}\log\int\mathbb{I}^{\boldsymbol{I}_{c\varepsilon}}\exp\left(f\left(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma}+\frac{1}{2}\log|\boldsymbol{Q}|+c(1+\frac{L}{N})\varepsilon.\end{aligned} \tag{4.24}$$

证明。固定 $\boldsymbol{Q}$，其中 $\boldsymbol{Q}>\delta\boldsymbol{I}$。令 $\mathbb{Q}$ 为如下测度：在该测度下，u_1、……、u_N 是 $\mathbb{R}^n$ 中协方差为 $N\boldsymbol{Q}$ 的相互独立的高斯向量，并令 $u^k=(u_{1,k},\ldots\ldots,u_{N,k})$，其中 $k=1$、……、n，如引理3.2的证明中所述。再记 $\boldsymbol{u}=(u^1,\ldots\ldots,u^n)\in\mathbb{R}^{n\times N}$，利用该引理中(3.7)所使用的同一测度变换，我们有

$$\begin{aligned}&\frac{1}{N}\log\mathbb{Q}\left(\exp\left(f\left(\frac{u^1}{|u^1|},\ldots,\frac{u^n}{|u^n|}\right)\right)\mathbb{I}^{\boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}\in N\boldsymbol{Q}_{\varepsilon c_1}}\right)+\frac{1}{2}\log|\boldsymbol{Q}|-c\left(\delta\right)\varepsilon\\&\le\frac{1}{N}\log\int\mathbb{I}^{\boldsymbol{Q}_\varepsilon}\exp\left(f\left(\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma}\\&\le\frac{1}{N}\log\mathbb{Q}\left(\exp\left(f\left(\frac{u^1}{|u^1|},\ldots,\frac{u^n}{|u^n|}\right)\right)\mathbb{I}^{\boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}\in N\boldsymbol{Q}_{\varepsilon c_2}}\right)+\frac{1}{2}\log|\boldsymbol{Q}|+c\left(\delta\right)\varepsilon,\end{aligned} \tag{4.25}$$

对于所有 N。此外，令 $\mathbb{E}$ 为独立同分布高斯向量的分布，我们有 $\mathbb{E}$ 中 $\boldsymbol{Q}^{1/2}\boldsymbol{u}$ 的分布是 $\mathbb{Q}$ 中 $\boldsymbol{u}$ 的分布，因此对于 $l=1,2$

$$\begin{aligned}&\mathbb{Q}\left(\exp\left(f\left(\frac{u^1}{|u^1|},\ldots,\frac{u^n}{|u^n|}\right)\right)\mathbb{I}^{\boldsymbol{u}\in\boldsymbol{Q}_{\varepsilon c_l}}\right)\\&=\mathbb{E}\left(\exp\left(f\left(\frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^1}{\left|\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^1\right|},\ldots,\frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^n}{\left|\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^n\right|}\right)\right)\mathbb{I}^{\boldsymbol{Q}^{1/2}\boldsymbol{u}\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^{\mathsf{T}}\in N\boldsymbol{Q}_{\varepsilon c_l}}\right).\end{aligned} \tag{4.26}$$

记作 $a\asymp b$ 如果存在仅依赖于 n 的常数 c 使得$c^{-1}\le\frac{a}{b}\le c$，并记作$a\asymp_\delta b$ 如果该常数还允许依赖于 δ，则我们有

$$\|\boldsymbol{Q}^{1/2}\boldsymbol{u}\boldsymbol{Q}^{1/2}\boldsymbol{u}^{\mathsf{T}}-N\boldsymbol{Q}\|_\infty\asymp\|\boldsymbol{Q}^{1/2}\boldsymbol{u}\boldsymbol{Q}^{1/2}\boldsymbol{u}^{\mathsf{T}}-N\boldsymbol{Q}\|_2\asymp_\delta\|\boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}-N\boldsymbol{I}\|_2\asymp\|\boldsymbol{u}\boldsymbol{u}^{\mathsf{T}}-N\boldsymbol{I}\|_\infty. \tag{4.27}$$

令 $\tilde{\sigma}^i=\frac{u^i}{|u^i|}$ 从而在 $\mathbb{E}$ 下，$\tilde{\sigma}^1,\ldots,\tilde{\sigma}^n$ 在单位球面上独立同分布。由不等式（4.27）可知，在（4.26）指示函数对应的事件上，我们有 $\left|\frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^i}{\left|\boldsymbol{Q}^{1/2}\boldsymbol{u}\right|}()^i-\boldsymbol{Q}^{1/2}\tilde{\sigma}^i\right|\le c\left(\delta\right)\varepsilon,$

并且（4.26）的最后一行有如下下界：

$$\mathbb{E}\left(\exp\left(f\left(\boldsymbol{Q}^{1/2}\tilde{\sigma}^1,\ldots,\boldsymbol{Q}^{1/2}\tilde{\sigma}^n\right)\right)\mathbb{I}_{\boldsymbol{u}\boldsymbol{u}^\intercal\in N\boldsymbol{I}_{c(\delta)^{-1}\varepsilon}}\right)e^{-Lc(\delta)\varepsilon}. \tag{4.28}$$

我们有

$$\left\{\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal\in\boldsymbol{I}_{c\varepsilon},\max_{i=1}^n\left||u^i|-N\right|\le\varepsilon N\right\}\subset\left\{\boldsymbol{u}\boldsymbol{u}^\intercal\in N\boldsymbol{I}_\varepsilon\right\}, \tag{4.29}$$

当 c 足够小，且对所有 $\varepsilon\in(0,c)$ 时。此外， $\tilde{\boldsymbol{\sigma}}$ 与 $|u^i|$相互独立，并假设 $\varepsilon\le\delta$ （不妨如此），我们有 $\mathbb{P}(\max_{i=1}^n||u^i|-1|\le\varepsilon)\to1$ 当 $N\to\infty$时，由（4.29）出发，将$c(\delta)^{-1}\varepsilon$ 代替 ε ，可得（4.28）至少为

$$\frac{1}{2}\mathbb{E}\left(\exp\left(f\left(\boldsymbol{Q}^{1/2}\tilde{\sigma}^1,\ldots,\boldsymbol{Q}^{1/2}\tilde{\sigma}^n\right)\right)\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal\in\boldsymbol{I}_{c(\delta)^{-1}\varepsilon}\right)e^{-Lc(\delta)\varepsilon},$$

对于 $N\ge c(\varepsilon)$，这推出了（4.24）的下界。（4.24）的上界可类似地推出，只需简化如下：将（4.29）替换为较简单的 $\left\{\boldsymbol{u}\boldsymbol{u}^\intercal\in N\boldsymbol{I}_\varepsilon\right\}\subset\left\{\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal\in\boldsymbol{I}_{c\varepsilon}\right\}$ 当 c 足够大，且对所有 $\varepsilon\in(0,c^{-1})$ 时，因此无需用到 $\tilde{\boldsymbol{\sigma}}$ 与 $|u^i|$ 的独立性。□

为将引理 4.2推广到非恒等约束，还需要下面的引理。设$\tilde{\beta}_1$，...，$\tilde{\beta}_n$ 表示 $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$ 的特征值。

引理4.4. 对于任意 δ、$C>0$ 和 $K\in\mathbb{N}$，存在常数 $L=c(\delta,C,K)$，使得 $(\beta,\boldsymbol{Q})\to\mathcal{F}_K(\tilde{\beta}_k)+\frac{1}{2}\log|\boldsymbol{Q}|$ 对满足 $\|\boldsymbol{Q}\|_\infty\le1$ 和 $|\beta|\le C$ 的 $\boldsymbol{Q}\ge\delta\boldsymbol{I}$ 是 L-利普希茨连续的。

证明。 $\boldsymbol{Q}$ 的特征值关于 $\boldsymbol{Q}$ 的元素是利普希茨连续的，其利普希茨常数仅取决于 n。由于 $\boldsymbol{Q}\ge\boldsymbol{I}\delta$ ，所有特征值都位于 $(\delta,n]$ 中，因此 1 2 log $|\boldsymbol{Q}|$ 对于此类 $\boldsymbol{Q}$ 关于 $\boldsymbol{Q}$ 的元素是利普希茨连续的。此外， $\tilde{\beta}_1$、...、$\tilde{\beta}_n$ 也是 $\boldsymbol{\beta}\boldsymbol{Q}$ 的特征值，因此它们作为 $\boldsymbol{Q}$ 和 β 的元素的函数是利普希茨连续的，其利普希茨常数取决于 n 和 C，并且它们可以由 C 控制。由于 $\mathcal{F}_K$ 在紧区间上是利普希茨连续的（回顾 (4.21)），结论得证。

□

现在我们可以计算极限自由能

$$\tilde{F}^\varepsilon_{N,K}(\beta,0,\boldsymbol{Q})=\frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma}, \tag{4.30}$$

无外场且带有一般约束的分箱模型 $\boldsymbol{Q}$。类似于（4.22）中， $\mathcal{F}_K$ 的光滑性意味着，对于所有有限的 K ， （4.30）中的配分函数对于所有 β都处于高温状态。

引理4.5（分箱模型的极限自由能）。对于每个 $\delta>0$ 和 $C>0$ ，我们有(4.31)

$$\lim_{\varepsilon\to0}\lim_{N\to\infty}\sup_{\boldsymbol{Q}\ge\delta\boldsymbol{I}}\sup_{|\beta|\le C}\left|\frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma}-\left(\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k)+\frac{1}{2}\log|\boldsymbol{Q}|\right)\right|=0,$$

~ ~其中最外层的上确界取遍对角线上为 1 的对称 $\boldsymbol{Q}\in[-1,1]^n$，而 β_1、.....、β_n 是 $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$的特征值。

证明。我们使用

$$f(\boldsymbol{\sigma}) = \sum_{k=1}^{n} \beta_k \tilde{H}_N^K(\sigma^k),$$

引理 4.3中。由式（4.8）并利用 $|\beta| \le C$，该 f的 Lipschitz 常数至多为 c（C） N，因此得到

$$\begin{aligned} &\tfrac{1}{N}\log\int_{\mathbb{I}^{\boldsymbol{I}_{c^{-1}\varepsilon}}} \exp\left(\textstyle\sum_{k=1}^n \beta_k \tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k)\right) d\sigma + \tfrac{1}{2}\log|\boldsymbol{Q}| - c\varepsilon \\ &\le \tfrac{1}{N}\log\int_{\mathbb{I}^{\boldsymbol{Q}_\varepsilon}} \exp\left(\textstyle\sum_{k=1}^n \beta_k \tilde{H}_N^K(\sigma^k)\right) \\ &\le \tfrac{1}{N}\log\int_{\mathbb{I}^{\boldsymbol{I}_{c\varepsilon}}} \exp\left(\textstyle\sum_{k=1}^n \beta_k \tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k)\right) d\sigma + \tfrac{1}{2}\log|\boldsymbol{Q}| + c\varepsilon, \end{aligned} \tag{4.32}$$

对于引理陈述中的任意 $\boldsymbol{Q}$。接下来，令 $\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2} = \boldsymbol{O}^\mathsf{T}\tilde{\boldsymbol{\beta}}\boldsymbol{O}$ 为一个 $n\times n$正交矩阵， $\tilde{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$ 的特征值对角矩阵（回忆式（3.23））， 由式（4.13）可得

$$\begin{aligned} \textstyle\sum_{k=1}^n \beta_k \tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k) &= N\mathrm{Tr}\left(\boldsymbol{\beta}(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})\boldsymbol{X}_K(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^\mathsf{T}\right) \\ &= N\mathrm{Tr}\left(\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\boldsymbol{X}_K\boldsymbol{\sigma}^\mathsf{T}\right) \\ &= N\mathrm{Tr}\left(\boldsymbol{O}^\mathsf{T}\tilde{\boldsymbol{\beta}}\boldsymbol{O}\boldsymbol{\sigma}\boldsymbol{X}_K\boldsymbol{\sigma}^\mathsf{T}\right) \\ &= N\mathrm{Tr}\left(\tilde{\boldsymbol{\beta}}\boldsymbol{O}\boldsymbol{\sigma}\boldsymbol{X}_K(\boldsymbol{O}\boldsymbol{\sigma})^\mathsf{T}\right) \\ &= \textstyle\sum_{k=1}^n \tilde{\beta}_k \tilde{H}_N^K((\boldsymbol{O}\boldsymbol{\sigma})^k). \end{aligned}$$

我们有 $\int g(\boldsymbol{O}\boldsymbol{\sigma})\,d\boldsymbol{\sigma} = \int g(\boldsymbol{\sigma})\,d\boldsymbol{\sigma}$ 对任意可测的 g，由对称性可知，因此我们可以使用引理 4.2估计式（4.32）的第一行和最后一行，从而得到，对于任意固定的 $\boldsymbol{Q}$ 和 β

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\left|\frac{1}{N}\log\int_{\mathbb{I}^{\boldsymbol{Q}_\varepsilon}}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma} - \left(\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k) + \frac{1}{2}\log|\boldsymbol{Q}|\right)\right| = 0. \tag{4.33}$$

与命题 3.10 中的做法相同，接着可通过构造两个由有限个 $\boldsymbol{Q}$ 和 β 在式（4.31中的网格，证明其关于 $\boldsymbol{A}^1$、...、 $\boldsymbol{A}^M \in [-1,1]^{n\times n}$ 和 b^1、...、 $b^L \in (0,\ C]^n$ 的一致性，然后利用 Lipschitz 连续性（见引理 4.4）。更具体地说，我们可以选取两个网格，使得对所有 $\boldsymbol{Q}$ ，均有

$$|\boldsymbol{Q} - \boldsymbol{A}^i| \le \varepsilon \quad \text{and} \quad \boldsymbol{A}^i_{\frac{\varepsilon}{2}} \subset \boldsymbol{Q}_\varepsilon \subset \boldsymbol{A}^i_{2\varepsilon}$$

对于某些 $i \in \{1,...,M\}$（参见（3.30）-（3.31）），以及对于所有 β

$$\max_k |\beta_k - b^{j_k}| \le \varepsilon \quad \text{and} \quad \left|\sum_{k=1}^n \beta_k \tilde{H}_N^K(\sigma^k) - \sum_{k=1}^n b^{j_k}\tilde{H}_N^K(\sigma^k)\right| \le c(C)\varepsilon N$$

对于某些 j_1，.....， $j_k \in \{1,...,L\}$（参见（3.32））。然后利用引理 4.4 完成证明（参见（3.33））。

□

~ ~为了从分箱版本 F_N^ε中恢复（4.4）的自由能$F_{N,\ K}^\varepsilon$我们需要令分箱数 $K\to\infty$趋于无穷。下一个引理表明，若 β 处于高温区域~HT（$\boldsymbol{Q}$）（回顾（3.20）），则（ $\sum_{k=1}^n \mathcal{F}_K(\beta_k)$ 中的和（4.31）收敛到退火自由能中出现的简单表达式（回顾（3.9）、（3.29））。

引理 4.6. 有如下结论。

$$\lim_{K\to\infty}\sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})}\left|\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k) - \frac{1}{2}\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta\right| = 0, \tag{4.34}$$

~ ~在对称正定矩阵 $\boldsymbol{Q}$上均匀分布，其中 β_1、...、 β_n 是$\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$的特征值。

证明。 回顾（4.20）。根据 [BK19，引理 14 + （4.28）]我们得到

$$\lim_{K\to\infty}\sup_{\tilde{\beta}\in[0,\frac{1}{\sqrt{2}}]}\left|\mathcal{F}_K(\tilde{\beta})-\frac{\tilde{\beta}^2}{2}\right|=0. \tag{4.35}$$

如果 $\beta\in\mathrm{HT}(\boldsymbol{Q})$，那么 $\tilde{\beta}_k\le\frac{1}{\sqrt{2}}$对所有 $k\in\{1,\dots,n\}$ 根据定义（3.20）。因此由（4.35）可得

$$\lim_{K\to\infty}\sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})}\left|\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k)-\sum_{k=1}^n\frac{\tilde{\beta}_k^2}{2}\right|=0.$$

~现在我们可以用 β 和 $\boldsymbol{Q}$ 来表示 β，使用

$$\sum_{k=1}^n\tilde{\beta}_k^2=\mathrm{Tr}\left(\left(\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}\right)^2\right)=\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{\beta}\boldsymbol{Q}\boldsymbol{\beta}\right)=\sum_{ij}\left(\boldsymbol{Q}\boldsymbol{\beta}\right)_{ij}\left(\boldsymbol{Q}\boldsymbol{\beta}\right)_{ji}=\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta.$$

□

4.2. 关于修正TAP自由能的上界。本小节中，我们证明了在外场存在的情况下，自由能关于修正TAP自由能的一个上界。

主要思路是将每个 n 球面分成两部分：一个 $\mathcal{M}_N$ 维数远小于 N的子空间，在该子空间中可以感受到外场的大部分作用；以及其补空间 $\mathcal{M}_N^{\perp}$，它几乎与所有外场正交（如[BK19，第 4]节所述）。我们将配分函数积分写成先对低维$\mathcal{M}_N$、再对高维 $\mathcal{M}_N$进行的二重积分，其中内层积分是重新中心化的哈密顿量的配分函数。内层积分本质上是一个无外场的配分函数，因此可以利用上一小节的结果对其进行估计。通过这种方式，我们得到一个配分函数的估计，其中剩余的外层积分现在是对 N 乘以修正TAP自由能的指数的积分，而其中的Onsager项是一个涉及 $\mathcal{F}_K$ 而非$\frac{N}{2}\beta^{\mathsf{T}}\left(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\right)^{\odot 2}\beta$的表达式。由于外层积分的维数远小于 N，因此我们可以使用拉普拉斯方法，根据修正TAP自由能的最大值对其进行估计。

以下引理构造空间 $\mathcal{M}_N$。回顾一下，外场记作 $\boldsymbol{h}\in\mathbb{R}^{n\times N}$，并且对于每个 $k\in\{1,...,n\}$，在固定值 $h_1,...,h_n\ge 0$ 时满足 $|h^k|=h_k$；此外，哈密顿量的对角化基中的外场记作 $\boldsymbol{h}=(h^1,...,h^n)$。

引理 4.7. 设 $N\ge 1$。对于任意的 $\beta_1,\dots,\beta_k$ 和 $h^1,\dots,h^n\in\mathbb{R}^N$，存在一个线性子空间序列 $\mathcal{M}_1,\ \mathcal{M}_2,\ ...,$ 使得 $\mathcal{M}_N\subset\mathbb{R}^N$，

$$\dim(\mathcal{M}_N)\le nN^{3/4}$$

且 $\mathcal{M}_N^n=(\mathcal{M}_N)^n$ 在该映射下近似不变

$$\boldsymbol{m}=(m^1,...,m^n)\to\left(\beta_1\tfrac{1}{N}\nabla\tilde{H}_N(m^1)+\tilde{h}^1,\dots,\beta_n\tfrac{1}{N}\nabla\tilde{H}_N(m^n)+\tilde{h}^n\right)$$

即

$$\lim_{N\to\infty}\sup_{\substack{m\in\mathcal{M}_N^n\\|m^1|,\dots,|m^n|\le 1}}\max_{k=1,\dots,n}\left|P^{\mathcal{M}_N^{\perp}}\left(\frac{\beta_k}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\right|=0. \tag{4.36}$$

证明。 根据 [BK19，引理 17] 在 $\beta = 1$ 的条件下，对于每个 k，都存在一个子空间 $\mathcal{M}_{N,\,k} \subset \mathbb{R}^N$ 使得

$$\lim_{N\to\infty} \sup_{\substack{m\in\mathcal{M}_{N,k}\\ |m^k|\le 1}} \left| P^{\mathcal{M}_{N,k}^{\perp}} \left(\frac{1}{N}\nabla \tilde{H}_N(m^k) + \tilde{h}^k \right) \right| = 0. \tag{4.37}$$

令 $\mathcal{M}_N := \mathcal{M}_{N,\,1} + \ldots + \mathcal{M}_{N,\,n}$，我们有 $\dim(\mathcal{M}_N) \le nN^{\frac{3}{4}}$，并且对于任意满足 $|m^k| < 1$ 的 k 和 $m^k \in \mathcal{M}_N$，可以分解为

$$m^k = v^1 + \ldots + v^n \tag{4.38}$$

对于某些 $v^l \in \mathcal{M}_{N,\,l}$、$|v^k| < 1$、$l = 1$、……、$n$。因此（利用 $\nabla \tilde{H}_N(m)$ 关于 m 和 $\mathcal{M}_{N,\,l} \subset \mathcal{M}_N$ 对所有 l 都是线性的）

$$\begin{aligned}
&\sup_{m^k\in\mathcal{M}_N, |m^k|<1} \left| P^{\mathcal{M}_N^{\perp}} \left(\tfrac{\beta_k}{N}\nabla \tilde{H}_N(m^k) + \tilde{h}^k \right) \right| \\
&= \sup_{\forall l: v^l\in\mathcal{M}_{N,l}, |m^l|<1} \left| P^{\mathcal{M}_N^{\perp}} \left(\tfrac{\beta_k}{N} \sum_{l=1}^n \nabla \tilde{H}_N(v^l) + \tilde{h}^k \right) \right| \\
&\le \beta_k \sum_{l=1}^n \sup_{v\in\mathcal{M}_{N,l}, |v|<1} \left| P^{\mathcal{M}_{N,l}^{\perp}} \left(\tfrac{1}{N}\nabla \tilde{H}_N(v^l) \right) \right| + \left| P^{\mathcal{M}_{N,k}^{\perp}} \tilde{h}^k \right| \\
&\le \beta_k \sum_{l=1}^n \sup_{v\in\mathcal{M}_{N,l}, |v|<1} \left| P^{\mathcal{M}_{N,l}^{\perp}} \left(\tfrac{1}{N}\nabla \tilde{H}_N(v^l) + \tilde{h}^l \right) \right| + c\,(\beta) \max_{l=1,\ldots,n} \left| P^{\mathcal{M}_{N,l}^{\perp}} \tilde{h}^l \right|
\end{aligned}$$

因此由（4.37）可得（4.36）。 □

下一个引理表明，在没有外场的情况下，限制在先前构造的子集补集上的配分函数，与不受限制的配分函数满足相同的近似关系。回顾第 3.2 小节开头的内容，E^U 表示关于 $\boldsymbol{\sigma} \in \mathcal{S}_{N-1}^n$ 在给定 $\boldsymbol{\sigma} \in U$ 条件下的期望，其中 U 是某个集合。

引理 4.8。对于任意 $C > 0, K > 0, \delta > 0$

$$\lim_{\varepsilon\to 0} \limsup_{N\to\infty} \sup_{\boldsymbol{Q}\ge\delta\boldsymbol{I}} \sup_{|\beta|\le C} \left| \frac{1}{N} \log E^{(\mathcal{M}_N^n)^{\perp}} \left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k \tilde{H}_N(\sigma^k)} \right] - \sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k) - \frac{1}{2}\log|\boldsymbol{Q}| \right| \le \frac{c}{K}$$

其中 $(\mathcal{M}_N^n)_{N\ge 1}$ 是引理 4.7 中的子空间序列，$\beta_1, \ldots, \beta_n$ 是 $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$ 的特征值。

证明。回顾 $N' = \dim(\mathcal{M}_N^n) \le nN^{\frac{3}{4}}$。与引理 3.11 的证明类似，令 $w_1, \ldots, w_N$ 为 $\mathbb{R}^N$ 的标准正交基，使得空间 $\mathcal{M}_N$ 由这些向量中的最后 $N - N'$ 个张成。令 $\boldsymbol{D}$ 为对角矩阵，满足 $D_{jj} = N\theta_{j/N}$，从而 $\tilde{H}_N(\sigma) = \sigma^{\mathsf{T}}\boldsymbol{D}\sigma$。令 $\boldsymbol{A}$ 为 $\boldsymbol{D}$ 在基 $w_1, \ldots, w_N$ 下表示时的 $(N - N') \times (N - N')$ 子式。根据特征值交错不等式和 (2.7)，$\boldsymbol{A}$ 的特征值 $Na_1, \ldots, Na_{N-N'}$ 满足 $Na_j = N\theta_{j/N} + o(1) = (N - N')\theta_{j/(N-N')} + o(N)$。我们有

$$\begin{aligned}
&E^{(\mathcal{M}_N^n)^{\perp}} \left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon} \exp\left(\sum_{k=1}^n \beta_k \tilde{H}_N(\sigma^k) \right) \right] \\
=&E^{N-N'} \left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon} \exp\left((N - N') \sum_{k=1}^n \beta_k \sum_{j=1}^{N-N'} \theta_{j/N-N'} (\sigma_j^k)^2 \right) \right] e^{o(N)}.
\end{aligned}$$

Also

$$|\tilde{H}_N^K(\sigma^i) - \tilde{H}_N(\sigma^i)| = \left| N \sum_{k=1}^{K} \sum_{j \in I_k} (x_k - \theta_{j/N})(\sigma_j^i)^2 \right| \leq \frac{2\sqrt{2}}{K} N \sum_{k=1}^{K} \sum_{j \in I_k} (\sigma_j^i)^2 = N \frac{2\sqrt{2}}{K},$$

因此，对于有界的 β，我们得到

$$\begin{aligned} & E^{N-N'} \left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon} \exp \left((N-N') \sum_{k=1}^{n} \beta_k \sum_{j=1}^{N-N'} \theta_{j/(N-N')} (\sigma_j^k)^2 \right) \right] \\ = & E^{N-N'} \left[\mathbb{I}^{\boldsymbol{Q}_\varepsilon} \exp \left(\sum_{k=1}^{n} \beta_k \tilde{H}_{N-N'}^K (\sigma^k) \right) \right] e^{\mathcal{O}(\frac{N}{K})}. \end{aligned}$$

该结论由引理4.5得出。 □

下一个引理将用于证明， $\boldsymbol{m}$ 中某些 $|m^k|$ 接近1的情况对配分函数的贡献可以忽略不计。

引理4.9。设 $\mathcal{U}_N \subset \mathbb{R}^N$ 是一个维数为 $N' = o\left({}^{N}\!/_{\log N} \right)$ 的线性子空间序列。对于所有 $\eta \in (0,1)$，都有

$$\limsup_{N \to \infty} \frac{1}{N} \log \int_{\mathcal{S}_{N-1}^n} \mathbb{I}^{\left\{ \boldsymbol{\sigma} :\ \exists j \in \{1,\ldots,n\} : \left| P^{\mathcal{U}_N}(\sigma^j) \right|^2 > 1-\eta \right\}} d\boldsymbol{\sigma} < \frac{1}{2} \log \eta.$$

证明。首先注意到

$$\begin{aligned} \int_{\mathcal{S}_{N-1}^n} \mathbb{I}^{\left\{ \boldsymbol{\sigma} :\ \exists j \in \{1,\ldots,n\} : \left| P^{\mathcal{U}_N}(\sigma^j) \right|^2 > 1-\eta \right\}} d\boldsymbol{\sigma} & \leq \sum_{j=1}^{n} \int_{\mathcal{S}_{N-1}^n} \mathbb{I}^{\left\{ \boldsymbol{\sigma} : \left| P^{\mathcal{U}_N}(\sigma^j) \right|^2 > 1-\eta \right\}} d\boldsymbol{\sigma} \\ & = \sum_{j=1}^{n} \int_{\mathcal{S}_{N-1}} \mathbb{I}^{\left\{ \sigma^j : \left| P^{\mathcal{U}_N}(\sigma^j) \right|^2 > 1-\eta \right\}} d\sigma^j. \end{aligned} \tag{4.39}$$

由 [BK19,(2.9)]

$$\int_{\mathcal{S}_{N-1}} \mathbb{I}^{\left\{ \sigma^j : \left| P^{\mathcal{U}_N}(\sigma^j) \right|^2 > 1-\eta \right\}} d\sigma^j = \frac{\Gamma\left(\frac{N}{2}\right)}{\pi^{\frac{N'}{2}} \Gamma\left(\frac{N-N'}{2}\right)} \int_{\mathcal{B}_{N'}} \mathbb{I}^{\{m : |m|^2 > 1-\eta\}} (1-|m|^2)^{\frac{N-N'-2}{2}} dm, \tag{4.40}$$

其中， dm 表示 $\mathcal{B}_{N'} = \{m \in \mathbb{R}^{N'}$ 上的勒贝格测度： $|m| < 1\}$。由于

$$\frac{1}{N} \log \left(\frac{\Gamma\left(\frac{N}{2}\right)}{\pi^{\frac{N'}{2}} \Gamma\left(\frac{N-N'}{2}\right)} \right) = o(1)$$

and

$$\begin{aligned} \frac{1}{N} \log \int_{\mathcal{B}_{N'}} \mathbb{I}^{\{m : |m|^2 > 1-\eta\}} (1-|m|^2)^{\frac{N-N'-2}{2}} dm & = \frac{1}{2} \log \eta + o(1) + \frac{1}{N} \log \int_{\mathcal{B}_{N'}} dm \\ & = \frac{1}{2} \log \eta + o(1) + \frac{1}{N} \underbrace{\log \left(\frac{\pi^{\frac{N'}{2}}}{\Gamma(\frac{N'}{2}+1)} \right)}_{=o(N)}, \end{aligned}$$

该结论由（4.39）和（4.40）得出。 □

现在我们证明，确定性哈密顿量（4.2）的自由能由相应的修正TAP自由能（4.8）从（4.18）给出的上界所约束。

命题4.10. 对于 $K\geq 2$，存在一个 $C=C(\beta,\boldsymbol{h},\boldsymbol{Q},K)$，使得当 $\varepsilon\in(0,C)$ 时， N足够大且 $c=c(\beta)$

$$\tilde{F}_N^{\varepsilon}(\beta,\boldsymbol{h},\boldsymbol{Q})\leq \frac{1}{N}\sup_{\boldsymbol{m}:\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}<\boldsymbol{Q}} \tilde{F}_{TAP}^{K}(\boldsymbol{m})+\frac{c}{K}. \tag{4.41}$$

证明。 令 $\mathcal{M}_N^n=\mathcal{M}_N\times\ldots\times\mathcal{M}_N$为引理 4.7中的空间，其中每个 n分量的维数均为 $N'\leq nN^{\frac{3}{4}}$。对于任意 $\boldsymbol{\sigma}=(\sigma^1,\ldots,\sigma^n)\in\mathbb{R}^{n\times N}$，令 $\boldsymbol{m}$为到$\mathcal{M}_N^n$的投影，即 $\forall i\in\{1,\ldots,n\}$, $m^i:=P^{\mathcal{M}_N}\sigma^i$且 $\boldsymbol{m}=(m^1,\ldots,m^n)$.

按照 ($\tilde{H}_N(\sigma^k)$ 中的方法，以 m^k 为中心重新调整 3.50)，得到

$$\begin{aligned}&\int_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\left(\beta_k\tilde{H}_N(\sigma^k)+Nh^k\cdot\sigma^k\right)\right)d\boldsymbol{\sigma}\\ &=E\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)}e^{\sum_{k=1}^n\left(N\left(\frac{\beta_k}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\cdot(\sigma^k-m^k)+\beta_k\tilde{H}_N(\sigma^k-m^k)\right)}\right].\end{aligned} \tag{4.42}$$

由于引理 4.7蕴含

$$\lim_{N\to\infty}\sup_{m\in\mathcal{M}_N^n}\sup_{\substack{\hat{\boldsymbol{\sigma}}\in(\mathcal{M}_N^n)^{\perp}\\ |\sigma^i|\leq 1,\forall i\in\{1,\ldots,n\}}}\left|\sum_{k=1}^n\left(\frac{\beta_k}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\cdot(\sigma^k-m^k)\right|=0,$$

有效外场消失，且 (4.42) 至多为

$$E\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}_N(\sigma^k-m^k)\right)\right]e^{o(N)}. \tag{4.43}$$

期望等于

$$E\left[e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)}E\left[\mathbb{I}_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}_N(\sigma^k-m^k)\right)\middle|\boldsymbol{m}\right]\right] \tag{4.44}$$

其中， $E[\cdot|\boldsymbol{m}]$分布（关于 $\boldsymbol{\sigma}-\boldsymbol{m}$ ）是在 n 个球面 $\mathcal{M}_N^{\perp}\cap\mathcal{S}_{N-1}(\sqrt{1-|m^k|^2})$ 的笛卡尔积上的均匀分布，适用于 $k\in\{1,\ldots,n\}$。

注意，对于所有的 k、 $\ell\in\{1,\ldots,n\}$

$$\begin{aligned}&(\sigma^k-m^k)\cdot(\sigma^\ell-m^\ell)-(Q_{k,\ell}-(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_{k,\ell})\\ =&\sigma^k\cdot\sigma^\ell-Q_{k,\ell}-\underbrace{(\sigma^k-m^k)\cdot m^\ell}_{=0}-\underbrace{(\sigma-m)^\ell\cdot m^k}_{=0}-\underbrace{(m^k\cdot m^\ell-(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_{k,\ell})}_{=0},\end{aligned}$$

由于 $m^k=P^{\mathcal{M}_N}\sigma^k\in\mathcal{M}_N$ 和 $\sigma^k-m^k\in\mathcal{M}_N^{\perp}$，所以

$$\{\boldsymbol{\sigma}:\boldsymbol{\sigma}\in\boldsymbol{Q}_\varepsilon\}=\{\boldsymbol{\sigma}:\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_\varepsilon\}. \tag{4.45}$$

令 $\hat{}\sigma^k=\frac{\sigma^k-m^k}{\sqrt{1-|m^k|^2}}$。 还利用 $\tilde{H}_N$ 是二次齐次的（回顾（4.2））， （4.44）中的表达式等于

$$E\left[e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)}E\left[\mathbb{I}_{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_\varepsilon\}}e^{\sum_{k=1}^n\beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)}\middle|\boldsymbol{m}\right]\right]. \tag{4.46}$$

令 $\eta > 0$并定义 $W_j(\eta) = \{\boldsymbol{\sigma}\colon \ |m^j|^2 \le 1-\eta\}$以及 $W(\eta) = \bigcap_{j=1}^n W_j(\eta)$。利用$N^{-1}\tilde{H}_N(\sigma)$ β_k, $|\tilde{h}^k|$均有界以及引理 4.9我们得到

$$
\begin{aligned}
(4.47)\quad & E\left[\mathbb{I}^{W_j(\eta)^c} e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)} E\left[\mathbb{I}^{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_\varepsilon\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)}\middle|\boldsymbol{m}\right]\right] \\
& \le e^{cN} E\left[\mathbb{I}^{W_j(\eta)^c}\right] \le e^{N(c+\log\eta)} \le \exp\left(N\tilde{F}_{\mathrm{TAP}}^K(0)\right),
\end{aligned}
$$

如果 η 取值足够小（具体取决于 $\boldsymbol{Q}$、β和 N 足够大），那么要证明（4.41），只需给出以下界

$$
E\left[\mathbb{I}^{W(\eta)} e^{\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)} E\left[\mathbb{I}^{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_\varepsilon\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)}\middle|\boldsymbol{m}\right]\right].
$$

ˆ 回顾矩阵 $\boldsymbol{Q}(\boldsymbol{m})$，其定义如下

$$
\hat{\boldsymbol{Q}}(\boldsymbol{m})_{ij} = \frac{Q_{ij} - m^i\cdot m^j}{\sqrt{1-|m^i|^2}\sqrt{1-|m^j|^2}}.
$$

设 $\varepsilon' = \varepsilon\eta^{-1}$。对于 $\boldsymbol{\sigma}\in W(\eta)$

$$
(4.48)\qquad \{\boldsymbol{\sigma}:\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})_\varepsilon\} \subset \{\boldsymbol{\sigma}:\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}\}.
$$

利用这一点，我们可以对（4.47）给出以下上界

$$
(4.49)\quad E\left[\mathbb{I}^{W(\eta)} e^{\sum_{k=1}^n \beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k} E\left[\mathbb{I}^{\{\hat{\boldsymbol{\sigma}}\in\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}\}}\exp\left(\sum_{k=1}^n\beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)\right)\middle|\boldsymbol{m}\right]\right].
$$

因为 ˆσ^k是 在 $\mathcal{S}_{N-1}^n\cap(\mathcal{M}_N^n)^\perp$ 下均匀分布于 $E[\cdot|\boldsymbol{m}]$，我们也可以将其写为 as

$$
(4.50)\quad E\left[\mathbb{I}^{W(\eta)} e^{\sum_{k=1}^n \beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k} E^{(\mathcal{M}_N^n)^\perp}\left[\mathbb{I}^{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}}\exp\left(\sum_{k=1}^n\beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)\right)\right]\right].
$$

注意到（4.50）有如下上界：

$$
\begin{aligned}
(4.51)\quad & E\left[\mathbb{I}^{W(\eta)}\,\mathbb{I}^{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}\}} e^{\sum_{k=1}^n \beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k} E^{(\mathcal{M}_N^n)^\perp}\left[\mathbb{I}^{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N\left(\hat{\sigma}^k\right)}\right]\right] \\
& + e^{cN} E\left[\mathbb{I}^{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}\}^c} E^{(\mathcal{M}_N^n)^\perp}\left[\mathbb{I}^{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}}\right]\right],
\end{aligned}
$$

对任意 $\delta>0$，其中我们粗略地将 exp 中的所有项都以上界 cN 取代，从而得到第二项。由于在 $E^{(\mathcal{M}_N^n)^{\mathsf{T}}}$ 下，ˆσ^k 在子空间 $\mathcal{M}_N^\perp$（其维数为 $N-N'$）中独立同分布，且均匀分布在半径为 1 的球面上，由 (3.4)（用 $N-N'$ 替代 N）和 (3.17) 可得 $E^{(\mathcal{M}_N^n)^\perp}[\mathbb{I}_{\varepsilon'}^{\hat{}}\boldsymbol{Q}(\boldsymbol{m})] \le e^{\frac{1}{2}\log(2\delta n^{n-1})(N-N')}$，对 $\varepsilon'\le\delta$ 成立。因此存在 $\delta>0$，使得 (4.51) 的第二项至多为 $\exp(N\tilde{F}^K\mathrm{TAP}(0))$。因此，只需对 (4.51) 的第一项给出上界，即可证明 (4.41)（参见 (4.47)）。

现在，我们可以应用引理 4.8，用 $(\beta_{\boldsymbol{m}})_k = \beta_k(1-|m^k|^2)$ 代替 β_k，$\boldsymbol{\beta}_{\boldsymbol{m}} =$对角矩阵 $\beta_{\boldsymbol{m}}\in\hat{}\mathbb{R}^{n\times n}$ 代替 $\boldsymbol{\beta}$，并用 $\boldsymbol{Q}$ 代替 $\boldsymbol{Q}$，从而将（4.51）的第一项界定为

$$
\begin{aligned}
(4.52)\quad & E\Bigg[\mathbb{I}^{W(\eta)}\,\mathbb{I}^{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}\}}\exp\Bigg(\sum_{k=1}^n\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k \\
& \qquad + N\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m}))+\frac{N}{2}\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\Bigg)\Bigg] e^{o(N)+\frac{cN}{K}},
\end{aligned}
$$

˜回顾（4.18）可知 $\beta_k(\boldsymbol{m})$ 是对称半正定矩阵的特征值

$$\boldsymbol{\beta}^{1/2}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{1/2}=\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac{1}{2}}\hat{\boldsymbol{Q}}(\boldsymbol{m})\boldsymbol{\beta}_{\boldsymbol{m}}^{\frac{1}{2}}.$$

由于每个 m^k 都是到 $\mathcal{M}_N$ 我们可以使用 [BK19, (2.9)]将期望写成

$$\begin{aligned}(4.53)\quad &\left(\frac{1}{\pi^{\frac{N'}{2}}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-N'}{2}\right)}\right)^n\int_{(\mathcal{B}_N(\sqrt{1-\eta})\cap\mathcal{M}_N)^n}\mathbb{I}^{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}\}}\prod_{k=1}^n(1-|m^k|^2)^{\frac{N-N'-2}{2}}\\&\times\exp\left(\sum_{k=1}^n\left(\beta_k\tilde{H}_N(m^k)+N\tilde{h}^k\cdot m^k\right)+N\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m}))+\frac{N}{2}\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\right)d\boldsymbol{m},\end{aligned}$$

其中， $\mathcal{B}_N(r)$表示 $\mathbb{R}^N$中的半径为 n的球， $d\boldsymbol{m}$是 $\mathcal{M}_N^n$上的 nN'维勒贝格测度。我们有

$$\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|+\sum_{k=1}^n\log(1-|m^k|^2)\overset{(3.65)}{=}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|.$$

因此，回顾（4.18），我们有（4.53）等于 4.54

$$\left(\frac{1}{\pi^{\frac{N'}{2}}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-N'}{2}\right)}\right)^n\int_{(\mathcal{B}_N(\sqrt{1-\eta})\cap\mathcal{M}_N)^n}\mathbb{I}^{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}\}}\exp\left(\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})-\tfrac{N'+2}{2}\sum_{k=1}^n\log(1-|m^k|^2)\right)d\boldsymbol{m}.$$

由于（4.54）中的前因子至多为 $e^{o(N)}$且 $\int_{\mathcal{M}_N^n}d\boldsymbol{m}\le\int_{B_{nN'}(n)}d\boldsymbol{m}=\frac{\pi^{\frac{nN'}{2}}}{\Gamma\left(\frac{nN'}{2}+1\right)}$

$n^{nN'}=e^{o(N)}$因此（4.52）中的期望上界为

$$(4.55)\qquad\exp\left(\sup_{\boldsymbol{m}\in\mathcal{M}_N^n,\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}}\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})+\frac{cN}{K}\right),$$

当 N 足够大时。由于 $\boldsymbol{D}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{D}>0$ 且 $\boldsymbol{D}$ 可逆，这意味着 $\boldsymbol{A}>0$ 我们有 ˆ $\boldsymbol{Q}>\delta\boldsymbol{I}$意味着 $\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}>0$，因此结论（4.41）得证。□

4.3. 最大化点的位置。 本小节将根据修正TAP自由能 $F_{\mathrm{TAP}}^K(\boldsymbol{m})$，从自由能的上界命题 4.10 推导关于TAP自由能 $\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})$ 的命题 4.1 。为此，我们将证明 $\tilde{F}^K{}_{\mathrm{TAP}}(\boldsymbol{m})$ 的最大值在某个 $\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)$ 处取得。类似于（3.61），有

$$(4.56)\qquad\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)\overset{(3.20)}{\Leftrightarrow}\beta\in\mathrm{HT}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\Leftrightarrow\tilde{\beta}(\boldsymbol{m})\in\mathrm{HT}(\boldsymbol{Q}),$$

也就是说， m 满足Plefka条件，当且仅当“重新中心化后的有效温度”˜$\beta(\boldsymbol{m})$ 位于高温区域 $\mathrm{HT}(\boldsymbol{Q})$。因此，一旦证明 $\tilde{F}^K{}_{\mathrm{TAP}}(\boldsymbol{m})$ 的最大化点满足 $\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)$，我们便可通过命题 4.1 ，从命题 F_N^ε 4.10 和引理 4.6 ，令 $K\to\infty$ 取极限，推导出自由能的上界。

为了得到 $\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})$ 的导数的简洁表达式，我们将对各项作如下解释

$$\sum_{k=1}^n\mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m}))\quad\text{and}\quad\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|$$

将（4.18）表示为初等矩阵函数的迹 [Hig08，第 1]章。

定义 4.11。给定标量函数 f和实对称矩阵 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\intercal}\in\mathbb{R}^{n\times n}$，我们定义与 f 相关联的初等矩阵函数 $f(\boldsymbol{A})$，如下：

$$f(\boldsymbol{A}):=\boldsymbol{U}f(\boldsymbol{D})\boldsymbol{U}^{\mathsf{T}}\qquad where\qquad f(\boldsymbol{D})=\operatorname{diag}(f(\lambda_1),\ldots,f(\lambda_n)).$$

当 $f(\lambda_i)$ 对所有 $i\leq n$ 都有定义时，这些矩阵值函数是定义良好的。

注释4.12. 关于 [Hig08，第1章。2] 主矩阵函数的定义更一般地基于约当标准形。不过，本文只讨论可对角化矩阵，因此定义得以简化。

由此可得

$$\sum_{k=1}^{n}\mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m}))=\operatorname{Tr}\big(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}})\big),$$

其中，$\mathcal{F}_K(\boldsymbol{A})$是与 $\mathcal{F}_K(x)$ 相关联的主矩阵函数，并且（对于 $\boldsymbol{m}<\boldsymbol{Q}$）

$$\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}|=\operatorname{Tr}\big(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\big),$$

其中 $\log(\boldsymbol{A})$ 是与 $\log(x)$ 相关联的主矩阵函数。将（4.18）中的相应项替换后，我们得到 $\tilde{F}^K_{\mathrm{TAP}}$

$$\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})=\sum_{k=1}^{n}\left(\beta_k\tilde{H}_N(m^k)+Nm^k\cdot\tilde{h}^k\right)+N\operatorname{Tr}(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}))$$

$$+\frac{N}{2}\operatorname{Tr}(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})).\tag{4.57}$$

我们希望研究该函数最大化点的临界点条件，这将需要一个用于对初等矩阵函数求导的公式。

引理 4.13。设 f 是定义域上光滑的标量函数，且 $\boldsymbol{A}(\alpha)$ 是从 $\mathbb{R}$ 的子集映射到 $\mathbb{R}^{n\times n}$ 的子集上的光滑映射，并且在该子集上 $f(\boldsymbol{A})$ 有定义。那么 $f_{ab}(\boldsymbol{A}(\alpha))$ $a,b=1,\ldots,n$ 是光滑的，且

$$\partial_\alpha\operatorname{Tr}(f(\boldsymbol{A}(\alpha)))=\operatorname{Tr}\left(f'(\boldsymbol{A}(\alpha))\,\partial_\alpha\boldsymbol{A}(\alpha)\right).\tag{4.58}$$

尤其，对于正定 $\boldsymbol{A}(\alpha)$

$$\partial_\alpha\operatorname{Tr}(\log(\boldsymbol{A}(\alpha)))=\operatorname{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha))\tag{4.59}$$

and

$$\partial_\alpha\operatorname{Tr}(\mathcal{F}_K(\boldsymbol{A}(\alpha)))=\operatorname{Tr}(v_{\mu_K}(2\boldsymbol{A}(\alpha))\partial_\alpha\boldsymbol{A}(\alpha))-\frac{1}{2}\operatorname{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha)).\tag{4.60}$$

证明。由线性性可得

$$\partial_\alpha\operatorname{Tr}(f(\boldsymbol{A}(\alpha)))=\operatorname{Tr}(\partial_\alpha f(\boldsymbol{A}(\alpha))).\tag{4.61}$$

为处理右侧，我们使用 [Hig08第3 章中的概念。2]。令 $L(\boldsymbol{A},\boldsymbol{C})$表示 ′在方向 $f(\boldsymbol{A}$上的弗雷歇导数，其中方向 $\boldsymbol{C}$ 定义见 [Hig08，式 (3.6)]。于是

$$\partial_\alpha f(\boldsymbol{A}(\alpha))=L(\boldsymbol{A}(\alpha),\partial_\alpha\boldsymbol{A}(\alpha)).\tag{4.62}$$

我们将 $\boldsymbol{A}(\alpha)=\boldsymbol{U}(\alpha)\boldsymbol{D}(\alpha)\boldsymbol{U}(\alpha)^{\mathsf{T}}$ 写成其特征分解，其中 $\boldsymbol{D}(\alpha)=\operatorname{diag}(\lambda_1(\alpha)\ldots,\lambda_n(\alpha))$ 是 $\boldsymbol{A}$的特征值。令 $\odot$ 表示哈达玛积。根据 [Hig08，推论 3.12（另见第 61 页顶部及公式 (3.13) 前的注释）]，我们有

$$L(\boldsymbol{A}(\alpha),\partial_\alpha\boldsymbol{A}(\alpha))=\boldsymbol{U}(\alpha)(\boldsymbol{\Delta}(\alpha)\odot\boldsymbol{U}(\alpha)^{\mathsf{T}}\partial_\alpha\boldsymbol{A}(\alpha)\boldsymbol{U}(\alpha))\boldsymbol{U}(\alpha)^{\mathsf{T}}$$

其中 $\boldsymbol{\Delta}$ 由下式给出

$$\boldsymbol{\Delta}=\boldsymbol{\Delta}_{f(\boldsymbol{A})}=[\Delta f(\lambda_i,\lambda_j)]_{i,j\le n}\quad\text{and}\quad \Delta f(\lambda,\lambda')=\begin{cases}\frac{f(\lambda)-f(\lambda')}{\lambda-\lambda'} & \lambda\neq\lambda'\\ f'(\lambda) & \lambda=\lambda'.\end{cases}\tag{4.63}$$

因此，利用迹在循环置换下的不变性

$$\mathrm{Tr}(\partial_\alpha f(\boldsymbol{A}(\alpha)))=\mathrm{Tr}(\boldsymbol{U}(\boldsymbol{\Delta}\odot\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}\boldsymbol{U})\boldsymbol{U}^\mathsf{T})=\mathrm{Tr}(\boldsymbol{I}(\boldsymbol{\Delta}\odot\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}\boldsymbol{U}))=\mathrm{Tr}(\boldsymbol{U}(\boldsymbol{I}\odot\boldsymbol{\Delta})\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}),$$

其中，最后一个不等式成立是因为对于任意对称矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$

$$\mathrm{Tr}(\boldsymbol{A}(\boldsymbol{B}\odot\boldsymbol{C}))=\mathrm{Tr}((\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{C})).$$

因此该论断成立，因为 $\boldsymbol{I}\odot\boldsymbol{\Delta}=$对角矩阵$(f'(\lambda_1)\dots,f'(\lambda_n))$ 所以

$$\boldsymbol{U}(\boldsymbol{I}\odot\Delta)\boldsymbol{U}^\mathsf{T}=f'(\boldsymbol{A}).$$

这证明了（4.58）。由（4.58）我们现在推导出（4.59）-（4.60）。回忆一下，标量函数 $\log(x)$ 和 $\mathcal{F}_K(x)$ 在 $(0,\infty)$ 上都是光滑的。为证明第一个公式，我们有$\frac{d}{dx}\log(x)=\frac{1}{x}$，因此（4.58）意味着

$$\partial_\alpha\mathrm{Tr}(\log(\boldsymbol{A}(\alpha)))=\mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha)),$$

其中，$\boldsymbol{A}^{-1}$ 是将 $=x^{-1}$ 应用于 $\boldsymbol{A}$ 后由 $f(x)$ 产生的主矩阵函数，它与 $\boldsymbol{A}$ 的通常矩阵逆相一致。

根据[BK19，引理 12]或 [GM05，定理 6] 可知

$$\mathcal{F}_K'(\beta)=v_{\mu_K}(2\beta)-\frac{1}{2\beta}\text{ for all }z>0,\tag{4.64}$$

因此（4.58）意味着

$$\partial_\alpha\mathrm{Tr}(\mathcal{F}_K(\boldsymbol{A}(\alpha)))=\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{A}(\alpha))\partial_\alpha\boldsymbol{A}(\alpha))-\frac{1}{2}\mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha)),$$

其中，我们仍可将由主矩阵函数 x^{-1} 产生的 $\boldsymbol{A}^{-1}$ 解释为通常意义下的逆矩阵。□

现在我们研究 $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ 的最大化点，该函数定义于（4.57）中。首先注意到集合

$$\{\boldsymbol{m}:\boldsymbol{m}\boldsymbol{m}^\mathsf{T}<\boldsymbol{Q}\}\tag{4.65}$$

是一个开集。此外，由于 $F^K{}_{\mathrm{TAP}}(\boldsymbol{m})$ 在 $|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T}|\to 0$ 时发散至 $-\infty$，全局最大值位于（4.65）中。我们将矩阵 $\boldsymbol{m}=(m_1^1,\dots,m_1^n,\dots,m_N^1,\dots,m_N^n)\in\mathbb{R}^{Nn}$向量化，并将$\tilde{F}^K{}_{\mathrm{TAP}}$视为从 $\mathbb{R}^{Nn}\mapsto\mathbb{R}$ 到的函数。通过这一向量化，$\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ 的梯度是 $\mathbb{R}^{Nn}$ 中的一个向量，其海森矩阵是一个 $Nn\times Nn$ 分块矩阵，由 $N\times N$ 个大小为$n\times n$的分块组成。也就是说，对于任意足够光滑的函数 $f\colon\mathbb{R}^{Nn}\to\mathbb{R}$,

$$\nabla f(\boldsymbol{m})=\left(\partial_{m_1^1}f(\boldsymbol{m}),\cdots,\partial_{m_1^n}f(\boldsymbol{m}),\cdots,\partial_{m_N^1}f(\boldsymbol{m}),\cdots,\partial_{m_N^n}f(\boldsymbol{m})\right)^\mathsf{T}\in\mathbb{R}^{Nn}$$

and

$$\nabla^2 f(\boldsymbol{m})=\begin{bmatrix}\boldsymbol{f}_{1,1}&\cdots&\boldsymbol{f}_{1,N}\\ \vdots&\ddots&\vdots\\ \boldsymbol{f}_{N,1}&\cdots&\boldsymbol{f}_{N,N}\end{bmatrix}\in\mathbb{R}^{Nn\times Nn},\ \boldsymbol{f}_{i,j}=\begin{bmatrix}\partial_{m_i^1}\partial_{m_j^1}f(\boldsymbol{m})&\cdots&\partial_{m_i^1}\partial_{m_j^n}f(\boldsymbol{m})\\ \vdots&\ddots&\vdots\\ \partial_{m_i^n}\partial_{m_j^1}f(\boldsymbol{m})&\cdots&\partial_{m_i^n}\partial_{m_j^n}f(\boldsymbol{m})\end{bmatrix}\in\mathbb{R}^{n\times n}.$$

由于 $\tilde{F}^K_{\mathrm{TAP}}$ 是光滑的，其局部最大化点 $\boldsymbol{m}^*$ 满足

$$\nabla\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}^*)=0\quad\text{and}\quad\nabla^2\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}^*)\le 0.$$

备注4.14.由于我们严格来说只证明了 $\mathcal{F}_K$ 在 $(0,\infty)$ 上光滑，而没有证明其在 $[0,\infty)$ 上光滑，因此严格地说，我们只能声称 $\tilde{F}^K{}_{\mathrm{TAP}}$ 中涉及 $\mathcal{F}_K$的项（因而 $\tilde{F}^K_{\mathrm{TAP}}$ 本身）在 β 的所有条目均为正数时是光滑的，从而 $\boldsymbol{\beta}^{1/2}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{1/2}$ 是正定的。在下面的证明中，我们假设 β 的条目为正数（这也简化了论证），随后通过逼近去除这一假设。再多做一些工作，可以证明 $\mathcal{F}_K$ 实际上在 $[0,\infty)$上是光滑的，并将下面的所有论证扩展到涵盖具有零分量的 β，但我们不作此处理。

该部分

$$f(\boldsymbol{m})=\sum_{\ell=1}^{n}\beta_\ell\tilde{H}_N(m^\ell),$$

依赖于哈密顿量的 $\tilde{F}^K{}_{\mathrm{TAP}}(\boldsymbol{m})$具有一个简单的海森矩阵，由 $Nn\times Nn$矩阵给出

$$\nabla^2 f(\boldsymbol{m})=2N\mathrm{diag}(\theta_1\boldsymbol{\beta},\theta_{\frac{N-1}{N}}\boldsymbol{\beta},\ldots,\theta_{\frac{1}{N}}\boldsymbol{\beta})=2N\begin{bmatrix}\theta_1\boldsymbol{\beta} & \cdots & \mathbf{0}_n\\ \vdots & \ddots & \vdots\\ \mathbf{0}_n & \cdots & \theta_{\frac{1}{N}}\boldsymbol{\beta}\end{bmatrix}\tag{4.66}$$

(回顾 (4.6)-(4.8) 和 (4.57)).

另一部分 of $\tilde{F}^K{}_{\mathrm{TAP}}(\boldsymbol{m})$ 是 $Ng(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})$ 的

$$g(\boldsymbol{A})=\mathrm{Tr}(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{A})\boldsymbol{\beta}^{\frac{1}{2}}))+\frac{1}{2}\mathrm{Tr}(\log(\boldsymbol{Q}-\boldsymbol{A})).$$

其海森矩阵由下一个引理给出。

引理 4.15。假设 $\beta_k>0$ 对于 $k=1$、.....、 n。那么对于 $\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}>0$，海森矩阵$\nabla^2 g(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})$是如下的 $Nm\times Nm$ 矩阵

$$-2\begin{bmatrix}\boldsymbol{\beta}^{\frac{1}{2}}v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} & \cdots & \mathbf{0}_n\\ \vdots & \ddots & \vdots\\ \mathbf{0}_n & \cdots & \boldsymbol{\beta}^{\frac{1}{2}}v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\end{bmatrix}+\boldsymbol{L},\tag{4.67}$$

其中

$$\boldsymbol{Q}_{\boldsymbol{m}}:=\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}},\tag{4.68}$$

并且 $\boldsymbol{L}$ 是一个秩至多为 n^4的矩阵。

证明。我们有

$$\frac{1}{2}\mathrm{Tr}(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}))=\frac{1}{2}\Big(\mathrm{Tr}(\log(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}))-\mathrm{Tr}(\log(\boldsymbol{\beta}))\Big).$$

根据引理4.13,

$$\partial_{m_i^\ell}\frac{1}{2}\mathrm{Tr}\Big(\log(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}\Big)=-\frac{1}{2}\mathrm{Tr}(\boldsymbol{Q}_{\boldsymbol{m}}^{-1}\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\boldsymbol{\beta}^{\frac{1}{2}}),$$

and

$$\partial_{m_i^\ell}\mathrm{Tr}\Big(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}})\Big)=-\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\boldsymbol{\beta}^{\frac{1}{2}})+\frac{1}{2}\mathrm{Tr}(\boldsymbol{Q}_{\boldsymbol{m}}^{-1}\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\boldsymbol{\beta}^{\frac{1}{2}}).$$

因此， $g(\boldsymbol{m})$ 的一阶导数等于

$$\partial_{m_i^\ell}g(\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})=-\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\boldsymbol{\beta}^{\frac{1}{2}}).$$

为求得二阶导数，令 $h:\ \mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}$ 由下式给出

$$h(\boldsymbol{A})=v_{\mu_K}(2\boldsymbol{\beta}^{1/2}(\boldsymbol{Q}-\boldsymbol{A})\boldsymbol{\beta}^{1/2}).$$

根据乘积法则和链式法则

(4.69)

$$\begin{aligned}\partial_{m_j^{\ell'}}\mathrm{Tr}(h(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}}) &= \textstyle\sum_{ab}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\partial_{m_j^{\ell'}}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}})_{ab})+\\ &\quad\textstyle\sum_{abcd}\partial_{\boldsymbol{A}_{cd}}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\partial_{m_j^{\ell'}}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_{cd}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}})_{ab}).\end{aligned}$$

因此

(4.70)
$$\nabla^2 g(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})=\boldsymbol{W}+\boldsymbol{L},$$

其中

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))}=-\sum_{ab}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\partial_{m_j^{\ell'}}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}})_{ab}),$$

and

(4.71)
$$\boldsymbol{L}_{((i,\ell),(j,\ell'))}=-\sum_{abcd}\partial_{\boldsymbol{A}_{cd}}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\partial_{m_j^{\ell'}}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_{cd}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}})_{ab}).$$

我们有

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))}=-\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\partial_{m_j^{\ell'}}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}}).$$

Also

$$\partial_{m_j^{\ell'}}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}=\begin{bmatrix}0&\cdots&m_j^1&\cdots&0\\ \vdots&\ddots&\vdots&\ddots&\vdots\\ m_j^1&\cdots&2m_j^{\ell'}&\cdots&m_j^n\\ \vdots&\ddots&\vdots&\ddots&\vdots\\ 0&\cdots&m_j^n&\cdots&0\end{bmatrix}$$

其中只有第 ℓ'行和列非零，且

(4.72)
$$\partial_{m_i^\ell}\partial_{m_j^{\ell'}}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}=(\delta_{i=j}(\delta_{(\ell,\ell')=(a,b)}+\delta_{(\ell,\ell')=(b,a)}))_{a,b\le n}$$

如果 $i\neq j$，则该矩阵为零矩阵；而如果 $i=j$，则除条目(ℓ,ℓ')和(ℓ',ℓ) 外，该矩阵为零矩阵：当 $\ell\neq\ell'$（在非对角线上）时取值为 1，当 $\ell=\ell'$（在对角线上）时取值为 2。利用这一点以及矩阵 $v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})$ 的对称性，我们得到

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))}=-\delta_{i=j}2v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})_{\ell,\ell'}\beta_\ell^{1/2}\beta_{\ell'}^{1/2}=-2\delta_{i=j}(\boldsymbol{\beta}^{\frac{1}{2}}v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}})_{\ell,\ell'}.$$

这给出了4.67 中的第一项。(　　)

至于 $\boldsymbol{L}$，我们可以将其各项写成

$$\boldsymbol{L}_{((i,\ell),(j,\ell'))}=-\left(\sum_{a,b,c,d=1}^{n}d_{a,b,c,d}v^{c,d}\left(w^{a,b}\right)^\mathsf{T}\right)_{((i,\ell),(j,\ell'))}$$

其中

$$d=\partial_{\boldsymbol{A}_{cd}}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T}),$$

且 $v^{c,d}\in\mathbb{R}^{Nn}$ 由下式给出

$$v_{j,\ell'}^{cd}=\left(\partial_{m_j^{\ell'}}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\right)_{cd},$$

和 $w^{a,b}\in\mathbb{R}^{Nn}$ 乘以

$$w_{i,\ell}^{ab}=(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\mathsf{T}\boldsymbol{\beta}^{\frac{1}{2}})_{ab},$$

使得 $v^{c,d}\left(w^{a,b}\right)^{\intercal}$ 是一个 $Nn\times Nn$ 矩阵。因此 $\boldsymbol{L}$ 是至多 n^4 个秩至多为 1 的项之和，所以其秩至多为 n^4。□

命题4.1 的证明余下部分涉及下述稍强的 Plefka条件:

$$\mathrm{Plef}_N^{\delta}(\boldsymbol{Q},\beta) \;=\; \left\{\boldsymbol{m}\in\mathbb{R}^{n\times N}:\boldsymbol{m}\boldsymbol{m}^{\intercal}<\boldsymbol{Q},\|\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\intercal})\boldsymbol{\beta}^{\frac{1}{2}}\|_2\le\tfrac{1}{\sqrt{2}}-\delta\right\}.$$

注意

$$\mathrm{Plef}_N^{\delta}(\boldsymbol{Q},\beta)\subset\mathrm{Plef}_N^{0}(\boldsymbol{Q},\beta)=\mathrm{Plef}_N(\boldsymbol{Q},\beta)\text{ for all }\delta\ge 0,\boldsymbol{Q},\beta. \tag{4.73}$$

这个更强的Plefka条件是一种技术手段，使我们可以在下文命题 β4.1 的证明中，利用连续性，从仅含非零元素的 β 的上界推出所有 β的上界。下一个引理是对 [BK19引理 13], 的略微加强，下面将用它证明，任何 $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$的最大化点都必须满足更强的Plefka条件。

引理 4.16. 对所有 $K\ge 2$，存在一个 $\varepsilon\in(0,\frac{2\sqrt{2}}{K})$ 和一个 $\delta_K>0$，使得

$$v_{\mu_K}(\beta)\ge\sqrt{2}-\varepsilon\;\Rightarrow\;\beta\le\frac{1}{\sqrt{2}}-\delta_K$$

证明. 我们可以令 $\varepsilon=\sqrt{2}-v_{\mu_K}(\frac{1}{\sqrt{2}}-\delta_K)$，因为

$$v_{\mu_K}(\beta)\ge v_{\mu_K}\left(\tfrac{1}{\sqrt{2}}-\delta_K\right)\Rightarrow\beta\le\tfrac{1}{\sqrt{2}}-\delta_K,$$

and

$$x_K<v_{\mu_K}\left(\tfrac{1}{\sqrt{2}}-\delta_K\right)<\sqrt{2}, \tag{4.74}$$

其中第二个不等式对某个充分 $\delta_K>0$小的值成立，因为 $x_K<v_{\mu_K}(\frac{1}{\sqrt{2}})<\sqrt{2}$由 [BK19，引理 13] 且 v_{μ_K} 是连续的。□

现在我们证明， $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ 的所有最大化点都必须满足更强的 Plefka条件。

引理4.17（临界点条件）。假设 $\beta_k>0$对于 $k=1,\ldots\ldots,n$。令 $K\ge 1$。存在常数 $c(K)$，使得如果 $N\ge c(K)$，则

$$\boldsymbol{m}\boldsymbol{m}^{\intercal}<\boldsymbol{Q}\ and\ \nabla^2\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})\le\boldsymbol{0}\implies\boldsymbol{m}\in\mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q},\beta). \tag{4.75}$$

对于 δ_K 如引理 *4.16*。

证明。 由（4.66）和引理 4.15可得

$$\nabla^2\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})=N\left(\boldsymbol{A}+\boldsymbol{L}\right)\le\boldsymbol{0}, \tag{4.76}$$

for

$$\boldsymbol{A}=2\begin{bmatrix}\theta_1\boldsymbol{\beta}&\cdots&\boldsymbol{0}_n\\ \vdots&\ddots&\vdots\\ \boldsymbol{0}_n&\cdots&\theta_{\frac{1}{N}}\boldsymbol{\beta}\end{bmatrix}-2\begin{bmatrix}\boldsymbol{\beta}^{\frac{1}{2}}v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}&\cdots&\boldsymbol{0}_n\\ \vdots&\ddots&\vdots\\ \boldsymbol{0}_n&\cdots&\boldsymbol{\beta}^{\frac{1}{2}}v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\end{bmatrix},$$

且 $\boldsymbol{L}$的秩至多为 n^4(回顾 $\boldsymbol{Q}_{\boldsymbol{m}}:=\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^{\intercal})\boldsymbol{\beta}^{\frac{1}{2}}$)。

由于 $\boldsymbol{A}$ 是分块对角矩阵，其特征值就是各个分块的特征值。根据 Weyl 不等式[HJ13，定理 4.3。1] 除 n^4+1 之外，矩阵 $\boldsymbol{A}$ 的所有特征值都不大于整个海森矩阵的最大特征值 $\nabla^2\tilde{F}^K_{\mathrm{TAP}}$。这意味着，在最后 $n^4+2\le 2n^4$ 个分块中，至少有一个分块的所有特征值都不大于 $\nabla^2\tilde{F}^K_{\mathrm{TAP}}$的最大特征值。

因此，如果 $\nabla^2 \tilde{F}^K_{\mathrm{TAP}} \le \mathbf{0}$，那么

$$2\theta_{1-\frac{2n^4}{N}}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} \le \mathbf{0}. \tag{4.77}$$

若 $\boldsymbol{\beta}^{1/2}\boldsymbol{B}\boldsymbol{\beta}^{1/2} \le \mathbf{0}$ 对于矩阵 $\boldsymbol{B}$ 则 $\boldsymbol{B} \le \mathbf{0}$，因为我们假设 $\boldsymbol{\beta}$ 是对角矩阵，且对角线上的元素为正。因此（4.77）意味着

$$\theta_{1-\frac{2n^4}{N}} \le v(\tilde{\beta}_i(\boldsymbol{m})) \qquad \forall i \le n.$$

引理中 $v_{\mu_K}(\cdot)$ 的性质 4.16 意味着存在一个 $C(K)$ 和 $\delta_K > 0$ 使得对所有 $\varepsilon \le C(K)$,

$$v_{\mu_K}(2\beta) \ge \sqrt{2} - \varepsilon \implies \beta \le \frac{1}{\sqrt{2}} - \delta_K.$$

由 $\theta_{1-\frac{2n^4}{N}} = \sqrt{2} + o_N(1)$可知，对于充分大的 N（其大小取决于 K），

$$\tilde{\beta}_i(\boldsymbol{m}) \le \frac{1}{\sqrt{2}} - \delta_K \quad \forall i \le n$$

˜（回顾一下，$\beta_i(\boldsymbol{m})$ 是 $\boldsymbol{Q}_m$ 的特征值，定义见（4.68））， 这意味着 $\boldsymbol{m} \in \mathrm{Plef}_{K,\,N}(\boldsymbol{Q},\beta) \subseteq \mathrm{Plef}_N(\boldsymbol{Q},\beta)$。

□

最后，我们给出命题 4.1的证明。

命题的证明 *4.1*。 我们首先假设 $\beta \in (0,\infty)^n$。由 (4.3)，只需研究确定性对角化哈密顿量 ˜ 的自由能 $H_N(\boldsymbol{\sigma})$。从上界命题 4.10 出发可知，对于任意 $K \ge 2$ 和 $0 < \varepsilon \le C(\beta, \boldsymbol{h}, \boldsymbol{Q}, K)$ 以及 $N \ge c(\varepsilon, K)$，有

$$\tilde{F}^{\varepsilon}_N(\beta, \boldsymbol{h}, \boldsymbol{Q}) \le \frac{1}{N} \sup_{\boldsymbol{m}:\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} < \boldsymbol{Q}} \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) + \frac{c(\beta)}{K}.$$

回顾下文的 (4.65)，可知 $\tilde{F}^K_{\mathrm{TAP}}$ 在集合 (4.65) 中存在一个全局最大化点。该最大化点必须满足

$$\nabla^2 \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) \le 0.$$

因此，由引理 4.17 可得

$$\tilde{F}^{\varepsilon}_N(\beta, \boldsymbol{h}, \boldsymbol{Q}) \le \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}^{\delta_K}_N(\boldsymbol{Q},\beta)} \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) + \frac{c(\beta)}{K}. \tag{4.78}$$

根据定义 (1.5)、(4.56) 中的等价关系以及引理 $\mathcal{F}_K$ 给出的 4.6 中 $m \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$

$$\left| \mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}) - \frac{1}{2}\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta \right| \le o_K(1),$$

其中，项 $o_K(1)$ 与任何参数无关，并且当 $K \to \infty$ 时趋于零。这使我们可以在上界中将 ($\mathcal{F}_K$ 的 $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ 中的 4.78 替换为 $\tilde{F}_{\mathrm{TAP}}$ 的 Onsager 修正项（见 (4.5)），因此由 (4.78) 可得

$$\tilde{F}^{\varepsilon}_N(\beta, \boldsymbol{h}, \boldsymbol{Q}) \le \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}^{\delta_K}_N(\boldsymbol{Q},\beta)} \tilde{F}_{\mathrm{TAP}}(\boldsymbol{m}) + o_K(1), \tag{4.79}$$

其中项 $o_K(1)$ 依赖于 β，并且当 $K \to \infty$ 对于固定的 β 时趋于零。此上界对所有 $0 < \varepsilon < C(\beta, \boldsymbol{h}, \boldsymbol{Q})$、所有 $K \ge 2$ 以及所有 $N \ge c(\varepsilon, K)$ 均成立。利用 (2.4)，我们可以界定归一化哈密顿量 $\frac{1}{N}H_N$ 与其对角化后的确定性形式之间的差

对应项 $\frac{1}{N}\tilde{H}_N$ 由任意 $\eta > 0$ 以概率趋于 1。因此，我们对 F_{TAP} 得到（回顾 (1.4)）

$$F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta)} F_{\mathrm{TAP}}(\boldsymbol{m}) + o_K(1) \tag{4.80}$$

以概率趋于 1。利用（4.73）并选取足够大的 K（其大小取决于 η），即可得到（4.1）。因此，只要 $\beta \in (0, \infty)^n$，我们便证明了（4.1）。

为处理 β 中趋于零的分量，注意若 $\beta_1, \beta_2 \in [0, \infty)^n$ 则由（1.3）和（2.1）可得

$$\left|F_N^{\varepsilon}(\beta_1, \boldsymbol{h}, \boldsymbol{Q}) - F_N^{\varepsilon}(\beta_2, \boldsymbol{h}, \boldsymbol{Q})\right| \leq c\left|\beta_1 - \beta_2\right|, \tag{4.81}$$

以趋于 1 的概率成立。记 $F_{\mathrm{TAP}}(\boldsymbol{m}; \beta)$ 为 F_{TAP}，并显式标明其对 β 的依赖（回顾（1.4））。类似地，我们有

$$\left|F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_1) - F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_2)\right| \leq c\left|\beta_1 - \beta_2\right|, \tag{4.82}$$

对于有界的 β_1, β_2，并利用

$$\boldsymbol{m} \to \beta^{\mathsf{T}}\left(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\right)^{\odot 2}\beta \text{ is Lipschitz on compact subsets of } \mathbb{R}^n \tag{4.83}$$

对 $\|\boldsymbol{Q}\|_\infty \leq 1$ 和 $\boldsymbol{m}$ 一致成立，其中 $\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} \leq \boldsymbol{Q}$。再次利用（4.83），可知存在一个足够小的常数 $\rho(K)$，它仅依赖于 K

$$\mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta_2) \subset \mathrm{Plef}_N^{0}(\boldsymbol{Q}, \beta_1) = \mathrm{Plef}_N(\boldsymbol{Q}, \beta_1) \text{ for } \left|\beta_1 - \beta_2\right| \leq \rho(K), \tag{4.84}$$

对于有界的 β_1、β_2。

因此，对于任意含有零元素的 β_1 和 $\eta > 0$，我们可以选取足够大的 K（其大小取决于β_1 和 η），以及所有元素均为正且足够接近 β_1的 β_2，使得

$$F_N^{\varepsilon}(\beta_1, \boldsymbol{h}, \boldsymbol{Q}) \overset{(4.81)}{\leq} F_N^{\varepsilon}(\beta_2, \boldsymbol{h}, \boldsymbol{Q}) + \tfrac{\eta}{3} \overset{(4.80)}{\leq} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta_2)} F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_2) + \tfrac{2\eta}{3} \overset{(4.82),(4.84)}{\leq} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta_1)} F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_1) + \eta$$

w以趋于 1的概率。这证明了4.1对于 β 其元素趋于零。

□

结合TAP下界命题 3.1 和TAP上界命题 4.1完成了定理 1.1的证明。

5. Ground State Energy

剩下的就是证明定理 1.2中的基态公式。为避免矩阵可逆性方面的技术问题，我们先假设 β 和 $\boldsymbol{h}$ 均不为零，然后利用连续性将结论推广至所有 β 及外场。

根据一致界（4.3），我们可以用其对角化基表示哈密顿量和外场，因此只需计算以下极限

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) := \sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} \left(\frac{1}{N}\sum_{k=1}^{n} \beta_k \tilde{H}_N(m^k) + \sum_{k=1}^{n} m^k \cdot \tilde{h}^k\right) \tag{5.1}$$

其中 $\tilde{H}_N$ 是 H_N 在（4.2）中定义的确定性对应物，并且 $\tilde{h}^k$ 是向量 h^k 在无序矩阵 J的对角化基下的表示，与上一节相同。我们定义如下的基态泛函变分形式

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{\Lambda}-\sqrt{2}\boldsymbol{I}\geq 0} \left(\frac{1}{4} h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big) \boldsymbol{\beta}^{-1/2} h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}) \right). \tag{5.2}$$

我们将 n ~ ~ ˜ow证明$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \boldsymbol{\beta}, \boldsymbol{h})$ 依概率收敛于高斯辛正交系综$(\beta, h, \boldsymbol{Q})$.

命题 5.1. 对于 $\beta_1, \ldots, \beta_n, \ldots, h_1, \ldots, h_n \neq 0$，我们有

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} \left(\frac{1}{N} \sum_{k=1}^{n} \beta_k \tilde{H}_N(m^k) + \sum_{k=1}^{n} m^k \cdot \tilde{h}^k \right) \xrightarrow{\mathbb{P}} \widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}). \tag{5.3}$$

证明。我们使用拉格朗日乘子显式求解这个约束最大化问题。考虑拉格朗日函数。

$$\begin{aligned} f(\boldsymbol{m}, \boldsymbol{\Lambda}) &= \sum_{k=1}^{n} \beta_k \sum_{i=1}^{N} \theta_{i/N} (m_i^k)^2 + \sum_{k=1}^{n} \tilde{h}^k \cdot m^k + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{\frac{1}{2}}(\tilde{\boldsymbol{Q}} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}) \\ &= \sum_{i=1}^{N} \left((\boldsymbol{\beta}^{\frac{1}{2}} m_i)^{\mathsf{T}} (\theta_{i/N}\boldsymbol{I} - \boldsymbol{\Lambda})(\boldsymbol{\beta}^{\frac{1}{2}} m_i) + (\boldsymbol{\beta}^{-\frac{1}{2}} \tilde{h}_i) \cdot \boldsymbol{\beta}^{\frac{1}{2}} m_i \right) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{\frac{1}{2}}). \end{aligned}$$

注意，

$$\partial_{\Lambda_{ij}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = (1 + \delta_{i\neq j}) \sqrt{\beta_i \beta_j} \left(\tilde{\boldsymbol{Q}} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} \right)_{ij}.$$

因此，如果 $\boldsymbol{\Lambda}^* > \sqrt{2}\boldsymbol{I}$ 和 $\boldsymbol{m}^*$ 是 $f(\boldsymbol{m}, \boldsymbol{\Lambda})$ 的临界点，那么

$$f(\boldsymbol{m}^*, \boldsymbol{\Lambda}^*) \leq \sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}).$$

Also

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) \leq \inf_{\Lambda>\sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) \leq \inf_{\Lambda>\sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}).$$

˜因此，由于对于固定的 N、h，f 是可微的，如果右端存在有限优化器，使得 $\boldsymbol{\Lambda}^* > \sqrt{2}\boldsymbol{I}$，那么

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) = \inf_{\Lambda>\sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}). \tag{5.4}$$

对于固定的 $\boldsymbol{\Lambda}$，考虑 $\sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda})$。我们有

$$\partial_{m_i} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = 2(\theta_{i/N}\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{\beta}^{1/2} m_i + \boldsymbol{\beta}^{-1/2}\tilde{h}_i.$$

如果 $\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}$，那么 $\boldsymbol{m} \to f(\boldsymbol{m}, \boldsymbol{\Lambda})$ 的唯一临界点因此为

$$m_i(\boldsymbol{\Lambda}) = \frac{1}{2}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda} - \theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i \tag{5.5}$$

a并且由于凹性，该临界点对应于一个局部最大化点，因此

$$\sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = \sum_{i=1}^{N} \left(\frac{1}{4} \tilde{h}_i^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} (\boldsymbol{\Lambda} - \theta_{i/N}\boldsymbol{I})^{-1} \boldsymbol{\beta}^{-1/2} \tilde{h}_i \right) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}). \tag{5.6}$$

注意，如果 $\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2} > 0$，那么由于 $\left(\boldsymbol{U}^{\mathsf{T}}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}\boldsymbol{U}\right)_{ii} = u_i^T\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}u_i > 0$，对于所有正交的 $\boldsymbol{U}$，我们有

$$\sup_{\boldsymbol{m}} f(\boldsymbol{m},\boldsymbol{\Lambda}) \geq \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}) \to \infty \text{ if } \boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I},\ \sup_k \lambda_k(\boldsymbol{\Lambda}) \to \infty.$$

此外，如果 $\boldsymbol{\Lambda} \to \sqrt{2}\boldsymbol{I}$ 且 $\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}$，则几乎必然

$$\begin{aligned}\sup_{\boldsymbol{m}} f(\boldsymbol{m},\boldsymbol{\Lambda}) &\geq \tfrac{1}{4}\tilde{h}_N^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_N \\ &\geq \tfrac{1}{4}\frac{\left|\tilde{h}_N\right|^2}{\lambda_{\max}\left(\boldsymbol{\beta}^{1/2}(\boldsymbol{\Lambda}-\sqrt{2}\boldsymbol{I})\boldsymbol{\beta}^{1/2}\right)} \\ &\to \infty,\end{aligned}$$

由于 $\tilde{h}_N \neq 0$ 几乎处处成立且 $\lambda_{\max}\left(\boldsymbol{\beta}^{1/2}(\boldsymbol{\Lambda}-\sqrt{2}\boldsymbol{I})\boldsymbol{\beta}^{1/2}\right) \to 0$。这表明，极小化器属于

$$\inf_{\boldsymbol{\Lambda}>\sqrt{2}\boldsymbol{I}} f(\boldsymbol{m}(\boldsymbol{\Lambda}),\boldsymbol{\Lambda}),$$

可达到在 $\{\boldsymbol{\Lambda}:\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{\beta}\}$ 中的一点处取得，因此存在一个优化器

$$\inf_{\boldsymbol{\Lambda}>\sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m},\boldsymbol{\Lambda}),$$

其为临界点： f，从而（5.4）成立。

现在我们证明， $f(\boldsymbol{m}(\boldsymbol{\Lambda})\ \boldsymbol{\Lambda})$ 收敛于 $\boldsymbol{\Lambda}$ 的极限函数，使得

$$\inf_{\boldsymbol{\Lambda}>\sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m},\boldsymbol{\Lambda}) \to \inf_{\boldsymbol{\Lambda}>\sqrt{2}\boldsymbol{I}} \left(\frac{1}{4}h^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}\left(\boldsymbol{\Lambda}-\sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}\right)\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2})\right).$$

根据命题 5.3 和（5.4），该收敛在 $\boldsymbol{\beta}$ 和 $\boldsymbol{h}$ 的紧子集上是均匀分布的，且极限是 $\sqrt{2}$-Lipschitz，因为左侧也是如此。

~~~回顾（5.6）。注意，由于 $\mathbb{E}[h_{i,k}h_{i,l}] = h_ih_l$ 有

$$\mathbb{E}\left[\frac{1}{4}\tilde{h}_i^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i\right] = \frac{1}{4}h^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}h,$$

~~~以及 h_i 彼此独立，因此根据大数定律

$$\begin{aligned}f(\boldsymbol{\Lambda},\boldsymbol{m}) &= \sum_{i=1}^{N}\left(\frac{1}{4}\tilde{h}_i^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i\right) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}) \\ &\to \frac{1}{4}h^{\mathsf{T}}\boldsymbol{\beta}^{-1/2}\left(\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\,d\mu_{\mathrm{sc}}(x)\right)\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}),\end{aligned}$$

依概率。对于 $\boldsymbol{\Lambda}$使得 $\lambda_{min}(\boldsymbol{\Lambda}) > \sqrt{2}$，我们可以显式地计算该积分。令$\boldsymbol{\Lambda} = \boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^{\mathsf{T}}$。我们看到

$$\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx = \int_{-\sqrt{2}}^{\sqrt{2}}\boldsymbol{U}(\boldsymbol{D}_\lambda-x\boldsymbol{I})^{-1}\boldsymbol{U}^{\mathsf{T}}\mu_{\mathrm{sc}}(x)\,dx = \boldsymbol{U}\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{D}_\lambda-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx\boldsymbol{U}^{\mathsf{T}}$$

而内部的积分很容易计算。事实上，使用一维情形的公式，我们可以看到。

$$\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx = \boldsymbol{U}\boldsymbol{D}_{\lambda-\sqrt{\lambda^2-4}}\boldsymbol{U}^{\mathsf{T}} = \boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^{\mathsf{T}} - \boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-4}}\boldsymbol{U}^{\mathsf{T}} = \boldsymbol{\Lambda} - \boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-4}}\boldsymbol{U}^{\mathsf{T}}.$$

由于

$$\boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-2}}\boldsymbol{U}^{\mathsf{T}} = \sqrt{\boldsymbol{U}\boldsymbol{D}_{\lambda^2-2}\boldsymbol{U}^{\mathsf{T}}} = \sqrt{\boldsymbol{U}\boldsymbol{D}_{\lambda^2}\boldsymbol{U}^{\mathsf{T}}-2\boldsymbol{I}} = \sqrt{(\boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^{\mathsf{T}})^2-2\boldsymbol{I}} = \sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}$$

我们有

$$\int_{-\sqrt{2}}^{\sqrt{2}} (\boldsymbol{\Lambda} - x\boldsymbol{I})^{-1} \mu_{\mathrm{sc}}(x)\, dx = \boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}.$$

由此公式可得

$$\sup_{\boldsymbol{m}} f(\boldsymbol{\Lambda}, \boldsymbol{m}) \to \frac{1}{4} h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big) \boldsymbol{\beta}^{-1/2} h + \mathrm{Tr}(\boldsymbol{\Lambda} \boldsymbol{\beta}^{1/2} \tilde{\boldsymbol{Q}} \boldsymbol{\beta}^{1/2})$$

并且该临界点对应于一个最大值。注意，对于 $\boldsymbol{\Lambda} \geq \sqrt{2}\boldsymbol{I}$，该公式有良好定义。由（5.4）可知，（5.1）的最大值在此处取得

$$\inf_{\boldsymbol{\Lambda} \geq \sqrt{2}\boldsymbol{I}} \left(\frac{1}{4} h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big) \boldsymbol{\beta}^{-1/2} h + \mathrm{Tr}(\boldsymbol{\Lambda} \boldsymbol{\beta}^{1/2} \tilde{\boldsymbol{Q}} \boldsymbol{\beta}^{1/2}) \right) \tag{5.7}$$

在极限下。 □

~现在，我们显式求解 GSE$(\beta, h, \boldsymbol{Q})$ 中的优化问题，从而得到（1.7）中的闭式表达式。

5.1. 一维情形。我们首先讨论 $n = 1$ 作为热身。在这种情形下求解变分问题要容易得多，因为我们不必担心矩阵的不可交换性。当 $n = 1$ 变分问题是

$$\inf_{\lambda > \sqrt{2}} \left(\frac{1}{4} \frac{h^2}{\beta} \big(\lambda - \sqrt{\lambda^2 - 2}\big) + \lambda \beta \tilde{q} \right). \tag{5.8}$$

令 $A = \frac{1}{4}\frac{h^2}{\beta}$ 且 $B = \beta\tilde{q}$。作变量替换 $\lambda = \frac{1}{\sqrt{2}}\left(x + \frac{1}{x}\right)$，$x \in (0, 1]$，并利用 $\sqrt{\lambda^2 - 2} = \frac{1}{\sqrt{2}}\left(\frac{1}{x} - x\right)$ 可得（5.8）等于

$$\inf_{x \in (0,1]} \left(A\sqrt{2}x + B\frac{1}{\sqrt{2}} \left(\frac{1}{x} + x \right) \right) = \sqrt{2B\,(2A + B)} = \sqrt{\tilde{q}h^2 + 2\beta^2\tilde{q}^2}$$

w这证明了（1.11），事实上这就是公式 [BK19，（1.6）和引理20].

5.2. n 维情形。这一变量替换的矩阵版本使我们能够~也求解情形 $n > 1$，从而由（β，h，$\boldsymbol{Q}$）得到高斯辛正交系综(1.7)。

~命题 5.2. 对于 $\boldsymbol{Q} > 0$ 和 $\boldsymbol{\beta} > 0$，有

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}). \tag{5.9}$$

证明。令

$$\boldsymbol{A} = \frac{1}{4} \boldsymbol{\beta}^{-1/2} h h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \text{ and } \boldsymbol{B} = \boldsymbol{\beta}^{1/2} \tilde{\boldsymbol{Q}} \boldsymbol{\beta}^{1/2}.$$

~将高斯辛正交系综$(\beta, h, \boldsymbol{Q})$的第一项写成 1×1 矩阵的迹，并利用迹的循环性质，我们得到

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{\Lambda} - \sqrt{2}\boldsymbol{I} \geq 0} \left(\frac{1}{4} \mathrm{Tr}\big(\big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big)\boldsymbol{A}\big) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{B}) \right).$$

我们使用变量替换$\boldsymbol{\Lambda} = \frac{1}{\sqrt{2}}(\boldsymbol{X} + \boldsymbol{X}^{-1})$，其中 $\boldsymbol{0} < \boldsymbol{X} \leq \boldsymbol{I}$ 和 $\boldsymbol{X}$ 是对称的（要看出 $\boldsymbol{\Lambda}$ 总能写成这种形式，请回忆 $\boldsymbol{\Lambda}$ 是对称的，并且其特征值大于或等于 $\sqrt{2}$）。由此可得 $(\boldsymbol{\Lambda}^2 - 2\boldsymbol{I})^{\frac{1}{2}} = \frac{1}{\sqrt{2}}(\boldsymbol{X}^{-1} - \boldsymbol{X})$。因此

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{0} < \boldsymbol{X} \leq \boldsymbol{I}} \left(\sqrt{2}\mathrm{Tr}\big(\boldsymbol{X}\boldsymbol{A}\big) + \frac{1}{\sqrt{2}} \mathrm{Tr}((\boldsymbol{X} + \boldsymbol{X}^{-1})\boldsymbol{B}) \right). \tag{5.10}$$

考虑下确界中该量的临界点方程:

$$\partial_{\boldsymbol{X}}\left(\sqrt{2}\mathrm{Tr}\big(\boldsymbol{X}\boldsymbol{A}\big)+\frac{1}{\sqrt{2}}\mathrm{Tr}(\boldsymbol{B}(\boldsymbol{X}+\boldsymbol{X}^{-1}))\right)=\sqrt{2}\boldsymbol{A}+\frac{1}{\sqrt{2}}\boldsymbol{B}-\frac{1}{\sqrt{2}}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}^{-1}=\mathbf{0}.$$

使用变量替换 $\boldsymbol{Y}=\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{X}^{-1}$，这等价于

$$\boldsymbol{Y}^{\intercal}\boldsymbol{Y}=2\boldsymbol{A}+\boldsymbol{B}.$$

如果我们将 $2\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{U}_1\boldsymbol{D}\boldsymbol{U}_1^{\intercal}$对角化，则它进一步等价于 $\boldsymbol{Y}=\boldsymbol{U}_2\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}_{\mathrm{T}1}$对于某个正交的 $\boldsymbol{U}_2$，因此等价于 $\boldsymbol{X}=\boldsymbol{U}_1\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}_{\mathrm{T}_2}\boldsymbol{B}^{\frac{1}{2}}=(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}$，其中 $\boldsymbol{U}=\boldsymbol{U}_{\mathrm{T}1}\boldsymbol{U}_{\mathrm{T}2}$因此，如果 $\boldsymbol{X}$ 是对称的，并且

$$\boldsymbol{X}=(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}, \tag{5.11}$$

对于某个正交 $\boldsymbol{U}$ 它是一个临界点。这样的 $\boldsymbol{U}$ 可以如下找到。对称条件等价于

$$(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}=\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{U}^{\intercal}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\iff\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}=(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{U}^{\intercal}.$$

将 $\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}=\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^{\intercal}$写成其奇异值分解。于是，该条件变为，

$$\boldsymbol{U}\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^{\intercal}=\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{S}^{\intercal}\boldsymbol{U}^{\intercal},$$

因此 $\boldsymbol{U}=\boldsymbol{T}\boldsymbol{S}^{\intercal}$ 是正交的，并使 $\boldsymbol{X}$ 对称。因此我们证明了，这个 $\boldsymbol{X}$ 是（5.10）中下确界表达式的临界点。

现在注意到 $\boldsymbol{X}\mapsto\mathrm{Tr}(\boldsymbol{B}\boldsymbol{X}^{-1})$ 是凸的，因为 $\boldsymbol{X}\mapsto\boldsymbol{X}^{-1}$ 是凸的，并且 $\boldsymbol{X}\mapsto\mathrm{Tr}(\boldsymbol{B}\boldsymbol{X})$ 对正定 $\boldsymbol{B}$是递增的，见 [Bha96，推论 V.2。6]。因此（5.10）中下确界表达式关于 $\boldsymbol{X}$是凸的，因此所示临界点是全局极小值点。

接下来我们计算该极小值点处的值。将其代入（5.10）并利用 $\boldsymbol{X}^{-1}=(\boldsymbol{X}^{-1})^{\intercal}=(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{-\frac{1}{2}}$可得

$$\begin{aligned}
&\widetilde{\mathrm{GSE}}(\beta,h,\tilde{\boldsymbol{Q}})\\
=&\frac{1}{\sqrt{2}}\left(2\mathrm{Tr}\big(\boldsymbol{A}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}\big)+\mathrm{Tr}((2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{B})+\mathrm{Tr}((2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{B})\right)\\
=&\frac{1}{\sqrt{2}}\left(2\mathrm{Tr}\big(\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\big)+\mathrm{Tr}(\boldsymbol{B}^{\frac{3}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U})+\mathrm{Tr}(\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U})\right)\\
=&\frac{1}{\sqrt{2}}\mathrm{Tr}\left(\boldsymbol{B}^{\frac{1}{2}}\left(2\boldsymbol{A}+\boldsymbol{B}+(2\boldsymbol{A}+\boldsymbol{B})\right)\left((2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\right)\right)\\
=&\sqrt{2}\mathrm{Tr}\big(\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\big)\\
=&\sqrt{2}\mathrm{Tr}\big(\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^{\intercal}\boldsymbol{T}\boldsymbol{S}^{\intercal}\big)=\sqrt{2}\mathrm{Tr}(\boldsymbol{\Sigma}).
\end{aligned}$$

注意， $\boldsymbol{\Sigma}$ 是 $\boldsymbol{B}^{1/2}(2\boldsymbol{A}+\boldsymbol{B})^{1/2}$的奇异值对角矩阵，即特征值平方根构成的对角矩阵。

$$\boldsymbol{B}^{1/2}\left(2\boldsymbol{A}+\boldsymbol{B}\right)^{1/2}\left(\boldsymbol{B}^{1/2}\left(2\boldsymbol{A}+\boldsymbol{B}\right)^{1/2}\right)^{T}.$$

反复利用 $\boldsymbol{C}\boldsymbol{D}$ 和 $\boldsymbol{D}\boldsymbol{C}$ 对于方阵 $\boldsymbol{C}$具有相同特征值这一性质，我们得到这些特征值与下式的特征值一致： $\boldsymbol{D}$

$$\left(2\boldsymbol{A}+\boldsymbol{B}\right)\boldsymbol{B}=\left(2\boldsymbol{A}+\boldsymbol{B}\right)\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}.$$

再次利用同一性质，右侧这一项反过来也具有相同的特征值 a s

$$\boldsymbol{\beta}^{1/2}\left(2\boldsymbol{A}+\boldsymbol{B}\right)\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}=\left(\frac{1}{2}hh^{\intercal}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)\tilde{\boldsymbol{Q}},$$

其特征值也与.....相同。

$$\left(\frac{1}{2}hh^{\intercal}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\left(\frac{1}{2}hh^{\intercal}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}.$$

这证明了

$$\operatorname{Tr}\left(\boldsymbol{\Sigma}\right)=\operatorname{Tr}\left(\sqrt{\left(\frac{1}{2}hh^{\intercal}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\left(\frac{1}{2}hh^{\intercal}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac{1}{2}}}\right),$$

并回顾 GSE 的定义（1.7），证明至此完成。 □

因此，对于仅具有非零分量的 β 和 h

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\intercal}=\tilde{\boldsymbol{Q}}}\left(\frac{1}{N}\sum_{k=1}^{n}\beta_k\tilde{H}_N(m^k)+\sum_{k=1}^{n}m^k\cdot\tilde{h}^k\right)\xrightarrow{\mathbb{P}}\mathrm{GSE}(\beta,h,\tilde{\boldsymbol{Q}}) \tag{5.12}$$

˜（结合命题5.1和5.2）。GSE$(\beta,h,\boldsymbol{Q})$ 的公式在 β 或 h 中某些元素为零时也有定义。为将（5.12）推广到这种情形，我们将利用下一引理所提供的连续性论证；该引理表明（5.1）关于 β 和 h是 Lipschitz 连续的。

˜引理 5.3. 若 $\boldsymbol{Q}$ 是正定的，且其元素均不超过 1，则

$$\left|\sup_{\boldsymbol{m}\boldsymbol{m}^{\intercal}=\tilde{\boldsymbol{Q}}}f(\boldsymbol{m},\beta^1,\boldsymbol{h}^1)-\sup_{\boldsymbol{m}\boldsymbol{m}^{\intercal}=\tilde{\boldsymbol{Q}}}f(\boldsymbol{m},\beta^2,\boldsymbol{h}^2)\right|\leq\sqrt{2}\|\beta^1-\beta^2\|_\infty+\|h^1-h^2\|_\infty.$$

证明. 这是因为

$$\begin{aligned}\sup_{\boldsymbol{m}\boldsymbol{m}^{\intercal}=\tilde{\boldsymbol{Q}}}\left|f(\boldsymbol{m},\beta^1,\boldsymbol{h}^1)-f(\boldsymbol{m},\beta^2,\boldsymbol{h}^2)\right|&\leq\sup_{|m^1|,\ldots,|m^n|\leq1}\left|f(\boldsymbol{m},\beta^1,\boldsymbol{h}^1)-f(\boldsymbol{m},\beta^2,\boldsymbol{h}^2)\right|\\&\leq\sqrt{2}\|\beta^1-\beta^2\|_\infty+\sup_{k\leq n}|(h^k)^1-(h^k)^2|\\&\leq\sqrt{2}\|\beta^1-\beta^2\|_\infty+\|h^1-h^2\|_\infty,\end{aligned}$$

因为

$$|(h^k)^1-(h^k)^2|=|h_k^1u-h_k^2u|^2=|h_k^1-h_k^2|^2.$$

□

另由式（1.7）可知

$$\text{for all }\tilde{\boldsymbol{Q}}>0\text{ the map }(h,\beta)\to\mathrm{GSE}(\beta,h,\tilde{\boldsymbol{Q}})\text{ is continuous.} \tag{5.13}$$

定理 1.2 由命题 5.1和 5.2 以及连续性可立即得出。

定理 *1.2* 的证明。 上述归约（5.1）以及命题 5.1 和 5.2 证明了当 β 和 h 的所有元素均非零时，命题（1.8）成立。利用引理5.3 和（5.13）作简单的逼近论证，即可将结论推广至所有 β， h。

□

References

[AJ21] G. B. Arous 和 A. Jagannath。《自旋玻璃中的碎裂与亚稳态》。2021年。[AK18] A. E. Alaoui 和 F. Krzakala。《尖峰维格纳模型中的估计：副本公式的简短证明》。载于：2018年 IEEE 国际信息论研讨会（ISIT）。2018年，第 1874–1878 页。[ASS03] M. Aizenman、R. Sims 和 S. L. Starr。《Sherrington-Kirkpatrick自旋玻璃模型的扩展变分原理》。载于：《物理评论B》68（2003年12月21日），第 214403 页。[AZ22] A. Auffinger 和 Y. Zhou。《球形混合向量 p-自旋模型的性质》。载于：《随机过程及其应用》146（2022年），第 382–413 页。[Bel22] D. Belius。《平均场自旋玻璃自由能的高温TAP上界》。2022年。[Bha96] R. Bhatia。《矩阵分析》。《数学研究生教材》。纽约施普林格，1996年。 ISBN：9780387948461。[BJ18] G. Ben Arous 和 A. Jagannath。《平均场自旋玻璃中的谱隙估计》。载于：《数学物理通讯》361.1（2018年），第 1–52 页。[BK19] D. Belius 和 N. Kistler。《球形SK模型的TAP-Plefka变分原理》。载于：《数学物理通讯》367.3（2019年），第 991–1017 页。[BLM13] S. Boucheron、G. Lugosi 和 P. Massart。《集中不等式：独立性理论的非渐近理论》，Michel Ledoux 作序。牛津大学出版社，牛津，2013年，第 x+481 页。 ISBN：978-0-19-953525-5。[Bol14] E. Bolthausen。《Sherrington–Kirkpatrick模型TAP方程解的迭代构造》。英文。载于：《数学物理通讯》325.1（2014年1月），第 333–366 页。[Bol19] E. Bolthausen。《SK副本对称公式的Morita型证明》。英文。载于：《经典与无序系统的统计力学》。V. Gayrard、L.-P. Arguin、N. Kistler 和 I. Kourkova 编。第 293 卷。丛书：《施普林格数学与统计学会议论文集》。查姆：施普林格国际出版，2019年，第 63–93 页。 ISBN：978-3-030-29077-1。[BY22] C. Brennecke 和 H.-T. Yau。《重新审视SK模型的副本对称公式》。载于：《数学物理学杂志》63.7（2022年），第 073302 页。[Che13] W.-K. Chen。《混合 p-自旋球面模型的Aizenman-Sims-Starr方案与Parisi公式》。载于：《电子概率学杂志》18（2013年），第 94 期，14。[CL04] A. Crisanti 和 L. Leuzzi。《球面 $2+p$ 自旋玻璃模型：一个可精确求解的玻璃态到自旋玻璃态转变模型》。载于：《物理评论快报》93（2004年21），第 217203 页。[CP17] W.-K. Chen 和 D. Panchenko。《某些球形混合 p-自旋模型中的温度混沌》。载于：《统计物理学杂志》166.5（2017年），第 1151–1162 页。[CPS22] W.-K. Chen、D. Panchenko 和 E. Subag。《广义TAP自由能》。英文。载于：《纯粹与应用数学通讯》（2022年1月），cpa.22040。[CS92] A. Crisanti 和 H. .-J. Sommers。《球形 p-自旋相互作用自旋玻璃模型：静力学》。英文。载于：Zeitschrift f ¨ur Physik B Condensed Matter 87.3（1992年），第 341–354 页。[Der80] B. Derrida。《随机能量模型：一族无序模型的极限》。英文。载于：《物理评论快报》45.2（1980年7月），第 79–82 页。

[EYY12] L. Erdos、H.-T. Yau 和 J. Yin。《广义 Wigner 矩阵特征值的刚性》。载于：《数学进展》229.3（2012），第 1435–1515 页。[FPV92] S. Franz、G. Parisi 和 M. Virasoro。《平衡态内外的复制方法》。载于：http://dx.doi.org/10.1051/jp1:1992115 2（1992 年 10 月）。[FR20] S. Franz 和 J. Rocchi。《玻璃态有效势的大偏差》。载于：《物理学杂志 A：数学与理论》53.48（2020），第 485002 页。[GH21] A. Guionnet 和 J. Husson。《k 维球积分的渐近行为及其应用》。2021年。[GM05] A. Guionnet 和 M. Maida。《 R-变换的傅里叶视角及球积分的相关渐近行为》。载于：J. Funct. Anal. 222.2（2005），第 435–490 页。[GM84] D. Gross 和 M. Mezard。《最简单的自旋玻璃》。英文。载于：《核物理 B》240.4（1984 年 11 月），第 431–452 页。[Gue03] F. Guerra。《平均场自旋玻璃模型中复制对称性破缺的界》。载于：《数学物理通讯》233.1（2003），第 1–12 页。[Har56] Harish-Chandra。《半单李代数上的不变微分算子》。载于：《美国国家科学院院刊》42（1956），第 252–253 页。[Hig08] N. J. Higham。《矩阵函数：理论与计算》。工业与应用数学学会（SIAM），费城，宾夕法尼亚州，2008年，第 xx+425 页。ISBN：978- 0-89871-646-7。[HJ13] R. A. Horn 和 C. R. Johnson。《矩阵分析》。第二版。剑桥大学出版社，剑桥，2013年，第 xviii+643 页。ISBN：978-0-521-54823-6。[HK22] J. Husson 和 J. Ko。《次线性秩球积分》。载于：arXiv 预印本 arXiv:2208.03642（2022）。[IZ80] C. Itzykson 和 J. B. Zuber。《平面近似。II》。载于：J. Math. Phys. 21（1980），第 411–421 页。[Jag19] A. Jagannath。《平均场自旋玻璃在短时间尺度与长时间尺度上的动力学》。载于：J. Math. Phys. 60.8（2019），第 083305 页，共 9 页。[JLM20] A. Jagannath、P. Lopatto 和 L. Miolane。《张量主成分分析的统计阈值》。载于：《应用概率年刊》30.4（2020），第 1910–1933 页。[Ko19] J. Ko。《具有矢量自旋的球形自旋玻璃的 Crisanti–Sommers 公式》。载于：arXiv 预印本 arXiv:1911.04355（2019）。[Ko20] J. Ko。《具有受约束重叠的多个球形自旋玻璃系统的自由能》。载于：《电子概率学杂志》25（2020），34 页。[KTJ76] J. M. Kosterlitz、D. J. Thouless 和 R. C. Jones。《自旋玻璃的球模型》。英文。载于：《物理评论快报》36.20（1976 年 5 月），第 1217–1220 页。[MPV87] M. M´ezard、G. Parisi 和 M. Virasoro。《自旋玻璃理论及其他：复制方法及其应用导论》。第 9 卷。世界科学出版公司，1987年。[Pan13] D. Panchenko。《Parisi 超度量猜想》。载于：《数学年刊》（第 2 辑）177.1（2013），第 383–393 页。[Pan14] D. Panchenko。《混合 p-自旋模型的 Parisi 公式》。载于：《概率年刊》42.3（2014），第 946–958 页。[Pan16] D. Panchenko。《一般 $2p$- 自旋模型中的温度混沌》。载于：《数学物理通讯》346.2（2016），第 703–739 页。[Pan18a] D. Panchenko。《具有矢量自旋的混合 p-自旋模型的自由能》。载于：《概率年刊》46.2（2018），第 865–896 页。[Pan18b] D. Panchenko。《Potts 自旋玻璃中的自由能》。载于：《概率年刊》46.2（2018），第 829–864 页。

[Par79] G. Parisi。《自旋玻璃的无限多个序参量》。英文。载于：《物理评论快报》43.23（1979年12月），第1754–1756页。[Par80] G. Parisi。《自旋玻璃SK模型的一系列近似解》。载于：《物理学杂志A：数学与一般物理》13.4（1980年）。出版商：IOP Publishing，第L115页。[Par87] B. N. Parlett。《对称特征值问题》。1987年。[Ple82] T. Plefka。《无限程伊辛自旋玻璃模型的TAP方程的收敛条件》。载于：《物理学杂志A：数学与一般物理》15.6（1982年），第1971页。[PT07] D. Panchenko和M. Talagrand。《多个球面SK模型中的重叠》。载于：《概率年刊》35.6（2007年），第2321–2355页。[SK75] D. Sherrington和S. Kirkpatrick。《可解自旋玻璃模型》。载于：《物理评论快报》35.26（1975年）。出版商：美国物理学会，第1792页。[Sub17] E. Subag。《纯球面自旋玻璃吉布斯测度的几何》。英文。载于：《数学创立》210.1（2017年10月），第135–209页。[Sub20] E. Subag。《球形自旋玻璃中的自由能景观》。载于：arXiv:1804.10576*[数学]*（2020年6月）。arXiv：1804.10576。[Sub21] E. Subag。《球面纯 p-自旋模型的自由能——通过TAP方法计算》。载于：arXiv:2101.*04352 [凝聚态物理]*（2021年4月）。arXiv：2101.04352。[Tal00] M. Talagrand。《多层次对称性破缺》。英文。载于：《概率论及相关领域》117.4（2000年8月），第449–466页。[Tal03] M. Talagrand。《自旋玻璃：数学家的挑战：腔方法与平均场模型》。第46卷。施普林格科学与商业媒体，2003年。[Tal06a] M. Talagrand。《球面平均场模型的自由能》。英文。载于：《概率论及相关领域》134.3（2006年3月），第339–382页。[Tal06b] M. Talagrand。《Parisi公式》。载于：《数学年刊》（第2辑）163.1（2006年），第221–263页。[TAP77] D. J. Thouless、P. W. Anderson和R. G. Palmer。《“可解自旋玻璃模型”的解》。英文。载于：《哲学杂志》35.3（1977年3月），第593–601页。