结构化噪声污染下非对称秩一矩阵的失配估计

Teng Fu*、YuHao Liu*、Jean Barbier、Marco Mondelli†、ShanSuo Liang‡ 和TianQi Hou‡
*清华大学数学科学系，中国北京
†奥地利科学技术研究所
‡华为技术有限公司中央研究院理论实验室，2012实验室。电子邮箱：{fut21, yh-liu21}@mails.tsinghua.edu.cn, jean.barbier.cs@gmail.com, marco.mondelli@ist.ac.at, liang.shansuo@huawei.com, thou@connect.ust.hk

arXiv:2302.03306v2 [cs.IT] 8 Feb 2023

摘要——**我们研究贝叶斯统计学家在非对称旋转不变噪声污染下估计秩一信号的性能，其中噪声的奇异值具有一般分布。由于信噪比和噪声结构未知，研究中错误地假设了高斯设定。我们推导出失配贝叶斯估计量误差的精确解析表达式，并分析一种近似消息传递（AMP）算法。第一个结果利用了矩形矩阵球面积分以及低秩矩阵扰动的渐近行为；第二个结果则基于辅助 AMP 的设计与分析。数值实验表明，AMP 估计量与贝叶斯估计量之间存在性能差距，这是由对信号范数的错误估计造成的。**

I. 引言与问题设定

从含噪观测中估计低秩矩阵，在统计学和机器学习中有许多应用：例如稀疏主成分分析（PCA） [1],[2]、社区检测 [3], [4] 或群同步 [5]。统计学 [6], [7] 和随机矩阵理论 [8]–[12] 的研究已确立谱算法的性质。与此同时，用于推断的近似消息传递（AMP）算法也蓬勃发展，其渐近性能可通过状态演化（SE） [13], [14],追踪：这类算法已用于（广义）线性模型 [13], [15]–[19] ，以及与本文设定更为接近的高斯噪声下的低秩矩阵恢复 [20]–[24]。

近来，人们对更“现实”、更丰富的矩阵估计模型的兴趣大幅增加。这些模型纳入了失配（与通常研究的贝叶斯最优设定相对）和/或噪声中的统计依赖（而非独立同分布高斯噪声），参见 [25]–[28]。本文将[28] 扩展至信号矩阵和噪声矩阵非对称的情形。我们的主要贡献是*(i)* 给出失配贝叶斯估计量均方误差（MSE）的表达式，以及*(ii)* 针对噪声统计失配情形的高斯 AMP 状态演化。比较贝叶斯、AMP 和谱估计量后，我们发现了一种与对称情形 [28] 相似的（令人惊讶的）现象；特别是，贝叶斯估计量与 AMP 估计量并不像预期那样一致。

现在我们介绍问题设置。设 $\boldsymbol{u}^* \in \mathbb{R}^n$ 和 $\boldsymbol{v}^* \in \mathbb{R}^m$ 分别在球面 $\mathbb{S}^{n-1}(\sqrt{n})$ 和 $\mathbb{S}^{m-1}(\sqrt{m})$ 上服从均匀分布。球面上的均匀测度 $\mathbb{S}^{n-1}(\sqrt{n})$ 将记为 P_{n° 我们专注于从按如下方式构造的观测数据 $\boldsymbol{Y}$ 中推断 $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ 的任务

$$\boldsymbol{Y} = \sqrt{\frac{\lambda_*}{mn}}\boldsymbol{u}^*\boldsymbol{v}^{*\mathsf{T}} + \boldsymbol{Z} \in \mathbb{R}^{n\times m}. \tag{1}$$

标量 $\lambda_* \geq 0$ 是信噪比（SNR）。我们假设 $n/m \to \alpha \in (0,1]$ 为 $n \to \infty$ （$\alpha \leq 1$，不失一般性，因为当$\alpha > 1$时我们可以考虑 $\boldsymbol{Y}^\mathsf{T}$ ）。略微滥用记号，我们也将$\alpha = \alpha_n = n/m$记为。表示噪声的矩阵 $\boldsymbol{Z}$ 具有双酉不变性，这意味着对于任意正交矩阵 $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ 和 $\boldsymbol{V} \in \mathbb{R}^{m\times m}$， $\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}\mathsf{T}$ 与 $\boldsymbol{Z}$同分布。其经验奇异值分布 $\mu_{\boldsymbol{Z}} = \frac{1}{n}\sum_{i\leq n}\delta_{\sigma_i}$（其中（$\sigma_i$）$_{i\leq n}$是 $\boldsymbol{Z}$的奇异值）在弱意义下收敛于μ_Z，当 $n \to \infty$时成立。$\boldsymbol{Z}\boldsymbol{Z}^\mathsf{T}$ 的渐近经验谱密度记为 ρ_Z，并通过密度变换 $\mu_Z(x)\,\mathrm{d}x = \rho_Z(y)\,\mathrm{d}y$ 从 μ_Z 得到，其中 $y = x^2$。我们还知道， $\boldsymbol{Z}^\mathsf{T}\boldsymbol{Z}$ 的渐近谱密度满足 $\rho_Z = \alpha\rho_Z + (1-\alpha)\delta_0$。注意， $\boldsymbol{Y}$与 $\boldsymbol{Z}$具有相同的极限奇异值分布，但其极限最大奇异值不一定相同，因此下文将省略下标 Z，而不会对 $\boldsymbol{Z}$ 和 $\boldsymbol{Y}$的渐近分布造成歧义，例如，对于 $\mu_Z = \mu_Y$ 和 ρ 均记为 μ，其中 $\rho_Z = \rho_{Y^\circ}$ 我们将 $\boldsymbol{Y}$ 的最大奇异值的极限记为 $\bar{\nu}$，并将 μ的支撑集的上确界记为 $\bar{\gamma}$。 $\boldsymbol{Z}$ 渐近地没有离群值，因此 μ 的支撑集端点 $\bar{\gamma}$ 也是其渐近最大奇异值。

对于一系列关于$\boldsymbol{u}^*\boldsymbol{v}^{*\mathsf{T}}$的估计量 $M(\boldsymbol{Y}) = M_n(\boldsymbol{Y}) \in \mathbb{R}^{n\times m}$，我们将相应的均方误差定义为

$$\mathrm{MSE}_n(M) := \frac{1}{2mn}\mathbb{E}\|\boldsymbol{u}^*\boldsymbol{v}^{*\mathsf{T}} - M(\boldsymbol{Y})\|_\mathrm{F}^2, \tag{2}$$

其中， $\mathbb{E}$ 表示关于（$\boldsymbol{Z}$、 $\boldsymbol{u}^*$、 $\boldsymbol{v}^*$）的期望。我们还考虑另一种对估计量 $\boldsymbol{u}$、 $\boldsymbol{v}$ 关于 $\boldsymbol{u}^*$、 $\boldsymbol{v}^*$的范数不敏感的性能度量，即重标度重叠（下文中， $\langle\cdot,\cdot\rangle$ 表示标准内积）：

$$\mathrm{Overlap}(\boldsymbol{u},\boldsymbol{v}) := \lim_{n\to\infty}\frac{|\langle\boldsymbol{u},\boldsymbol{u}^*\rangle|}{\|\boldsymbol{u}\|\|\boldsymbol{u}^*\|}\frac{|\langle\boldsymbol{v},\boldsymbol{v}^*\rangle|}{\|\boldsymbol{v}\|\|\boldsymbol{v}^*\|}. \tag{3}$$

如果估计量的范数趋于零，我们就将重叠度设为零。

II. 失配贝叶斯估计量

我们考察一位统计学家的表现：他错误地假设噪声为高斯噪声，且其条目独立同分布，服从 $\mathcal{N}(0,1/m)$，信噪比为 λ，该值可能与真实值不同λ_*。因此，用于推断的失配后验分布为

$$P_{\text{mis}}(\mathrm{d}\boldsymbol{u},\mathrm{d}\boldsymbol{v}\mid\boldsymbol{Y})=\frac{1}{Z_n(\boldsymbol{Y})}e^{\sqrt{\frac{\lambda}{\alpha}}\langle\boldsymbol{u},\boldsymbol{Y}\boldsymbol{v}\rangle}P_n(\mathrm{d}\boldsymbol{u})P_m(\mathrm{d}\boldsymbol{v}) \quad (4)$$

an以及配分函数（即后验归一化 常数）为

$$Z_n(\boldsymbol{Y}):=\int P_n(\mathrm{d}\boldsymbol{u})P_m(\mathrm{d}\boldsymbol{v})\exp\Big(\sqrt{\frac{\lambda}{\alpha}}\langle\boldsymbol{u},\boldsymbol{Y}\boldsymbol{v}\rangle\Big). \quad (5)$$

我们分析的相关失配贝叶斯估计量为

$$M_{\text{mis}}(\boldsymbol{Y}):=\int\boldsymbol{u}\boldsymbol{v}^{\mathsf{T}}P_{\text{mis}}(\mathrm{d}\boldsymbol{u},\mathrm{d}\boldsymbol{v}\mid\boldsymbol{Y}). \quad (6)$$

我们将分析式 (2) 的 $n\to\infty$ 极限，记为 MSE(M_{mis})。然而，对于贝叶斯情形，定义 (3) 并无意义（由于符号对称性，$\boldsymbol{u}$、$\boldsymbol{v}$ 的后验均值可能为 0）。因此，在贝叶斯情形下，重叠度定义为

$$\text{Overlap}_{\text{mis}}:=\lim_{n\to\infty}\Big(\frac{1}{mn}\frac{1}{\|M_{\text{mis}}(\boldsymbol{Y})\|_{\text{F}}^2}\times\int P_{\text{mis}}(\mathrm{d}\boldsymbol{u},\mathrm{d}\boldsymbol{v}\mid\boldsymbol{Y})\langle\boldsymbol{u},\boldsymbol{u}^*\rangle^2\langle\boldsymbol{v},\boldsymbol{v}^*\rangle^2\Big)^{1/2}. \quad (7)$$

定义。这里给出一些有用的定义和变换，它们直接借鉴自 [12], [29]。对于任意对称概率测度 μ，比率为 α 的 μ 的 D 变换定义为

$$D_\mu^{(\alpha)}(z):=\int\mu(\mathrm{d}t)\frac{z}{z^2-t^2}\times\Big[\alpha\int\mu(\mathrm{d}t)\frac{z}{z^2-t^2}+\frac{1-\alpha}{z}\Big]$$

对于 $z>\bar{\gamma}$，$D^{(\alpha)}$,$_\mu$ $^{-1}(z)$ 表示其在 $(\bar{\gamma},+\infty)$ 上的函数反函数。令 $T^{(\alpha)}(z):=(\alpha z+1)(z+1)$，其从$[-1,\infty)$ 到 $[0,\infty)$ 的反函数记为 $T^{(\alpha)}$,$_{-1}(z)$，因为它在该区间内递增。最后，比例为 α 的 μ 的矩形 R 变换定义为 $C_\mu^{(\alpha)}(z)=T^{(\alpha)}$,$_{-1}(z(D^{(\alpha)}$,$_\mu$ $^{-1}(z))^2)$，其中 $z\neq0$ 且 $C_\mu^{(\alpha)}(0)=0$。有时我们会省略上标 α 而不会造成歧义，例如 $D^{(\alpha)}$时的 $D_\mu^{-1}(z)$,$_\mu$ $^{-1}(z)$。

A. 对数配分函数

我们的第一个结果涉及对数配分函数 $\ln Z_n(\boldsymbol{Y})$，其表达式可用于确定推断中的相变，并推导贝叶斯估计器的均方误差。我们的计算受到 [25], [26], [28] 的启发。关键在于，由于信号先验分布的球面对称性（$\boldsymbol{u}^*,\boldsymbol{v}^*$），配分函数 (5) 是一个矩形球面积分 I_n [29]。事实上，我们可以将配分函数识别为 $Z_n(\boldsymbol{Y})=I_n(\sqrt{\lambda/\alpha},\boldsymbol{Y})$，其中

$$I_n(\theta,\boldsymbol{Y}):=\int P_n(\mathrm{d}\boldsymbol{u})P_m(\mathrm{d}\boldsymbol{v})\exp(\theta\langle\boldsymbol{u},\boldsymbol{Y}\boldsymbol{v}\rangle). \quad (8)$$

I_n of [29] 的严格渐近公式仅适用于较小的 θ。相比之下，[30], [31] 中给出的非严格公式适用于任意 θ，其表达式为，当 $n\to\infty$时，

$$\frac{1}{n}\ln I_n(\theta,\boldsymbol{Y})\to\frac{1}{2\alpha}\operatorname*{extr}_{(z_1,z_2)\in\mathcal{D}(\theta,\bar{\nu})}\phi(z_1,z_2)-\frac{1+\alpha}{2\alpha}, \quad (9)$$

其中，$\phi=\phi(z_1,z_2)$的表达式为

$$\phi:=-\int\hat{\rho}(\mathrm{d}t)\ln(z_1z_2-\theta^2t)+z_1\alpha+z_2-(\alpha-1)\ln z_1 \quad (10)$$

以及 $\mathcal{D}(\theta,\ \bar{\nu}):=\{(z_1,\ z_2)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}\colon\ z_1z_2\geq\theta^2\bar{\nu}^2\}$。极值 extr ϕ 表示在满足$\nabla\phi=\mathbf{0}$的情况下，于 $(z_1,\ z_2)$ 处计算 ϕ，前提是该驻点 $(z_1,\ z_2)$（可以证明其唯一）存在。如果 $\nabla\phi=\mathbf{0}$ 不存在解，极值就取 $\mathcal{D}$的唯一边界。

我们的目标是得到式 (9) 更明确的表达式。关键在于分析矩形球面积分 (8) 中的“高温区间和低温区间”$\theta=\sqrt{\lambda/\alpha}$，并确定两者之间的转变点。高温区间对应于满足 $\nabla\phi=\mathbf{0}$ 有解的 θ 取值；在低温区间则无解，且 $(z_1,\ z_2)$ 粘滞于其边界值。分隔这两个区间的“粘滞转变”发生在

$$\bar{\theta}:=\Big(\lim_{\nu\downarrow\bar{\nu}}D_\mu^{(\alpha)}(\nu)\Big)^{1/2}. \quad (11)$$

此外，还存在第二个相变，这次由 λ_* 的取值控制。它将这样一个区域与另一个区域区分开来：在前一个区域中，数据 $\boldsymbol{Y}$ 的最大奇异值与奇异值主体分离（因此可以进行非平凡估计）；而在后一个区域中则不存在离群值。我们将其称为“BBP 转变”，以纪念 [8]。这一转变在当前设定下由 [12] 进行了研究。因此，视 λ_* 和 $\theta=\sqrt{\lambda/\alpha}$ 的取值而定，会出现四种可能的区间（其中两种对应相同的对数配分函数表达式）。两个温度区间下的矩形球面积分推导见附录 B；附录 A 回顾了 BBP 转变的背景。

广义观测模型。我们将针对一个比式 (1) 中的数据模型稍微更一般的模型陈述第一个结果，该模型之后将用于推导贝叶斯估计器的均方误差。该模型由 $\boldsymbol{Y}_\epsilon=\boldsymbol{Y}+\sqrt{\epsilon}\,\boldsymbol{W}$ 定义，其中我们在原始数据 (1) 中加入了一个标准 Wigner 矩阵 $\boldsymbol{W}$，且 $\epsilon\ll1$。随后，我们可以定义对数配分函数 $\ln Z_n(\boldsymbol{Y}_\epsilon)$，其中 $Z_n(\boldsymbol{Y}_\epsilon)$ 的定义与式 (5) 相同，只是将 $\boldsymbol{Y}$ 换成了 $\boldsymbol{Y}_\epsilon$。当 $\epsilon=0$ 时，该模型与初始模型 (1) 一致。我们将 $\bar{\nu}_\epsilon$ 定义为当 $n,\ m\to\infty$ 时，$\boldsymbol{Y}_\epsilon$ 的最大奇异值的极限；$\bar{\gamma}_\epsilon$ 是（广义）噪声 $\boldsymbol{Z}+\sqrt{\epsilon}\boldsymbol{W}$ 的最大奇异值的极限，其奇异值极限密度为 μ_ϵ；$\boldsymbol{Y}_\epsilon^{\mathsf{T}}\boldsymbol{Y}_\epsilon$ 的特征值极限密度记为 $\hat{}\rho_\epsilon$，并且 $h_\epsilon:=\lim_{x\downarrow\bar{\gamma}_\epsilon^2}D_{\mu_\epsilon}(x)$。当 $\epsilon=0$ 时，我们省略下标 0，即 $\bar{\nu}_0$ 时记为 $\bar{\nu}$，$\bar{\gamma}_0$ 时记为 $\bar{\gamma}$，等等。我们对这一广义模型的对数配分函数作如下猜想。

猜想 1（对数配分函数）。定义广义模型的矩形球面积分在高温区间和低温区间之间的粘滞转变：$\lambda_\epsilon:=\alpha\lim_{z\downarrow\bar{\nu}_\epsilon}D_{\mu_\epsilon}(z)$。我们几乎必然有

$$\lim_{n\to\infty}\frac{1}{n}\ln Z_n(\boldsymbol{Y}_\epsilon)=\lim_{n\to\infty}\frac{1}{n}\ln I_n\Big(\sqrt{\frac{\lambda}{\alpha}},\boldsymbol{Y}_\epsilon\Big)=f_\epsilon^{(\alpha)}(\lambda,\lambda_*),$$

其中

$$f_\epsilon^{(\alpha)}(\lambda,\lambda_*) := \begin{cases} g_{\lambda,\epsilon}^{(\alpha)}((D_{\mu_\epsilon}^{-1}(\frac{1}{\lambda_*}))^2) & \bar{h}_\epsilon\lambda_* \geq 1 \cap \lambda\lambda_* > \alpha, \\ g_{\lambda,\epsilon}^{(\alpha)}(\bar{\gamma}_\epsilon^2) & \bar{h}_\epsilon\lambda_* < 1 \cap \lambda > \alpha\bar{h}_\epsilon, \\ \int_0^{\sqrt{\frac{\lambda}{\alpha}}} \frac{C_{\mu_\epsilon}^{(\alpha)}(t^2)}{t}\,\mathrm{d}t & \textit{otherwise}. \end{cases}$$

with

$$g_{\lambda,\epsilon}^{(\alpha)}(x) := -\frac{1}{2\alpha}\Big[\int \hat{\rho}_\epsilon(\mathrm{d}t)\ln(x-t) - 2\alpha T^{(\alpha),-1}\Big(\frac{\lambda x}{\alpha}\Big) + (\alpha-1)\ln\Big(T^{(\alpha),-1}\Big(\frac{\lambda x}{\alpha}\Big)+1\Big) + \ln\frac{\lambda}{\alpha}\Big].$$

此外，我们还有 $\frac{1}{n}\mathbb{E}\ln Z_n(\boldsymbol{Y}_\epsilon) \to f_\epsilon^{(\alpha)}(\lambda,\lambda_*)$。

带 $\alpha = 1$ 的非对称情形的对数配分函数与对称情形的对数配分函数之间存在联系。事实上，根据 [28]，对称情形下的配分函数也可以表示为 $f_\epsilon^{\mathrm{sym}}(\lambda,\lambda_*;\,\rho_\epsilon)$。若设 $\mu_\epsilon = \rho_\epsilon$，则得到 $f_\epsilon^{(1)}(\lambda,\lambda_*;\,\mu_\epsilon) = 2f_\epsilon^{\mathrm{sym}}(\lambda,\lambda_*;\,\rho_\epsilon)$。我们强调，非对称情形中的噪声 $\boldsymbol{Z}$ 是高斯矩阵，而非 Wigner 矩阵，并且信号向量 $\boldsymbol{u}^* \neq \boldsymbol{v}^*$。上式中系数 2 的含义是：与非对称情形相比，对称情形中的“信息”只有一半，因为当 $\boldsymbol{Z}$ 是 Wigner 矩阵时，$Z_{ij} = Z_{ji}$ 成立。

我们指出，与对称情形的 [28] 相比，本文的结果并不严谨。这是因为，相较于自由概率，人们对球面积分的矩形版本 [29]，以及更一般的矩形自由概率 [32] 所知甚少。

B. 从对数配分函数到均方误差

将对数配分函数与均方误差联系起来，意味着要为失配的（广义）模型推导出一个“广义 I-MMSE 关系”，类似于 [25], [26] 中针对高斯噪声的关系，或者更贴近本工作的 [28] 中针对旋转不变噪声模型的关系。下面推导均方误差（定义见式 (2)）关系式的过程，正是广义模型中引入 Wigner 矩阵而非直接考虑式 (1) 的原因。还需注意，其（基本）证明独立于猜想 1，且完全严谨。

引理 1（联系对数配分函数与均方误差）。设 $f_{\epsilon,n}^{(\alpha)}(\lambda,\lambda_*) := \mathbb{E}\ln Z_n(\boldsymbol{Y}_\epsilon)/n$。对于模型 (1) 有

$$MSE_n = 1 - \frac{\alpha}{\lambda}\frac{\partial f_{\epsilon,n}^{(\alpha)}(\lambda,\lambda_*)}{\partial\epsilon}\Big|_{\epsilon=0} - 2\alpha\sqrt{\frac{\lambda_*}{\lambda}}\frac{\partial f_{0,n}^{(\alpha)}(\lambda,\lambda_*)}{\partial\lambda_*},$$

其中 $MSE_n := MSE_n(M_{mis}(\boldsymbol{Y}))$。

证明。首先，与 [28], 一样，我们定义以下对象

$$M_n := \frac{\langle \boldsymbol{u}^*,\boldsymbol{u}\rangle\langle \boldsymbol{v}^*,\boldsymbol{v}\rangle}{mn},\quad Q_n := \frac{\langle \boldsymbol{u}^{(1)},\boldsymbol{u}^{(2)}\rangle\langle \boldsymbol{v}^{(1)},\boldsymbol{v}^{(2)}\rangle}{mn},$$

其中 $(\boldsymbol{u},\ \boldsymbol{v}) = (\boldsymbol{u}^{(1)},\ \boldsymbol{v}^{(1)})$ 和 $(\boldsymbol{u}^{(2)},\ \boldsymbol{v}^{(2)})$ 是来自失配后验分布 $P_{\mathrm{mis}}(\,\cdot\mid \boldsymbol{Y}_\epsilon)$ 的两个条件独立样本，该失配后验分布的定义与 (4) 相同，但使用 $\boldsymbol{Y}_\epsilon$ 代替 $\boldsymbol{Y}$。我们有

$$\mathrm{MSE}_n = \frac{1}{2}\big(1 - 2\mathbb{E}\langle M_n\rangle_0 + \mathbb{E}\langle Q_n\rangle_0\big).$$

这里，$\langle\cdot\rangle_\epsilon$ 表示在所有后验样本的联合分布 $P_{\mathrm{mis}}(\,\cdot\,|\boldsymbol{Y}_\epsilon)^{\otimes\infty}$ 下的期望；因此，$\langle\cdot\rangle_0$ 是在 $P_{\mathrm{mis}}(\,\cdot\mid\boldsymbol{Y})^{\otimes\infty}$下的期望。我们直接得到

$$\frac{\partial}{\partial\sqrt{\lambda_*}}f_{0,n}^{(\alpha)}(\lambda,\lambda_*) = \frac{\sqrt{\lambda}}{\alpha}\mathbb{E}\langle M_n\rangle_0,$$

$$\frac{\partial}{\partial\sqrt{\epsilon}}f_{\epsilon,n}^{(\alpha)}(\lambda,\lambda_*) = \frac{1}{n}\sqrt{\frac{\lambda}{\alpha}}\mathbb{E}\langle \boldsymbol{u}^\mathsf{T}\boldsymbol{W}\boldsymbol{v}\rangle_\epsilon = \frac{\lambda\sqrt{\epsilon}}{\alpha}(1-\mathbb{E}\langle Q_n\rangle_\epsilon),$$

其中，最后一个等式使用了简单的高斯分部积分。结合上述方程可得引理 1。 □

现在，通过结合猜想 1 和引理 1，我们可以得到失配贝叶斯估计量的均方误差的显式表达式；如果能对猜想 1 给出严格证明，该表达式将成为一个定理。为此，我们将使用为分析 Dyson-Bessel 过程 [33], [34]所发展出的技术，求 $\int\hat{}\rho_\epsilon(\mathrm{d}t)\ln(x-t)$ 关于 ϵ 的导数，其中 $g_{\lambda\ \epsilon}^{(\alpha)}{}_{,}(x)$。完整推导见附录 C。

猜想 2（失配贝叶斯估计量的性能）。考虑模型 (1)。则有

$$\lim_{n\to\infty} MSE_n(M_{mis}(\boldsymbol{Y})) = \frac{1}{2}\big(1-2M(\lambda,\lambda_*)+Q(\lambda,\lambda_*)\big), \quad (12)$$

其中 M、Q 由 (13) 给出（其中 μ 是渐近噪声奇异值密度，$\mathbb{1}(\cdot)$是指示函数）。

根据猜想 2，与 [28], 类似，我们得到关于式 (7) 中定义的重叠度的以下结果：

$$\mathrm{Overlap}_{\mathrm{mis}} = \frac{M(\lambda,\lambda_*)}{Q(\lambda,\lambda_*)^{1/2}}, \quad (14)$$

只要分母非零，该结果就成立。在数值实验中比较不同方法的重叠度时，我们将对失配贝叶斯估计量使用该结果。

III. 近似消息传递与谱估计器

本节研究近似消息传递估计量和谱估计量。具体而言，我们在高维极限下对近似消息传递的性能进行了确定性刻画，称为状态演化。

A. 近似消息传递估计量

我们从一个与 $\boldsymbol{Z}$独立、且与 $\boldsymbol{u}^*$正相关的初始化 $\boldsymbol{u}^1 \in \mathbb{R}^n$ 开始进行近似消息传递迭代：

$$(\boldsymbol{u}^*,\boldsymbol{u}^1) \xrightarrow{W_2} (U_*,U_1),\ \mathbb{E}[U_*U_1] := \epsilon > 0,\ \mathbb{E}[U_1^2] = 1, \quad (15)$$

其中，$(\boldsymbol{u}^*,\boldsymbol{u}^1) \xrightarrow{W_2} (U_*,U_1)$ 表示 $(\boldsymbol{u}^*,\boldsymbol{u}^1)$ 的联合经验分布在 Wasserstein-2（W_2）距离下收敛到 (U_*,U_1) 的分布。我们注意到，这种初始化并不实用，但可以根据数据的左奇异向量设计一种实用的初始化方法 $\boldsymbol{Y}$ [24], [35], [36]。那么，近似消息传递迭代为

$$\begin{aligned} \boldsymbol{g}^t &= \boldsymbol{Y}^\mathsf{T}\boldsymbol{u}^t - \alpha\cdot\beta_t\boldsymbol{v}^{t-1}, \quad \boldsymbol{v}^t = v_t(\boldsymbol{g}^t), \\ \boldsymbol{f}^t &= \boldsymbol{Y}\boldsymbol{v}^t - \alpha_t\boldsymbol{u}^t, \qquad \boldsymbol{u}^{t+1} = u_{t+1}(\boldsymbol{f}^t), \end{aligned} \quad (16)$$

其中我们假设 $\boldsymbol{v}^0 = \boldsymbol{0}$。这里，非线性函数$v_t,\ u_{t+1}:\ \mathbb{R}\mapsto\mathbb{R}$ 是逐分量应用的；$\beta_1 = 0$ 和

$$M(\lambda,\lambda_*)=\alpha\sqrt{\tfrac{1}{\lambda\lambda_*}}\frac{C_\mu^{(\alpha)}(\frac{1}{\lambda_*})-T^{(\alpha),-1}(\frac{\lambda\lambda_*}{\alpha}T^{(\alpha)}(C_\mu^{(\alpha)}(\frac{1}{\lambda_*})))}{T^{(\alpha)}(C_\mu^{(\alpha)}(\frac{1}{\lambda_*}))}$$
$$\cdot\Big[\tfrac{1}{\lambda_*}C_\mu^{(\alpha)\prime}(\tfrac{1}{\lambda_*})(2\alpha C_\mu^{(\alpha)}(\tfrac{1}{\lambda_*})+\alpha+1)-T^{(\alpha)}(C_\mu^{(\alpha)}(\tfrac{1}{\lambda_*}))\Big]\cdot\mathbb{1}\big(\bar{h}\lambda_*\geq 1\cap\lambda\lambda_*>\alpha\big),$$
$$Q(\lambda,\lambda_*)=\Big\{1-\tfrac{\alpha}{\lambda\lambda_*}\Big[1-\frac{C_\mu^{(\alpha)}(\frac{1}{\lambda_*})-T^{(\alpha),-1}(\frac{\lambda\lambda_*}{\alpha}T^{(\alpha)}(C_\mu^{(\alpha)}(\frac{1}{\lambda_*})))}{T^{(\alpha)}(C_\mu^{(\alpha)}(\frac{1}{\lambda_*}))}(2\alpha C_\mu^{(\alpha)}(\tfrac{1}{\lambda_*})+\alpha+1)\Big]\Big\}\cdot\mathbb{1}\big(\bar{h}\lambda_*\geq 1\cap\lambda\lambda_*>\alpha\big)$$
$$+\Big\{1-\tfrac{\alpha}{\lambda}\Big[\bar{h}-\frac{T^{(\alpha),-1}(\bar{\gamma}^2\bar{h})-T^{(\alpha),-1}(\frac{\lambda\bar{\gamma}^2}{\alpha})}{\bar{\gamma}^2}(2\alpha T^{(\alpha),-1}(\bar{\gamma}^2\bar{h})+\alpha+1)\Big]\Big\}\cdot\mathbb{1}\big(\bar{h}\lambda_*<1\cap\lambda>\alpha\bar{h}\big).\tag{13}$$

对于 $t\geq 2$, $\beta_t=\langle u_t'(\boldsymbol{f}_{t-1})\rangle$; 对于 $t\geq 1$, $\alpha_t=\langle v_t'(\boldsymbol{g}_t)\rangle$, 其中$u_t'$ 和 v_t' 表示导数。于是, $(\boldsymbol{u}^*,\ \boldsymbol{v}^*)$ 的近似消息传递估计量为 $(\boldsymbol{u}^t,\ \boldsymbol{v}^t)$, 而尖峰 $\boldsymbol{u}^*\boldsymbol{v}^{*\mathsf{T}}$ 的估计量为$M_{\text{AMP}}^t=\boldsymbol{u}^t\boldsymbol{v}^{t\mathsf{T}}$。我们将此算法称为高斯 *AMP*, 与对称情形中的做法一样 [28], , 因为这是通常用于高斯噪声的近似消息传递算法。

现在给出该高斯 AMP 的状态演化。我们初始化 $\bar{\nu}_1=\theta\mathbb{E}[U_*U_1]=\theta\epsilon$ 和 $(\bar{\boldsymbol{\Delta}})_{1,1}=\mathbb{E}[U_1^2]=1$, 其中 $\theta:=\sqrt{\lambda_*\alpha}$。然后, 我们为 $t\geq 1$ 定义如下状态演化:

$$(Z_1,\dots,Z_t)\sim\mathcal{N}(0,\bar{\boldsymbol{\Omega}}_t),$$
$$G_t=Z_t+\bar{\nu}_tV_*-\alpha\cdot\bar{\beta}_tV_{t-1}+\sum_{i=1}^{t-1}(\bar{\boldsymbol{B}}_t)_{t,i}V_i,\ \ V_t=v_t(G_t),$$
$$(Y_1,\dots,Y_t)\sim\mathcal{N}(0,\bar{\boldsymbol{\Sigma}}_t),\tag{17}$$
$$F_t=Y_t+\bar{\mu}_tU_*-\bar{\alpha}_tU_t+\sum_{i=1}^{t}(\bar{\boldsymbol{A}}_t)_{t,i}U_i,\ \ U_{t+1}=u_{t+1}(F_t),$$

其中, 我们定义 $V_0=0$, $\beta_1=0$, 并且对于 $t\geq 2$, $\beta_t=\mathbb{E}[u_t'(F_{t-1})]$; 对于 $t\geq 1$, $\bar{\alpha}_t=\mathbb{E}[v_t'(G_t)]$。矩阵 $\boldsymbol{\Delta}_t$、$\boldsymbol{\Gamma}_t$, $\boldsymbol{\Phi}_t$, $\boldsymbol{\Psi}_t$ 以及向量 $\bar{\boldsymbol{\nu}}_t$, $\bar{\boldsymbol{\mu}}_t$ 定义如下:

$$(\bar{\boldsymbol{\mu}}_t)_i=\frac{\theta}{\alpha}\mathbb{E}[V_*V_i],\ \ (\bar{\boldsymbol{\nu}}_t)_i=\theta\mathbb{E}[U_*U_i],\ \ 1\leq i\leq t,$$
$$(\bar{\boldsymbol{\Delta}}_t)_{ij}=\mathbb{E}[U_iU_j],\ \ (\bar{\boldsymbol{\Gamma}}_t)_{ij}=\mathbb{E}[V_iV_j],\ \ 1\leq i,j\leq t,$$
$$(\bar{\boldsymbol{\Phi}}_t)_{ij}=\mathbb{E}[\partial_jU_i],\ \ 1\leq i<j\leq t,$$
$$(\bar{\boldsymbol{\Psi}}_t)_{ij}=\mathbb{E}[\partial_jV_i],\ \ 1\leq i\leq j\leq t.\tag{18}$$

我们强调, $\partial_jU_i=\partial_{y_j}U_i$和 $\partial_jV_i=\partial_{z_j}V_i$。然后, 我们定义矩阵$\boldsymbol{\Omega}_t$, $\boldsymbol{B}_t$, $\boldsymbol{\Sigma}_t$, $\boldsymbol{A}_t$ 如下

$$\bar{\boldsymbol{\Omega}}_t=\alpha\sum_{j=0}^{2t-2}\bar{\kappa}_{2(j+1)}\bar{\boldsymbol{\Theta}}_t^{(j)},\ \ \bar{\boldsymbol{B}}_t=\alpha\sum_{j=0}^{t-1}\bar{\kappa}_{2(j+1)}\bar{\boldsymbol{\Phi}}_t(\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Phi}}_t)^j,$$
$$\bar{\boldsymbol{\Sigma}}_t=\sum_{j=0}^{2t-1}\bar{\kappa}_{2(j+1)}\bar{\boldsymbol{\Xi}}_t^{(j)},\ \ \bar{\boldsymbol{A}}_t=\sum_{j=0}^{t}\bar{\kappa}_{2(j+1)}\bar{\boldsymbol{\Psi}}_t(\bar{\boldsymbol{\Phi}}_t\bar{\boldsymbol{\Psi}}_t)^j,\tag{19}$$

其中

$$\bar{\boldsymbol{\Theta}}_t^{(j)}=\textstyle\sum_{i=0}^{j}(\bar{\boldsymbol{\Phi}}_t\bar{\boldsymbol{\Psi}}_t)^i\bar{\boldsymbol{\Delta}}_t(\bar{\boldsymbol{\Psi}}_t^{\mathsf{T}}\bar{\boldsymbol{\Phi}}_t^{\mathsf{T}})^{j-i}$$
$$+\textstyle\sum_{i=0}^{j-1}(\bar{\boldsymbol{\Phi}}_t\bar{\boldsymbol{\Psi}}_t)^i\bar{\boldsymbol{\Phi}}_t\bar{\boldsymbol{\Gamma}}_t\bar{\boldsymbol{\Phi}}_t^{\mathsf{T}}(\bar{\boldsymbol{\Psi}}_t^{\mathsf{T}}\bar{\boldsymbol{\Phi}}_t^{\mathsf{T}})^{j-1-i},$$
$$\bar{\boldsymbol{\Xi}}_t^{(j)}=\textstyle\sum_{i=0}^{j}(\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Phi}}_t)^i\bar{\boldsymbol{\Gamma}}_t(\bar{\boldsymbol{\Phi}}_t^{\mathsf{T}}\bar{\boldsymbol{\Psi}}_t^{\mathsf{T}})^{j-i}$$
$$+\textstyle\sum_{i=0}^{j-1}(\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Phi}}_t)^i\bar{\boldsymbol{\Psi}}_t\bar{\boldsymbol{\Delta}}_t\bar{\boldsymbol{\Psi}}_t^{\mathsf{T}}(\bar{\boldsymbol{\Phi}}_t^{\mathsf{T}}\bar{\boldsymbol{\Psi}}_t^{\mathsf{T}})^{j-1-i},$$

且 $\{\bar{\kappa}_{2j}\}_{j\geq 1}$ 表示与 $\boldsymbol{Z}$的奇异值分布极限相关联的矩形自由累积量序列。

我们假设非线性函数 v_t 和 u_{t+1} 是 2 阶伪利普希茨函数, 其中函数 ψ 是 2 阶伪利普希茨函数, 当且仅当存在常数 $C>0$, 使得$\|\psi(\boldsymbol{x})-\psi(\boldsymbol{y})\|\leq C(1+\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)\|\boldsymbol{x}-\boldsymbol{y}\|$。我们还假设偏导数

$$\partial_{z_j}v_t(Z_t+\bar{\nu}_tV_*-\alpha\cdot\bar{\beta}_tV_{t-1}+\sum_{i=1}^{t-1}(\bar{\boldsymbol{B}}_t)_{t,i}V_i),$$
$$\partial_{y_j}u_{t+1}(Y_t+\bar{\mu}_tU_*-\bar{\alpha}_tU_t+\sum_{i=1}^{t}(\bar{\boldsymbol{A}}_t)_{t,i}U_i)$$

在式 (17) 给出的 $(Z_1,\dots,Z_t)$、$(V_1,\dots,V_t)$、$(Y_1,\dots,Y_t)$ 和 $(U_1,\dots,U_t)$ 的分布下, 这些函数在概率为 1 的集合上连续。于是, 我们得到高斯 AMP 的如下严格状态演化。请注意, 由于失配, 即使噪声实际上是高斯噪声, 该定理也不能由高斯 AMP 的标准状态演化分析推出 [23],。

定理 1 (高斯 AMP 的状态演化)。考虑模型 (1)、高斯 *AMP* 及其上述状态演化。令 $\phi:\mathbb{R}^{2t+1}\mapsto\mathbb{R}$ 和 $\psi:\ \mathbb{R}^{2t+2}\mapsto\mathbb{R}$ 为任意二阶伪 *Lipschitz* 函数。对于每个 $t\geq 1$, 我们几乎必然有

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\phi\big((\boldsymbol{g}^1)_i,\dots,(\boldsymbol{g}^t)_i,(\boldsymbol{v}^1)_i,\dots,(\boldsymbol{v}^t)_i,(\boldsymbol{v}^*)_i\big)$$
$$=\mathbb{E}\phi\big(G_1,\dots,G_t,V_1,\dots,V_t,V_*\big),\tag{20}$$
$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi\big((\boldsymbol{f}^1)_i,\dots,(\boldsymbol{f}^t)_i,(\boldsymbol{u}^1)_i,\dots,(\boldsymbol{u}^{t+1})_i,(\boldsymbol{u}^*)_i\big)$$
$$=\mathbb{E}\psi\big(F_1,\dots,F_t,U_1,\dots,U_{t+1},U_*\big).\tag{21}$$

证明思路。我们仔细选择非线性函数 $\tilde{v}_t$ 和 $\tilde{u}_{t+1}$, 以设计一种辅助 AMP 算法, 使得 *(i)*存在状态演化刻画, 且 *(ii)* 其迭代结果接近高斯 AMP。完整证明见附录 D。 □

B. 谱估计器

谱估计量的形式为 $J\boldsymbol{u}_1\boldsymbol{v}_1^{\mathsf{T}}$, 其中 $\boldsymbol{u}_1$ 和 $\boldsymbol{v}_1$是与最大奇异值 σ_1 ($\boldsymbol{Y}$) 对应的奇异向量, 其范数分别为 $\sqrt{n}$ 和 $\sqrt{m}$。我们考虑以下两种谱估计量:

- 最优谱估计量 (*OptSpec*) 为 $M_{\text{OS}}=J_{\text{OS}}\boldsymbol{u}_1\boldsymbol{v}_{\mathsf{T}1}$, 其中 $J_{\text{OS}}=J(\mu,\lambda_*)$ 取决于 $\boldsymbol{Z}$ 奇异值分布的极限和信噪比 λ_*, 它是最优的缩放方式, 因为不存在失配, 因此在所有谱估计量中均方误差最小。

- 高斯失配谱估计器（*GauSpec*）为 $M_{\mathrm{GS}} = J_{\mathrm{GS}}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathsf{T}}$，其中 $J_{\mathrm{GS}} = J(\mu_{\mathrm{G}}\lambda)$ 是在噪声为高斯矩阵且真实信噪比为 λ 时的最优缩放方式。因此，它取决于高斯矩阵的奇异值分布 μ_{G} 以及 λ，而不是 (μ, λ_*)。

缩放常数 $J(\mu, \lambda_*)$ 的显式表达式见附录 A 的定理 3。在这两种情况下，$\boldsymbol{u}^*$ 和 $\boldsymbol{v}^*$ 的估计量分别为 $\boldsymbol{u}_1$ 和 $\boldsymbol{v}_1$。由 [12], 可知，重叠度 $\langle \boldsymbol{u}_1, \boldsymbol{u}^* \rangle \langle \boldsymbol{v}_1, \boldsymbol{v}^* \rangle / (mn)$ 收敛至 J_{OS}。由此可推出，$\mathrm{MSE}(M_{\mathrm{OS}}) = \frac{1}{2}(1 - J_{\mathrm{OS}}^2)$，且 $\mathrm{MSE}(M_{\mathrm{GS}}) = \frac{1}{2}(1 + J_{\mathrm{GS}}^2 - 2J_{\mathrm{GS}}J_{\mathrm{OS}})$。由 $\mathrm{MSE}(M_{\mathrm{GS}}) = \mathrm{MSE}(M_{\mathrm{OS}}) + \frac{1}{2}(J_{\mathrm{OS}} - J_{\mathrm{GS}})^2$, 可知，$\mathrm{MSE}(M_{\mathrm{GS}}) \geq \mathrm{MSE}(M_{\mathrm{OS}})$（高斯估计器存在失配，而最优估计器不存在失配）。

请注意，当 $\lambda = \lambda_*$ 时，$\mathrm{MSE}(M_{\mathrm{mis}})$ 与 $\mathrm{MSE}(M_{\mathrm{GS}})$ 之间似乎没有特殊关系。事实上，如果 $h > 1/\sqrt{\alpha}$，那么 $\mathrm{MSE}(M_{\mathrm{mis}})$ 与 $\mathrm{MSE}(M_{\mathrm{GS}})$ 并不相同（尽管在我们的实验中数值上十分接近）。有趣的是，这一性质与方阵情形不同 [28], 即：当 $\lambda = \lambda_*$ 且 $h > 1$ 时，$\mathrm{MSE}(M_{\mathrm{mis}}) = \mathrm{MSE}(M_{\mathrm{GS}})$。

IV. 数值实验

在所有数值实验中，我们确保密度 μ 的方差为 1。我们分别考察噪声统计失配和信噪比失配的影响。具体而言，在第一个例子中，我们考虑噪声统计失配但信噪比匹配的情形 $\lambda = \lambda_*$；在第二个例子中，噪声匹配而信噪比不匹配。

在第一个例子中，噪声的渐近经验奇异值密度是参数为 c 的对称化泊松分布的矩形对应形式。该分布可视为 $((1-c/n)\delta_0 + (c/(2n))(\delta_{-1} + \delta_1))^{*n}$ 的弱极限（f^{*n} 表示函数 f 与自身卷积 n 次），其矩形 R 变换为 $\frac{cz}{1-z}$，参见 [32, 第 4 节 3]。我们设定 $c = 1$ 以确保方差为 1。此外，参数为 $c = 1$ 的对称化泊松分布的矩形对应形式，是随机矩阵 $M(d, d') = \sum_{k=1}^{d} u_d(k)\, v_{d'}(k)^{\mathsf{T}}$ 的奇异值分布的极限，其中 $d \to \infty$、$d/d' \to \alpha$ 和 $u_d(k)$ $v_{d'}(k)$ $(k \geq 1)$ 是分别从 $\mathbb{R}^d, \mathbb{R}^{d'}$ 的单位球面上独立均匀抽取的随机向量，参见 [32, 命题 61]。这一刻画使我们能够模拟对称化泊松分布的矩形对应形式，并计算其理论贝叶斯均方误差。我们将贝叶斯估计器与在 [37] 中实现的正确的近似消息传递进行比较（基于先前工作 [38]），并称其为“正确的”，因为我们猜想它在低

近期关于结构化噪声下秩矩阵估计的研究见 [39]。然而，我们强调，当信号 $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ 并非在球面上均匀分布或服从高斯分布时，[37]中的近似消息传递算法已被证明不再是最优的。在这种情况下，针对对称情形的猜想贝叶斯最优近似消息传递算法近期在 [39] 中被提出。

我们观察到以下现象，其中一些与对称情形 [28] 相对应。首先，高斯 AMP 的表现不如失配贝叶斯估计量，尤其是在 λ_* 较小时。当 λ_* 为

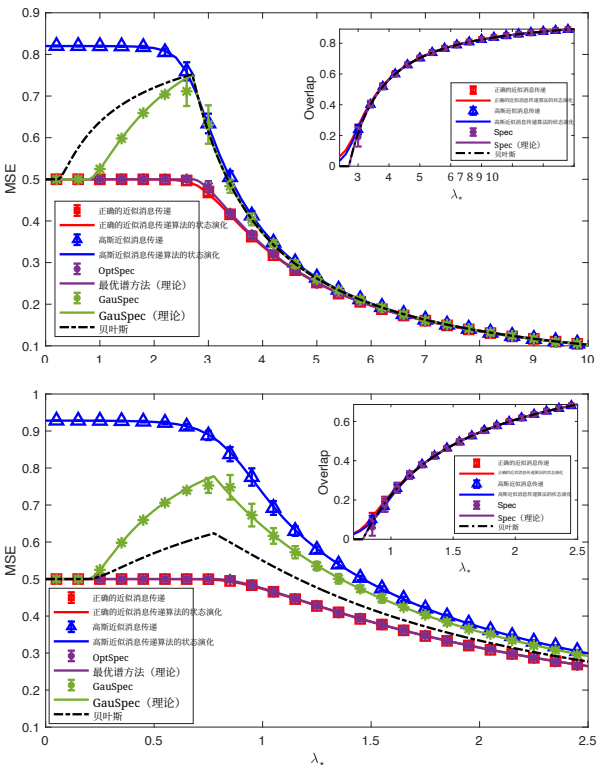

图 1. 两种失配设置下，均方误差和重叠度随真实信噪比 λ_* 的变化。第一幅图对应于信噪比匹配但噪声失配的设置，其中噪声根据对称化泊松分布的矩形对应形式 [32]生成。第二幅图则相反：噪声服从高斯分布（因此噪声匹配），而信噪比失配，取值为 $\lambda = 4\lambda_*$。对于近似消息传递算法，我们使用 $m = 20000$；对于谱估计器，我们使用 $m = 10000$。我们设定 $\alpha = n/m = 0.6$。近似消息传递算法的结果取 100 次试验的平均值，谱估计器的结果取 20 次试验的平均值。

足够大时，所有估计量的重叠度都相匹配，并达到最优估计器（正确的近似消息传递和 OptSpec）的重叠度。因此，均方误差上的差距源于对信号范数的不正确估计，而非重叠度（即其“方向”）。第三，当不存在信噪比失配且 $h > 1/\sqrt{\alpha}$ 时，我们观察到 $\mathrm{MSE}(M_{\mathrm{mis}})$ 几乎与 $\mathrm{MSE}(M_{\mathrm{GS}})$ 相匹配。然而，我们强调，这两个值并不完全相同。第四，所有估计量的表现都不如两个最优估计器。最后，贝叶斯和高斯谱均方误差曲线随真实信噪比 λ_* 单调不减。

V. 展望

与 [28], 本文一道，我们揭示了秩一矩阵估计中失配所导致的各种算法行为中一种令人惊讶的现象。澄清这些观察结果的普适性将十分有趣；特别是，当噪声为非齐次时，它们是否会扩展到矩阵估计的多视图版本 [36], [40], , 或者表现出何种差异 [41], , 以及在张量估计中是否同样如此 [42], [43]。另一个方向是，将本文所量化的失配代价与贝叶斯最优性能进行比较，而这需要将 [39]扩展到非对称设置。

参考文献

[1] I. M. Johnstone 和 A. Y. Lu，《高维主成分分析中的一致性与稀疏性》，美国统计学会期刊，第 104 卷，第 486 期，第 682–693 页，2009 年。[2] H. Zou、T. Hastie 和 R. Tibshirani，《稀疏主成分分析》，计算与图形统计学期刊，第 15 卷，第 2 期，第 265–286 页，2006 年。[3] E. Abbe，《社区检测与随机块模型：最新进展》，机器学习研究期刊，第 18 卷，第 1 期，第 6446–6531 页，2017 年。[4] C. Moore，《社区检测的计算机科学与物理学：景观、相变与困难性》，arXiv 预印本 arXiv:1702.00467，2017 年。[5] A. Perry、A. S. Wein、A. S. Bandeira 和 A. Moitra，《紧群上同步问题的消息传递算法》，纯粹与应用数学通讯，第 71 卷，第 11 期，第 2275–2322 页，2018 年。[6] I. M. Johnstone，《主成分分析中最大特征值的分布》，《统计年鉴》，第 29 卷，第 2 期，第 295–327 页，2001 年。[7] D. Paul，《大维度尖峰协方差模型的样本特征结构渐近性》，中华统计学志，第 1617–1642 页，2007 年。[8] J. Baik、G. B. Arous 和 S. Péché，《非空复样本协方差矩阵最大特征值的相变》，概率年刊，第 33 卷，第 5 期，第 1643–1697 页，2005 年。[9] J. Baik 和 J. W. Silverstein，《尖峰总体模型的大型样本协方差矩阵特征值》，多元分析期刊，第 97 卷，第 6 期，第 1382–1408 页，2006 年。[10] Z. Bai 和 J. Yao，《广义尖峰总体模型中的样本特征值》，多元分析期刊，第 106 卷，第 167–177 页，2012 年。[11] F. Benaych-Georges 和 R. R. Nadakuditi，《大型随机矩阵有限秩低秩扰动的特征值与特征向量》，数学进展，第 227 卷，第 1 期，第 494–521 页，2011 年。[12] ——，《大型矩形随机矩阵低秩扰动的奇异值与奇异向量》，多元分析期刊，第 111 卷，第 120–135 页，2012 年。[13] M. Bayati 和 A. Montanari，《稠密图上的消息传递动力学及其在压缩感知中的应用》，IEEE 信息论汇刊，第 57 卷，第 2 期，第 764–785 页，2011 年。[14] E. Bolthausen，《Sherrington–Kirkpatrick模型的 TAP 方程解的迭代构造》，数学物理通讯，第 325 卷，第 1 期，第 333–366 页，2014 年。[15] D. L. Donoho、A. Maleki 和 A. Montanari，《压缩感知的消息传递算法》，美国国家科学院院刊，第 106 卷，第 45 期，第 18 914–18 919 页，2009 年。[16] J. Barbier、F. Krzakala、N. Macris、L. Miolane 和 L. Zdeborová，《高维广义线性模型中的最优误差与相变》，美国国家科学院院刊，第 116 卷，第 12 期，第 5451–5460 页，2019 年。[17] A. Maillard、B. Loureiro、F. Krzakala 和 L. Zdeborová，《高维相位恢复：统计相变与计算相变》，《神经信息处理系统进展》，第 33 卷，第 11 071–11 082 页，2020 年。[18] M. Mondelli 和 R. Venkataramanan，《带谱初始化的广义线性模型近似消息传递》，载于《人工智能与统计国际会议》，PMLR，2021 年，第 397–405 页。[19] S. Rangan，《用于随机线性混合估计的广义近似消息传递》，载于 IEEE国际信息论研讨会（ISIT），2011 年，第 2168–2172 页。[20] Y. Deshpande 和 A. Montanari，《信息论最优稀疏主成分分析》，载于 2014 年 IEEE国际信息论研讨会，IEEE，2014 年，第 2197–2201 页。[21] J. Barbier、N. Macris 和 C. Rush，《稀疏尖峰矩阵估计中的全有或全无统计相变与计算相变》，《神经信息处理系统进展》，第 33 卷，第 14 915–14 926 页，2020 年。[22] M. Dia、N. Macris、F. Krzakala、T. Lesieur、L. Zdeborová 等，《对称秩一矩阵估计中的互信息：副本公式的证明》，《神经信息处理系统进展》，第 29 卷，2016 年。

[23] A. K. Fletcher 和 S. Rangan，“噪声中秩一矩阵的迭代重建”，《信息与推断：IMA期刊》，第 7 卷，第 3 期，第 531–562 页，2018 年。[24] A. Montanari 和 R. Venkataramanan，“通过近似消息传递估计低秩矩阵”，《统计年鉴》，第 49 卷，第 1 期，第 321–345 页，2021 年。[25] F. Pourkamali 和 N. Macris，“高斯噪声下对称秩一矩阵的失配估计”，载于《国际苏黎世信息与通信研讨会》（IZS 2022）论文集。苏黎世联邦理工学院，2022 年，第 84–88 页。[26] ——，“高斯噪声下非对称秩一矩阵的失配估计”，载于《2022 IEEE国际信息论研讨会》（ISIT）。IEEE，2022 年，第 1288–1293 页。[27] F. Camilli、P. Contucci 和 E. Mingione，“失配设定下的推断问题：具有马蒂斯相互作用的自旋玻璃模型”，《SciPost物理》，第 12 卷，第 4 期，第 125 页，2022 年。[28] J. Barbier、T. Hou、M. Mondelli 和 M. Sáenz，“无知的代价：在低秩矩阵估计中忽略噪声结构要付出多少代价？”，载于《神经信息处理系统进展》，2022 年。[29] F. Benaych-Georges，“矩形 R 变换作为矩形球面积分的极限”，《理论概率杂志》，第 24 卷，第 4 期，第 969–987 页，2011 年。[30] Y. Kabashima，“从相关模式进行推断：感知机学习与线性向量信道的统一理论”，载于《物理学杂志：会议系列》，第 95 卷，第 1 期。IOP出版社，2008 年，第 012001 页。[31] A. Maillard、L. Foini、A. L. Castellanos、F. Krzakala、M. Mézard 和 L. Zdeborová，“高温展开与消息传递算法”，《统计力学杂志：理论与实验》，第 2019 卷，第 11 期，第 113301 页，2019 年。[32] F. Benaych-Georges，“矩形自由卷积的无限可分布：分类与矩阵解释”，《概率论及相关领域》，第 139 卷，第 1 期，第 143–189 页，2007 年。[33] A. Guionnet 和 J. Huang，“矩形球面积分的大偏差渐近性质”，arXiv 预印本 arXiv:2106.07146，2021 年。[34] M. Potters 和 J.-P. Bouchaud，《随机矩阵理论入门：面向物理学家、工程师和数据科学家》。剑桥大学出版社，2020 年。[35] M. Mondelli 和 R. Venkataramanan，“旋转不变模型中近似消息传递的主成分分析初始化”，载于《神经信息处理系统进展》，第 34 卷，2021 年，第 29 616–29 629 页。[36] X. Zhong、T. Wang 和 Z. Fan，“正交不变系综的近似消息传递：多元非线性函数与谱初始化”，arXiv 预印本 arXiv:2110.02318，2021 年。[37] Z. Fan，“旋转不变矩阵的近似消息传递算法”，《统计年鉴》，第 50 卷，224，2022 年 1，197第。–期。[38] M. Opper 和 O. Winther，“用于概率建模的自适应、自平均 Thouless-Anderson-Palmer 平均场理论”，《物理评论E》，第 64 卷，第 5 期，第 056131 页，2001 年。[39] J. Barbier、F. Camilli、M. Mondelli 和 M. Saenz，“结构化主成分分析中的贝叶斯最优极限及其实现方法”，arXiv 预印本 arXiv:2210.01237，2022 年。[40] J. Barbier 和 G. Reeves，“多视图低秩对称尖峰矩阵模型的信息论极限”，载于 IEEE国际信息论研讨会（ISIT），2020 年，第 2771–2776 页。[41] A. Guionnet、J. Ko、F. Krzakala 和 L. Zdeborová，“非均匀噪声下的低秩矩阵估计”，arXiv 预印本 arXiv:2208 05918，2022 年。[42] J. H. de Morais Goulart、R. Couillet 和 P. Comon，“随机张量的随机矩阵视角”，《机器学习研究杂志》，第 23 卷，第 1–36 页，2022 年。[43] M. E. A. Seddik、M. Guillaud 和 R. Couillet，“随机张量遇上随机矩阵”，arXiv 预印本 arXiv:2112.12348，2021 年。[44]A. Guionnet、M. Maı 等，“R 变换及球面积分相关渐近性质的傅里叶视角”，《泛函分析杂志》，第 222 卷，第 2 期，第 435–490 页，2005 年。[45] F. Benaych-Georges 和 M. Debbah，“自由反卷积：从理论到实践”，《生物启发式自主网络与服务范式》，第 201–224 页，2008 年。

附录 A 旋转不变矩阵的低秩扰动

在本附录中，我们回顾旋转不变矩阵低秩扰动的一些已知结果 [12]。第一个结果涉及存在秩一扰动时 $\boldsymbol{Y}$ 的最大奇异值。

定理 2 （ [12]中的定理 2.8。考虑尖峰模型（1）。那么，当 $n \to \infty$ 时， $\boldsymbol{Y}$ 的最大奇异值 $\bar{\nu}$ 几乎必然收敛到

$$\bar{\nu} = D_{\mu}^{-1}\big(\frac{1}{\lambda_*}\big)\mathbb{1}\big(\bar{h}\lambda_* \geq 1\big) + \bar{\gamma}\mathbb{1}\big(\bar{h}\lambda_* < 1\big). \tag{22}$$

第二个结果涉及真实信号与对应于最大奇异值的奇异向量之间的重叠度。

定理 3 （ [12]中的定理 2.9。考虑尖峰模型（1）。那么，当 $n \to \infty$ 时，对应于 $\boldsymbol{Y}$ 的最大奇异值 $\bar{\nu}$ 的奇异向量 $\boldsymbol{u}_1$ 和 $\boldsymbol{v}_1$ 几乎必然满足（23）。

$$J(\mu, \lambda_*) := \lim_{n\to\infty} \frac{|\langle \boldsymbol{u}_1, \boldsymbol{u}^* \rangle \langle \boldsymbol{v}_1, \boldsymbol{v}^* \rangle|}{mn} = \frac{|T^{(\alpha)}(C_{\mu}^{(\alpha)}(\frac{1}{\lambda_*})) - \frac{1}{\lambda_*} C_{\mu}^{(\alpha)\prime}(\frac{1}{\lambda_*})(2\alpha C_{\mu}^{(\alpha)}(\frac{1}{\lambda_*}) + \alpha + 1)|}{\sqrt{T^{(\alpha)}(C_{\mu}^{(\alpha)}(\frac{1}{\lambda_*}))}} \cdot \mathbb{1}(\bar{h}\lambda_* \geq 1) \tag{23}$$

附录 B 对数配分函数的计算

在本节中，我们将具体分析矩形球面积分的“高温和低温” θ 区间，并确定两者之间的转变点。

A. 高温区间

既然我们已经认识到配分函数是一个矩形球面积分，就可以使用[29] 中的主要结果，该结果严格证明了对于足够小的 θ,

$$\lim_{n\to\infty} \frac{1}{n} \ln I_n(\theta, \boldsymbol{Y}) = \int_0^{\theta} \frac{C_{\mu}^{(\alpha)}(t^2)}{t}\, \mathrm{d}t. \tag{24}$$

函数 $C_{\mu}^{(\alpha)}$是所谓的带比率 α [29]的矩形 R 变换。因此，本节将证明，在某个“温度”区间内，矩形球面积分的表达式 (24) 和 (9) 是等价的。该区间对应于满足以下条件的 θ 取值： $\nabla\phi = \mathbf{0}$ 确实存在一个解。我们称之为矩形球面积分的高温区间。

我们明确写出驻点所满足的方程。利用 $\hat{}\rho$ 的 R 变换（例如参见 [34]），可得

$$\frac{\partial\phi}{\partial z_2} = 0 \quad \Rightarrow \quad z_2 = \frac{\theta^2}{z_1} R_{\hat{\rho}}\left(\frac{\theta^2}{z_1}\right) + 1. \tag{25}$$

它满足的一个有用恒等式（对等式两边关于 θ 求导，并注意到该式在 $\theta = 0$ 处成立即可得到）为 [44]

$$\theta R_{\rho}(\theta) - \int \rho(\mathrm{d}t) \ln(\theta R_{\rho}(\theta) - \theta t + 1) = \int_0^{\theta} R_{\rho}(x)\, \mathrm{d}x.$$

因此，

$$\phi(z_1, z_2(z_1)) = \int_0^{\frac{\theta^2}{z_1}} R_{\hat{\rho}}(t)\, \mathrm{d}t + \alpha z_1 - \alpha \ln z_1 + 1. \tag{26}$$

于是， z_1 的驻点条件满足

$$\frac{\mathrm{d}}{\mathrm{d}z_1}\phi(z_1, z_2(z_1)) = -\frac{\theta^2}{z_1^2} R_{\hat{\rho}}\left(\frac{\theta^2}{z_1}\right) + \alpha - \frac{\alpha}{z_1} = 0. \tag{27}$$

我们希望证明，当 (z_1, z_2) 是上述鞍点方程 (25) 和 (27) 的解时，

$$\frac{1}{2\alpha}[\phi(z_1, z_2(z_1)) - (1+\alpha)] = \int_0^{\theta} \frac{C_{\mu}^{(\alpha)}(t^2)}{t}\, \mathrm{d}t. \tag{28}$$

注意，当 $\theta = 0$ 时，该方程成立：将 $\theta = 0$代入 (27) 得到 $z_1 = 1$；对 (26) 进行同样的操作得到$\phi(z_1, z_2(z_1))|_{\theta=0} = \alpha + 1$，因此 (28) 变为 $0 = 0$。接下来，对 θ 求导，然后将等式两边分别乘以 θ （因为我们处于极值 $\partial_{z_1}\phi = 0$处）。这给出

$$z_1 - 1 = \frac{1}{\alpha}\frac{\theta^2}{z_1} R_{\hat{\rho}}\left(\frac{\theta^2}{z_1}\right) = \frac{1}{\alpha} K_{\hat{\rho}}\left(\frac{\theta^2}{z_1}\right) \overset{?}{=} C_{\mu}^{(\alpha)}(\theta^2) =: \gamma(\theta), \tag{29}$$

其中 $K_{\hat{\rho}}(z) := zR_{\hat{\rho}}(z)$。第一个等式直接由 (27) 得出，而问号部分正是我们需要证明的内容。简单记 $\gamma = \gamma(\theta)$, 若定义 $M_{\rho}(z) = \sum_{k\geq 1} m_k(\rho) z^k$ ，其中 $m_k(\rho) := \int \rho(\mathrm{d}t) t^2$，根据 [29]的引理 3.2，我们有

$$M_{\rho}\left(\frac{\theta^2}{(\alpha\gamma+1)(\gamma+1)}\right) = \gamma. \tag{30}$$

根据自由累积量与矩之间的关系，我们还有 $K_{\hat{\rho}}(z(\alpha M_{\rho}(z)+1)) = \alpha M_{\rho}(z)$。这里我们使用了这样一个事实：

$$M_{\hat{\rho}}(z) = \alpha M_{\rho}(z) \qquad K_{\hat{\rho}}(z\, M_{\hat{\rho}}(z) + 1) = \alpha M_{\rho}(z),$$

以及来自 [45] 的关系 $z = \theta^2/(\alpha\gamma+1)(\gamma+1)$。将 $z = \theta^2/(\alpha\gamma+1)(\gamma+1)$ 代入上式即可得到

$$K_{\hat{\rho}}\left(\frac{\theta^2}{\gamma+1}\right) = \alpha M_{\rho}\left(\frac{\theta^2}{(\alpha\gamma+1)(\gamma+1)}\right) = \alpha\gamma. \tag{31}$$

最后，将其与 (29) 和 (31) 的左侧进行比较，我们得到

$$\gamma(\theta) = z_1(\theta) - 1 \tag{32}$$

这表明式 (29) 成立。因此，只要 (z_1, z_2) 是驻点条件 (25) 和 (27) 的解，式 (28) 就成立，并且 $C_{\mu}^{(\alpha)}(\theta^2) = z_1 - 1$。

为完成论证，我们现在证明：当驻定解存在时，它是唯一的。我们再次写出两个临界条件：

$$\frac{\partial\phi}{\partial z_1}=-\int\hat{\rho}(\mathrm{d}t)\frac{z_2}{z_1z_2-\theta^2t}+\alpha-\frac{\alpha-1}{z_1}=0, \quad (33)$$

$$\frac{\partial\phi}{\partial z_2}=-\int\hat{\rho}(\mathrm{d}t)\frac{z_1}{z_1z_2-\theta^2t}+1=0. \quad (34)$$

注意，$z_1>0$，$z_2>0$。结合这两个方程可得

$$z_2=z_2(z_1)=\alpha z_1-\alpha+1. \quad (35)$$

该方程与 (34) 一起唯一确定了满足驻定方程的解 $(z_1,\ z_2)$。事实上，将 (35) 代入 (34)，我们可得

$$\int\hat{\rho}(\mathrm{d}t)\Big(z_2(z_1)-\frac{\theta^2t}{z_1}\Big)^{-1}=1. \quad (36)$$

该积分关于 z_1严格递减。因此，如果它有关于 z_1的解，那么该解唯一。在这种情况下，由于 (35)，z_2 也唯一。

总而言之：如果方程组 (33)-(34)，或等价地 (35)-(36)，存在解（我们现在知道该解唯一），那么必须将其代入 (9)，从而得到用矩形 R 变换表示的等价表达式 (24)。

B. 低温区间

现在考虑不存在驻定解的情形。这意味着，不存在任何 $(z_1,z_2)\in\mathcal{D}(\theta,\bar{\nu})$ 能满足方程组 (33)-(34)（或等价地，满足 (34)-(35)）。我们将这种情形称为矩形球面积分的“低温区间”。此时，极值位于定义域 $\mathcal{D}(\theta,\bar{\nu})$ 的边界上。这意味着，需要选取由 (35) 给出的 z_2，而 z_1 则由下述定义 $\mathcal{D}(\theta,\bar{\nu})$ 边界的方程给出（回顾一下，$\bar{\nu}$ 是 $\boldsymbol{Y}$ 的最大奇异值）：

$$z_1z_2=\theta^2\bar{\nu}^2. \quad (37)$$

现在解释为何当 θ 不小时，(35) 和 (37) 能确定 (9) 的值，并且此时不存在 $\nabla\phi=\mathbf{0}$ 的驻定解。应将 (10) 中的函数 ϕ 理解为以下函数序列在 $n\to\infty$ 时的极限：

$$\phi_n(z_1,z_2):=-\frac{1}{n}\sum_{k\le n}\ln(z_1z_2-\theta^2\lambda_k)+z_1\alpha+z_2-(\alpha-1)\ln z_1,$$

其中，$(\lambda_k)_{k\le n}$ 是矩阵 $\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}}$ 的特征值，我们按从大到小的顺序排列，即从最大值 λ_1 到最小值 λ_n。函数 ϕ_n 表示一个需要进行极值化的作用量，以通过最速下降法得到矩形球面积分的表达式。因此，对于较大但有限的 n，我们需要寻找 $\nabla\phi_n=\mathbf{0}$ 的驻定解（而不是直接求解 $\nabla\phi=\mathbf{0}$）。在计算大 n 时球面积分所需代入作用量 ϕ_n 的，实际上是 $\nabla\phi_n=\mathbf{0}$ 的唯一解；该解仅在高温区间内渐近地与 $\nabla\phi=\mathbf{0}$ 的解相匹配。

驻点条件 $\nabla\phi_n=\mathbf{0}$ 的形式类似于 (33)-(34)，但将对渐近密度 $\hat{\rho}(t)$ 的积分替换为对特征值 $(\lambda_k)_{k\le n}$ 的经验期望。因此，条件 (35) 仍然成立。相反，驻点条件 (36) 现在写作

$$\frac{1}{n}\sum_{2\le k\le n}\Big(z_2(z_1)-\frac{\theta^2\lambda_k}{z_1}\Big)^{-1}+\frac{1}{n}\Big(z_2(z_1)-\frac{\theta^2\lambda_1}{z_1}\Big)^{-1}=1.$$

上述求和 $n^{-1}\sum_{2\le k\le n}(\cdots)^{-1}$（随 z_1 增大而减小）与式 (36) 中的积分具有相同的极限。根据不存在 $\nabla\phi=\mathbf{0}$ 驻定解这一假设，对于 $\mathcal{D}(\theta,\bar{\nu})$ 中 z_1 的任意取值，该积分在 $n\to\infty$ 时始终小于 1。因此，第二项必须足够大以进行补偿并满足上述驻点条件，也就是说，它必须满足

$$z_2(z_1)z_1-\theta^2\lambda_1=\Theta(n^{-1}).$$

$\lambda_1(n)$ 的极限为 $\bar{\nu}^2$，因此在 n 趋于无穷的极限下，我们得到 (37)。因此，当 $\nabla\phi=\mathbf{0}$ 无解时，对于所有足够大的 θ，$\nabla\phi_n=\mathbf{0}$ 在 n 较大时的解 (z_1,z_2) 会“停留”在一个常数值上，这与标准球面积分 [34], [44] 中的情形类似。在这个“低温区间”内，我们因此通过求解 (35)、(37) 得到 (z_1,z_2) 的值：

$$z_1=z_1(\alpha,\theta,\bar{\nu})=T^{(\alpha),-1}(\theta^2\bar{\nu}^2)+1, \quad (38)$$

$$z_2=z_2(z_1)=\alpha z_1(\alpha,\theta,\bar{\nu})-\alpha+1. \quad (39)$$

将 (37)、(38)、(39) 代入 (9)、(10)，低温区间下的对数矩形球面积分为

$$-\frac{1}{2\alpha}\int\hat{\rho}(\mathrm{d}t)\ln(\bar{\nu}^2-t)+z_1(\alpha,\theta,\bar{\nu})-\frac{\alpha-1}{2\alpha}\ln z_1(\alpha,\theta,\bar{\nu})-\frac{\ln|\theta|}{\alpha}-1. \quad (40)$$

C. 求相界

我们的目标是求出温度 θ 的取值 θ，它将上述高温相与低温相分隔开。根据前面的所有说明，我们知道，解 (z_1,z_2) 关于 θ 连续（显然，它在高温相 $z_1z_2>\theta^2\bar{\nu}^2$ 中连续，进入低温相时则会停留在边界 $z_1z_2=\theta^2\bar{\nu}^2$ 上）。因此，转变点 θ 由以下条件确定：若 $\theta<\theta$，则 (35)、(36) 成立，且球面积分可由 (24) 或 (9) 给出，二者结果相同；若 $\theta>\theta$，则 (35)、(38) 成立，并且必须采用变分形式 (9)，而且恰在 $\theta=\theta$ 时，所有这些条件同时成立，球面积分的两种表达式也相同。因此，将 (38)、(37) 代入 (36) 后，关于 θ 的条件可写为

$$\lim_{z\downarrow\bar{\nu}^2}\frac{\alpha-1+((\alpha-1)^2+4\alpha\bar{\theta}^2z)^{1/2}}{2\alpha\bar{\theta}^2}\int\hat{\rho}(\mathrm{d}t)\frac{1}{z-t}=1,$$

其中，取 lim 是为了确保左侧的值有意义。回忆一下，$H_{\hat{\rho}}(z):=\int\hat{\rho}(\mathrm{d}t)(z-t)^{-1}$，并假设 $\lim_{z\downarrow\bar{\nu}^2}H_{\hat{\rho}}(z)$ 有限（此类密度的一个标准例子是马尔琴科-帕斯图尔定律）。那么，由上文可得

$$\bar{\theta}=\sqrt{D_\mu(\bar{\nu}^+)}.$$

这里， $D_\mu > 0$ 是 μ的 D 变换。如果 $z \geq \bar{\nu}^2$，那么 $z \geq \bar{\gamma}^2$ ，因此我们可以使用这个定义。由此，我们得到了分隔这两种状态的相变发生时的温度，见式 (11)。

D. 汇总各项以得到对数配分函数

现在，我们已经分别得到了两个温度区间下的对数矩形球面积分，并找到了它们之间的转变点 θ ，因此可以着手求出对数配分函数。第一个转变涉及数据 $\boldsymbol{Y}$ 的最大奇异值 $\bar{\nu}$ 的极限行为。这个问题已在 [12] 中进行过研究，我们在附录 A 的定理 2 中回顾了相关内容。另一个将发挥作用的转变，是由式（11）所决定的矩形球面积分在高温和低温区间行为之间的转变，其表达式为：

$$\text{Sticking transition}: \quad \lambda/\alpha = D_\mu^{(\alpha)}(\bar{\nu}^+). \tag{41}$$

我们对广义观测模型的猜想见猜想 1；当令 $\epsilon = 0$ 时，该猜想与最初的猜想 (1) 相符。猜想 1 可由我们此前的所有结果直接汇总得到，尤其包括：矩形球面积分在高温区间 (24) 和低温区间 (40) 下的表达式之区别，以及对这两种表达式之间转变点 (41) 的确定。

附录 C 均方误差的计算

既然我们已经得到了对数配分函数的公式，就可以推导失配贝叶斯统计学家的均方误差（MSE）表达式。在本节中，由于我们只考虑失配贝叶斯估计量，将省略估计量符号 $M_{\text{mis}}(\boldsymbol{Y})$。根据引理 1 的细节，我们可以定义 $M(\lambda, \lambda_*)$ 和 $Q(\lambda, \lambda_*)$ 如下

$$M(\lambda,\lambda_*) := \lim_{n\to\infty} \mathbb{E}\langle M_n\rangle_0 = 2\alpha\sqrt{\frac{\lambda_*}{\lambda}}\frac{\partial f_0(\lambda,\lambda_*)}{\partial\lambda_*}, \tag{42}$$

$$Q(\lambda,\lambda_*) := \lim_{n\to\infty} \mathbb{E}\langle Q_n\rangle_0 = 1 - \frac{2\alpha}{\lambda}\frac{\partial f_\epsilon(\lambda,\lambda_*)}{\partial\epsilon}\bigg|_{\epsilon=0}. \tag{43}$$

当 $n\to\infty$ 时，MSE 的极限为

$$\lim_{n\to\infty} \text{MSE}_n = \big(1 - 2M(\lambda,\lambda_*) + Q(\lambda,\lambda_*)\big)/2. \tag{44}$$

因此，我们需要根据猜想 1 所规定的不同温度（即信噪比）范围，计算导数 (42)、(43)，从而完成猜想 2 的证明。

A. Low temperature regime 1: $\bar{h}\lambda_* \geq 1$ and $\lambda\lambda_* > \alpha$

1) 计算 $M(\lambda, \lambda_*)$：从现在起，我们将 D_{μ_ϵ} 简记为 D_ϵ，将 C_{μ_ϵ} 简记为 C_ϵ，等等。首先，当 $h\lambda_* \geq 1$ 和 $\lambda\lambda_* > \alpha$ 时，我们得到

$$\frac{\partial f_0(\lambda,\lambda_*)}{\partial\lambda_*} = \frac{\mathrm{d}g_{\lambda,\epsilon}^{(\alpha)}(x)}{\mathrm{d}x}\bigg|_{\epsilon=0,x=(D_0^{-1}(\frac{1}{\lambda_*}))^2} \frac{\partial(D_\epsilon^{-1}(\frac{1}{\lambda_*}))^2}{\partial\lambda_*}\bigg|_{\epsilon=0}. \tag{45}$$

利用连接 D-变换与矩形 R-变换的关系，我们很容易得到

$$\frac{\partial(D_\epsilon^{-1}(\frac{1}{\lambda_*}))^2}{\partial\lambda_*}\bigg|_{\epsilon=0} = -\frac{1}{\lambda_*}C_0'\left(\frac{1}{\lambda_*}\right)\left(2\alpha C_0\left(\frac{1}{\lambda_*}\right)+\alpha+1\right) + \left(\alpha C_0\left(\frac{1}{\lambda_*}\right)+1\right)\left(C_0\left(\frac{1}{\lambda_*}\right)+1\right). \tag{46}$$

我们需要计算的另一项是当 $\epsilon = 0$ 时的 $\mathrm{d}_x g_{\lambda,\epsilon}^{(\alpha)}(x)$。根据 $xD(\sqrt{x}) = T^{(\alpha)}\big(\int \rho(\mathrm{d}t)\frac{x}{x-t} - 1\big)$ 这一事实，我们可以得到 $\int \hat{\rho}(\mathrm{d}t)\frac{x}{x-t} = \alpha T^{(\alpha),-1}(xD(\sqrt{x})) + 1$。因此，$g_{\lambda,\epsilon}^{(\alpha)}(x)$ 关于 x 的导数可以写成

$$\frac{\mathrm{d}g_{\lambda,\epsilon}^{(\alpha)}(x)}{\mathrm{d}x} = -\frac{1}{2x}\left(T^{(\alpha),-1}(xD_\epsilon(\sqrt{x})) - T^{(\alpha),-1}\left(\frac{\lambda x}{\alpha}\right)\right). \tag{47}$$

结合 (42)、(45)、(46) 和 (47)，可知当 $h\lambda_* \geq 1$ 且 $\lambda\lambda_* > \alpha$ 时（注意，根据猜想 1，在当前温度区间内，对数配分函数仅依赖于 λ_* ，因此由 (42) 可知，在其他温度区间， $M(\lambda, \lambda_*)$ 为零）， $M(\lambda, \lambda_*)$ 满足 (13)。

2) 计算 $Q(\lambda, \lambda_*)$：其次，我们将计算 $f_\epsilon^{(\alpha)}(\lambda, \lambda_*)$ 关于 ϵ 的导数。注意， $f_\epsilon^{(\alpha)}(\lambda, \lambda_*)$ 的三个部分都与 ϵ 有关，因此我们的计算也分为三个部分。当 $h\lambda_* \geq 1$ 且 $\lambda\lambda_* > \alpha$ 时，我们知道

$$\frac{\partial f_\epsilon^{(\alpha)}(\lambda,\lambda_*)}{\partial\epsilon} = \frac{\mathrm{d}g_{\lambda,\epsilon}^{(\alpha)}(x)}{\mathrm{d}x}\bigg|_{\epsilon=0,x=(D_0^{-1}(\frac{1}{\lambda_*}))^2} \frac{\partial(D_\epsilon^{-1}(\frac{1}{\lambda_*}))^2}{\partial\epsilon}\bigg|_{\epsilon=0} + \frac{\partial g_{\lambda,\epsilon}^{(\alpha)}(x)}{\partial\epsilon}\bigg|_{\epsilon=0,x=(D_0^{-1}(\frac{1}{\lambda_*}))^2}. \tag{48}$$

第一项很容易得到，因为

$$\frac{\partial(D_\epsilon^{-1}(\frac{1}{\lambda_*}))^2}{\partial\epsilon}\bigg|_{\epsilon=0} = 2\alpha C_0\left(\frac{1}{\lambda_*}\right)+\alpha+1.$$

我们来关注第二项。注意到，在 $g_{\lambda,\epsilon}^{(\alpha)}(x)$ 中，与 ϵ 相关的唯一一项是

$$\ell(x,\epsilon) = -\frac{1}{2}\int \rho_\epsilon(\mathrm{d}t)\ln(x-t) + \text{const},$$

其中我们使用了关系式 $\hat{\rho}_\epsilon = \alpha\rho_\epsilon + (1-\alpha)\delta_0$，并且最后一个常数可以忽略，因为我们只需计算 $\ell(x, \epsilon)$ 关于 ϵ 的导数。

令 $\boldsymbol{H}(t) = \boldsymbol{A} + \frac{1}{\sqrt{m}}\boldsymbol{G}(t) \in \mathbb{R}^{n\times m}$，其中 $\boldsymbol{A}$ 是确定性的，且 $\boldsymbol{G}(t)$ 的元素相互独立，均为布朗运动。特征值的动力学 $\lambda_1(t) \geq \lambda_2(t) \geq \cdots \geq \lambda_n(t)$ 已得到深入研究，被称为 Laguerre 过程或威沙特过程。具体而言， [33] 中的定理 2.1 给出，对于 $1 \leq i \leq n$，

$$\mathrm{d}\lambda_i(t) = 2\sqrt{\lambda_i}\frac{\mathrm{d}B_i}{\sqrt{m}} + \left(\frac{1}{m}\sum_{j:j\neq i}^{n}\frac{\lambda_i+\lambda_j}{\lambda_i-\lambda_j}+1\right)\mathrm{d}t, \tag{49}$$

其中 $B_1, B_2, \ldots, B_n$ 是相互独立的布朗运动。令

$$\ell_n(x,\{\lambda_i\}) = -\frac{1}{n}\sum_{i=1}^{n}\ln(x-\lambda_i).$$

那么，

$$\frac{\partial \ell_n}{\partial \lambda_i} = \frac{1}{n}\frac{1}{x-\lambda_i}, \quad \frac{\partial^2 \ell_n}{\partial \lambda_i^2} = \frac{1}{n}\frac{1}{(x-\lambda_i)^2}.$$

运用伊藤引理（参见 [34]中的（8.15）—（8.17））， 我们有

$$\begin{aligned}\mathrm{d}\ell_n =& \sum_{i=1}^{n}\frac{\partial \ell_n}{\partial \lambda_i}\mathrm{d}\lambda_i + \left(\sum_{i=1}^{n}\frac{2\lambda_i}{m}\frac{\partial^2 \ell_n}{\partial \lambda_i^2}\right)\mathrm{d}t \\ =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x-\lambda_i}\mathrm{d}t + \frac{1}{mn}\sum_{i,j:i\neq j}^{n}\frac{\lambda_i+\lambda_j}{(x-\lambda_i)(\lambda_i-\lambda_j)}\mathrm{d}t \\ &+ \frac{2}{mn}\sum_{i=1}^{n}\frac{\lambda_i}{(x-\lambda_i)^2}\mathrm{d}t + \frac{4}{n\sqrt{m}}\sum_{i=1}^{n}\frac{\sqrt{\lambda_i}}{x-\lambda_i}\mathrm{d}B_i. \quad (50)\end{aligned}$$

然后，第二项可简化为

$$\frac{1}{2mn}\sum_{i,j=1}^{n}\frac{\lambda_i+\lambda_j}{(x-\lambda_i)(x-\lambda_j)}\mathrm{d}t - \frac{1}{mn}\sum_{i=1}^{n}\frac{\lambda_i}{(x-\lambda_i)^2}\mathrm{d}t.$$

我们可以很容易地计算出

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda_i}{x-\lambda_i} = -1 - x\frac{\partial \ell_n}{\partial x}, \quad \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x-\lambda_i} = -\frac{\partial \ell_n}{\partial x}.$$

将该方程代入（50），可得

$$\begin{aligned}\mathrm{d}\ell_n =& \frac{4}{n\sqrt{m}}\sum_{i=1}^{n}\frac{\sqrt{\lambda_i}}{x-\lambda_i}\mathrm{d}B_i + \frac{2}{mn}\sum_{i=1}^{n}\frac{\lambda_i}{(x-\lambda_i)^2}\mathrm{d}t \\ &- \frac{\partial \ell_n}{\partial x}\mathrm{d}t + \alpha\frac{\partial \ell_n}{\partial x}\left(1 + x\frac{\partial \ell_n}{\partial x}\right)\mathrm{d}t. \quad (51)\end{aligned}$$

现在我们对等式两边取期望，并注意到第一项已经为零。然后，我们还注意到，当 $n\to\infty$时，第二项将消失，因此可得

$$\mathbb{E}[\mathrm{d}\ell_n] = \mathbb{E}\left[\alpha x\left(\frac{\partial \ell_n}{\partial x}\right)^2 + (\alpha-1)\frac{\partial \ell_n}{\partial x}\right]\mathrm{d}t + O\left(\frac{1}{n}\right).$$

因此，当 $\ell = \lim_{n\to\infty}\ell_n$时，可得

$$\frac{\partial \ell}{\partial t} = \alpha x\left(\frac{\partial \ell}{\partial x}\right)^2 + (\alpha-1)\frac{\partial \ell}{\partial x}. \quad (52)$$

利用该方程，我们可以计算得出

$$\left.\frac{\partial \ell(x,\epsilon)}{\partial \epsilon}\right|_{\epsilon=0, x=(D_0^{-1}(\frac{1}{\lambda_*}))^2} = \left.\frac{1}{2}D_0(\sqrt{x})\right|_{x=(D_0^{-1}(\frac{1}{\lambda_*}))^2} = \frac{1}{2\lambda_*}.$$

因此，将上述方程代入（48），可得（13）中的 $Q(\lambda,\ \lambda_*)$，其中 $h\lambda_*\geq 1$ 且 $\lambda\lambda_*>\alpha$。

B. 低温区间 2：$h\lambda_* < 1$ 且 $\lambda > \alpha h$

接下来，当 $h\lambda_*<1$ 且 $\lambda>\alpha h$ 时，我们可以得到

$$\begin{aligned}\frac{\partial f_\epsilon^{(\alpha)}(\lambda,\lambda_*)}{\partial \epsilon} =& \left.\frac{\mathrm{d}g_{\lambda,\epsilon}^{(\alpha)}(x)}{\mathrm{d}x}\right|_{\epsilon=0,x=\bar{\gamma}_0^2}\cdot\left.\frac{\partial \bar{\gamma}_\epsilon^2}{\partial \epsilon}\right|_{\epsilon=0} \\ &+ \left.\frac{\partial g_{\lambda,\epsilon}^{(\alpha)}(x)}{\partial \epsilon}\right|_{\epsilon=0,x=\bar{\gamma}_0^2}. \quad (53)\end{aligned}$$

我们需要计算的唯一一项是 $\partial_\epsilon\bar{\gamma}_\epsilon^2$。由式（49）可知

$$\frac{\mathrm{d}\lambda_1}{\mathrm{d}t} = \frac{1}{m}\sum_{j=2}^{n}\frac{\lambda_1+\lambda_j}{\lambda_1-\lambda_j} + 1.$$

因此，我们可以得到

$$\begin{aligned}\left.\frac{\mathrm{d}\bar{\gamma}_\epsilon^2}{\mathrm{d}\epsilon}\right|_{\epsilon=0} &= \lim_{m\to\infty}\left(\frac{1}{m}\sum_{j=2}^{n}\frac{\lambda_1+\lambda_j}{\lambda_1-\lambda_j}+1\right) \\ &= \lim_{z\to\bar{\gamma}_0^2}\int\hat{\rho}(\mathrm{d}t)\frac{z+t}{z-t} + \alpha \\ &= 2\alpha T^{(\alpha),-1}(\bar{\gamma}_0^2 D_0(\bar{\gamma}_0^+)) + \alpha + 1. \quad (54)\end{aligned}$$

因此，结合（53）和（54），在该区间下，即当 $h\lambda_*<1$ 且 $\lambda>\alpha h$时，我们得到（13）中的 $Q(\lambda,\ \lambda_*)$，其中 $\bar{\ }$ $\bar{\ }$我们回顾一下，$h=h_0=\lim_{z\downarrow\bar{\gamma}_0}D_0(z)$，并且我们已经假设其有限。

C. 高温区间：其他情况

最后一种情况很容易计算。在广义观测模型中使用矩形自由卷积（见[29]中的定理 3.12 和 3.13），并注意到 $\sqrt{\epsilon}\boldsymbol{W}$的矩形 R 变换为 ϵt，我们可以得到$C_\epsilon^{(\alpha)}(t) = C_0^{(\alpha)}(t) + \epsilon t$。因此，

$$\frac{\partial f_\epsilon^{(\alpha)}(\lambda,\lambda_*)}{\partial \epsilon} = \int_0^{\sqrt{\frac{\lambda}{\alpha}}}\frac{t^2}{t}\mathrm{d}t = \frac{\lambda}{2\alpha}. \quad (55)$$

这表明，在高温区间内，$Q(\lambda,\ \lambda_*) = 0$。

附录 D
近似消息传递的证明

A. 辅助 AMP 及其状态演化

为了证明高斯 AMP 的状态演化结果，我们定义如下所谓的辅助 AMP：

$$\begin{aligned}\tilde{\boldsymbol{z}}^t &= \boldsymbol{Z}^\mathsf{T}\tilde{\boldsymbol{u}}^t - \sum_{i=1}^{t-1}b_{ti}^{\mathrm{A}}\tilde{\boldsymbol{v}}^i, \qquad \tilde{\boldsymbol{v}}^t = \tilde{v}_t(\tilde{\boldsymbol{z}}^1,\ldots,\tilde{\boldsymbol{z}}^t), \\ \tilde{\boldsymbol{y}}^t &= \boldsymbol{Z}\tilde{\boldsymbol{v}}^t - \sum_{i=1}^{t}a_{ti}^{\mathrm{A}}\tilde{\boldsymbol{u}}^i, \quad \tilde{\boldsymbol{u}}^{t+1} = \tilde{u}_{t+1}(\tilde{\boldsymbol{y}}^1,\ldots,\tilde{\boldsymbol{y}}^t), \quad (56)\end{aligned}$$

其中 $\tilde{\boldsymbol{u}}^1 = \boldsymbol{u}^1$ 见式（16）。非线性函数 $\tilde{v}_t$ 和 $\tilde{u}_{t+1}$递归定义如下

$$\begin{aligned}\tilde{v}_t(z_1,\ldots,z_t,v_*) =& v_t\Big(-\alpha\bar{\beta}_t\tilde{v}_{t-1}(z_1,\ldots,z_{t-1},v_*) \\ &+ z_t + \bar{\nu}_t v_* + \sum_{i=1}^{t-1}(\bar{\boldsymbol{B}}_t)_{t,i}\tilde{v}_i(z_1,\ldots,z_i,v_*)\Big), \quad (57)\end{aligned}$$

$$\begin{aligned}\tilde{u}_{t+1}(y_1,\ldots,y_t,u_1,u_*) =& u_{t+1}\Big(-\bar{\alpha}_t\tilde{u}_t(y_1,\ldots,y_{t-1},u_1,u_*) \\ &+ y_t + \bar{\mu}_t u_* + \sum_{i=1}^{t}(\bar{\boldsymbol{A}}_t)_{t,i}\tilde{u}_i(y_1,\ldots,y_{i-1},u_1,u_*)\Big), \quad (58)\end{aligned}$$

$\bar{\ }$ $\bar{\ }$其中 $\tilde{u}_1(u_1,u_*) := u_1$。参数 $(\bar{\nu}_t,\beta_t,\boldsymbol{B}_t)$ 和 $\bar{\ }$$(\bar{\mu}_t,\bar{\alpha}_t,\boldsymbol{A}_t)$ 来自真实的高斯 AMP。我们可以用与 (19) 相同的方法得到系数 $\{b_{t,i}^{\mathrm{A}}\}_{i=1}^{t-1}$ 和 $\{a_{t,i}^{\mathrm{A}}\}_{i=1}^{t}$

将 $\bar{\boldsymbol{\Delta}}_t$、$\bar{\boldsymbol{\Gamma}}_t$、$\bar{\boldsymbol{\Phi}}_t$、$\bar{\boldsymbol{\Psi}}_t$ 替换为 $\boldsymbol{\Delta}_t^{\mathrm{A}}$、$\boldsymbol{\Gamma}_t^{\mathrm{A}}$、$\boldsymbol{\Phi}_t^{\mathrm{A}}$、$\boldsymbol{\Psi}_t^{\mathrm{A}}$，它们定义如下:

$$
\begin{aligned}
&(\boldsymbol{\Delta}_t^{\mathrm{A}})_{ij} = \langle \tilde{\boldsymbol{u}}^i \tilde{\boldsymbol{u}}^j \rangle, \ (\boldsymbol{\Gamma}_t^{\mathrm{A}})_{ij} = \langle \tilde{\boldsymbol{v}}^i \tilde{\boldsymbol{v}}^j \rangle, \ 1 \le i, j \le t, \\
&(\boldsymbol{\Phi}_t^{\mathrm{A}})_{ij} = \langle \partial_j \tilde{\boldsymbol{u}}^i \rangle, \ 1 \le i < j \le t, \\
&(\boldsymbol{\Psi}_t^{\mathrm{A}})_{ij} = \langle \partial_j \tilde{\boldsymbol{v}}^i \rangle, \ 1 \le i \le j \le t.
\end{aligned} \tag{59}
$$

这里，$\tilde{\boldsymbol{u}}^i \tilde{\boldsymbol{u}}^j$，$\partial_j \tilde{\boldsymbol{u}}^i \in \mathbb{R}^n$ 分别表示逐元素乘积和关于 z_j 的偏导数，而 $\langle \boldsymbol{u} \rangle = (\sum_{i=1}^n u_i)/n$ 表示一个向量的均值。因此，由[37], 可得辅助 AMP 的状态演化如下:

$$
\begin{aligned}
&(\tilde{Z}_1, \dots, \tilde{Z}_t) \sim \mathcal{N}(0, \bar{\boldsymbol{\Omega}}_t^{\mathrm{A}}), \ \tilde{V}_t = \tilde{v}_t(\tilde{Z}_1, \dots, \tilde{Z}_t, V_*), \\
&(\tilde{Y}_1, \dots, \tilde{Y}_t) \sim \mathcal{N}(0, \bar{\boldsymbol{\Sigma}}_t^{\mathrm{A}}), \ \tilde{U}_{t+1} = \tilde{u}_{t+1}(\tilde{Y}_1, \dots, \tilde{Y}_t, \tilde{U}_1, U_*),
\end{aligned} \tag{60}
$$

其中，随机变量 $U_1 = U_1$ 见式 (17)。此外，系数可表示为

$$
\begin{aligned}
&(\bar{\boldsymbol{\Delta}}_t^{\mathrm{A}})_{ij} = \mathbb{E}[\tilde{U}_i \tilde{U}_j], \ (\bar{\boldsymbol{\Gamma}}_t^{\mathrm{A}})_{ij} = \mathbb{E}[\tilde{V}_i \tilde{V}_j], \ 1 \le i, j \le t, \\
&(\bar{\boldsymbol{\Phi}}_t^{\mathrm{A}})_{ij} = \mathbb{E}[\partial_j \tilde{U}_i], \ 1 \le i < j \le t, \\
&(\bar{\boldsymbol{\Psi}}_t^{\mathrm{A}})_{ij} = \mathbb{E}[\partial_j \tilde{V}_i], \ 1 \le i \le j \le t.
\end{aligned} \tag{61}
$$

通过在式 (19) 中将 $\bar{\boldsymbol{\Delta}}_t$、$\bar{\boldsymbol{\Gamma}}_t$、$\bar{\boldsymbol{\Phi}}_t$、$\bar{\boldsymbol{\Psi}}_t$ 分别替换为 $\bar{\boldsymbol{\Delta}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{\Gamma}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{\Phi}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{\Psi}}_t^{\mathrm{A}}$，即可定义矩阵 $\bar{\boldsymbol{\Omega}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{\Sigma}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{A}}_t^{\mathrm{A}}$、$\bar{\boldsymbol{B}}_t^{\mathrm{A}}$。由 [37], 可得以下关于上文定义的辅助 AMP 的状态演化命题。

命题 1. （辅助 *AMP* 的状态演化）。考虑(56) 中的辅助 *AMP* 及其(60)中的状态演化。令 $\tilde{\psi}: \mathbb{R}^{t+2} \to \mathbb{R}$ 和 $\tilde{\phi}: \mathbb{R}^{t+1} \to \mathbb{R}$ 为任意的二阶伪 *Lipschitz* 函数。则对于每个 $t \ge 1$，我们几乎必然有

$$
\begin{aligned}
&\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^m \tilde{\phi}\big((\tilde{\boldsymbol{z}}^1)_i, \dots, (\tilde{\boldsymbol{z}}^t)_i, (\tilde{\boldsymbol{v}}^1)_i, \dots, (\tilde{\boldsymbol{v}}^t)_i, (\boldsymbol{v}^*)_i\big) \\
&\qquad = \mathbb{E}\tilde{\phi}\big(\tilde{Z}_1, \dots, \tilde{Z}_t, \tilde{V}_1, \dots, \tilde{V}_t, V_*\big),
\end{aligned} \tag{62}
$$

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \tilde{\psi}\big((\tilde{\boldsymbol{y}}^1)_i, \dots, (\tilde{\boldsymbol{y}}^t)_i, (\tilde{\boldsymbol{u}}^1)_i, \dots, (\tilde{\boldsymbol{u}}^{t+1})_i, (\boldsymbol{u}^*)_i\big) \\
&\qquad = \mathbb{E}\tilde{\psi}\big(\tilde{Y}_1, \dots, \tilde{Y}_t, \tilde{U}_1, \dots, \tilde{U}_{t+1}, U_*\big).
\end{aligned} \tag{63}
$$

B. 证明细节

我们将利用辅助 AMP 及其状态演化来完成证明。我们的策略与方阵情形[28],类似，因此只给出证明概要。第一步是证明

$$
(\tilde{Z}_1, \dots, \tilde{Z}_t) \overset{\mathrm{d}}{=} (Z_1, \dots, Z_t), \ (\tilde{Y}_1, \dots, \tilde{Y}_t) \overset{\mathrm{d}}{=} (Y_1, \dots, Y_t), \tag{64}
$$

其中，左侧的随机变量定义于 (60)，右侧的随机变量定义于 (17)。结合两个状态演化以及非线性函数 $\tilde{v}_t$、$\tilde{u}_{t+1}$ 的定义，(64) 可通过与 [28] 中方阵情形类似的归纳论证得出。

然后，第二步是证明，对于任意两个阶数为 2 的伪 Lipschitz 函数 $\psi: \mathbb{R}^{t+2} \to \mathbb{R}$ 和 $\phi: \mathbb{R}^{t+1} \to \mathbb{R}$，以下极限对 $t \ge 1$ 几乎必然成立:

$$
\begin{aligned}
&\lim_{m\to\infty} \Big| \frac{1}{m} \sum_{i=1}^m \phi\big((\tilde{\boldsymbol{g}}^1)_i, \dots, (\tilde{\boldsymbol{g}}^t)_i, (\tilde{\boldsymbol{v}}^1)_i, \dots, (\tilde{\boldsymbol{v}}^t)_i, (\tilde{\boldsymbol{v}}^*)_i\big) \\
&- \frac{1}{m} \sum_{i=1}^m \phi\big((\boldsymbol{g}^1)_i, \dots, (\boldsymbol{g}^t)_i, (\boldsymbol{v}^1)_i, \dots, (\boldsymbol{v}^t)_i, (\boldsymbol{v}^*)_i\big) \Big| = 0,
\end{aligned} \tag{65}
$$

$$
\begin{aligned}
&\lim_{n\to\infty} \Big| \frac{1}{n} \sum_{i=1}^n \psi\big((\tilde{\boldsymbol{f}}^1)_i, \dots, (\tilde{\boldsymbol{f}}^t)_i, (\tilde{\boldsymbol{u}}^1)_i, \dots, (\tilde{\boldsymbol{u}}^{t+1})_i, (\tilde{\boldsymbol{u}}^*)_i\big) \\
&- \frac{1}{n} \sum_{i=1}^n \psi\big((\boldsymbol{f}^1)_i, \dots, (\boldsymbol{f}^t)_i, (\boldsymbol{u}^1)_i, \dots, (\boldsymbol{u}^{t+1})_i, (\boldsymbol{u}^*)_i\big) \Big| = 0,
\end{aligned} \tag{66}
$$

其中，我们对 $s \in \{1, 2, \dots, t\}$ 定义:

$$
\begin{aligned}
\tilde{\boldsymbol{g}}^s &= \tilde{\boldsymbol{z}}^s + \bar{\nu}_s \boldsymbol{v}^* - \alpha \bar{\beta}_s \tilde{\boldsymbol{v}}^{s-1} + \sum_{i=1}^{s-1} (\bar{\boldsymbol{B}}_s)_{s,i} \tilde{\boldsymbol{v}}^i, \\
\tilde{\boldsymbol{f}}^s &= \tilde{\boldsymbol{y}}^s + \bar{\mu}_s \boldsymbol{u}^* - \bar{\alpha}_s \tilde{\boldsymbol{u}}^s + \sum_{i=1}^{s} (\bar{\boldsymbol{A}}_s)_{s,i} \tilde{\boldsymbol{u}}^i.
\end{aligned}
$$

该结论可通过应用柯西-施瓦茨不等式及另一个归纳论证得出，这也与 [28] 中分析的方阵情形类似。

最后一步中，利用命题 1 以及第二步的结论 (65) 和 (66)，我们得到

$$
\begin{aligned}
&\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^m \phi\big((\boldsymbol{g}^1)_i, \dots, (\boldsymbol{g}^t)_i, (\boldsymbol{v}^1)_i, \dots, (\boldsymbol{v}^t)_i, (\boldsymbol{v}^*)_i\big) \\
&\qquad = \mathbb{E}\phi\big(\tilde{G}_1, \dots, \tilde{G}_t, \tilde{V}_1, \dots, \tilde{V}_t, V_*\big),
\end{aligned} \tag{67}
$$

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \psi\big((\boldsymbol{f}^1)_i, \dots, (\boldsymbol{f}^t)_i, (\boldsymbol{u}^1)_i, \dots, (\boldsymbol{u}^{t+1})_i, (\boldsymbol{u}^*)_i\big) \\
&\qquad = \mathbb{E}\psi\big(\tilde{F}_1, \dots, \tilde{F}_t, \tilde{U}_1, \dots, \tilde{U}_{t+1}, U_*\big),
\end{aligned} \tag{68}
$$

其中，我们定义了 $s \in \{1, \dots, t\}$,

$$
\begin{aligned}
\tilde{G}_s &= \tilde{Z}_s + \bar{\nu}_s V_* - \alpha \bar{\beta}_s \tilde{V}_{s-1} + \sum_{i=1}^{s-1} (\bar{\boldsymbol{B}}_s)_{s,i} \tilde{V}_i, \\
\tilde{F}_s &= \tilde{Y}_s + \bar{\mu}_s U_* - \bar{\alpha}_s \tilde{U}_s + \sum_{i=1}^{s} (\bar{\boldsymbol{A}}_s)_{s,i} \tilde{U}_i.
\end{aligned}
$$

最后，利用（64），我们得到

$$
\begin{aligned}
&\mathbb{E}\phi\big(G_1, \dots, G_t, V_1, \dots, V_t, V_*\big) \\
&\qquad = \mathbb{E}\phi\big(\tilde{G}_1, \dots, \tilde{G}_t, \tilde{V}_1, \dots, \tilde{V}_t, V_*\big),
\end{aligned} \tag{69}
$$

$$
\begin{aligned}
&\mathbb{E}\psi\big(F_1, \dots, F_t, U_1, \dots, U_{t+1}, U_*\big) \\
&\qquad = \mathbb{E}\psi\big(\tilde{F}_1, \dots, \tilde{F}_t, \tilde{U}_1, \dots, \tilde{U}_{t+1}, U_*\big).
\end{aligned} \tag{70}
$$

结合上述等式（67）—（70），最终得到（17），证明至此完成。