arXiv:2302.03682v1 [math.ST] 7 Feb 2023

从随机初始化开始的近似消息传递及其在 $\mathbb{Z}_2$ 同步中的应用

GenLi Wei Fan Yuting Wei

宾夕法尼亚大学沃顿商学院统计与数据科学系，宾夕法尼亚州费城

2023年2月8日

摘要

本文研究如何利用带噪观测和先验结构信息重构未知的秩一矩阵。由于贝叶斯最优估计量具有非凸性，通常难以计算，而近似消息传递（AMP）则成为一种高效的一阶方法，可用于近似贝叶斯最优估计量。然而，当 AMP 从随机初始化开始时，其理论基础仍鲜有研究，而这种方案在实践中具有重要价值。我们聚焦于一种称为 $\mathbb{Z}_2$ 同步的典型模型，刻画了 AMP 从随机初始化开始时的有限样本动力学，并揭示了其快速全局收敛性。我们的理论首次在无需信息性初始化（例如谱初始化）或样本拆分的情况下，对该模型中的 AMP 给出了非渐近刻画。

关键词： 近似消息传递，随机初始化，非渐近分析， $\mathbb{Z}_2$ 同步，尖峰 Wigner 模型，全局收敛

目录

1 引言 **2**

1.1 本文：用于 $\mathbb{Z}_2$ 同步的随机初始化 AMP.3 1.2 主要贡献与技术挑战.4 1.3 符号.4

2 主要结果 **5**

3 定理1的证明 **7**

3.1 预备知识.7 3.2 AMP动力学的非渐近分析.9

4 讨论 19

辅助引理和断言的证明 19

引理的证明 2 .19 引理的证明 3 .22 引理的证明 4 .25 命题的证明 (43).32 命题的证明 (53).36 命题的证明 (56).39 命题的证明 (59).39 引理的证明 5 .40 不等式的证明 (70).43

1 引言

在能够获得带噪观测的情况下，估计未知结构化信号 $v^\star \in \mathbb{R}^n$ 的问题

$$M = \lambda v^\star v^{\star\top} + W \in \mathbb{R}^{n\times n} \qquad \text{with } \lambda > 0 \tag{1}$$

具有根本性的研究意义，并已在各种不同背景下得到广泛探讨（Abbe等人,2020;Candesand Plan, 2010; Chi 等人, 2019; Johnstone, 2001;Keshavan 等人, 2009; Singer, 2011; Zhong和Boumal,2018）。当噪声矩阵 $W = [W_{ij}]_{1\le i,j\le n}$ 的各个元素独立地服从高斯分布时，更确切地说， $W_{ii} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{2}{n})$ 和 $W_{ji} = W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{1}{n})$，其中 $i \neq j$——该模型通常称为变形 Wigner 模型或尖峰 Wigner 模型，是理解低秩矩阵估计可行性及其基本极限的典型模型。

观测矩阵 M 的谱性质已得到广泛研究（例如，Baik et al. (2005); Capitaine et al. (2009); Cheng et al. (2021); Féral and Péché (2007); Péché (2006)），这推动了在 $v^\star$ 不含结构信息时设计谱方法（Cai and Zhang, 2018; Chen et al., 2021; Keshavan et al., 2009; Singer, 2011; Yan et al., 2021）。在实践中，有许多应用场景可事先获得关于 $v^\star$ 的额外结构信息，例如有限群结构（Perry et al., 2018）、锥约束（Deshpande et al., 2014; Lesieur et al., 2017）、稀疏性（Berthet and Rigollet, 2013; Johnstone and Lu, 2009）等。先验结构的存在进一步加剧了计算最大似然估计或贝叶斯最优估计时的非凸性问题，因此迫切需要设计能够高效执行的算法。值得注意的是，近似消息传递（AMP）算法作为一种高效的非凸范式应运而生，能够应对上述挑战（Donoho et al., 2009; Feng et al., 2022）。AMP 最初是在压缩感知背景下提出的，它不仅是一类具有快速收敛特性的一阶迭代算法（Bayati and Montanari, 2011a; Celentano and Montanari, 2022; Donoho et al., 2010; Fan, 2022; Rangan, 2011），也是一种强有力的工具，可用于确定其他统计程序在高维渐近下的表现（Bayati and Montanari, 2011b; Bu et al., 2020; Donoho and Montanari, 2016; Javanmard et al., 2016; Lelarge and Miolane, 2019; Li and Wei, 2021; Ma et al., 2018; Sur and Candès, 2019; Sur et al., 2019; Zhang et al., 2022）。

然而，尽管现有的近似消息传递理论涵盖了大量应用，但至少在两个方面仍显不足。首先，现有近似消息传递理论中占主导地位的部分具有渐近性，其含义是：对于任意 n 固定的 迭代 次数 t，它预测的是近似消息传递动力学在大规模极限下的表现。因此，已有的近似消息传递理论无法描述经过不断增加的迭代次数后近似消息传递的行为，这与其他基于优化的过程形成对比；后者通常具有能够适应大量迭代次数的非渐近分析（Chen 和 Wainwright，2015年；Chi 等人，2019年；Keshavan 等人，2009；Ma 等人，2020）。另一个问题源于对信息性初始化的要求，即现有针对低秩估计的近似消息传递理论通常要求从一个与真实信号已经具有非消失相关性的点开始（Celentano 等人，2021 年；Montanari 和 Venkataramanan，2021 年；Zhong 等人，2021 年）。尽管信息性的初始估计（如谱初始化）有时是合理且可分析的，但这一要求阻碍了我们理解其他广泛采用的替代方案（如随机初始化）的影响。正如我们即将明确说明的那样，解决这一问题可能还需要为近似消息传递建立一种新的非渐近框架，因为当各次迭代仅与真实向量表现出极弱的相关性时，追踪近似消息传递动力学十分困难。

受上述问题的启发，人们对理解近似消息传递（AMP）的有限样本性能日益关注。此类研究中的首项工作是Rush 和 Venkataramanan (2018年)（另见后续工作Cademartori 和 Rush (2023)），该研究考察了用于稀疏回归的 AMP，并允许迭代总次数高达 $o\big(\frac{n}{\log n \log\log n}\big)$。近期，Li 和 Wei (2022年)针对尖峰 Wigner 模型建立了一个非渐近分析框架，刻画了 AMP 至多进行$O\big(\frac{n}{\mathsf{poly}(\log n)}\big)$次迭代时的行为。尽管其中的理论非常适用于谱初始化近似消息传递，但随机初始化的 AMP 能否达到与信息性初始化的 AMP 相同的性能，仍是一个悬而未决的问题。

1.1 本文：用于 $\mathbb{Z}_2$ 同步的随机初始化 AMP

在本文中，我们尝试通过研究一个称为 $\mathbb{Z}_2$ 同步的具体模型来应对上述挑战。具体而言， $\mathbb{Z}_2$ 同步是尖峰 Wigner 模型的一个特例，其中真实值具有遵循 $v^\star \in \{\pm \frac{1}{\sqrt{n}}\}^n$ 的离散结构。在本文及下文中，我们对 $v^\star = [v_i^\star]_{1\le i\le n}$ 施加先验分布，使得

$$v_i^\star \overset{\text{i.i.d.}}{\sim} \mathsf{Unif}\Big(\pm\frac{1}{\sqrt{n}}\Big), \qquad 1\le i\le n.$$

目标是根据观测值重构 $v^\star$，这些观测值 M (见（1）)。该问题可视为一个更一般问题的基本示例——紧群上的同步（Abbe等人，2020；Chen和Candès，2018年；Gao和Zhang，2022年；Perry等人，2018年；Singer，2011年；Zhong和Boumal，2018年），并且与随机块模型密切相关（Deshpande 等人，2017 年；Lelarge和Miolane，2019年）。

近似消息传递算法。 由于底层优化问题具有组合性质，通常难以直接计算贝叶斯最优解。因此，我们需要寻找计算上可行的替代方案，而近似消息传递（AMP）自然成为一种有效的选择（Celentano 等人，2021 年；Deshpande 等人，2017 年；Fan等人，2021 年；Li 和 Wei，2022年）。具体而言，给定初始化点 x_0， $x_1\in\mathbb{R}^n$，针对 $\mathbb{Z}_2$ 同步的近似消息传递算法采用以下更新规则：

$$x_{t+1} = M\eta_t(x_t) - \langle \eta_t'(x_t)\rangle \eta_{t-1}(x_{t-1}), \qquad t\ge 1, \tag{2}$$

其中，对于任意向量 $x\equiv[x_i]_{1\le i\le n}$，我们记 $\langle x\rangle := \frac{1}{n}\sum_{i=1}^n x_i$，去噪函数定义为[1]

$$\begin{aligned} \eta_t(x) &= \gamma_t \tanh(\pi_t x), \qquad \text{for } t\ge 1 \\ \text{with}\quad \pi_t &:= \sqrt{\max\big\{n(\|x_t\|_2^2-1),1\big\}} \quad \text{and} \quad \gamma_t := \|\tanh(\pi_t x_t)\|_2^{-1}. \end{aligned} \tag{3}$$

这里约定，若输入参数是向量，则 $\eta_t(\cdot)$、 $\eta_t'(\cdot)$ 和双曲正切$(\cdot)$ 均逐元素作用。

迄今为止，有两种分析策略能够在最具挑战性的情形（即 $\lambda>1$）下适用于迭代次数不断增加的情况。其中一种尝试由 Celentano 等人(2021 年) 提出，他们设计了一种三阶段混合算法：先运行谱初始化近似消息传递，再进行自然梯度下降（NGD）。他们当时猜想，第三阶段（即 NGD）并非必要。最近，Li 和 Wei (2022年) 提出了另一种策略来验证这一猜想，并表明仅靠谱初始化近似消息传递就已足够，无需第三个细化阶段。尽管底层优化问题具有非凸性，采用谱初始化的近似消息传递仍能达到几乎与贝叶斯最优估计相同的效果。

随机初始化。如前所述，针对这一问题，现有的所有近似消息传递理论（Celentano 等人，2021 年；Li 和 Wei，2022 年；Mondelli 和 Venkataramanan，2021 年；Montanari 和 Venkataramanan，2021 年）都要求采用信息性初始化，例如通过谱方法获得。这引出了以下自然问题：

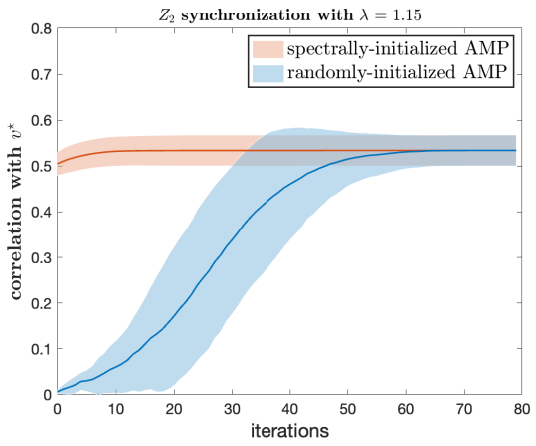
> *AMP*要取得成功，是否必须进行谱初始化？我们能否采用更简单的初始化方案，同时仍能达到与谱初始化同样好的效果？

一种广泛采用的初始化策略是对AMP进行随机初始化；例如，

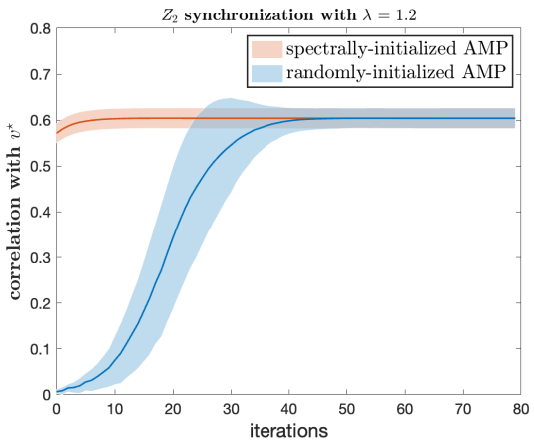
$$x_1 \sim \mathcal{N}\Big(0, \frac{1}{n}I_n\Big) \qquad (\text{independent of } M) \qquad \text{and} \qquad \eta_0(x_0)=0. \tag{4}$$

为回答上述问题，我们先使用（4）开展一系列数值实验，结果如图1所示。令人振奋的是，随机初始化的AMP似乎表现出乎意料地好：只需几十次迭代，便能达到几乎与谱初始化近似消息传递相同的性能（请注意，谱初始化本身也包含几十次幂迭代）。这些令人鼓舞的数值结果促使我们深入探究随机初始化对AMP收敛的影响；这一问题在现有文献中尚未得到充分理解。

[1]请注意，为便于分析，我们采用的缩放方式与 Deshpande 等人 (2017 年) 略有不同，但二者仅相差一个全局缩放因子，因而等价。



(a) 使用 $\lambda = 1.15$ 的近似消息传递



(b) 使用 $\lambda = 1.2$ 的近似消息传递

图 1. 相关性 $\eta_t(x_t)$ 与$v^\star$ (即 $|\langle \eta_t(x_t), v^\star \rangle|/(\|\eta_t(x_t)\|_2 \|v^\star\|_2)$) 随迭代次数 t 的变化，其中近似消息传递分别采用随机初始化和谱初始化。这里，$n = 10000$ 和 $v_i^\star \overset{\text{i.i.d.}}{\sim}$ 均匀分布$(\pm \frac{1}{\sqrt{n}})$ $(1 \le i \le n)$。我们根据 (1) 生成 20 个 的独立副本并报告平均结果，阴影区域的宽度表示（两倍）标准差。图 (a) 和 (b) 分别对应 和 。

1.2 主要贡献与技术挑战

本文提供了一种非渐近分析方法，可以预测近似消息传递如何从随机初始化出发随时间演化。我们的理论能够跟踪近似消息传递迭代与真实向量之间的相关性 $v^\star$。具体而言，我们在定理 1 中证明，在初始阶段，近似消息传递迭代中的信号分量会以指数速度增长，从 $O(\frac{\log n}{\lambda - 1})$ 增长到 $\widetilde{O}(\frac{1}{\sqrt{n}})$，所需迭代次数不超过 $O(\sqrt{\lambda^2 - 1})$；后者与谱初始化和真实向量之间的相关性一致。此外，一旦信号分量的幅值超过 $\frac{1}{2}\sqrt{\lambda^2 - 1}$，渐近分析中推导出的状态演化递推便能很好地预测有限样本AMP动力学。据我们所知，本文首次给出了刻画近似消息传递从随机初始化出发时性能的结果，为实践中采用随机初始化提供了依据并予以支持。

基于 Li 和 Wei (2022年) 最近提出的分析方法，我们的理论发展需要在该框架之外提出新思路，以追踪从随机初始化开始的近似消息传递。在继续之前，我们先指出需要克服的关键挑战。

- 基于状态演化分析的既有理论，无法对与真实向量的相关性趋于零时的迭代过程提供“细粒度”的理解。更确切地说，过去的理论无法衡量近似消息传递在初始阶段的进展，此时其信号分量为 $o(1)$（事实上，初始化时甚至小至 $\widetilde{O}(\frac{1}{\sqrt{n}})$），而是在大-$n$ 极限下将信号强度视为 0。

- 另一个技术难点源于迭代之间复杂的统计依赖；当算法从随机初始化开始，且迭代次数随维度增长时，这一问题尤其难以处理 n。以往文献通过精巧的留一法去耦论证（例如参见Chen 等人 (2019年)）或全局景观分析（例如参见Ge 等人 (2017 年）来处理其他非凸优化方法中的这一问题，但这些方法在分析近似消息传递时仍然无法使用。

1.3 符号说明

本文中，我们用 $\varphi(\cdot)$（分别用 $\varphi_n(\cdot)$）表示标准高斯随机变量（分别表示高斯随机向量 $\mathcal{N}(0, I_n)$）的概率密度函数（p.d.f.）。对于任意矩阵 M，我们令$\|M\|$和 $\|M\|_{\mathsf{F}}$分别表示 M的谱范数和弗罗贝尼乌斯范数。除非另有说明，

对于任意向量 $x \in [x_i]_{1\le i\le n} = \mathbb{R}^n$，我们用 $|x|_{(i)}$（分别用 $x_{(i)}$）表示 x 中绝对值（分别是数值）第 i 大的元素。我们将 $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$ 记作 $\mathbb{R}^d$中的单位球面。此外，对于任意两个向量 x，$y \in \mathbb{R}^n$，我们将 $x \circ y$记作它们的克罗内克积，即 $x \circ y = (x_1y_1、\ldots\ldots、x_ny_n)^\top \in \mathbb{R}^n$. 当一个函数作用于向量时，应理解为以逐分量方式作用；例如，对于任意向量 $x = [x_i]_{1\le i\le n}$，我们令 $x + 1 := [x_i + 1]_{1\le i\le n}$。

此外，给定两个函数 $f(n)$ 和 $g(n)$，我们写作 $f(n) \lesssim g(n)$ 或 $f(n) = O(g(n))$，表示对于某个与 n 无关的普适常数 $c_1 > 0$，有 $|f(n)| \le c_1 g(n)$；类似地，$f(n) \gtrsim g(n)$ 表示~ $f(n) \ge c_2|g(n)|$，其中的常数 $c_2 > 0$ 为普适常数。若 $f(n) = O(g(n))$ 在对数因子意义下成立，我们写作 $f(n) = O(g(n))$。我们还采用记号 $f(n) \asymp g(n)$，表示 $f(n) \lesssim g(n)$ 和 $f(n) \gtrsim g(n)$ 同时成立。此外，当我们写作 $f(n) \ll g(n)$ 或 $f(n) = o(g(n))$ 时，表示 $f(n)/g(n) \to 0$ as$n \to \infty$；我们也写作 $f(n) \gg g(n)$，若 $g(n)/f(n) \to 0$as $n \to \infty$。我们用 c, C 表示与 n 无关的普适常数；注意，常数可能随行而变。

2 主要结果

本节将精确陈述随机初始化的 AMP 的主要理论保证。为方便记号，定义

$$\alpha_{t+1} := \lambda v^{\star\top}\eta_t(x_t), \tag{5}$$

它刻画了第 t 次迭代（去噪后）在真实向量 $v^\star$ 方向上的投影。从某种意义上说，这个量反映了第 t 次迭代所携带的信号分量的大小。基于这一记号，我们定义一个关键阈值如下:

$$\varsigma := \min\left\{t : \alpha_t \ge \frac{1}{2}\sqrt{\lambda^2 - 1}\right\}, \tag{6}$$

它反映了 AMP 迭代携带显著信号分量所需的时间（注意，随机初始猜测满足 $|v^{\star\top}x_1| \lesssim \widetilde{O}(\frac{1}{\sqrt{n}})$，这意味着初始信号分量极小）。此外，我们定义从第 ς 次迭代开始的状态演化递推如下:

$$\alpha_\varsigma^\star = |\alpha_\varsigma| \qquad \text{and} \qquad \alpha_{t+1}^\star = \lambda\left[\int \tanh\left(\alpha_t^\star\left(\alpha_t^\star + x\right)\right)\varphi(\mathrm{d}x)\right]^{1/2} \qquad (\forall t \ge \varsigma). \tag{7}$$

有了上述定义，我们的主要结果概括在以下定理中。

定理1. 考虑 $\mathbb{Z}_2$ 同步问题，其中 $n^{-1/9}\log n \lesssim \lambda - 1 \le 0.2$。假设我们采用随机初始化 $n^{-1/9}$ 运行 *AMP*（参见 (2) 和 (3)）。考虑任意满足 $1 \le t \le {}^{cn}/_{(\lambda-1)^5\log^2 n}$ 的 t，其中 $c > 0$ 是某个普适常数。那么，以至少 $1 - O(n^{-10})$ 的概率，以下结果成立:

- （分解与误差界）*AMP* 迭代可分解为

$$x_t = \alpha_t v^\star + \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k + \xi_{t-1}, \tag{8a}$$

其中，α_t 定义见（5），ϕ_k是满足 $\phi_k \overset{\text{独立同分布}}{\sim} \mathcal{N}(\mathbf{0}, \frac{1}{n}I_n)$ 的独立同分布高斯向量,_ 并且

$$\|\eta_t(x_t)\|_2 = \|\beta_t\|_2 := \left\|(\beta_t^1, \beta_t^2, \ldots, \beta_t^t)\right\|_2 = 1, \tag{8b}$$

$$\|\xi_t\|_2 \lesssim \sqrt{\frac{t\log n}{n(\lambda-1)^2}} + \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}; \tag{8c}$$

- （穿越时间）由 ς 定义的（6）阈值满足

$$\varsigma = O\left(\frac{\log n}{\lambda - 1}\right); \tag{9}$$

- （非渐近状态演化）对于满足 $\varsigma \le t \le {}^{cn}(\lambda-1)^5$ 对数$^2 n$ 的任意 t，我们有

$$\alpha_t^2 = \left(1 + O\left(\sqrt{\frac{\left(t + \frac{\log^3 n}{\lambda-1}\right)\log n}{n(\lambda-1)^5}}\right)\right)\alpha_t^{\star 2}, \tag{10}$$

其中 $\{\alpha_t^\star\}$ 表示在(7)中定义的渐近状态演化参数。

备注 1（ λ的范围）。 定理 1 仅涵盖 λ 大于但接近 1的情形。事实上，$\lambda = 1$ 表示 $\mathbb{Z}_2$ 同步的相变点（Deshpande 等人，2017 年），具体而言：（i）当 $\lambda < 1$时，没有任何估计量在渐近意义上优于 0 估计量；（ii）当 λ 严格大于 1 时，则有可能与 $v^\star$实现非平凡的相关性。我们通过考虑一个更精细但极具挑战性的情形 $\lambda - 1 \gtrsim n^{-1/9}\log n$ （使得 λ 可能非常接近 1），聚焦于可行区域。虽然可以改进指数 $1/9$，但这超出了本文的范围。上界 $\lambda \le 1.2$ 并非关键；作出这一假设仅仅是为了简化表述。

接下来，我们将对定理1 作一些解释，并讨论其含义。以下假设 $\lambda > 1$。

高斯近似。定理1中的第一个结果（18a）指出，每个 AMP 迭代由三个分量组成：（i）信号分量 $\alpha_t v^\star$，它与真实信号 $v^\star$对齐；（ii）噪声分量$\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k$，它是独立同分布高斯向量的线性组合；以及（iii）残差分量 ξ_{t-1}。尽管这种分解与Li 和 Wei（2022年）的分解相似，但我们证明，即便没有精心设计的谱初始化，这种分解仍然成立。下面给出几点说明。

- 关于噪声分量，其分布（记作$\mu\left(\sum_{k=1}^t \beta_t^k\phi_k\right)$）与高斯分布 $\mathcal{N}\left(0\text{之间的 }1\text{-瓦瑟斯坦距离 }\frac{1}{n}I_n\right)$ 至多为（Li 和 Wei，2022年，引理 9）

$$W_1\left(\mu\left(\sum_{k=1}^t \beta_k\phi_k\right), \mathcal{N}\left(0, \frac{1}{n}I_n\right)\right) \lesssim \sqrt{\frac{t\log n}{n}}. \tag{11}$$

当 t不太大时，噪声分量可以很好地近似为高斯向量 $\mathcal{N}(0, \frac{1}{n}I_n)$。

- 关于信号分量 $\alpha_t v^\star$，显而易见，α_t 决定了 AMP 在恢复真实信号方面的有效程度。值得注意的是，一旦 α_t 超过阈值 $\frac{1}{2}\sqrt{\lambda^2-1}$，它就遵循一种非渐近状态演化，与渐近对应形式 $\alpha_t^\star$ （见（10））非常相似。这一结果得益于我们分析的非渐近性质。

总而言之，除至多为 $\widetilde{O}\left(\sqrt{n}^{\,t}(\lambda-1)^2 + \sqrt{n}^{\,1}(\lambda-1)^3\right)$的小误差项外，AMP迭代近似为

$$x_t \approx \alpha_t v^\star + \mathcal{N}\left(0, \frac{1}{n}I_n\right), \qquad t < O\left(\frac{n(\lambda-1)^5}{\log^2 n}\right)$$

即使从随机值初始化也是如此。Montanari 和 Venkataramanan （2021 年）给出了这一观察结果的渐近版本，不过其中的结果要求既有信息性初始化，又有一个不会随 t增长的固定 n。

随机初始化后的动力学。 与 Li 和 Wei (2022年, 定理 4) 相比，定理 1 最具挑战性的部分在于分析随机初始化后的初始阶段。正如我们的分析将要阐明的，我们可以将近似消息传递轨迹划分为三个阶段来理解。

- **阶段 #1：摆脱随机初始化**。 在使用 $x_1 \sim \mathcal{N}(0, \frac{1}{n}I_n)$ 进行随机初始化时，AMP 起始时的信号分量极小，约为 $\widetilde{O}(\frac{1}{\sqrt{n}})$ 的量级，此时规范的状态演化变得空泛。为克服这一技术障碍，我们对这一阶段（在 $|\alpha_t|$ 超过 $\sqrt{\lambda-1}\,n^{-1/4}$之前） α_t 如何演化进行了细粒度刻画，即

$$\alpha_{t+1} \approx \lambda\alpha_t + \lambda g_{t-1}, \qquad \text{with } g_{t-1} \sim \mathcal{N}\left(0, \frac{1}{n}\right); \tag{12}$$

详见第A.4节。从这一近似的含噪递推关系可以看出，尽管信号分量起初可能被噪声项掩盖，但只需经过 $O(\frac{\log n}{\lambda-1})$ 次迭代，信号分量便会超过噪声规模，并达到 $\sqrt{\lambda-1}\, n^{-1/4}$ 的量级（见第 3.2.2节）。

• **阶段#2：指数增长**。一旦信号分量的大小超过 $\sqrt{\lambda-1}n^{-1/4}$，AMP迭代便与真实信号产生非平凡相关。有趣的是，信号强度 α_t 开始呈指数增长，直至达到 $\sqrt{\lambda^2-1}$. 的量级。正如我们将在第3.2.2节中论证的， α_{t+1} 满足

$$\alpha_{t+1} \geq \sqrt{1+\frac{1-o(1)}{3}(\lambda-1)}\,\alpha_t \tag{13}$$

在此阶段，最多经过 $O(\frac{\log n}{\lambda-1})$ 次迭代。

• **阶段#3：局部精化**。达到 $\sqrt{\lambda^2-1}$的量级后， α_t 进入局部精化阶段，在此期间，随机初始化的 AMP 的行为与采用谱初始化或其他信息性初始化的 AMP 类似。在此阶段，渐近状态演化（7）也开始能够有效预测 α_t 的演化（见（10））。正如我们将在第3.2.4节中证明的，信号强度 α_t 满足

$$|\alpha_{t+1}^2-\alpha^{\star 2}| \lesssim \big(1-(\lambda-1)\big)^{t-\varsigma} + \widetilde{O}\left(\sqrt{\frac{t+\frac{1}{\lambda-1}}{n(\lambda-1)^5}}\right), \tag{14}$$

其中 $\alpha^\star$ （由 λ确定）表示 $\alpha_t^\star$ 的极限（参见（7）），并且是唯一解 of

$$\alpha^{\star 2} = \lambda^2 \mathbb{E}\big[\tanh\left(\alpha^\star\left(\alpha^\star+G\right)\right)\big], \qquad \text{for } G\sim\mathcal{N}(0,1). \tag{15}$$

渐近最优性。 正如先前所示（例如参见 Celentano 等人 (2021 年，引理 A.7）以及 Li 和 Wei (2022年，注 4）），我们可以构造一个基于近似消息传递的估计量，其风险与贝叶斯最优估计量的风险一致 $\widehat{X}^{\text{bayes}} := \mathbb{E}[v^\star v^{\star\top} \mid M]$。更具体地，将基于近似消息传递的估计量取为

$$u_t := \frac{1}{\lambda\sqrt{n(\alpha_t^2+1)}}\tanh(\pi_t x_t), \tag{16}$$

其渐近风险满足（参见第 B 节以及 Deshpande 等人 (2017 年））：

$$\lim_{t\to\infty}\lim_{n\to\infty}\mathbb{E}\big[\big\|v^\star v^{\star\top}-u_t u_t^\top\big\|_{\mathsf{F}}^2\big] = \lim_{n\to\infty}\mathbb{E}\big[\|v^\star v^{\star\top}-\widehat{X}^{\text{bayes}}\|_{\mathsf{F}}^2\big] = 1-\frac{\alpha^{\star 4}}{\lambda^4}, \tag{17}$$

其中 $\alpha^\star$ 是极限状态演化的不动点（参见（15））。结合定理 1中的非渐近结果，我们可以得到更精细的风险刻画，具体证明见第 B节。

推论 1。 以至少 $1-O(n^{-10})$ 的概率，存在某个 $t=O(\frac{\log n}{\lambda-1})$，使得 t

$$\big\|v^\star v^{\star\top}-u_t u_t^\top\big\|_{\mathsf{F}}^2 = 1-\frac{\alpha^{\star 2}}{\lambda^4}+O\left(\sqrt{\frac{\log^4 n}{n(\lambda-1)^6}}\right). \tag{18}$$

3 定理证明1

在本节中，我们给出主要结果——定理1的证明。为使表述更加简洁，我们先引入以下记号：

$$v_t := \alpha_t v^\star + \sum_{k=1}^{t-1}\beta_{t-1}^k \phi_k. \tag{19}$$

3.1 预备知识

在开始证明主定理之前，我们先整理几个有用的结果，这些结果将在整个证明过程中频繁用到。

集中结果。 我们首先记录 Li 和 Wei 提出的几个有用的集中结果(2022年)。此处及下文中，我们令 $|x|_{(i)}$ 表示 i中按绝对值从大到小排列的第 $x \in \mathbb{R}^n$个元素的大小。

引理 1. 考虑一组随机向量 $\{\phi_k\}_{1\le k<t}$ *in* $\mathbb{R}^n$。假设对于每个 $1 \le k \le t-1 < n$， $\phi_k = (\phi_{k,\,1},\ \phi_{k,\,2},\ \ldots\ldots,\ \phi_{k,\,n})$ 均为从 $\mathcal{N}\left(0, \frac{1}{n} I_n\right)$ 中独立同分布抽取。考虑以下集合

$$\mathcal{E}_s := \left\{\{\psi_k\}_{k=1}^{t-1} : \max_{1\le k\le t-1} \|\psi_k\|_2 < 1 + C\sqrt{\frac{\log\frac{n}{\delta}}{n}}\right\} \bigcap \left\{\{\psi_k\}_{k=1}^{t-1} : \sup_{a\in\mathcal{S}^{t-2}} \left\|\sum_{k=1}^{t-1} a_k\psi_k\right\|_2 < 1 + C\sqrt{\frac{t\log\frac{n}{\delta}}{n}}\right\}$$
$$\bigcap \left\{\{\psi_k\}_{k=1}^{t-1} : \sup_{a\in\mathcal{S}^{t-2}} \sum_{i=1}^{s} \left|\sum_{k=1}^{t-1} a_k\psi_k\right|_{(i)}^2 < \frac{C(t+s)\log\frac{n}{\delta}}{n}\right\} \bigcap \left\{\{\psi_k\}_{k=1}^{t-1} : \max_{1\le k<t, 1\le i\le n} |\psi_{k,i}| < C\sqrt{\frac{\log\frac{n}{\delta}}{n}}\right\},$$

并记为 $\mathcal{E} := \bigcap_{s=1}^{n} \mathcal{E}_s$。那么存在一个足够大的常数 $C > 0$，使得对于每个 $\delta > 0$,

$$\mathbb{P}\left(\{\phi_k\}_{k=1}^{t-1} \in \mathcal{E}\right) \ge 1 - \delta.$$

特别地，令 $\delta = n^{-11}$，可知以下事件发生的概率至少为 $1-O(n^{-11})$:

$$\left|\max_{1\le k\le t-1} \|\phi_k\|_2 - 1\right| \lesssim \sqrt{\frac{\log n}{n}}, \qquad \textit{and} \quad \max_{1\le k\le t, 1\le i\le n} |\phi_{k,i}| \lesssim \sqrt{\frac{\log n}{n}}.$$

该引理可由关于高斯随机向量的标准测度集中结论推出（Massart，2007）；证明可参见 Li 和 Wei(2022年，第 D.1.1 节），为简洁起见，此处省略。

关于 η_t、 π_t和 γ_t 的性质（参见(3)）。 接下来，我们总结式（3）中定义的三组关键量的一些基本性质。首先，汇总我们在式（3）中对 π_t 和 γ_t的选取所具有的若干基本性质；证明见第A.1节。

引理 2。 假设分解(8a) 在 $\|\xi_{t-1}\|_2 \lesssim 1$条件下成立。以至少 $1 - O(n^{-10})$的概率，以下性质成立:

$$\frac{1}{\sqrt{n}}\pi_t = |\alpha_t| + O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2} \wedge \frac{1}{|\alpha_t|}\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right); \tag{20a}$$

$$\gamma_t^{-2} = n\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x) + \pi_t^2 O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right). \tag{20b}$$

此外，还有

$$\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x) = \frac{\pi_t^2}{n}(\alpha_t^2 + 1) + O\left(\frac{\pi_t^4}{n^2}\right), \tag{20c}$$

这进而意味着

$$\gamma_t^{-2} = \pi_t^2\left(\alpha_t^2 + 1 + O\left(\frac{\pi_t^2}{n} + \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right). \tag{20d}$$

接下来，我们来列出下述引理中关于量 η_t 的若干性质，其证明推迟至第A.2节。

引理 3。 考虑任意 $1 \le t \le n$，并假设 $\|\xi_{t-1}\|_2$ 满足

$$\|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{t^3\log n}{n}} \qquad \textit{for all } t \lesssim \frac{\log n}{\lambda - 1}. \tag{21}$$

则以下性质以至少 $1 - O(n^{-11})$ 的概率成立:

• 如果 $t \lesssim \frac{n}{(\lambda-1)^4 \log^2 n}$ 且 $\|\xi_{t-1}\|_2 \lesssim \frac{1}{\sqrt{\log n}}$，则任意 $x \in \mathbb{R}$ 均满足

$$|\eta_t(x)| \lesssim |x|, \qquad |\eta_t'(x)| \lesssim 1 =: \rho, \qquad |\eta_t''(x)| \lesssim \sqrt{n} =: \rho_1, \qquad |\eta_t^{(\prime\prime\prime)}(x)| \lesssim n =: \rho_2; \tag{22a}$$

• 若 $t \lesssim \frac{\text{对数} n}{\lambda-1}$ 且 $\alpha_t \lesssim \sqrt{\lambda-1}\, n^{-0.1}$，则有 $\pi_t \lesssim \sqrt{\lambda-1}\, n^{-0.4}$；对于任意满足 $|x| \lesssim \sqrt{\text{对数} n/n}$ 的 x，我们有

$$\eta_t'(x) = 1 + O\big((\lambda-1)n^{-0.2}\log n\big) \qquad \textit{and} \qquad |\eta_t''(x)| \lesssim (\lambda-1)n^{0.8}|x|; \tag{22b}$$

• 若 $t \lesssim \frac{\text{对数} n}{\lambda-1}$ 且 $\alpha_t \lesssim \sqrt{\lambda-1}\, n^{-1/4}$，则有 $\pi_t \lesssim (\lambda-1)^{-3/4} n^{1/4}$ 对数 n；对于任意 x，都可以找到一个量 $c_0 \lesssim (\text{对数}^2 n)/\sqrt{n(\lambda-1)^3}$，它与 x 无关；以及另一个量 $|c_x| \lesssim n|x|^5 (\text{对数}^4 n)/(\lambda-1)^3$，它依赖于 x，使得

$$\eta_t(x) = (1-c_0)\left(x - \frac{1}{3}\pi_t^2 x^3 + c_x\right). \tag{22c}$$

3.2 AMP 动力学的非渐近分析

现在我们可以给出主定理的证明。下面概述证明的结构。

• 首先，聚焦于满足 $t \le \varsigma \wedge \frac{n}{\log c\ \lambda-1()}$ 的初始阶段，我们在第 $\|\xi_t\|_2$ 3.2.1 节中推导出 的一个上界，如下：

$$\|\xi_t\|_2 \lesssim \sqrt{\frac{t^3 \log n}{n}}; \tag{23}$$

其中，ς 是式（6）中定义的阈值，而 $c > 0$ 是一个足够小的常数。我们通过归纳论证来证明这一点。

• 其次，我们在第3.2.2 节研究信号强度 α_t 在 AMP 执行过程中的演化。关键的是，回顾 ς 反映了首次满足 t 的时刻$|\alpha_t| \gtrsim \sqrt{\lambda^2-1}$ （参见式（6））， 我们证明

$$\varsigma \lesssim \frac{\log n}{\lambda-1}; \tag{24}$$

换言之，尽管初始化是随机的（因而不具信息性），AMP 至多经过 $O\big(\frac{\log n}{\lambda-1}\big)$ 次迭代便能找到一个信息性估计。

• 第三，在上述对 ς 的控制到位后，我们继续推导关于$\|\xi_t\|_2$ 的一个更完整的上界，涵盖 ς之后的迭代，即

$$\|\xi_t\|_2 \lesssim \sqrt{\frac{t\,\mathbb{1}(t>\varsigma)\log n}{n(\lambda-1)^2}} + \sqrt{\frac{\min\{t,\varsigma\}^3 \log n}{n}} \tag{25}$$

对于任意 $t \le \frac{cn}{(\lambda-1)^5 \log^2 n}$。这是第 3.2.3 节的主要内容，同样通过归纳论证完成。

• 最后，当迭代次数超过阈值 ς后，我们在第3.2.4 节证明，渐近状态演化（刻画大系统极限的状态演化）在有限样本/有限时间范围内也相当准确。特别是，我们建立了非渐近状态演化与其渐近对应物之间的密切联系，这对于刻画近似消息传递的有限样本收敛行为起着关键作用。

接下来将详细解释这四个步骤。

3.2.1 控制 ξ_t 当 $t \le \varsigma \wedge \log^{c} \overset{n}{\lambda^{-1}}()$ (命题 (23)的证明)

在本小节中，我们证明所声称的界（23），适用于 $\|\xi_t\|_2$，其思路借鉴了Li 和 Wei（2022年）的工作。首先，我们重述Li 和 Wei（2022年，定理 2），并做了适当简化，以适用于 $\mathbb{Z}_2$ 同步问题（即利用性质 $\|\beta_t\|_2^2 = 1$、 $E_t = 0$和（22a））。为方便记号，定义 $\kappa_t > 0$ 使得

$$\kappa_t^2 := \max\left\{\left\langle \int \left[x\eta_t'\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right) - \frac{1}{\sqrt{n}}\eta_t''\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right)\right]^2 \varphi_n(\mathrm{d}x)\right\rangle, \left\langle \int \left[\eta_t'\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right)\right]^2 \varphi_n(\mathrm{d}x)\right\rangle\right\}, \tag{26}$$

这里回顾一下， $\varphi_n(\cdot)$ 是 $\mathcal{N}(0, I_n)$ 的概率密度函数；对于任意向量 $x = [x_i]_{1\le i\le n}$，我们记作 $\langle x\rangle := \frac{1}{n}\sum_{i=1}^n x_i$ 和 $x^2 = [x_i^2]_{1\le i\le n}$。我们将采用以下假设。

假设 1。 对于任意 $1 \le t \le n$，考虑任意向量 $\mu_t \in \mathcal{S}^{t-1}$、 $\xi_{t-1} \in \mathbb{R}^n$以及系数$(\alpha_t, \beta_{t-1}) \in \mathbb{R}\times\mathbb{R}^{t-1}$，它们可能都与 ϕ_k存在统计依赖关系。令 v_t按 (19)中的定义。我们假设存在（可能是随机的）量 A_t、 B_t、 D_t，使得以至少 $1 - O(n^{-11})$的概率，以下不等式成立：

$$\left|\sum_{k=1}^{t-1}\mu_t^k\left[\langle\phi_k, \eta_t(v_t)\rangle - \langle\eta_t'(v_t)\rangle\beta_{t-1}^k\right]\right| \le A_t, \tag{27a}$$

$$\left|v^{\star\top}\eta_t(v_t) - v^{\star\top}\int \eta_t\Big(\alpha_t v^\star + \frac{\|\beta_{t-1}\|_2}{\sqrt{n}}x\Big)\varphi_n(\mathrm{d}x)\right| \le B_t, \tag{27b}$$

$$\left\|\sum_{k=1}^{t-1}\mu_t^k\phi_k\circ\eta_t'(v_t) - \frac{1}{n}\sum_{k=1}^{t-1}\mu_t^k\beta_{t-1}^k\eta_t''(v_t)\right\|_2^2 - \kappa_t^2 \le D_t. \tag{27c}$$

在这些假设下，Li 和 Wei (2022年，定理 2）针对 AMP 迭代给出了一般性的非渐近刻画，如下所示。

定理 2。 *[改编自Li 和 Wei(2022年，定理 2)]*假设*1*成立，并考虑任意的 $t \le n$。以至少 $1 - O(n^{-11})$的概率，针对(2）的 AMP 迭代在 $\mathbb{Z}_2$-同步问题中满足分解式(8a)，其中 $\|\beta_t\|_2^2 = 1$ 且

$$\alpha_{t+1} = \lambda v^{\star\top}\int \eta_t\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right)\varphi_n(\mathrm{d}x) + \Delta_{\alpha,t} \tag{28}$$

其中残差项满足

$$|\Delta_{\alpha,t}| \lesssim B_t + \left|v^{\star\top}\eta_t(x_t) - v^{\star\top}\eta_t(v_t)\right| \lesssim B_t + \|\xi_{t-1}\|_2, \tag{29a}$$

$$\|\xi_t\|_2 \le \sqrt{\kappa_t^2 + D_t}\,\|\xi_{t-1}\|_2 + O\left(\sqrt{\frac{t\log n}{n}} + A_t + \sqrt{(1 + t\,\mathbb{1}_{t\le\varsigma})\log n}\,\|\xi_{t-1}\|_2^2 + \sqrt{\frac{t\log n}{n}}\|\xi_{t-1}\|_2\right). \tag{29b}$$

备注 2。 关于上述29b）中对 $\|\xi_t\|_2$的界，直接应用Li 和 Wei (2022年，定理 2）会得到一项 $\sqrt{t\log n}\,\|\xi_{t-1}\|_2^2$（而不是$\sqrt{(1 + t\,\mathbb{1}_{t\le\varsigma})\log n}\,\|\xi_{t-1}\|_2^2$（见29b））。在此我们做了些许修改，使其更适合当前设置。

(i) 当 $t \le \varsigma$ 时，这样的项 $\sqrt{t\log n}\|\xi_{t-1}\|_2^2$ 足以满足我们的需求；

(ii) 当 $t > \varsigma$ (也就是说 α_t 超过 $\sqrt{\lambda^2 - 1}$的阶时)，只需直接运用 Li 和 Wei (2022年，展示公式 (249))，就可以将 $\|\xi_{t-1}\|_2^2$ 前面的因子从 $\sqrt{t}\log n$ 改进为 $\sqrt{\log n}$.

将这些结果结合起来，就得到了所声称的界（29b）。值得注意的是，这一看似微小的改动对于将迭代次数提升至 $O(n/$多项式$(\log n))$ 而非 $O(\sqrt{n}/$多项式$(\log n))$至关重要。

考虑到定理 2，为了控制 $|\Delta_{\alpha,\,t}|$ 和 $\|\xi_t\|_2$，归根结底分别需要确定 A_t，B_t，D_t 和 κ_t。

- **对 A_t、B_t 和 D_t 的界。** 重复 Li 和 Wei (2022年，D.2 节中的相同分析，我们得到

$$A_t \lesssim \sqrt{\frac{t\log n}{n}}, \qquad B_t \lesssim \sqrt{\frac{t\log n}{n}}, \qquad D_t \lesssim \sqrt{\frac{t\log^2 n}{n}}. \tag{30}$$

唯一需要进一步讨论的项是 A_t，因为 Li 和 Wei (2022年) 仅证明了，对于采用独立初始化的近似消息传递，$A_t \lesssim \frac{1}{\alpha_t}\sqrt{\frac{t\log n}{n}}$ (其中取 $s=1$)。为消去前因子 $1/\alpha_t$，我们依靠对 $\eta_t(x)$ 的更精细控制(一参见（22a）)。特别地，性质（22a）告诉我们

$$\|\eta_t(v_t)\|_2 \lesssim \|v_t\|_2 = \Big\|\alpha_t v^\star + \sum_{k=1}^{t-1}\beta_{t-1}^k \phi_k\Big\|_2 \lesssim 1,$$

最后一个不等式可见展示公式（87）。由此可得

$$\|\nabla_\Phi f_\theta(\Phi)\|_2 \lesssim \frac{1}{\sqrt{n}} \tag{31}$$

通过采用与 Li 和 Wei(2022年，D.2.1 节不等式（229）附近所详述的相同分析。这里，$\nabla_\Phi f_\theta(\Phi)$ 是控制 A_t 的关键量，见 Li 和 Wei (2022年，D.2.1 节)，因此我们所需的 A_t 界可立即得到。鉴于这只涉及对Li 和 Wei (2022年，D.2.1 节所作的细微且直接的改动，为简洁起见，我们省略细节，并请读者参阅 Li 和 Wei(2022年，D.2.1 节以了解更多详情。

- **界定 κ_t。** 因此，主要步骤归结为界定 κ_t。为此，我们声称对于 κ_t，以下关系成立，其证明推迟到第A.3节。

引理 4。 以至少 $1-O(n^{-10})$ 的概率，以下结论成立：

– 在归纳假设 （23）对 ξ_{t-1}成立的条件下，有

$$\kappa_t \le 1 + o\Big(\frac{\lambda-1}{\log n}\Big) \tag{32a}$$

前提是 $t \le \varsigma \wedge \frac{\log n}{c(\lambda-1)}$;

– 在归纳假设 (25) 成立时，ξ_{t-1}，可得

$$\kappa_t \le 1 - \frac{1}{15}(\lambda-1), \tag{32b}$$

provided that $t \le \frac{cn(\lambda-1)^5}{\log^2 n}$ *and* $|\alpha_t| \gtrsim \sqrt{\lambda^2-1}$.

有了上述对 A_t、B_t、D_t、κ_t的估计，我们现在可以应用定理 2。根据归纳假设 (23)，将定理 29) 中的递归公式与定理 2 中的 (30) 结合，可得

$$\|\xi_t\|_2 \le \sqrt{\kappa_t + \sqrt{\frac{t\log^2 n}{n}}}\,\|\xi_{t-1}\|_2 + O\Bigg(\sqrt{\frac{t\log n}{n}} + \sqrt{(1+t\,\mathbb{1}_{t\le\varsigma})\log n}\,\|\xi_{t-1}\|_2^2 + \sqrt{\frac{t\log n}{n}}\|\xi_{t-1}\|_2\Bigg), \tag{33}$$

结合（32）进一步意味着

$$\|\xi_t\|_2 \le \left(1 + o\Big(\frac{\lambda-1}{\log n}\Big) + O\Bigg(\sqrt{\frac{t^4\log^2 n}{n}}\Bigg)\right)\|\xi_{t-1}\|_2 + O\Big(\sqrt{\frac{t\log n}{n}}\Big), \tag{34}$$

前提是 $t \le \varsigma \wedge \overset{n}{\text{对数}\, c\ \lambda-1}()$。

现在我们可以通过归纳法证明关系式（23）。为验证其在基例（即 $t=1$）中的有效性，我们注意到，根据构造（例如参见（Li 和 Wei，2022年，步骤3，定理 1 的证明）），ξ_1 具有如下形式

$$\xi_1 = \Big(\frac{\sqrt{2}}{2} - 1\Big) z_1 z_1^\top W z_1, \quad \text{where } z_1 = \eta_1(x_1) \text{ is independent of } W.$$

初等计算表明，以至少 $1 - O(n^{-11})$ 的概率，

$$\|\xi_1\|_2 = \Big|\frac{\sqrt{2}}{2} - 1\Big| \cdot \|z_1\|_2 \cdot |z_1^\top W z_1| \lesssim \sqrt{\frac{\log n}{n}}, \tag{35}$$

已知 $\|z_1\|_2 = 1$ 且 $z_1^\top W z_1 \sim \mathcal{N}(0, \overset{2}{\ } I_n)$。这就已经证明了当 $t=1$时，关系式（23）对基例成立。接下来，考虑 $\overset{n}{c\ \lambda-1}\log()$ 的情形，其中 $\sqrt{\overset{n}{t^4}}$ 是一个足够小的常数。已知在关于 o的假设下，$\overset{\lambda-1}{\ }\log n(\log)$，递归关系式（34）立即推出

$$\begin{aligned}
\|\xi_t\|_2 &\le \Big(1 + o\Big(\frac{\lambda-1}{\log n}\Big)\Big)\|\xi_{t-1}\|_2 + O\Big(\sqrt{\frac{t\log n}{n}}\Big) \\
&\le \Big(1 + o\Big(\frac{\lambda-1}{\log n}\Big)\Big)^{t-1}\|\xi_1\|_2 + \sum_{j=0}^{t-2}\Big(1 + o\Big(\frac{\lambda-1}{\log n}\Big)\Big)^{j} O\Big(\sqrt{\frac{(t-j)\log n}{n}}\Big) \\
&\lesssim \sqrt{\frac{t^3\log n}{n}}
\end{aligned} \tag{36}$$

对所有 $t \le \varsigma \wedge \overset{n}{\log c\ \lambda-1}()$，结论得证。

注 3. 细心的读者可能会注意到，在 (34) 中建立的关于 $t \le \varsigma \wedge \overset{n}{\log c\ \lambda-1}()$ 的递归公式并不依赖于关系 (24)（该关系将在下一小节中建立）。

3.2.2 α_t 的演化及 ς 的上界（命题 (24) 的证明）

现在我们继续证明命题 (24)，该命题给出了阈值 ς的上界，这需要仔细分析信号强度 α_t 在初始阶段的演化。为此，我们根据 α_t的大小分为两种情况，并将在介绍若干预备事实后详细说明。

预备事实。 继续之前，我们先回顾 Li 和 Wei Li 和 Wei (2022年) 中已建立的一些额外预备事实。根据 Li 和 Wei (2022年，定理1），我们知道：由构造可知，

$$\xi_t \in \mathsf{span}(U_{t-1}) = \mathsf{span}\Big\{\eta_1(x_1), \ldots, \eta_{t-1}(x_{t-1})\Big\}, \tag{37}$$

其中 $U_{t-1} \in \mathbb{R}^{n\times(t-1)}$ 是一个矩阵，其列由一组标准正交基 $\{z_1, \ldots, z_{t-1}\}$构成。事实上，我们可以用更明确的方式指定 U_t。沿用 Li 和 Wei （2022年，第4.1节）的做法，定义

$$z_1 := \frac{\eta_1(x_1)}{\|\eta_1(x_1)\|_2} \in \mathbb{R}^n \qquad \text{and} \qquad W_1 := W \in \mathbb{R}^{n\times n}, \tag{38a}$$

它们彼此统计独立；然后，对于任意 $2 \le t \le n$，我们可以递归地定义以下对象：

$$U_{t-1} := [z_k]_{1\le k\le t-1} \in \mathbb{R}^{n\times(t-1)}, \tag{38b}$$

以及

$$z_t := \frac{\big(I_n - U_{t-1}U_{t-1}^\top\big)\,\eta_t(x_t)}{\big\|\big(I_n - U_{t-1}U_{t-1}^\top\big)\,\eta_t(x_t)\big\|_2}, \tag{38c}$$

$$W_t := \big(I_n - z_{t-1}z_{t-1}^\top\big)\,W_{t-1}\,\big(I_n - z_{t-1}z_{t-1}^\top\big), \tag{38d}$$

其中， $\{x_t\}$ 是由 AMP 更新（2）生成的序列。因此，该过程为 $\{U_t\}$ 和标准正交基 $\{z_t\}$ （关于Li 和 Wei (2022年，第4.1节）中$\{z_t\}$的标准正交性，参见 $\{z_t\}$ 的标准正交性揭示了分解

$$\eta_t(x_t) = \sum_{k=1}^{t} \beta_t^k z_k, \qquad \text{with } \beta_t^k := \langle \eta_t(x_t),\, z_k \rangle, \tag{39}$$

满足 $\|\eta_t(x_t)\|_2 = \|\beta_t\|_2$，其中 $\beta_t = [\beta_t^1, \ldots, \beta_t^t]$。此外，我们发现生成

$$\phi_k \coloneqq W_k z_k + \zeta_k, \qquad \text{where } \zeta_k \coloneqq \Big(\frac{\sqrt{2}}{2} - 1\Big) z_k^\top W_k z_k \cdot z_k + \sum_{i=1}^{k-1} g_i^k z_i, \qquad 1 \le k \le n, \tag{40}$$

其中 g_i^k是从以下分布中独立抽取的 $\mathcal{N}(0, \frac{1}{n})$。Li 和 Wei Li 和 Wei (2022年，引理 2）中证明了以下性质，这些性质在后续分析中起着关键作用：

- ϕ_k 独立同分布。独立同分布。独立同分布$\sim \mathcal{N}(0, \frac{1}{n} I_n)$，对于 $1 \le k \le n$；
- ϕ_k 的随机性仅来自 W_k，且 ϕ_k 与 x_1 和 $\{z_i\}_{i<k}$相互独立。
- x_k z_k 在给定 $\{z_i\}_{i<k}$ 和 x_1的条件下，与 W_k 条件独立。
- ϕ_k 与 $\{x_j\}_{j \le k}$ 和 $\{z_j\}_{j \le k}$相互独立。

阶段 I：较小的相关性（$|\alpha_t| \lesssim \sqrt{\lambda - 1}\, n^{-1/4}$）。让我们从最开始说起，此时相关系数 α_t 相当小。为此，我们定义阈值 τ_0，使得

$$\tau_0 := \max\big\{\tau : |\alpha_t| \lesssim \sqrt{\lambda - 1}\, n^{-1/4} \text{ for all } t \le \tau\big\}; \tag{41}$$

换言之， $\tau_0 + 1$ 表示第一个超过 $\sqrt{\lambda - 1}\, n^{-1/4}$水平的项。接下来，我们希望证明，以至少 $1 - O(n^{-10})$ 的概率，该阈值不会太大，即

$$\tau_0 \lesssim \frac{\log n}{\lambda - 1}. \tag{42}$$

命题 (42)的证明。为了证明这一结果（42），我们先陈述一个重要命题：从随机点初始化时，近似消息传递迭代以高概率满足以下递归关系：

$$\alpha_{t+1} = \lambda^{t-k+1} \alpha_k + \sum_{i=1}^{t-k+1} \lambda^i g_{t-i} + O\left(\lambda^{t-k} \frac{\log^4 n}{n^{3/4} (\lambda - 1)^{1.5}}\right) \tag{43}$$

对任意 $1 \le k \le t$，其中我们记

$$g_k \coloneqq v^{\star\top} \phi_k \quad (1 \le k \le t) \qquad \text{and} \qquad g_0 = 0. \tag{44}$$

这一命题中的关系是阶段 I 分析的核心，在这一阶段，近似消息传递迭代与 $v^\star$ 之间的相关性持续增大，直至达到非平凡值。为使表述更简洁，我们将此命题的证明推迟到第 A.4节。

利用上述递归公式（43），我们现在来证明关系（42）。定义$t_i := C' i \log n$，其中某个量为 $C' = \frac{C''}{\lambda - 1}$，其中 C'' 是一个足够大的常数。注意到——

$$\begin{aligned}
\mathbb{P}\Big(|\alpha_k| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}},\ \text{for all } k \le 201 C' \log n\Big) &\le \mathbb{P}\Big(|\alpha_{t_i + 1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}},\ \text{for all } 1 \le i \le 200\Big) \\
&= \prod_{i=1}^{200} \mathbb{P}\Big(|\alpha_{t_i+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} \;\Big|\; |\alpha_{t_j+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}},\ \forall 1 \le j < i\Big).
\end{aligned}$$

To 为了控制上述关系右端的值，考虑以下随机变量 le

$$X_i := \sum_{j=1}^{C' \log n} \lambda^j g_{t_i - j} \sim \mathcal{N}\left(0, \frac{\lambda^{2C' \log n + 2} - \lambda^2}{n(\lambda^2 - 1)}\right) \tag{45}$$

对于每个 $1 \le i \le 200$。借助这一记号，并应用关系式（43），得到

$$\alpha_{t_i+1} = \lambda^{t_i - t_{i-1}} \alpha_{t_{i-1}+1} + X_i + O\left(\lambda^{t_i - t_{i-1} - 1} \frac{\log^4 n}{n^{3/4}(\lambda - 1)^{1.5}}\right). \tag{46}$$

如上述预备事实中所述，每个 ϕ_j 都与 AMP迭代 x_i 相互独立，其中 $i \le j$，因此 α_{i+1}，因为 $\alpha_{i+1} = v^{\star\top} \eta_i$（$x_i$）。因此，上文定义的随机变量 X_i 与 α_{t_j+1} 对所有 $j \le i-1$都相互独立。结合关系式（43），可得

$$\begin{aligned}
&\mathbb{P}\Big(|\alpha_{t_i+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} \;\Big|\; |\alpha_{t_j+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}}, 1 \le j < i\Big) \\
&\le \mathbb{P}\Big(|\lambda^{t_i - t_{i-1}} \alpha_{t_{i-1}+1} + X_i| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} + \lambda^{t_i - t_{i-1} - 1} \frac{\log^4 n}{n^{3/4}(\lambda - 1)^{1.5}} \;\Big|\; |\alpha_{t_j+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}}, 1 \le j < i\Big) \\
&= \mathbb{P}\Big(|\lambda^{t_i - t_{i-1}} \alpha_{t_{i-1}+1} + X_i| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} + \lambda^{C' \log n} \frac{\log^4 n}{n^{3/4}(\lambda - 1)^{1.5}} \;\Big|\; |\alpha_{t_{i-1}+1}| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}}\Big) \\
&\le \mathbb{P}\Big(|X_i| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} + \lambda^{C' \log n} \frac{\log^4 n}{n^{3/4}(\lambda - 1)^{1.5}}\Big).
\end{aligned} \tag{47}$$

这里，倒数第二行由上文所述的独立关系得出，而最后一行则由以下初等事实得出：

$$\mathbb{P}_{X \sim \mathcal{N}(\mu, \sigma^2)}(|X| < x) \le \mathbb{P}_{X \sim \mathcal{N}(0, \sigma^2)}(|X| < x), \qquad \forall x > 0.$$

综合以上各点，可得

$$\begin{aligned}
\mathbb{P}\Big(|\alpha_k| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}}, \text{ for all } k \le 201 C' \log n\Big) &\le \prod_{i=1}^{200} \mathbb{P}\left(|X_i| \lesssim \frac{\sqrt{\lambda - 1}}{n^{1/4}} + \frac{\lambda^{C' \log n} \log^4 n}{n^{3/4}(\lambda - 1)^{1.5}}\right) \\
&= \prod_{i=1}^{200} \mathbb{P}\left(\sqrt{\frac{n(\lambda^2 - 1)}{\lambda^{2C' \log n + 2} - \lambda^2}} |X_i| \lesssim \frac{(\lambda - 1) n^{1/4}}{\lambda^{C' \log n}} + \frac{\log^4 n}{n^{1/4}(\lambda - 1)}\right) \\
&\lesssim \left(\frac{\log^4 n}{n^{1/4}(\lambda - 1)}\right)^{200} \lesssim n^{-11},
\end{aligned} \tag{48}$$

其中，我们在表达式（45）中使用了 X_i 的分布，而最后一个不等式则由假设 $\lambda - 1 \gtrsim n^{-1/9}$ 得出。因此，上述不等式保证，以至少 $1 - O(n^{-10})$ 的概率，存在某个 $k \lesssim$ 对数$_{\lambda-1}^{n}$，使得

$$|\alpha_k| \gtrsim \sqrt{\lambda - 1}\, n^{-1/4}. \tag{49}$$

因此，这意味着 $\tau_0 \lesssim$ 对数$_{\lambda-1}^{n}$（见定义（41）），正如（42）中所述。换言之，至多经过$O($ $_{\lambda-1}^{\log n}$ $)$ 次迭代后，$|\alpha_t|$ 将超过 $\sqrt{\lambda - 1}\, n^{-1/4}$的阶。 □

第二阶段：中等至较大的相关性（$\sqrt{\lambda - 1}\, n^{-1/4} \lesssim |\alpha_t| \le \frac{1}{2}\sqrt{\lambda^2 - 1}$）。接下来，我们来看一下这样的时间区间：在 $|\alpha_t|$ 超过 $\sqrt{\lambda - 1}\, n^{-1/4}$的水平之后，但在还到 $\frac{1}{2}\sqrt{\lambda^2 - 1}$的水平之前。数学上，这指的是区间（$\tau_0$, ς], 其中 τ_0 和 ς 分别定义于（41）和（6）中。事实上，我们将首先考察

$$t \in \left(\tau_0, \varsigma \wedge \frac{c_5 \log n}{\lambda - 1}\right] \tag{50}$$

对于某个常数 $c_5 > 0$；我们很快将证明 $\varsigma \lesssim \frac{n}{\log_{\lambda-1}}$。

根据定理 2 以及界（30），我们可以写成

$$\alpha_{t+1} = \lambda v^{\star\top} \int \eta_t \left(\alpha_t v^\star + \frac{1}{\sqrt{n}} x \right) \varphi_n(\mathrm{d}x) + \Delta_{\alpha,t}, \tag{51}$$

其中残差项满足

$$|\Delta_{\alpha,t}| \lesssim \sqrt{\frac{t \log n}{n}} + \left| v^{\star\top} \eta_t(x_t) - v^{\star\top} \eta_t(v_t) \right|. \tag{52}$$

我们首先针对 α_{t+1} 的精细递归关系提出一个断言：

$$|\alpha_{t+1}| \geq \frac{\lambda |\alpha_t|}{\sqrt{\alpha_t^2 + 1}} + o\big((\lambda - 1)|\alpha_t|\big) + O(|\Delta_{\alpha,t}|); \tag{53}$$

该结果的证明推迟到第 A.5 节。注意，每当 $\alpha_t < \frac{1}{2}\sqrt{\lambda^2 - 1}$ 时，都有

$$\left(\frac{\lambda}{\sqrt{1 + \alpha_t^2}} \right)^2 \geq \frac{\lambda^2}{\frac{1}{4}\lambda^2 + \frac{3}{4}} > 1 + \frac{1}{3}(\lambda - 1), \quad \text{for } \lambda \in (1, 1.2], \tag{54}$$

将其与表达式（53）结合起来，可得

$$|\alpha_{t+1}| \geq \left(\sqrt{1 + \frac{1}{3}(\lambda - 1)} + o(\lambda - 1) \right) |\alpha_t| + O(|\Delta_{\alpha,t}|). \tag{55}$$

此外，我们断言，对于每个 $t \leq \tau'$ where $\tau' := \min\{t : \alpha_t \geq (\lambda - 1)^{-3/4} n^{-1/4}\}$，都有

$$|\Delta_{\alpha,t}| \ll (\lambda - 1)|\alpha_t|, \tag{56}$$

我们将在第A.6节中证明这一点。结合关系式（55）和（56），它满足 $|\alpha_{\tau'+1}| \geq |\alpha_{\tau'}|$。此外，注意到界（52）与（36）结合可确保

$$|\Delta_{\alpha,\tau'+1}| \lesssim \sqrt{\frac{(\tau'+1)^3 \log n}{n}} \ll (\lambda - 1)|\alpha_{\tau'+1}|. \tag{57}$$

递归地应用这一论证，我们便得到

$$|\alpha_{t+1}| \geq \left(\sqrt{1 + \frac{1}{3}(\lambda - 1)} + o(\lambda - 1) \right) |\alpha_t|.$$

现在将上述递归关系与假设 $\lambda - 1 \gtrsim n^{-1/9}$ 结合起来可知，$|\alpha_t|$ 在至多 $O\big(\log_{\lambda-1} n\big)$ 次迭代内超过了 $\frac{1}{2}\sqrt{\lambda^2 - 1}$。因此，回顾我们对（6）中 ς 的定义，我们可以立即得出

$$\varsigma = O\left(\frac{\log n}{\lambda - 1} \right). \tag{58}$$

3.2.3 对 ξ_t 的更完整界（证明命题（25））

现在我们继续通过归纳论证，为任意满足 t 的 $\varsigma < t < \frac{cn}{(\lambda-1)^5 \log^2 n}$ 的（回顾（58），有 $\varsigma = O\big(\frac{\log n}{\lambda-1}\big)$）证明命题（25）。在此过程中，我们还需要证明

$$|\alpha_t| \geq \frac{1}{2}\sqrt{\lambda^2 - 1} \tag{59}$$

在此阶段内（n 也就是说，一旦 $|\alpha_t|$ 超过 $\frac{1}{2}\sqrt{\lambda^2 - 1}$，信号强度就永远不会低于这个水平）。

首先，基例 $t=\varsigma$ 的断言 (36) 已在表达式 (36)中得到验证；当 $t=\varsigma$时，条件 (59)也以高概率成立。接下来，假设断言 (25) 在迭代 $t-1$之前一直成立，我们希望证明它在时刻 t仍然成立。为此，不等式 (33)结合引理 4可知

$$\begin{aligned}\|\xi_t\|_2 &\le \sqrt{\kappa_t+\sqrt{\frac{t\log^2 n}{n}}}\|\xi_{t-1}\|_2+O\left(\sqrt{\frac{t\log n}{n}}+\sqrt{(1+t\mathbb{1}_{t\le\varsigma})\log n}\,\|\xi_{t-1}\|_2^2+\sqrt{\frac{t\log n}{n}}\|\xi_{t-1}\|_2\right)\\ &\le \sqrt{1-\frac{1}{15}(\lambda-1)+\sqrt{\frac{t\log^2 n}{n}}}\|\xi_{t-1}\|_2+O\left(\sqrt{\frac{t\log n}{n}}\right)\\ &\quad+O\left(\sqrt{(1+t\mathbb{1}_{t\le\varsigma})\log n}\left(\sqrt{\frac{t\mathbb{1}_{t>\varsigma}\log n}{n(\lambda-1)^2}}+\sqrt{\frac{\min\{t,\varsigma\}^3\log n}{n}}\right)+\sqrt{\frac{t\log n}{n}}\right)\|\xi_{t-1}\|_2\\ &\le \left(1-\frac{1}{15}(\lambda-1)+O\left(\sqrt{\frac{\left(\frac{t}{(\lambda-1)^2}+\varsigma^3\right)\log^2 n}{n}}\right)\right)\|\xi_{t-1}\|_2+O\Big(\sqrt{\frac{t\log n}{n}}\Big),\end{aligned}\tag{60}$$

其中，第二行由32b）以及针对 $t-1$的归纳假设（25）得到。此外，对于 α_{t+1} 的第$(t+1)$次迭代，（A.7节中也将详细说明如何证明（

鉴于上述递归关系，注意到对于某个足够小的常数 $c>0$，~~$\sqrt{(t/(\lambda-1)^2+\varsigma^3}$~~ $^2 n \le 2c^)\log(\lambda-1)$，我们可以直接推导出

$$\begin{aligned}\|\xi_t\|_2 &\le \left(1-\frac{1}{20}(\lambda-1)\right)\|\xi_{t-1}\|_2+O\Big(\sqrt{\frac{t\log n}{n}}\Big)\\ &\le \left(1-\frac{1}{20}(\lambda-1)\right)^{t-\varsigma}\|\xi_\varsigma\|_2+\sum_{j=0}^{t-\varsigma-1}\left(1-\frac{1}{20}(\lambda-1)\right)^j O\Big(\sqrt{\frac{(t-j)\log n}{n}}\Big)\\ &\lesssim \sqrt{\frac{\varsigma^3\log n}{n}}+\sqrt{\frac{t\log n}{n(\lambda-1)^2}}\end{aligned}\tag{61}$$

对于所有 $\varsigma\le t\le {}^{cn}_{(\lambda-1)^5\log^2 n}$。将此结果与界（36）（针对 $t\le\varsigma$）结合，便可立即证明：对于所有 $t\le {}^{cn}_{(\lambda-1)^5\log^2 n}$，命题（25）成立。

3.2.4 近似状态演化分析（性质(10)的证明）

当信号强度 α_t 达到 $\sqrt{\lambda^2-1}$的量级时，近似消息传递进入局部精化阶段。根据（25）和（24），对于任意 $t\le {}^{cn}_{(\lambda-1)^5\log^2 n}$，AMP迭代~~$x_t$~~可以分解为（8a），其误差项有如下界：

$$\|\xi_t\|_2\lesssim\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^2}}.\tag{62}$$

在均值　　时间上，为了描述 α_t 的演化，我们根据 $\Delta_{\alpha,\,t}$ 利用关系（29a）给出如下界　　s:

$$|\Delta_{\alpha,t}|\lesssim B_t+\|\xi_{t-1}\|_2\lesssim\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^2}},\tag{63}$$

其中最后一个不等式由（30）得出。结合关系（28），可得

$$\alpha_{t+1}=\lambda v^{\star\top}\int\eta_t\left(\alpha_t v^\star+\frac{1}{\sqrt{n}}x\right)\varphi_n(\mathrm{d}x)+O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^2}}\right).$$

接下来，我们将刻画 α_{t+1} 与其渐近对应量之间的距离，以进一步理解 α_{t+1} 的演化。具体而言，回顾一下，渐近状态演化定义为

$$\alpha_{t+1}^{\star} = \lambda \left[\int \tanh\left(\alpha_t^{\star}\left(\alpha_t^{\star} + x\right)\right) \varphi(\mathrm{d}x) \right]^{1/2}, \tag{64}$$

假设我们从某个 $\varsigma = O\big(\log_{\lambda^{-1}} n\big)$ 的 $\alpha_{\varsigma}^{\star} = |\alpha_{\varsigma}|$ 开始。我们的目标是控制 α_{t+1} 与 $\alpha_{t+1}^{\star}$ 之间的差异。为简化表述，不失一般性地假设 $\alpha_t > 0$，并采用以下记号

$$\tau_t \coloneqq \left(\alpha_t^{\star}\right)^2.$$

首先，与命题（53）的证明中相同的分析（此处采用不同的误差界（63））给出

$$\begin{aligned}
\alpha_{t+1} &= \lambda \left[\int \tanh\left(\alpha_t^2 + \alpha_t x\right) \varphi(\mathrm{d}x) \right]^{1/2} + O\left(\left| \frac{\pi_t^2}{\alpha_t^2 n} - 1 \right| \alpha_t^3 + \sqrt{\frac{(t + \frac{\log^3 n}{\lambda - 1}) \log n}{n(\lambda - 1)^2}} \right) \\
&= \lambda \left[\int \tanh\left(\alpha_t^2 + \alpha_t x\right) \varphi(\mathrm{d}x) \right]^{1/2} + O\left(\sqrt{\frac{(t + \frac{\log^3 n}{\lambda - 1}) \log n}{n(\lambda - 1)^2}} \right).
\end{aligned} \tag{65}$$

这里，最后一行由不等式（20a）得出，该不等式表明

$$\left(\frac{\pi_t}{\sqrt{n}}\right)^2 = \alpha_t^2 + O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t \log n}{n}} \right) = \alpha_t^2 + O\left(\sqrt{\frac{(t + \frac{\log^3 n}{\lambda - 1}) \log n}{n(\lambda - 1)^2}} \right).$$

由关系（65）和（64）可得

$$\frac{\alpha_{t+1}^2 - \tau_{t+1}}{\tau_{t+1}} = \frac{\int \left[\tanh\left(\alpha_t^2 + \alpha_t x\right) - \tanh\left(\tau_t + \sqrt{\tau_t} x\right)\right] \varphi(\mathrm{d}x)}{\int \tanh\left(\tau_t + \sqrt{\tau_t} x\right) \varphi(\mathrm{d}x)} + O\left(\sqrt{\frac{(t + \frac{\log^3 n}{\lambda - 1}) \log n}{n(\lambda - 1)^3}} \right). \tag{66}$$

这里，提醒读者（另见（113）以及 Deshpande 等人（2017 年，附录 B.2））

$$\int \tanh\left(\alpha^2 + \alpha x\right) \varphi(\mathrm{d}x) = \int \tanh^2\left(\alpha^2 + \alpha x\right) \varphi(\mathrm{d}x) \asymp \alpha^2, \qquad \text{for } \alpha \in (0, \lambda],$$

其中最后一个不等式由关系式（113）得出。递归公式（66）量化了 α_t 与 $\alpha_t^{\star}$ 之间的差异随时间的变化，这在我们接下来的分析中发挥着关键作用。

为了更好地理解上述递归关系，我们对每个 $\tau \in [0, \lambda^2]$ 定义

$$h(\tau) \coloneqq \int \tanh\left(\tau + \sqrt{\tau} x\right) \varphi(\mathrm{d}x).$$

借助这个函数，可以写成

$$\alpha_{t+1}^{\star 2} = \lambda^2 h(\alpha_t^{\star 2}). \tag{67}$$

此外，直接计算可得

$$h'(\tau) \coloneqq \int \left(1 + \frac{x}{2\sqrt{\tau}}\right) \left(1 - \tanh^2\left(\tau + \sqrt{\tau} x\right)\right) \varphi(\mathrm{d}x) \in (0, 1),$$

其取值范围可由 Li 和 Wei（2022年，D.3.3 节）中的展示公式（263）得到。我们指出上述结果的几个直接推论。

- 注意到 $h'(\tau) > 0$，可得 $\alpha_{t+1}^{\star} > \alpha_t^{\star} \gtrsim \sqrt{\lambda^2 - 1}$ 当 $t \geq \varsigma$.

• 根据 Li 和 Wei (2022年，D.3.3 节中的展示公式（264），我们有 $0 \le \lambda^2 h'(\tau) \le 1-(\lambda-1)$。若定义 $\alpha^\star$ 为（64）（当 $t \to \infty$））的极限点，则可以看出

$$|\alpha_{t+1}^{\star 2} - \alpha^{\star 2}| \le \big(1-(\lambda-1)\big) \cdot |\alpha_t^{\star 2} - \alpha^{\star 2}|, \qquad \text{for } t \ge \varsigma. \tag{68}$$

换言之，渐近状态演化参数 $\alpha_t^{\star 2}$以指数速度收敛到某个不动点 $\alpha^{\star 2}$.

• 根据上述记号，我们也可以写成

$$\begin{aligned}\frac{|\alpha_{t+1}^2 - \tau_{t+1}|}{\tau_{t+1}} &= \frac{|h(\alpha_t^2) - h(\tau_t)|}{h(\tau_t)} + O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^3}}\right) \\ &= \frac{h'(\tau)}{h(\tau_t)/\tau_t} \cdot \frac{|\alpha_t^2 - \tau_t|}{\tau_t} + O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^3}}\right)\end{aligned} \tag{69}$$

对于某个满足 $\min\{\tau_t, \alpha_t^2\} \le \tau \le \max\{\tau_t, \alpha_t^2\}$ 的 τ。

我们首先证明 $\alpha_t^2 = (1+o(1))\tau_t$。根据定义，$\alpha_\varsigma^\star = \alpha_\varsigma \gtrsim \sqrt{\lambda^2-1}$，因此该结论对于 $t=\varsigma$ 显然成立。接下来，假设归纳假设 $\alpha_t^2 = (1+o(1))\tau_t$ 成立，我们希望证明它对于第 $(t+1)$ 步也成立。为此，我们首先声称存在一个足够小的普适常数$c>0$，使得

$$\frac{h'(\tau)}{h(\tau_t)/\tau_t} \le 1 - c(\lambda-1), \tag{70}$$

其证明推迟至第A.9节。这进一步使我们能够推导出

$$\frac{|\alpha_t^2 - \tau_t|}{\tau_t} \lesssim \sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^5}}, \tag{71}$$

我们将通过归纳论证来证明这一点。鉴于当 $t=\varsigma$时，式（71）显然成立，我们打算证明第$(t+1)$次迭代时式（71）成立，并假设它对所有 $s \le t$都成立。为此，我们结合式（70）和式（69）来证明

$$\begin{aligned}\frac{|\alpha_{t+1}^2 - \tau_{t+1}|}{\tau_{t+1}} &\le \big(1-c(\lambda-1)\big) \cdot \frac{|\alpha_t^2 - \tau_t|}{\tau_t} + O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^3}}\right) \\ &= \big(1-c(\lambda-1)\big)^{t+1-\varsigma} \cdot \frac{|\alpha_\varsigma^2 - \tau_\varsigma|}{\tau_\varsigma} + O\left(\sum_{k=\varsigma}^{t-1} \big(1-c(\lambda-1)\big)^{t-k} \sqrt{\frac{(k+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^3}}\right) \\ &\le \sum_{k=\varsigma}^{t-1} \big(1-c(\lambda-1)\big)^{t-k} \cdot O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^3}}\right) \\ &\lesssim \sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^5}},\end{aligned} \tag{72}$$

其中，第二行是通过递归应用（71）得到的。因此，它也推出了 $\alpha_{t+1}^2 =$（$1+o$（1））τ_{t+1}，由此完成了第（$t+1$）次迭代的归纳假设。将上述结果与表达式（68）结合，还可得到

$$\begin{aligned}|\alpha_{t+1}^2 - \alpha^{\star 2}| &= \big(1-(\lambda-1)\big)^{t-\varsigma} \cdot |\alpha_\varsigma^{\star 2} - \alpha^{\star 2}| + \alpha_{t+1}^{\star 2} O\left(\sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^5}}\right) \\ &\lesssim \big(1-(\lambda-1)\big)^{t-\varsigma} + \sqrt{\frac{(t+\frac{\log^3 n}{\lambda-1})\log n}{n(\lambda-1)^5}}.\end{aligned} \tag{73}$$

4 讨论

本文明确刻画了随机初始化时 AMP 的有限样本收敛行为，重点研究了原型问题 $\mathbb{Z}_2$ 同步问题。已证明，该算法具有快速的全局收敛性，只需不超过 $O(\frac{\log n}{\lambda-1})$ 次迭代，即可达到风险为 $O(\sqrt{\text{对数}n\ \lambda-1\ ()^6}\frac{4}{n})$、接近贝叶斯最优的点。据我们所知，我们的理论首次为随机初始化的 AMP 在低秩矩阵估计中的有效性提供了严格证据。尽管本文聚焦于一种针对 $\mathbb{Z}_2$ 同步问题设计的特定去噪函数，但我们预计，我们的分析框架可以推广到更广泛的一类可分且满足 Lipschitz 连续性的去噪函数。

展望未来，值得探索的研究方向不胜枚举。一个自然的拓展方向是研究关于 $v^\star$ 的其他结构先验；例如，若已知 $v^\star$ 满足一般的锥约束（参见 Bandeira 等人（2019）；Wei 等人（2019）），随机初始化的 AMP 会表现如何，便很值得探究。另一个值得关注的方向是超越尖峰 Wigner 模型。近期一项相关研究（Wu 和 Zhou，2022年）探讨了随机初始化在张量分解问题的幂迭代中的作用，并借助 AMP 类型的分析来研究张量幂方法。我们能否进一步拓展这些方法，以理解（随机初始化的）AMP 如何解决低秩矩阵补全和张量补全等更具挑战性的问题？此外，请注意，当前理论刻画了 AMP 至多进行 $O(^n\text{poly}(\log\ _n))$ 步迭代时的行为。尽管这在实践中已经很有用，但从技术角度而言，若迭代次数继续增加，AMP 最终是否会收敛到一个不动点，也值得研究。最后，虽然 AMP 是理解高维中各种统计程序的一种通用工具，但也有若干其他分析框架，例如凸高斯极小极大定理（CGMT）（Celentano 等人，2020年；Miolane 和 Montanari，2021 年；Thrampoulidis 等人，2018年）以及留一法分析（Abbe 等人，2020年；Chen 等人，2020年；El Karoui，2018年），它们同样行之有效，也各有优势。有没有有效的方法将这些框架结合起来，从而同时发挥它们各自的优势？这些问题留待未来研究。

致谢

本研究部分由美国国家科学基金会资助项目 DMS 2147546/2015447 和美国国家科学基金会职业奖 DMS-2143215 资助。

A 辅助引理和断言的证明

A.1 引理 2 的证明

性质 (20a) 的证明。回顾一下， $\pi_t := \sqrt{n(\|x_t\|_2^2-1)\vee 1}$。为证明性质 (20a)，第一步是计算 $\|x_t\|_2$。注意，对于相互独立的高斯向量 $\phi_k \sim \mathcal{N}\big(0,\ \frac{1}{n}I_n\big)$，有

$$v^{\star\top}[\phi_1,\ldots,\phi_{t-1}] \sim \mathcal{N}\Big(0,\frac{1}{n}I_{t-1}\Big),$$

由于 $\|v^\star\|_2=1$。因此，很容易看出

$$\Big|\Big\langle v^\star,\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\rangle\Big| = \Big|\Big\langle v^{\star\top}[\phi_1,\ldots,\phi_{t-1}],[\beta_{t-1}^1,\ldots,\beta_{t-1}^{t-1}]\Big\rangle\Big| \le \Big\|v^{\star\top}[\phi_1,\ldots,\phi_{t-1}]\Big\|_2\|\beta_{t-1}\|_2 \lesssim \sqrt{\frac{t\log n}{n}} \tag{74}$$

概率至少为 $1-O(n^{-11})$。结合不等式（74）和（86），得到

$$\|v_t\|_2^2 = \Big\|\alpha_t v^\star+\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\|_2^2 = \|\alpha_t v^\star\|_2^2+\Big\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\|_2^2+2\Big\langle\alpha_t v^\star,\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\rangle$$

$$= \alpha_t^2 + 1 + O\left(\sqrt{\frac{t\log n}{n}}\right) \asymp 1, \tag{75}$$

进而得到

$$\|x_t\|_2^2 = \left(\|v_t\|_2 + O(\|\xi_{t-1}\|_2)\right)^2 = \alpha_t^2 + 1 + O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right). \tag{76}$$

基于上述性质，我们还可以推出

$$\begin{aligned}\sqrt{n(\|x_t\|_2^2 - 1)} &= \sqrt{n\left(\alpha_t^2 + O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right)} \le \sqrt{n}\,|\alpha_t| + O\left(\sqrt{n\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)}\right) \\ &= \sqrt{n}\left(|\alpha_t| + O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\right)\right). \end{aligned} \tag{77}$$

其中，第一个不等式基于以下事实：$\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ 对于 $a, b \ge 0$。特别地，若 $\alpha_t^2 \lesssim \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}$，则基本关系 $\sqrt{a+b} = \sqrt{a} + O(\sqrt{b})\,(0 \le a \lesssim b)$ 使我们能够将式（≤）中的“78”替换为“=”，从而得到

$$\sqrt{n(\|x_t\|_2^2 - 1)} = \sqrt{n}\left(\alpha_t + O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\right)\right). \tag{78}$$

相比之下，若 $\alpha_t^2 \gtrsim \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}$，则基本关系 $\sqrt{a+b} = \sqrt{a} + O\left(\frac{b}{\sqrt{a}}\right)\,(0 \le b \lesssim a)$ 得出

$$\begin{aligned}\sqrt{n(\|x_t\|_2^2 - 1)} = \sqrt{n\left(\alpha_t^2 + O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right)} &= \sqrt{n}|\alpha_t|\sqrt{1 + \frac{1}{\alpha_t^2}O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)} \\ &= \sqrt{n}\left(|\alpha_t| + \frac{1}{|\alpha_t|}O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right). \end{aligned} \tag{79}$$

综合前两个界可得

$$\sqrt{n(\|x_t\|_2^2 - 1)} = \sqrt{n}|\alpha_t| + \sqrt{n}\left\{\frac{1}{|\alpha_t|}O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right) \wedge O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\right)\right\}.$$

上述界还可导出所需形式（π_t 当 $\sqrt{n(\|x_t\|_2^2 - 1)} \ge 1$ 时的所需形式 $\sqrt{n(\|x_t\|_2^2 - 1)} < 1$）。因此，接下来只需考察

$$c_{10}|\alpha_t| \le \left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}$$

对于某个充分小常数 $c_{10} > 0$。但在这种情况下，我们仍有

$$\begin{aligned}\pi_t = 1 &\lesssim \sqrt{n}\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2} \asymp \sqrt{n}\left(|\alpha_t| + O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\right)\right) \\ &\asymp \sqrt{n}|\alpha_t| + \sqrt{n}\left\{O\left(\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\right) \wedge \frac{1}{|\alpha_t|}O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)\right\}, \end{aligned}$$

因此，所声称的界仍然成立。

性质 (20b) 的证明。 首先回顾定义 $\gamma_t^{-2} := \|$双曲正切$(\pi_t x_t)\|_2^2$。为证明性质 (20b)，考虑以下差值

$$\begin{aligned}
&\left| \|\tanh(\pi_t x_t)\|_2^2 - \int \left\| \tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x) \right) \right\|_2^2 \varphi_n(\mathrm{d}x) \right| \\
&\quad \le \left| \|\tanh(\pi_t x_t)\|_2^2 - \|\tanh(\pi_t v_t)\|_2^2 \right| + \left| \|\tanh(\pi_t v_t)\|_2^2 - \int \left\| \tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x) \right) \right\|_2^2 \varphi_n(\mathrm{d}x) \right|
\end{aligned}$$

其中 $\varphi_n \sim \mathcal{N}(0,\ I_n)$。为估计上式右端，注意到双曲正切 gi 的利普希茨性质 ves

$$\|\tanh(\pi_t x_t) - \tanh(\pi_t v_t)\|_2 \le \|\tanh(\pi_t (x_t - v_t))\|_2 \le \pi_t \|\xi_{t-1}\|_2.$$

由此可得

$$\begin{aligned}
\left| \|\tanh(\pi_t x_t)\|_2^2 - \|\tanh(\pi_t v_t)\|_2^2 \right| &= \left| \|\tanh(\pi_t x_t)\|_2 + \|\tanh(\pi_t v_t)\|_2 \right| \cdot \left| \|\tanh(\pi_t x_t)\|_2 - \|\tanh(\pi_t v_t)\|_2 \right| \\
&\le \left(2 \|\tanh(\pi_t v_t)\|_2 + \pi_t \|\xi_{t-1}\|_2 \right) \pi_t \|\xi_{t-1}\|_2 \\
&\lesssim \left(\pi_t \|v_t\|_2 + \pi_t \|\xi_{t-1}\|_2 \right) \pi_t \|\xi_{t-1}\|_2 \asymp \pi_t^2 \|\xi_{t-1}\|_2,
\end{aligned}$$

其中最后 一行用到了假设 $\|\xi_{t-1}\|_2 \lesssim 1$ 和方程 (75)。因此，我们得到 at

$$\begin{aligned}
&\left| \|\tanh(\pi_t x_t)\|_2^2 - \int \left\| \tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x) \right) \right\|_2^2 \varphi_n(\mathrm{d}x) \right| \\
&\quad \lesssim \pi_t^2 \|\xi_{t-1}\|_2 + \left| \|\tanh(\pi_t v_t)\|_2^2 - \int \left\| \tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x) \right) \right\|_2^2 \varphi_n(\mathrm{d}x) \right|.
\end{aligned} \tag{80}$$

接下来，我们推导对 (80) 中第二项的界，这可由一致集中性质推出。具体而言，考虑如下形式的函数

$$f_\theta(\Phi) := \|\tanh(\pi v)\|_2^2 - \int \left\| \tanh\left(\frac{\pi}{\sqrt{n}}(\alpha + x) \right) \right\|_2^2 \varphi_n(\mathrm{d}x), \qquad \text{where } v = \alpha v^\star + \sum_{k=1}^{t-1} \beta^k \phi_k;$$

这里，我们定义

$$\Phi = \sqrt{n} \left[\phi_1, \phi_2, \ldots, \phi_{t-1} \right], \quad \beta = \left[\beta^1, \beta^2, \ldots, \beta^{t-1} \right], \quad \theta = [\alpha, \beta, \pi] \in \mathbb{R}^{t+1}.$$

那么，只需对以下集合中的所有 θ，统一地界定 $f_\theta(\Phi)$

$$\Theta := \left\{ \theta = (\alpha, \beta, \pi) \mid \|\beta\|_2 = 1, \alpha \lesssim \sqrt{\lambda^2 - 1}, \pi \lesssim \sqrt{n} \right\}. \tag{81}$$

为此，首先考虑 f 关于 Φ 的导数，直接计算可得

$$\begin{aligned}
\|\nabla_\Phi f_\theta(\Phi)\|_2 &\le \frac{2\pi \|\beta\|_2}{\sqrt{n}} \|\tanh(\pi v) \circ \tanh'(\pi v)\|_2 \\
&\le \frac{2\pi \|\beta\|_2}{\sqrt{n}} \|\tanh(\pi v)\|_2 \lesssim \frac{\pi^2}{\sqrt{n}} \|v\|_2 \lesssim \frac{\pi^2}{\sqrt{n}},
\end{aligned} \tag{82}$$

其中我们注意到 $\|v\|_2 \le \alpha + \frac{1}{\sqrt{n}} \|\Phi\| \lesssim 1$。此外，由于函数 $f_\theta(\Phi)$ 一致有界于 2，若我们取 $\delta = n^{-300}$ 作为引理 1中定义的集合 $\mathcal{E}$（关于这一性质的证明，请参见 Li 和 Wei (2022年，D.1.1 节）

$$\mathbb{E}\left[|f(\Phi) - f(\mathcal{P}_{\mathcal{E}}(\Phi))| \right] \lesssim n^{-100}.$$

其中，$\mathcal{P}_{\mathcal{E}}(\cdot)$ 表示到集合 $\mathcal{E}$ 的欧几里得投影。结合上述不等式以及函数 $f_\theta(\Phi)$ 的以下性质，

1. 对任意 $\Phi \in \mathcal{E}$, $\|\nabla_\theta f_\theta(\Phi)\|_2 \lesssim n^{100}$;
2. 对任意固定的 θ, 有 $\mathbb{E}[f_\theta(\Phi)] = 0$,

我们可以应用Li 和 Wei (2022年, 推论 3) 得出

$$\sup_{\theta\in\Theta}\left|\frac{1}{\pi^2}f_\theta(\Phi)\right| \lesssim \sqrt{\frac{t\log n}{n}}. \tag{83}$$

综合以上各点, 我们得出

$$\left|\|\tanh(\pi_t x_t)\|_2^2 - \int\left\|\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right\|_2^2\varphi_n(\mathrm{d}x)\right| \lesssim \pi_t^2\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right), \tag{84}$$

注意到20b) 可得性质（

$$\int\left\|\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right\|_2^2\varphi_n(\mathrm{d}x) = n\int\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\varphi(\mathrm{d}x).$$

性质 (20c)**的证明**。 我们只需计算 $\int$双曲正切$^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\varphi(\mathrm{d}x)$, 具体如下。首先, 注意到 双曲正切(0) $=$ 双曲正切$''(0) = 0$, 双曲正切$'(0) = 1$, |双曲正切$'''(x)| \le 4$ 对于$x \in \mathbb{R}$, 我们得到

$$|\tanh(x) - x| \le \frac{2}{3}|x|^3,$$

根据中值定理。将这一关系与 |双曲正切$(x) - x| \le |x|$ 这一事实结合起来, 我们还可以进一步得到

$$|\tanh(x) - x| \le |x| \wedge |x|^3.$$

因此, 我们看到

$$|\tanh(x) + x| = |2x + \tanh(x) - x| = 2|x| + O(|x| \wedge |x|^3) \lesssim |x|,$$

这进而导致

$$\tanh^2(x) = x^2 + O(|\tanh(x) - x||\tanh(x) + x|) = x^2 + O(x^4).$$

现在我们可以计算 $\int$双曲正切$^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\varphi(\mathrm{d}x)$ 的值。根据上面得到的表达式, 可得

$$\begin{aligned}\int\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\varphi(\mathrm{d}x) &= \int\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)^2\varphi(\mathrm{d}x) + O\left(\int\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)^4\varphi(\mathrm{d}x)\right)\\ &= \frac{\pi_t^2}{n}(\alpha_t^2+1) + O\left(\frac{\pi_t^4}{n^2}\right).\end{aligned} \tag{85}$$

至此, 我们完成了所述结果的证明。

A.2 引理的证明 3

性质 (22a) **的证明**。性质 (22a) 涉及 η_t 及其导数的大小。根据 η_t的定义, 我们分别对参数 π_t 和 γ_t 进行界定。为此, 回顾引理 1 : 以至少 $1 - O(n^{-11})$ 的概率,

$$\left\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\right\|_2 = 1 + O\left(\sqrt{\frac{t\log n}{n}}\right) \tag{86}$$

对所有 $\beta_{t-1} = [\beta_{t-1}^k]_{1\le k<t} \in \mathcal{S}^{t-2}$同时成立。由此结果以及 19) 中 v_t 的定义可得

$$\|v_t\|_2 = \left\|\alpha_t v^\star + \sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k\right\|_2 \le \|\alpha_t v^\star\|_2 + \left\|\sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k\right\|_2 = |\alpha_t| + 1 + O\left(\sqrt{\frac{t\log n}{n}}\right) \lesssim 1, \tag{87}$$

鉴于 $t \lesssim {}^{n}$ 对数n 以及

$$|\alpha_t| = \left|\lambda v^{\star\top} \eta_{t-1}(x_{t-1})\right| \le \lambda \lesssim 1. \tag{88}$$

这些性质 与关于 ξ_t 的假设相结合，我们可以如下控制 ℓ_2 范数的 x_t s:

$$\|x_t\|_2 = \|v_t + \xi_{t-1}\|_2 \le \|v_t\|_2 + \|\xi_{t-1}\|_2 \lesssim 1. \tag{89}$$

鉴于 $\|x_t\|_2 \lesssim 1$， π_t 的值可控制为

$$\pi_t \coloneqq \sqrt{n(\|x_t\|_2^2 - 1)} \vee 1 \lesssim \sqrt{n}. \tag{90}$$

此外，注意到双曲正切$(0) = 0$、双曲正切$'(x) = 1 -$ 双曲正切$^2(x) \in [0,1]$ 以及 |双曲正切$(x)| \le 1$，可得

$$|\tanh(x)| \asymp |x| \wedge 1. \tag{91}$$

我们声称，这将导出以下结论：

$$\left\|\tanh(\pi_t x_t)\right\|_2 \asymp \pi_t; \tag{92}$$

暂且先将此视为已知，待证明性质（22a）后，我们再回过头来证明它。根据这一断言（92），可得

$$\gamma_t \coloneqq \|\tanh(\pi_t x_t)\|_2^{-1} \asymp \pi_t^{-1}. \tag{93}$$

为了证明性质（22a），只需回顾 η_t 在表达式（3）中的定义。对于任意 $x \in \mathbb{R}$，直接计算可得

$$\eta_t(x) = \gamma_t \tanh(\pi_t x) \tag{94a}$$
$$\eta_t'(x) = \gamma_t \pi_t \big(1 - \tanh^2(\pi_t x)\big) \tag{94b}$$
$$\eta_t''(x) = -2\gamma_t \pi_t^2 \tanh(\pi_t x)\big(1 - \tanh^2(\pi_t x)\big) \tag{94c}$$
$$\eta_t'''(x) = -2\gamma_t \pi_t^3 \big(1 - \tanh^2(\pi_t x)\big)\big(1 - 3\tanh^2(\pi_t x)\big). \tag{94d}$$

将这些恒等式与（91）、（93）以及 |双曲正切（x）| ≤ 1这一事实结合起来，便可轻松验证表达式（22a）。接下来只需证明断言（92），下文将对此进行证明。

关系（92）的证明。 注意，由展示公式（91）可知 $\left\|$双曲正切（$\pi_t x_t$）$\right\|_2 \asymp \left\||\pi_t x_t| \wedge 1\right\|_2$，，其中两个算子 $|\cdot|$ 和 $\wedge$ 均按逐项方式作用；此外，我们沿用记号 1 表示全 **1** 向量。为了建立关系（92），只需证明 $\left\||\pi_t x_t| \wedge 1\right\|_2 \asymp \pi_t$。为此，首先利用（90）观察到

$$\left\||\pi_t x_t| \wedge 1\right\|_2 \le \|\pi_t x_t\|_2 \wedge \|1\|_2 = \|\pi_t x_t\|_2 \wedge \sqrt{n} \asymp \pi_t \wedge \sqrt{n} \asymp \pi_t.$$

此外，我们按如下方式引入指标集 $\mathcal{I}$：

$$\mathcal{I} \coloneqq \left\{ i \in [n] \;\middle|\; |\xi_{t-1,i}| \le 0.9|v_{t,i}| \text{ and } |v_{t,i}| \lesssim \frac{1}{\sqrt{n}} \right\}, \tag{95}$$

这显然满足

$$0.1|\pi_t v_{t,i}| \le |\pi_t x_{t,i}| \le 1.9|\pi_t v_{t,i}|, \qquad \forall i \in \mathcal{I}.$$

由此可得：

$$\big\||\pi_t x_t| \wedge 1\big\|_2 \geq \big\||\pi_t x_t \circ \mathbb{1}_{\mathcal{I}}| \wedge 1\big\|_2 \asymp \big\||\pi_t v_t \circ \mathbb{1}_{\mathcal{I}}| \wedge 1\big\|_2. \tag{96}$$

为了进一步控制展示公式（96）的右侧，我们断言

$$\big\||\pi_t v_t \circ \mathbb{1}_{\mathcal{I}}| \wedge 1\big\|_2 \overset{\text{(i)}}{\asymp} \big\|\pi_t v_t \circ \mathbb{1}_{\mathcal{I}}\big\|_2 \overset{\text{(ii)}}{\asymp} \pi_t. \tag{97}$$

为说明这一点，关系 (i) 可利用表达式（90）和（95）的定义加以验证，因为它们保证

$$|\pi_t v_{t,i}| \leq |\pi_t| \cdot |v_{t,i}| \lesssim 1, \qquad \forall i \in \mathcal{I}.$$

为验证 (ii)，注意在指标集 $\mathcal{I}^c$ 上，有 $|v_{t,\,i}| \lesssim |\xi_{t-1,\,i}|$。因此，有

$$\|\pi_t v_t \circ \mathbb{1}_{\mathcal{I}^c}\|_2 \lesssim \|\pi_t \xi_{t-1} \circ \mathbb{1}_{\mathcal{I}^c}\|_2 \lesssim \pi_t \|\xi_{t-1}\|_2 \lesssim \pi_t \sqrt{\frac{1}{\log n}}, \tag{98}$$

这里回顾一下我们的假设 $\|\xi_{t-1}\|_2 \lesssim \sqrt{1/\log n}$。利用上述计算，我们可以回顾 $\|v_t\|_2 \asymp 1$ 表达式（87）以得到

$$\|\pi_t v_t \circ \mathbb{1}_{\mathcal{I}}\|_2^2 = \|\pi_t v_t\|_2^2 - \|\pi_t v_t \circ \mathbb{1}_{\mathcal{I}^c}\|_2^2 \asymp \left(1 - O\left(\frac{1}{\log n}\right)\right)\pi_t^2 \asymp \pi_t^2, \tag{99}$$

正如（97）第(ii)部分所声称的。

性质 (22b) 的证明。 为了研究 $\eta_t(x_t)$ 的导数，我们首先考虑参数 π_t 和 γ_t。鉴于 $\|\xi_{t-1}\|_2$ 满足表达式（21），当 $t \lesssim \log_{\lambda-1} n$ 时，有

$$\|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{t^3 \log n}{n}} \lesssim \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}. \tag{100}$$

在假设 $\lambda - 1 \gtrsim n^{-1/9} \log n$ 下，我们看到 $\|\xi_{t-1}\|_2$ 也满足

$$\|\xi_{t-1}\|_2 \lesssim (\lambda-1) \cdot \sqrt{\frac{\log^4 n}{n} \cdot \frac{1}{(\lambda-1)^5}} \lesssim (\lambda-1) \cdot \sqrt{\frac{\log^4 n}{n} \cdot \frac{n^{5/9}}{\log^5 n}} \lesssim (\lambda-1) n^{-0.2}.$$

同样，也很容易看出

$$\left(\frac{t \log n}{n}\right)^{1/2} \lesssim (\lambda-1) n^{-0.2}.$$

将这些结果与引理 78 证明中的关系 (2 以及假设 $\alpha_t \lesssim \sqrt{\lambda - 1}\, n^{-0.1}$ 结合起来可得

$$\pi_t \leq \sqrt{n}\alpha_t + O\left(\sqrt{n}\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t \log n}{n}}\right)^{1/2}\right) \lesssim \sqrt{\lambda - 1}\, n^{0.4}. \tag{101}$$

同样，引理 20d 的关系 (2 结合假设 $\alpha_t \leq \sqrt{\lambda - 1} n^{-0.1}$ 及上述界可得

$$(\gamma_t \pi_t)^{-2} = \alpha_t^2 + 1 + O\left(\frac{\pi_t^2}{n} + \|\xi_{t-1}\|_2 + \sqrt{\frac{t \log n}{n}}\right) = 1 + O\left((\lambda-1) n^{-0.2}\right),$$

通过直接计算还可得 $\gamma_t \pi_t = 1 + O((\lambda-1) n^{-0.2})$ 在我们关于 $\lambda - 1$ 的假设下。利用上述性质，结合一些代数运算以及 (94，进一步得到

$$\eta_t'(x) = \gamma_t \pi_t (1 - \tanh^2(\pi_t x)) = 1 + O\big((\lambda-1) n^{-0.2} \log n\big),$$
$$|\eta_t''(x)| \lesssim \pi_t \cdot |\pi_t x| \lesssim (\lambda-1) n^{0.8} |x|,$$

其中我们对任意 $|x| \lesssim \sqrt{\text{对数}/n}$ 调用关系 $|1 - \text{双曲正切}^2(\pi_t x)| \asymp 1$。

性质 (22c)的证明。 同样，当 $t \lesssim \text{对数}_{\lambda-1} n$ 时，$\|\xi_{t-1}\|_2$ 满足不等式 (100)。将其与假设 $\alpha_t \lesssim \sqrt{\lambda-1} n^{-1/4}$ 以及引理 2证明中的关系 (78) 综合起来，我们得到

$$\begin{aligned}\pi_t &\le \sqrt{n}\alpha_t + O\bigg(\sqrt{n}\Big(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\Big)^{1/2}\bigg) \lesssim \sqrt{\lambda-1}\, n^{1/4} + (\lambda-1)^{-3/4} n^{1/4}\log n \\ &\asymp (\lambda-1)^{-3/4} n^{1/4}\log n. \end{aligned} \tag{102}$$

类似地，引理 2 的关系 (20d 与我们关于 $\lambda-1$ 的假设结合起来，可得

$$\gamma_t \pi_t = 1 + O\left(\frac{\log^2 n}{\sqrt{n(\lambda-1)^3}}\right). \tag{103}$$

为推导性质 (22c)，注意以下一些简单事实：双曲正切(0) $=$ 双曲正切$''(0)$ $=$ 双曲正切$''''(0)$ $=0$，双曲正切$'(0)$ $=1$，双曲正切$'''(0)$ $=-2$ 以及 $|$双曲正切$^{(5)}(x)| \le K$，其中常数 K满足条件。于是，中值定理保证，对于任意 x，存在一个量 c 使得

$$\tanh(x) = x - \frac{1}{3}x^3 + cx^5 \qquad \text{for some} \quad 0 \le c \le K',$$

对于 $K' = K/120$。根据上述计算，我们可以得出

$$\eta_t(x) = \gamma_t \tanh(\pi_t x) = (1-c_0)\pi_t^{-1}\tanh(\pi_t x) = (1-c_0)\Big(x - \frac{1}{3}\pi_t^2 x^3 + c_x\Big),$$

其中 c_0 和 c_x 是满足以下条件的某些量

$$|c_0| \lesssim \frac{\log^2 n}{\sqrt{n(\lambda-1)^3}} \qquad \text{and} \qquad |c_x| \lesssim \pi_t^4 |x|^5 \lesssim \frac{n|x|^5\log^4 n}{(\lambda-1)^3}.$$

至此，所需性质的证明完成。

A.3 引理 4 的证明

不失一般性，在整个证明过程中，我们假设 $\alpha_t > 0$. 在开始对 κ_t 进行界定之前，我们先稍微简化一下所关注的项。首先，由于每个条目都服从 $v_i^\star \overset{\text{i.i.d}}{\sim}$均匀分布$\{\frac{1}{\sqrt{n}}, -\frac{1}{\sqrt{n}}\}$，因此可以很容易地推导出 $=$

$$\begin{aligned}&\left\langle \int \left[x\eta_t'\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right) - \frac{1}{\sqrt{n}}\eta_t''\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right)\right]^2 \varphi_n(\mathrm{d}x)\right\rangle \\ &\quad = \int \left[x\eta_t'\left(\alpha_t v_1^\star + \frac{1}{\sqrt{n}}x\right) - \frac{1}{\sqrt{n}}\eta_t''\left(\alpha_t v_1^\star + \frac{1}{\sqrt{n}}x\right)\right]^2 \varphi(\mathrm{d}x) \\ &\quad = \int \left[x\eta_t'\left(\frac{1}{\sqrt{n}}(\alpha_t + x)\right) - \frac{1}{\sqrt{n}}\eta_t''\left(\frac{1}{\sqrt{n}}(\alpha_t + x)\right)\right]^2 \varphi(\mathrm{d}x),\end{aligned}$$

其中，$\varphi(\cdot)$ 是 $\mathcal{N}(0,1)$ 的概率密度函数；我们利用了 $\eta_t(\cdot)$ 关于 0 对称这一事实，并且积分是在 $\mathcal{N}(0, I_n)$ 分布下进行的。同理，我们得到

$$\left\langle \int \left[\eta_t'\left(\alpha_t v^\star + \frac{1}{\sqrt{n}}x\right)\right]^2 \varphi_n(\mathrm{d}x)\right\rangle = \int \left[\eta_t'\Big(\frac{1}{\sqrt{n}}(\alpha_t + x)\Big)\right]^2 \varphi(\mathrm{d}x).$$

因此，有

$$\kappa_t^2 = \max\left\{\int |I_1(x)|^2\varphi(\mathrm{d}x),\ \int |I_2(x)|^2\varphi(\mathrm{d}x)\right\}, \tag{104}$$

其中我们定义

$$I_1(x) \coloneqq x\eta_t'\Big(\frac{1}{\sqrt{n}}(\alpha_t+x)\Big) - \frac{1}{\sqrt{n}}\eta_t''\Big(\frac{1}{\sqrt{n}}(\alpha_t+x)\Big),$$
$$I_2(x) \coloneqq \eta_t'\Big(\frac{1}{\sqrt{n}}(\alpha_t+x)\Big).$$

现在我们继续证明引理 4，首先将注意力限制在范围 $t < \varsigma \wedge \log_c \lambda^{-1}()$内。根据 α_t 的取值，我们分为两种情形。

情形 I: $\alpha_t \lesssim \sqrt{\lambda^2-1}\, n^{-0.1}$。引入一个事件 $\mathcal{A} \coloneqq \{x\colon |x| \le \sqrt{24\log n}\}$。对于 $x \sim \mathcal{N}(0,1)$，容易验证 $P(\mathcal{A}) = 1 - O(n^{-12})$。在给定 $\mathcal{A}$ 且假设 $\alpha_t \lesssim \sqrt{\lambda^2-1}\, n^{-0.1}$ 的条件下，有

$$\Big|\frac{1}{\sqrt{n}}(\alpha_t+x)\Big| \lesssim \sqrt{\frac{\log n}{n}}.$$

同时，回想一下，$\sqrt{\lambda+1} \asymp 1,$，因此 $\alpha_t \lesssim \sqrt{\lambda^2-1}\, n^{-0.1}$ 等价于 $\alpha_t \lesssim \sqrt{\lambda-1}\, n^{-0.1}$。于是，根据引理 3 中得到的性质 (22b)，可知：当 $|z| \lesssim \sqrt{\log_n n}$ 时，有

$$\eta_t'(z) = 1 + O\big((\lambda-1)n^{-0.2}\log n\big) \qquad \text{and} \qquad |\eta_t''(z)| \lesssim (\lambda-1)n^{0.8}|z|. \tag{105}$$

因此，对于所有 x r 位于 $\mathcal{A}$ 内的 es，我们可以将 $I_1(x)$ 与 x 之间的差异一致地界定如下 s:

$$\begin{aligned}
|I_1(x)-x| &\le \Big|\eta_t'\Big(\frac{1}{\sqrt{n}}(\alpha_t+x)\Big) - 1\Big|\,|x| + \frac{1}{\sqrt{n}}\Big|\eta_t''\Big(\frac{1}{\sqrt{n}}(\alpha_t+x)\Big)\Big| \\
&\lesssim \big((\lambda-1)n^{-0.2}\log n\big)\,|x| + (\lambda-1)n^{0.3}\Big|\frac{1}{\sqrt{n}}(\alpha_t+x)\Big| \\
&\lesssim \big((\lambda-1)n^{-0.2}\log n\big)\,|x| + (\lambda-1)n^{0.3}\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}}|x|\right) \\
&\lesssim \big((\lambda-1)n^{-0.2}\log n\big)\,(1+|x|).
\end{aligned} \tag{106}$$

此外，考虑任意 $x \in \mathbb{R}$。回顾引理 3 中的关系（22a）——该关系表明 $|\eta_t'(x)| \lesssim 1$，$|\eta_t''(x)| \lesssim \sqrt{n}$ ——我们注意到粗略界 $|I_1(x)| \lesssim |x| + 1$。结合前两个界，可得

$$\begin{aligned}
\int |I_1(x)|^2\varphi(\mathrm{d}x) &= \int x^2\varphi(\mathrm{d}x) + \int (I_1(x)-x)(I_1(x)+x)\varphi(\mathrm{d}x) \\
&= 1 + \int_{\mathcal{A}} (I_1(x)-x)(I_1(x)+x)\varphi(\mathrm{d}x) + \int_{\mathcal{A}^c} (I_1(x)-x)(I_1(x)+x)\varphi(\mathrm{d}x) \\
&= 1 + O\left(\int_{\mathcal{A}} (\lambda-1)n^{-0.2}\log n\cdot(|x|+1)^2\varphi(\mathrm{d}x) + \int_{\mathcal{A}^c}(|x|+1)^2\varphi(\mathrm{d}x)\right) \\
&= 1 + O\left((\lambda-1)n^{-0.2}\log n\right),
\end{aligned} \tag{107}$$

其中，最后一个等式利用了以下事实：对于 $c_n = \sqrt{24\log n}$，有

$$\int_{\mathcal{A}^c} x^2\varphi(\mathrm{d}x) = 2\int_{c_n}^{\infty} x^2\varphi(\mathrm{d}x) \lesssim \int_{c_n}^{\infty} x^2 \exp\left(-\frac{1}{2}x^2\right)\mathrm{d}x \lesssim \frac{\log n}{n^{12}}.$$

关于另一项 $I_2(x)$，引理 22a 中的关系 (3) 意味着 $|I_2(x)| \lesssim 1$。此外，若 $|x| \le \sqrt{24\log n}$，根据关系 (22b) 可得 $|I_2(x)| = 1 + O((\lambda-1)n^{-0.2}\log n)$。将这些结果合并，可得

$$\begin{aligned}
\int |I_2(x)|^2\varphi(\mathrm{d}x) &= \int_{\mathcal{A}} |I_2(x)|^2\varphi(\mathrm{d}x) + \int_{\mathcal{A}^c} |I_2(x)|^2\varphi(\mathrm{d}x) \\
&= 1 + O\big((\lambda-1)n^{-0.2}\log n\big) + O\big(P(\mathcal{A}^c)\big) \\
&= 1 + O\big((\lambda-1)n^{-0.2}\log n\big).
\end{aligned} \tag{108}$$

结合不等式 (107) 和 (108) 可得

$$\kappa_t^2 = 1 + O\left((\lambda-1)n^{-0.2}\log n\right) = 1 + o\left(\frac{\lambda-1}{\log n}\right).$$

情形 II： $\sqrt{\lambda^2-1}\,n^{-0.1} \lesssim \alpha_t \lesssim \sqrt{\lambda^2-1}$。我们首先注意到，在假设（23）下，当 $\|\xi_{t-1}\|_2\ t \lesssim \log_{\lambda-1} n$ 时，以下关系成立：

$$\|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{t^3\log n}{n}} \lesssim \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}. \tag{109}$$

我们回顾性质（94）中得到的基本事实，因此，

$$\int |I_1(x)|^2\varphi(\mathrm{d}x) = \int \left[\left(\gamma_t\pi_t x + \frac{2}{\sqrt{n}}\gamma_t\pi_t^2\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right)\cdot\left(1-\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right)\right]^2\varphi(\mathrm{d}x) \tag{110}$$

$$\int |I_2(x)|^2\varphi(\mathrm{d}x) = \int \left[\gamma_t\pi_t\left(1-\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right)\right]^2\varphi(\mathrm{d}x). \tag{111}$$

因此，问题归结为在 $\sqrt{\lambda^2-1}n^{-0.1} \lesssim \alpha_t \le \sqrt{\lambda^2-1}$. 条件下控制上述两个表达式的右侧

为方便记号，我们引入以下额外的简写符号：

$$\begin{aligned}
J_1(x) &:= \gamma_t\pi_t x + \frac{2}{\sqrt{n}}\gamma_t\pi_t^2\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right),\\
J_2(x) &:= \mu_t\alpha_t x + 2\mu_t\alpha_t^2\tanh\Big(\alpha_t(\alpha_t+x)\Big),\\
K_1(x) &:= 1-\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right),\\
K_2(x) &:= 1-\tanh^2\Big(\alpha_t(\alpha_t+x)\Big)\\
\mu_t &:= \left[\int \tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)\right]^{-\frac{1}{2}}.
\end{aligned}$$

有了这组记号，由（110）及简单的代数运算可得

$$\begin{aligned}
\int |I_1(x)|^2\varphi(\mathrm{d}x) &= \int [J_1(x)K_1(x)]^2\varphi(\mathrm{d}x)\\
&= \int [J_2(x)K_2(x)]^2\varphi(\mathrm{d}x) + O\left(\int |J_2(x)|^2\,|K_1(x)-K_2(x)|\,|K_1(x)+K_2(x)|\,\varphi(\mathrm{d}x)\right)\\
&\quad + O\left(\int |J_1(x)-J_2(x)||J_1(x)+J_2(x)||K_1(x)|^2\varphi(\mathrm{d}x)\right).
\end{aligned} \tag{112}$$

要对 $\int |I_1(x)|^2\varphi(\mathrm{d}x)$进行界定，只需分别控制（112）右侧的这三项。继续之前，我们先说明几个有用的初步性质。

- 首先，我们想证明

$$\mu_t \asymp \alpha_t^{-1}. \tag{113}$$

为了解释这一点，请注意初等事实：双曲正切$^2(x)\ \le x^2$ 意味着

$$\int \tanh^2(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x) \le \int \alpha_t^2(\alpha_t+x)^2\varphi(\mathrm{d}x) \asymp \alpha_t^2, \qquad \text{for } \alpha_t \le \lambda.$$

另一方面，当 $|x| \le 1/2$, $\alpha_t \in (0, \lambda]$ 且 $\lambda \in (1, 1.2]$ 时，有 $\alpha_t(\alpha_t + x) \in [-0.0625, 2.04]$。显然，对于任意 $z \in [-0.0625, 2.04]$，都有双曲正切$^2(z)/z^2 \ge 0.22$，因此，

$$\begin{aligned}\int \tanh^2(\alpha_t(\alpha_t + x))\varphi(\mathrm{d}x) &\ge \int \tanh^2(\alpha_t(\alpha_t + x))\mathbf{1}(|x| \le 0.5)\varphi(\mathrm{d}x) \\ &\gtrsim \int \alpha_t^2(\alpha_t + x)^2\mathbf{1}(|x| \le 0.5)\varphi(\mathrm{d}x) \asymp \alpha_t^2.\end{aligned}$$

前两个界合在一起证明了这一结论（113）。

- 此外，根据引理2中的方程（20a）和界（109），我们有

$$\pi_t = \sqrt{n}\alpha_t + \frac{\sqrt{n}}{\alpha_t} O\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right) = \alpha_t\sqrt{n} + O\left(\sqrt{\frac{\log^4 n}{\alpha_t^2(\lambda-1)^3}}\right) \tag{114}$$

$$= \alpha_t\sqrt{n}\left(1 + O\left(\sqrt{\frac{\log^4 n}{n\alpha_t^4(\lambda-1)^3}}\right)\right) \asymp \alpha_t\sqrt{n} \tag{115}$$

只要 $t \lesssim \log_{\lambda-1} n$。这里，最后一个不等式成立是因为

$$\sqrt{\frac{\log^4 n}{(\lambda-1)^3}} \lesssim (\lambda-1)\cdot\frac{\log^2 n}{(\lambda-1)^{5/2}} = o\big((\lambda-1)n^{0.3}\big) = o\big(\alpha_t^2\sqrt{n}\big),$$

只要 $\alpha_t \gtrsim \sqrt{\lambda^2 - 1}n^{-0.1}$ 且 $\lambda - 1 \gtrsim n^{-1/9}\log n$。

- 此外，结合引理 2 中的方程（20b）与（114）、（115）和（109），可得

$$\begin{aligned}\gamma_t^{-2} &= n\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x) + O\left(\alpha_t^2\sqrt{\frac{n\log^4 n}{(\lambda-1)^3}}\right) \\ &= n\int \tanh^2\left(\alpha_t(\alpha_t + x)\right)\varphi(\mathrm{d}x) + O\left(\sqrt{\frac{n\log^4 n}{(\lambda-1)^3}}\right) \\ &= n\mu_t^{-2} + O\left(\sqrt{\frac{n\log^4 n}{(\lambda-1)^3}}\right).\end{aligned}$$

这里，第二个等式来自以下事实（另见（146））：

$$\begin{aligned}&\left|\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right) - \tanh^2\left(\alpha_t(\alpha_t + x)\right)\varphi(\mathrm{d}x)\right| \\ &\le \left|1 - \frac{\alpha_t^2 n}{\pi_t^2}\right|\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x) + \left|\int \frac{\alpha_t^2 n}{\pi_t^2}\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right) - \tanh^2\left(\alpha_t(\alpha_t + x)\right)\varphi(\mathrm{d}x)\right| \\ &\lesssim \left|1 - \frac{\alpha_t^2 n}{\pi_t^2}\right|\alpha_t^2 + \left|\frac{\pi_t^2}{\alpha_t^2 n} - 1\right|\alpha_t^4 = O\left(\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right).\end{aligned} \tag{116}$$

因此，我们可以将 γ_t 表示为 μ_t 的形式：

$$\gamma_t = \left[n\mu_t^{-2} + O\left(\sqrt{\frac{n\log^4 n}{(\lambda-1)^3}}\right)\right]^{-1/2} = \frac{\mu_t}{\sqrt{n}}\left(1 + O\left(\mu_t^2\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right)\right) \asymp \frac{\mu_t}{\sqrt{n}}. \tag{117}$$

- 利用（115）和（117），再经过简单的代数运算可得

$$|\gamma_t\pi_t - \mu_t\alpha_t| \lesssim \mu_t\alpha_t\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\left(\frac{1}{\alpha_t^2}+\mu_t^2\right) \lesssim \frac{1}{\alpha_t^2}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}, \tag{118}$$

$$\left|\frac{1}{\sqrt{n}}\gamma_t\pi_t^2 - \mu_t\alpha_t^2\right| \lesssim \mu_t\alpha_t^2\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\left(\frac{1}{\alpha_t^2}+\mu_t^2\right) \lesssim \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}. \tag{119}$$

此外，根据引理 2中的关系式（20d），可得

$$\gamma_t^2\pi_t^2 = (\alpha_t^2+1)^{-1} + O\left(\frac{\pi_t^2}{n} + \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right),$$

并采用与推导（20c）类似的分析，可得

$$\mu_t^{-2} = \int \tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x) = \alpha_t^2(\alpha_t^2+1) + O\left(\alpha_t^4\right).$$

将这两个界结合起来，经过简单的代数运算可得

$$|\gamma_t\pi_t - \mu_t\alpha_t| \lesssim \frac{\pi_t^2}{n} + \alpha_t^2 + \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}} \lesssim \alpha_t^2 + \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}} \lesssim \alpha_t^2, \tag{120}$$

注意到范围 $\sqrt{\lambda^2-1}\,n^{-0.1} \lesssim \alpha_t \lesssim \sqrt{\lambda^2-1}$ 以及我们的假设 $\lambda - 1 \gtrsim n^{-1/9}\log n$. 将上述界与（118）结合，得到

$$|\gamma_t\pi_t - \mu_t\alpha_t| \lesssim \alpha_t^2 \wedge \frac{1}{\alpha_t^2}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}} \lesssim \left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4}, \tag{121}$$

其中，最后一个不等式由以下初等事实得出：$\min\{a,b\} \le \sqrt{ab}$.

- 取不等式　　性（115）、（118）和（119）以及这一事实 $\mu_t\alpha_t^2 \asymp \alpha_t$ （参见（113））可得

$$|K_1(x) - K_2(x)| \lesssim \left|\frac{\pi_t}{\sqrt{n}} - \alpha_t\right||\alpha_t + x| \lesssim \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\cdot(|x|+\alpha_t), \tag{122}$$

and

$$\begin{aligned}
|J_1(x) - J_2(x)| &\le |\gamma_t\pi_t - \mu_t\alpha_t|\,|x| + 2\left|\frac{1}{\sqrt{n}}\gamma_t\pi_t^2 - \mu_t\alpha_t^2\right|\left|\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\right| \\
&\quad + \mu_t\alpha_t^2\left|\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right) - \tanh\left(\alpha_t(\alpha_t+x)\right)\right| \\
&\lesssim \left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4}|x| + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}} + \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}(|x|+\alpha_t) \\
&\lesssim \left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4}|x| + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}.
\end{aligned} \tag{123}$$

利用上述关系，我们可以控制表达式（112）右侧的值。直接结合（122）和（123）即可得到

$$\int |I_1(x)|^2\varphi(\mathrm{d}x) = \int [J_2(x)K_2(x)]^2\varphi(\mathrm{d}x) + O\left(\frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right) + O\left(\left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4} + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right)$$

$$= \int \left[\mu_t\alpha_t x + 2\mu_t\alpha_t^2 \tanh\left(\alpha_t(\alpha_t+x)\right)\right]^2 \left[1-\tanh^2\left(\alpha_t(\alpha_t+x)\right)\right]^2 \varphi(\mathrm{d}x) + O\left(\left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4} + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right), \tag{124}$$

其中，第一行用到了简单观测：$|J_i(x)| \lesssim |x| + \alpha_t$ 和 $|K_i(x)| \lesssim 1$，其中 $i = 1, 2$。类似地，可以用相同的方法推导出

$$\int |I_2(x)|^2 \varphi(\mathrm{d}x) = \mu_t^2\alpha_t^2 \int \left[1-\tanh^2\left(\alpha_t(\alpha_t+x)\right)\right]^2 \varphi(\mathrm{d}x) + O\left(\left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4} + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right). \tag{125}$$

事实证明，上述两个恒等式中的主要项满足

$$\int \left[\mu_t\alpha_t x + 2\mu_t\alpha_t^2 \tanh\left(\alpha_t(\alpha_t+x)\right)\right]^2 \left[1-\tanh^2\left(\alpha_t(\alpha_t+x)\right)\right]^2 \varphi(\mathrm{d}x) \le 1 - c\alpha_t^2, \tag{126a}$$

$$\mu_t^2\alpha_t^2 \int \left[1-\tanh^2\left(\alpha_t(\alpha_t+x)\right)\right]^2 \varphi(\mathrm{d}x) \le 1 - c\alpha_t^2, \tag{126b}$$

我们马上会对此给出论证。结合这些结果与（104），可得：若 $\sqrt{\lambda^2-1}\, n^{-0.1} \lesssim \alpha_t \lesssim \sqrt{\lambda^2-1}$，则

$$\kappa_t^2 \le 1 - c\alpha_t^2 + O\left(\left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4} + \frac{1}{\alpha_t}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}}\right) = 1 + o\left(\frac{\lambda-1}{\log n}\right), \tag{127}$$

其中最后一个不等式可通过注意到以下关系得出

$$\left(\frac{\log^4 n}{n(\lambda-1)^3}\right)^{1/4} + \frac{1}{\sqrt{\lambda-1}n^{-0.1}}\sqrt{\frac{\log^4 n}{n(\lambda-1)^3}} = o\left(\frac{\lambda-1}{\log n}\right)$$

在我们的假设 $\lambda - 1 \gtrsim n^{-1/9}\log n$ 下。

关系（126）的证明。接下来，考虑根据带噪观测 $v^\star$（其服从 $v_i^\star \sim$均匀分布$\{\pm\frac{1}{n}\}$）来估计 $Y = \alpha_t v^\star + g$ 的问题，其中 $g \sim \mathcal{N}(0, \frac{1}{n}I_n)$。如前所述，贝叶斯最优估计（或最小均方估计量（MMSE））为

$$\mathbb{E}[v^\star \mid Y] = \tanh(\sqrt{n}\alpha_t Y),$$

根据其最优性，它满足

$$\mathrm{Cor}\left(v^\star, f(Y)\right) \le \mathrm{Cor}\left(v^\star, \mathbb{E}[v^\star \mid Y]\right), \tag{128}$$

对于任意可测函数 f；其中$\mathrm{Cor}(\cdot,\cdot)$表示两个随机向量的相关性。特别地，贝叶斯最优估计量优于恒等估计量（即，$f(Y) = Y$），因此（128）可改写为

$$\frac{\mathbb{E}[\langle v^\star, \alpha_t v^\star + g\rangle]}{\sqrt{\mathbb{E}[\|\alpha_t v^\star + g\|_2^2]}} \le \frac{\mathbb{E}[\langle v^\star, \tanh(\sqrt{n}\alpha_t(\alpha_t v^\star + g))\rangle]}{\sqrt{\mathbb{E}[\|\tanh(\sqrt{n}\alpha_t(\alpha_t v^\star + g)\|_2^2]}} = \frac{\int \tanh(\alpha_t(\alpha_t+x))\varphi(dx)}{\sqrt{\int \tanh^2(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x)}}$$
$$= \sqrt{\int \tanh^2(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x)}, \tag{129}$$

第一个等号成立是由于 $\varphi(\cdot)$的对称性；第二个等号成立是因为 $\int$ 双曲正切$(\alpha^2+\alpha x)\varphi(\mathrm{d}x) = \int$ 双曲正切$^2(\alpha^2+\alpha x)\varphi(\mathrm{d}x)$ （参见Deshpande 等人(2017 年，附录 B.2））。因此，上述关系意味着

$$\frac{\alpha_t}{\sqrt{\alpha_t^2+1}} = \frac{\mathbb{E}[\langle v^\star,\, \alpha_t v^\star + g\rangle]}{\sqrt{\mathbb{E}[\|\alpha_t v^\star + g\|_2^2]}} \le \sqrt{\int \tanh^2(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x)} = \frac{1}{\mu_t}, \tag{130}$$

这反过来揭示了

$$\mu_t\alpha_t \le \sqrt{\alpha_t^2+1} =: \gamma.$$

借助这一关系，我们可以得出

$$\begin{aligned}
&\max\left\{\int\Big[\mu_t\alpha_t x + 2\mu_t\alpha_t^2\tanh\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]\Big[1-\tanh^2\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]\varphi(\mathrm{d}x),\right.\\
&\qquad\left.\mu_t^2\alpha_t^2\int\Big[1-\tanh^2\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]^2\varphi(\mathrm{d}x)\right\}\\
&\le \gamma^2\max\left\{\int\Big[x+2\alpha_t\tanh\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]\Big[1-\tanh^2\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]\varphi(\mathrm{d}x),\right.\\
&\qquad\left.\int\Big[1-\tanh^2\Big(\alpha_t\big(\alpha_t+x\big)\Big)\Big]^2\varphi(\mathrm{d}x)\right\}\\
&=: \kappa^2(\gamma,\alpha_t^2).
\end{aligned}$$

事实证明，函数 $\kappa^2(\cdot,\ \cdot)$ 已在 Li 和 Wei （2022年）中得到研究；更具体地说，Li 和 Wei（2022年，关系式 (272)）与 $\gamma := \sqrt{\alpha_t^2+1}$ 表明 $\kappa(\gamma,\ \alpha_t^2) \le 1-\frac{\gamma-1}{12}$，因此

$$\kappa^2(\gamma,\alpha_t^2) \le 1-\frac{\gamma-1}{12} \le 1-c\alpha_t^2 \tag{131}$$

对于某个常数 $c>0$。这里请注意，在 γ 和 α_t^2 中，我们分别将其视为 λ 和 τ，分别对应于 Li 和 Wei （2022年，关系式 (272)）。将所有内容汇总起来，即完成了所需关系式（126）的证明。

情形 III： $\alpha_t \gtrsim \sqrt{\lambda^2-1}$ **和** $t < \frac{cn}{(\lambda-1)^5\,\text{对数}^2\,n}$。本情形下 κ_t 的计算可由与 Li 和 Wei （2022年，D.3.4 节）中类似的论证得到。唯一的区别在于参数 π_t 和 γ_t的计算；该计算在其中由 Li 和 Wei （2022年，引理 14）完成，但在此处需要不同的证明。具体而言，我们旨在证明

$$\pi_t = \big(1+o(\lambda-1)\big)\alpha_t\sqrt{n} \qquad \text{and} \qquad \gamma_t^{-2} = \big(1+o(\lambda-1)\big)n\int\tanh(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x). \tag{132}$$

如果这两个关系成立，那么就可以原封不动地沿用（Li 和 Wei，2022年，D.3.4 节）中的论证来证明

$$\kappa_t \le 1-\frac{1}{15}(\lambda-1),$$

如所声称的那样。

我们现在证明 说明如何证明关系（132）。根据引理 2中的方程（20a），我们有

$$\begin{aligned}
\pi_t &= \alpha_t\sqrt{n}\left(1+\frac{1}{\alpha_t^2}O\Big(\|\xi_{t-1}\|_2+\sqrt{\frac{t\log n}{n}}\Big)\right) = \alpha_t\sqrt{n}\left(1+O\left(\sqrt{\frac{(t+\log^3 n/(\lambda-1))\log n}{n(\lambda-1)^2}}\right)\right)\\
&= \alpha_t\sqrt{n}\left(1+O\left(\sqrt{\frac{(\lambda-1)^2}{\log n}}\right)\right) = \alpha_t\sqrt{n}(1+o(\lambda-1))
\end{aligned}$$

在条件 $t \lesssim {}^{n}{}_{(\lambda-1)^{5}}{}_{\text{对数}^{2}n}$ 以及假设（25）下。此外，采用与不等式（116）相同的分析，我们可以保证

$$\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)\varphi(\mathrm{d}x) = \int \tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x) + O\left(\sqrt{\frac{(t+\log^3 n/(\lambda-1))\log n}{n\alpha_t^2(\lambda-1)^2}}\right)$$
$$= \int \tanh\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x) + O\left(\sqrt{\frac{(\lambda-1)^2}{\log n}}\right) = \int \tanh\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x) + o(\lambda-1).$$

然后，根据方程 (20b)，可得

$$\gamma_t^{-2} = n\left(\int \tanh(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x) + o(\lambda-1)\right) + n\alpha_t^2\big(1+o(\lambda-1)\big)O\left(\sqrt{\frac{(t+\log^3 n/(\lambda-1))\log n}{n}}\right)$$
$$= \big(1+o(\lambda-1)\big)n\int \tanh(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x),$$

其中，最后一个等号由以下事实得出：$\int$双曲正切$(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x) \asymp \alpha_t^2 \asymp 1$ (见关系 (113))。由此证明了论断 (132)。

A.4 论断的证明(43)

本小节旨在证明所述分解 (43)。为此，回顾一下，$\{\eta_i(x_i)\}_{1\le i\le t-1}$与以下对象张成相同的线性空间：$\{z_i\}_{1\le i\le t-1}$ (见 (37) 和 (38b))。需要注意的是，$\{\eta_1(x_1)\ldots,\eta_{t-1}(x_{t-1})\}$彼此近乎正交，因此构成一组近乎标准正交基；这一性质总结于下方的引理中，其证明见第 A.8 节。

引理 5. 假设　　定理 *1* 的假设成立。以至少 $1-O(n^{-11})$的概率，我们有　　*e*

$$\Big\|\sum_{i=1}^{t} w_i\eta_i(x_i)\Big\|_2 = \big(1+o(1)\big)\|w\|_2 \tag{133}$$

对所有 $t\le\tau_0$ 和所有 $w=[w_i]_{1\le i\le t}\in\mathbb{R}^t$同时成立，其中 τ_0 定义见 (42)。

鉴于引理 5 以及 $\xi_t\in\mathsf{span}(U_{t-1}) = \mathsf{span}\{\eta_1(x_1)\ldots,\eta_{t-1}(x_{t-1})\}$ (参见 (37))，可以将 ξ_t 表示为 $\{\eta_i(x_i)\}_{1\le i\le t-1}$ 的线性组合，如下所示：

$$\xi_t = \sum_{k=1}^{t-1}\gamma_t^k\eta_k(x_k), \qquad \text{with } \gamma_t=[\gamma_t^k]_{1\le k<t}\in\mathbb{R}^{t-1} \text{ obeying } \|\gamma_t\|_2\asymp\|\xi_t\|_2. \tag{134}$$

有了这个分解，我们打算证明

$$\alpha_{t+1} = \lambda v^{\star\top}\eta_t(x_t) = \lambda v^{\star\top}\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(x_k)\Big)$$
$$= \lambda v^{\star\top}\eta_t(v_t) + O\left(\frac{\log^{2.5} n}{n^{3/4}(\lambda-1)^{1.5}}\right), \tag{135}$$

具体证明如下。

- 为了看出这一点，首先注意到 $\eta_t(\cdot)$ 是一个利普希茨函数，其利普希茨常数为 $O(1)$（见引理 3）。因此，对于每个 $t\lesssim{}^{n}_{\text{对数}\lambda-1}$，我们有 ——

$$\left|v^{\star\top}\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(x_k)\Big) - v^{\star\top}\eta_t\left(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right)\right|$$

$$\leq \Big\|\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(x_k)\Big)-\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\Big)\Big\|_2 \lesssim \sum_{k=1}^{t-1}\big|\gamma_{t-1}^k\big|\,\|\eta_k(x_k)-\eta_k(v_k)\|_2\,.$$

鉴于分解（134）以及柯西-施瓦茨不等式，我们还可以进一步得到

$$\begin{aligned}
&\Big|v^{\star\top}\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(x_k)\Big)-v^{\star\top}\eta_t\Big(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\Big)\Big| \lesssim \sum_{k=1}^{t-1}\big|\gamma_{t-1}^k\big|\,\|\eta_k(x_k)-\eta_k(v_k)\|_2\\
&\quad\lesssim \sum_{k=1}^{t-1}\big|\gamma_{t-1}^k\big|\,\|\xi_{k-1}\|_2 \leq \|\gamma_{t-1}\|_2\Big(\sum_{k=1}^{t-1}\|\xi_{k-1}\|_2^2\Big)^{1/2}\\
&\quad\asymp \|\xi_{t-1}\|_2\Big(\sum_{k=1}^{t-1}\|\xi_{k-1}\|_2^2\Big)^{1/2} \lesssim \sqrt{\frac{t^3\log n}{n}}\cdot\left(\sum_{k=1}^{t-1}\frac{k^3\log n}{n}\right)^{1/2} \lesssim \frac{\log^{4.5}n}{n(\lambda-1)^{3.5}},
\end{aligned}$$

其中最后一行调用了 $\|\xi_t\|_2 \lesssim \sqrt{t^3\log_n n}$（参见（36））以及 $t \lesssim \log_{\lambda-1} n$。

- 此外，当 $|\alpha_t| \lesssim \sqrt{\lambda-1}n^{-1/4}$时，我们知道 $\big\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\big\|_\infty \lesssim \sqrt{t\log_n n}$ 在事件 $\{\phi_k\}_{k=1}^{t-1}\in\mathcal{E}$（在引理 1 中定义，其中 $\delta = O(n^{-10})$）的条件下。因此可保证

$$|v_{t,i}| = \left|\alpha_t v_i^\star+\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_{k,i}\right| \leq \frac{|\alpha_t|}{\sqrt{n}}+\left|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_{k,i}\right| \lesssim \sqrt{\frac{t\log n}{n}}, \tag{136a}$$

$$\begin{aligned}
\left|v_{t,i}+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_{k,i})\right| &\lesssim |v_{t,i}|+\sum_{k=1}^{t-1}\big|\gamma_{t-1}^k\big|\,|v_{k,i}| \lesssim |v_{t,i}|+\|\gamma_{t-1}\|_2\|\widetilde{v}_{t-1,i}\|_2\\
&\lesssim \sqrt{\frac{t\log n}{n}}+\sqrt{\frac{t^3\log n}{n}}\cdot\sqrt{\frac{t^2\log n}{n}} \lesssim \sqrt{\frac{t\log n}{n}},
\end{aligned} \tag{136b}$$

对于每个 $1\leq i<t$，其中我们记 $\widetilde{v}_{t-1,i} := (v_{1,i}, v_{2,i}, \ldots, v_{t-1,i})$。为说明（136b）为何成立，我们注意到，第一个不等式应用了引理 3，第二个不等式源自柯西-施瓦茨不等式，而最后一行使用了（136a）以及 $\|\gamma_{t-1}\|_2 \asymp \|\xi_{t-1}\|_2 \lesssim \sqrt{t^3\log_n n}$ 这一事实（参见（134）和（36））。此外，鉴于 $t \lesssim \log n/(\lambda-1)$ 且 $\lambda-1 \gtrsim n^{-1/9}\log n$，我们有$t^8 \lesssim n/\log n$。对引理 5 的证明中不等式（166）重复上述论证，可以确保对于任意 $2\leq k\leq 14$,

$$\sum_{i=1}^n |v_{t,i}|^k = \sum_{i=1}^t \big|v_{t,(i)}\big|^k + \sum_{i=t+1}^n \big|v_{t,(i)}\big|^k \lesssim \left(\frac{\log n}{n}\right)^{k/2-1}, \tag{137a}$$

and

$$\sum_{i=1}^n\left|v_{t,i}+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_{k,i})\right|^k \lesssim \left(1+t^{2k}\left(\frac{\log n}{n}\right)^{k/2}\right)\left(\frac{\log n}{n}\right)^{k/2-1} \lesssim \left(\frac{\log n}{n}\right)^{k/2-1}. \tag{137b}$$

这里，我们注意到

$$\begin{aligned}
&\sum_{i=1}^n\left|\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_{k,i})\right|^k\\
&\lesssim \sum_{i=1}^n\|\gamma_{t-1}\|_2^k\Big(\sum_{k=1}^{t-1}\eta_k^2(v_{k,i})\Big)^{k/2} \lesssim \sum_{i=1}^n\left(\frac{t^3\log n}{n}\right)^{k/2}\Big(\sum_{k=1}^{t-1}v_{k,i}^2\Big)^{k/2}\\
&\lesssim \left(\frac{t^3\log n}{n}\right)^{k/2}\sum_{i=1}^n\Big(\sum_{k=1}^{t-1}v_{k,(i)}^2\Big)^{k/2}
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \left(\frac{t^3\log n}{n}\right)^{k/2}\cdot t\left(\frac{t^2\log n}{n}\right)^{k/2}+\left(\frac{t^3\log n}{n}\right)^{k/2}\left(\sum_{k=1}^{t-1}v_{k,(t+1)}^2\right)^{k/2-1}\cdot\sum_{i=t+1}^{n}\left(\sum_{k=1}^{t-1}v_{k,(i)}^2\right)\\
&\lesssim t^{\frac{5k}{2}+1}\left(\frac{\log n}{n}\right)^{k}+t^{2k}\left(\frac{\log n}{n}\right)^{k-1}\\
&\lesssim t^{2k}\left(\frac{\log n}{n}\right)^{k-1},
\end{aligned}
$$

其中第二行利用了这一事实 $\|\gamma_{t-1}\|_2 \asymp \|\xi_{t-1}\|_2 \lesssim \sqrt{t^3 \text{对数}_n n}$（参见（134）和（36））以及引理 3，倒数第三行使用了不等式（136）；倒数第二行源于以下事实 $\|v_k\|_2 \lesssim 1$（例如参见（87））并且在事件 $\{\phi_k\}_{k=1}^{t-1}\in\mathcal{E}$ 条件下，

$$
|v_{k,(t+1)}| \le \frac{|\alpha_k|}{\sqrt{n}}+\Big|\sum_{i=1}^{k-1}\beta_{k-1}^i\phi_i\Big|_{(t+1)} \lesssim \sqrt{\frac{\log n}{n}};
$$

最后一行由以下事实得出：$t^{k/2+1}\lesssim t^8\lesssim n/\text{对数}\ n$。因此，结合引理 3 和表达式（136），可得

$$
\begin{aligned}
&\eta_t\left(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right)-\eta_t\left(v_t\right)\\
&=(1-c_0)\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)+O\left(\pi_t^2\right)\cdot\left[\left(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right)^3-\left(v_t\right)^3\right]+c_x
\end{aligned}
\tag{138}
$$

对于某些向量 $c_x\in\mathbb{R}^n$，参数满足 $\pi_t\lesssim(\lambda-1)^{-3/4}n^{1/4}\text{对数}\ n$ 且 $c_0\lesssim \text{对数}\sqrt[4]{n(\lambda-1)^3}$。这里，最后一个方程用到了以下事实：

$$
\begin{aligned}
\|c_x\|_2 &\lesssim \left\|\frac{n\left|v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right|^5\log^4 n}{(\lambda-1)^3}\right\|_2+\left\|\frac{n|v_t|^5\log^4 n}{(\lambda-1)^3}\right\|_2\\
&\lesssim \frac{n\log^4 n}{(\lambda-1)^3}\sqrt{\left(\frac{\log n}{n}\right)^4}\lesssim\frac{\log^6 n}{n(\lambda-1)^3},
\end{aligned}
\tag{139}
$$

其中，性质（137）应用于 $k=10$. 接下来，注意到

$$
\begin{aligned}
&\left\|\left(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right)^3-\left(v_t\right)^3\right\|_2\\
&=\left\|\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\circ\left((v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k))^2+\left(v_t\right)^2+(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k))\circ v_t\right)\right\|_2\\
&\lesssim\left\|\max_{1\le k\le t-1}\eta_k(v_k)\circ\left((v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k))^2+\left(v_t\right)^2\right)\right\|_2\cdot\sqrt{t}\|\xi_{t-1}\|_2\\
&\lesssim\max_{1\le k\le t-1}\|\eta_k(v_k)\|_\infty\cdot\left(\left\|(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k))^2\right\|_2+\left\|\left(v_t\right)^2\right\|_2\right)\cdot\sqrt{t}\|\xi_{t-1}\|_2\\
&\lesssim\frac{t\log n}{n}\cdot\|\xi_{t-1}\|_2\lesssim\frac{\log^4 n}{n^{1.5}(\lambda-1)^{2.5}},
\end{aligned}
$$

其中，最后一行可通过将 $k=4$ 应用于性质（136）和（137）得到。这里，我们用到了以下事实：$\|\xi_t\|_2\lesssim\sqrt{t^3\text{对数}_n n}$（参见（36））， $t\lesssim\text{对数}_{\lambda-1}n$ 以及

$$
\left\|v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right\|_2\le\|v_t\|_2+\left\|\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right\|_2\le 1+\sqrt{t}\|\gamma_{t-1}\|_2\le 1+\sqrt{\frac{t^4\log n}{n}}\lesssim 1.
$$

将这些结果结合起来，我们得到

$$\left\|\eta_t\left(v_t+\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right)-\eta_t\left(v_t\right)-(1-c_0)\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)\right\|_2=O\left(\frac{\log^6 n}{n(\lambda-1)^3}\right).$$

- 最后，只需考虑 $v^{\star\top}\sum_{k=1}^{t-1}\gamma_{t-1}^k\eta_k(v_k)$，我们将如下进行控制：

$$\begin{aligned}\left|\sum_{k=1}^{t-1}\gamma_{t-1}^k v^{\star\top}\eta_k\left(v_k\right)\right| &\le \left|\sum_{k=1}^{t-1}\gamma_{t-1}^k\left[v^{\star\top}\eta_k\left(x_k\right)+O(\|\xi_{k-1}\|_2)\right]\right| \\ &= \left|\sum_{k=1}^{t-1}\gamma_{t-1}^k\left(\frac{\alpha_{k+1}}{\lambda}+O\left(\sqrt{\frac{k^3\log n}{n}}\right)\right)\right| \\ &\lesssim \sqrt{t}\|\gamma_{t-1}\|_2\cdot\frac{\sqrt{\lambda-1}}{n^{1/4}} \asymp \sqrt{t}\|\xi_{t-1}\|_2\cdot\frac{\sqrt{\lambda-1}}{n^{1/4}} \lesssim \frac{\log^{2.5} n}{n^{3/4}(\lambda-1)^{1.5}}.\end{aligned}$$

将上述三个不等式合并，便得到所需的界（135）。

基于表达式（135），我们现在继续证明断言（43）。首先，回顾一下$v_t := \alpha_t v^\star+\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k$。若定义 $g_{t-1} := v^{\star\top}\phi_{t-1}\sim\mathcal{N}\left(0,\ \frac{1}{n}\right)$ 以及 $\widetilde{\beta}_{t-1} := [\beta_{t-1}^1,\ldots,\beta_{t-1}^{t-2}]$，经简单的代数运算可得

$$\begin{aligned}\left|v^{\star\top}\left(v_t-\alpha_t v^\star-\phi_{t-1}\right)\right| &= \left|v^{\star\top}\Big[\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k-(1-\beta_{t-1}^{t-1})\phi_{t-1}\Big]\right| \\ &\le \Big|\sum_{k=1}^{t-2}\beta_{t-1}^k g_k\Big|+|1-\beta_{t-1}^{t-1}||g_{t-1}| \\ &= \Big|\sum_{k=1}^{t-2}\beta_{t-1}^k g_k\Big|+\frac{1-\left(\beta_{t-1}^{t-1}\right)^2}{1+|\beta_{t-1}^{t-1}|}|g_{t-1}| \\ &\lesssim \|\widetilde{\beta}_{t-1}\|_2\sqrt{\sum_{k=1}^{t-2}(g_k)^2}+\|\widetilde{\beta}_{t-1}\|_2^2|g_{t-1}| \lesssim \|\widetilde{\beta}_{t-1}\|_2\sqrt{\frac{t\log n}{n}} \\ &\lesssim \sqrt{\frac{t\log n}{n}}\cdot\left(\frac{t\sqrt{\lambda-1}}{n^{1/4}}+\frac{t\log^4 n}{\sqrt{n(\lambda-1)^3}}\right) \lesssim \frac{\log^2 n}{n^{3/4}(\lambda-1)}+\frac{\log^6 n}{n(\lambda-1)^3},\end{aligned}$$

其中，最后一个不等式由引理 5 证明中的界（162）以及条件 $t\lesssim_{\text{对数}}\lambda_{-1}^n$ 得出。根据上述计算，我们可以推导出

$$v^{\star\top}v_t=\alpha_t+g_{t-1}+O\left(\frac{\log^2 n}{n^{3/4}(\lambda-1)}+\frac{\log^6 n}{n(\lambda-1)^3}\right).$$

事实上，直接应用引理 3 还可进一步得到以下结论：

$$\begin{aligned}\left|v^{\star\top}\left(\eta_t(v_t)-v_t\right)\right| &\lesssim \left|v^{\star\top}\left[c_0 v_t+\pi_t^2(v_t\circ v_t\circ v_t)+O\left(\frac{n\log^4 n}{(\lambda-1)^3}|v_t|^5\right)\right]\right| \\ &\lesssim \frac{\log^4 n}{n^{3/4}(\lambda-1)},\end{aligned}\tag{140}$$

最后一个不等式的证明推迟到本小节末尾。

综上所述，综合考虑上述结果并利用关系式（135），我们得到

$$\alpha_{t+1}=\lambda v^{\star\top}v_t+\lambda v^{\star\top}\left(\eta_t\left(v_t\right)-v_t\right)+O\left(\frac{\log^{2.5} n}{n^{3/4}(\lambda-1)^{1.5}}\right)$$

$$= \lambda\alpha_t + \lambda g_{t-1} + O\left(\frac{\log^4 n}{n^{3/4}(\lambda-1)^{1.5}}\right). \tag{141}$$

因此，递归地应用上述关系式可得到所需的分解：

$$\alpha_{t+1} = \lambda^{t-k+1}\alpha_k + \sum_{i=1}^{t-k+1}\lambda^i g_{t-i} + O\left(\sum_{i=1}^{t-k+1}\lambda^i \frac{\log^4 n}{n^{3/4}(\lambda-1)^{1.5}}\right)$$

对于任意 $1 \le k \le t$。

不等式的证明 (140)。 为了证明不等式 (140)，我们先注意以下简单性质：以至少 $1 - O(n^{-11})$的概率，

$$\begin{aligned}
v^{\star\top}(v^\star \circ v^\star \circ v^\star) &= \frac{1}{n};\\
v^{\star\top}\Big(v^\star \circ v^\star \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) &= \frac{1}{n}v^{\star\top}\Big(\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) \lesssim \frac{1}{n};\\
v^{\star\top}\Big(v^\star \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) &= \frac{1}{n}\Big\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\|_2^2 \asymp \frac{1}{n};\\
v^{\star\top}\Big(\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) &\lesssim \frac{t\log n}{n}v^{\star\top}\Big(\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) \lesssim \sqrt{\frac{t^3\log^3 n}{n^3}}.
\end{aligned}$$

提醒读者， $v_i^\star \sim$ 均匀分布$(\pm\frac{1}{\sqrt{n}})$，并且我们用到了引理 1。

接下来，回顾 $v_t := \alpha_t v^\star + \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k$ 以得到

$$\begin{aligned}
\big|v^{\star\top}(v_t \circ v_t \circ v_t)\big| = \Big|&\alpha_t^3 v^{\star\top}(v^\star \circ v^\star \circ v^\star) + 3\alpha_t^2 v^{\star\top}\Big(v^\star \circ v^\star \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big)\\
&+3\alpha_t v^{\star\top}\Big(v^\star \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big) + v^{\star\top}\Big(\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k \circ \sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big)\Big|\\
&\lesssim \frac{\alpha_t^3}{n} + \frac{\alpha_t^2}{n} + \frac{\alpha_t}{n} + \sqrt{\frac{t^3\log^3 n}{n^3}}\\
&\lesssim \frac{1}{n^{5/4}},
\end{aligned}$$

其中，只要满足 $\alpha_t \lesssim \sqrt{\lambda-1}\,n^{-1/4}$，最后一行就成立。因此，为了推出 (140)，只需注意到 (137)、 $c_0 \lesssim$对数$\sqrt[4]{n(\lambda-1)^3}$、 $\pi_t \lesssim \frac{n^{1/4}}{(\lambda-1)^{3/4}}$以及

$$\big|v^{\star\top}v_t\big| = \Big|\alpha_t + v^{\star\top}\Big(\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k\Big)\Big| \lesssim \frac{\sqrt{\lambda-1}}{n^{1/4}} + \sqrt{\frac{t\log n}{n}} \lesssim \frac{\sqrt{\lambda-1}}{n^{1/4}}.$$

A.5 命题的证明(53)

为简化记号，我们不失一般性地假设 $\alpha_t > 0$ 在整个证明过程中成立。在深入证明命题（53）之前，让我们回顾引理2 以得到

$$\frac{\pi_t}{\alpha_t\sqrt{n}} = 1 + O\bigg(\frac{1}{\alpha_t^2}\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right) \wedge \frac{1}{\alpha_t}\left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2}\bigg)$$

$$= 1 + O\left(\frac{\log^2 n}{\alpha_t^2\sqrt{(\lambda-1)^3 n}} \wedge \left(\frac{\log^2 n}{\alpha_t^2\sqrt{(\lambda-1)^3 n}}\right)^{1/2}\right),$$

其中我们使用了 $\|\xi_t\|_2 \le \sqrt{t^3 \log n}$（见（36））以及 $t \lesssim \log_{\lambda-1} n$。进而，这意味着

$$\left|\frac{\pi_t^2}{\alpha_t^2 n} - 1\right| \alpha_t^2 \lesssim \frac{\log^2 n}{\sqrt{(\lambda-1)^3 n}}. \tag{142}$$

现在，我们来建立 α_t的递归关系。回顾3中 η_t 的定义以及定理2，可知

$$\begin{aligned}
\alpha_{t+1} &= \lambda v^{\star\top} \int \eta_t\left(\alpha_t v^\star + \frac{1}{\sqrt{n}} x\right) \varphi_n(\mathrm{d}x) + \Delta_{\alpha,t} \\
&= \lambda\gamma_t v^{\star\top} \int \tanh\left(\pi_t\big(\alpha_t v^\star + \frac{1}{\sqrt{n}} x\big)\right) \varphi_n(\mathrm{d}x) + \Delta_{\alpha,t} \\
&= \lambda\gamma_t\sqrt{n} \int \tanh\left(\frac{\pi_t}{\sqrt{n}}\left(\alpha_t + x\right)\right) \varphi(\mathrm{d}x) + \Delta_{\alpha,t},
\end{aligned} \tag{143}$$

最后一个等号由 $\varphi(\cdot)$的对称性得出，即

$$\frac{1}{\sqrt{n}} \int \tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right) \varphi(\mathrm{d}x) = -\frac{1}{\sqrt{n}} \int \tanh\left(\frac{\pi_t}{\sqrt{n}}(-\alpha_t + x)\right) \varphi(\mathrm{d}x).$$

与关系（113）类似的分析可得 $\int$双曲正切$^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x)$。$\asymp \frac{\pi_t^2}{n}$结合这一结果、引理2 和（36），得到

$$\gamma_t^{-2} = n\left(1 + O\left(\sqrt{\frac{t^3\log n}{n}}\right)\right) \int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}\left(\alpha_t + x\right)\right) \varphi(\mathrm{d}x). \tag{144}$$

结合（143）和（144），得到

$$\alpha_{t+1} = \left(1 + O\left(\sqrt{\frac{t^3\log n}{n}}\right)\right) \frac{\lambda \int \tanh\left(\frac{\pi_t}{\sqrt{n}}\left(\alpha_t + x\right)\right) \varphi(\mathrm{d}x)}{\left[\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}\left(\alpha_t + x\right)\right) \varphi(\mathrm{d}x)\right]^{1/2}} + \Delta_{\alpha,t}. \tag{145}$$

为证明命题（53），只需控制 $\int$双曲正切$\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x)$ 以及 $\int$双曲正切$^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x)$。为此，我们先来作几项观察会有所帮助。定义两个函数：

$$\begin{aligned}
f(z) &:= \frac{1}{z}\tanh(zy) - \tanh(y), \\
g(z) &:= \frac{1}{z^2}\tanh^2(zy) - \tanh^2(y).
\end{aligned}$$

双曲正切(zy)的泰勒展开为

$$\begin{aligned}
f'(z) &= -\frac{1}{z^2}\big[\tanh(zy) - zy + zy\tanh^2(zy)\big] = -\frac{2}{3}zy^3 + O(z^3y^5), \\
g'(z) &= -\frac{2\tanh(zy)}{z^3}\big[\tanh(zy) - zy + zy\tanh^2(zy)\big] = \frac{1}{3}zy^4 + O(z^3y^6),
\end{aligned}$$

通过直接计算可得以下关系

$$f(z) = \int_1^z f'(t)dt = -\frac{1}{3}(z^2-1)y^3 + O((z^4-1)y^5),$$

$$g(z)=\int_1^z g'(t)dt=\frac{1}{6}(z^2-1)y^4+O((z^4-1)y^6).$$

取 $z=\frac{\pi_t}{\alpha_t\sqrt{n}}$, ~~$y=\alpha_t$~~ (α_t+x)，我们可以看到

$$\frac{\alpha_t\sqrt{n}}{\pi_t}\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)-\tanh\left(\alpha_t(\alpha_t+x)\right)=-\frac{1}{3}\left(\frac{\pi_t^2}{\alpha_t^2n}-1\right)\alpha_t^3(\alpha_t+x)^3+O\left(\frac{\pi_t^4}{\alpha_t^4n^2}-1\right)\alpha_t^5(\alpha_t+x)^5,$$
$$\frac{\alpha_t^2n}{\pi_t^2}\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)-\tanh^2\left(\alpha_t(\alpha_t+x)\right)=\frac{1}{6}\left(\frac{\pi_t^2}{\alpha_t^2n}-1\right)\alpha_t^4(\alpha_t+x)^4+O\left(\frac{\pi_t^4}{\alpha_t^4n^2}-1\right)\alpha_t^6(\alpha_t+x)^6.$$

因此，我们可以得出

$$\begin{aligned}&\left|\int\frac{\alpha_t\sqrt{n}}{\pi_t}\tanh\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)-\tanh\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)\right|\\&=\left|\int\frac{1}{3}\left(\frac{\pi_t^2}{\alpha_t^2n}-1\right)\alpha_t^3(\alpha_t+x)^3\varphi(\mathrm{d}x)\right|+\int O\left(\frac{\pi_t^4}{\alpha_t^4n^2}-1\right)\alpha_t^5(\alpha_t+x)^5\varphi(\mathrm{d}x)\\&\lesssim\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^4.\end{aligned}$$

同样，我们可以证明

$$\left|\int\frac{\alpha_t^2n}{\pi_t^2}\tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t+x)\right)-\tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)\right|\lesssim\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^4.\tag{146}$$

将这些关系代入（145），得到

$$\begin{aligned}\alpha_{t+1}&=\lambda\Big(1+O\Big(\sqrt{\frac{t^3\log n}{n}}\Big)\Big)\frac{\int\tanh\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^4\right)}{\left[\int\tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^4\right)\right]^{1/2}}+\Delta_{\alpha,t}\\&=\lambda\Big(1+O\Big(\sqrt{\frac{t^3\log n}{n}}\Big)\Big)\frac{\int\tanh\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)}{\left[\int\tanh^2\left(\alpha_t(\alpha_t+x)\right)\varphi(\mathrm{d}x)\right]^{1/2}}\cdot\frac{1+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^2\right)}{1+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^2\right)}+\Delta_{\alpha,t}\\&=\lambda\left[\int\tanh^2\left(\alpha_t^2+\alpha_tx\right)\varphi(\mathrm{d}x)\right]^{-1/2}\int\tanh\left(\alpha_t^2+\alpha_tx\right)\varphi(\mathrm{d}x)+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^3+\sqrt{\frac{t^3\log n}{n}}\alpha_t+\Delta_{\alpha,t}\right),\end{aligned}\tag{147}$$

其中我们利用了如下事实（见（113））

$$\int\tanh\left(\alpha_t^2+\alpha_tx\right)\varphi(\mathrm{d}x)=\int\tanh^2\left(\alpha_t^2+\alpha_tx\right)\varphi(\mathrm{d}x)\asymp\alpha_t^2.$$

此外，（130）保证

$$\frac{\alpha_t}{\sqrt{\alpha_t^2+1}}\le\sqrt{\int\tanh^2(\alpha_t(\alpha_t+x))\varphi(\mathrm{d}x)}=\left[\int\tanh^2\left(\alpha_t^2+\alpha_tx\right)\varphi_n(\mathrm{d}x)\right]^{-1/2}\int\tanh\left(\alpha_t^2+\alpha_tx\right)\varphi(\mathrm{d}x),$$

进而得到

$$\alpha_{t+1}\ge\frac{\lambda\alpha_t}{\sqrt{\alpha_t^2+1}}+O\left(\left|\frac{\pi_t^2}{\alpha_t^2n}-1\right|\alpha_t^3+\sqrt{\frac{t^3\log n}{n}}\alpha_t+|\Delta_{\alpha,t}|\right).$$

最后，将上述结果与（142）结合起来即可得到

$$\alpha_{t+1}\ge\frac{\lambda\alpha_t}{\sqrt{\alpha_t^2+1}}+o\big((\lambda-1)\alpha_t\big)+O(\Delta_{\alpha,t})\tag{148}$$

其中我们再次　采用假设：$\lambda-1\gtrsim n^{-1/9}$ 对数 n. 至此，命题（53）得证。

A.6 命题的证明(56)

考虑以下情形:

$$|\alpha_t| < (\lambda-1)^{-3/4} n^{-1/4} \lesssim \sqrt{\lambda-1}\, n^{-0.1}.$$

首先，利用引理3中的性质 (22b) 来确保

$$\left|v^{\star\top}\eta_t(x_t) - v^{\star\top}\eta_t(v_t)\right| \lesssim |v^{\star\top}\xi_{t-1}| + (\lambda-1)n^{-0.2}(\log n)\|\xi_{t-1}\|_2. \tag{149}$$

进一步　界定 (149)，注意到 ξ_{t-1} 可以按以下方式用 $\{\eta_k(x_k)$进行分解　　$\}$:

$$\xi_{t-1} = \sum_{k=1}^{t-2} \gamma_{t-1}^k \eta_k(x_k), \qquad \text{with } \gamma_{t-1} = [\gamma_{t-1}^k]_{1\le k\le t-2} \in \mathbb{R}^{t-2} \text{ obeying } \|\gamma_{t-1}\|_2 \lesssim \|\xi_{t-1}\|_2;$$

这一命题的证明可见第 A.4 节（见其中的引理 5 及其证明）。根据这一关系，我们可以应用 (149) 和柯西-施瓦茨不等式，得到

$$\begin{aligned}
\left|v^{\star\top}\eta_t(x_t) - v^{\star\top}\eta_t(v_t)\right| &\lesssim \left|\sum_{k=1}^{t-2} \gamma_{t-1}^k v^{\star\top}\eta_k(x_k)\right| + (\lambda-1)n^{-0.2}(\log n)\|\xi_{t-1}\|_2 \\
&\lesssim \|\gamma_{t-1}\|_2 \left(\sum_{k=1}^{t-2}\left(v^{\star\top}\eta_k(x_k)\right)^2\right)^{1/2} + (\lambda-1)n^{-0.2}(\log n)\|\xi_{t-1}\|_2 \\
&\lesssim \sqrt{t}\|\xi_{t-1}\|_2 \max_{\tau_0\le s\le t}|\alpha_s| + (\lambda-1)n^{-0.2}(\log n)\|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{t\log n}{n}},
\end{aligned}$$

条件是 $t \lesssim_{\log} \lambda^{-1} n$ 且 $\max_{\tau_0\le s\le t}|\alpha_s| \lesssim \sqrt{\lambda-1}n^{-0.1}$。这里，最后一个不等式用到了 $\|\xi_t\|_2 \lesssim \sqrt{t^3 \log n}\frac{n}{}$（见 (36)）。代入 (52) 得到

$$|\Delta_{\alpha,t}| \lesssim \sqrt{\frac{t\log n}{n}} \ll (\lambda-1)|\alpha_t|, \tag{150}$$

给定 $|\alpha_t| \gtrsim \sqrt{\lambda-1}n^{-1/4}$ 并且 $\lambda - 1 \gtrsim n^{-1/9}$ 对数 n。由此完成了关系（56）的证明　　.

A.7 命题（59）<style>

在本节中，不失一般性地假设 $\alpha_t > 0$. 如关系（53 ）所述，在第 A.5 节中），应用引理2可得

$$\begin{aligned}
\left|\frac{\pi_t^2}{\alpha_t^2 n} - 1\right|\alpha_t^2 &\lesssim \left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right) \wedge \alpha_t \left(\|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}}\right)^{1/2} \lesssim \alpha_t \left(\frac{(\lambda-1)^3}{\log n}\right)^{1/4}, \\
\gamma_t^{-2} &= n\left(1 + O\left(\sqrt{\frac{(\lambda-1)^3}{\log n}}\right)\right)\int \tanh^2\left(\frac{\pi_t}{\sqrt{n}}(\alpha_t + x)\right)\varphi(\mathrm{d}x), \\
|\Delta_{\alpha,t}| &\lesssim \|\xi_{t-1}\|_2 + \sqrt{\frac{t\log n}{n}} \lesssim \sqrt{\frac{(\lambda-1)^3}{\log n}},
\end{aligned}$$

根据归纳假设 $\|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{(\lambda-1)^3}{n}}$ 对数 当 $t \lesssim \frac{n}{(\lambda-1)^5}$ 对数^2 n 且 $\alpha_t \gtrsim \sqrt{\lambda^2-1}$。现在，鉴于对（147）进行了类似计算，我们可以推导出

$$\begin{aligned}
\alpha_{t+1} &= \lambda\left[\int \tanh\left(\alpha_t^2 + \alpha_t x\right)\varphi_n(\mathrm{d}x)\right]^{1/2} + O\left(\left|\frac{\pi_t^2}{\alpha_t^2 n} - 1\right|\alpha_t^3 + \frac{\lambda-1}{\sqrt{\log n}}\alpha_t + \Delta_{\alpha,t}\right) \\
&\ge \frac{\lambda\alpha_t}{\sqrt{1+\alpha_t^2}} + o\left((\lambda-1)^{3/4}\alpha_t^2 + (\lambda-1)\alpha_t\right)
\end{aligned} \tag{151}$$

其中我们利用了这样一个事实：$\sqrt{(\lambda-1)^3 \log n} \ll (\lambda-1)\alpha_t$。

接下来，我们证明这一关系（151）结合一些简单的代数运算表明 $\alpha_{t+1} \geq \frac{1}{2}\sqrt{\lambda^2-1}$。具体而言，分别考虑以下两种情况。

- 首先，考虑 $\alpha_t \leq \frac{2}{3}\sqrt{\lambda^2-1}$的情形。类似于不等式（54），关系（151）意味着存在某个常数 $c>0$ 使得

$$\alpha_{t+1} \geq \big(1+c(\lambda-1)\big)\alpha_t + o\big((\lambda-1)\alpha_t\big) \geq \alpha_t \geq \frac{1}{2}\sqrt{\lambda^2-1}.$$

- 否则，考虑 $\alpha_t > \frac{2}{3}\sqrt{\lambda^2-1}$ 的情形。注意到 $\frac{\lambda\alpha_t}{\sqrt{1+\alpha_t^2}}$ 关于 α_t 单调递增，我们得到

$$\alpha_{t+1} \geq \big(1+c(\lambda-1)\big)\frac{2}{3}\sqrt{\lambda^2-1} + o\big((\lambda-1)^{3/4}\big) \geq \frac{1}{2}\sqrt{\lambda^2-1}.$$

因此，我们所需的界（25）得证。

A.8 引理5的证明

在整个证明过程中，我们考虑事件 $\{\phi_k\}_{k=1}^{t-1} \in \mathcal{E}$（在引理1 中定义，其中 $\delta = O(n^{-11})$），该事件以至少 $1-O(n^{-11})$ 的概率成立。在此事件上，有 $\|\phi_t\|_\infty \lesssim \sqrt{\log n}$ 以及

$$\|x_t\|_\infty \leq \|\alpha_t v^\star\|_\infty + \Big\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\|_\infty + \|\xi_{t-1}\|_2 \lesssim \frac{|\alpha_t|}{\sqrt{n}} + \sqrt{\frac{t\log n}{n}} + \sqrt{\frac{t^3\log n}{n}} \lesssim \sqrt{\frac{\log^4 n}{n(\lambda-1)^3}} \tag{152}$$

对于任意 $t \leq \tau_0$，这里提醒读者（参见（36），性质 $\|\beta_{t-1}\|_2 = 1$、$\mathcal{E}_1$ 的定义以及 41）中 τ_0 的定义）

$$|\alpha_t| \lesssim \sqrt{\lambda-1}\, n^{-1/4}, \qquad \Big\|\sum_{k=1}^{t-1}\beta_{t-1}^k\phi_k\Big\|_\infty \lesssim \sqrt{\frac{t\log n}{n}} \quad \text{and} \quad \|\xi_{t-1}\|_2 \lesssim \sqrt{\frac{t^3\log n}{n}} \tag{153}$$

只要 $t \lesssim \log_{\lambda-1} n$（参见（42））。

为了证明 $\{\eta_1(x_1)\ldots,\ \eta_t(x_t)\}$ 构成一个近似正交基，一种策略是证明对于每个 $1 \leq i \leq t \leq \tau_0$，都有 $\eta_i(x_i) \approx \phi_{i-1}$，进而可推出

$$\Big\|\sum_{i=1}^t w_i\eta_i(x_i)\Big\|_2 \approx \Big\|\sum_{i=1}^t w_i\phi_{i-1}\Big\|_2 \asymp \|w\|_2, \tag{154}$$

鉴于 $\phi_k \sim \mathcal{N}(0,\ \frac{1}{n}I_n)$ 是相互独立的高斯向量；这里，为方便记号，我们引入 $\phi_0 := x_1 \sim \mathcal{N}(0,\ \frac{1}{n}I_n)$。受此直觉启发，我们首先利用三角不等式推导出

$$\begin{aligned}&\Big|\Big\|\sum_{i=1}^t w_i\eta_i(x_i)\Big\|_2 - \Big\|\sum_{i=1}^t w_i\beta_{i-1}^{i-1}\phi_{i-1}\Big\|_2\Big| \\ &\qquad \leq \Big\|\sum_{i=1}^t w_i\Big[\eta_i(x_i) - \eta_i(\beta_{i-1}^{i-1}\phi_{i-1})\Big]\Big\|_2 + \Big\|\sum_{i=1}^t w_i\Big[\eta_i(\beta_{i-1}^{i-1}\phi_{i-1}) - \beta_{i-1}^{i-1}\phi_{i-1}\Big]\Big\|_2\end{aligned} \tag{155}$$

对于任意向量 $w \in \mathbb{R}^t$。为了界定这些项，我们采用以下三个步骤。

- 我们首先考虑 $\eta_t(x_t)$ 与 x_t之间的差异。利用引理 22c）中的性质（3，我们可以将其表示为

$$\eta_t(x_t) = (1-c_0)\bigg(x_t - \frac{\pi_t^2}{3}x_t\circ x_t\circ x_t + c_{x_t}\bigg),$$

其中 $c_0 \lesssim \frac{\log^2 n}{\sqrt{n(\lambda-1)^3}}$，$\pi_t \lesssim \frac{n^{1/4}\log n}{(\lambda-1)^{3/4}}$，且 c_{x_t} 是一个满足（参见（152））的向量

$$\|c_{x_t}\|_\infty \lesssim \frac{n\|x_t\|_\infty^5 \log^4 n}{(\lambda-1)^3} \lesssim \frac{\log^{14} n}{n^{3/2}(\lambda-1)^{10.5}}. \tag{156}$$

于是有

$$\|\eta_t(x_t) - x_t\|_2 \lesssim c_0\|x_t\|_2 + \frac{1}{3}\pi_t 2\|x_t \circ x_t \circ x_t\|_2 + \sqrt{n}\,\|c_{x_t}\|_\infty. \tag{157}$$

接下来我们声称

$$\|\eta_t(x_t) - x_t\|_2 \lesssim \frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}; \tag{158}$$

为使论述更简洁，这一断言的证明推迟到本节末尾。再次应用同样的论证，也可得到

$$\left\|\eta_t(\beta_{t-1}^{t-1}\phi_{t-1}) - \beta_{t-1}^{t-1}\phi_{t-1}\right\|_2 \lesssim \frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}. \tag{159}$$

- 结合关系（158）和（159），并利用三角不等式，可得

$$\begin{aligned}
\left\|\eta_t(x_t) - \eta_t\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right\|_2 &\le \|\eta_t(x_t) - x_t\|_2 + \left\|x_t - \beta_{t-1}^{t-1}\phi_{t-1}\right\|_2 + \left\|\eta_t(\beta_{t-1}^{t-1}\phi_{t-1}) - \beta_{t-1}^{t-1}\phi_{t-1}\right\|_2 \\
&= \left\|x_t - \beta_{t-1}^{t-1}\phi_{t-1}\right\|_2 + O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right) \\
&= \left\|\alpha_t v^\star + \sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k + \xi_t\right\|_2 + O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right) \\
&\le \|\alpha_t v^\star\|_2 + \left\|\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k\right\|_2 + \|\xi_t\|_2 + O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right) \\
&\le \left\|\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k\right\|_2 + |\alpha_t| + O\left(\sqrt{\frac{t^3\log n}{n}}\right) + O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right),
\end{aligned} \tag{160}$$

其中最后一行用到了（153）。记 $\widetilde{\beta}_{t-1} := (\beta_{t-1}^1, \ldots\ldots, \beta_{t-1}^{t-2}) \in \mathbb{R}^{t-2}$，它由移除 β_{t-1} 的最后一个元素得到。根据引理 1，我们注意到，以概率 $1 - O(n^{-11})$，

$$\left|\left\|\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k\right\|_2 - \|\widetilde{\beta}_{t-1}\|_2\right| \le \|\widetilde{\beta}_{t-1}\|_2 \cdot \sup_{a=[a_k]_{1\le k<t-2}\in\mathcal{S}^{t-3}}\left|\left\|\sum_{k=1}^{t-2}a_k\phi_k\right\|_2 - 1\right| \lesssim \sqrt{\frac{t\log n}{n}}\|\widetilde{\beta}_{t-1}\|_2,$$

这进而意味着

$$\left(1 - O\left(\sqrt{\frac{t\log n}{n}}\right)\right)\|\widetilde{\beta}_{t-1}\|_2 \le \left\|\sum_{k=1}^{t-2}\beta_{t-1}^k\phi_k\right\|_2 \le \left(1 + O\left(\sqrt{\frac{t\log n}{n}}\right)\right)\|\widetilde{\beta}_{t-1}\|_2.$$

因此，我们可以进一步控制（160）的右侧项：

$$\left\|\eta_t(x_t) - \eta_t\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right\|_2 \le \left(1 + O\left(\sqrt{\frac{t\log n}{n}}\right)\right)\|\widetilde{\beta}_{t-1}\|_2 + O\left(|\alpha_t| + \frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right). \tag{161}$$

- 接下来，我们要对这一项进行界定。$\|\beta_{t-1}\|_2$. 首先，回顾一下，β_t 对应于 $\eta_t(x_t)$ 投影到线性空间 U_t 时的线性系数（见（39）和（38b））。因此，我们可以写成

$$\|\widetilde{\beta}_t\|_2 = \left\|U_{t-1}^\top \eta_t(x_t)\right\|_2$$

$$
\begin{aligned}
&\le \left\|U_{t-1}^\top\left[\eta_t\left(x_t\right)-\eta_t\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right]\right\|_2+\left\|U_{t-1}^\top\left[\eta_t\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)-\beta_{t-1}^{t-1}\phi_{t-1}\right]\right\|_2+\left\|U_{t-1}^\top\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right\|_2\\
&\le\left(1+O\Big(\sqrt{\frac{t\log n}{n}}\Big)\right)\|\widetilde{\beta}_{t-1}\|_2+O\left(|\alpha_t|+\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)+O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)+\left\|U_{t-1}^\top\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right\|_2\\
&=\left(1+O\Big(\sqrt{\frac{t\log n}{n}}\Big)\right)\|\widetilde{\beta}_{t-1}\|_2+O\left(|\alpha_t|+\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)+O\left(\sqrt{\frac{t\log n}{n}}\right)\\
&=\left(1+O\Big(\sqrt{\frac{t\log n}{n}}\Big)\right)\|\widetilde{\beta}_{t-1}\|_2+O\left(|\alpha_t|+\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right).
\end{aligned}
$$

这里，第三行由（159）和（161）得出；倒数第二行成立，是因为 ϕ_{t-1} 与 U_{t-1} 相互独立（参见展示公式（40）下方的性质），因此 $U_{t-1}^\top\phi_{t-1}\sim\mathcal{N}(0,\ \frac{1}{n}I_t)$；只要 $t\lesssim\log_{\lambda-1}n$，最后一行便成立。注意到 $\|\widetilde{\beta}_1\|_2=\sqrt{1-\|\beta_1\|_2^2}=0$，我们可以递归地应用上述关系，得到

$$
\begin{aligned}
\|\widetilde{\beta}_t\|_2 &\le \sum_{\tau=1}^{t}\left(1+O\Big(\sqrt{\frac{t\log n}{n}}\Big)\right)^{t-1-\tau}\cdot O\left(|\alpha_\tau|+\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)\\
&\lesssim \sum_{\tau=1}^{t}\left(|\alpha_\tau|+\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)\lesssim \frac{t\sqrt{\lambda-1}}{n^{1/4}}+\frac{t\log^3 n}{\sqrt{n(\lambda-1)^3}},
\end{aligned}
$$

其中倒数第二个不等式由以下事实得出：$\sqrt{t\log n}\lesssim\frac{1}{t}$ as $t\lesssim\log_{\lambda-1}n$，而最后一个不等式用到了 $|\alpha_t|\lesssim\sqrt{\lambda-1}\,n^{-1/4}$ 对于 $t\le\tau_0$（参见 τ_0 在（41）中的定义）。在假设 $\lambda-1\gtrsim n^{-1/9}\log n$ 成立的条件下，我们还可以进一步得到

$$
\|\widetilde{\beta}_t\|_2\lesssim\frac{t\sqrt{\lambda-1}}{n^{1/4}}. \tag{162}
$$

将这一关系代入（161），并利用条件 $|\alpha_t|\lesssim\sqrt{\lambda-1}\,n^{-1/4}$ （$\forall t\le\tau_0$），可得

$$
\left\|\eta_t\left(x_t\right)-\eta_t\left(\beta_{t-1}^{t-1}\phi_{t-1}\right)\right\|_2\lesssim\frac{t\sqrt{\lambda-1}}{n^{1/4}}. \tag{163}
$$

值得注意的是，不等式（162）还意味着

$$
|\beta_t^t|=\sqrt{\|\beta_t\|_2^2-\|\widetilde{\beta}_t\|_2^2}=\sqrt{1-\|\widetilde{\beta}_t\|_2^2}=1-O\left(\frac{t^2(\lambda-1)}{n^{1/2}}\right).
$$

最后，将上述界与表达式（155）结合起来，我们得出结论 at

$$
\begin{aligned}
&\left|\Big\|\sum_{i=1}^{t}w_i\eta_i(x_i)\Big\|_2-\Big\|\sum_{i=1}^{t}w_i\phi_{i-1}\Big\|_2\right|\\
&\le(1-\beta_{t-1}^{t-1})\Big\|\sum_{i=1}^{t}w_i\phi_{i-1}\Big\|_2+\Big\|\sum_{i=1}^{t}w_i\Big[\eta_i(x_i)-\eta_i(\beta_{i-1}^{i-1}\phi_{i-1})\Big]\Big\|_2+\Big\|\sum_{i=1}^{t}w_i\Big[\eta_i(\beta_{i-1}^{i-1}\phi_{i-1})-\beta_{i-1}^{i-1}\phi_{i-1}\Big]\Big\|_2\\
&=O\left(\frac{t^2(\lambda-1)}{n^{1/2}}\right)\Big\|\sum_{i=1}^{t}w_i\phi_{i-1}\Big\|_2+O\left(\frac{t\sqrt{\lambda-1}}{n^{1/4}}\right)\cdot\sqrt{t}\|w\|_2+O\left(\frac{\log^3 n}{\sqrt{n(\lambda-1)^3}}\right)\cdot\sqrt{t}\|w\|_2\\
&=O\left(\frac{t^{3/2}\sqrt{\lambda-1}}{n^{1/4}}+\sqrt{\frac{t\log^6 n}{n(\lambda-1)^3}}\right)\|w\|_2=o(\|w\|_2),
\end{aligned}
$$

其中，最后一个关系由事实 $t<\tau_0\lesssim\log n/(\lambda-1)$ 和假设 $\lambda-1\ge n^{-1/9}$ 得出。最后，观察到 $\|\sum_{i=1}^{t}w_i\phi_{i-1}\|_2=(1+O(\sqrt{t\log n}))\|w\|_2$，我们得到

$$
\Big\|\sum_{i=1}^{t}w_i\eta_i(x_i)\Big\|_2=\big(1+o(1)\big)\|w\|_2,
$$

由此完成了对所需界的证明。

不等式的证明 (158). 对于任意固定的整数 $k \geq 2$，其不随 n缩放，我们可以写成

$$\sum_{i=1}^{n} |x_{t,i}|^k \lesssim \sum_{i=1}^{n} |\alpha_t v_i^\star|^k + \sum_{i=1}^{n} |u_{t,i}|^k + \sum_{i=1}^{n} |\xi_{t-1,i}|^k, \qquad \text{with} \quad u_t \coloneqq \sum_{k=1}^{t-1} \beta_{t-1}^k \phi_k.$$

分别对每一项进行界定。首先，回想 $\|\xi_{t-1}\|_2 \lesssim \sqrt{t^3 \log n}$ (参见（36）)，可得我们来

$$\sum_{i=1}^{n} |\xi_{t-1,i}|^k \leq \|\xi_{t-1}\|_2^k \lesssim \left(\frac{t^3 \log n}{n}\right)^{k/2}. \tag{164}$$

其次，在事件 $\{\phi_k\}_{k=1}^{t-1} \in \mathcal{E}$(见引理 1) 上，我们看到 $\|u_t\|_2 \lesssim 1$ 且 $\|u_t\|_\infty \leq \sqrt{(t \log n)/n}$。由此可得

$$\begin{aligned}
\sum_{i=1}^{n} |u_{t,i}|^k = \sum_{i \leq t} |u_{t,(i)}|^k + \sum_{i>t} |u_{t,(i)}|^k &\overset{(*)}{\lesssim} t \left(\frac{t \log n}{n}\right)^{k/2} + \left(\frac{\log n}{n}\right)^{k/2-1} \sum_{i>t} |u_{t,(i)}|^2 \\
&\lesssim t \left(\frac{t \log n}{n}\right)^{k/2} + \left(\frac{\log n}{n}\right)^{k/2-1} \\
&\lesssim \left(1 + \frac{t^{k/2+1} \log n}{n}\right) \left(\frac{\log n}{n}\right)^{k/2-1},
\end{aligned} \tag{165}$$

其中 $x_{(i)}$ 表示 i的第 x 大元素（按绝对值计）。这里，为了说明不等式$(*)$ 成立，我们回想在事件 $\{\phi_k\}_{k=1}^{t-1} \in \mathcal{E}$ (见引理 1) 上，有

$$\sup_{a \in \mathcal{S}^{t-2}} \sum_{i=1}^{t} \left| \sum_{k=1}^{t-1} a_k \phi_k \right|_{(i)}^2 \lesssim \frac{t \log n}{n},$$

这也保证了 $|u_{t,\ (t+1)}| \lesssim \sqrt{\log n / n}$. 综合以上各部分，我们得到

$$\sum_{i=1}^{n} |x_{t,i}|^k \lesssim n \left(\frac{1}{n(\lambda-1)}\right)^{3k/4} + \left(1 + \frac{t^{k/2+1} \log n}{n}\right) \left(\frac{\log n}{n}\right)^{k/2-1} + \left(\frac{t^3 \log n}{n}\right)^{k/2} \lesssim \left(\frac{\log n}{n}\right)^2, \tag{166}$$

其中，若取 $k = 6$, $t \lesssim \log n/(\lambda - 1)$ 并假设 $\lambda - 1 \gtrsim n^{-1/9} \log n.$，则最后一个不等式成立。因此得到

$$\pi_t^2 \|x_t \circ x_t \circ x_t\|_2 = \pi_t^2 \Big(\sum_{i=1}^{n} |x_{t,i}|^6 \Big)^{1/2} \lesssim \frac{n^{1/2} \log^2 n}{(\lambda-1)^{3/2}} \cdot \frac{\log n}{n} = \frac{\log^3 n}{\sqrt{n(\lambda-1)^3}},$$

其中我们使用了引理 π_t 中的界3 （第 3 种情况）。结合（156）、（157）以及 $c_0 \lesssim \log^2 n / \sqrt{n(\lambda-1)^3}$ 这一事实，即可完成不等式（158）的证明。

A.9 不等式（70）的证明

在继续之前，我们先对 $h^{\tau_t}(\tau_t)$ 作几点观察。如 Li 和 Wei （2022年，展示公式（254））附近所述，序列 τ_t （其中 $\tau_{t+1} = \lambda^2 h(\tau_t)$）单调递增，这意味着 $h^{\tau_t}(\tau_t) \leq \lambda^2$。此外，贝叶斯估计量的最优性（参见（~~130~~））意味着 $h^{\tau_t}(\tau_t) \leq \tau_t + 1$。结合这两点观察，我们得到

$$\frac{\tau_t}{h(\tau_t)} \leq (\tau_t + 1) \wedge \lambda^2.$$

根据归纳假设，我们有 $\alpha_t^2 = (1+o(1))\tau_t$ 当 $t \gtrsim$ ς时。因此，对于每个 τ 满足最小值$\{\tau_t, \alpha_t^2\} \le \tau \le$最大值$\{\tau_t, \alpha_t^2\}$ 的 $\tau = (1+o(1))\tau_t$，有 $\tau_t \gtrsim \lambda^2 - 1$。按照 Li 和 Wei（2022年）展示公式（263）的定义，定义 $\mathcal{T}_2$，使得

$$\mathcal{T}_2(s,\tau) := s^2 h'(\tau) = s^2 \int \left(1 + \frac{x}{2\sqrt{\tau}}\right)\left(1 - \tanh^2\left(\tau + \sqrt{\tau}x\right)\right)\varphi(\mathrm{d}x). \tag{167}$$

有了这些记号，我们可以将目标量界定为

$$\frac{\tau_t}{h(\tau_t)} h'(\tau) \le \mathcal{T}_2(\sqrt{\tau_t + 1} \wedge \lambda, \tau).$$

因此，只需将上述不等式右侧的上界进一步界定为 $1 - c(\lambda - 1)$。为此，直接计算可得

$$\frac{\mathcal{T}_2(\sqrt{\tau_t+1}\wedge\lambda,\tau) - \mathcal{T}_2(\sqrt{\tau+1}\wedge\lambda,\tau)}{\mathcal{T}_2(\sqrt{\tau+1}\wedge\lambda,\tau)} = (\sqrt{\tau+1}\wedge\lambda)^2 - (\sqrt{\tau_t+1}\wedge\lambda)^2 = o(\lambda-1). \tag{168}$$

此外，数值结果已证明（见 Li 和 Wei (2022年) 中的图 1），有

$$\mathcal{T}_2(\lambda,\tau) \le 1 - (\lambda - 1), \quad \text{for } \lambda \in (0, 1.2] \text{ and } \tau > \sqrt{\lambda^2 - 1}.$$

注意到 $\tau = (1 + o(1))\tau_t \gtrsim \lambda^2 - 1$，由上述关系可推得 t

$$\mathcal{T}_2(\sqrt{\tau+1}\wedge\lambda,\tau) \le 1 - \left((\sqrt{\tau+1}\wedge\lambda) - 1\right) = 1 - c_1(\lambda - 1) \tag{169}$$

某个通用 常数 $c_1 > 0$。最后，结合关系式 (168) 和 (169)，可得 t

$$\frac{\tau_t}{h(\tau_t)} h'(\tau) \le \mathcal{T}_2(\sqrt{\tau_t+1}\wedge\lambda,\tau) = (1+o(\lambda-1))\mathcal{T}_2(\sqrt{\tau+1}\wedge\lambda,\tau) \le 1 - c(\lambda-1)$$

对于某个普适常数 $c > 0$。至此，我们完成了关系（70）的证明。

B 表达式（17）的证明及推论 1

首先，根据 u_t 的定义（16），可得

$$\|u_t\|_2 = \left\| \frac{1}{\lambda\sqrt{n(\alpha_t^2+1)}} \tanh(\pi_t x_t) \right\|_2 \le \frac{1}{\lambda},$$

这里我们用到了 |双曲正切(πₜxₜ)| < 1 这一事实。因此，所关注的量 $\|v^\star v^{\star\top} - u_t u_t^\top\|_{\mathsf{F}}^2$ 一致地有一个常数上界 $1 + \frac{1}{\lambda^4}$.。此外，我们发现以下观察很有帮助：

$$\begin{aligned}
\|v^\star v^{\star\top} - u_t u_t^\top\|_{\mathsf{F}}^2 &= \|v^\star\|_2^4 - 2(v^{\star\top} u_t)^2 + \|u_t\|_2^4 \\
&= 1 - 2\frac{1}{n\lambda^2(\alpha_t^2+1)}\left(v^{\star\top}\tanh(\pi_t x_t)\right)^2 + \frac{1}{n^2\lambda^4(\alpha_t^2+1)^2}\|\tanh(\pi_t x_t)\|_2^4 \\
&= 1 - \frac{1}{\lambda^4}\left(\frac{2\alpha_{t+1}^2}{(\alpha_t^2+1)n\gamma_t^2} + \frac{1}{(\alpha_t^2+1)^2 n^2\gamma_t^4}\right).
\end{aligned} \tag{170}$$

要验证表达式（17），只需注意到 $\gamma_t^{-2} - n\alpha_t^2(\alpha_t^2+1) = o(1)$ 以至少 $1 - O(n^{-10})$（根据引理 2），进而可得

$$\lim_{t\to\infty}\lim_{n\to\infty} \mathbb{E}\left[\|v^\star v^{\star\top} - u_t u_t^\top\|_{\mathsf{F}}^2\right] = 1 - \frac{\alpha^{\star 4}}{\lambda^4}.$$

为证明推论 1，我们再次运用引理 2 来证明

$$
\begin{aligned}
\frac{1}{n}\gamma_t^{-2} &= \frac{1}{n}\pi_t^2\left(\alpha_t^2+1+O\left(\frac{\pi_t^2}{n}+\|\xi_{t-1}\|_2+\sqrt{\frac{t\log n}{n}}\right)\right)\\
&= \left(\alpha_t^2+O\left(\left(\|\xi_{t-1}\|_2+\sqrt{\frac{t\log n}{n}}\right)^{1/2}\right)\right)\left(\alpha_t^2+1+O\left(\frac{\pi_t^2}{n}+\|\xi_{t-1}\|_2+\sqrt{\frac{t\log n}{n}}\right)\right)\\
&= (\alpha_t^2+O(\delta^{1/2}))(\alpha_t^2+1+O(\delta))\\
&= \alpha_t^2(\alpha_t^2+1)+O(\delta^{1/2})
\end{aligned}
\tag{171}
$$

其中 $\delta := \sqrt{\frac{t}{n}\frac{\text{对数}n}{\lambda-1}(\cdot)^2}+\sqrt{\frac{\text{对数}n}{\lambda-1}(\cdot)^3}$，其中代入 $\|\xi_t\|_2$ 中的界，如表达式（8c）所示。将表达式（171）代入（170）可得

$$
\|v^\star v^{\star\top}-u_t u_t^\top\|_{\mathsf{F}}^2 = 1-\frac{\alpha_t^2}{\lambda^4}\left(2\alpha_{t+1}^2-\alpha_t^2+O(\delta^{1/2})\right). \tag{172}
$$

经过大约 $\frac{\text{对数}n}{\lambda-1}$ 次迭代后，性质（14）确保 $\alpha_t^2-\alpha^{\star 2}=O(\sqrt{\frac{\text{对数}n}{\lambda-1}(\cdot)^6})$。综合以上结果，我们得到

$$
\|v^\star v^{\star\top}-u_t u_t^\top\|_{\mathsf{F}}^2 = 1-\frac{\alpha^{\star 4}}{\lambda^4}+O\left(\sqrt{\frac{\log^4 n}{n(\lambda-1)^6}}\right),
$$

该结论以至少 $1-O(n^{-10})$ 的概率成立。

参考文献

Abbe, E.、Fan, J.、Wang, K. 和 Zhong, Y.（2020）。低期望秩随机矩阵的逐元素特征向量分析。《统计年鉴》，48(3)：1452。

Baik, J.、Arous, G. B. 和 Péché, S.（2005）。非零复样本协方差矩阵最大特征值的相变。《概率年刊》，33(5)：1643–1697。

Bandeira, A. S., Kunisky, D. 和 Wein, A. S.（2019年）。约束主成分分析问题中界限认证的计算难度。*arXiv*预印本 *arXiv:1902.07324*。

Bayati, M. 和Montanari, A.（2011a）。密集图上的消息传递动态及其在压缩感知中的应用。*IEEE*信息论汇刊，57(2):764–785。

Bayati, M. 和 Montanari, A.（2011b）。高斯矩阵的 LASSO 风险。*IEEE*信息论汇刊，58(4):1997–2017。

Berthet, Q. 和 Rigollet, P.（2013）。高维稀疏主成分的最优检测。统计年鉴，41(4):1780–1815。

Bu, Z., Klusowski, J. M., Rush, C. 和 Su, W. J.（2020）。通过近似消息传递对 SLOPE 进行算法分析与统计估计。*IEEE*信息论汇刊，67(1):506–537。

Cademartori, C. 和 Rush, C.（2023）。针对右旋转不变设计的广义近似消息传递算法的非渐近分析。*arXiv* 预印本 *arXiv:2302.00088*。

Cai, T.T. 和 Zhang, A.（2018年）。奇异子空间的速率最优扰动界及其在高维统计学中的应用。统计年鉴，46(1)：60–89。

Candes, E. J. 和 Plan, Y.（2010）。带噪声的矩阵补全。《*IEEE*会刊》，98(6):925–936。

Capitaine, M., Donati-Martin, C. 和 Féral（2009）。大规模 Wigner 矩阵有限秩变形的最大特征值：涨落的收敛性与非普适性。《概率年刊》，37(1):1–47。Celentano, M., Fan, Z. 和 Mei, S.（2021）。Z2同步的 TAP 自由能局部凸性与近似消息传递收敛。*arXiv* 预印本 *arXiv:2106.11428*。Celentano, M. 和 Montanari, A.（2022）。高维回归中凸惩罚的基本障碍。统计年鉴，50(1):170–196。Celentano, M., Montanari, A. 和 Wei, Y.（2020）。一般高斯设计下的 LASSO 及其在假设检验中的应用。*arXiv* 预印本 *arXiv:2007.13716*。Chen, Y. 和 Candès, E. J.（2018）。投影幂方法：一种基于成对差异进行联合对齐的高效算法。纯粹与应用数学通讯，71(8):1648–1714。Chen, Y., Chi, Y., Fan, J. 和 Ma, C.（2019）。随机初始化的梯度下降：非凸相位恢复的快速全局收敛。数学规划，176(1):5–37。Chen, Y., Chi, Y., Fan, J. 和 Ma, C.（2021）。数据科学的谱方法：统计视角。基础与趋势® 机器学习，14(5):566–806。Chen, Y., Chi, Y., Fan, J., Ma, C. 和 Yan, Y.（2020）。噪声矩阵补全：通过非凸优化理解凸松弛的统计保证。*SIAM*优化杂志，30(4):3098– 3121。Chen, Y. 和 Wainwright, M. J.（2015）。通过投影梯度下降快速进行低秩估计：一般性的统计保证与算法保证。*arXiv* 预印本 *arXiv:1509.03025*。Cheng, C., Wei, Y. 和 Chen, Y.（2021）。应对微小特征值间隙：异方差噪声下精细化的特征向量估计与推断。*IEEE*信息论汇刊，67(11):7380–7419。Chi, Y., Lu, Y. M. 和 Chen, Y.（2019）。非凸优化与低秩矩阵分解的结合：综述。*IEEE*信号处理汇刊，67(20):5239–5269。Deshpande, Y., Abbe 和 Montanari, A.（2017）。平衡二元随机块模型的渐近互信息。信息与推断：*IMA*期刊，6(2):125–170。Deshpande, Y., Montanari, A. 和 Richard, E.（2014）。锥约束主成分分析。神经信息处理系统进展，27。Donoho, D. 和 Montanari, A.（2016）。高维稳健 M 估计：通过近似消息传递求渐近方差。概率论及相关领域，166(3):935–969。Donoho, D. L., Maleki, A. 和 Montanari, A.（2009）。压缩感知的消息传递算法。美国国家科学院院刊，106(45):18914–18919。Donoho, D. L., Maleki, A. 和 Montanari, A.（2010）。压缩感知的消息传递算法：I. 动机与构造。载于*2010*年*IEEE*信息论研讨会（*ITW 2010*，开罗），第1–5页。IEEE。El Karoui, N.（2018）。预测变量几何对高维岭正则化广义稳健回归估计量性能的影响。概率论及相关领域，170(1):95–175。Fan, Z.（2022）。旋转不变矩阵的近似消息传递算法。统计年鉴，50(1):197–224。Fan, Z., Mei, S. 和 Montanari, A.（2021）。TAP自由能、自旋玻璃与变分推断。《概率年刊》，49(1):1–45。

Feng, O. Y.、Venkataramanan, R.、Rush, C. 和 Samworth, R. J.（2022年）。近似消息传递统一教程。*Foundations and Trends® in Machine Learning*，15(4):335–536。Féral, D. 和 Péché, S.（2007年）。大型 Wigner 矩阵秩一变形的最大特征值。数学物理通讯，272(1):185–228。Gao, C. 和 Zhang, A. Y.（2022年）。半定规划在相位同步中达到精确极小极大最优性。*IEEE*信息论汇刊。Ge, R.、Jin, C. 和 Zheng, Y.（2017 年）。非凸低秩问题中不存在伪局部极小值：统一的几何分析。载于国际机器学习会议，第1233–1242页。PMLR。Javanmard, A.、Montanari, A. 和 Ricci-Tersenghi, F.（2016年）。半定松弛中的相变。美国国家科学院院刊，113(16):E2218–E2223。Johnstone, I. M.（2001年）。关于主成分分析中最大特征值的分布。统计年鉴，29(2):295–327。Johnstone, I. M. 和 Lu, A. Y.（2009年）。关于高维主成分分析的一致性与稀疏性。美国统计学会会刊，104(486):682–693。Keshavan, R.、Montanari, A. 和 Oh, S.（2009年）。从含噪条目进行矩阵补全。神经信息处理系统进展，22。Lelarge, M. 和 Miolane, L.（2019年）。对称低秩矩阵估计的基本极限。概率论及相关领域，173(3):859–929。Lesieur, T.、Krzakala, F. 和 Zdeborová, L.（2017 年）。约束低秩矩阵估计：相变、近似消息传递与应用。统计力学杂志：理论与实验，2017(7):073403。Li, G. 和 Wei, Y.（2022年）。尖峰模型中近似消息传递的非渐近框架。*arXiv preprint arXiv:2208.03313*。Li, Y. 和 Wei, Y.（2021年）。最小 ℓ_1-范数插值器：精确渐近分析与多重下降。*arXivpreprint arXiv:2110.09502*。Ma, C.、Wang, K.、Chi, Y. 和 Chen, Y.（2020年）。非凸统计估计中的隐式正则化：梯度下降在线性速率下收敛于相位恢复、矩阵补全和盲反卷积。计算数学基础，20(3):451–632。Ma, J.、Xu, J. 和 Maleki, A.（2018年）。用于相位恢复的基于优化的近似消息传递：初始化和 ℓ_2-正则化的影响。*arXiv preprint arXiv:1801.01170*。Massart, P.（2007年）。集中不等式与模型选择：圣弗卢尔概率论暑期学校 *XXXIII-2003*。施普林格。Miolane, L. 和 Montanari, A.（2021 年）。LASSO 的分布：对稀疏球的一致控制与自适应参数调优。统计年鉴，49(4):2313–2335。Mondelli, M. 和 Venkataramanan, R.（2021 年）。旋转不变模型中近似消息传递的主成分分析初始化。神经信息处理系统进展，34:29616–29629。Montanari, A. 和 Venkataramanan, R.（2021 年）。通过近似消息传递进行低秩矩阵估计。统计年鉴，49(1):321–345。Péché, S.（2006年）。厄米随机矩阵的小秩扰动的最大特征值。概率论及相关领域，134(1):127–173。Perry, A.、Wein, A. S.、Bandeira, A. S. 和 Moitra, A.（2018年）。紧群上同步问题的消息传递算法。纯粹与应用数学通讯，71(11):2275–2322。

Rangan, S.（2011年）。用于随机线性混合估计的广义近似消息传递。载于*2011 IEEE*国际信息论研讨会论文集，第2168–2172页。IEEE。Rush, C. 和 Venkataramanan, R.（2018年）。近似消息传递算法的有限样本分析。*IEEE*信息论汇刊，64(11):7264–7286。Singer, A.（2011年）。通过特征向量和半定规划进行角同步。应用与计算调和分析，30(1):20–36。Sur, P. 和 Candès, E. J.（2019年）。高维逻辑回归的现代最大似然理论。美国国家科学院院刊，116(29):14516–14525。Sur, P.、Chen, Y. 和 Candès, E. J.（2019年）。高维逻辑回归中的似然比检验渐近服从重缩放卡方分布。概率论及相关领域，175(1-2):487–558。Thrampoulidis, C.、Abbasi, E. 和 Hassibi, B.（2018年）。高维正则化 m-估计量的精确误差分析。*IEEE*信息论汇刊，64(8):5592–5628。Wei, Y.、Wainwright, M. J. 和 Guntuboyina, A.（2019年）。凸锥上假设检验的几何：广义似然比检验与极小极大半径。统计年鉴，47(2):994–1024。Wu, Y. 和 Zhou, K.（2022年）。随机超完备模型中张量幂迭代收敛性的下界。*arXiv preprint arXiv:2211.03827*。Yan, Y.、Chen, Y. 和 Fan, J.（2021 年）。带缺失数据的异方差主成分分析推断。*arXiv preprintarXiv:2107.12365*。Zhang, Y.、Mondelli, M. 和 Venkataramanan, R.（2022年）。混合广义线性模型中谱方法的精确渐近分析。*arXiv preprint arXiv:2211.11368*。Zhong, X.、Wang, T. 和 Fan, Z.（2021 年）。正交不变系综的近似消息传递：多变量非线性与谱初始化。*arXiv preprint arXiv:2110.02318*。Zhong, Y. 和 Boumal, N.（2018年）。相位同步的近乎最优界。*SIAM*优化杂志，28(2):989–1016。